

ACTA UNIVERSITATIS UPSALIENSIS
Uppsala Studies in Education No 117

Monica Langerth Zetterman

Innehållsdesign

– principer, metoder och verktyg samt
tillämpningar inom utbildningshistorisk forskning
och undervisning



UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Auditorium Minus, Museum Gustavianum, Akademigatan 3, Uppsala, Friday, March 14, 2008 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Langerth Zetterman, M. 2008. Innehållsdesign. Principer, metoder och verktyg samt tillämpningar inom utbildningshistorisk forskning och undervisning. (Content design. Principles, methods, tools, and applications in history of education). Acta Universitatis Upsaliensis. Uppsala Studies in Education 117. 272 pp. Uppsala. ISBN 978-91-554-7091-3.

This thesis explores content design – an area which encompasses the practices and the conceptions of the description, organisation and manipulation of digital content. The overall aim was to identify and examine principles, methods and tools appropriate for content design within the humanities and the social sciences. Another purpose was to investigate the limitations and opportunities of the identified methods and tools by means of modelling and applications of prosopographical materials, designed for research and teaching in history of education.

The prosopographical collection consists of three different kinds of sources: transcriptions from biographical reference books, written biographical accounts and digitalised archival sources, such as enrolment registers. These resources were encoded according to the Text Encoding Initiative (TEI) guidelines with the purpose to denote specific structures and semantic features of the content. The thesis demonstrates how the prosopographical collection, stored in a master file in TEI/XML format, was encoded and organised and then further transformed, migrated and manipulated by other tools and to other platforms. This resulted in several examples of applications demonstrating a broad range of uses for research and teaching in history of education and alike.

One conclusion is that the TEI guidelines serve well as a valuable tool for the markup of rather complex historical materials designed for multiple purposes: for qualitative analyses, and as input to multivariate statistical analyses, and for migration into relational databases. Another conclusion is that such digital collections, provided with markup, could be treated as research tools themselves, because they lend themselves much more than simply access, retrieval or reading. In this prosopographical collection, the markup contributes to make explicit the underlying theories and thus provides scholars, teachers and students with tools to reuse and rearrange the content for other kinds of uses in other areas.

Keywords: Text Encoding Initiative, markup, markup languages, text encoding, history of education, prosopography, educational science, educational uses, methods in education, methodology, historical sources, digital collection, collective biography, electronic texts

*Monica Langerth Zetterman, Department of Education, Box 2109
Uppsala University, SE-750 02 Uppsala, Sweden*

© Monica Langerth Zetterman 2008

ISSN 0347-1314

ISBN 978-91-554-7091-3

urn:nbn:se:uu:diva-8464 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-8464>)

Printed in Sweden by Edita Västra Aros, Västerås 2008

Distributor: Uppsala University Library, Box 510, SE-751 20 Uppsala

www.uu.se, acta@ub.uu.se

Till Agnes, Alice & Erik

Innehållsförteckning

Termlista	13
Förord	17
DEL I Inledning.....	19
1 Utgångspunkter.....	21
1.1 Problembeskrivning.....	24
1.2 Syfte och frågor	36
1.3 Tillvägagångssätt.....	37
1.4 Avgränsningar.....	40
1.5 Något om terminologi	42
1.6 Disposition	48
Del II Innehållsdesign – avgränsning och inventering av principer, metoder och verktyg	49
2 Innehållsdesign – några utgångspunkter	51
2.1 Innehållsdesign – definition och avgränsning.....	51
2.2 Grundläggande principer för märkning av innehåll.....	55
2.3 Open Source – öppna standarder, metoder och verktyg	60
3 Principer och metoder för innehållsdesign.....	70
3.1 Inledning.....	70
3.2 Märkspråkens utveckling.....	73
3.3 Innehållets ordning	77
3.4 SGML – ett märkspråk för industrin	80
3.5 XML – en verktygslåda för webben	82
3.6 Alternativ till den hierarkiska dokumentmodellen	85
3.7 Metoder och verktyg för metadata på webben.....	91
4 Verktøy för innehållsdesign – Text Encoding Initiative.....	102
4.1 Inledning.....	102
4.2 Riktlinjerna tar form.....	103
4.3 Från projekt till konsortium.....	107
4.4 Riktlinjernas utgångspunkter och principer	113
4.5 Modulerna.....	118
4.6 TEI P5 – från pizza till smörgåsbord.....	126

4.7	Avslutning.....	130
Del III	Tillämpad innehållsdesign.....	131
5	Prosopografisk metod, verktyg och material.....	133
5.1	Om metod	133
5.2	Prosopografins utgångspunkter och karaktäristika	135
5.3	Prosopografiska tillämpningar – en översikt.....	138
5.4	Introduktion till projektet Formering för offentlighet.....	143
5.5	Modellerings utgångspunkter	145
5.6	Introduktion till det prosopografiska materialet.....	151
5.7	Kapitalbeskrivningar som verktyg för insamling av material	157
5.8	Biografiska uppslagsböcker som källmaterial i prosopografin	162
5.9	Arkivmaterial som källor i prosopografin.....	167
5.10	Avslutning	171
6	Innehållsdesign av prosopografiskt material	172
6.1	Beskrivande märkning av kapitalbeskrivningar	172
6.2	Beskrivande märkning av arkivmaterial	184
6.3	Verktyg för semi-automatiserad märkning.....	187
6.4	Metoder för migrering av filformat	189
6.5	Ordning och nyordning – klassificering av indikatorer.....	196
7	Tillämpningar av det prosopografiska materialet	201
7.1	Beskrivande statistik.....	201
7.2	Nätverksanalys	208
7.3	Interaktiv visualisering.....	213
7.4	Att förena bilder med text.....	216
7.5	Webbarkivet	220
7.6	Sammanfattande kommentarer om tillämpningar.....	225
Del IV	Avslutning	229
8	Sammanfattande diskussion.....	231
8.1	Varför innehållsdesign?	232
8.2	Återanvändningens aspekter.....	233
8.3	Innehållets ordning	236
8.4	Verktygens styrkor och svagheter.....	240
8.5	Innehållsdesign för forskning och undervisning.....	244
8.6	Avslutande kommentarer.....	246
Summary		249
	Content design. Principles, methods, tools and applications in history of education	249
Innehållsförteckning för digitala bilagor.....		260
Bilageförteckning		261

Figurförteckning.....	261
Bildförteckning.....	263
Filförteckning.....	263
Dokumentförteckning.....	264
Verktügsförteckning.....	264
Referenser.....	265
Otryckt material.....	265
Digitalt material	265
Bearbetat digitalt material	266
Övriga tryckta referenser	267

Figur- och tabellförteckning

Figurförteckning

<i>Figur 1.</i> Exempel på den deskriptiva märkningens grundläggande principer: elementet.....	55
<i>Figur 2.</i> Exempel på den deskriptiva märkningens hierarkiska struktur med hjälp av nästlade element.....	56
<i>Figur 3.</i> XML: hierarkisk dokumentmodell.....	56
<i>Figur 4.</i> Exempel på märkning av en viss kvinnas egenskaper och tillgångar.....	58
<i>Figur 5.</i> Urklipp ur nätverksgraf med Ellen Keys egenskaper (attribut) accentuerade ...	59
<i>Figur 6.</i> Exempel på procedurell märkning med LaTeX.....	74
<i>Figur 7.</i> Exempel på märkning av föreningsaktiviteter i FFO-KAPITAL	79
<i>Figur 8.</i> Märkspråket COCOA, exempel på syntax	85
<i>Figur 9.</i> Exempel på COCOA märkning av 1600-talsdrama.....	86
<i>Figur 10.</i> Jämförelse av bruket av sluttagg för nästlade element mellan SGML och XML	87
<i>Figur 11.</i> Periodisk sökning av ”semantic web”, antal träffar per angiven tidsperiod.....	93
<i>Figur 12.</i> Exempel på Topic Maps datamodell och grundläggande principer.....	99
<i>Figur 13.</i> Översikt TEI:s överordnade moduler (<i>Pizza Chef Model</i>) i TEI P4.....	119
<i>Figur 14.</i> TEI-dokumentets grundläggande struktur och de obligatoriska elementen i <teiHeader>	123
<i>Figur 15.</i> Exempel på grundläggande struktur och element i TEI-dokumentets kropp.....	124
<i>Figur 16.</i> Illustration av FFO-databasens utformning.....	152
<i>Figur 17.</i> Översikt av kapitalbeskrivningsmallens logiska ordning och hierarkiska struktur	159
<i>Figur 18.</i> Utdrag ur inskrivningsmatrikeln för läsåret 1861-1862 (Serie D1, vol. 1).....	170
<i>Figur 19.</i> Principer för beskrivning av innehåll i kapitalbeskrivningarna	177
<i>Figur 20.</i> Exempel på poster som värdet av ett <i>key</i> -attribut relaterar till	179
<i>Figur 21.</i> Märkning av en kvinnas kontakt med andra kvinnor i materialet	181
<i>Figur 22.</i> Exempel på post i inskrivningsmatrikel för perioden 1914–1928 (Serie D1, vol. 2)	185
<i>Figur 23.</i> Märkning av arkivpost (urklipp).....	185
<i>Figur 24.</i> Märkning av unikt id samt länkning till externa dokument.....	186
<i>Figur 25.</i> Semiautomatisk märkning av föreningsnamn	188
<i>Figur 26.</i> Översikt för migrering av filformat	191
<i>Figur 27.</i> Migrering från TEI/XML masterfil till andra filformat och användningsområden	192
<i>Figur 28.</i> Extrahering av önskad information med XSLT, från masterfil till tabell.....	194
<i>Figur 29.</i> FFO-LUNDSTROM registrets 4 nivåer.	199
<i>Figur 30.</i> Kvinnornas sociala ursprung (boendeort och sfär). Procent (N=1251)	205
<i>Figur 31.</i> Lärarinneelever vid Kungl. Högre Lärarinneseminariet; årsvis jämförelse av födelseorter.....	207

<i>Figur 32. Nätverksanalys av 101 kvinnors interna kontakter i FFO-KAPITAL</i>	<i>210</i>
<i>Figur 33. Nätverksanalys av FFO-KAPITAL: flest kontakter och mest omnämnd</i>	<i>211</i>
<i>Figur 34. Exempel på sökning i en nätverksanalys.....</i>	<i>212</i>
<i>Figur 35. Exempel på en kvinnas vy</i>	<i>214</i>
<i>Figur 36. Visualisering av XML-dokumentets struktur</i>	<i>215</i>
<i>Figur 37. Illustrering av trädstrukturens yttersta gren och dess innehåll.....</i>	<i>216</i>
<i>Figur 38. Exempel på hur bild och text kan användas tillsammans.....</i>	<i>218</i>
<i>Figur 39. Exempel på relaterande märkning som förenar bild och transkriberad text..</i>	<i>219</i>
<i>Figur 40. Databasschema för FFO-KAPITAL</i>	<i>223</i>
<i>Figur 41. Webbaserat gränssnitt för sökning i FFO-KAPITAL.....</i>	<i>224</i>
<i>Figur 42. Webbgränssnitt för visning av resultat efter sökning i FFO-KAPITAL.....</i>	<i>225</i>

Tabellförteckning

<i>Tabell 1. Jämförelse TEI:s medlemmar per land och typ av medlemsinstitution. Absoluta tal.....</i>	<i>110</i>
<i>Tabell 2. Fördelning av världsländer för digitala samlingar som använder TEI:s riktlinjer under tidsperioden 1996-2007. Absoluta tal</i>	<i>112</i>
<i>Tabell 3. Basmodulens (base tag set) elementuppsättningar i TEI P4.</i>	<i>121</i>
<i>Tabell 4. Tilläggsmodulens elementuppsättningar i TEI P4.....</i>	<i>122</i>
<i>Tabell 5. Jämförelse av indikatorer och registrerade värden för kvinnornas egenskaper och tillgångar från två biografiska uppslagsböcker.</i>	<i>164</i>
<i>Tabell 6. Översikt över webbarkivets moduler och funktioner.....</i>	<i>221</i>
<i>Tabell 7. Uppskattad tidsåtgång för olika moment i arbetet med den prosopografiska databasen..</i>	<i>226</i>

Förkortningar

ASCII	American Standard Code for Information Interchange
COCOA	Word Count and Concordance on Atlas
CSS	Cascading Style Sheets
DTD	Document Type Definition
EAD	Encoded Archival Description
FFO-HLS	Databasen för Kungl. Högre Lärarinneseminarier i Formering för offentlighet
FFO-KAPITAL	Databasen med kapitalbeskrivningar i Formering för offentlighet
FFO-LUNDSTROM	Databasen Lundströms biografiska uppslagsbok i Formering för offentlighet
FSF	Free Software Foundation
GIF	Graphics Interchange Format
GIS	Geographic Information System
GNU	GNU's Not Unix
HTML	Hypertext Markup Language
ISO	International Organization for Standardization
JPEG/JPG	Joint Photographic Experts Group
MECS	Multi-Element Code System
NAD	Nationella Arkivdatabasen (vid Riksarkivet)
OCF	Oxford Concordance Program
ODD	One Document Does it All
ODF	Open Document Format
OHCO	Ordered Hierarchy of Content Objects
OSI	Open Source Initiative
OTA	Oxford Text Archive
PHP	Hypertext Preprocessor
PDF	Portable Document Format
RDBMS	Relational Database Management System
RDF	Resource Description Framework
RELAXNG	REgular LAnguage for XML Next Generation
SGML	Standard Generalized Markup Language
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
SVG	Scalable Vector Graphics
TACT	Textual Analysis Computing Tools
TEI	Text Encoding Initiative
TIFF/TIF	Tagged Image File Format
URI	Uniform Resource Identifier
URL	Universal Resource Locator
UTF	Unicode Transformation Format
W3C	World Wide Web Consortium
XML	Extensible Markup Language
XSL-FORSKNING	Extensible Stylesheet Language Formatting Objects,
XSLT	Extensible Stylesheet Language Transformations

Termlista

Term	Beskrivning
absolut referens	referensen anges i sin helhet, fullständig sökväg, t. ex: < http://www.uu.se/ >
analog	data representeras av kontinuerligt varierande värden, i kontrast till digital som är diskreta värden
annotera	infoga märkning eller noter
attribut	används tillsammans med ett element för att beteckna typ av utmärkande egenskap, kännetecken eller beskrivning, ett attribut innehåller alltid ett värde, se attributvärde
attributvärde	beskriver en egenskap hos ett objekt
bastext	den text man utgår från vid skapande av nya versioner
binär	tvåställiga värden, eller andra diskreta värden, t.ex. det binära talsystemet
data	vanligen avses någon typ av information, i datorsammanhang kan data även ses som mönster utan meningsinnehåll
databas	viss mängd av data sammanställda i specifikt syfte, ordnade i ett eller flera dataregister
datamodell	beskriver hur data är strukturerat, i t.ex. ett schema eller en databas
datatyp	beskriver formatet eller typen på värden ett dataelement kan ha, t.ex. tecken, text, heltal, decimaltal
deklarativ	beskrivande (märkning, skriptspråk)
digital	data representeras med siffror, t.ex. binärt genom att använda 1 eller 0.
diplomatarisk utgåva	texten skrivs av (transkriberas) såsom den är, inklusive fel-skrivningar och allt annat.
divisioner	ett sätt att särskilja hela och delar av innehållsobjekt/dokument utan att på förhand ange genre eller typ
dokument	bärare av skriftlig men även annan information
dokumentformat	se filformat
dokumenthierarki	dokumentets interna struktur och delar och dessas inbördes relationer
dokumentmodell	beskrivning av de objekt, deras egenskaper och inbördes relationer som utgör dokument(en). Jfr en bok som består av kapitel, rubriker, stycken, meningar, ord, interpunktion etc.
dokumentstruktur	organisering av dokumentets delar (t.ex. kapitel, stycken, meningar)
dokumenttypsdefinition (DTD)	den modell eller det schema som används för att ange hur innehållet skall beskrivas och struktureras samt på vilka sätt och på vilka hierarkiska nivåer en viss sorts element kan användas, se även datamodell
dynamisk tesaurus	dynamisk ämnesordbok
elektronisk	sådant som är eller nyttjar elektronik

Term	Beskrivning
element	en behållare som består av en 1) ”starttagg” och en 2) ”sluttagg” vilka tillsammans markerar, beskriver och innesluter ett visst 3) innehåll; består av samtliga dessa tre delar, innehållet kan bestå av specifika värden eller andra element
elementuppsättning	grupper av element som har något gemensamt eller som avses användas för en viss genre
entitet	objekt med en varaktig identitet eller värde utöver dess attribut
exekvera	körning av program, dvs. utförande eller verkställande av instruktioner uttryckta i något program- eller skriptspråk
extern annotering	noteringar och beskrivningar som inte infogas i dokumentet utan lagras utanför, t.ex. i en separat fil
faksimil	en avkopierad eller avfotograferad avbildning av en utgåva, måldokumentet strävar efter att efterlikna startdokumentet
faktoid (eng. factoid)	inom prosopografi: en viss källa som hävdar något om en viss person vid en viss tidpunkt
filformat	avser den interna strukturen för lagring av information i digitala datafiler, filformatet beskriver hur det digitala innehållet i en fil ska kodas/tolkas
filologi	texttolkning och textkritik av innehåll i dokument
format	kodning av data i en datafil för konvertering av 1:or och 0:or till läsbar information för människor, olika typer av format används för olika sorters information, textbaserad information använder inte samma metoder som bildbaserad information, se även <i>filformat</i>
formatera	skapande av strukturer för att lagra data, t.ex. ett filsystem eller ett system för organisering av bibliografisk information
inbäddad annotering	märkning eller annan notering som infogas i samma dokument som det beskrivna innehållet
innehållsobjekt	de logiska, semantiska eller strukturella delar som en text (ett dokument) består av
interoperabilitet	förmågan hos olika system eller program att fungera tillsammans och kommunicera med varandra
intertextualitet	studiet av relationer mellan texter
kodning	översättning av data från ett format till ett annat
kompilera	översättning av instruktioner skrivna i programspråk till maskinkod
konkordans	lista eller register över ord som finns i ett eller flera verk eller texter, ofta visas utvalt ord mitt på raden omgiven av ett antal ord före och efter
korpus	på särskilt sätt sammanställa en viss avgränsad textmängd, används oftast inom språkvetenskap
källkod	instruktioner, data och kommentarer i ett givet programspråk eller skriptspråk, är läsbar för människa och kan kompileras till maskinkod
lagringsformat	se <i>format</i> , <i>filformat</i>
maskinkod	instruktioner som en dator kan utföra direkt utan att tolka dem
masterfil	speciellt iordningställt basdokument eller förlaga som används i sin helhet eller i delar för vidare bearbetning eller presentation

Term	Beskrivning
materialitet	analytiskt begrepp för betydelsen av det textbärande mediet (papper, digitalt etc.), kan variera utan att det språkliga/semantiska innehållet förändras
metadata	beskrivande information om data/innehåll
metaspråk	ett språk som används för att beskriva ett annat språk (t.ex. SGML, XML)
migrera	att överföra data mellan olika lagringsformat och/eller lagringsplatser.
modularisera	dela in, avgränsa eller sammanfoga fristående moduler till en mindre eller större helhet
måltext	den text som produceras utifrån en starttext
märkning, märka	metod för att kombinera innehåll (text, bild etc.) med information om detta innehåll
märkningsschema	det schema som används för att beskriva och bestämma hur märkningen skall/bör utföras
märkspråk	språk som används för att beskriva och strukturera text eller grafik (t.ex. TEI, GraphML, HTML)
namnrymd	en abstrakt behållare som används för att identifiera kontexten för och särskilja begrepp och namn från varandra, ett elementnamn får inte ha två värden i samma kontext
normalisera/normalisering	textens stavning enhetligas utifrån någon princip, t.ex. moderniserad stavning
notation	system som används för att ge konventionella eller speciella tecken en specifik innebörd
nästla	ungefär det som är inuti och kommer närmast efter i en viss ordning
onomastik	namnforskning, ofta person- och ortsnamn
ontologi	avgränsad samling begrepp och sammanhang inom ett visst system inklusive begreppens definitioner som fastställer betydelsen inom detta område och beskriver hur termer kan kombineras och användas
operativsystem	består av flera program som ansvarar för datorns grundläggande funktioner, t.ex. för inmatning av data, att starta körning av program, att stänga av datorn, ett operativsystem påverkar datorns funktion mer än maskinen, exempel på o.s. är Unix, Debian (Linux) & Windows
operator	symbol som indikerar att en operation skall utföras på en eller flera tal eller begrepp (operander), t.ex. AND, OR och NOT används för att innesluta, utesluta eller kombinera data t.ex. vid sökning på webben eller i en databas
parser	en tolk som analyserar en dataström
plattform	grundläggande förutsättning för att köra datorprogram, består av en viss sorts hårdvara, operativsystem och program eller kombination av dessa
post	sammansatt (data)typ bestående av flera sammanhängande element, t.ex. en personpost kan bestå av namn, födelsedatum, bostadsort
procedurell	kod skrivs in i den ordning som den ska köras (exekveras), funktionsinriktat sätt att utföra operationer
programspråk	språk som använder en strikt syntax och semantik för att uttrycka beräkningar som skall utföras av en dator, kompileras normalt till maskinkod före exekvering

Term	Beskrivning
proprietära program	program som kontrolleras av företag eller annan rättighets-innehavare och försetts med restriktioner för användning, kopiering, vidareutveckling etc.
prosopografi	samling av uppgifter om individer, kan ofta liknas vid en kollektivbiografi
referensmarkör	ett tecken eller annat som anger en specifik insättningspunkt i t.ex. en text eller på en dataskärm
relativ referens	sökväg eller länk relativt den aktuella hierarkiska kontexten i en filstruktur, ex. /docs/doc.xml
rendering	den process som en dator utför för att framställa text, bild eller dylikt
rotelement	det överordnade elementet i en hierarkisk dokumentstruktur
schema	beskriver struktur för data/innehållsobjekt samt dessas format
scriptons	grundläggande textelement/innehållsobjekt som de framträder för en användare
skript	textfil med instruktioner eller kommandon till datorprogram eller ett operativsystem
skriptspråk	skriptspråk är ett tolkat programspråk med "egen" exekveringsmiljö som kan användas på olika operativsystem, kompileras normalt inte till maskinkod
syntax	regler för hur man sätter samman enskilda enheter till komplexa strukturer i ett specifikt språk (programspråk eller mänskligt språk)
tagg	en markör som används för att markera början och slut på ett märkord
teckenkodning (textkodning)	metod för att representera eller koda en vald uppsättning tecken
textkritik	vetenskapliga studier av text(er) i syfte att åstadkomma t.ex. vetenskapliga utgåvor som innehåller variantapparat, språkliga kommentarer, tidsbestämning etc.
textons	grundläggande textelement/innehållsobjekt som de är arrangerade i texten
textsträng	(även sträng eller teckensträng) en datatyp för lagring av tecken/text oftast avsedd för läsning av människor.
texttypologi	systematiserad klassificering av textens karaktär baserat på ett bestämt urval av väldefinierade egenskaper (t.ex. textons/scriptons)
textualitet	hur ett textinnehåll är utformat och materialiserat
transformering	omvandling av innehållsobjektens form, format, ordning eller andra egenskaper.
transversal	den mekanism som framställer och ordnar information: som ordnar scriptons, genererade ur textons, till text

Förord

Mitt avhandlingsprojekt är nu fullföljt och jag vill rikta ett varmt tack till alla er som hjälpt mig att genomföra och slutföra detta arbete.

Först och främst vill jag tacka Donald Broady som varit min handledare. Donald har uppmuntrat mig på alla sätt och styrt mig lagom mycket, gett mig viktig kritik och hjälpt till med språktvätt. Han har introducerat mig i nya sammanhang som lett till att jag skaffat mig kontakter både i Sverige och utomlands. Tiden som doktorand har varit en lärorik period fylld med spännande utmaningar. Det känns också fint att vi är två som betraktar tjocka manualer för märkspråk med beskrivningar om element och attribut som intressant läsning.

Jag vill också rikta ett stort tack till alla andra medlemmar i forskningsgruppen för utbildnings- och kultursociologi (SEC). Ett särskilt varmt tack till er på Seminariet för inspirerande gemenskap och gott samarbete kring projektansökningar, workshops, bokutgivning, registeranalyser och annat som en forskningsgrupp sysslar med.

Jag har haft förmånen att delta i projektet ”Formering för offentlighet. En kollektivbiografi över Stockholmskvinnor 1880-1920”.¹ Ett varmt tack till alla medarbetare i projektet för ert engagemang och tålamod med mitt tidsödande arbete med det prosopografiska materialet. Tack också för att jag har haft förmånen att delta i, och lära av, en gemenskap där offentliga kvinnor, källmaterial, klassificering och många andra frågor livligt har diskuterats.

Tack till alla på Pedagogiska institutionen för gott samarbete och trevliga stunder med och skratt i fikarummet. Ett särskilt varmt tack till doktorandgruppen – en härlig och varm gemenskap där vi har delat stort och smått under årens lopp.

Avhandlingsarbetet har kunnat genomföras tack vare en doktorandanställning på Pedagogiska institutionen, finansierad av Samhällsvetenskapliga fakulteten. Jag har därtill erhållit bidrag från Riksbankens Jubileumsfond genom projektet ”Formering för offentlighet”. Mina lärorika vistelser vid Oxford Text Archive, University of Oxford och besök vid Centre for Computing in the Humanities, King’s College London, har kunnat genomföras tack vare resebidrag från Göransson-Sandviken resestipendium som förvaltas av Gästrike-Hälsinge nation, samt med ett bidrag från projektet ”Märkning av utbildningsinnehåll” finansierat av Vinnova. Den ”Exploratory Workshop” om prosopo-

¹ Förteckning över medarbetare i projektet Formering för offentlighet, se <www.skeptron.uu.se/broadly/sec/ffo.htm>

grafiska metoder som anordnades vid Uppsala universitet i maj 2007 och finansierades av European Science Foundation och Riksbankens Jubileumsfond har också haft betydelse för avhandlingsarbetet.

De som stöttat och tålmodigt hjälpt mig att färdigställa avhandlingen genom att läsa och kritisera under slutskedet av avhandlingsskrivandet vill jag särskilt uppmärksamma. Stort tack till Mats Dahlström vid Bibliotekshögskolan i Borås som gav mig mycket hjälp med avhandlingsmanuset hösten 2007. Utan din initierade och välstrukturerade kritik hade det blivit mycket svårare att färdigställa avhandlingen.

Till er andra som hjälpt till med läsning och handfast stöd i slutfasen av avhandlingsarbetet vill jag rikta min stora tacksamhet; tack till Esbjörn Larsson, Gunilla Roos, Ulla Riis, John Davies och Emil Bertilsson för er hjälp med språktvätt, korrektur och kloka synpunkter. Tack också Ida Lidegran och Ingrid Nordqvist för synpunkter på och tålmodig läsning av mina ofärdiga avhandlingskapitel.

Sist men inte minst alla mina hjärtligaste och varmaste tack till min älskade och underbara familj. Min man, Erik Zetterman, har tålmodigt hjälpt mig med att sammanställa de digitala bilagorna och med korrekturläsning samt gett mig initierade synpunkter på termlistan. Mina döttrar Agnes och Alice Langerth har varit till stor hjälp vid fotografering och transkribering av Kungl. Högre Lärarinneseminariets arkivmaterial. Min syster, Kerstin Käis, har läst delar av manuset. Tack också för att ni har stöttat, hejat på och uppmuntrat mig under hela tiden. Under de senaste månaderna har jag dessutom haft full service. Erik, Agnes och Alice, ni är bäst i hela världen!

DEL I Inledning

1 Utgångspunkter

Användning av digitala material, arkiv och databaser är en viktig förutsättning för såväl forskning som undervisning. Det blir dessutom allt vanligare att forsknings- och undervisningsmaterial blir tillgängligt på webben. En av förutsättningarna för utvecklingen av internet och webben är användningen av öppna standarder och öppen källkod. Själva begreppet öppen källkod – *Open Source* – innebär att informationen är fri vilket gör att fler kan hjälpa till att granska och utveckla program. Det vill säga ungefär som det fungerar inom forskning och undervisning där arbetet ofta går ut på att använda, utforska, utveckla, förfina, sammanställa eller bygga vidare på vad andra gjort. Genom att kombinera forskningens och undervisningens behov av tillgänglighet och öppenhet med tillämpningar som härrör ur *Open Source*, ges möjlighet att erbjuda digitala tillämpningar som utformats med program som är lätt använda och där arbetsprocessen är transparent.

Valet att använda historiskt material i den föreliggande studien, grundas i såväl mitt eget ämnesområde som intresset för det historiska materialets komplexitet. Ett utbildningshistoriskt material är dels av pedagogiskt intresse, dels tillhör historiskt material den mer komplexa typen av texter och avancerade tillämpningar man kan använda för forskning och undervisning. Materialet i den här studien härstammar från en period kring sekelskiftet 1900 och erbjuder många möjligheter för studier av kvinnors väg till offentlighet, deras verksamhetsområden, lärarinneutbildning, sociala kontaktnät och föreningsliv. Avhandlingen handlar dock inte primärt om borgerliga kvinnors väg till offentlighet utan om att skapa förutsättningar för sådana studier och en mångsidig användning av digitalt material för forskning och undervisning.²

I avhandlingen undersöker jag hur man kan gå tillväga för att digitalisera och sammanställa uppgifter hämtade ur olika slags källor samt ordna, bearbeta och använda dessa uppgifter för att utforma historiska digitala material med hjälp av i första hand öppen källkod och programvara. Arbetsuppgifterna kan i praktiken bestå av att beskriva, ordna och bearbeta digitalt innehåll med hjälp av relevanta standarder och i plattformsoberoende filformat.³ Den som utför dessa

² Några resultat av statistiska analyser som gjorts på det prosopografiska materialet presenteras i Monica Langerth Zetterman, *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet*, Rapporter från Forskningsgruppen för utbildnings- och kultursociologi, 1 ed., vol. 41 (Uppsala: Forskningsgruppen för utbildnings- och kultursociologi, 2008).

³ Filformat som inte är beroende av en viss sorts operativsystem eller programvara.

uppgifter kan kallas teknisk redaktör, kunskapsarkitekt eller innehållsarkitekt, och själva verksamhetsområdet kallas här för *innehållsdesign*.⁴

Användningen av digitala arkiv, samlingar och databaser utgör en stor del av forskares, lärares och studenters vardagliga arbete. I dag är det knappast någon ovanlighet att använda webblesurser i undervisningen eller för forskning. Inom högskola och universitet används webbplatser, bibliotekskataloger, digital publikationer, webbaserade databaser och lärplattformar i stora delar av verksamheten. Utvecklingen av metoder och verktyg för internet medför flera nya möjligheter för användning av digitalt material för forskning och undervisning.

Forskare, lärare och studerande är dock fortfarande ofta beroende av hur digitalt material presenteras och framställs på den webbplats som för tillfället besöks. Många webbplatser presenterar innehållet med en snygg inramning men att arbeta med innehållet brukar inte vara möjligt. Det saknas emellertid inte teknik för att åstadkomma detta. Det finns flera slag av lämpliga metoder och verktyg för att utforma digitalt innehåll så att materialet kan brukas och återanvändas på flera sätt. Med hjälp av märkning kan man infoga information om själva innehållet – i detta sammanhang text – för att beskriva fysiska, logiska eller semantiska strukturer. Därmed skapas förutsättningar för att särskilja innehåll från form, något som senare kan nyttjas för vidare bearbetning och som medför att innehållet kan ordnas och omordnas, presenteras och användas på skilda sätt och i olika tillämpningar.⁵ Arbetsuppgifterna kan antingen utföras av forskare, lärare och studenter eller av tekniska redaktörer. Konkret kan det handla om att tillhandahålla delar av större material på webben och samtidigt ge användare möjlighet att studera eller ladda ner hela materialet, till exempel en databas, som kan användas för andra typer av analyser än de som presenteras på just den webbplatsen. Det kan även handla om att upprätta ett personligt register över sina bibliografiska referenser som inte är bundet till en specifik programvara. Man kan också tänka sig att lärares undervisningsmaterial skulle kunna delas in i moduler och fördes med beskrivningar som medför att detta undervisningsmaterial kan anpassas till olika kurser och skilda målgrupper.

För att utforska dessa metoders möjligheter och begränsningar används historiskt material om kvinnor i offentlig verksamhet från 1880-talet fram till 1940-talet.⁶ En del av det material som används i den här studien har producerats av medarbetare i projektet ”Formering för offentlighet. En kollektivbiogra-

⁴ Begreppet innehållsdesign presenteras i Donald Broady, *Content design. Methods and tools for the creation of portable hypermedia archive s. Notes for a proposed CID project. Version 2* (Nada/CID, KTH, 1997) s. 18-19; Donald Broady, ”Digitala arkiv och portföljer” *IT i skolan: mirakelmedicin eller sockerpiller?* Observatoriet för IT lärande kunskap och kompetens, Rapport 45/2001 (Stockholm: IT-kommissionen, 2001).

⁵ Här avses till exempel olika slags analyser, transformering, extrahering, samt presentation till exempel publicering på webben.

⁶ Materialet har samlats in med den prosopografiska metoden, som är ett slags ett slags kollektivbiografier. Se vidare s. 133 ff.

fi över stockholmskvinnor 1880-1920”.⁷ Den ena delen av detta material består av både nyskrivna biografiska texter och avskrifter ur en biografisk uppslagsbok. Min roll som teknisk redaktör i projektet *Formering för offentlighet* har varit att iordningställa, märka och analysera det material som producerats och samlats in i projektet. Målsättningen har varit att kunna utföra statistiska analyser på materialet samt att tillgängliggöra och presentera materialet på webben. Den andra delen av det historiska material som används i avhandlingens tillämpningar samlades in och iordningställdes av mig i ett senare skede i avhandlingsarbetet. Detta material består av handlingar i Kungl. Högre Lärarinneseminiariets arkiv som fotograferats och registrerats samt avskrifter av dessa.

Genom att utforska metoder och verktyg för innehållsdesign är min förhoppning med denna studie inte bara ska skriva om hur det skulle kunna vara, eller granska hur andra har gjort, utan även visa på och analysera innehållsdesign teoretisk och praktiskt.

När avhandlingsarbetet påbörjades hade jag vissa kunskaper i HTML och databashantering i övrigt har jag lärt mig nödvändiga tekniska kunskaper under arbetets gång. Med detta vill jag ha sagt att det inte på något sätt är omöjligt för en intresserad lärare, forskare eller student att lära sig tillämpa de metoder och verktyg som utforskas och presenteras i avhandlingen.

Användningar av digitala samlingar⁸ och andra datamängder har blivit en allt viktigare förutsättning för forskning men även för undervisning. Inom samhällsvetenskaperna finns en tradition att bygga register av olika slag. Här används till exempel longitudinella databaser, individbaserade register eller register av specifika sociala fenomen.⁹ Humanister arbetar oftare med digitaliserat arkivmaterial, äldre tryck och språkkorpusar.¹⁰

Tillgången till internationella digitala arkiv med historiskt källmaterial, klassiska utgåvor, vetenskapliga skrifter och textkorpusar är idag relativt god.¹¹

⁷ Donald Broady, et al., *Formering för offentlighet. En kollektivbiografi över Stockholmskvinnor 1880-1920* (Uppsala: SEC/ILU, Uppsala universitet, 1998). Förteckning över medarbetare, information om projekt och bibliografi, se <www.skeptron.uu.se/broad/SEC/ffo.htm>

⁸ I avhandlingen används, om inget annat anges, digitala arkiv och digitala samlingar som samlingsbegräpp för olika typer av digitaliserade samlingar inom olika genrer, t.ex. textkritiska editioner, manuskriptsamlingar, fulltextarkiv och digitaliserat arkivmaterial.

⁹ Som exempel kan nämnas de register som används inom Forskningsgruppen för utbildnings- och kultursociologi (SEC) och som innehåller omfattande datamaterial från SCB över elever i svensk grund- och gymnasieskola, studerande i högskolan samt sociogeografiska data för den svenska befolkningen, för information om SEC:s forskningsprojekt där dataregistren används se: <www.skeptron.uu.se/broad/SEC/research.htm>, en översikt av dataregistren finns tillgänglig vid: <www.skeptron.uu.se/monicalz/esf/slides/Borjesson-Mikael.pdf>. Den demografiska databasen vid Umeå universitet är ett annat exempel på en omfattande forskningsdatabas, se <www.ddb.umu.se/>.

¹⁰ Vid Stockholms stadsarkiv finns ett antal databaser och register tillgängliga på webben, <www.ssa.stockholm.se/>. Vid den nationella arkivdatabasen finns det samlade nationella arkivbeståndet, <www.nad.ra.se/>. Stockholm Umeå Corpus (SUC) är en (morfosyntaktiskt) annoterad språkkorpus som utvecklats sedan 1992, se <<http://spraakbanken.gu.se/parole/>>.

¹¹ För Sveriges del se exemplen i noten ovan, internationellt sett finns en mängd fritt tillgängliga projekt och databaser inom samhällsvetenskap och humaniora vara flera av dem hänvisas till i avhandlingen.

Trots detta för de digitala arkiven en relativt undanskymd tillvaro som verktyg och hjälpmedel inom utbildningshistorisk forskning och utbildning i Sverige.

För utbildningshistorisk forskning och undervisning kan man tänka sig att tillgång till mer komplext och rikhaltigt innehållsmärkt material skulle kunna vara av intresse. Sökningar i biblioteksdataserna resulterar dock i ett tämligen magert resultat för svensk utgivning av texter om användning av digitala arkiv och samlingar eller vetenskapliga utgåvor om specifika arkiv i utbildningshistorisk forskning och undervisning.¹² Flertalet av titlarna som finns är dessutom inte utgivna i Sverige. Det finns dock undantag, bland annat en metodbok med digitala historiska dokument som handlar om händelserna i Ådalen 1931, som jag beskriver närmare senare i detta kapitel.¹³ Nedan ger jag exempel på webbplatser som dels gjort material tillgängligt och dels tillåter användaren att ladda ner material.

1.1 Problembeskrivning

1.1.1 Digitala samlingar för forskning och undervisning

De tekniska förutsättningarna öppnar som sagt upp för en rad olika möjligheter för forskning och undervisning. Det finns nu verktyg för att explicit beskriva, ordna, bearbeta, manipulera och presentera innehåll efter önskemål och behov. Digitala arkiv har under den senaste tjugoårsperioden utvecklats i rask takt, framför allt under de senaste tio åren. Det finns en mängd olika typer av digitala samlingar tillika många olika aktörer. Nedan ger jag exempel på några samlingar som är av direkt eller indirekt intresse för avhandlingen. Vissa samlingar är organiserade som volontärprojekt som publicerar upphovsrättsfria klassiska texter på webben.¹⁴ Problemet med dessa digitala utgåvor är dels den ibland bristfälliga korrekturen, den rudimentära märkningen och att de textkritiska aspekterna, till exempel versionsval, inte är i fokus, vilket innebär att dessa utgåvor inte alltid är lämpade för forskningsändamål.¹⁵

Ett exempel på en webbplats som nyttjar webbtekniken väl är demografiska databasen vid Umeå universitet där man producerar och tillhandahåller historiska data av material i befolkningsdatabaser med uppgifter baserade på kyrkböcker från 1700- och 1800-talen. Demografiska databasen uppstod i början av 1970-talet ur pedagogen Egil Johanssons forskning. Han kombinerade informa-

¹² Sökningar har gjorts med sökorden: utbildningshistoria | historia + digitala arkiv | digitala samlingar | digitala medier | datorstött undervisning.

¹³ Roger Johansson, *Ådalen 1931 - minuterna som blev historia: metodbok till en digital dokumentsamling* (Lund: Studentlitteratur, 2006).

¹⁴ Projekt Runeberg publicerar sedan 1992 digitala utgåvor av klassisk nordisk litteratur, se <<http://runeberg.org/>>. Project Gutenberg publicerar framför allt utgåvor av anglosaxiska texter, se <www.gutenberg.org/>.

¹⁵ Johan Svedjedal, *En svensk Litteraturbank? Utredning för Svenska Akademien, avlämnad i januari 2003* (2003) s. 17.

tion från husförhör med den tidens moderna statistiska metoder och kunde härigenom studera läskunnighetens framväxt hos den svenska befolkningen.¹⁶ Användare av den demografiska databasen har fri tillgång till delar av det inskannade källmaterial, samt möjlighet att göra sökningar i vissa databaser såsom ”Tabellverket”.¹⁷ Resultaten av sökningen visas i tabellform i HTML-format vilket gör det möjligt för användaren att ladda ner och använda materialet lokalt på sin egen dator.¹⁸ Det material som finns tillgängligt i den demografiska databasen introduceras och beskrivs relativt fylligt och arbetsmetoderna är väl dokumenterade. Dessutom finns förslag på uppsatsämnen, litteraturguider och länklister, något som är användbart för undervisning. Denna tidiga databas var länge unik, inte bara inom pedagogikforskningen, utan har även använts av historiker, kulturgeografer och medicinare. Redan under 1980-talet nyttjades databasens material för att skapa läromedel för ungdomsskolans bruk.¹⁹

Ett annat exempel är de kvinnohistoriska samlingarna vid Göteborgs universitetsbibliotek.²⁰ Där finns för närvarande tre kvinnohistoriska portaler vid de kvinnohistoriska samlingarna:²¹ ”Kvinnors kamp för kunskap”,²² ”Kvinnors kamp för rösträtt”²³ samt ”Kärlek, makt och systerskap”.²⁴ De två första ämnesportalerna innehåller, förutom introduktioner och översikter till ämnet, även inskannat arkivmaterial av pamfletter, politiska beslut och protokoll, medlemsmatriklar, tidningar och andra texter. Allt material som kan laddas ner är i PDF-format vilket ger en avbildning av originaldokumentet men erbjuder inte en vidare användning än presentation på skärm eller utskrift. Den tredje portalen är en webbutställning om svensk kvinnorörelse på 1970-talet. Portalen innehåller inte källmaterial utan framför allt introducerande texter och länkar till föreningar och organisationer.

¹⁶ Egil Johansson, *En studie med kvantitativa metoder av folkundervisningen i Bygdeå socken 1845-1873*, Akademiska avhandlingar, vol. 2 (Umeå: Pedagogiska institutionen, Umeå universitet, 1972).

¹⁷ <www.ddb.umu.se/tabellverk/Atkomst/tabnet.htm>

¹⁸ Detta kräver förstås att användare respekterar sedvanlig forskningsmoral och anger källor i eventuella publikationer. Vid webbplatsen anges också tydligt att databasen är deras egendom och att det inte per automatik är: [...] tillåtet att överföra informationen till en egen databas eller att använda denna information för kommersiella syften.

<www.ddb.umu.se/tabellverk/Atkomst/tabnet.htm>

¹⁹ Bengt Grensjö & Berthel Sutter, ”Forskande lärande i skolans och arkivens värld” *Arkiven i forskningens tjänst: 20 år med Forskningsarkivet vid Umeå universitet*, red. Göran Larsson, (Umeå: Forskningsarkivet Umeå universitet, 2003).

²⁰ År sedan 1997 nationellt ansvarsbibliotek för kvinno-, mans- och genusforskning, se <www.ub.gu.se/kvinn/>.

²¹ Verksamheten består av, förutom de ovan nämnda ämnesportalerna, att förteckna kvinnohistoriskt handskriftsmaterial samt ge referensservice samt att tillhandahålla databaserna KVINNSAM (Nordens största databas inom genusområdet) och GENA (GENusavhandlingar).

²² <www.ub.gu.se/kvinn/portaler/kunskap/>.

²³ <www.ub.gu.se/kvinn/portaler/rostratt/>.

²⁴ <www.ub.gu.se/kvinn/portaler/systerskap/>.

Vid Göteborgs universitet finns Språkbanken som tillhandahåller olika slags digitaliserade texter, lexikala material och parallellkorpusar.²⁵ Till den omfattande samlingen, som innehåller 75 miljoner ord, finns olika slags verktyg att tillgå för analyser av texterna.²⁶ Ett annat exempel på en svensk digital utgivning som vänder sig till forskare och studerande samt litterärt allmänintresserade och som erbjuder viss flexibilitet är, "Litteraturbanken: en digital resurs för svensk skönlitteratur och humaniora", som nämndes även i inledningen. Vid Litteraturbanken samlar man in, arkiverar och publicerar skönlitteratur och andra texter.²⁷ Användare kan, vid själva webbplatsen, även hantera och arbeta med materialet, såsom att infoga egna kommentarer eller att skapa egna antologier. Litteraturbanken målsättning är att skapa ett arkiv för läsning och studier av digitala versioner av klassiska verk och andra texter.²⁸ Möjligheterna att med hjälp av den här sortens digitala samlingar uppmärksammas av Lars Borin som framhåller att:

Jag vill hävda att Litteraturbankens verksamhet kan sägas utgöra ett specialfall av *användning av text* i humanistisk forskning och utbildning. Jag tror den sortens textanvändning kan berikas av de möjligheter som språkteknologin erbjuder för att skapa nya verktyg för forskning och undervisning i humaniora.²⁹

Vad Borin avser med nya möjligheter är bland annat texters tillgänglighet och de möjligheter till dynamik som texttekniker kan erbjuda. Textteknikens möjligheter till automatiserad märkning av innehåll i samma stund som dokumenten blir digitalt tillgängliga bör inte heller underskattas menar han. I detta avseende finns många möjligheter att samordna kunskaper och resurser från det litterära fältet med det språkteknologiska, inte minst med tanke på snabbare och bättre tillgänglighet och verklig användning av digitala material.

Bland internationella resurser kan särskilt Oxford Text Archive (OTA)³⁰ framhållas. OTA är det äldsta arkivet för lagring av elektroniska resurser och de handhar nu mer än 2000 digitala resurser, framför allt anglosaxiska kanoniska

²⁵ Språkbanken har funnits sedan 1975, presentation om språkbanken tillgänglig <<http://spraakbanken.gu.se/>>.

²⁶ Språkbanken tar också fram en språklig "verktygslåda" som omfattar verktyg som "tokeniserare" (program som renser textord från bl.a. interpunktion), "lemmatiserare" (program som återför böjningsformer till en huvudform), ett morfosyntaktiskt taggningsprogram (som automatiskt markerar ordklass på orden i en text) och ett textlänkingsprogram (som skapar parallella texter). Citat från <<http://spraakbanken.gu.se/>> under rubriken "Om Språkbanken".

²⁷ Svenska Akademien har ansvaret för Litteraturbanken som sedan 2006 drivs som en ideell förening, se vidare <<http://litteraturbanken.se/>> samt Svedjedal, *En svensk Litteraturbank? Utredning för Svenska Akademien, avlämnad i januari 2003*.

²⁸ De format som erbjuds är e-text, faksimil och PDF. PDF är det enda format som går att "ladda hem" till en lokal dator.

²⁹ Lars Borin, "Vi som går köksvägen. Språkteknologer och korpuslingvister i Litteraturbanken" *Fältanteckningar: utbildnings- och kultursociologiska texter tillägnade Donald Broadly*, red. Mikael Börjesson, et al., (Uppsala: Forskningsgruppen för utbildnings- och kultursociologi ILU Uppsala universitet, 2006) s. 403.

³⁰ Se <<http://ota.ahds.ac.uk/>>.

texter, iordningställda av forskare men fritt tillgängliga att använda för vidare forskning eller undervisning. Arkivet har som främsta syfte att samla och lagra digitaliserade material. Kvaliteten bland verken skiftar och de textkritiska kraven varierar. Men texterna är kodade och märkta av forskare och andra specialister och har iordningställts för filologiska, lingvistiska och stilistiska analyser eller genrebestämningar. Alla OTA:s digitala resurser ackompanjeras av beskrivning om innehållet som förutom den sedvanliga informationen om förlagan i form av författare, titel, förlag, utgivningsår och edition, även innehåller uppgifter om genre, filformat, ansvarig för deposition och märkning, samt vilka konventioner och märkningsscheman som använts. I OTA:s digitala arkiv finns möjlighet att söka i texterna samt göra enkla konkordanser men även spara dokumenten lokalt på egen dator.³¹ OTA erbjuder dock i övrigt inga relaterade verktyg för att arbeta med och bearbeta arkivets resurser. Detta får man som användare leta efter på annat håll.

Det finns exempel på andra digitala arkiv som erbjuder fler möjligheter att hantera och arbeta med innehållet. Men oftast begränsas användare och brukare av det digitala arkivets lokala funktionalitet snarare än en generell funktionalitet med möjlighet att kombinera olika byggstenar utifrån aktuellt behov. Ett exempel på denna typ av generella funktionalitet är *Text Analysis Portal for Research* (TAPoR) där användare erbjuds en rad olika verktyg direkt via portalen för att arbeta med sina egna lokala texter.³² Användaren behöver alltså inte installera eller lära sig särskilt mycket om verktygen. Vet man vad man vill göra eller vilken typ av verktyg man letar efter är det bara att använda sökfunktionen för att leta efter ett lämpligt verktyg. Vet man däremot inte vad man vill göra eller vilken typ av verktyg man kan tänkas behöva, tillhandahålls exempeltexter att utforska.

Vad vi kan se är att utbudet i de digitala arkiven spänner från volontärprojekt som publicerar digitala texter ofta i HTML- eller PDF-format som saknar textkritiska eller vetenskapliga krav, till mer avancerade digitala vetenskapliga utgåvor som genomgått textkritisk granskning och där resurser och material ackompanjeras av verktygslådor för forskning och undervisning.

1.1.2 Behov av tillgänglighet och kompetens

Digitalisering är även viktigt för att göra vårt gemensamma kulturarv tillgängligt. I en utredning från 2003 om Kungl. Bibliotekets verksamhet och arbetsformer, den så kallade KB-utredningen, inleds sammanfattningen med följande uppmaning:

³¹ Konkordans är en lista över ord som finns i ett eller flera verk eller texter, ofta illustreras utvalt ord mitt på raden omgiven av ett antal ord före och efter.

³² "[t]ools for sophisticated analysis and retrieval, along with representative texts for experimentation." Se vidare <<http://tapor1.mcmaster.ca/>>.

Det är nu hög tid att göra en kraftsamling för informationsförsörjningen till forskning och utbildning.³³

I utredningen konstateras även att flera länder i Europa³⁴ satsat stort på digitalisering av kulturarvet och att skälen till detta är många, till exempel av tillgänglighets- och bevarandeskäl. För att Sverige skall kunna hävda sig internationellt är det viktigt att prioritera forsknings- och utbildningssektorn.

För att vi ska kunna ligga i täten måste staten satsa på att utveckla den digitala informationsförsörjningen till utbildning och forskning. Insatserna ska bidra till att regeringens mål om Sverige som ett informationssamhälle för alla uppfylls.³⁵

Digitala material iordningställs numer ofta i samarbete mellan flera forskargrupper eller andra intressenter vilket i sin tur påverkar behovet av en god tillgänglighet. I en kartläggning utförd av Vetenskapsrådet år 2005 gjordes en översikt av den nationella infrastrukturen inom samhällsvetenskap och humaniora, med fokus på material för forskning och som ska finnas tillgängligt, nu och i framtiden. Slutsatserna av kartläggningen tyder på att materialen allt oftare utformas i samverkan mellan flera olika parter, att innehållet tenderar bli alltmer komplext samt att behovet av bättre tillgänglighet ökar. I rapporten föreslås en rad åtgärder, men först och främst betonas behovet av att:

[...] materialet dokumenteras och digitaliseras enligt gemensamma standarder (en förutsättning för integrering av olika material och komparationer).³⁶

I rapporten framhålls humanistiska laboratorier som särskilt viktiga för framtida resursutveckling.³⁷ Därtill nämns att verksamheten vid humanistlaboratorier med stor sannolikhet kommer att ha betydelse inom humanistisk forskning eftersom digitalisering, märkning och statistisk bearbetning spelar en allt större roll för forskning såväl som för undervisning. Det finns alltså all anledning till att studera vad vi kan lära av hur man inom humaniora utvecklat metoder och verktyg för att beskriva och ordna digitala material.³⁸

Kunskaper i hur man iordningställer och utformar digitalt material behöver också förbättras. I rapporten från Vetenskapsrådet noteras att humanistiska

³³ KB-utredningen, *KB - ett nav i kunskapssamhället: Kungl. biblioteket - Sveriges nationalbibliotek: verksamhet och visioner: betänkande*, Statens offentliga utredningar, 2003:129 (Stockholm: Fritzes offentliga publikationer, 2003) s. 17.

³⁴ I utredningen nämns bl.a. Frankrike och Skottland, Ibid. s. 177.

³⁵ Ibid. s. 17.

³⁶ Vetenskapsrådet, *Om forskningens infrastrukturer inom humaniora och samhällsvetenskap i Sverige* (Vetenskapsrådet, 2005) s. 3.

³⁷ Här framhålls HUMlab vid Umeå universitet där informationstekniken studeras, brukas och utvecklas. I rapporten nämns även Språk- och Litteraturcentrum vid Lunds universitet.

³⁸ Till exempel Text Encoding Initiative (TEI), de ledande riktlinjerna för beskrivning, lagring och utbyte av humanistiskt och samhällsvetenskapligt innehåll. Se vidare kapitel 4 s. 102.

forskare ofta saknar kunskaper i bearbetning av forskningsmaterial.³⁹ Ingen liknande rapport finns att tillgå rörande liknande förhållanden inom samhällsvetenskaplig forskning och undervisning. Med tanke på den relativt ringa produktionen av fritt tillgängliga svenska digitala material för forskning och undervisning inom dessa områden kan man anta att forskare, lärare och studenter antingen inte känner till dessa möjligheter eller saknar nödvändiga kunskaper om metoder och verktyg för hantering och utformning av plattformsoberoende digitalt forsknings- och undervisningsmaterial. Istället används ofta proprietära lösningar som medför att det digitala innehållet inte kan flyttas, integreras eller återanvändas i andra sammanhang än vad som var tänkt från början.⁴⁰

Frågan om tillgänglighet är också beroende av hur informationen ordnas och beskrivs. Tillgänglighet och flyttbarhet är i sig inget nytt eller revolutionerande, det formulerades för svensk del bland annat i forskningsprogrammet ”Det nya handbiblioteket” för mer än tio år sedan.⁴¹ Frågan nu är snarare hur metoderna skall användas. De verktyg som behövs är relativt sett lättillgängliga jämfört med tidigare mycket tack vare utvecklingen inom öppna standarder och öppna programvara.

Med uppgiften att beskriva och ordna ett visst innehåll följer nödvändigheten av att somligt inkluderas och accentueras samtidigt som annat dämpas eller exkluderas. Hur man klassificerar, beskriver och ordnar forskningsmaterial styrs av en rad olika faktorer, däribland forskningsområden. Beroende på hur ett innehåll beskrivs och ordnas ges också olika möjligheter till användning. I KB:s utredning betonas vikten av *ordning och reda* och användningen av gemensamma standarder för informationshantering.⁴² Även om KB-utredningen mer fokuserar på nationell samordning och bevarande av kulturarvet i ett vidare perspektiv än vad föreliggande studie gör, är resultaten i utredningen synnerligen relevanta för verksamhetsområdet innehållsdesign. Digitalisering och standardisering är även föremål för utredning av EU-kommissionen. I en rekommendation från EU-kommissionen från augusti 2006 används begreppet *kulturellt innehåll* och även där betonas vikten av gemensamma standarder och normer för effektiv åtkomst och användning av digitalt innehåll.⁴³

Dessutom påverkas beskrivningen och ordningen, dels av teoretiska utgångspunkter och antaganden med därav tillgängliga metoder och verktyg, dels av frågor av mer praktisk natur såsom källmaterialets kvalitet, forskningsgruppens sammansättning och kunskapsområden. Något som också kräver breda

³⁹ Här avses svenska förhållanden, se Vetenskapsrådet, *Om forskningens infrastruktur inom humaniora och samhällsvetenskap i Sverige* ssk. s. 8.

⁴⁰ Flera av de digitala samlingar som finns i och utanför Sverige kommer att nämnas nedan.

⁴¹ Donald Broady, ”Det nya handbiblioteket” *Biblioteken, Kulturen och den sociala intelligensen. Aktuell forskning inom biblioteks- och informationsvetenskap*, red. Lars Höglund, (Göteborg: Forskningsrådsnämnden/Valfrid, 1995).

⁴² KB-utredningen, KB - ett nav i kunskapssambället: *Kungl. biblioteket - Sveriges nationalbibliotek: verksamhet och visioner: betänkande*, s. 149.

⁴³ *Kulturellt innehåll - digitalisering, elektronisk tillgång och digitalt bevarande* (Europeiska unionens officiella tidning, 2006).

och goda kunskaper om vilka valmöjligheter som finns. Det som är obekant eller upplevs som krångligt kommer inte att efterfrågas. För att nyttja fördelar och möjligheter med innehållsdesign kanske det behövs en bättre ”beställar-kompetens” hos lärare och studerande samt tillämpningar där metoder och verktyg kan prövas.

1.1.3 Behov av standarder och verktyg

Inledningsvis nämnde jag att öppen källkod och öppna standarder är en viktig komponent för innehållsdesign. Användningen av IT och digitala material i undervisning på alla nivåer har idag en stor roll och kommer troligen att få en ännu större roll i framtiden. I IT-standardiseringsutredningen slås det fast att användningen av öppna standarder bör ske i större utsträckning och att formell standarder skall användas av svenska myndigheter.⁴⁴ Vidare poängteras det att Sverige bör verka för att använda öppna och brett accepterade *de facto* standarder.⁴⁵ Även om man i utredningen inte ägnar utrymme åt högskoleundervisning och forskning framhålls ändå att användningen av öppna standarder kan vara av stor betydelse för att skapa goda lärmiljöer och förutsättningar för en individualiserad undervisning.⁴⁶ I flertalet skolor används proprietära programvaror⁴⁷ i undervisningen.⁴⁸ Många lärare anser att tillgången till digitala läromedel skulle kunna vara bättre än vad den för närvarande är. Den grupp som var mest nöjd med tillgången är skolledarna, det vill säga de som är längst från själva undervisningen.⁴⁹

Tillgången till digitalt undervisningsmaterial behöver förstås inte ha något direkt samband med vilket operativsystem som används i skolan. I många fall används de proprietära programvarorna för till exempel dokumenthantering och kalkylering i undervisningen för att de följer med det operativsystem som används i verksamheten. Dessa proprietära programvaror ger inte de bästa förutsättningar som behövs för att flytta material till andra plattformar, andra

⁴⁴ IT-standardiseringsutredningen, *Den osynliga infrastrukturen: om förbättrad samordning av offentlig IT-standardisering: betänkande*, Statens offentliga utredningar, 2007:47 (Stockholm: Fritze, 2007). Utredningens uppdrag var att föreslå hur IT-standardiseringen i offentlig förvaltning bättre skulle kunna samordnas samt om och hur användningen av öppen programvara bör främjas.

⁴⁵ Exempel här är W3C: s standarder för webben och OASIS standarder för digitala dokument, se vidare avsnitt 2.3.2 s. 65 f.

⁴⁶ IT-standardiseringsutredningen, *Den osynliga infrastrukturen: om förbättrad samordning av offentlig IT-standardisering: betänkande* s. 395 f.

⁴⁷ Med proprietär avses sådana program som kontrolleras av t.ex. företag och som har försetts med restriktioner för användning, kopiering och vidareutveckling.

⁴⁸ I skolektorn används företrädesvis Microsofts operativsystem, ca 90 procent av de medverkande skolorna uppger att de använder det och endast en liten del använder operativsystem med öppen källkod. *IT och skola 2005*. s. 17.

⁴⁹ *IT och skola 2005. En sammanfattning av en enkätstudie riktad mot skolledare, lärare och elever under våren 2005*. (Skolverket, Myndigheten för skolutveckling, Sveriges kommuner och landsting, Stiftelsen för kunskap- och kompetensutveckling, Microsoft, 2005) s. 23 & 60.

format eller andra typer av program. Risken är att man stannar kvar på den väg som från början valts – istället för att ta alternativa vägar.

En annan viktig aspekt vid digitalisering är även modularitet där märkningen, som en av andra möjliga metoder, kan medverka till att materialet blir utbyggbart, modifierbart och flyttbart så att andra kan bruka materialet utifrån egna önskemål:

Ett väl strukturerat och väl beskrivet innehåll behöver inte vara bundet till vissa presentationsformer eller undervisningsformer. Innehållet kan vara utspjitt på nätet, det kan vara synnerligen heterogent och det behöver inte alls vara avsett enbart för undervisningsbruk. Om det är väl strukturerat och väl beskrivet kan lärare, studenter och elever överblicka och genomsöka det och sammanfoga material hämtat från olika håll.⁵⁰

Ett led i den riktningen är användningen av vedertagna märkningsscheman och riktlinjer för märkning, där innehållet, snarare än formen är i centrum.⁵¹ Om digitala arkiv var konstruerade utifrån en generell funktionalitet, som ett slags råämne, skulle forskare, lärare och studerande kunna använda ett och samma underlag för olika användningsområden.⁵² En digitaliserad textsamling som utvecklas för forskningsändamål skulle samtidigt kunna användas för undervisning genom att presentera endast valda delar eller sammanfoga det med nytt material, utan att behöva ändra i det ursprungliga innehållet. Något som skulle kunna underlätta samarbete mellan lärare och forskargrupper. Det skulle även finnas möjligheter till att utföra avancerade sökningar, kollationera, konkordanser, utvinna dokumentation om innehållet, samt tillföra nya annoteringar och kommentarer.⁵³ Sådana arkiv kan göras plattformsberoende, och modulariserade, vilket ger möjligheter att flytta hela eller delar av arkiven, förutsatt att de upphovsrättsliga frågorna är lösta.⁵⁴

⁵⁰ Broady, "Digitala arkiv och portföljer" s. 11-16.

⁵¹ Det bör betonas att märkning inte är den allenarådande metoden för att strukturera och ordna innehåll. De mest utvecklade och väl fungerande metoderna är relations- och objektdatabaser som passar bra för väl avgränsade och mer strikta modeller för innehållsmodellering än i detta fall.

⁵² Att använda innehållsdesign och digitala arkiv för flexibelt forskningsmaterial har tidigare presenterats i Donald Broady & Monica Langerth Zetterman. "Formation for the public sphere – a collective biography". *DRH2003 (Digital Resources in the Humanities 2003)*, 31 aug-3 sept, University of Gloucestershire, Cheltenham, England, 2003; Monica Langerth Zetterman. "Using Markup for Multivariate Analyses in the Prosopographical Study "Formation for the Public Sphere"." *ACH/ALLC Conference*, June 15 - June 18, Victoria University, Canada, 2005; Monica Langerth Zetterman, "Märkvärt. Noteringar om en prosopografi över kvinnor i offentlighet kring sekelskiftet 1900." *Fältanteckningar: utbildnings- och kultursociologiska texter tillägnade Donald Broady*, red. Mikael Börjesson, et al., Skeptronserien, (Uppsala: Forskningsgruppen för utbildnings- och kultursociologi ILU Uppsala universitet, 2006).

⁵³ En konkordans i detta sammanhang är en lista eller register över ord som finns i ett verk eller en text. Ofta visas de valda sökorden mitt på raderna omgivna av ett antal ord före och efter.

⁵⁴ Detta kan lösas genom användning av de s. k. *copyleft*-licenserna som t.ex. GNU General Public License (GPL) eller Creative Commons, se s. 60 ff. för diskussion om *Open Source*-rörelsen, *copyleft* och öppna standarder.

Prosopografiska samlingar som kräver tillgång till ett stort underlag och insamling av relevanta uppgifter från olika typer av källor. Dessa uppgifter måste sedan beskrivas, ordnas och bearbetas för att senare kunna analyseras.⁵⁵ Mitt val är den hierarkiska dokumentmodellen för beskrivning och organisering av det prosopografiska materialet.⁵⁶

Användare har dessutom behov av olika sorters verktyg, anpassade efter användargruppernas specifika behov och förkunskaper. Problemet är också att det finns många verktyg och metoder att välja på ibland. Vilka verktyg kan användas till vad? En lärare som vill använda digitala resurser i undervisningen har inte behov av samma typ av verktyg som forskaren, som själv producerar digitalt material. Vid vissa tillfällen behövs verktyg för att hantera och interagera med digitala resurser i undervisningen. Andra tillfällen kräver helt andra sorters verktyg; sådana som forskare eller speciellt intresserade lärare behöver för att beskriva, ordna och bearbeta digitala material.

1.1.4 Pedagogisk relevans

När det aktuella problemområdet ställs mot en pedagogisk praktik är det framför allt tre aspekter som kommer i fokus. Den första aspekten är tillgängligheten. Förknippat med tillgänglighet är frågor om innehålls urval, funktionalitet och generalitet. Den andra handlar om förutsättningar för att utforska det okända och det som inte är på förhand givet. Den tredje handlar om förutsättningar för användning av forskningsmaterial i undervisningen, det som i policydokument brukar benämnas forskningsanknytning, det ges dock inga konkreta exempel på hur man skulle kunna använda exempelvis arkivmaterial, digitala samlingar eller databaser i sådana undervisningssituationer.⁵⁷

I pedagogiska sammanhang är tillgång till information och kunskapsorganisation en nödvändig förutsättning då textbaserad information ofta utgör basen för mycket av undervisningen. Frågor man kan ställa sig är hur vi organiserar och framställer innehåll; vad som uppfattas som viktigt, hur vi kan ta till oss informationen eller vad vi kan använda innehållet till? Presenteras innehållet fragmentariskt eller i en kontext? Vad väljs ut? Vem gör urvalet?⁵⁸ Därtill har det bärande mediet betydelse för hur vi kan interagera med en text eller ett visst

⁵⁵ Exempel på en svensk kollektivbiografisk och utbildningshistorisk studie är Esbjörn Larsson, *Från adlig uppfostran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866*, Studia historica Upsaliensia, 220 (Uppsala: Acta Universitatis Upsaliensis: Uppsala University Library distributör, 2005).

⁵⁶ Den hierarkiska dokumentmodellen introduceras på s. 55.

⁵⁷ Jfr Roger Säljö & Maria Södling, *Utbildning på vetenskaplig grund: röster från fältet* (Stockholm: Högskoleverket, 2006) s. 11.

⁵⁸ Veronica Johansson hävdar, i sin kunskapsöversikt och informationspolitiska analys av elektroniska dokument, att de aktörer och de maktintressen som påverkar informationshantering och tillgång relaterar till mer generella aspekter på tillgänglighet, tillgång och kunskapssyn i allmänhet. Veronica Johansson, "Elektroniska dokument i informationspolitisk belysning. En fråga om makt, kontroll och förhandlingar", *Human IT* 7.2 (2004).

innehåll, en digitaliserad samling historiskt material ger andra förutsättningar för analyser och bearbetning än en tryckt bok.

Förutom de möjligheterna som hjälpmedel för forskning och undervisning är tillgängligheten för studerande utanför universitets- och högskoleorterna såväl som för andra användargrupper också en viktig fråga. Johan Svedjedal gör i sin utredning en koppling mellan satsningen på Litteraturbanken och utbildningens nytta med tillgången på digitaliserat kulturmaterial:

Med tanke på de olika satsningar som nu görs på "nätuniversitet" bör Litteraturbanken också kunna bli en viktig faktor för de universitet som satsar på nya former av utbildning, ofta riktade till elever som studerar på mindre hemorter med begränsade biblioteksresurser. Samma nytta kan materialet naturligtvis göra för användare inom andra delar av utbildningssystemet: gymnasister och elever på folkhögskolor kan använda Litteraturbanken som underlag för uppsatser, redovisningar etc. Bakom detta ligger förhållandet att Litteraturbanken gör ett kulturellt viktigt textmaterial nationellt tillgängligt.⁵⁹

Tillgängligheten och nyttoaspekterna diskuteras även i KB-utredningen men skillnaden att där prioriteras tillgängliggörande av unika och svårtillgängliga verk för digitalisering, sådant som äldre tryck och handskrifter.⁶⁰ Tillgänglighet förväxlas ibland med att undervisningens kvalitet automatiskt skulle förbättras. Man kan naturligtvis hävda att om lärare och studenter får tillgång till mer och bättre digitalt material så skapas bättre förutsättningar för att kvalitativt förbättra undervisning. För att ytterligare komplicera frågan så är ju materialets kvalitet, användarnas kompetens och verktygens användbarhet andra viktiga förutsättningar. Ett sätt att undvika förvirring är att skilja mellan kvalitet på innehåll och på undervisning. Gunnar Grepperud menar att man i högre utsträckning bör tala om vilken sorts kvalitet som digitalisering och IT kan tillföra undervisningen, istället för att tala om kvalitetshöjning på någon sorts allmän nivå.⁶¹

En annan faktor är själva ämnet. Naturvetare som exempelvis studenter och forskare i bioinformatik, skulle inte klara av att studera och analysera sitt material utan hjälp av datorer. I samhällsvetenskapliga och humanistiska ämnen är det inte alltid så enkelt att efterlikna ett laboratorium eller att utföra experimentet. Digitala samlingar med tillgång till källmaterial och lämpliga verktyg kan ge studerande tillfälle att bearbeta och studera primärkällor och forskningsmaterial, söka information, tolka resultaten och formulera frågeställningar.

När det gäller pedagogisk forskning eller styrdokument om IT-användning hamnar sällan texters funktion i forsknings- och undervisningssammanhang

⁵⁹ Svedjedal, *En svensk Litteraturbank? Utredning för Svenska Akademien, avlämnad i januari 2003* s. 5.

⁶⁰ KB-utredningen, *KB - ett nav i kunskapssambället: Kungl. biblioteket - Sveriges nationalbibliotek: verksamhet och visioner: betänkande* s. 213 f.

⁶¹ Gunnar Grepperud, *Fleksibel utdanning på universitets- og høyskolenivå: forventninger, praksis og utfordringer. 3: Fleksibel utdanning som utfordring* (Tromsø: UVETT Universitetets videre- og etteruddannelse Universitetet i Tromsø, 2005) ssk. s. 66.

eller en infrastruktur som tillåter användare att arbeta med digitalt material överst på dagordningen.

En infrastruktur för forskning och undervisning med en bredare funktionalitet skulle kunna tillföra mer än att enbart presentera och tillgängliggöra olika slags digitalt material. Tekniken är nu så väl utvecklad att det är dags att börja fundera över de innehållsliga frågorna. Hur vi kan underlätta studenters tillgång till, analyser av och samarbete kring ett specifikt innehåll? I Litteraturbankens digitala samlingar finns en uttalad målsättning att förutom att tillgängliggöra själva innehållet förse det med användbara verktyg. Förutom presentation av digitala versioner av litterära verk, tillhandhålls som tidigare nämnts även verktyg som tillåter användare att arbeta med det digitala materialet i webbläsaren.⁶² Det finns även möjlighet att skapa egna antologier av Litteraturbankens e-texter.⁶³ Genom att markera en eller flera ord i texten skapas automatiskt en kommentar med det markerade ordet som rubrik. Där kan sedan användaren skriva in sina egna kommentarer. Dessa går sedan att klicka på, visa som lista och redigera. Användarna ges här möjlighet att verkligen arbeta med materialet.⁶⁴ Användares anteckningar sparas på Litteraturbankens server och finns kvar vid nästa inloggning.⁶⁵

En annan aspekt på tillgänglighet är de digitala samlingar som publicerar verk som inte är svårtillgängliga på samma sätt som handskrifter och äldre tryck. Här avses opublicerade och oredigerade texter, det vill säga sådant som snarare kan ses som arkivmaterial och som tidigare inte fallit inom ramen för de traditionella litterära verk som ingår i kurslitteraturen, då texterna haft otillräcklig legitimitet på grund av exempelvis författarens kön.⁶⁶ Andra exempel på svårtillgängligt material är sådant som varit hemligstämpat i arkiven. Roger Johanssons metodbok för studier av digitaliserade samlingar om händelserna i Ådalen 1931 är ett bra exempel på hur man kan nyttja digitala medier för att göra källmaterial tillgängligt och användbart för olika målgrupper.⁶⁷ Den digitala dokumentsamling som skapats har sammanställts tematiskt utifrån material som

⁶² I Johan Svedjedals sammanfattning i utredning om Litteraturbanken är detta en av huvudpunkterna: [...] att samla in, arkivera och publicera material samt göra det möjligt för användare att arbeta med det. Svedjedal, *En svensk Litteraturbank? Utredning för Svenska Akademien, avlämnad i januari 2003* s. 42.

⁶³ Se <<http://litteraturbanken.se/>>. Annoteringsfunktionen fungerar än så länge på ett urval av texter – de texter som finns tillgängliga som e-text men inte på de texter som publiceras i PDF-format. Funktionen kräver inloggning. Det är både enkelt och gratis att bli registrerad användare. Funktionen för antologier är tillfälligt avstängd fr.o.m. februari 2008 för uppdatering och förbättringar (e-post från Litteraturbanken 2008-01-31).

⁶⁴ Jfr. Borin, ”Vi som går köksvägen. Språkteknologer och korpuslingvister i Litteraturbanken”.

⁶⁵ Litteraturbankens annoteringsfunktion illustreras i bilaga 13, se <CD:/bilagor/bilaga13.pdf>.

⁶⁶ Även om flera kvinnliga författare blev uppmärksammade och lästa under 1800-talet kunde de inte vara alltför frispråkiga. Det finns förstås en anledning till att vissa kvinnliga författare (t.ex. Victoria Benedictsson) publicerade sig under pseudonym. En diskussion om maktförhållanden, tolkningsföreträde och idealiserade texter som måttstock för legitimitet av tidigare opublicerade litterära texter av kvinnor förs i: Julia Flanders, ”Gender and the Electronic text” *Electronic text: investigations in method and theory*, red. Kathryn Sutherland, (Oxford: Clarendon, 1997) ssk. s. 135 f.

⁶⁷ Johansson, *Ådalen 1931 - minuterna som blev historia: metodbok till en digital dokumentsamling*.

samlats in från olika arkiv och källor. Samlingen innehåller även tidigare hemligstämplade dokument. Boken är en introduktion till historisk forskning och källkritik och innehåller övningar som vänder sig till olika målgrupper, från grundskolelever till lärare inom högskolan men även personer som sysslar med släkt- och regionforskning.

Andra exempel på hur digitala samlingar för forskning även kan nyttjas i undervisning är att använda dem som hjälpmedel för att förnya undervisningen. Julia Flanders, vetenskaplig ledare för *The Brown University Women Writers Project*,⁶⁸ anser att en av de främsta pedagogiska utmaningarna med vetenskapliga digitala samlingar är möjligheterna att införa ett annat arbetssätt i undervisningen. Istället för ett komprimerat urval av texter, till exempel i en antologi, kan studenter ges möjlighet att konfronteras med excerpter såväl som hela verk, enskilda objekt såväl som kontexten och därtill bekanta sig med ett okänt innehåll. En utgångspunkt för detta, menar Flanders, är att studenterna ges möjlighet gå i närkamp med originaltexterna:

[...] huge, unmanageable texts; difficult texts; long, beautiful texts; absurd texts; poignantly misguided texts; fascinatingly dull texts; texts whose footnotes are the most interesting thing of all. The collection thus encourages students to think and act like their teachers, at least qualitatively – by asking the kinds of questions that only the full textual world can answer.⁶⁹

I *Women Writers Project* finns en uttalad målsättning att få lärandesituationer att likna forskarens arbetssätt och att studenterna, med viss hjälp, skall ta sig an ett *utforskat material*. Denna något idealistiska metod undgår dock inte ett visst mått av självkritik och pragmatisk realism. Att skapa ett digitalt arkiv med vetenskapliga utgåvor, versioner och textkritisk apparat kräver väldigt mycket arbete och resurser. Förutom det finns alltid risken att arkivet blir omodernt efter några år, då nya tekniker och metoder introduceras.

Detta innebär att man i det här projektet istället valt en medelväg mellan de stora universitetens generella digitala samlingar, med många genrer och tidsperioder och en uppsjö av kringmaterial, och de kursspecifika webbaserade undervisningsmaterialen. *Women Writers Project* tillhandhåller därför bredvidmaterial, som relateras till de digitala texterna i samlingen, i form av kortare essäer om författarna och vissa ämnen.⁷⁰ Detta menar man, kan hjälpa studenterna att få en uppfattning om textens sociala sammanhang, historia och tillkomst.

⁶⁸ The Brown University Women Writers Project (WWP), digitala samlingar och fulltextarkiv med texter av kvinnliga, ofta tidigare opublicerade, författare från perioden 1400-1850. Förutom uppbyggnad av digitala samlingar har man sedan tidigt 1990-tal utvecklat och utforskat verktyg, textteoretiska frågor, handledningar och "best practices" för textmärkning samt kursplaneutveckling. De digitala samlingarna är dock inte fritt tillgängliga utan kräver licens för åtkomst. Se <www.wwp.brown.edu/>.

⁶⁹ Julia Flanders, "Learning, Reading, and the Problem of Scale: Using Women Writers Online", *Pedagogy: Critical Approaches to Teaching Literature, Language, Composition, and Culture* 2.1 (2002) s. 50.

⁷⁰ Se <www.wwp.brown.edu/about/index.html>.

En annan aspekt är att skapa förutsättningar för en forskningsanknuten högskoleundervisning genom användning av konkret forskningsmaterial i undervisningen. Den digitala dokumentsamlingen om Ådalen 1931 är ett exempel på att skapa sådana förutsättningar. Istället för att studera andra forskares tolkningar ges genom sådana här samlingar tillfälle att ta del av själva källmaterialet:

Syftet med att sammanställa ett arkiv på en cd-skiva, att så att säga flytta in arkivet i biblioteket, är att ge läsaren en möjlighet att själva [sic!] möta och tolka källorna.⁷¹

Jag har i detta avsnitt problematiserat förutsättningar för och behov av tillgång till och användning av digitalt material i forskning och undervisning. Sammanfattningsvis handlar mycket av det som ovan beskrivits, om att bevara och tillgängliggöra vårt kulturarv. Det vill säga sådant som påverkas av strategier hos kulturinstitutioner, som till exempel Kungl. Biblioteket. Hur kan långsiktiga förutsättningar ges för utbyte av innehåll mellan olika plattformar samt för tillgång till digitala material i framtiden? Problemet ligger även i att ta fram verktyg för skilda målgrupper med olika behov och förkunskaper så att de kan nyttja de resurser som faktiskt finns att tillgå. Inte minst för att mer medvetna användare kan ställa högre krav på hur kulturarvet och annat värdefullt material görs tillgängligt för forskare, lärare och studerande och för andra användargrupper.

1.2 Syfte och frågor

Ett övergripande syfte är att bidra till metodutveckling genom att föreslå tillvägagångssätt för innehållsdesign och utformning av digitala material för forskning och undervisning. Förutsättningar för de grundläggande momenten i innehållsdesign, att beskriva, ordna och bearbeta digitala material, behandlas ofta alltför lättvindigt i arbetsprocessen och riskerar därmed resultera i att materialet blir låst till särskilda presentationsformer. Ett delsyfte är därför att avgränsa och inventera dessa grundläggande moment: beskriva, ordna och bearbeta. Ett annat delsyfte är att genom ett explorativt tillvägagångssätt undersöka de inventerade metodernas och verktygens begränsningar och möjligheter i tillämpningar för utbildningshistoriska studier. Med utgångspunkt i dessa syften formuleras följande frågor:

- Hur kan de moment, rubricerade som innehållsdesign, avskiljas? Vad kännetecknar dessa moment?
- Vilka slags principer, metoder och verktyg är ändamålsenliga för innehållsdesign inom samhällsvetenskap och humaniora?

⁷¹ Johansson, *Ådalen 1931 - minuterna som blev historia: metodbok till en digital dokumentsamling* s. 13.

- På vilka sätt kan dessa principer, metoder och verktyg användas i tillämpningar för utbildningshistoriska studier?
- Vilka möjligheter och begränsningar kan urskiljas för dessa metoder, verktyg och tillämpningar?

Dessa frågor är vägledande för avhandlingens disposition. De två första frågorna används för att inventera och analysera principer, metoder och verktyg för innehållsdesign. De sista två frågorna vägleder framför allt avhandlingens tillämpningar men även till viss del de delar som rör inventering och analys av principer, metoder och verktyg för utformning av digitala material inom samhällsvetenskap och humaniora.

De tillämpningar som utforskas i studien baseras på historiskt material som sammanställts och utformats för prosopografiska studier för projektet *Formering för offentlighet. En kollektivbiografi över stockholmskvinnor 1880-1920*. I detta projekt har borgerliga kvinnors vägar till offentligheten studerats. Denna prosopografiska databas innehåller uppgifter om 1250 offentligt verksamma kvinnor som samlats in genom biografiska uppslagsverk⁷² och för 101 av dessa kvinnor finns nyskrivna kapitalbeskrivningar (biografiska beskrivningar).⁷³ Den andra tillämpningen är prosopografiskt material som samlats in och utformats för utbildningshistoriska studier. Det prosopografiska materialet består av uppgifter om samtliga (1980 kvinnor) inskrivna lärarinnestudenter vid Kungl. Högre Lärarinneseminariet under hela verksamhetsperioden 1861-1943.⁷⁴ Dessa två prosopografiska material sträcker sig över en relativt lång historisk period, en tid när borgerliga kvinnor började få möjligheter till högre utbildning och att träda ut i de offentliga sammanhangen.

1.3 Tillvägagångssätt

I avhandlingens syfte presenteras ambitionen att inte bara beskriva och utforska ett antal principer, metoder och verktyg utan även att bidra med kunskap om hur vi kan utforma och använda digitalt material för forskning och undervisning. I detta ingår också ett bakomliggande instrumentellt syfte med tanke på att tillämpningarna, genom utforskandet, förhoppningsvis kan sammanföra kunskap från olika vetenskapsområden till andra forsknings- och undervisningspraktiker.

⁷² Uppgifterna är avskrifter ur framför allt Nils Styrbjörn Lundström, *Svenska kvinnor i offentlig verksamhet: porträtt och biografier* (Uppsala: 1924). Men även ur Valborg Hedberg & Louise Arosenius, *Svenska kvinnor från skilda verksamhetsområden: biografisk uppslagsbok* (Stockholm: Bonnier, 1914).

⁷³ Detta material benämns FFO-LUNDSTROM (1251 kvinnor) och den delmängd som utgörs av de 101 kvinnor som det samlats in mer detaljerade uppgifter i kapitalbeskrivningarna om benämns FFO-KAPITAL.

⁷⁴ Detta material benämns FFO-HLS.

För att undersöka hur man kan gå tillväga för att digitalisera och sammanställa uppgifter hämtade ur olika slags källor samt ordna, bearbeta och använda dess uppgifter använder jag mig av ett explorativt arbetssätt. Arbetssättet är explorativt i den bemärkelsen att i modelleringen av materialet kombineras metoder för märkning av innehåll, statistiska och andra metoder för studier av tillämpningar för utbildningshistoriska studier. Metoden som använts för att sammanställa samma sorts uppgifter för en avgränsad grupp av individer kallas för *prosopografi*, ibland även kallad kollektivbiografi. Det prosopografiska materialet i den här studien består av digitala foton av arkivposter som transkriberats, författade kapitalbeskrivningar⁷⁵ samt avskrifter ur biografiska uppslagsböcker.

Materialet kan inte förknippas med ett specifikt kunskapsområde. Delar av det prosopografiska materialet är inte i första hand insamlat med utgångspunkt i strikta källkritiska principer. Detta skulle kunna medföra att delar av detta material har en relativ begränsad användning för renodlade historiska studier.

Det första arbetsmomentet innebar att avgränsa och inventera de principer metoder och verktyg som är ändamålsenliga för innehållsdesign inom samhällsvetenskap och humaniora. Denna inventering utfördes genom att studera olika riktlinjer och föreskrifter för märkning av innehåll. Resultaten av inventeringen presenteras i kapitel 3 som handlar om märkspråkens utveckling och hierarkiska dokumentmodeller som kontrasteras mot alternativa metoder och verktyg. Inventeringens resultat fortsätter i kapitel 4 där *Text Encoding Initiative* (TEI), presenteras.⁷⁶

I den tillämpade delen av avhandlingen (kapitel 1, 6 och 7) har jag använt mig av flera olika tillvägagångssätt. Den största utmaningen har varit att bringa någon slags ordning i detta disparata material för att kunna utforska det med hjälp av metoder som härrör från olika kunskapsområden och ämnen.

Jag har tidigare nämnt att studien är explorativ och syftet med utforskandet är trefaldigt: För det första att utforska lämpliga metoder och verktyg för innehållsdesign. Här ingår även att testa på vilka sätt verktyg som mestadels används inom samhällsvetenskaperna kan kombineras med humanistiska metoder. För det andra är syftet att testa den hierarkiska dokumentmodellen på den här typen av material – som var en blandning av nyskrivna biografiska texter, avskrifter ur självbiografiska uppslagsböcker och arkivmaterial i form av inskrivnings- och betygsliggare och annat. För det tredje finns det explorativa i de

⁷⁵ Benämningen "kapitalbeskrivningar" i betydelsen levnadsteckningar avser den mall med vars hjälp uppgifterna har samlats in. Begreppet "kapital" är Bourdieus kapitalbegrepp och är ett av nyckelbegreppen i Bourdieus sociologiska verktygslåda. Kapital avser här relationerna mellan individers (eller grupper) egenskaper och tillgångar relativt det sammanhang där dessa egenskaper och tillgångar har något värde. Donald Broady, *Sociologi och epistemologi: om Pierre Bourdieus författarskap och den historiska epistemologin*, Skeptronserien, 2., korr. uppl. ed. (Stockholm: HLS Förlag, 1991) s. 169-181.

⁷⁶ *Text Encoding Initiative* (TEI) är en internationell organisation som utvecklar riktlinjer för digitala utgåvor av litterära och lingvistiska texter för forskning och undervisning samt för utbyte, lagring och bevarande av digitala texter. Se vidare <www.tei-c.org/>. TEI:s riktlinjer är det verktyg som huvudsakligen använts för beskrivning av det material som används i studien.

konkreta uppgifterna som ingår de arbetsmoment och uppgifter som först skulle identifieras och sedan genomföras för utformning och bearbetning av materialet.⁷⁷

En översiktlig beskrivning av arbetsgången enligt följande:

1. *Sammanställning av kapitalbeskrivningar* genom att ta emot och lagra de dokument som skickades till mig, ett register upprättades där noteringar om författare, datum, version, namn på den kvinna som biografierades samt antal ord i dokumentet noterades.
2. *Migrering av kapitalbeskrivningarnas format* genom att spara dokumenten i rent textformat, importera dem till en editor och förse dem med grundläggande märkning och sammanfoga dem i en förlaga, en masterfil, som varit utgångspunkt för den grundligare semantiska märkningen.
3. *Iordningställande av uppgifter som registrerats* från biografiska uppslagsböcker. Dessa uppgifter var lagrade i kalkylblad samt försågs med märkning innan migrering till statistikprogrammen.
4. *Konstruktion av datamodell* med utgångspunkt i kapitalbeskrivningarna och med sikte på bearbetningarna, detta moment innehåller många ställningstaganden och överväganden om klassificering och innehållsobjektens inbördes relationer.
5. *Märkning av kapitalbeskrivningar* genom att använda TEI där märkningen utfördes med olika XML-editorer.⁷⁸ Detta har varit det mest tidsödande momentet, men det är en nödvändig utgångspunkt för vidare bearbetningar. Momentet innehåller många ställningstaganden och överväganden om hur man kan anpassa de uppgifter som skall registreras i prosopografin till den hierarkiska dokumentmodellen. Här har jag gjort många revideringar framför allt i början. När materialet och verktygen blir mer bekanta går det snabbare att infoga märkningen.
6. *Halv-automatisering* för märkning (beskrivning) av återkommande egenskaper av samma typ, till exempel personnamn. Ett eget verktyg utvecklades för detta ändamål då det inte fanns några lämpliga verktyg tillgängliga.⁷⁹
7. *Transformering* av specifika uppgifter från masterfilen, i XML-format, till andra filformat och program. Detta moment har varit relativt tidsödande, då jag har konstruerat egna skript för transformering av detta material⁸⁰.

⁷⁷ I detta sammanhang kan nämnas att alla webblänkar är kontrollerade 2008-01-25 och datum för senaste besök anges alltså inte för var och en av webblänkarna.

⁷⁸ Detta har varierat under åren, i början användes Emacs som är en samling program för redigering av text och programkod, konstruerades ursprungligen under 1970-talet av Richard Stallman för GNU-projektet, se 60 ff. för en presentation av GNU. De senaste åren har programmet Oxygen använts för redigering, märkning och programmering.

⁷⁹ En av utgångspunkt för innehållsdesign är att använda generella metoder och verktyg som redan finns tillgängliga, ibland måste dock denna princip frångås.

⁸⁰ Det skriptspråk som använts är XSLT som är designat för transformering av innehåll i XML-filer, för tillvägagångssätt se s. 193 f.

8. *Klassificering* av uppgifter i materialet; ett nödvändigt moment för senare bearbetning inför statistiska analyser, geometrisk dataanalys, nätverksanalyser och lagring av vissa data i databasen.
9. *Bearbetning* av innehållet i masterfilen, vilket innebär att omvandla den till olika filformat så att den kan importeras till aktuellt program. Principerna har framför allt inneburit att antingen har specifika uppgifter extraherats ur masterfilen, som exempelvis personnamn eller så har uppgifterna omordnas, för exempelvis nätverksanalyser. Bearbetning av det från XML-formatet migrerade materialet har sedan skett i olika kalkyl- och statistikprogram, i program för nätverksanalyser samt i relationsdatabasen.
10. *Revideringar* av märkning, klassificering och variabler är ett återkommande moment som förekommer i de flesta arbetsmomenten och som tar mycket tid i anspråk.
11. *Fotografering* av arkivmaterialet har skett med hjälp av en ”vanlig” digitalkamera med relativt god upplösning.
12. *Transkribering* av det fotograferade arkivmaterialet har utförts med ett diplomatariskt tillvägagångssätt. Detta innebär att texten avskrivits såsom den är förtecknad i originalhandlingarna, med stavfel eller andra fel. Till att börja med upprättades ett komplett register med utgångspunkt i det samlade materialet där grundläggande uppgifter om samtliga kvinnor förtecknades. Detta register används för att koppla ihop det material som tillkommer.
13. *Relationsdatabasen* som innebär arbete med flera moment, bland annat konstruktion av datamodell (se punkt 4), utformning av utmatning och sökparametrar, konstruktion av webbaserat sökgränssnitt.
14. *Webbarkiv* som innehåller dokument, masterfiler, register, dokumentation, analyser och en relationsdatabas.
15. *Testning av programvaror* för olika slags analyser av materialet som nätverksanalyser, dynamiska grafer, interaktiva kartor etcetera. Hit räknas även samtliga program som använts för statistiska bearbetningar.
16. *Dokumentation*, sist men inte minst är det viktigt att dokumentera arbetsprocessens överväganden, ställningstaganden, tillvägagångssätt och problem. Dokumentationen ligger även till grund för eventuella användarhandledningar och tillämpningsexempel.

De moment som ovan översiktligt presenterats kommer samtliga att illustreras, analyseras och diskuteras i den tillämpade delen III i avhandlingen.

1.4 Avgränsningar

Ramen för avhandlingen är dels metoder och verktyg som används för att beskriva, ordna och bearbeta digitalt innehåll, dels det prosopografiska materialet som tillämpningarna baseras på. Det prosopografiska materialet som sådant

utgör således en avgränsande faktor, de metoder och verktyg jag använder en annan och slutligen den prosopografiska metoden en tredje avgränsande faktor. Därav följer att avhandlingen är mer praktikorienterad och resultatfokuserad än många andra pedagogiska studier. Resultaten, det vill säga verktygslådan och annat material relaterat till studien publiceras på den i avhandlingen bifogade CD:n.⁸¹ Förutom detta finns databaser och annat material tillgängligt i webbar-kivet, fritt tillgängligt för andra att modifiera och använda.

Avsikten med min studie är alltså att utforska en uppsättning verktyg och metoder för att iordningställa utbildningshistoriskt material för forskning och undervisning. Däremot har jag inte haft för avsikt att studera till exempel undervisning på webben eller lärande i webbaserade utbildningssituationer.

Jag har heller inte studerat kunskapens sociala dimension i den forskningspraktik jag deltagit i. Jag är förstas medveten om att kunskapsprocessen inte är individuell eller historielös, utan ett resultat av ett kollektivt arbete. Det kollektiva arbete och de beslut som påverkat det prosopografiska materialets beskrivning och utformning presenteras och diskuteras i anslutning till aktuella avsnitt. Det enstaka experimentet har en begränsad betydelse relativt den sammansatta erfarenheten då ett experiment inte kan förstås utan hjälp av tankekollektivet.⁸²

Studien har en interdisciplinär utgångspunkt, något som känns nödvändigt av tämligen pragmatiska skäl. Den praktik som är en förutsättning för avhandlingens metodinriktade empiriska del (del II) är framför allt ett resultat av forskning och utveckling inom biblioteks- och informationsvetenskap, datavetenskap, filologi, litteraturvetenskap, språkvetenskap och historia. Medan den praktik som är en förutsättning för den tillämpade empiriska delen (del III) av avhandlingen är sociologisk och historisk.

Jag avgränsar undersökningen till att presentera och analysera sådant som direkt, och indirekt, relaterar till utformning och märkning av digitalt material i allmänhet och prosopografiskt material i synnerhet.

Sammanfattningsvis är det så att den metodinriktade empiriska delen av avhandlingen (del II) behandlar principer, metoder och verktyg för märkning av humanistiskt och samhällsvetenskapligt innehåll. Det material jag studerat för att avgränsa och inventera dessa metoder och verktyg består här av artiklar, riktlinjer, diskussionslistor och medlemslistor. I del III tillämpar jag de metoder och verktyg som inventerats och presenterats i del II. Slutligen sammanställs detta arbete i form av förslag till en verktygslåda för innehållsdesign av utbildningshistoriskt material.

⁸¹ Se innehållsförteckning för digitala bilagor (CD:n) s. 260.

⁸² Ludwik Fleck, *Uppkomsten och utvecklingen av ett vetenskapligt faktum: inledning till läran om tankestil och tankekollektiv* (Eslöv: B. Östlings bokförl. Symposion, 1997) s. 102 f.

1.5 Något om terminologi

I avhandlingen förekommer en del termer från olika vetenskapsområden. Vissa av termerna har en mer specifik teknisk innebörd än andra och för dessa hänvisas läsaren till anförda noter och referenser. Begreppet innehållsdesign tas inte upp i detta avsnitt då det introducerades i kapitel 2, s. 51.

1.5.1.1 Innehåll och design

En central aspekt av innehållsdesign är förstås själva innehållet, eller mer specifikt skillnaden mellan det intellektuella och semantiska innehållet och det som är form, struktur och ordning. Med innehåll avses här den information eller det stoff, tankemängd eller ämne som framförs och presenteras i skriftlig eller muntlig form – i motsats till innehållets yttre utformning och presentation. Termen ”innehåll” avser även de uppfattningar och betydelser som kan inrymmas i ett begrepp. Detta är särskilt viktigt då märkningen erbjuder möjligheter att explicit tillskriva ett visst begrepp en eller flera specifika betydelser.⁸³

Ett begrepp som är närliggande men inte riktigt överensstämmer med min ansats är den litteraturvetenskapliga termen innehållsfält (*content space*) för beskrivning och avgränsning av de innehållsliga delar som binder samman och bygger upp ett verk.⁸⁴ Termen innehållsfält används av vissa litteratursociologer för att beskriva alla slags innehåll, inte bara för skriven text, utan även för bilder och ljud.⁸⁵ I den här studien används dock en annan utgångspunkt, nämligen den att ett visst innehåll inte enbart avgränsas med utgångspunkt i semantiska innebörder utan även av innehållets ordning och struktur. Därför används termen *innehållsobjekt* för att avgränsa de logiska, strukturella eller semantiska delarna ett dokument består av.

I avhandlingen används oftast innehåll synonymt med text särskilt i de sammanhang då jag talar om avhandlingens empiriska material. En del av materialet består av registerposter och andra av mer prosaliknande text. En stor del av materialet består dock av digitala foton, det vill säga avbildningar av arkivposter. Därför används i första hand termen innehåll, förutom där det är särskilt befogat eller passande att använda text. Nämnas bör att merparten av de verktyg och metoder jag använder är konstruerade med utgångspunkt i att beskriva och strukturera textbaserat innehåll. En ganska oproblematiskt utgångspunkt kan man tycka. Peter Shillingsburg påpekar dock att en text har olika innebörder beroende på vem du frågar. För forskaren är texten det fysiska objektet, den

⁸³ Metadata är inte innehåll utan beskrivningar av aspekter om ett visst innehåll.

⁸⁴ Anna Gunder, ”Berättelsens spel - berättarteknik och ergodicitet i Michael Joyces afternoon, a story”, *Human IT*.3 (1999). Termen introducerades, enligt Gunder, av Johan Svedjedal.

⁸⁵ Ibid; Anna Gunder, ”Forming the Text, Performing the Work - Aspects of Media, Navigation, and Linking”, *Human IT*.2-3 (2001); Johan Svedjedal, *The literary web: literature and publishing in the age of digital production: a study in the sociology of literature*, Skrifter utgivna av Avdelningen för litteratursociologi vid Litteraturvetenskapliga institutionen i Uppsala, 42 (Stockholm: Kungl. bibl., 2000) s. 57. Svedjedal föredrar termen *content space* snarare än *writing space* (som introducerades av David Bolter mfl) just för att dessa fält kan innehålla mer än text.

andra anser att text är serier av tecken eller symboler medan den tredje menar att text är abstrakta begrepp.⁸⁶

Vad gäller begreppet design så skiljer sig designprocessen i innehållsdesign från annan design till exempel grafisk formgivning. En grafisk formgivare arbetar med själva presentationsaspekten av innehållet. Till exempel hur texter och bilder skall kombineras och placeras eller hur rubrikerna skall formges med god estetik eller läsbarhet. En innehållsdesigner sysslar med de bakomliggande förutsättningarna där själva presentationsnivån är sekundär. För en innehållsdesigner är istället innehållets beskrivning, ordning och bearbetning det primära.

1.5.1.2 Material och resurs

Vid vissa tillfällen används termen resurs medan jag vid andra tillfällen använder termen material. Jag tänker mig en skillnad mellan dem som består i att en resurs är något, en viss form av innehåll, som är tillverkat, färdigställt eller åtminstone sammanställt och som är tillgängligt för användning och bearbetning. Man skulle också kunna säga att en digital resurs är en sorts produkt och därmed resultatet av design. Detta kan jämföras med ett material som är utgångspunkten för denna designprocess, det vill säga det innehåll som används som utgångspunkt för att konstruera något nytt eller något annat. Ett material behöver heller inte vara en representation av en originalutgåva eller ett helt nyproducerat innehåll, utan kan bestå av en sammanställning eller omorganisering av andra resurser. Givetvis är gränserna ofta flytande mellan vad som utgör en resurs och ett material, kriteriet för material är dock att det används som utgångspunkt för att skapa något annat.⁸⁷ Resurs är det resultat som på ett eller annat sätt har skapats för ett visst användningsområde eller något syfte. Bruket av termerna material och resurs skulle möjligen kunna bytas ut mot *startdokument* och *måldokument* där *transponering* betecknar själva överförings- och omvandlingsprocessen mellan vad som varit och det som blir nytt.⁸⁸

⁸⁶ Peter L. Shillingsburg, *From Gutenberg to google: electronic representations of literary texts* (Cambridge: Cambridge Univ. Press, 2006) s. 12.

⁸⁷ För vetenskapliga utgivning, speciellt angloamerikansk, skulle man istället för resurs och material möjligen kunna tala om *bastext* där enskilda texter är det centrala. Men i en dokument- eller innehållsorienterad praktik är det inte relevant att tala om *bastext* då syftet inte är att representera alternativa versioner utan det primära är att representera ett visst dokument. Bastext-teorin diskuteras och analyseras ingående i: Mats Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion*, Skrifter från Valfrid, 34 (Borås: Valfrid, 2006) ssk. s. 176 f.

⁸⁸ "[ö]verföra delar av ett eller flera startdokument till ett eller flera måldokument (av samma eller skilda medieslag). Processen benämns här *transponering*, en term hämtad från musiken, där den betecknar den överföring och omgestaltung av notation som görs mellan olika tonarter." Ibid. s. 103.

1.5.1.3 Digital och elektronisk

Digital är när värden anges med hjälp av sifferuttryck – oftast i form av det binära talsystemet⁸⁹ – i motsats till analog där värden anges som punkter på en skala.⁹⁰ Betydelsen av elektronisk avser sådant som är eller nyttjar elektronik. Den största skillnaden är alltså inte mellan elektronisk och digital, utan mellan digital och analog. Digital information har stora möjligheter att ge en exakthet i kopian och störs inte av brus jämfört med analog information som alltid påverkas av brus vid överföring.

Det handlar även om vilken lagringsmetod som används. Datorer behöver både elektronik och digital teknik för att fungera, en DVD behöver elektronik för att användas, däremot inte för att behålla den lagrade informationen. Information lagrad med hjälp av analog teknik behöver också ofta elektronik, kassettband lagrar ljud och data analogt och magnetiskt men behöver elektroniska verktyg för att användas.

Susan Hockey, framhåller att skillnaden mellan elektroniska texter och digitala avbildningar av texter framför allt ligger i funktionaliteten. Elektroniska texter är avskrifter och har alltså överförts med hjälp av transkribering. Elektroniska texter kan därför manipuleras på många olika sätt exempelvis ges möjligheter till olika typer av sökningar, sammanställning av (ord)listor, ta bort och lägga till eller ändra ordföljd. Digitala bilder, av exempelvis text, är just avbildningar och tillåter därför inte manipulering på samma sätt som elektroniska texter.⁹¹ Poängen med att särskilja elektroniska texter från digitala texter skulle då vara att de elektroniska texterna väsentligen förbättrar möjligheterna att manipulera innehållet jämfört med digitala avbildningar.⁹² Jag håller med om att överföringsmetoden, mellan ett medium och ett annat, påverkar graden av manipulerbarhet. Men mot bakgrund i diskussionen ovan om skillnader mellan digital och analog anser jag att man inte bör skilja elektroniska texter och digitala bilder på det sätt som Hockey gör då dessa tekniker är varandras förutsättning.⁹³ Metoderna för överföring är inte beroende av antingen elektronik eller digitala tekniker utan för det mesta av båda delarna.

En bättre förklaringsmodell, som inte skiljer mellan elektronisk och digital, är de två överföringsmetoder som Mats Dahlström beskriver. En *textorienterad*

⁸⁹ Det binära talsystemet är ett positionssystem med basen två i motsats till vårt vanliga decimalsystem som har tio som bas. I det binära systemet används vanligen siffrorna 0 eller 1 för att beteckna positioner och de decimala talen betecknas baserat på potenser av 2. Till exempel talet 45 betecknas 101101 som då betecknar: $32+0+8+4+1$. Siffrorna 0 och 1 kallas ofta för bitar, från engelskans *bit* (binary digit).

⁹⁰ I matematisk betydelse är analoga värden kontinuerliga till skillnad mot digitala värden som är diskreta.

⁹¹ Detta ger förstås en större säkerhet i fråga om närhet till originalets visuella egenskaper.

⁹² Se Susan M. Hockey, *Electronic texts in the humanities: principles and practice* (Oxford: Oxford University Press, 2000) s. 1 ff.

⁹³ Detta är giltigt utom för själva lagringen, där behövs ingen elektronik. På en hårddisk lagras data på en skiva med magnetiskt material och bevaras utan elektronik, för att skriva in och läsa det data som lagrats på hårddisken behövs dock elektronik.

som ytrar sig i transkription och en *bildorienterad* som ytrar sig i faksimilering med hjälp av bildfångst.⁹⁴ Transkription handlar alltså inte bara om avskrifter av text från ett dokument till ett annat utan även om att textbaserat innehåll kan överföras med hjälp av optisk teckenigenkänning.⁹⁵

Digital teknik är grunden för datorers funktionalitet och används oavsett om man producerar PDF-filer eller om man transkriberar en text i ett ordbehandlingsprogram. Elektroniska texter förutsätter därmed digital teknik och skillnaden i graden av manipulerbarhet beror på i vilket filformat innehållet lagras samt med vilken metod innehållet överförs från ett medium till ett annat. Lagras innehållet med hjälp av PDF-formatet finns mindre möjligheter till senare manipulation, dock ger formatet möjligheter till versionshantering och långtidsbevarande.⁹⁶ Om innehållet däremot överförs och lagras i XML-format ges många fler möjligheter till att senare välja format för lagring och presentation, till exempel PDF, HTML eller något annat. Sammanfattningsvis hävdar jag att den metod som används för överföring och textfångst påverkar hur innehållet, texten eller dokumenten senare kan lagras, användas och återanvändas.

Bruket av termen elektronisk text, vill jag hävda, härstammar från den tidiga datoranvändningen då det blev praxis att säga e-post, e-lärande, e-texter etcetera. Termen digital verkar oftare användas i sammanhang som specifikt rör hur data representeras, lagras och presenteras.

Med utgångspunkt i ovanstående diskussion använder jag i de allra flesta fall *digital* när jag talar om texter, bilder eller annat material som produceras, lagras och manipuleras med hjälp av digital teknik oavsett på vilket sätt överföringen mellan medier har skett. Ibland används elektronisk text som en översättning av den vanligt förekommande termen *electronic text*.

1.5.1.4 Metod och verktyg

I avhandlingens titel används termerna metod och verktyg. Ett verktyg beskrivs ofta i konkreta ordalag som ett redskap som kan användas ”för att bearbeta material” och anses ofta fungera som vår förlängda arm.⁹⁷ En metod har däremot en mer övergripande betydelse och kan beskrivas som ett ”planmässigt tillvägagångssätt för att uppnå ett resultat”.⁹⁸ Hur ett verktyg utformas och brukas kan sägas vara beroende av både vilken metod som används och av materialets eller objektens beskaffenhet. Ett verktyg är därmed mer än bara ett verktyg, då det står i relation till användning, sammanhang, intentioner och bakgrund –

⁹⁴ Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 202 ff.

⁹⁵ Optical Character Recognition (OCR). På samma sätt kan talat språk överföras genom tekniker för taligenkänning. Se Ibid. s. 208.

⁹⁶ En metod för långtidsbevarande och unika ID:nr av digitala publikationer har implementerats bland annat i DiVA systemet, utvecklat vid Uppsala universitet. Se t.ex. Eva Müller, et al., ”The DiVA Project - Development of an Electronic Publishing System”, *D-Lib Magazine* 9.11 (2003).

⁹⁷ Nationalencyklopedin: uppslagsord ”verktyg”.

⁹⁸ Nationalencyklopedin: uppslagsord ”metod”.

det vill säga de(t) system som är grunden till verktyget och i vissa fall i sig självt även kan vara föremål för verktygets användning.

Ett exempel på tvetydigheten mellan metoder och verktyg är XML-formatet, som utformades som ett verktyg för att beskriva och utbyta innehåll på webben, som en delmängd av det mycket mer generella och komplexa metaspråket SGML. Metod i detta exempel är de strukturella principer som SGML står för och som anger hur en godtycklig mängd textobjekt och deras egenskaper kan beskrivas och ordnas. Varken SGML eller XML säger dock något om till exempel vilka element som bör användas för att beskriva ett godtyckligt innehåll utan istället anges generella regler för hur ett välformulerat dokumentets format måste se ut. Detta medför, anser jag, att varken SGML eller XML kan anses vara verktyg. SGML är ett metaspråk och kan å ena sidan ses som en metod som föreslår hur systematisk beskrivning och ordning av innehåll kan gå till. Å andra sidan bygger SGML på formella principer om objektens relationer och dokumentens grundläggande egenskaper. Det bör även betonas att SGML-rekommendationen inte enbart beskriver beskrivning av textbaserat innehåll, utan kan användas för olika slags notationer exempelvis musiknotation eller kemiska formler.⁹⁹ Det faktum att standarden SGML- och XML-specifikationen har utformats för att ange generella regler och principer för beskrivning av innehållets struktur och semantiska innehåll är i sig en viktig förutsättning för en rad olika verktyg (applikationer) som bygger på dessa principer.¹⁰⁰

Som exempel på hur dessa metoder påverkat ett specifikt verktyg för presentation kan nämnas att det sätt som webbläsare tolkar och visar innehåll baseras på SGML:s hierarkiska dokumentstruktur.¹⁰¹ Webbläsarna är dock programmerade att vara överseende med felaktigheter i HTML-koden för den i grunden strikt hierarkiska dokumentstrukturen. Innehållet kommer oftast att visas för användaren även om det bakomliggande dokumentet har felaktigheter i sin struktur.¹⁰²

Eftersom HTML är applikation som baseras på SGML, är HTML inte heller en metod, utan ett verktyg.¹⁰³ Enligt denna logik kan man även säga att riktlinjerna framtagna av *Text Encoding Initiative*, TEI, också är ett verktyg, ursprungligen med utgångspunkt i SGML och numer av XML.¹⁰⁴ Men med samma logik

⁹⁹”It [SGML] is a notation for any kind of structure.” Charles F. Goldfarb & Yuri Rubinsky, *The SGML handbook* (Oxford: Clarendon, 1990) s. xiii.

¹⁰⁰ Distinktionen här är viktig då SGML, som är en ISO-standard, inte har lika många föreskrivande och ”tvingande” moment som XML-specifikationen har. Själva SGML-standardens föreskriver inte den mer strikta syntax som senare blev praxis utan där är valmöjligheterna fler. Det finns många möjligheter att variera och förenkla SGML-syntaxen, t.ex. genom att utelämna en ”slut-taggar” av ett element som ingår i ett annat och överordnat element som markerar slutet på just detta innehållsobjekt. Ibid. s. 71 f. I XML är det inte tillåtet att utelämna ”slut-taggar”.

¹⁰¹ Hierarkiska dokumentstrukturer presenteras även på s. 55 f. samt s. 77 f.

¹⁰² Den nya standarden för webbdokument, XHTML, är dock inte lika förlåtande för felaktiga innehållsstrukturer och kommer att ”tvinga” innehållsproducenter att skriva välformulerade webbdokument.

¹⁰³ HTML är en dokumentmodell, en dokumenttypsdefinition (DTD) av SGML.

¹⁰⁴ TEI presenteras närmare i kapitel 4 som börjar på s. 102.

kan man även hävda att TEI är en uppsättning principer som använder sig av metoderna SGML och XML av andra potentiellt möjliga metoder. Mer självklara exempel på verktyg är de program som jag använder för att arbeta med det digitala materialet.¹⁰⁵

I avhandlingen analyserar och granskar jag ett antal för avhandlingen relevanta metoder och verktyg utifrån deras epistemologiska grunder och deras funktionalitet. Att försöka klargöra verktygens förutsättningar är även viktigt för att kunna se skillnaden mellan de ”verkliga” objekten och det sätt vi beskriver och manipulerar objekten.¹⁰⁶ De metoder och verktyg som ligger till grund för mina antaganden är förstås i fokus, övriga verktyg har ett mer sekundärt intresse och kommer inte att ges samma utrymme. TEI:s riktlinjer är ett för den här studien primärt verktyg som har utvecklats med hjälp av metoden SGML, som i sin tur grundas i principer om innehållsobjektens strukturer, relationer och beskaffenhet.

Avslutningsvis kan nämnas att forskning och praktik kring märkning huvudsakligen utvecklats i anglosaxiska sammanhang. Min uppfattning är att den svenska terminologin saknar en tydlig och stringent begreppsanvändning. Ett sådant exempel är ordet märkning, som i vissa sammanhang benämns kodning, annotering eller uppmärkning. I engelskan används *encoding* och *markup* ofta synonymt. *Encoding* (kodning) är den process som gör att information kan transformeras till ett annat format. Till exempel på det sätt tecken kodas digitalt genom binär notation. En text är alltså alltid kodad enligt något format utan att för den skull innehålla någon explicit märkning.¹⁰⁷ *Markup* används i praktiken ofta med liknande betydelse som *encoding* men avser snarare själva aktiviteten att tillföra koder till den redan teckenkodade texten. Märkning (*markup*) innebär alltid att man kodar något, men kodning (*encoding*) behöver inte innehålla märkning. Jag använder märkning synonymt med (*text*) *encoding* och *markup*.¹⁰⁸

Avslutningsvis vill jag nämna att jag i avhandlingen använder begrepp som kan tänkas vara obekanta inom pedagogiken, därför har en termlista sammanställts som innehåller korta beskrivningar av ett antal centrala termer, den återfinns i början av avhandlingen, efter listan med förkortningar.¹⁰⁹

¹⁰⁵ För exempel på dessa verktyg se kapitel 7.

¹⁰⁶ Willard McCarty, *Humanities computing* (Basingstoke; New York: Palgrave Macmillan, 2005) s. 37-41.

¹⁰⁷ Textens tecken måste representeras av ett specifikt schema, tidigare ofta Latin-1 (ISO-8859) eller det numera vanligare Unicode-formatet som klarar av att representera många fler tecken än tidigare teckenkodningar. Därtill är en text alltid ”kodad” i bemärkelsen att använda styckesindelningar, punkter, mellanrum, versaler etc. för att strukturera och ordna en text.

¹⁰⁸ I TEI står E för *encoding* och i XML betecknar M *markup*. Se HUMANIST-listans arkiv (Vol. 14) för diskussionstråd om kodning (*encoding*) vs märkning (*markup*), <http://lists.village.virginia.edu/lists_archive/Humanist/v14/0319.html>

¹⁰⁹ Termlistan återfinns i början av avhandlingen direkt efter ”Förkortningar”.

1.6 Disposition

Avhandlingen är skriven för en publik som inte kan antas ha kännedom om innehållsdesign, märkning, öppna program och annat relaterat till innehållet i avhandlingen. Därför har jag inledningsvis givit en översiktlig introduktion till digitala samlingar för forskning och undervisning samt till innehållsdesign.

Del II, *Innehållsdesign – avgränsning och inventering av principer, metoder och verktyg*, ägnas framför allt åt utgångspunkter och principer. Kapitel 2 redogör för avgränsning och utgångspunkter för innehållsdesign samt öppna standarder och metoder. Därefter ger jag i kapitel 3 en historisk överblick över märkspråkens utveckling som följs av en redogörelse över den hierarkiska innehållsmodellens principer och utgångspunkter. I detta kapitel beskrivs även alternativ till denna modell samt en översikt av metoder och verktyg för webben. I nästa kapitel, 4, presenteras *Text Encoding Initiatives*, TEI:s, bakgrund, utgångspunkter och riktlinjer.

Därefter tar del III, *Tillämpad innehållsdesign* vid och denna del inleds med kapitel 5 om prosopografiska metoder och tillämpningar. I samma kapitel introduceras det empiriska materialet, databasen samt de metoder som använts vid insamling och sammanställning av dessa material. I kapitel 6 presenteras tillvägagångssätt för beskrivning och bearbetning av detta material jämte illustreringar av de metoder och verktyg som använts. I kapitel 7 presenteras sedan de tillämpningar som utformats med utgångspunkt i det material som presenterats i föregående kapitel. Den avslutande delen IV ägnas åt sammanfattande diskussion och förslag till en verktygslåda för innehållsdesign. Där finns även en innehållsförteckning för de digitala bilagor och filer som finns på CD:n som är en del av avhandlingen. På denna CD bifogas samtliga digitala bilagor, digitala färgbilder för de svartvita figurerna i avhandlingen, samt andra bilder med exempel på tillämpningar eller metoder. Därtill en rad filer, verktyg och exempel på tillämpningar samt exempel på det digitala material som har försetts med beskrivande märkning.

Del II Innehållsdesign – avgränsning och inventering av principer, metoder och verktyg

2 Innehållsdesign – några utgångspunkter

Jag kommer i detta kapitel att inleda med en definition och beskrivning av verksamhetsområdet innehållsdesign samt grundläggande principer för märkning av innehåll. Inledningsvis har jag presenterat några exempel på utformning av digitalt material där vare sig plattform eller specifika framtida användningsområden behöver förutses eller föregripas.¹¹⁰ En viktig del av detta arbete är bruket av plattformsoberoende format, standarder och program. Därför tillägnas några avsnitt i kapitlet åt öppen källkod och öppna standarder, det vill säga de verksamheter och ideologier som är en förutsättning för öppna program, format och standarder.

2.1 Innehållsdesign – definition och avgränsning

Vad kännetecknar innehållsdesign? På vilket sätt skiljer sig denna form av informationshantering från låt säga webbdesign eller elektronisk publicering?

Peter Shillingsburg¹¹¹ presenterade under 1990-talet ett antal principer för elektronisk publicering av vetenskapliga utgåvor för *Modern Language Association* (MLA) i USA.¹¹² Shillingsburgs principer är i högsta grad aktuella för innehållsdesign och på flera punkter är de samstämmiga. Shillingsburg sätter användbarhet i första rummet. Näst viktigast är flyttbarheten mellan olika system. Därefter vikten av att innehållet i ett digitalt arkiv bör vara så interaktivt som möjligt och försett med textkritiska kommentarer samt ha rikligt med interna (*korsreferenser*) och externa länkar (*intertextualitet*). Därtill ägnas uppmärksamhet åt vikten av att bevara integriteten hos originalmaterialet, vilket är väsentligt för veten-

¹¹⁰ Se Broady, *Content design. Methods and tools for the creation of portable hypermedia archives. Notes for a proposed CID project. Version 2*; Broady, "Digitala arkiv och portföljer".

¹¹¹ Shillingsburg är professor i engelska och förestår The Centre for Textual Scholarship at De Monfort University in Leicester (<www.cts.dmu.ac.uk/> Innan dess arbetade han i många år vid Mississippi State University samt vid University of North Texas. Han har bland annat utvecklat CASE (Computer-Assisted Scholarly Edition). CASE är en samling verktyg avsedda för studier av digitala texter i hela forskningsprocessen, från analyser till presentation, se <<http://peter.shillingsburg.net/>>.

¹¹² Se Peter Shillingsburg, "Principles for Electronic Archives, Scholarly Editions, and Tutorials" *The literary text in the digital age*, red. Richard J. Finneran, Editorial theory and literary criticism (Ann Arbor: University of Michigan Press, 1996).

skapliga utgåvor. Vidare framhålls att arkivet bör vara utbyggbart och ge möjlighet att infoga nytt material.¹¹³

Några år efter Schillingsburg presenterade Donald Broady riktlinjer för ett nytt verksamhetsområde som gavs benämningen *content design*¹¹⁴ eller innehållsdesign.¹¹⁵ Broady urskiljer ett antal grundläggande principer för innehållsdesign som bör innehålla möjligheter att:

- separera innehåll (intellektuellt/semantiskt) från dess form (struktur och dokumenthierarkier) och presentation,
- infoga information (genom märkning/textkodning) om innehållet,
- använda internationella standarder för informationshantering,
- lagra digitalt innehåll i plattformsoberoende format,
- med hjälp av metoder och verktyg för innehållsdesign nyttja dessa möjligheter för att iordningställa modulariserat, utbyggbart och portabelt material.

De idéer som formulerades under 1990-talet realiserades i olika projekt vid Centrum för användarorienterad IT-design (CID) på KTH.¹¹⁶ Detta resulterade bland annat i märkning av ett antal utgåvor av nationalutgåvan av Strindbergs samlade verk¹¹⁷ och projektet AntiLoop; ett system och verktyg som skapades

¹¹³ Se Modern Language Association (MLA) där finns det första utkastet till dagens principer, vid: <<http://sunsite.berkeley.edu/MLA/principles.html>>. För aktuella riktlinjer från MLA, se <www.mla.org/resources/documents/rep_scholarly/cse_guidelines>. MLA som har sitt säte i USA är den främsta och största internationella organisationen för litteraturvetenskap och litteraturkritik och bildades 1883. Syftet är att stärka studier av, och undervisning om, litteratur och moderna språk. MLA ansvarar för den vanligaste stilistiska manualen inom humaniora, MLA Style Manual. Se <www.mla.org/>. Se även MLA:s riktlinjer för elektronisk vetenskaplig utgivning i Lou Burnard, Katherine O'Brien O'Keeffe & John Unsworth, red, *Electronic textual editing* (New York, N.Y.: Modern Language Association of America, 2006).

¹¹⁴ Broady, *Content design. Methods and tools for the creation of portable hypermedia archives. Notes for a proposed CID project. Version 2.*

¹¹⁵ Om argument för innehållsdesign i utbildningssammanhang, se Broady, "Digitala arkiv och portföljer". För en beskrivning av några av de tidiga projekten se, Hans Melkersson, "Innehållsdesign. Om principer för märkning och lagring av innehåll" *Fältanteckningar: utbildnings- och kultursociologiska texter tillägnade Donald Broady*, red. Mikael Börjesson, et al., Skeptronserien (Uppsala: Forskningsgruppen för utbildnings- och kultursociologi ILU Uppsala universitet, 2006).

¹¹⁶ Dåvarande CID, hemmahörande vid Institutionen för Numerisk analys och datalogi, NADA, vid Kungl Tekniska Högskolan. Se, <<http://cid.nada.kth.se/>> för historik och information om CID. [2007-03-20]

¹¹⁷ Donald Broady, "Hur kodar man Röda Rummet?" *ABM, IT och forskningen. Rapport från en konferens på Kungliga Biblioteket den 17 november 1999*, red. Mats Rolén, (Stockholm: Riksbankens Jubileumsfond, 2000). samt Donald Broady & Hasse Haitto. "Internet and the humanities: the promises of Integrated Open Hypermedia". *Contemporary computer and network technologies*, 17-18 Jan., Moskva, 1996. Se även <www.skeptron.uu.se/broady/dl/dl-proj.htm> för fler projekt och publikationer med anknytning till innehållsdesign.

för hantering och bearbetning av digitala resurser, till exempel det märkta Strindbergsmaterialet.¹¹⁸

Termen ”innehållsdesign” är, enligt Broady, en lämplig benämning för att skilja verksamhetsområdet från andra mer etablerade designverksamheter, såsom grafisk design och systemdesign. En innehållsdesigner behöver inte vara upphovsperson till själva innehållet utan kan vara den som sammanställer, beskriver och ordnar materialet för vidare bearbetning och manipulering men även förbereda för senare presentation. Det kan finnas anledningar till att skilja på upphovspersonen och de mer redaktionella uppgifterna för att en innehållsdesigners arbetsuppgifter liknar den tekniska redaktörens uppgifter. En innehållsdesigners verksamhet går alltså ut på att:

[...] åstadkomma material som kan användas och återanvändas i sammanhang som inte kan förutses och som inte bör föregripas.¹¹⁹

Däriigenom kan man skapa möjligheter för att ett och samma innehåll kan filtreras, kombineras, organiseras och presenteras på flera olika sätt för att passa olika syften och behov hos skilda kategorier av användare. Detta arbete utförs av en teknisk redaktör eller en informationsarkitekt som har kunskaper om samtliga moment i arbetskedjan. Noteras bör dock att även lärare, forskare och studerande givetvis kan tillämpa vissa moment och verktyg för innehållsdesign utan att för den skull behärska samtliga moment i arbetsprocessen.

Materialets funktionalitet är en viktig aspekt av innehållsdesign. Användningen av innehållet i digitala arkiv är lika beroende av arkivets konstruktion som det bärande mediets funktionalitet. Det är en väsentlig skillnad om innehållet presenteras som en avbildning, till exempel ett faksimil, jämfört med en digitaliserad text som tillåter manipulering av olika slag. Olika uppfinningar har därför under historiens gång underlättat för användare att orientera sig i och använda texter. Möjligen är det dock ett uttryck för historisk arrogans, som Johan Svedjedal formulerar det, att tro att det senaste halvseklets förändringar och rationaliseringar av former för skriftlig kommunikation i förstone medför någon genomgripande förändring av vårt sätt att använda texter på det sätt som ovan skisserats.¹²⁰ Likaså konstaterar Svedjedal att övergången från papyrusrullar till *codex* var en process som tog många år och då med anledning av att nya tekniska förutsättningar, nämligen pergamentmaterialet, bättre lämpade sig för att konstruera ark och blad än vad papyrus gjorde.¹²¹

Brukarna konfronteras ju heller aldrig, enligt Roger Chartier, med något abstrakt innehåll frigit från materiella betingelser, något som han menar påverkar innehållets mening:

¹¹⁸ Se Melkersson, ”Innehållsdesign. Om principer för märkning och lagring av innehåll”. för mer information om AntiLoop systemet.

¹¹⁹ Broady, ”Digitala arkiv och portföljer” s. 18.

¹²⁰ Johan Svedjedal, *Den sista boken* (Stockholm: Wahlström & Widstrand, 2001) s. 20.

¹²¹ Ibid. s. 22 f.

Mot en rent semantisk definition av texten – som inte bara förekommer inom den strukturalistiska kritiken i alla dessa varianter, utan också i de litteraturteorier som är allra mest angelägna om att rekonstruera receptionen av verken – måste man hävda att formerna producerar mening och att en text som är oförändrad till bokstaven förses med en helt ny mening och status när de mönster förändras som bjuder ut den till tolkning.¹²²

Digitala texter är ju inte fångna i sin ursprungliga materiella form och därmed blir innehållet heller inte detsamma i nya formella arrangemang, som genom övergång från *codex* till skärm.¹²³ Det medverkar till nya möjligheter för brukaren som till exempel att själv kunna manipulera texten. Men det innebär även att vi behöver förstå dessa verktygs och metoders funktionalitet och historisering och inte minst deras underliggande antaganden, samt begrepps- och teorbildning.

Mediets betydelse för texten är inte alltid självklar. Inom till exempel lingvistik har en immateriell tolkning av text varit dominerande sedan länge. Som en motreaktion på det lingvistiska immateriella synsättet har andra modeller för textbegreppet arbetats fram av forskare inom andra områden än lingvistik, bland annat de som ovan nämnts.¹²⁴ Den kritik som framförs handlar dels om mediets påverkan på textens ”beteende”, dels om att texten *per se* inte kan likställas med den information den förmedlar, något som bland annat Espen Aarseth hävdar i sin studie om ergodisk litteratur. Han framhåller att texter inte kan separeras från sina materiella betingelser och texter är objekt, vilka som helst, vars primära funktion är att vidarebefordra information.¹²⁵

Även Mats Dahlström påminner oss om de medietypiska och materiella förutsättningarnas betydelse för det sätt man ser på och studerar skriftlig kommunikation i form av verk och dokument.¹²⁶ Han menar att vi riskerar trivialisera det bärande mediets betydelse för innehållet om vi betraktar digitala arkiv som immateriella instanser av tryckta texter. Att enbart se digital text som lingvistiska teckensekvenser medför också att vi missar en dimension av den komplexa hanteringen av innehåll.¹²⁷

Med de argument som ovan presenterats om hur material och form, dels sätter upp ramar för, dels påverkar det semantiska innehållet, skulle man kunna hävda att det inte går att separera innehållet från dess form. Broady ställer dess-

¹²² Roger Chartier, *Böckernas ordning: läsare, författare och bibliotek i Europa mellan 1300-tal och 1700-tal*, 1. uppl. ed. (Göteborg: Anamma, 1995) s. 15.

¹²³ Ibid. Men även Roger Chartier, *Forms and meanings: texts, performances, and audiences from codex to computer* (Philadelphia: Univ. of Pennsylvania Press, 1995).

¹²⁴ Här hänvisas till Mats Dahlströms redogörelse om textbegreppets materiella betingelser. Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 70-76. Se även Johan Svedjedals framställning om olika typer av textbärare och dess olika förutsättningar för strövtåg i texten. Svedjedal, *Den sista boken* s. 11-32.

¹²⁵ Espen Aarseth, *Cybertext: perspectives on ergodic literature* (Bergen: Univ. of Bergen, 1995) ssk. s. 70.

¹²⁶ Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 70 ff. Även Mats Dahlström, ”Nya medier, gamla verktyg”, *Human IT* 6.4 (2002).

¹²⁷ Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* ssk. s. 71-72.

utom frågan om att eftersom all kunskapsutveckling sker i ett sammanhang är det naivt att tro att det skulle vara möjligt att skilja på innehåll och form.¹²⁸ Det är möjligt, men det som bör hållas i åtanke är att eftersom kunskapsutveckling såväl som de metoder som finns för kunskapsorganisation är relaterade påverkas även metoderna av sammanhanget. Därför: att skilja på innehåll och form är att vara explicit i två avseenden: dels beskriva vad som utgör form och vad som utgör ett intellektuellt och semantiska innehåll, dels klargöra de praktiker, underliggande antaganden och perspektiv som ligger till grund för innehållets beskrivning, ordning och användbarhet.

Efter denna inledning om vad innehållsdesign kan vara presenteras i nästa avsnitt de grundläggande principerna för märkning av innehåll i syfte att introducera läsaren i beskrivande märkning som presenteras närmare i kapitel 3 och 4.

2.2 Grundläggande principer för märkning av innehåll

Det vanligaste filformatet för innehållsmärkning idag är XML, *Extensible Markup Language*, som är ett beskrivande märkspråk för lagring och utbyte av data.¹²⁹ För att beskriva ett visst innehåll med hjälp av så kallade märkspråk används märkord (element) som är inneslutna i avskiljare som markerar elementets början (<), benämning (**starttagg**) och slut (>), sammantaget blir det en <**starttagg**>. För att markera var det innehåll som skall inneslutas av märkningen sedan slutar används en specifik identifierare (/) som markerar att detta är en <**sluttagg**>.¹³⁰ Ett element består alltså av en ”starttagg” och en ”sluttagg” som markerar, beskriver och innesluter det valda innehållet.

<element> innehåll </element>

Figur 1. Exempel på den deskriptiva märkningens grundläggande principer: elementet.

¹²⁸ Broady, ”Digitala arkiv och portföljer” s. 14.

¹²⁹ *Extensible* för att det kan utvidgas och anpassas samt *Markup Language* för att det använder märkord för avgränsning och strukturering av innehåll. XML-formatet är plattformsberoende. XML är förenklat och härlett ur den mycket mer omfattande och äldre SGML-standarden (ISO 8879:1986). Se vidare <www.w3.org/XML/>. Se även standardverket Goldfarb & Rubinsky, *The SGML handbook*.

¹³⁰ I exemplet i *Figur 1* ovan visas principerna för XML där sluttagg alltid måste förekomma. Märkningens principer är till viss del schemaspecifika och i vissa märkspråk behövs t.ex. ingen sluttagg, se vidare avsnitt 3.2 s. 73 f. För introduktion beskrivning av den deskriptiva märkningens grundläggande principer se, *Ibid.* s. 5-17. Se även kapitel 1: ”A Gentle Introduction to XML” i Lou Burnard & Michael Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange*, XML Version (Oxford, Providence, Charlottesville, Bergen: Text Encoding Initiative Consortium, 2002).

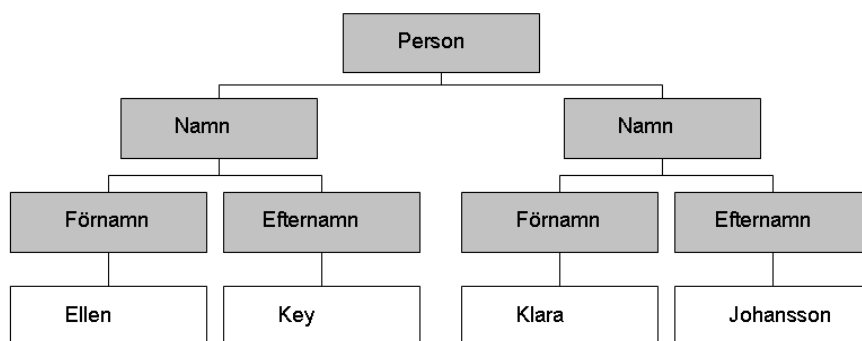
Elementen kan sedan nästlas, det vill säga inneslutas i varandra, så att en hierarkisk struktur bildas som visas i figuren nedan.¹³¹

```
<rot>
  <element1>Innehåll 1
    <element1a>Delmängd 1a av Innehåll 1</element1a>
    <element1b>Delmängd 1b av Innehåll 1</element1b>
  </element1>
  <element2>Innehåll 2</element2>
</rot>
```

Figur 2. Exempel på den deskriptiva märkningens hierarkiska struktur med hjälp av nästlade element

I Figur 2 i ovan kan vi se hur de hierarkiska principerna utgår från att ordningen av den i detta fall sekventiella informationen, till exempel löpande text, måste ha ett rotelement (<rot> i Figur 2) som innesluter alla förekommande element. ”Inuti” rotelementet nästlas sedan de efterföljande elementen, i exemplet ovan är elementen <element 1a> och <element 1b> nästlade i <element 1>. Relationen mellan dessa två element kan beskrivas som att de är syskon till varandra och barn till <element 1> som i sin tur således är förälder. Enligt samma princip är även <element 1> och <element 2> syskon då de befinner sig på samma nivå i hierarkin. Elementet <element 2> är dock inte förälder då detta element inte innesluter några ytterligare element.¹³²

I XML beskrivs och ordnas innehållet på det hierarkiska sätt som jag gett exempel på ovan, vilket innebär att innehållets struktur kan liknas vid ett träd.



Figur 3. XML: hierarkisk dokumentmodell.

¹³¹ Termen ”nästla” betecknar ungefär det som är inuti och kommer närmast efter i en viss ordning.

¹³² Termerna ”föräldrar”, ”barn” och ”syskon” är en direkt översättning av engelska vedertagna benämningar för beskrivning av hierarkiska strukturer i XML.

Deskriptiva metaspråk använder en hierarkisk dokumentmodell som utgår från att varje innehållsobjekt relaterar till, och ingår i, andra objekt.¹³³ Innehållsobjekt avser används här för att beteckna de logiska, semantiska eller strukturella delar ett dokument består av. I *Figur 3* illustreras den hierarkiska dokumentmodellen som ett träd med elva noder av två skilda typer, som benämns ”elementnoder” respektive ”textnoder”. De grå noderna är element och de vita noderna representerar text och därmed det innehåll, som elementen **<förnamn>** och **<efternamn>** beskriver. Det måste alltid finnas ett, och endast ett, överordnat element i ett XML-dokument. Under roten, **<person>**, kan det dock förekomma fler nästlade element, i detta fall är kvinnornas för och efternamn nästlade i ett överordnat namn-element. Vanligtvis används ett schema som specificerar regler för vilka element som får användas och var de får användas. Detta schema beskriver då en viss dokumentmodell.¹³⁴ Flera element kan förekomma på samma nivå i en hierarki enligt samma princip som de två ”syskonelementen” **<namn>** i exemplet ovan. Hur elementen kan nästlas eller förekomma på olika hierarkiska nivåer är beroende av dokumentmodellens regler och restriktioner.¹³⁵

I strikt bemärkelse avser termen *element* en kombination av taggarna och det innehåll de innesluter och inte enbart starttaggen **<namn>** och sluttaggen **</namn>** som man kan föranledas att tro utifrån exemplet som presenterades inledningsvis. I praktiken används dock *element* i båda betydelserna. Elementen är, som vi ovan sett, egentligen benämningen på de logiska delar innehållet byggs upp av. En bok består ofta av kapitel som i sin tur innehåller stycken som består av meningar osv. Dessa element skiljer man åt genom att tillfoga markörer, som i figurerna ovan.¹³⁶

Principerna för en hierarkisk dokumentmodell utgår från att större innehållsobjekt består av mindre delar, som till exempel kapitel, stycken, meningar och ord, ned till minsta meningsbärande enhet.¹³⁷ Den hierarkiska strukturen medför att innehållsobjekten och relationerna mellan dessa bestäms av en specifik ordning, något som ibland kan vara en nackdel. Speciellt om man i ett och samma dokument behöver beskriva olika aspekter, som i sin tur har olika struktur jämfört med den ordnande principen. Blandningen av XML:s notation och den faktiska märkta texten kan dock ge upphov till konflikter mellan denna mer

¹³³ I dessa sammanhang vanligen benämnd Ordered Hierarchy of Content Object (OHCO). Modellen presenteras och jämförs med andra dokumentmodeller bland annat i: Steven DeRose, et al., ”What is Text, Really?” *Journal of Computing in Higher Education* 1.2 (1990). Modellen har kritiserats av ett flertal forskare, se vidare avsnitt 4.4 s. 113.

¹³⁴ I avhandlingen används schema och dokumentmodell synonymt.

¹³⁵ Dokumentmodeller kan i sin tur definieras i olika format och med olika typer av notationer. Se vidare avsnitt 3.5 s. 82.

¹³⁶ Goldfarb & Rubinsky, *The SGML handbook* s. 18-19.

¹³⁷ Den minsta meningsbärande enheten varierar beroende på hur innehållet beskrivs. En lingvistisk analys beskriver oftast andra aspekter än en materialhistorisk analys.

abstrakta klassifikationsidé och de konkreta texter som oftast innehåller flera komplexa innehållsliga strukturer.¹³⁸

Det är dock möjligt, om än inte önskvärt, att utforma en näst intill ”platt” organisation av innehållet, med endast ett rotelement och med samtliga efterföljande element på nästa, och samma, hierarkiska nivå.¹³⁹ Nedan beskrivs ett annat exempel på hur en sådan ”platt” märkning kan se ut. Exemplet är hämtat ur det prosopografiska material som tillämpas i avhandlingen som består av uppgifter som samlats in om kvinnornas sociala bakgrund, utbildning, verksamhetsområden, yrkesliv, civilstatus, bostadsort med mera.¹⁴⁰

I figuren nedan beskrivs ett urval av en persons egenskaper, i detta fall Ellen Key.¹⁴¹

```
<node id="Ellen Key">
  <data id=N_23>
    <birthplace value="landsort">
      Född på landsorten
    </birthplace>
    <mentioned value="26">
      Omnämnad av 26 andra kvinnor
    </mentioned>
    <sphere value="PUB">Publicist </sphere>
    <education value="Lärarinneseminarief">
      Lärarinnekursen
    </education>
  </data>
</node>
```

Figur 4. Exempel på märkning av en viss kvinnas egenskaper och tillgångar

Denna beskrivning kan sedan användas och bearbetas på olika sätt. Man kan skapa index med hjälp av märkorden, sortera om ordningen av innehållet, utföra strukturerade sökningar, extrahera delar av det märkta innehållet eller infoga nya delar. Med hjälp av märkningen blir det även möjligt att flytta utvalda data till en databas. Det går även att förbereda för presentation av innehållet i olika format, till exempel i HTML eller PDF-format. Märkningen medför även migrera av innehållet till andra filformat eller applikationer i syfte att bearbeta eller analysera det.¹⁴² Genom att tillfoga en explicit beskrivning av innehållet ges även förutsättningar för andra att använda innehållet med den aktuella beskrivningen eller att kritisera, ändra och göra tillägg i beskrivningen av innehållet. Ett exempel på hur märkningen kan användas är beskrivningen av sociala kontaktnät för en viss grupp kvinnor i materialet. Då denna aspekt inte kunde inordnas

¹³⁸ Dessa konflikter diskuteras närmare på s. 113 f.

¹³⁹ Om dokumentmodellen utformas så, förloras också den hierarkiska ordningen av innehållet och texten representeras istället snarare i form av sekventiellt förekommande innehållsobjekt.

¹⁴⁰ Se vidare s. 196 f. för beskrivning av indikatorer på kvinnornas egenskaper och tillgångar.

¹⁴¹ Exemplet nedan visar en förenklad märkning med GraphML där en hel del information är borttagen.

¹⁴² Migrering är en term för att överföra av data mellan olika lagringsformat och/eller lagringsplatser.

i den dokumentmodell som användes, löste jag det genom att ge varje annan kvinnas namn som nämndes i varje kvinnas biografi en unik identitet för att på så sätt relatera varje nämnt namn till en specifik kvinnas kontaktnät.¹⁴³

I *Figur 5* visas ett exempel på hur den relativt sett ”platta” beskrivningen kan användas.¹⁴⁴ Här har märkningen i *Figur 4* använts som utgångspunkt för en nätverksanalys av de 101 kvinnornas kontakter inom gruppen.¹⁴⁵



Figur 5. Urklipp ur nätverksgraf med Ellen Keys egenskaper (attribut) accentuerade

De kvinnor som hade många kontakter med övriga kvinnor i urvalet befinner sig i centrum av nätverksgrafen och kvinnorna med få kontakter återfinns således i periferin. Konstruktionen av graferna genom noder (som representerar kvinnorna) och de riktade länkarna (som representerar deras kontakter) baseras delvis på binära relationer.¹⁴⁶ Denna information har i sin tur utvunnits ur det

¹⁴³ Metoden för att beskriva kvinnornas kontaktnät presenteras närmare på s. 208 f.

¹⁴⁴ Grafen tillgänglig i vid <CD:/figurer/figur5.jpg>.

¹⁴⁵ Exemplet är ett urklipp, alla kvinnor är därför inte representerade i grafen. Se kapitel 7.2 s. 208 om nätverksanalyser.

¹⁴⁶ Varje kvinna har antingen kontakt eller inte kontakt med varje annan kvinna i gruppen.

digitaliserade och märkta innehållet. I analysen finns möjligheter att inkludera ytterligare uppgifter om kvinnornas egenskaper som anger till exempel utbildning, verksamhetsområde, födelseort etcetera (det vill säga den märkning som exemplifierades i *Figur 4*). Uppgifter kan sedan användas för olika slags analyser, exempelvis filtrering av undergrupper i nätverket eller annan strukturering av grafen.¹⁴⁷

De exempel som visats ovan ger en inblick i den deskriptiva märkningens grundläggande principer. I kapitel 3 beskrivs och analyseras de metoder och de verktyg som baseras på principerna som ovan lagts fram. Nedan ges en översiktlig introduktion om *Open Source* – en praktik som haft stor betydelse för utvecklingen av metoder och verktyg för innehållsdesign.

2.3 Open Source – öppna standarder, metoder och verktyg

Syftet med de följande två avsnitten är att ge en översikt över vad som är innebörden i öppen källkod och öppna standarder.¹⁴⁸ Avsikten är att belysa några väsentliga utgångspunkter för innehållsdesign och därmed de metoder och verktyg som är föremål för min studie. De program och de metoder som är ett resultat av utvecklingen inom öppen källkodsrörelsen och öppna standarder är inte bara relevanta för den här studien utan för stora delar av internets infrastruktur. Som exempel kan nämnas att merparten av webbservrar använder öppen programvara och öppna operativsystem samt att merparten av e-posten skickas med hjälp av öppen programvara.¹⁴⁹

Syftet med dessa verksamheter går i huvudsak ut på att användare skall kunna bygga vidare på varandras verk och därigenom få tillgång till en ständigt växande källa med information och material, öppen för andra att nyttja och vidareutveckla.

2.3.1 Öppen källkod och öppna program

Idéerna som ligger till grund för öppen källkod utvecklades ursprungligen av programmerare för programmerare i datorernas barndom under 1960-talet.¹⁵⁰ Utvecklingen startade bland *hackers*, en begränsad skara människor som behärs-

Den XML-fil som analysen i *Figur 5* baseras på finns tillgänglig vid <CD:/xml/ffo-network1.xml>.

¹⁴⁷ Se kapitel 7.2 s. 208 för exempel på filtrering och strukturering av grafer i nätverksanalyser.

¹⁴⁸ Översättning av *Open Source*. I sammanhanget används även *öppen programvara* vilket avser sådana program som bygger på öppen källkod.

¹⁴⁹ Webbservern Apache används till drygt 50 procent av samtliga webbplatser på Internet (dec. 2007), och är därmed störst trots en minskning från 2005 då andelen Apache servrar var nästan 70 procent. Se <http://news.netcraft.com/archives/web_server_survey.html>. Linux används alltmer som operativsystem för webbservrar och Sendmail är vanligt för hantering av e-post.

¹⁵⁰ Under 1960-talet var datorer inte lika med PC eller MAC.

kade den tidens stordatorer.¹⁵¹ Inom hackerkulturen delade man med sig av information och program och skapade därmed en gemenskap där tillgång till information var fri och detta arbetssätt gynnade den fortsatta utvecklingen.¹⁵² Under 1970-talet när utvecklingen av datorer gjorde stora framsteg började företagen visa större intresse för utvecklingen vilket ledde till att det blev viktigt att hålla källkoden hemlig för konkurrenter. Detta var en av anledningarna till att Richard Stallman i början av 1980-talet startade GNU-projektet som arbetade för fri programvara. Vissa av de fria program som spreds och utvecklades med hjälp av internet under 1980-talet blev mer framgångsrika än de kommersiella dito.¹⁵³ Den tredje generationens *hackers* fick nu en ny framtid bland annat i framväxande spelföretag och gruppen kunniga, dataintresserade ökade i takt med att datortekniken blev tillgängliga även utanför institutionerna.¹⁵⁴

Rörelsen är tydligt ideologiskt driven och som namnet antyder är öppenhet ett viktigt begrepp. Det handlar inte om att allt ska vara gratis utan om förutsättningar för att dokumentation, regler och källkod, skall kunna distribueras fritt och vara öppna för insyn. Förespråkare för öppen källkod framhåller även distribuerade samarbetsformer som en del av konceptet. Vem som helst skall kunna använda och läsa, eller modifiera och vidare distribuera programmen. På så vis kan användare modifiera och anpassa program efter sina behov, förutsatt att nödvändiga programmeringskunskaper finns. Modifikationer, förbättringar och tillägg publiceras och görs vanligen till en del av programmets officiella version. TEI:s riktlinjer är tillgängliga under GNU-licensen och tillåter användare att fritt använda, kopiera, distribuera och förändra riktlinjerna under förutsättning att källan refereras och att information anges om att riktlinjerna är tillgängliga under licensen GNU-GPL.¹⁵⁵

Intresseorganisationen *Open Source Initiative* (OSI) bildades 1997 och arbetar med riktlinjer och licenser samt att sprida information om projektet.¹⁵⁶ Bildandet av OSI var en brytpunkt med den redan existerande och hackerorienterade organisationen *Free Software Foundation* (FSF).¹⁵⁷ Richard Stallman bildade FSF

¹⁵¹ Benämningen *hacker* myntades vid Massachusetts Institute of Technology (MIT) under det tidiga 1960-talet inom en grupp framstående AI-programmerare, där man utvecklade de programmeringsspråk, verktyg och den hackerslang, som fortfarande präglar den s. k. hackerkulturen. De första svenska undersökningarna av hackerkulturen var Eva Erson, "*Det är månen att nå*": en studie i några datorintresserade pojars språk och föreställningsvärld (Umeå: Institutionen för nordiska språk, Umeå Universitet, 1992); Jörgen Nissen, *Pojkarna vid datorn: unga entusiaster i datateknikens värld*, Linköping studies in arts and science, 89 (Stockholm; Stehag: Symposion Graduale, 1993)..

¹⁵² För ordlista och teknisk introduktion till utveckling av program samt allmänt om hackerkulturen, se: <www.catb.org/jargon/html/>. Genom utvecklingen av ARPANet, operativsystemet UNIX och språket C fick idéerna och verktygen snabb spridning inom akademien.

¹⁵³ Ett exempel på det är X Windows System, ett grafiskt gränssnitt som introducerades 1984 för UNIX-datorer och nätverk, numer även för Linuxsystem. Se <www.x.org/>.

¹⁵⁴ För en grundlig och intressant historisk genomgång av hackerkulturen, se Nissen, *Pojkarna vid datorn: unga entusiaster i datateknikens värld* s. 83-119.

¹⁵⁵ General Public License (version 2), för TEI:s riktlinjer för åtkomst och referenshantering, se <www.tei-c.org/Guidelines/access.xml>.

¹⁵⁶ Se <www.opensource.org/> för mer information om verksamhet, metod och riktlinjer.

¹⁵⁷ Se <www.fsf.org/>.

1985 för att finansiera utvecklingen av GNU-projektet som han startade några år tidigare för att utveckla ett fritt och oberoende operativsystem.¹⁵⁸

Uppmärksamheten för den organiserade verksamheten för öppen källkodsmetoden ökade genom den text som Eric Raymond publicerade 1997 *The Cathedral and the Bazaar*, som handlar om olika metoder och arbetssätt vid systemutveckling.¹⁵⁹ Raymond ställer två olika utvecklingsmetoder mot varandra. I katedralmetoden, bedrivs utveckling som ett mer slutet och hierarkiskt styrt arbetssätt med begränsad insyn, där information enbart är tillgänglig för projektets medarbetare. Proprietära program, som till exempel Word och andra program i *Microsofts Office*, utvecklas ofta på detta sätt.¹⁶⁰ Genom basarmetoden – katedralens motsats – är information och källkoder öppet tillgängliga och man arbetar enligt en distribuerad samarbetsmodell. De som vill delta i ett projekt kan hjälpa till med utvecklingen medan andra enbart använder programmen utan att bidra. I basarmetoden har användare och inte enbart programmerare en viktig roll i systemutvecklingen. Ett centralt antagande Raymond gör är att om dokumentation och källkod finns tillgänglig och därmed exponeras för ett större antal människor kommer felaktigheter och dålig funktionalitet upptäckas snabbare än i katedralmetoden där få personer ägnar mycket tid åt att leta efter fel och förbättra programmen. En stor del av utveckling av system går nämligen ut på att testa funktionalitet och hitta fel, något som med fördel utförs av användare snarare än systemutvecklarna och här anser Raymond att:

The next best thing to having good ideas is recognizing good ideas from your users. Sometimes the latter is better.¹⁶¹

Raymond använder bland annat Linus Torvalds och Linux-projektet som exempel på en fungerande distribuerad och öppen arbetsmetod. Raymond framhåller dessutom vikten av modularitet som här innebär att programmen och systemen kan byggas ut och förbättras genom att nya byggstenar infogas eller gamla förbättras. Modulariteten medför även att projekten blir väl avgränsade och inte för stora något och därmed svåra att överblicka. Dessutom anser Ray-

¹⁵⁸ Se <www.gnu.org/home.sv.html> för mer information om GNU, GNU-GPL (licensen), fri programvara och ideologiska ståndpunkter.

¹⁵⁹ Eric S. Raymond, *The cathedral and the bazaar. Musings on Linux and open source by an accidental revolutionary*, Rev. ed. (Sebastopol, Calif.: O'Reilly, 2001). Nätopplagan tillgänglig vid: <www.catb.org/~esr/writings/cathedral-bazaar/>. Första versionen publicerades som paper vid en konferens 1997, se <<http://opensource.org/history>>.

¹⁶⁰ Liknelsen med proprietära program och katedraler är vanlig. Raymond talar dock enbart om fri mjukvara och nämner inte proprietära program i sin bok. Där gör han istället katedralliknelsen med bland annat Emacs, en mycket populär plattform för editering och programmering för UNIX och Linuxanvändare.

¹⁶¹ I Raymond, *The cathedral and the bazaar. Musings on Linux and open source by an accidental revolutionary*. Citat från nätopplagan vid <www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/ar01s07.html> .

mond att personer som får syssla med det de är intresserade av ofta är benägna att investera mycket tid och kunskap i ett projekt.

De begrepp som Raymond använder har senare utvidgats för beskrivning av olika metoder och projekt som inte i första hand handlar om programmering. *Wikipedia*, TEI och inte minst *Sourceforge* kan beskrivas som basarmetoder.¹⁶² *Sourceforge* som är en samlingsplats för utveckling av öppen programvara är exempel på hur projekt och program kan utvecklas med den distribuerade samarbetsmodellen. I verkligheten är det förstås inte vattentäta skott mellan katedral- och basarmetoden, många projekt har en hierarkisk struktur, mer eller mindre explicit formulerad.

I *Open Source Initiative*, OSI:s, definition av öppen källkod fokuserar på utgångspunkterna: frihet, öppenhet och integritet (i det sistnämnda ingår riktlinjer för upphovsrätt och användningen).¹⁶³ För att distribuera program med en öppen källkodslicens skall distributionen vara fri, det vill säga ingen licens får förhindra att sälja eller ge bort det man utvecklat. Källkoden måste vara tillgänglig och möjlig att modifiera och licensen får inte begränsas till att gälla en specifik produkt. Därtill får ingen diskriminering av vare sig personer eller insatsområden förekomma, vilket medför att programmen även kan användas kommersiellt.¹⁶⁴

Inom *Free Software Foundation* (FSF) skapades en licens för källkoden till de fria program som utvecklades. Användare skulle ha oinskränkt rätt att förutom använda även kunna ändra och förbättra programmet. FSF:s riktlinjer, som publicerades flera år tidigare än OSI:s, liknar kriterierna för öppen källkod men FSF anses ofta vara en mer strikt och ideologisk färgad organisation än öppen källkodsrörelsen som är mer pragmatisk i sin framtoning och som framhåller fördelar med öppen källkod i motsats till mer stängda metoder. Dessutom anses från OSI:s håll att program skall kunna spridas utan att för den skull behöva ansluta sig till GNU-licensen.¹⁶⁵

Creative Commons grundades i USA år 2001 och är ett annat exempel på licenser som använder sig av liknande principer som fri programvara och öppen källkodsrörelsen.¹⁶⁶ I *Creative Commons* licenser för ”kreativa verk” används upphovsrätten i kombination med särskilda villkor för användning och spridning.

¹⁶² <<http://sourceforge.net/>> är den mest populära webbplatsen för öppna källkodsprojekt, i juni 2007 fanns drygt 1,6 miljoner användare och 153 000 projekt registrerade.

¹⁶³ OSI:s definition författades till stora delar av Bruce Perens, medarbetare till Raymond och liknar licensen för operativsystemet Debian Linux. Debian innehöll mjukvara som inte distribuerades under GNU-licensen, men som kunde anses vara fri. I ett inlägg 1999 i e-postlistan för Debian riktar Perens kritik mot hur OSI utvecklats i jämförelse med FSF. Tillgängligt vid <<http://list.s.debian.org/debian-devel/1999/02/msg01641.html>>.

¹⁶⁴ En svensk översättning av definitionen av öppen källkod finns tillgänglig vid: <www.free-soft.org/mirrors/www.opensource.org/docs/osd-swedish.php>.

¹⁶⁵ Olika slags licenser ryms inom öppen källkod, däribland GNU-GPL. Kommersiell försäljning av programmen är tillåten samt att *copyleft* inte är ett krav. *Copyleft* är en del av GNU GPL-licensen med avsikt att se till att fri programvara är fri även i modifierade eller utökade versioner. Se vidare Statskontoret, *Öppen programvara* (Stockholm: Statskontoret, 2003) s. 5 ff.

¹⁶⁶ Här ingår idéerna bakom *copyleft*-rörelsen.

Upphovspersonen väljer själv vilka rättigheter hon eller han vill behålla enligt vissa principer. Det kan vara att verket får spridas för icke-kommersiell användning eller att upphovspersonen alltid måste omnämnas.¹⁶⁷

Den ideologiska logiken i öppen källkodsrörelsen, som är av akademiskt ursprung, följer inte den ekonomiska marknadens värdering av tillgångar. I de här intressegemenskaperna värdesätts i första hand sådana sorters tillgångar som är specifika för den intressegemenskap man ingår i. Dessa tillgångar förvärvas och ackumuleras inom intressegemenskapen, till exempel genom goda programmeringskunskaper och intelligenta problemslösningar, en nödvändighet för att skaffa sig högt anseende inom intressegemenskapen.

Kreativiteten och aktiviteten i vissa projekt är hög och många deltagare tillbringar fritiden att utveckla program medan andra är anställda för att göra just detta.¹⁶⁸ Att synas och tillerkännas värde för det man åstadkommer kan vara en viktig förutsättning för att inkluderas i gruppens inre kärntrupp. Denna mindre grupp är ofta väldigt aktiv och driver utvecklingen framåt medan den större gruppen perifera deltagare kan vara relativt inaktiv. Dessa intressegemenskaper är dock beroende även av de perifera deltagarna, som testar och marknadsför produkten och som kan utgöra grund för nyrekrytering till den inre kärntruppen.

I Statskontorets rapport från 2003 om öppen programvara i statlig förvaltning presenteras såväl fördelar som nackdelar med öppen programvara. Bland annat nämns möjligheter att modifiera källkod, oberoendet av stora företag och möjligheterna att skraddarsy och anpassa produkter för olika målgrupper av användare – det vill säga ett funktionellt angreppssätt snarare än produktstyrkt. Som nackdelar nämns initialt stora kostnader vid en övergång från proprietära till öppna system, på grund av omfattande migrationsarbete samt kostnader för kompetensutveckling för användarna.¹⁶⁹ En annan nackdel är den relativt krångliga tillgången till underhåll och support. Eftersom många av programmen utvecklas med hjälp av volontära insatser existerar inte samma typer av stödorganisation som för kommersiell programvara. Användare måste själva hitta till de rätta e-postlistorna och frågelådorna, dessutom bör man klara av att ställa frågorna på rätt sätt för att få svar. När man väl är inne i gemenskapen är tillgången på hjälp mycket god och instruktiv. Detta är återigen ett exempel på en annan sorts ekonomi. Man ”betalar” för support genom att investera i att ta reda på var den finns, vilka versioner som är de senaste och på vilket sätt den kan användas.¹⁷⁰ Under senare år har alltfler företag börjat specialisera sig på att sälja

¹⁶⁷ Se <www.creativecommons.se/licenser.html>.

¹⁶⁸ All som deltagit i någon e-post lista har säker någon gång reflekterat över graden av aktivitet och engagemang hos vissa deltagare som ofta (dagligen) svarar på mer eller mindre intelligenta frågor och som ofta levererar åsikter i många olika frågor.

¹⁶⁹ Statskontoret, *Öppen programvara*.

¹⁷⁰ Vid webbplatser som <<http://sourceforge.net/>> finns inbyggda funktioner för versionskontroll av program.

support och kommersiella förpackningar av öppen källkod, som exempelvis operativsystemet RedHat.¹⁷¹

Som jag tidigare nämnt är det inte omöjligt att en låg användarkompetens kan vara en orsak till den ringa användningen av öppna standarder och öppen programvara inom många områden av samhällsvetenskaplig och humanistisk forskning och undervisning. Andra orsaker kan vara vanans makt. Det är svårt att bryta invanda mönster, en slutsats som Statskontoret presenterar i sin användarstudie av öppen programvara.¹⁷²

2.3.2 Öppna standarder

Relaterat till definitionen av öppen källkod är även öppna standarder, det vill säga regelverk och definitioner av filformat och innehåll för informationslagring och informationsutbyte. Med öppna standarder avser jag sådana standarder där protokoll och dokumentation är offentliga och där regelverket och riktlinjerna är allmänt tillgängliga. Jag använder termen standard i samma vida mening som den definition som tagits fram av EU:s organ för interoperabilitet mellan statliga tjänster och organisationer och den beskrivs närmare nedan.¹⁷³

I Statskontorets rapport framhålls att öppna standarder dessutom innebär att de produceras i en ”allmänt accepterad och öppen miljö”. Därtill att standarden utvecklas i en konsensusmodell och inte av en enskild intressent.¹⁷⁴ Exempel på öppna standarder och filformat är just XML och DocBook (se nedan) – motsatser till proprietära filformat såsom *Microsofts Word*.

Användningen av termen standard i detta sammanhang kan avse en standard fastslagen av ett standardiseringsorgan, det vill säga *de jure*. I praktiken avses ofta *de facto* standarder eller rekommendationer, det vill säga sådana som inte är fastställda av något officiellt standardiseringsorgan utan produceras inom etablerade intressegemenskaper och som har en bred användning. *Text Encoding Initiative*, TEI:s, riktlinjer är ett exempel på en *de facto* standard för beskrivning, lagring och utbyte av samhällsvetenskapligt och humanistiskt digitalt material. Även om man inom TEI-konsortiet ursprungligen avstod från att skapa och tillhandahålla en standard, utan istället tog beslut om att skapa en samling riktlinjer för litterära och lingvistiska texter, har TEI:s riktlinjer nått sådan spridning att det är relevant att tala om en *de facto* standard.¹⁷⁵

Bruce Perens, som tillsammans med Eric Raymond formulerade principerna i definitionen för öppen källkod, har även formulerat sex principer för öppna

¹⁷¹ RedHat är en distribution under licensen GNU/Linux och har delvis anpassats för att locka användare från Microsoft Windows, se <www.redhat.com/>.

¹⁷² Statskontoret, *Öppen programvara. Erfarenheter av produkter som bygger på öppen källkod inom förvaltningen* (Stockholm: Statskontoret, 2004).

¹⁷³ Punkterna något förkortade från: *European Interoperability Framework for pan-European eGovernment Services* (Luxembourg: Office for Official Publications of the European Communities, 2004) s. 9.

¹⁷⁴ Statskontoret, *Öppen programvara*.

¹⁷⁵ Nancy Ide & C. M. Sperberg-McQueen, ”The TEI: History, Goals, and Future” *Text encoding initiative: background and context* (Dordrecht; London: Kluwer Academic, 1995).

standarder som till stora delar liknar OSI:s definition av öppen källkod. Perens' principer handlar om fri tillgänglighet, frihet för användare, frihet för val av användningsområden och program. Definitionen anger i de två sista punkterna att tillägg och utvidgning är tillåtna, dessa måste dock publiceras och inte vara begränsade att gälla enbart för en viss tillämpning.¹⁷⁶

I EU:s definition av öppna standarder fastslås i den första punkten att standarden skall ägas och underhållas av en icke-kommersiell organisation samt att utveckling skall ske offentligt och med en genomskinlig beslutsprocess. Nästa punkt fastställer att standarden skall publiceras offentligt och att specifikationen ges ut gratis eller till självkostnadspris samt att den skall kunna kopieras, spridas vidare och användas. Vidare framhålls att immateriell egendom¹⁷⁷ gällande standarden nu och för evigt hålls tillgängliga utan avgift. Därtill framhåller EU att det inte skall finnas några begränsningar för återanvändning av standarden.¹⁷⁸

Den största skillnaden mellan Perens och EU:s definition består i att EU-policyn framhåller att standarden skall handhas av icke-kommersiella organisationer samt vara offentlig och gratis tillgänglig, något som inte uttrycks explicit i de principer som Perens formulerat. Inom EU används förstås frågor om standardisering som policy medan man inom öppen källkods rörelsen och rörelsen för fria program har ideologiska utgångspunkter, FSF mer uttalat ideologiska och OSI mer pragmatiska.

World Wide Web Consortium, W3C, är ett annat exempel på ett intresseorgan som producerar och förvaltar många riktlinjer och *de facto* standarder för webben, däribland specifikationerna för HTML och XML. W3C bildades 1994 och är ett konsortium med drygt 500 medlemmar från forsknings- och utvecklingsområdet samt industrin.¹⁷⁹ *World Wide Web Consortium* tillämpar en arbetsordning för standardiseringsarbetet som följer fyra steg.¹⁸⁰

1. Arbetsdokument som föreslår en eventuell standard (*Working draft, WD*).
2. Förslag till specifikation som skickas ut för granskning i olika arbetsgrupper (*Candidate recommendation, CR*).
3. Förslag till standard som har stöd och förankring och publiceras för granskning (*Proposed recommendation, PR*).
4. Slutlig rekommendation godkänd och antagen av W3C:s medlemsorganisationer (*Recommendation, REC*).

¹⁷⁶ Se *Open Standards. Principles and Practice*, <<http://perens.com/OpenStandards/>>.

¹⁷⁷ Som till exempel patent, varumärken, upphovsrätter.

¹⁷⁸ *European Interoperability Framework for pan-European eGovernment Services* s. 9. <<http://ec.europa.eu/idabc/servlets/Doc?id=19528>>.

¹⁷⁹ Se <www.w3.org/Consortium/>.

¹⁸⁰ Se <www.w3.org/2005/10/Process-20051014/tr.html#q73>.

Denna arbetsprocess premierar öppna beslutsgångar och väl dokumenterade projekt något som även genererar en ansevärd mängd dokument.¹⁸¹ Förutom specifikationer på standarder publicerar W3C en mängd tekniska rapporter och kommentarer och webbplatsen är en viktig referens och kunskapskälla vid användning av webbt tekniker.

I avhandlingen används eller berörs flera standarder och riktlinjer utarbetade av W3C, den viktigaste i detta sammanhang är XML, en standard och en arkitektur eller snarare en verktygslåda.¹⁸² Som namnet, *Extensible Markup Language* antyder är XML tänkt att vara utbyggbart. XML som blev en standard 1998 är som tidigare nämnts härlett ur SGML. Det togs från början fram för just digital publicering på webben men används i stor omfattning för utbyte av olika sorters data.¹⁸³ En central person i detta sammanhang är James Clark som programmerat en rad viktiga verktyg och tillämpningar för SGML och XML, bland annat standarden och programmeringsspråket *Document Style Semantics and Specification Language* (DSSSL) för transformering av SGML samt *Extensible Stylesheet Language and Transformation* (XSLT) för transformering av XML, den senare av dem används i den här studien.¹⁸⁴

Världens största standardiseringsorgan heter *International Organization for Standardization* (ISO) och består av 154 medlemsländer. ISO har sedan 1947 publicerat över 16000 internationella standarder inom skilda verksamhetsområden.¹⁸⁵ Cirka hälften av alla standarder är inom mekanik- och elektronikteknik (*Engineering technology*) och materialteknik (*Materials Technology*), en sjättedel inom informationsteknik (*Electronics, information technology and telecommunications*) och en tiondel av standarderna avser forskning, infrastruktur och administration (*Generalities, infrastructures and sciences*). En av de mer allmänt kända standarderna är ISO 9000 om kvalitetsstyrning. SGML blev en ISO standard 1986 och den ligger även till grund för en rad andra ISO standarder. Exempelvis SGML-tillämpningen HyTime som anger principer för länkning mellan SGML dokument¹⁸⁶ samt Topic Maps som är en standard för beskrivning av relationer mellan olika resurser och begrepp.¹⁸⁷

Standardiseringsarbetet sker förstås på olika nivåer och i Sverige är det *Svensk Standardiseringsinstitut* (SIS) som utarbetar standarder inom de flesta samhällsområden.¹⁸⁸ Standardiseringsorganet för Europa som heter *European Committee for*

¹⁸¹ Man kan med visst belägg även hävda att W3C:s webbplats är svårnavigerad och att dokumentation ibland är alltför excessiv.

¹⁸² I del 2 tillämpas flera av verktygen som tillhör XML-arkitekturen: XPath, XSLT, XPointer, XLink m fl.

¹⁸³ <www.w3.org/XML/>.

¹⁸⁴ Hur XSLT kan användas för transformering exemplifieras i avhandlingen bland annat i avsnitt 6.4.2 s. 193. DSSSL är en ISO-standard ISO/IEC 10179:1996, se <www.jclark.com/dsssl/>. James Clark är redaktör för W3C:s XSLT rekommendationen, <www.w3.org/TR/xslt>.

¹⁸⁵ <www.iso.org/>.

¹⁸⁶ Hypermedia/Time-based Structuring Language, ISO/IEC 10744:1997, (2:a utgåvan).

¹⁸⁷ Topic maps ISO/IEC 13250:2003, ISO standard för Topic Maps data model publicerades 2006 och Topic Maps i XML syntax version 2.0 publicerades 2007.

¹⁸⁸ SIS är en ideell förening med medlemmar privat och offentlig sektor. <www.sis.se/>.

Standardization/Comité Européen de Normalisation, CEN, en paraplyorganisation för de nationella standardiseringsorganen inom EU.¹⁸⁹

Konsortiet *Organization for the Advancement of Structured Information Standards* (OASIS) är ett annat standardiseringsorgan som stödjer utveckling och tillämpning av öppna standarder.¹⁹⁰ OASIS har bland annat arbetat fram den öppna standarden och filformatet Open Document (ODF) som bygger på XML och som används av Open Office.¹⁹¹ Förutom det utvecklar och underhåller OASIS Doc Book, en standard för digital publicering med XML, men även SGML, av olika slags böcker och tekniska manualer.¹⁹²

2.3.3 Avslutning

Utan standarder och gemensamma riktlinjer skulle digitalisering, lagring och utbyte av material bli komplicerat. Forskare, lärare och studenter skulle bli hänvisade till proprietära lösningar där innehållet kanske inte kan återanvändas eller utvecklas när de tekniska lösningarna blivit ålderdomliga. Ett problem med standarder är att de tvunget är resultat av förhandlingar och kompromisser samt kräver tillräckligt starka incitament för att verkligen användas. Bara en sådan sak som datum, visst vore det enklare om datum beskrevs på ett konsekvent sätt? Nu finns det visserligen en ISO standard för notation av datum men långt ifrån alla aktörer eller länder använder sig av den.¹⁹³

Man kan argumentera för anledningar till att använda XML standarden för beskrivning och lagring av digitalt material, såsom interoperabilitet och ett plattformsoberoende filformat. Anledningar till att inte använda XML är sådana argument som att det tar för lång tid att lära sig de verktyg som behövs eller att det är enklare att fortsätta arbeta med de verktyg man är van vid, såsom MS Word.

Genom att använda öppna standarder för beskrivning och lagring av innehåll i något plattformsoberoende format¹⁹⁴ ges möjligheter att beskriva och lagra digitalt material på ett stringent sätt som samtidigt tillåter framtida utbyte och migrering av innehållet. Material som produceras för forskning och undervisning skulle på så sätt kunna användas om säg 20-30 år tillsammans med de metoder och verktyg som då kommer att finnas tillgängliga. Utvecklingen inom öppen källkod och öppna standarder är viktig för såväl forskning som utbild-

¹⁸⁹ <www.cen.eu>

¹⁹⁰ OASIS hette från början SGML Open när de bildades 1993, men bytte namn 1998 för att bättre spegla hela verksamhetsområdet och inte minst arbetet med XML, se <www.oasis-open.org/> samt en omfattande webbportal för tillämpningar av XML och information om metoder och verktyg för digitalisering, öppna standarder etc, <<http://xml.coverpages.org/>>.

¹⁹¹ ISO/IEC 26300:2006, Open Document Format for Office Applications.

¹⁹² För information om DocBook se <www.docbook.org/>.

¹⁹³ ISO 8601:2004 (Data elements and interchange formats) för konsekvent datumnotation (YYYY-MM-DD) av tidsräkning i den Gregorianska kalendern.

¹⁹⁴ Vanliga textfiler är plattformsoberoende och då XML lagras i textformat, så kommer XML-formatet att kunna läsas och användas på olika plattformar och av olika verktyg även i framtiden.

ning inte minst för att det ger alternativ till proprietära plattformar men även på grund av de ofta relativt sett begränsade ekonomiska resurserna inom offentlig sektor. Dessutom är verksamheten inom öppen källkod och öppna standarder ett resultat av utvecklingsarbete inom akademien och är därmed, som jag ovan visat, även utsatt för intern såväl som extern insyn och granskning.

I detta kapitel har jag beskrivit utgångspunkterna för innehållsdesign och förutsättningar för att utforma innehåll där vare sig plattform eller specifika framtida användningsområden behöver förutses eller föregripas. Därtill har utvecklingen av öppen källkod och öppna standarder diskuterats. Denna utveckling och verksamhet ligger till grund för de flesta av de metoder och verktyg som nedan kommer att beskrivas och/eller användas.

I de två följande kapitlen presenteras avhandlingens teoretiska utgångspunkter genom att beskriva och analysera de principer, metoder och verktyg som kan användas för innehållsdesign.

3 Principer och metoder för innehållsdesign

Detta kapitel ägnas åt att identifiera och avgränsa de metoder och verktyg som kan vara lämpliga för innehållsdesign av prosopografiska material och inledningsvis presenteras därför en historisk översikt av märkspråkens utveckling. Den historiska översikten följs av en översikt över den beskrivande märkningens principer och dess teori- och begreppsbildning som avslutas med en översikt av föregångarna till dagens märkspråk. Därefter följer beskrivningar av de märkspråk, metoder och verktyg som används för innehållsdesign och i denna framställning läggs den största tonvikten på verktygslådor för webben.

3.1 Inledning

Som tidigare nämnts har studien en pragmatisk utgångspunkt där principer och metoder för beskrivning av innehåll utforskas och analyseras. Det jag valt att benämna textorienterade metoder för innehållsdesign innebär att själva beskrivningen, ordnandet och bearbetningen utgår från textens struktur och det semantiska innehållet. För beskrivning av innehållet används beskrivande märkspråk som ger möjlighet att kombinera innehållet med explicit om information innehållets olika egenskaper.¹⁹⁵ Det kan vara beskrivningar av textens logiska struktur, fysiska egenskaper eller det semantiska innehållet och ibland även instruktioner om innehållets presentation. Huvudsakligen används dock beskrivande märkspråk för att särskilja innehållet från dess form.¹⁹⁶

Den extra information som tillförs innehållet benämns märkning (*markup*). Märkspråk utgör förutsättningarna för flera av de verktyg jag använder i den här studien och dessa har sitt ursprung i metaspråket, SGML. Utvecklingen av märkspråken har, tillsammans med datorteknikens utveckling, varit viktig för utvecklingen inom strukturerad informationshantering och innehållsmärkning. För inventeringen av de textorienterade metoderna kan det vara värdefullt att få en inblick i praktiska och teoretiska bakgrunder till dagens metoder och verktyg för vetenskaplig litterär utgivning, avancerade metoder för textanalys, digital publicering och digitala arkiv. Jag inleder därför med en kort beskrivning av

¹⁹⁵ I avhandlingen används beskrivande och deskriptiva märkspråk synonymt.

¹⁹⁶ Oavsett om det semantiska innehållet egenskaper eller en texts fysiska egenskaper beskrivs kan ändå den beskrivande märkningen av innehållet särskiljas från beskrivningen av dess form, genom att den explicita (meta)information som infogas identifierar just de egenskaper som avses beskrivas.

märkspråkens historia och praktik. Denna praktik är till stora delar driven av konkreta forskningsproblem och experimentellt arbete. Susan Hockey beskriver detta arbete och samtidigt syftet med märkning på följande sätt:

Syftet med kodning [att infoga märkning] inuti en text är att tillhandahålla information som kan hjälpa ett datorprogram att utföra arbete med denna text.¹⁹⁷

De pragmatiska utgångspunkterna har inte hindrat att tvärvetenskapliga teorier om märkning, märkspråk och elektroniska texter har växt fram parallellt med praktiken under årens lopp, ibland som resultat av strider och konflikter mellan olika intressegrupper och discipliner.¹⁹⁸

Dagens märkspråk härstammar från de traditionella praktiker för märkning i marginalen av manuskript som gjordes inför sättningen av en bok. Under århundraden utfördes detta arbete av en yrkeskår med specialister som utformade bokens komposition, avgjorde språkliga korrigeringar samt vilka typer som skulle användas för tryckning. För att undvika missförstånd användes, och används, specifika korrekturtecken enligt ett väl utarbetat system som innehåller teckengruppering för lokalisering, åtgärder och återtagande.

Det tryckta ordet har som vi alla vet inte alltid varit så självklart, utan mycket har hänt sedan tryckerikonstens början. Förändringarna har varit många och nya metoder och uppfinningar har under seklernas gång tillkommit. Skrivandets teknik som ett intellektuellt verktyg är en relativt sen uppfinning.¹⁹⁹ Med bakgrund i detta kan vi således betrakta elektronisk publicering och digital utgivning som en del i denna utveckling i likhet med övergången från produktion av handskrifter i skriptorium till böcker producerade i tryckeriet.

Roger Chartier, fransk bokhistoriker, skriver i sin bok om böcker, läsare och bibliotek under 1300- till 1700-talen, om de bortglömda rummen, mellan texten och objektet, där meningen konstrueras och materialiseras, om hur typografi och komposition i stor utsträckning bidrar till texten. Boken, menar Chartier, kan inte ses som en abstrakt text som existerar oberoende av den praktik där de blivit till eller där de används:

Författarna skriver inte böcker. Nej, de skriver texter som blir skrivna objekt, handskrivna, graverade eller tryckta (och idag databehandlade).²⁰⁰

¹⁹⁷ Min översättning av "The purpose of encoding within a text is to provide information which will assist a computer program to perform functions on that text.", i Hockey, *Electronic texts in the humanities: principles and practice* s. 24.

¹⁹⁸ Mer om detta i kapitlet om TEI s. 102.

¹⁹⁹ Walter J. Ong & Lars Fyhr, *Muntlig och skriftlig kultur: teknologiseringen av ordet*, 2:a uppl. ed. (Göteborg: Anthropos, 1991). s. 100 f.

²⁰⁰ Chartier, *Böckernas ordning: läsare, författare och bibliotek i Europa mellan 1300-tal och 1700-tal*. s. 21.

Chartier framhåller här texternas materiella aspekter och det faktum att texter inte existerar oberoende av det medium eller de förhållanden som medverkar till texternas reception och användningssätt.

I Adrian Johns' omfattande studie om böcker, publicering och naturvetenskaplig kunskapspraktik under 1500-, 1600- och 1700-talen i Storbritannien²⁰¹ kan vi läsa om de professionella praktiker och om de yrkesmän som medverkade till dåtidens moderna kunskapsproduktion: *pressmen, compositors & correctors*. Deras verksamhet jämfördes ibland med arkitektkonsten och i ett samtida verk om tryckningens konst framläggs bland annat uppfattningen att typernas formgivning skall utgå från geometrin.²⁰² Johns driver också tesen att de kunskaper som utvecklades i tryckerierna hade långt större implikationer för vetenskapliga kunskapers spridning i Europa under dessa århundraden än vad som framkommit i tidigare materialbibliografiska studier. Dessa studier har inte, till skillnad från Johns, studerat tryckningens praktik med utgångspunkt i själva tryckerikonsten, genom dess egen praktik.²⁰³ Tryckerikonstens funktion har dock sedan digitaliseringen av publiceringsprocessen radikalt förändrats och ett nytt möjligt arbetsfält för till exempel typografer är istället professionell digital märkning och design av texter.²⁰⁴

Inom textsociologin utgår man alltså från att text är något mer än dess abstrakta och immateriella bemärkelse och beaktar även materiella förutsättningar och grafiska egenskaper. Textsociologen Don McKenzie framhåller att "former påverkar betydelser".²⁰⁵ Även Chartier uppmanar oss att ha formens betydelse i åtanke och framhåller felaktigheten i att reducera texter till enbart deras semantiska innehåll då de materiella betingelserna och själva utformningen har en stor betydelse för hur och i vilken praktik texternas sedan används.²⁰⁶

Även om syftet med märkspråk är att särskilja innehåll från dess form, betyder inte det att formen fråntas sin betydelse för reception, användningar eller betydelser. Tvärtom, jag hoppas att jag kan visa hur form och framställning påverkar innehållets betydelser. Inte heller kan de materiella förutsättningarna för innehållets användningar förbises. Och som Mats Dahlström poängterar medför mediet – bok, webb eller annan digital form – att text och innehåll lever under olika förutsättningar då mediernas egen logik kan förstärka vissa av innehållets egenskaper och utesluta andra.²⁰⁷ De materiella förutsättningarna för

²⁰¹ Adrian Johns, *The nature of the book: print and knowledge in the making* (Chicago: Univ. of Chicago Press, 1998). ssk. s. 74-108.

²⁰² Här avses tryckerimästarens Joseph Moxons omfattande verk (*printing master*) på över 400 sidor om tryckerikonsten: *Mechanik Exercises on the Whole Art of Printitng*, 1683-84, som Johns i detalj studerat och ofta hänvisar till i kapitlet "Literatory Life" i Ibid. ssk. s. 79-105.

²⁰³ Ibid. s. 1-5.

²⁰⁴ Se t.ex. John B. Thompson, *Books in the digital age: the transformation of academic and higher education publishing in Britain and the United States* (Cambridge: Polity, 2005), en studie om digitalisering i och transformering av fältet av akademisk publicering i Storbritannien och USA.

²⁰⁵ "Forms effect meanings" Donald McKenzie, *Bibliography and the sociology of texts* (Cambridge: Cambridge University Press, 1999) s. 4.

²⁰⁶ Chartier, *Böckernas ordning: läsare, författare och bibliotek i Europa mellan 1300-tal och 1700-tal*.

²⁰⁷ Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion*. s. 73 f.

innehållsdesign kan förstås nyttjas som en möjlighet. I den här studien handlar det bland annat om att nyttja datorernas förmåga att systematiskt, och dessutom utan att tröttna, bearbeta stora mängder innehåll exakt på det sätt vi vill eller att manipulera och analysera innehåll på ett sätt som inte är möjligt i tryckt text.

3.2 Märkspråkens utveckling

Jag kommer i detta avsnitt att koncentrera min framställning på de specifika steg i de textorienterade märkspråkens utveckling som varit av betydelse för de metoder och de verktyg som används i min studie. Detta innebär att framställningen i vissa avseenden haltar och att jag gör några logiska hopp mellan olika aspekter av informationshantering, märkning och principer för ordnande av innehåll. De läsare som är intresserade av detaljerade och tekniska principer, beskrivningar och riktlinjer om text- och dokumenthantering hänvisas framför allt till de referenser som tas upp i kapitlet.²⁰⁸

Märkning av text går tillbaka till begynnelsen av den digitala textteknikens dagar i början av 1960-talet. Som ovan nämnts härstammar märkning från typografin och i början märktes därför främst innehållets typografiska egenskaper. Dåtidens texthantering använde sig av olika kommandofiler,²⁰⁹ där en fil innehöll koder med datorspecifika kommandon för formatering av innehållet och en annan fil innehöll koder för att representera själva innehållet, det vill säga bokstäver och punktering. Denna fil behandlades sedan med ett formateringsprogram som genererade ytterligare en fil som innehöll både den formaterade texten och instruktioner för utskrift.²¹⁰

Vid en övergång från gamla till nya tekniker är det inte helt ovanligt att man i början imiterar det äldre tillvägagångssättet och den typografiska märkningen är ett bra exempel på detta. Då anges helt enkelt regler för önskad styckesindelning, vilka ord som skall accentueras med fet eller kursiv stil etcetera, allt i syfte att förstärka specifika egenskaper och därmed öka läsbarheten av texten. Alla som läst text producerad av äldre datorer vet vilken skillnad layouten kan göra.²¹¹

²⁰⁸ För den historiskt intresserade se TEI:s bibliografi för P4 (Appendix A) i Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange*. tillgänglig vid: <www.tei-c.org/P4X/P2BIB.html>. Se även antologin Susan Schreibman, Raymond George Siemens & John Unsworth, red, *A companion to digital humanities* (Malden, MA; Oxford: Blackwell Pub., 2004).

²⁰⁹ Filer som innehåller en rad kommandon kallas för batch-filer i Windows och shell-skript i UNIX.

²¹⁰ Introduktioner till märkspråkens utveckling ges bl.a. i Hockey, *Electronic texts in the humanities: principles and practice*. Samt i Allen H. Renear, "Out of Praxis: Three (Meta)Theories of Textuality" *Electronic text: investigations in method and theory*, red. Kathryn Sutherland, (Oxford: Clarendon, 1997).

²¹¹ Tidiga elektroniska texter består endast av versaler, något som i hög grad motverkar läsbarheten.

Innebörden av en typografisk märkning är ofta tvetydig; kursiv stil kan vara avsedd att beteckna antingen främmande ord, titlar, scenanvisningar eller något annat. Problemet med typografisk märkning är därför svårigheter att skapa förutsättningar för textanalyser, såsom att skapa listor av alla titlar eller indexera främmande ord med hjälp av automatiserade metoder (datorer).²¹² Typografiskt märkta texter kan dock förberedas för analys genom att migreras till ett annat och lämpligare filformat som möjliggör strukturell märkning. Det är möjligt att nyttja de typografiska egenskaperna för att semi-automatiskt tillföra enklare strukturell märkning ett dokument i textformat.²¹³

Den första typen av märkning var procedurell märkning anger hur dokumentets struktur skall bearbetas och presenteras. Märkningen tolkas och innehållet bearbetas av applikationen i den ordning den framträder. Nedan presenteras ett exempel på det procedurella märkspråket LaTeX,²¹⁴ och där märkning, eller snarare formateringskommandon, infogas i texten för information om hur storlek, titel, rubriker och brödtext skall skapas.

```
\documentclass[a4paper]{article}

\title{LaTeX exempel}
\author{Monica Langerth Zetterman}
\date{30 oktober 2006}

\begin{document}

\maketitle                                % Skriver ut rubriken som
                                           % deklarerades ovan med \title, \author
                                           % och eventuellt \date

\section{Vanlig Text}                     % Detta kommando gör en rubrik
```

Figur 6. Exempel på procedurell märkning med LaTeX

Nackdelen med den procedurella märkningen var dels att den inte beskriver dokumentens logiska struktur, dels att den inte var generisk vilket medförde att den infogade märkningen inte kunde återanvändas då den inte var giltig för olika typer av dokument.²¹⁵

²¹² Hockey, *Electronic texts in the humanities: principles and practice*.

²¹³ I en pilotstudie utförd vid Oxford Text Archive visas hur typografisk märkning kan, om den är konsekvent utförd, användas för migrering av textdokument till dokument med enklare XML-märkning. Se vidare James Cummings & Monica Langerth Zetterman, "Legacy Data Migration: A pilot study on the methodological feasibility of conversion and enhancement of electronic resources." *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities*. (Gothenburg, Sweden: 2004).

²¹⁴ Se LaTeX projektets webb, <www.latex-project.org/>.

²¹⁵ För en mycket intressant och grundlig genomgång samt diskussioner om 1980-talets state-of-the-art teknik om strukturerad dokumenthantering och publicering, se antologin Jacques André, et al., *Structured documents*, The Cambridge series on electronic publishing, (Cambridge: Cambridge University Press, 1989).

Nästa steg i utvecklingen var program utvecklade för textfångst (*text retrieval*)²¹⁶ som kunde tolka markeringar och i texten infogade tecken (*embedded markup*) för att underlätta för programmet såväl som användaren att identifiera träffar av de termer som eftersöktes. Dessa enklare märkningsscheman var specifika för varje program, användare och dokument och definierades i början av varje dokument. Programmen användes för att till exempel hitta alla dokument som innehåller en viss term och producerade sedan en lista på antalet träffar. Problemet är att avgränsa vad som utgör ett dokument. Utgör ett dokument en hel roman eller ett kapitel, hela pjäsen eller en akt?²¹⁷ Beroende på hur innehållet är strukturerat och ordnat blir resultatet av en sådan analys inte jämförbart mellan olika genrer eller olika elektroniska versioner av samma drama. Metoden var alltså inte särskilt användbar för mer systematiskt litterära och lingvistiska textanalyser och därför utvecklades sedermera mer sofistikerade system för märkning och bearbetning av text. Tack vare de tidiga textfångst-programmen utvecklades dock alltmer sofistikerade metoder för att räkna rader, och referera till specifika ord eller passager i texten. I slutet av 1960 talet kom ett av de första egentliga deskriptiva märkspråken, COCOA, som användes för lingvistiska, stilistiska och litterära studier.²¹⁸

Jag har tidigare introducerat de grundläggande principerna för deskriptiva, eller beskrivande, märkningen som är den dominerande metod som används idag för att beskriva och ordna innehåll utan att för den skull indikera hur innehållet sedan skall formateras eller bearbetas. Märkspråken kan även beskrivas på olika sätt. I flera sammanhang används de två kategorierna ”procedurrell” och ”deskriptiv” märkning för att belysa skillnaden mellan märkspråk och TEI använder deskriptiv i sina riktlinjer. Den deskriptiva märkningen har sitt ursprung i SGML och termen introducerades av Goldfarb – SGML:s upphovsman.²¹⁹ Ibland även benämnd generisk märkning (*generalized markup*) för att markera skillnaden gentemot procedurrell märkning. I beskrivande märkspråk infogas alltså inte instruktionerna om kommande bearbetning eller presentation tillsammans med märkningen i texten, och denna skillnad är väsentlig:

With generalized markup, the markup process stops at the first step: the user locates each significant element of the document and marks it with the mnemonic name (”generic identifier”) that he feels best characterizes it.²²⁰

²¹⁶ Här används textfångst som benämning för uppgiften att inhämta den information i form av dokument, eller delar av dokument, som motsvarar vissa specificerade aspekter.

²¹⁷ Se Hockey, *Electronic texts in the humanities: principles and practice*. s. 26 för exempel på studier med olika implementering av denna metod.

²¹⁸ *Count and Concordance Generation on Atlas (COCO A)*.

²¹⁹ Jämförelsen används bland annat i TEI:s riktlinjer i andra introduktionskapitlet som handlar om märkspråken SGML och XML, se <www.tei-c.org/P4X/SG.html>.

²²⁰ Goldfarb & Rubinsky, *The SGML handbook* s. 8.

För att styra presentationen används istället instruktioner för detta i efterhand, i form av stilmallar eller skript, som medför att samma innehåll kan bearbetas och renderas på olika sätt.²²¹ Man skiljer helt enkelt på innehåll och form och istället kan man fokusera på att beskriva innehållets struktur, logik och semantiska innehållsobjekt. SGML utvecklades just för denna typ av funktionalitet. I praktiken frångås dock emellanåt denna princip. HTML är till exempel en blandning av rent beskrivande märkning (exempelvis <a> för att markera plats för en hänvisning) och märkning för presentation (exempelvis <i> för markering av kursiv stil). Jämfört med HTML är XML:s riktlinjer mer inriktade på möjligheter till utvidgning, strikta syntaxregler²²² och en inbyggd logik som fokuserar på det semantiska innehållet.

Vid Oxford Text Archives gav man för några år sedan ut en skrift med goda råd om hur innehållsproducenter kan gå tillväga.²²³ Där klassificeras märkspråken enligt indelningen märkning för formatering, proprietär märkning (t.ex. Word och PDF) samt icke proprietära generella märkspråk (SGML, XML). Den senare kategorin innehåller mest utrymme för kontroll i och med att dessa märkspråk är plattformsberoende. Dessutom tillåter denna typ märkning av det semantiska innehållet snarare än formatet, det vill säga en annan typ av kontroll än de övriga två typerna av märkspråk.²²⁴

Till skillnad från många andra översikter om märkspråkens utveckling använder Susan Hockey inte i första hand den deskriptiva/procedurella uppdelningen i sin historiska genomgång av textmärkning. Den nämns dock i slutet av den historiska översikten. Susan Hockey men även Allen Renear strukturerar istället beskrivningen genom märkspråkens kronologiska utveckling och funktionalitet snarare än med utgångspunkt i dikotomin procedurella och deskriptiva märkspråk.

Dikotomin procedurell och deskriptiv har ibland kritiserats, bland annat av Allan Renear²²⁵ och Wendel Piez²²⁶ som anser att den är förlegad och tvetydig. Renear menar att det i princip är riktigt att deskriptiv märkning avser identifiera

²²¹ Stilmall är ibland en dålig översättning av engelskans *styelsheet*, speciellt för skriptspråket XSLT som används för avancerade transformeringar och bearbetningar av XML-filer och som inte alls enbart formaterar XML-filer för presentation på webben eller papper. Däremot är det mer passande att använda termen stilmall för CSS-stilmallar då CSS beskriver hur strukturerat dokument skall presenteras på webben, med kapitel, stycken, listor etc.

²²² Om märkningen inte följer den föreskrivna XML syntaxen, det vill säga inte är välformulerat (*well formed*) fungerar inte verktygen för bearbetning och presentation. På så sätt "tvingas" innehållsproducenter att hålla sig till föreskriven syntax.

²²³ Alan Morrison, Michael Popham & Karen Wikander, *Creating and Documenting Electronic Texts: A Guide to Good Practice* (Oxford: Oxford Text Archive, 2000).

²²⁴ Ibid. ssk. kapitel 4.

²²⁵ Se Allen H. Renear, "The Descriptive/Procedural Distinction is Flawed", *Markup Languages: Theory and Practice* 2.4 (2000); Allen H. Renear, "Text from Several Different Perspectives, the Role of Context in Markup Semantics". *Proceedings of the 2003 Conference on Computers, Literature, and Philology*, Florence: University of Florence, 2003.

²²⁶ Se Wendell Piez. "Beyond the "descriptive vs procedural" distinction". *Extreme Markup Languages* 2001, 2001.

objekt som tillhör en specifik domän och att procedurell märkning påbjuder en viss sorts representation av ett objekt, men gränsen mellan vad som kan ses som identifikation och vad som är en påbjudande akt är otydlig. Beroende på kontexten och därmed den möjliga tolkningshorisonten kan ett objekt också uppfattas på olika sätt. Renear menar att även om terminologin som kontrasterar deskriptiv och procedurell märkning varit användbar för att definiera och teoretisera märkning och märkspråk *per se*, bör en beskrivning av märkspråken själva och dess funktioner i större utsträckning beaktas relativt användningens sammanhang.²²⁷

3.3 Innehållets ordning

Ovan har de deskriptiva märkspråken introducerats. Denna metod är förstås inte given, det finns även andra sätt att beskriva och kategorisera innehåll, något jag tar upp i ett senare avsnitt om alternativ till de hierarkiska märkspråken.

Idén med deskriptiva märkspråk är som sagt att ett godtyckligt innehåll alltid har en struktur som kan beskrivas enligt logisk hierarkisk ordning.²²⁸ Den modell som används här är som tidigare nämnts den hierarkiska dokumentmodellen som är utgångspunkt för de metaspråk som i sin tur ligger till grund för de beskrivande märkspråken.²²⁹ En hierarkisk dokumentmodell reflekterar den lingvistiska uppfattningen att text består av ordnade innehållsobjekt. I den hierarkiska dokumentmodellen är själva innehållsobjekten av primär betydelse. De kan vara av olika slag: ord, meningar, stycken, citat eller semantiskt avgränsade objekt såsom beteckningar för yrkestitel, utbildningsnivå eller verksamhetsområde. Samma uppsättning innehållsobjekt kan framställas och formas på olika sätt – något som också påverkar betydelsen.²³⁰ De flesta innehållsobjekten ingår i större objekt, en mening är en del av ett stycke som är en del av ett avsnitt etcetera. Denna modell representerar ett synsätt på text som ett medel för överföring av information, bestående av lingvistiska sekvenser som är hierarkiskt ordnade.²³¹

I ett beskrivande märkspråk tas ingen hänsyn de olika komponenternas semantiska innehåll eller hur innehållet sedan ska presenteras utan avsikten är att:

²²⁷ Renear. "Text from Several Different Perspectives, the Role of Context in Markup Semantics".

²²⁸ I inledningen till SGML:s riktlinjer anges i och för sig att SGML är ett databasspråk för text men även att det är en notation för alla typer av strukturer. Goldfarb & Rubinsky, *The SGML handbook* s. xiii.

²²⁹ Min användning av: Ordered Hierarchy of Content Objects, OHCO, se även s. 56 f. i avhandlingen. Dokument används inte för beskrivning av materiella dokument, utan som ett begrepp för att avgränsa en viss mängd information, i detta sammanhang vanligen text. Begreppet dokument och dokumentets materiella betingelser studeras inom biblioteks- och informationsvetenskapen, se t.ex. Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 2 f.

²³⁰ DeRose, et al., "What is Text, Really?"

²³¹ Renear, "Out of Praxis: Three (Meta)Theories of Textuality". Kritik mot den hierarkiska dokumentmodellen, se kapitel 4.4 s. 113 f.

identifiera de komponenter, eller innehållsobjekt som utgör en helhet eller åtminstone en större mängd innehåll,
fastställa hur dessa komponenter relaterar till varandra genom en hierarkisk ordning vilket innebär att bestämma var en viss komponent kan befinna sig i förhållande till andra komponenter,
fastställa hur dessa komponenter skall markeras

Ett dokument är alltid impregnerat av implicit strukturell och semantisk märkning genom punktering, styckesindelning och andra former för att särskilt framhäva betydelser eller strukturer i texten.²³² Detta är en effekt av de bärande medierna såväl som sociala konventioner för att underlätta för läsare att se var ord och meningar börjar och slutar.²³³ Utmärkande för märkspråksmetoden är att tolkningen av den logiska hierarkiska strukturen explicit kan beskrivas istället för att vara underförstådd och endast ett hjälpmedel för presentation.

Espen Aarseth presenterar i sin studie om ergodisk litteratur²³⁴ en modell som skiljer sig från den hierarkiska dokumentmodellen genom att han ger de materiella förutsättningarna och de funktionella effekterna stor betydelse i sin pragmatiska och breda definition av text:

At text, then, is any object with the primary function to relay verbal information. Two observations follow from this definition: 1) a text cannot operate independently of some material medium, and this influences its behavior; 2) a text is not equal to the information it transmits.²³⁵

I definitionen särskiljer Aarseth mellan text och information, där informationen består av teckensträngar med två modaliteter. Dels teckensträngarna såsom de framträder för läsarna (*scriptons*), dels såsom de är arrangerade i texten (*textons*)²³⁶. Därtill har text en transversal funktion – den mekanism som framställer och ordnar information – som ordnar *scriptons*, genererade ur *textons*, till text.²³⁷ I

²³² Se bl.a. Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 21 f; DeRose, et al., "What is Text, Really?"; Allen H. Renear, David G. Durand & Ellen Mylonas, "Refining our notion of what text really is" *Research in Humanities Computing 4: Selected papers from the 1992 ALLC/ACH Conference.*, red. Susan M. Hockey & Nancy Ide, (Oxford: Oxford University Press, 1997).

²³³ När bokmaterial var en bristvara och bokläsning (oftast genom högläsning) var förbehållet några få utvalda rådde andra principer för tryckning av text. Styckesindelningar och mellanslag användes inte för att spara på material. Se till exempel i Chartier, *Böckernas ordning: läsare, författare och bibliotek i Europa mellan 1300-tal och 1700-tal*.

²³⁴ Ergodisk avser den icke-triviala ansträngningen att läsa en tvärgående text där läsare förflyttar sig rumsligt och på tvären och som kräver mer av en läsare än att "bara" läsa in sekventiell ordnad text (löpande text).

²³⁵ Aarseth, *Cybertext: perspectives on ergodic literature* s. 70.

²³⁶ *Textons* är de grundläggande elementen för uppbyggnaden av själva "byggnadsställningen".

²³⁷ "In addition to textons/scriptons, a text consists of what we shall call a *traversal function*; the mechanism by which scriptons are revealed or generated from textons and presented to the user of the text." Aarseth, *Cybertext: perspectives on ergodic literature* s. 70.

en statisk text är *scripts* alltid konstanta. I en dynamisk text kan de ändras medan *textons* antingen är konstanta eller deras antal och innehåll kan ändras.

Jag anser att Aarseths modell är användbar för att förstå idén bakom skillnaden mellan en texts innehåll och form men även för att den ger mediet, och därmed innehållets materiella förutsättningar, samt graden av manipulation (mänsklig eller maskinbaserad) en avgörande roll. En statisk webbsida med länkar till olika innehållsobjekt har ett fixt antal *scripts* och *textons*, medan ett datorspel eller en databas kan generera olika uppsättningar *scripts* med utgångspunkt i en uppsättning *textons*. I en chat som tillåter inmatning av text är antalet ingående *textons* förstås okänt. Och vid en dynamisk webbplats, som t.ex. en wiki som tillåter inmatning av text, är *textons* såväl som *scripts* föränderliga. Digitala texter kan därmed också sägas inneha mer eller mindre dynamik för *textons* såväl som *scripts*. Digitala texter i XML-formatet är mer dynamiska än digitala texter i PDF-formatet.

Nedan ett exempel på hur ett avsnitt annoterad text i FFO-databasen beskriver innehållets logiska ordning samt kvinnornas föreningsaktiviteter.

```
<div type="capdesc">
  <head>KAPITALSAMMANSTÄLLNING
    <name>Johansson, Klara Elisabet</name>
  </head>
[...]
```

```
<div type="orgcap">
  <head>DELTAGANDE I FÖRENINGAR, SAMMANSLUTNINGAR, STIFTELSE</head>:
  <p>
    <l> sekreterare i <name type="org" id="FBF.055" reg="FBF" key="kvinnopol">Fredrika Bremer-
      förbundet</name> från oktober 1898 och en tid framåt, förmodl okt-dec 1898. I stället för ordinarie
      sekr frk S Ulrich, som var tjänstledig från oktober. Källa: <bibl type="1-source" id="FBF.1899">Fr
      Bremerförbundets verksamhetsberättelse 1899</bibl>.
    </l>
    <l> <name type="org" id="SNI.078" reg="Sällskapet Nya Idun" key="sallsk">Sällskapet Nya Idun</name> 1898
      (Källa: <bibl type="2-source" id="Svensson.97">Svensson -97</bibl>.)
    </l>
  </p>
[...]
```

Figur 7. Exempel på märkning av föreningsaktiviteter i FFO-KAPITAL

I figuren ovan kan vi se exempel på hur Klara Johanssons medlemskap i Fredrika Bremer Förbundet och Sällskapet Nya Idun beskrivs på flera olika sätt genom att tillföra märkord med tillhörande attribut. Märkningen ger här möjlighet för en relativt mångsidig användning av innehållet genom den explicita beskrivningen av de innebörder som tillskrivs ett visst begrepp.

Med hjälp av elementet **<name>** i Figur 7 beskrivs det valda innehållet i detta exempel dels genom det som taggarna innesluter (Fredrika Bremer Förbundet), dels genom de värden som anges i attributen. Attributet **type="org"** innehåller information om att det är en organisation och **id**-attributet (**id="FBF.055"**), som alltid måste vara unikt, talar om för oss att detta är den 55:e förekomsten av

Fredrika Bremer Förbundet i FFO-KAPITAL.²³⁸ Attributet **reg** (**reg**="FBF") anger den normaliserade stavningen eller benämningen av begreppet och i **key** (**key**="kvinnopol") lagras uppgifter om projektets klassificering av organisationen.

Vi har i FFO-databasen klassificerat Fredrika Bremer Förbundet som en kvinnopolitisk organisation och Sällskapet Nya Idun som en sällskapsförening.²³⁹ Den normaliserade stavningen och klassificeringen kan förstås ändras genom att ändra attributvärden i **reg**, **key** eller **type**, men även genom att infoga ytterligare märkning som beskriver en annan sorts klassificering av organisationer och föreningar. Informationen som lagras i **id**-attributet anger en unik identifikation för just denna förekomst av begreppet och detta oavsett om modulen som den ingår i flyttas till en annan plats i samma textsamling. I exemplet ovan (Figur 7) kan vi se att förutom det semantiska innehållet beskrivs även innehållets dokumentstruktur genom att infoga information om rubriker med elementet **<head>** och beskrivning av stycken med elementet **<p>**. En av de viktigaste uppgifterna för utformningen av detta material är att beskriva innehållets ordning. Denna ordning upprättas genom att använda märkordet **<div>**²⁴⁰ som beskriver vilken typ av modul som beskriver innehållet.²⁴¹

När jag nu ovan egentligen föregripit den beskrivning av de principer som nedan kommer att presenteras och som varit en viktig utgångspunkt till de metoder och verktyg som utforskas i den här studien.

3.4 SGML – ett märkspråk för industrin

I slutet av 1960-talet tog man på IBM fram riktlinjer för möjligheten att definiera egna dokumenttyper. Arbetet med specifikationen utfördes av en forskningsgrupp där bland annat Charles Goldfarb ingick. Den första versionen kallades Generalized Markup Language (GML) och hade stöd för vad Goldfarb kallar generell märkning (*generalized markup*) som jag kort introducerat ovan.²⁴²

Den generella märkningsmetoden var ett steg framåt från den procedurella märkningen som enligt honom hade ett antal nackdelar, icke flexibelt, det blev

²³⁸ Benämningen "Fredrika Bremer Förbundet", i olika stavningsvarianter, förekommer ofta i kapitalbeskrivningarna.

²³⁹ Se avsnitt 5.3 s. 138 om klassificeringar i projektet, se även Monica Langerth Zetterman, *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet*, Rapporter från Forskningsgruppen för utbildnings- och kultursociologi, 1 ed., vol. 41 (Uppsala: Forskningsgruppen för utbildnings- och kultursociologi, Under utgivning).

²⁴⁰ **<div>** är en förkortning av engelskans *division*, och används för att beskriva innehållets logiska eller fysiska delar eller enheter oavsett vilken genre texten representerar.

²⁴¹ I kapitel 1 s. 133 presenteras datamodellen och hur den beskrivande märkningen används för bearbetning av innehållet.

²⁴² Föregångaren till SGML var alltså ett proprietärt format som ägdes av IBM, se Charles F. Goldfarb, "SGML: The Reason Why and the First Published Hint", *Journal of the American Society for Information Science* 48.7 (1997).

ofta fel, tidsödande att lära sig, komplexa kommandon med mera.²⁴³ Det generella i metoden sammanfattar Goldfarb i två principer:

- a) Märkning bör beskriva ett dokumentets struktur och andra egenskaper snarare än att specificera den bearbetning som skall utföras på det [dokumentet], då en beskrivande märkning endast behöver göras en gång och kommer att räcka till för all framtida bearbetning.
- b) Märkning bör vara noggrann så att tillgängliga verktyg för bearbetning, noggrant definierade objekt som program och databaser även kan användas för att bearbeta dokument.²⁴⁴

Principerna för det texthanteringssystem som Goldfarb och medarbetare skisserar i den första publicerade artikeln liknar till stor del de principer som vägleder dagens märkspråk. Man skiljde även då på funktionerna för lagring, innehållsproduktion och formatering, det vill säga precis de funktioner som dagens märkspråk har. SGML utvecklades i syfte att lösa konkreta problem inom industrin såsom att underlätta återanvändning och utbyte av information mellan olika plattformar samt i syfte att konstruera principer för strukturering av information och förutsättningar för intelligent informationshantering.²⁴⁵

Den generella märkningen innebar att man använde namngivna generiska identifierare (GI), på samma sätt som beskrivs ovan i avsnittet om COCOA.²⁴⁶ Detta räcker mycket väl till för att namnge vilken sorts objekt som skall beskrivas: ett stycke, en rad, en lista etcetera. Men för att beskriva och bearbeta innehållet behövs ytterligare en dimension i märkningen, nämligen funktionen att kunna typbestämma och ge objektet en unik identitet. Detta kan åstadkommas genom att använda attribut, (primära och sekundära) där typer kan anges och värden kan lagras.²⁴⁷

För att märka alla namn i ett dokument kan vi infoga elementet **<name>** vilket ger oss möjlighet att till exempel ordna namnen alfabetiskt. Om vi vill beskriva vilka olika typer av namn det finns i ett dokument kan vi lägga till ett attribut som anger typ **<name type="surname">** vilket gör det möjligt att ordna namnen utifrån typer och bokstavsordning. Om vi vill identifiera alla unika instanser av namnen infogar vi ett ID-attribut som endast kan ha ett (unikt) värde **<name type="surname" id="name.001">**. I SGML är ID-attributet ett primärattribut för att alla element kan innehålla dem, och enligt samma logik är ett sekundärt attribut sådana som är förbehållna endast vissa element.²⁴⁸

²⁴³ Detta faktum gäller även standard SGML som är ett mycket omfattande och komplext metaspråk.

²⁴⁴ Översättning från Goldfarb & Rubinsky, *The SGML handbook* s. 7-8.

²⁴⁵ Intelligent i betydelsen möjligheter att effektivisera och maximera möjligheter till lagring, produktion och användningsområden.

²⁴⁶ Se s. 85 f.

²⁴⁷ Goldfarb & Rubinsky, *The SGML handbook* s. 11.

²⁴⁸ Primära attribut kallas även för globala attribut, bland annat i TEI:s riktlinjer.

SGML klarar av det mesta vad det gäller modularitet, strukturerad informationshantering och funktioner för länkning av information. Trots det fick SGML:s omfattande och generella standard inte någon större spridning utanför industrin – utom bland de forskare som använde TEI:s riktlinjer – troligen på grund av just komplexiteten och omfånget. Program och verktyg för användning av SGML i verksamheten krävde stora investeringar vilket gjorde att endast stora organisationer hade råd att implementera SGML plattformen. Därför används SGML oftast inom industrin och större akademiska eller statliga organisationer.²⁴⁹

3.5 XML – en verktygslåda för webben

Det plattformsoberoende märkspråket XML utvecklades, som ovan nämnts, av W3C särskilt för webben. Genom att XML är en förenkling av SGML, och dessutom har en striktare syntax, medför detta förbättrade möjligheter att kontrollera om ett dokument är välformat, det vill säga om det överensstämmer med XML:s föreskrivna syntax. Grammatikkontrollen förenklas genom den strikt hierarkiska struktur varje XML dokument måste följa för att vara grammatiskt riktigt, vilket jag gav exempel på inledningsvis.²⁵⁰

En skillnad mellan XML och SGML är att den senare inte föreskriver någon specifik teckenkodning. I XML är dock användning av Unicode föreskriven, något som innebär att man kan använda tecken från alla språk.²⁵¹ Eftersom XML-filer är vanliga textfiler kan de läsas och skrivas utan särskilda hjälpmedel eller program men även bearbetas och struktureras med hjälp av olika verktyg. Dessutom är XML-formatet ett filformat som kan definieras av vem som helst. De ovan beskrivna egenskaper av XML framhålls ofta som fördelar jämfört med andra märkspråk då det är utbyggbart och därmed flexibelt. XML har, i likhet med SGML, inte en uppsättning fördefinierade element som på förhand bestämmer hur innehållet kan beskrivas. Innehållets struktur och ordning, det vill säga vilka element som är tillåtna att använda, bestäms av en dokumentmodell. En sådan används för att beskriva begränsningar och föreskrifter över till exempel dokumentets interna struktur, hur modellens komponenter skall kombineras, samt vilka datatyper och attribut som får användas. Det finns två sorters metoder för beskrivning av dokumentmodeller, en äldre men fortfarande förekommande som heter dokumenttypsdefinition (DTD) och en modernare variant som kallas för schema. En DTD definieras vanligen i en annan fil och

²⁴⁹ Precis som andra system som kräver stora initiala investeringar har SGML framför allt använts för komplex information med lång livslängd som till exempel flygplansmanualer och lagtexter.

²⁵⁰ Se avsnitt 2.2 s. 55.

²⁵¹ Unicode (ISO/IEC 10646): *Universal Multiple-Octet Coded Character Set* är en internationell standard för teckenkodning.

använder en specifik typ av notation.²⁵² I ett schema används XML-syntax för beskrivning av dokumentmodellen. Scheman ger dessutom möjligheter att fastställa vilka datatyper som är tillåtna i dokumentet, något som inte är möjligt att definiera i en DTD.²⁵³ I

I XML finns möjligheter att definiera olika namnrymder, något som medför att man kan kombinera olika XML-scheman i en fil utan att råka ut för konflikter i tolkningen, som skulle uppstå om två element med samma benämning betecknade olika typer av objekt.²⁵⁴ Metoden gör det möjligt att använda element från exempelvis TEI och DocBook i samma dokument, förutsatt att de deklarerats som tillhörande olika namnrymder.²⁵⁵

Det är inte obligatoriskt att använda vare sig DTD eller schema. Om ingen förutbestämd dokumentmodell används går det, som ovan nämnts, bra att kontrollera om dokumentet är välformat relativt XML-syntax, däremot är det då inte möjligt att validera innehållets beskrivning relativt en viss dokumentmodell. För att göra det möjligt att utbyta och migrera innehåll används vedertagna definitioner och dokumentmodeller med specifikationer på hur data skall tolkas inom ett visst kunskapsområde.²⁵⁶ Det finns DTD:er och scheman för till exempel lingvistik, poesi, drama, prosa med mera.²⁵⁷ Detta gör att man sedan kan tolka, lagra och utbyta data enligt en gemensam och genrespecifik dokumentmodell. Förutom de ovan nämnda märkspråken, TEI och DocBook, finns andra märkscheman för olika genrer och egenskaper. Det finns även specifika märkspråk för annotering av grafik, musik, matematisk och kemisk notation och många flera.²⁵⁸

Det är även möjligt att ge element "lokala benämningar" för att beskriva ett visst innehåll. Nackdelarna med detta förfaringssätt är att möjligheterna till användning av generella verktyg som till exempel stilmallar försvåras. Detta gäller även för utbyte av XML-filerna då en lokal terminologi blir ett hinder för utbyte och överföring till andra system på grund av att den inte enkelt kan förstås eller översättas. Användning av en lokal terminologi kan dock vara motive-

²⁵² En DTD kan dock definieras i samma XML-dokument som själva innehållet. Vanligtvis särskiljer man dock på dokumentets modell och dess innehåll.

²⁵³ RelaxNG är ett sådant schema, <<http://relaxng.org/>>. TEI:s dokumentmodell beskrivs i både DTD och med schema som t.ex. RelaxNG.

²⁵⁴ Se riktlinjer vid <www.w3.org/TR/REC-xml-names/>.

²⁵⁵ Sebastian Rahtz, Norman Walsh & Lou Burnard. "A unified model for text markup: TEI, Docbook, and beyond". *XML Europe 2004*, 18-21 April, Amsterdam, The Netherlands, 2004.

²⁵⁶ Ett exempel på det är TEI-schemat.

²⁵⁷ Se vidare i kapitel 4.5 s. 118 om de olika modulerna i TEI-schemat.

²⁵⁸ DocBook, skapades ursprungligen för teknisk dokumentation men kan användas för märkning av dokument som följer en boklig struktur, <www.docbook.org/>. För märkning av grafik kan *Scalable Vector Graphics*, SVG, användas se <www.w3.org/Graphics/SVG/>. *Encoded Archival Description* (EAD) används för beskrivning av arkivposter, se <www.loc.gov/ead/>. För märkning av sociala nätverk kan GraphML användas, se <<http://graphml.graphdrawing.org/>>. MusicXML används till notation av musik, se <www.musicxml.org/xml.html>. MathML är skapat för matematisk notation, se <www.w3.org/Math/>, *Chemical Markup Language* (CML) för kemisk notation, se <<http://cml.sourceforge.net/>>. Se även <<http://xml.coverpages.org/>> för en översikt över XML-applikationer, historik etc.

rad för att förenkla själva märkningsarbetet. För märkning av FFO materialet användes ofta temporära och lokala element som senare ersattes med TEI-element.

3.5.1 Tillämpningar av XML

XML är egentligen en hel familj med tekniker där märkspråket XML används för att beskriva innehållet. För bearbetning och presentation i olika medier används med olika metoder och verktyg, som på engelska benämns stilmallar men som mer har karaktären av skriptspråk.²⁵⁹ Med hjälp av familjen *Extensible Stylesheet Language* (XSL)²⁶⁰ kan ett och samma material manipuleras, bearbetas, omordnas och migreras till andra format samt presenteras på olika sätt. Följande tre verktyg ingår i XSL-familjen:

XSL Transformations (XSLT): används för att transformera XML²⁶¹

XML Path Language (XPath): används för åtkomst och länkar till delar av XML document.

XSL Formatting Objects (XSL-FO): en vokabulär som används för formateringsinstruktioner.

Genom att kombinera de ovan beskrivna metoderna och verktygen kan ett och samma innehåll – förutsatt att det är märkt med XML – presenteras på olika sätt och överföras till andra plattformar. I vissa fall har transformering skett stegvis, till exempel importen till statistikprogrammet SPSS via kalkylblad och HTML-tabeller, hur transformeringen gått till illustreras närmare i kapitel 6.4, s. 189.

Genom märkning av innehållet ges förutsättningar för bearbetning och transformering till flera tillämpningar och användningsområden. Arbetet och resultaten av detta presenteras och analyseras i del II. I del III presenteras med vilka av de ovan presenterade verktygen dessa transformeringar har utförts, hur det gått till. Där diskuteras även verktygens nackdelar och förtjänster samt konkreta lösningar på manipulering och bearbetning av innehåll.

I avsnitten ovan har den historiska utvecklingen av märkspråk och den hierarkiska dokumentmodellen behandlats. Jag har även visat hur innehållets egenskaper och strukturer kan beskrivas samt gett en översiktlig inledning de verktyg som använts för tillämpningarna. I nedanstående avsnitt presenterar jag alternativa metoder till en hierarkisk dokumentmodell.

²⁵⁹ Ett skriptspråk är ett tolkat programspråk med en egen körningsmiljö (exekveringsmiljö) och kan användas på olika operativsystem.

²⁶⁰ Se <www.w3.org/Style/XSL/>.

²⁶¹ Standardverken för XSLT 1.0 och 2.0 är Michael Kay, *XSLT: programmer's reference*, 2. ed. (Birmingham: Wrox Press, 2001); Michael Kay, *XSLT 2.0 programmer's reference*, 3rd ed. (Indianapolis, IN: Wiley Pub., 2004).

3.6 Alternativ till den hierarkiska dokumentmodellen

I följande avsnitt presenteras alternativ till den hierarkiska dokumentmodellen. Avsikten är att problematisera den hierarkiska datamodellen men även att diskutera möjliga alternativ. Först beskrivs ett av de äldre märkspråken som introducerades ungefär samtidigt som SGML i slutet av 1960-talet. Den andra metoden handlar om extern märkning och hur man kan gå tillväga för att placera märkningen utanför texten.

3.6.1 COCOA

I ett av de första deskriptiva märkspråken, som benämndes COCOA, infogas referenser – eller markörer – i texten för att markera en variabel som beskriver valda egenskaper såsom egennamn, titlar, ordklasser och liknande. COCOA användes först på 1960-talet och implementerades senare bland annat i textanalysprogrammen TACT (*Textual Analysis Computing Tools*)²⁶² och OCP (*Oxford Concordance Program*).²⁶³ Detta märkspråk har alltså varit en viktig del för utvecklingen av datorbaserade textanalyser. I COCOA används nedanstående principer för markörernas start- och slutpunkter samt för namngivning av variabeln:



Figur 8. Märkspråket COCOA, exempel på syntax

Till skillnad från XML och SGML infogas variabelns värde, i detta fall tre namn, mellan start- och slutmarkörerna. Detta förhindrar istället en hierarkisk beskrivning av innehållet. Märkningen blir så att säga endimensionell eller ”platt” då beskrivningsmodellen, eller märkspråket, inte erbjuder stöd för nästlade element. I exemplet nedan visas exempel på hur en pjäs från tidigt 1600-tal annoterats för stilistisk analys med hjälp av COCOA.²⁶⁴

²⁶² I TACT:s extensiva manual ges introduktion till märkning med COCOA samt många exempel på och användning av konkordanser och fulltextanalyser för forskning och undervisning. Ian Lancashire, *Using TACT with electronic texts: a guide to text-analysis computing tools* (New York: Modern Language Association of America, 1996).

²⁶³ Hockey, *Electronic texts in the humanities: principles and practices*. 27 ff.

²⁶⁴ Exemplet är ett urklipp från dokument nr: 0603 Oxford Text Archive och är ett drama annoterat med COCOA, *A new way to pay old debts* (1633) av Philip Massinger. Det extra mellanslag som syns i variabeln <Q TAPWELL. > finns även i originaldokumentet.

```

<SN ACTUS PRIMUS, SCENA PRIMA:>
<SSD WELBORNE. TAPWELL. FROTH.>
<Q WELBORNE.> NO BOUZE? NOR NO TOBACCO?
<Q TAPWELL. > NOT A SUCKE SIR,
NOR THE REMAINDER OF A SINGLE CANNE
LEFT BY A DRUNKEN PORTER, ALL NIGHT PALDE TOO.

```

Figur 9. Exempel på COCOA märkning av 1600-talsdrama

Markörerna i figuren ovan markerar, i nämnd ordning: vilken akt och scen det är, vilka karaktärer som medverkar i denna scen, samt vilka av dessa talare replikerna tillhör. Användaren kunde namnge variablerna efter eget tycke. Dock utvecklades så småningom konventioner för namngivning inom olika genrer.²⁶⁵ I COCOA används ingen sluttagg utan första aktens första scen förutsätts behålla sitt ”värde” tills en ny variabel (tagg) markerar något annat, i detta fall första aktens andra scen. Detta medför att referenser som logiskt och semantiskt överlappar varandra kan annoteras utan att kollidera med varandra, något som är till stor fördel om flera parallella strukturer behöver beskrivas.

Exempelvis pjäsen i exemplet ovan; om man där skulle vilja lägga till beskrivning av bokens fysiska karaktär samtidigt med beskrivningen av innehållets struktur, skulle det kunna bli problematiskt att åstadkomma med SGML eller XML.²⁶⁶ I dessa märkspråk kan det vara svårt att beskriva överlappande hierarkier på grund av den strikta och nästlade dokumentstrukturen. Även om en hierarkisk ordning av innehållet kan vara nödvändig, exempelvis för transformering och bearbetning, kan det orsaka problem vid märkning av komplexa dokumentet eller olika typer av egenskaper. Ett innehållsobjekt som skall beskrivas är inte alltid ordnad i någon ”naturlig” hierarki och den logiska ordningen, såsom vi uppfattar den, överensstämmer helt enkelt inte alltid med en hierarkisk dokumentmodell.²⁶⁷

För att lösa problemet med flera logiska beskrivningsnivåer kan man, i den hierarkiska datamodellen, infoga andra typer av markörer och hänvisningar. Det är fullt möjligt att där använda samma typ av referensmarkörer som COCOA

²⁶⁵ I COCOA var bokstaven ”L” dedicerad till radnumrering och kunde användas till att åsidosätta den automatiska radnumreringen som fanns i vissa textanalysprogram, som t.ex. i *Oxford Concordance Program*.

²⁶⁶ Hur man löser beskrivning av olika logiska ordningar vid märkning av olika slags texter är ett av de områden som det förkommer strider inom TEL. Mer om överlappande hierarkier och exempel på hur jag har hanterat beskrivningens hierarkiska nivåer i FFO-databasen i kapitel 6.1.2.3 s. 182.

²⁶⁷ Problemet med överlappande hierarkier diskuteras bland annat i David Barnard, et al., *Hierarchical Encoding of Text. Technical Problems and SGML Solutions*, Computers and the Humanities, The Text Encoding Initiative: Background and Contents. , red. Nancy Ide & Jean Vâronis, vol. 29 (1995); Steven DeRose. ”Markup Overlap: A review and a Horse”. *Extreme Markup Languages*, August 2-6, 2004, Montréal, Québec, 2004; Renear, ”Out of Praxis: Three (Meta)Theories of Textuality”.

använder.²⁶⁸ Detta kallas för tomma element, det vill säga element som inte innesluter någon text och där ingen sluttagg behövs. Det medför dock att det inte är möjligt att i ett och samma element beteckna ett givet område av innehållet utan att infoga ett nytt element som markerar starten på ett nytt eller annan typ av område. Denna funktionalitet tyder på arv från den typ av märkning som används med t.ex. COCOA. Elementet **<milestone>** i TEI är ett exempel på den sortens endimensionell märkning som ärvts från märkspråk som COCOA: värdet av ett **<milestone>** gäller tills ett nytt **<milestone>** element appliceras.²⁶⁹

I SGML är det möjligt att vid vissa tillfällen utesluta sluttagg, förutsatt att elementen nästlats ordentligt. Om ett element innehåller fler och därmed nästlade element räcker det att ”stänga” det överordnade elementet med en sluttagg för att de underordnade elementen skall anses ”stängda”.²⁷⁰

SGML	XML
<pre> <name> <persName id=KEY.001> <foreName>Ellen <surName>Key </name> </pre>	<pre> <name> <persName id="KEY.001"> <foreName>Ellen</foreName> <surName>Key</surName> </persName> </name> </pre>

Figur 10. Jämförelse av bruket av sluttagg för nästlade element mellan SGML och XML

I exemplet ovan ser vi skillnaden mellan användning av sluttagg i SGML och i det mer strikta märkspråket XML. I XML måste alla element stängas med en sluttagg för att dokumentet skall anses vara välformulerat (enligt XML:s syntax). Undantaget är som ovan nämnts vissa tomma element, det vill säga sådana som inte har något innehåll utan enbart fungerar som lagring för attributvärden i detta fall ett unikt ID-nummer för egennamnet ”Ellen Key”.²⁷¹

²⁶⁸ Detta tillvägagångssätt föreslås i Barnard, et al., *Hierarchical Encoding of Text. Technical Problems and SGML Solutions*.

²⁶⁹ Se Barnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 874. Se även <www.tei-c.org/P4X/ref-MILEST.html>.

²⁷⁰ Steven J. DeRose, *The SGML FAQ book: understanding the foundation of HTML and XML* (Boston: Kluwer Academic, 1997) s. 37.

²⁷¹ Från och med TEI P5 anges ”id” det av W3C fördefinierade attributet ”xml:id”. Lou Burnard. ”New tricks from an old dog: An overview of TEI P5”. *Dagstuhl Seminar Proceedings. Digital Historical Corpora - ARchitecture, Annotation, and Retrieval*, Dagstuhl, Germany, 2007. I sådana här attribut betecknar ”xml” den aktuella namnrymden. Namnrymder (namespaces) används för att vara specifik med vilket märkspråk som används samt för att kunna integrera olika märkspråk i ett och

Detta är inte nödvändigt i SGML. Som vi kan se i figuren ovan är SGML syntax mer komprimerad jämfört med den mer strikta XML syntaxen. En anledning till den stora valfrihet som SGML erbjuder, som exempelvis möjligheter att utelämnas sluttaggar för att stänga element – om de istället ersätts av en ny tagg som markerar slutet på den förra – är enligt Goldfarb ”en eftergift för slappa maskinskrivare”.²⁷² I XML måste dock även ett tomt element, det vill säga som inte betecknar något innehåll utan enbart tjänar som platshållare, innehålla markören "/" för att stänga elementet som illustrerades i *Figur 10* ovan.

En annan anledning till valfriheten är bland annat SGML:s inriktning mot professionella användningsområden som *de facto* standard för manualer och teknisk dokumentation., jämfört med den mer förenklade och populariserade versionen XML som är designad med en målsättning för ökad förutsägbarhet, enkelhet och webbanvändning.

3.6.2 Extern annotering

En annan möjlighet till märkning av innehåll är att tillföra beskrivning i digitala dokument genom extern märkning, så kallad *stand-off markup*, vilket innebär att beskrivningen lagras utanför själva texten och använder pekare som anger vilket innehåll samt var någonstans i dokumentet beskrivningen hör hemma.²⁷³ Möjligheter som den externa märkningen (*stand-off markup*), jämfört med den XML märkning som infogas i själva dokumentet (*embedded markup*), medför är att förse innehållet med flera olika lager av märkning. Det erbjuder en annan sorts funktionalitet än den strikt hierarkiska dokumentmodellen, till exempel att tillhandahålla multipla beskrivningar av ett och samma innehåll genom att flera olika logiska hierarkier kan beskrivas oberoende av varandra. Ett beskrivningslager för syntax, ett annat för narrativet, ett tredje lager för dokumentstrukturen etcetera. Dessa noteringar kan lagras utanför det ursprungliga innehållet (källfilen) och utgör därmed ett eget och fristående dokument.²⁷⁴ Vi skall dock komma ihåg att det är stor skillnad på extern annotering som görs av en redaktör eller forskare med full kontroll över innehållet och mellan system konstruerade för att många användare skall kunna tillfoga annoteringar för samarbete eller undervisning. I detta avsnitt kommer jag framför allt att diskutera annoteringar

samma dokument. Se även diskussionen i TEI-listan som föregick beslutet att ändra från ”id” till ”xml:id” i den nya utgåvan av TEI:s riktlinjer, <<http://listserv.brown.edu/archives/cgi-bin/wa?S2=TEI-L&D=0&q=xml%3Aid+P5&0=S&s=&f=&a=&b=>>>

²⁷² *The unclosed end-tag is the same concession to sloppy typists as the unclosed start-tag – the closing delimiter can be left out if it is followed by another tag.* Goldfarb & Rubinsky, *The SGML handbook* s. 318.

²⁷³ En W3C standard för detta ändamål är Annotea, som är en del av den semantiska webben, se <www.w3.org/2001/Annotea/>. I projektet ”Märkning av utbildningsinnehåll” utvecklades en prototyp baserad på Annotea, närmare beskriven i Jonas Engman, ”Handling fine-meshed filtering from archive to resource level. Combining RDF-metadata with XML content markup,” Umeå University, 2004 s. 10.

²⁷⁴ Det är även möjligt att extrahera annoteringar och sammanställa i ett nytt dokument från en ursprunglig text med inbäddad märkning.

som görs i syfte att samarbeta om och kommentera innehåll, snarare än att beskriva innehållet.²⁷⁵ Extern annotering kan vara användbart i alla sammanhang där samarbete om ett innehåll är värdefullt eller viktigt, såsom i undervisningen eller för forskare som utformar digitala material tillsammans.²⁷⁶

Ett problem med extern annotering är bland annat att när beskrivningen tas ur sitt sammanhang, till exempel om delar av ett dokument flyttas till ett annat dokument, ändras även den kontext som annoteringen initialt avsåg beskriva. Andra nackdelar med den externa märkningen är att de olika lagren relativt snabbt skapar oreda och att det därmed blir svårt att skapa en överblick av annoteringarna. Engman hävdar i sin rapport att det är just den externa märkningen som gör det möjligt att förse valfria delar av innehållet med metadata, där varje stycke skulle kunna beskrivas.²⁷⁷ Detta håller jag inte alls med om, det finns flera möjligheter att annotera även i TEI, dock med noteringarna infogade i dokumentet istället för utanför. Genom att till exempel använda elementet **<note>** som är avsett för annoteringar av olika slag kan noter som redan finns i den ursprungliga texten beskrivas eller nya noter skapas och som då infogas av den som kodar texten.²⁷⁸ Detta förutsätter att man genom typning, det vill säga använder attributet **type**, skiljer mellan olika slags noter.

Externa annoteringar på webben skulle dock kunna vara användbart för forskare, lärare och studenter för samarbete eller för att kollektivt skapa kommentarer om ett visst innehåll. I undervisningen skulle lärare och studenter kunna skapa egna kommentarer och tolkningar av en given textmängd utan att för den skull behöva ha rättigheter att ändra i det ursprungliga dokumentet.

Vanligtvis när vi annoterar en text, genom att anteckna i marginalen eller på ett särskilt papper, är vägen från tanke till text inte särskilt lång. Likaså är annoteringarna särskiljda från ursprungliga texten. Omvänt förhåller det sig på samma sätt: annoteringarna finns kvar nästa gång texten tas fram och läsaren får snabbt en överblick över alla annoteringar, åtminstone på samma uppslag. Noteringar upplevs ofta som något personligt som vi använder oss av för att strukturera och organisera innehållet eller för att underlätta att komma ihåg, hitta tillbaka, strukturera idéer och frågor.²⁷⁹ När användare däremot annoterar en text på webben är vägen från tanke till text oftast längre; en särskild funktion

²⁷⁵ Den beskrivande märkningens olika lager och logiker behandlas i kapitlet om TEI:s riktlinjer.

²⁷⁶ En funktion för externa annoteringar har som tidigare nämnts införts för Litteraturbankens e-texter, se s. 34.

²⁷⁷ Engman, "Handling fine-meshed filtering from archive to resource level. Combining RDF-metadata with XML content markup," s. 10.

²⁷⁸ Möjligheter att beskrivna existerande noter (t.ex. fotnoter) eller att infoga redaktörens annoteringar eller andra kommentarer av det ursprungliga innehållet finns med i basnivån av TEI:s riktlinjer och kan användas oavsett vilken modul som används. Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 884-885. Se även <www.tei-c.org/P4X/CO.html#CONONO>.

²⁷⁹ Hur forskare inom humaniora gör när de forskar, hur de använder texter och hur de integrerar teknik i sitt arbete studeras i rapporten, William Brockman, et al., *Scholarly Work in the Humanities and the Evolving Information Environment* (Washington, D.C.: Digital Library Federation, Council on Library and Information Resources, 2001) ssk. s. 7 ff.

skall aktiveras, det ska väljas om annotering skall vara privat eller offentlig och kunna läsas av alla som besöker webbplatsen innan det är dags att fästa tankarna på pränt.²⁸⁰

De webbaserade annoteringar, som är menade att göras publika, har en inbyggd logik i sig att de bör vara meningsfulla även för andra, kunna användas i ett längre tidsperspektiv och användas för samarbete. Därför menar man en studie om skillnaderna mellan privata och publika annoteringar, tenderar annoteringar att bearbetas och omarbetas i högre grad än diskussionsinlägg.²⁸¹ Resultaten av studien känns dock förutsägbara, det torde vara givet att sådant som är ämnat för publik publicering ägnas mer möda och omtanke än sådant som skrivs ned för privat bruk.

Tillvägagångssättet att använda sig av externa annoteringar för samarbete, forskning och undervisning påminner om de alltmer vanliga wikisystemen, där registrerade användare kan bidra till att skriva och bearbeta innehåll på ett enklare sätt än med vanligt HTML.²⁸² Den i särklass största wikin torde vara Wikipedia även om Sourceforge har en imponerande samling projekt, tillämpningar och program.²⁸³ I webbens barndom, det vill säga tidigt 1990-tal, fanns möjlighet att i webbläsaren Mosaic infoga privata annoteringar. Det fanns planer på att göra funktionen publik, men av detta blev det inget.²⁸⁴

I en användarstudie om externa annoteringar där externa annoteringar infogades i varianter av Chopins preludier framkommer det att publika annoteringar kan vara användbara för att underlätta för andra användare av digitala vetenskapliga utgivningar.²⁸⁵ Det är inte ovanligt att användare skriver ut ett digitalt vetenskapligt verk för att kunna tillfoga egna kommentarer. I det här fallet finns de deltagande forskarnas annoteringar tillgängliga för visning på webben och nya användare slipper bläddra fram och tillbaka i en omfattande notapparat som skulle ha varit fallet ifall utgåvan skrivits ut.

I Chopin-studien fanns möjlighet att hierarkiskt ordna annoteringarna och att gruppera annoteringar i en överordnad topologi. Möjligheter att ordna och gruppera annoteringar är nödvändigt om många användare skall ha samtidig

²⁸⁰ Tankens långa väg till webbnöteringen diskuteras i John Bradley & Paul Vetch, "Supporting Annotation as a Scholarly Tool - Experiences From the Online Chopin variorum Edition", *Literary and Linguistic Computing* 22.2 (2007).

²⁸¹ Skillnaderna mellan privata och publika annoteringar diskuteras i en studie där man analyserat studenternas användning av 1700 annoteringar, gjort intervjuer och diskussionsinlägg i en kurs om människa-dator interaktion, se Catherine Marshall & Bernheim B. Alice J. "Exploring the Relationship between Personal and Public Annotations". *ACM/IEEE Joint Conference on Digital Libraries (JCDL04)*, June 7–11, Tucson, Arizona, 2004.

²⁸² Namnet wiki betyder, enligt Wikipedia, "snabb" på hawaiiska. En wiki utmärks av versionshantering och kollektiv redigering av innehåll direkt via webbläsaren.

²⁸³ Se <<http://wikipedia.org>> och <<http://sourceforge.net/>>.

²⁸⁴ Uppgift hämtad från <<http://1997.webhistory.org/www.lists/www-talk.1993q3/0791.html>> Webbläsaren Mosaic utvecklas inte längre, den sista versionen släpptes i mitten av 1990-talet. Den finns dock fortfarande för nedladdning vid <<http://archive.ncsa.uiuc.edu/mosaic.html>>.

²⁸⁵ Bradley & Vetch, "Supporting Annotation as a Scholarly Tool - Experiences From the Online Chopin variorum Edition".

åtkomst till ett system. Frågan är hur gruppering och ordnande skall ske, distribuerat och valfritt eller med en kontrollerad vokabulär?

Ett hinder för effektiv överblick och struktur av externa annoteringar, som Bradley och Vetch diskuterar, är webbläsarnas nuvarande funktionalitet.²⁸⁶ Webbläsare är konstruerade för visning av innehåll på Internet och även om man kan förbättra funktionerna med tilläggsprogram fungerar de inte särskilt bra som verktyg för att manipulera material som inte "ägs" av personen som annoterar och som inte är lagrad på den lokala datorn. Externa annoteringar måste lagras på en server om de skall kunna göras publika, och inte på användarens lokala dator. Det medför förstås begränsade möjligheter för användarna att manipulera och organisera sina egna och men även andras annoteringar utan att annoteringarna skiljs från sitt sammanhang, speciellt om det annoterade innehållet förändras.²⁸⁷

3.7 Metoder och verktyg för metadata på webben

I de följande avsnitten presenterar jag metoder och verktyg används för information om dokument eller andra typer av resurser. Dessa metoder och verktyg används ofta tillsammans med de hierarkiska dokumentmodellerna och är en viktig del av utvecklingen av webbens mångsidighet och funktionalitet. Inledningsvis beskriver jag den historiska utvecklingen av "den semantiska webben" och andra verksamheter som bidragit till de metoder och verktyg som finns tillgängliga idag.

Det specifika med metoderna är att själva beskrivningen av data, snarare än att både beskriva och hierarkiskt ordna, är i fokus, därtill är metoderna mer ämnesorienterade än textorienterade. En förklaring till det kan vara att man inom detta område fokuserar på automation och maskinläsbar beskrivning av innehåll. De ämnesorienterade metoderna utvecklas främst inom datavetenskaperna och området artificiell intelligens till skillnad från text- och dokumentorienterade metoder som främst utvecklas inom humaniora och biblioteks- och informationsvetenskap.

Intresset för att förbättra åtkomst och användning av den information som finns tillgänglig på webben finns som ovan nämnts inom många verksamhetsområden. Metoder, verktyg och applikationer har utvecklats i snabb takt och sedan slutet av 1990-talet har talet om "den semantiska webben"²⁸⁸ florerat i

²⁸⁶ Ibid. s. 239.

²⁸⁷ Att ladda ner annoteringarna på sin egen dator är en möjlighet. Men utan samtidig åtkomst till det innehåll som annoteringen avser beskriva förloras till viss del vitsen med att lagra annoteringarna externt. Dessa blir då så att säga tagna ur sitt sammanhang och kan mista sin betydelse.

²⁸⁸ The Semantic Web, notera versalerna och att det benämns *den* istället för *en*.

medier och inom webbutveckling. Det talas om metadata för webben och standarder som *Dublin Core* och *Resource Description Framework* (RDF).²⁸⁹

Tim Berners-Lee formulerar visionen om den semantiska webben i ett utkast 1998²⁹⁰ samt 1999 i boken *Weaving the Web*.²⁹¹ Året efter publicerar W3C²⁹² en mer formell rapport om webbarkitektur och metoder för att beskriva och utbyta data²⁹³ och 2000 kommer en rapport om teknisk utveckling för den semantiska webben där visionen i stora drag skisseras.²⁹⁴ Ambitionen var att utveckla beskrivningsformat för att förbättra användbarheten av den enorma mängd stoff i form av dokument, bild, ljud, video och andra typer av data som webben erbjuder. Beskrivningsformatet skulle ligga som ett lager ovanpå webben och förstås av maskiner i första hand och inte nödvändigtvis av människor, som med den beskrivande märkningen. Genom att försöka representera innebörden hos de objekt som lagras skulle man kunna stödja system som tillåter självständigt regelbaserat agerande, slutsatsdragningar och automatiska analyser. Som stöd för detta behövs standardiserade modelleringspråk, modeller och ontologier för maskinerna, dessa bör dessutom vara generellt accepterade av aktörer inom olika domäner. Det är viktigt att komma ihåg att vad man talar om här är maskinläsbar semantik och inte i första hand för människor.

Intresset tog fart när en artikel om den semantiska webben publicerades under 2001.²⁹⁵ I Berners-Lee et al. artikel från 2001 skisseras storslagna visioner om hur nästa generation av webben skulle komma att formas, artikeln summeras bland annat med visioner om hur webben kan medverka till utveckling av alla former av kunskap:

The semantic web is not "merely" the tool for conducting individual tasks that we have discussed so far. In addition, if properly designed, the Semantic Web can assist the evolution of human knowledge as a whole.²⁹⁶

Media i alla former hängde på och den semantiska webben framstod som nästa stora utvecklingstrend för webben²⁹⁷ och stora ord som webbens framtid före-

²⁸⁹ Dublin Core är en standard för metadata om webblesurser som togs fram 1995, se <<http://dublincore.org/>>. RDF är W3C:s metod för att beskriva webblesurser och relationer mellan element från samma eller olika metadatastrukturer. Se <www.w3.org/RDF/>. Mer information om dessa metoder och verktyg på s. 97.

²⁹⁰ Se <www.w3.org/DesignIssues/Semantic.html>.

²⁹¹ Tim Berners-Lee & Mark Fischetti, *Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor*, 1. ed. (San Francisco: Harper, 1999).

²⁹² W3C arbetar med att utveckla standarder, programvara och tekniska protokoll för webben. Grundades 1994 och är ett konsortium med drygt 500 medlemmar från forsknings- och utvecklingsområdet samt industrin. Se <www.w3.org/Consortium/>.

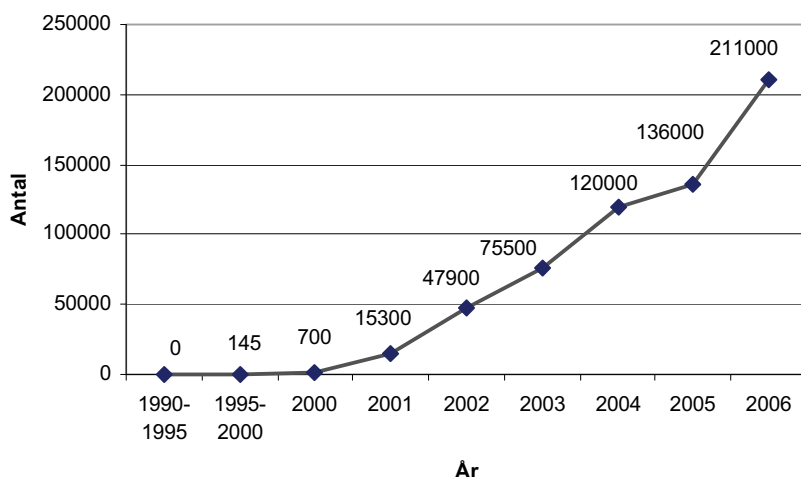
²⁹³ Web Architecture: Describing and Exchanging Data. Daterad den 7 juni 1999. Se <www.w3.org/1999/04/WebData>.

²⁹⁴ Se <www.w3.org/2000/01/sw/DevelopmentProposal>.

²⁹⁵ Tim Berners-Lee, James Hendler & Ora Lassila, "The Semantic Web", *Scientific American*. May (2001).

²⁹⁶ Tillgänglig vid: <www.sciam.com/issue.cfm?issuedate=May-01>.

kom flitigt. I *Figur 11* nedan visas en periodisk sökning efter webbsidor som handlar om den semantiska webben med sökmotorn Yahoo som illstruerar den ökade användningen av begreppet ”semantiska webben” från och med år 2001.²⁹⁸ Den första sökträffen är en artikel daterad 1997, *Semantic Web vision paper*, och är författad av Alexander Chislenko. Artikeln handlar om bristen på strukturerat och väl beskrivet innehåll på webben samt användare som inte vet hur de kan nyttja det material som finns tillgängligt. Den handlar även om hur man kan skapa distribuerade kunskapssystem, med dåtidens relevanta tekniker.²⁹⁹ Chislenko introducerar i artikeln termen den semantiska webben, det vill säga innan Berners-Lee som annars brukar anges som den som myntade detta begrepp.³⁰⁰



Figur 11. Periodisk sökning av ”semantic web”, antal träffar per angiven tidsperiod

Berners-Lee menar att den semantiska webben inte bara är ett verktyg för forskning och utbildning utan ett helt nytt sätt att tänka och samarbeta med

²⁹⁷ Ett typiskt exempel var projektet Personalized Access to Distributed Learning Repositories (PADLR), finansierat av Wallenberg Global Learning Network under perioden 2001-2004. Utvecklingen inriktades på att skapa verktyg för semantisk beskrivning av allehanda objekt (s. k. *learning objects*) och särskilt intresse inriktades mot utbyte av information inom tekniskt naturvetenskapliga utbildningsområdet.

²⁹⁸ Sökningen utfördes med hjälp av sökmotorn Yahoo.se 2007-05-20 på frasen ”semantic web”, med kalenderår som periodicitet förutom under 1990-talet då antalet artiklar var så få.

²⁹⁹ Min översättning av *Distributed knowledge-processing systems*. Tillgänglig vid: <www.ethologic.com/sasha/articles/SemanticWeb.html>.

³⁰⁰ I det första utkastet om webbdesign på W3C:s webb från 1998, <www.w3.org/DesignIssues/Semantic.html>.

hjälp av vad han benämner ”verksam information”.³⁰¹ Den semantiska webben är tänkt att vara dynamisk och kunna erbjuda avancerade former av öppen, global samverkan, alltifrån enklare informationsutbyten till mycket komplexa processer mellan många databaser och flera parter.³⁰² Det innebär även att alla användare, expert som novis, skall kunna hitta och få tillgång till precis rätt information genom att formulera frågor med hjälp av sunt förnuft och naturligt språk.

Visionen om den semantiska webben har växt fram ur flera olika pragmatiska önskningar och visionära influenser – alla med målsättningen att förbättra webben som den sett ut och ser ut idag. En sådan influens är önskan om att lösa bekymmer med den uppenbara oredan av digitala dokument, nya genrer, nya teknologier eller nya vägar för att ordna och publicera innehåll.³⁰³ En annan influens är forskningsområdet artificiell intelligens (AI) där man strävar efter att förstå och konstruera intelligenta system. Inom AI-området sysslar man med formella teorier om logik, beslutsfattande och sannolikheter och AI har genom den semantiska webben fått lämpligt underlag att lösa sådana problem med metoder och verktyg som utvecklades för mer än 30 år sedan.³⁰⁴ Ytterligare en influens är en mer utopisk önskan att överföra bördan av hantering av den hela tiden ackumulerade mängden information till datorer och intelligenta system genom att använda webben som bakgrund för att hjälpa oss hantera dessa informationsmängder.³⁰⁵ Dessa tre önskningar är försvarbara, frågan är dock: Är de realistiska? Vissa kritiker menar att den semantiska webben bara är en illusion och ett innehållslöst och orealistiskt *buzzword* som florerar inom vissa delar av datavärlden att användas för att söka pengar för nya forskningsprojekt.³⁰⁶ Andra menar att kritikerna varit för försiktiga på grund av lärdomen från kritiken mot webben vid introduktionen 1994. Då höjdes kritiska röster för att

³⁰¹ Min översättning av *actionable information* s. 96 i Nigel Shadbolt, Wendy Hall & Tim Berners-Lee, ”The Semantic Web Revisited”, *IEEE Intelligent Systems* 21.3 (2006).

³⁰² Ibid. s. 100 f.

³⁰³ Se till exempel David M. Levy, *Scrolling forward: making sense of documents in the digital age*, 1. ed. (New York: Arcade Pub, 2001).

³⁰⁴ Se t.ex. Drew McDermott & Dejing Dou. ”Representing Disjunction and Quantifiers in RDF”. *First International Semantic Web Conference*, June 9-12, 2002; Shadbolt, Hall & Berners-Lee, ”The Semantic Web Revisited” s. 100 f.

³⁰⁵ Ett exempel är utveckling av mer eller mindre specifika ontologier i syfte att förbättra informationshantering och åtkomst på webben eller i programvara. Exempel på verksamhet i Sverige är Knowledge Management Research Group (KMR) på KTH, där man strävar efter att utveckla en begreppslig webb, som ett komplement till en maskinnära semantisk webb. Ett exempel är bland annat en så kallad concept browser (Conzilla) där innehåll kan utforskas med hjälp av väl strukturerade grafer och visuella hjälpmedel. Verktöget är baserat på gamla välkända modelleringstekniker, som t ex UML (Unified Modelling Language). Se <<http://kmr.nada.kth.se/>>.

³⁰⁶ Bland annat i McCarty, *Humanities computing*. En diskussion om den semantiska webbens innehållslöshet fördes i den internationella diskussionslistan och humanistiska seminarier Humanist Discussion Group i oktober 2003 se <www.princeton.edu/humanist/>, (subject: *reification; or, a complaint*).

webben endast vara ett förbättrat gränssnitt för *Gopher*³⁰⁷ men som alla vet utvecklades webben till ett globalt och använt digitalt medium för information och kommunikation.³⁰⁸

Sett ur W3C:s perspektiv finns ett stort behov att försöka bringa ordning på de löst spridda nätverken av filer och dokument som utgör webben. Standarder för metadata och kommunikation är tänkta att bringa ordning på innehållet. Ett flertal standarder för metadata har tagits fram, men har inte fått någon större spridning.

Denna utveckling reflekterar webbens brist på struktur och ordning som leder till svårigheter för åtkomst av innehåll och material – i kontrast till bibliotekens strukturerade system.

Som ett led i W3C:s riktning med att ordna innehåll är att möjliggöra för distribuerade kunskapsbaser genom att lära av *Cyc* som är en av de största tillämpningarna av formaliserad kunskapsrepresentation.³⁰⁹ *Cyc* är ett AI-projekt som försöker samla ihop en allomfattande databas bestående av ontologier avsedda att beteckna all världens kunskap.³¹⁰ En ontologi är den stomme med begrepp och sammanhang som representerar ett system av begrepp inklusive deras definitioner. I en ontologi fastställs betydelsen för begrepp inom ett område som beskriver hur dess termer kan kombineras och användas. Målsättningen med projektet är att möjliggöra för AI-applikationer att kunna resonera på ett mer "mänskligt" sätt.³¹¹ Terry Winograd menar dock att förhoppningar inom området om att överföra mänskliga förmågor till maskiner kräver en nästan barnslig hoppfullhet.³¹² Willard McCarty instämmer och framhåller att sådana projekt bygger på en föreställning om att man genom ackumulering av omfattande mängder kunskap kan uppnå bättre förståelse.³¹³

En annan uppgift för den semantiska webben är att tillhandahålla infrastruktur för koordination av utbyte och delning av data. I denna vision kommer

³⁰⁷ Gopher är en föregångare till webben, som är ett nätverksprotokoll och ett system för sökning och hämtning av dokument på internet. Gopher använder ett textbaserat gränssnitt med en strikt hierarkisk struktur för lagring av dokument. För information om Gopher, se <<http://gopher.quux.org:70/Software/Gopher>> för de webbläsare som stödjer gopher-protokollet se även <[gopher://gopher.quux.org/1/Software/Gopher](http://gopher.quux.org/1/Software/Gopher)>.

³⁰⁸ Catherine Marshall & Frank Shipman. "Which Semantic Web?" *Fourteenth ACM Conference on Hypertext and Hypermedia*, Nottingham UK, 2003.

³⁰⁹ Benämns vanligen *Knowledge Representation* (KR). För introduktion till ämnet, se John F. Sowa, *Knowledge representation: logical, philosophical, and computational foundations* (Pacific Grove: Brooks/Cole, 2000).

³¹⁰ Av "encyklopedi". I Cyc särskiljs *common sense knowledge* och *formalized common sense*. För mer information och exempel se <www.opencyc.org/>.

³¹¹ Kritiken mot AI, intelligenta system och liknande företeelser behandlas inte närmare här, jag hänvisar istället till Terry Winograd och Willard MacCarty (se nedan).

³¹² *It takes an almost childish leap of faith to assume that the modes of explanation that work for the details of block manipulation will be adequate for understanding conflict, consciousness, genius, and freedom of will.* s. 207 i Terry Winograd, "Thinking machines: Can there be? Are we?" *The Boundaries of Humanity: Humans, Animals, Machines*, red. James Sheehan & Morton Sosna, (Berkeley: University of California Press, 1991).

³¹³ McCarty, *Humanities computing* ssk. s. 20-53.

distribuerade databaser utvecklas inom varje domän ämnade för tillämpningar och specifika behov. Sedan sköter den semantiska webben utbyte, kommunikation och åtkomst mellan databaserna. Förespråkarna för den semantiska webben hävdar att agenter kommer att kunna sköta sökning, urval, sammanställning och beslut av data som sedan kan sorteras och användas på olika sätt.³¹⁴ Det låter ju bra, men det ägnas inte mycket uppmärksamhet åt hur man skall lösa svårigheter med att utbyta data mellan olika kunskapsdomäner eller hur man skall hantera storskalighet. För datorer är det lätt att utbyta information om okomplicerade och väl strukturerade data som filnamn, filstorlek, lösenord etcetera, men betydelsen av termer såsom ”handling” eller ”värde” är förstås inte lika självklara och förutsätter antaganden om gemensam förståelse.

De ovan skisserade perspektiven exemplifierar de olika förväntningar om vad den semantiska webben egentligen kan åstadkomma och vad som levereras. Kunskapsrepresentation och kunskapsanvändning har varit fokus för AI i nära 50 år utan att några stora allmängiltiga dekontextualiserade kunskapssystem ännu kommit till användning. Man undrar om inte visionen om den semantiska webben borde lära av erfarenheterna från *Cyc*? Berners-Lee framhåller i utkastet från 1998 att sådana generella kunskapssystem redan testats i just *Cyc*-projektet och slår samtidigt fast att den utvecklingen kommit långt. I en artikel från 2006 har förväntningarna tonats ned och nu framhåller man istället att *Cyc* är intressant men har mycket begränsade användningsområden.³¹⁵ Det är lätt att instämma med Winograd i kritiken mot AI:s förhoppningar om användningen av en allomfattande kunskapsbas.

Jämför vi denna vision med TEI:s riktlinjer kan vi konstatera att TEI började arbetet med sin ontologi med en motsatt strategi som innebar att arbeta med frågor om hur vi kan beskriva innehåll, eller information, för utbyte, lagring och bearbetning. Genom att samla expertis från olika vetenskapsområden försökte, och försöker, man inom TEI först komma överens om minsta gemensamma nämnare, nämligen allmängiltiga beskrivningar som flertalet användare har bruk för. Ett exempel är att innehåll kan beskrivas som ordnat i olika delar, så kallade divisioner, oavsett genre.³¹⁶ Dessa allmängiltiga beskrivningar kan sedan kombineras med mer specifika beskrivningar som är speciellt utformade för att passa en genre eller ett forskningsområde. Om inte detta räcker för att beskriva ett visst innehåll är det fullt möjligt att konstruera egna och därmed mer lokalt

³¹⁴ Det vill säga agenter i form av program.

³¹⁵ ”Such [Cyc] a wide-ranging and all encompassing ontology might well have interesting applications, but it clearly won’t scale and its use can’t be enforced.” Shadbolt, Hall & Berners-Lee, ”The Semantic Web Revisited” s. 100.

³¹⁶ För detta ändamål finns elementet <div>, se Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 785 f. <www.tei-c.org/P4X/ref-DIV.html>.

användbara element, då detta tillvägagångssätt inte ger lika stora möjligheter utbyttbarhet som de generella riktlinjerna har.³¹⁷

3.7.1 Metadata för webben

För att den semantiska webben ska kunna erbjuda de förbättringar som önskas och för att förbättra sökning med hjälp av metainformation krävs förstås att dokument förses med metadata enligt fastställda standarder. Nedan tar jag översiktligt upp några av de vanligaste verktygen för metadata på webben.

Prefixet meta används i allmänhet som beteckning för en högre abstraktionsnivå av en beskrivning av något, metadata är alltså information om data och något vi dagligen använder. Exempelvis betyder sifferserien 0704987234 inget speciellt, men genom att lägga till ett bindestreck 070-4987234 kan sifferserien uppfattas som ett mobilnummer. Bibliotekskatalogen är ett annat exempel på stora mängder metadata, där bibliotekets bok- och tidskriftsbestånd ordnas och beskrivs och utan bibliotekskatalogen skulle det bli mycket svårt och tidsödande att hitta det man söker.³¹⁸

På webben finns inget motsvarighet till bibliotekskatalogen och därför kan det vara svårt att hitta det man söker. Det vanligaste sättet att navigera är att använda de sökmotorer som samlar in och indexerar webbsidor och erbjuder dem i olika söktjänster.³¹⁹ Bristen på metadata medför dock att det ofta blir onödigt många sökträffar och en låg precision i sökningen. En mer precis sökning förutsätter dock att den som publicerar på webben tillför metadata till resursen, något som fortfarande är ganska ovanligt.³²⁰

År 1995 startades ett initiativ för att förbättra åtkomst till informationen på webben. Målsättningen var att skapa samförstånd kring ett generellt renodlat och enkelt metadata-system. Systemet kom sedermera att få namnet *Dublin Core Metadata Initiative*, DCMI.³²¹ Här används ett antal specificerade metadataelement för att beskriva de mest centrala aspekterna av webbresurser. För att kun-

³¹⁷ I TEI:s applikation för framtagande av ett schema kan egendefinerade element läggas till de fördefinierade element som ingår i de moduler man väljer att använda. Se <<http://tei.oucs.ox.ac.uk/Roma/>> samt i avsnittet om TEI P5 s. 126 f.

³¹⁸ Ulf Kronman & John Parnefjord, "Resource Description Framework. Metadata för Internet", *Human IT*.4 (1999).

³¹⁹ Den i särklass vanligaste sökmotorn och webbnavigatorn idag är Google <www.google.com>. Google skapades av två studenter vid Stanford universitetet 1998. Sökmotorn baseras på avancerad sidrankningsteknik och visar de mest relevanta träffarna först. Förutom sökmotorn erbjuder Google en rad specialiserade tjänster som t ex Google Scholar och Google Maps. Googles motto är att "organisera hela världens information och göra den användbar" (min övers.) Google används även i vardagligt tal som ett verb, "att googla".

³²⁰ Det enklaste sättet att tillföra metadata är att använda sig av HTML:s grundläggande stöd för metainformation <title> och <meta>. Dessa element är avsedda för beskrivning av dokumentets titel, innehåll, upphovsman och nyckelord och är enbart avsedda för sökning.

³²¹ Namnet Dublin Core för att det första mötet hölls i Dublin, Ohio, USA 1995, DCMI är en officiell standard ANSI/NISO Z39.85 -2007.

<www.niso.org/standards/standard_detail.cfm?std_id=725>.

na tillgodose olika intressegruppers behov är elementen brett definierade. Elementen ingår i tre huvudkategorier: *innehåll* (subject, title etcetera), *upphovsrätt* (creator, publisher etcetera) samt *manifestation* (type, date etcetera).³²² Det är valfritt vilka element som skall användas samt deras inbördes ordning och modellen kan nog lösa en del problem, inte minst för att intressenter kommit överens om en generell standard. Problemet med fördefinierade element kvarstår dock, en sådan icke-generell standard är inte så enkel att utöka eller modifiera och passar därför kanske inte för att beskriva webbresurser som inte kan beskrivas med de viktigaste elementen.

Ett verktyg som utvecklats för den semantiska webben är *Resource Description Framework* (RDF). I RDF:s modell utgår man ifrån att det finns en resurs som är unik och identifierbar. En resurs har ett antal egenskaper och varje egenskap har ett värde. I en grafisk datamodell kallas detta för en *triplett* som innehåller subjekt, predikat och objekt. Vad som helst kan anses vara en resurs: en bok, en tidskrift, en webbsida eller en hel webbplats. En resurs kan beteckna allt som kan tillskrivas en unik identitet, en *Uniform Resource Identifier*, förkortat URI.³²³ Utvecklingen av RDF specifikationen slutfördes 2004 och har därefter integrerats i W3C: s samlingsplats för den semantiska webben.³²⁴

3.7.2 Topic Maps – ämneskartor

En motsvarighet till RDF är ISO-standarden *Topic Maps*³²⁵ som använder sig av en liknande princip för beskrivning av resurser eller ämnen. *Topic Maps* är ett standardiserat verktyg för beskrivning och länkning av innehåll och uttrycks i ett eget filformat: XML *Topic Maps* (XTM).³²⁶ *Topic Maps* har stora likheter med bruket av ”mind maps” och semantiska nätverk för visualisering av kunskapsorganisation. Principerna för *Topic Maps* grundas i praktiken på metoder att använda indexering för gruppering av och referering till valda begrepp:

A traditional index is in fact a map of the *knowledge* contained in a book; it lists the topics covered, by whatever name users might be expected to want to look them up, and includes salient (and *only* salient) references to those topics.³²⁷

Datamodellen är därför, menar Pepper, en anpassning till mediets, det vill säga webbens, förutsättningar för kunskapsorganisation då traditionella index är

³²² Se <<http://dublincore.org/>> samt Kronman & Parnefjord, ”Resource Description Framework. Metadata för Internet”.

³²³ En webbadress, URL, (Uniform Resource Locator) är unika identifierare för webbsidor.

³²⁴ *Semantic Web Activity*, här samlas metoder och verktyg för den semantiska webben, se <www.w3.org/2001/sw/>.

³²⁵ ISO/IEC 13250:2003, <www.y12.doe.gov/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf>.

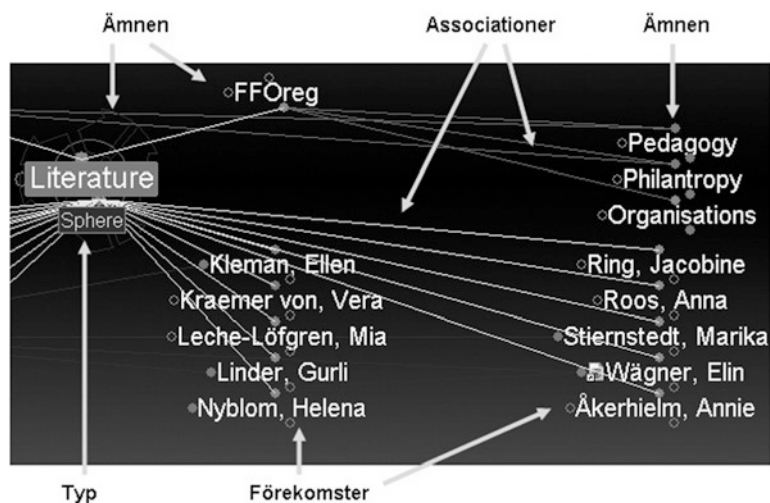
³²⁶ Specifikation vid: <<http://topicmaps.org/xtm/>>.

³²⁷ Steve Pepper, *The TAO of topic maps* (2002). Kursivering i original, tillgänglig vid: <www.ontopia.net/topicmaps/materials/tao.html#d0e222>.

bundna till enskilda dokument.³²⁸ Likheter finns även med så kallade dynamiska tesaurer, där användare kan navigera bland termerna, deras beskrivningar och relaterade termer genom att klicka på dem.³²⁹

Principerna för beskrivning och organisation i *Topic Maps* går i korthet ut på att definiera ett eller flera ämnen (*topics*) som länkas ihop med associationer (*associations*) som relaterar till en godtycklig mängd förekomster (*occurrences*) och principen är alltså snarlik RDF:s datamodell.³³⁰

Ämnen, associationer och förekomster kan vara av en viss sort, det finns även möjlighet att integrera ämnen från olika kartor till en ny karta. Resultatet blir en sorts ”ämneskartor” som beskriver och relaterar valda begrepp. Figuren nedan exemplifierar principen för en sådan ”ämneskarta” där kvinnorna i FFO-databasen har associerats med ett antal ämnen (*topics*).



Figur 12. Exempel på Topic Maps datamodell och grundläggande principer

I figuren ovan kan vi se att ett antal ”ämnen” definierats.³³¹ Det är verksamhetsområden eller sfärer (litteratur, pedagogik, filantropi) som kvinnorna i FFO-databasen associeras med. Dessa ”ämnen” är av typen ”sfär” och varje kvinna, som i modellen illustreras som en förekomst, associeras med en sfär

³²⁸ Ibid.

³²⁹ Se t.ex. <www.visualthesaurus.com/>.

³³⁰ De båda datamodellerna samexisterar, men på behörigt avstånd, framför allt på grund av olikheter i den grundläggande begreppsanvändningen, enligt Lars Marius Garshol. ”Living with topic maps and RDF”. *XML Europe 2003*, May 5-8, London, 2003.

³³¹ Figur 12. avser illustrera utgångspunkterna och idén för Topic Maps, bilden är konstruerad med utgångspunkt i innehåll som märkts med Topic Maps datamodell och som sedan använts i ett program (TheBrain) avsett för ”mindmaps”. Programmet tillgängligt för nedladdning vid:

som i detta sammanhang representerar kvinnans huvudsakliga verksamhetsområde. I exemplet ovan accentueras litteratursfären och de förekomster (kvinnor) som associerats med den. Associationerna mellan förekomster och ämnen kan även typas, i exemplet ovan är de av typen ”primär sfär”. I denna icke-hierarkiska datamodell är det möjligt att associera kvinnorna till fler än en sfär, till exempel en ”sekundär sfär”. I figuren ovan finns även ”organisationer” som också är ett ämne. Därtill bestäms placeringen av ämnen, associationer och förekomster av den applikation som använts för att producera grafen och inte genom datamodellen i sig.³³²

Topic Maps och RDF:s datamodeller har hittills haft en tämligen begränsad spridning av skarpa tillämpningar inom forskning och utbildning.³³³ Jämfört med TEI, som utarbetat riktlinjer och dokumentmodeller, är dessa verktyg yngre, vilket kan vara en orsak till att tillämpningarna ännu inte fått en mer allmän spridning.

Fördelarna med denna datamodell är att den är icke-hierarkisk och därmed tillåter beskrivning och strukturering i fler dimensioner än den hierarkiska dokumentmodellen. Det är dessutom valfritt att antingen tillämpa sin egen ontologi eller en domänspecifik ontologi, något som kan vara en fördel. Denna valfrihet och den icke-hierarkiska strukturen kan även framstå som en nackdel. I likhet med andra icke-hierarkiska modeller växer mängden strukturerad information snabbt och det kan vara svårt att ordna innehållet enligt en överordnad princip. En annan nackdel är att *Topic Maps* datamodell inte stödjer en explicit beskrivning av spatiala aspekterna av innehållet. Det går alltså inte att med hjälp av *Topic Maps* modell beskriva avstånd eller intensitet i en association. Verktöget framstår dock som ett fullgott komplement för att visualisera och strukturera mindre såväl som något större mängder metadata.

3.7.3 Avslutning

I detta kapitel har utvecklingen av märkspråk och den semantiska webben tecknats, framför allt dessas teoretiska utgångspunkter som framställts i en historisk och kritisk belysning. Avsikten har varit att ge en bakgrund till de metoder och verktyg mot vilken studiens explorativa ansats använder sig av. Den hierarkiska dokumentmodellen har även använts som fond för beskrivning och analys av

³³² Jfr diskussionen, vad en datamodell kan beskriva och vilka aspekter av innehållet modellen inte förmår beskriva, i: John Bradley, "Finding a Middle Ground between 'Determinism' and 'Aesthetic Indeterminacy': a Model for Text Analysis Tools", *Literary and Linguistic Computing* 18.2 (2003).

³³³ Ett exempel på en skarp tillämpning för forskning och utbildning är New Zealand Electronic Text Centre, som använder Topic Maps, tillsammans med andra verktyg, för att strukturera och beskriva arkivets textresurser. Se <www.nzetc.org/tm/scholarly/tei-NZETC-About-technology.html>. Med skarpa tillämpningar avses sådana som används i konkret drift och inte enbart tjänar som exempel på användningsområden. På Topic Maps webb <www.ontopia.net> listas sju tillämpningar och på motsvarande webb för RDF <www.w3.org/2001/sw/> listas 24 användarfall samt 19 kommersiella tillämpningar.

andra märkspråk. Framställningen i kapitlet visar att den hierarkiska dokumentmodellen är den modell vars principer är mest genomarbetad, analyserad och kritiskt granskad av de metoder och verktyg jag har presenterat. Den hierarkiska dokumentmodellen är också den som verkligen satts i arbete och har flest tillämpningar inom flera skilda kunskapsområden. Detta är en anledning till att TEI:s riktlinjer valts ut som verktyg för modellering och som huvudsaklig metod för beskrivning av det prosopografiska materialet. I nästa kapitel utforskas därför *Text Encoding Initiatives'* principer och riktlinjer lite närmare.

4 Verktyg för innehållsdesign – Text Encoding Initiative

I detta kapitel presenterar jag *Text Encoding Initiative* som valts som verktyg och utgångspunkt för beskrivning och modellering av det prosopografiska materialet. Kapitlet inleds med en organisatorisk och historisk översikt. Därefter diskuteras principer, utgångspunkter och TEI-schemats moduler. Slutligen beskriver jag kort några nyheter i den senaste utgåvan av riktlinjerna och där är den prosopografiska modulen av särskilt intresse för den här studien.

4.1 Inledning

Den främsta *de facto* standarden för märkning och utbyte av textmaterial inom humaniora och samhällsvetenskap är, som tidigare nämnts, *Text Encoding Initiative*, TEI, som utarbetat och omarbetat riktlinjer sedan 1987. Då träffades ett trettioåtal ledande forskare inom *Humanities Computing* för att utarbeta metodologiska principer för beskrivning och utbyte av data inom humanistisk forskning.³³⁴ TEI:s riktlinjer innehåller rekommendationer för en rad olika discipliner och genrer: prosa, poesi, drama, textkritiska utgåvor, lingvistiska analyser, transkribering av tal, filologiska analyser samt transkribering av otryckta historiska källor såsom manuskript.³³⁵ Men TEI är mer än ett dataformat och en samling riktlinjer. TEI är också ett exempel på ett nytt forskningsområde för studier om textbaserad kommunikation, ett omfattande interdisciplinärt arbete och en forskningsgemenskap. Ellen Mylonas och Allen Renear beskriver utvecklingen efter 1987:

This new community appeared almost from out of nowhere, and spread rapidly around the world, infiltrating a wide variety of professions, disciplines, and institutions in many countries. Like any research community it defines itself with its

³³⁴ Se "About these guidelines" i Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange*. För en översikt av TEI: s historiska bakgrund se, Ide & Sperberg-McQueen, "The TEI: History, Goals, and Future" s. 5-15. Se även Text Encoding Initiative, *The TEI Consortium: A Membership Prospectus* (2001). En sammanfattande översikt ges i: Edward Vanhoutte, "An introduction to the TEI and the TEI Consortium", *Literary and Linguistic Computing* 19.1 (2004).

³³⁵ TEI:s riktlinjer är indelade i moduler som avser täcka en rad olika forskningsområden, se innehållsförteckningen vid <www.tei-c.org/release/doc/tei-p5-doc/en/html/index-toc.html>.

shared interests, concepts, tools, and techniques, and has developed, or is developing, the characteristic organizational apparatus: conferences, journals, research centers, email lists, and the like.

The subject matter of this new research community is textual communication, with the principal goal of improving our general theoretical understanding of textual representation, and the auxiliary practical goal of using that improved understanding to develop methods, tools, and techniques that will be valuable to other fields and will support practical applications in publishing, archives, and libraries.³³⁶

Det är med anledning av det jag ovan framhållit som jag valt att använda TEI som verktyg för beskrivning av, och ordning för, det prosopografiska materialet. Från detta forskningsområde hämtar jag, som inledningsvis nämnts, dessutom den begreppsapparat, de metoder och de verktyg som används i den här studien.

TEI:s riktlinjer har i första hand använts för beskrivning och strukturering av redan existerande texter snarare än som ett verktyg för nyproducerat innehåll. Däri ligger också utmaningen i den här studien. På vilket sätt kan användningen av dessa verktyg, det vill säga TEI:s riktlinjer, som delvis är ämnat för andra ändamål än de jag har, bidra till utveckling av metoder för design av utbildningshistoriskt material? En annan aspekt relaterar mer direkt till själva verktyget. I metodkapitlet redogörs för konflikten med att förena en hierarkisk dokumentmodell med beskrivning av andra egenskaper ett innehåll kan ha. Ett av avhandlingens syften är också att undersöka vilka möjligheter och begränsningar dessa verktyg ger för beskrivning av materialets specifika och generella egenskaper och strukturer. Detta syfte försöker jag uppnå genom att modellera och därigenom sätta verktygen i arbete. I mina försök att undersöka dessa möjligheter men även begränsningar ägnar jag i detta kapitel utrymme åt att granska TEI, dels själva verksamheten och verktygen. Föregående kapitel ägnades därför åt att identifiera och avgränsa de metoder och verktyg som kan vara lämpliga för innehållsdesign av prosopografiska material och i nästa del av avhandlingen undersöks möjliga tillvägagångssätt genom tillämpningar av prosopografiska databaser för utbildningshistoriska studier.

4.2 Riktlinjerna tar form

Initiativet till gemensamma riktlinjer för märkning och lagring av texter, inom humaniora och samhällsvetenskap, växte fram ur önskemål om gemensamma riktlinjer för produktion och handhavande av digitala texter. I kapitel 3 beskrevs de märkspråk som utvecklades under perioden 1960-talet till slutet av 1980-

³³⁶ Ellen Mylonas & Allen H. Renear, "The Text Encoding Initiative at 10: Not Just an Interchange Format Anymore – But a New Research Community", *Computers and the Humanities* 33.1-2 (1997) s. 7.

talet. Dessa avspeglade ofta upphovsmännens specifika forskningsintressen. Därtill var märkspråken vanligen proprietära och därmed inkompatibla med andra applikationer och program, vilket medförde att digitala texterna inte kunde migreras mellan olika plattformar. Några försök under 1970- och -80-talet bland humanistiska forskare att nå överenskommelse om gemensamma riktlinjer misslyckades.³³⁷ 1987 anordnade *Association for Computers and the Humanities* (ACH)³³⁸ en sammankomst där ett trettiotal representanter från textarkiv och humanistiska datacentra kom fram till ett gemensamt förslag. Även om ACH var värdar för mötet bör poängteras att TEI bildades och organiserades av tre organisationer som då företrädde den forskning som fanns inom området: det var den nyss nämnda ACH samt *Association for Literary and Linguistic Computing* (ALLC)³³⁹ och *Association for Computational Linguistics* (ACL)³⁴⁰.

Under det konstituerande mötet 1987 mötet lyckades man inte bara komma överens om nödvändigheten av en gemensam praktik utan även att formulera de viktigaste grundläggande principerna som de kommande riktlinjerna skulle bygga på.³⁴¹ Anledningen att man just vid detta möte lyckades komma överens var, enligt Ide & Sperberg-McQueen, för det första att man hade hunnit skaffat sig tillräckligt mycket erfarenhet om textmärkningens problem och metoder för att över huvud taget kunna identifiera problemen. För det andra deltog fler företrädare för de viktigaste organisationerna och forskningscentra inom området än vid tidigare möten. Den tredje orsaken var att man nu hade tillgång till en metod, nämligen SGML, som kunde användas för den verktygslåda man behövde. Slutligen var situationen för den datorbaserade humanistiska forskningen så kaotisk att det bidrog till att nå konsensus, något som misslyckats vid

³³⁷ Ide & Sperberg-McQueen, "The TEI: History, Goals, and Future"; Allen H. Renear, "Text Encoding" *A companion to digital humanities*, red. Susan Schreiban, et al., (Malden, MA; Oxford: Blackwell Pub., 2004).

³³⁸ *Association for Computers and the Humanities* (ACH) är en av de två ledande intresseorganisationerna inom TEI, *Association for Literary and Linguistic Computing* (ALLC) är den andra. ACH som bildades 1978 är baserat i USA, men har medlemmar från övriga anglosaxiska världen. ACH är även värd för den äldsta diskussionslistan på Internet: HUMANIST <www.princeton.edu/~mccarty/humanist/>. Vidare arrangerar ACH och ALLC tillsammans en av de ledande konferenserna inom området, är utgivare av en av de ledande tidskrifterna inom området (Computers and the Humanities) och samarbetar med liknande organisationer för att främja arbetet med datorbaserad humanistisk forskning. Se vidare <www.ach.org/>.

³³⁹ Den något äldre organisationen ALLC bildades 1973 i syfte att stödja och utveckla datorstött arbete för språk- och litterära studier. ALLC arrangerar som ovan nämnts även konferenser samt ger ut den världsledande vetenskapliga tidskriften (*Literary and Linguistic Computing*) för datorbaserade litterära och lingvistiska studier. ALLC är europeiskt och flera av ledamöterna i organisationens styrelse är från andra länder än de anglosaxiska. Se vidare <www.allc.org/>.

³⁴⁰ *Association for Computational Linguistic* är en intresseorganisation för datorlingvistik, se vidare <www.aclweb.org/>.

³⁴¹ En introduktion till riktlinjernas grundläggande principer ges i Ide & Sperberg-McQueen, "The TEI: History, Goals, and Future" s. 5 f. Se även Lou Burnard, "Report of Workshop on Text Encoding Guidelines", *Literary and Linguistic Computing* 3 (1988). Samt TEI Guidelines (P4), "1.3 Historical background" Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 10 f. Se även <www.tei-c.org/P4X/AB.html#ABTEI>.

tidigare möten.³⁴² Under mötet 1987 enades man om nio grundläggande principer för kodning och utbyte av litterära och lingvistiska texter.³⁴³ Dessa principer kan sammanfattas i fyra punkter:

- tillhandahålla ett standardformat för kodning och utbyte av data³⁴⁴
- tillhandahålla riktlinjer för kodning av texter i detta format
- stödja kodning av alla typer av kännetecken och särdrag i alla sorters texter som används av forskare
- tillhandahålla ett plattformsoberoende filformat³⁴⁵

Den grannliga uppgiften att samla forskare från flera olika vetenskapliga områden och traditioner var en stor utmaning, vilket också avspeglas i de omfattande och komplexa riktlinjerna. Riktlinjer är ett resultat av att TEI-projektet har sitt ursprung i forskningspraktiker inom humaniora och de grundläggande första arbetsinsatserna utfördes med hjälp av företrädare för de ovan nämnda intresseorganisationerna, företrädare som från början positionerade sig och därmed fick stort inflytande.³⁴⁶ Därutöver har projektet hela tiden haft som målsättning att bistå forskningspraktiker snarare än att tillgodose administrativa och publicistiska verksamheter, vilket inneburit att stor hänsyn tagits till forskningens krav på att riktlinjerna skall vara mångsidiga, flexibla och utbyggbara.³⁴⁷ Detta

³⁴² Ide & Sperberg-McQueen, "The TEI: History, Goals, and Future" s. 6. Även Mylonas & Renear, "The Text Encoding Initiative at 10: Not Just an Interchange Format Anymore – But a New Research Community".

³⁴³ Dessa principer brukar benämnas "Poughkeepsie Principles" för att det första mötet ägde rum vid Vassar College i Poughkeepsie, New York.

³⁴⁴ Här avses TEI:s implementering av metaspråket SGML. XML implementerades år 2002 i TEI P4.

³⁴⁵ I minnesanteckningar från ett styrelsemöte i Mars 1991 (Dokument: TEI SCM17, avsnitt AI 3 Literature) hade man på agendan att TEI skulle utveckla program som fungerar med Word och Word Perfect, något som mig veterligen aldrig realiserades. Istället har tillämpningar och integrering skett i Open Office och tidigare framför allt i den UNIX baserade arbetsmiljön Emacs. Under senare år har, förutom integrering av TEI schemat i Open Office, skett en rad plattformsoberoende och TEI-kompatibla program för editering, transformering bearbetning och analyser blivit tillgängliga. Utveckling av, och tillgången till, verktyg har enligt min mening påskyndats och underlättats av öppen källkod och programvara.

³⁴⁶ I det här sammanhanget, anser jag, räckte det inte att vara en av de personer som från början startade projektet för att inta en position. De som skaffade sig positioner var dessutom tongivande, eller åtminstone väl insatta i den organisation eller den genre de företrädde. Därtill behövde de vara beredda att utföra mycket av det praktiska arbetet med att skissa på utkast, sitta i möten skriva minnesanteckningar, marknadsföra metoderna, vara flitig i diskussionslistan, presentera på konferenser etc. De två i särklass mest aktiva personer på listan (Lou Burnard och Sebastian Rahtz, båda vid *Oxford University*) står för nästan en fjärdedel av den totala andelen inlägg (24 procent av totalt 10 000 meddelanden skickade till TEI:s diskussionslista från dess start 1990). Burnard har varit med sedan starten och Rahtz postade sitt första inlägg till listan 1999. E-postlistan tillgänglig vid <<http://listserv.brown.edu/archives/cgi-bin/wa?A0=TEI-L>>.

³⁴⁷ Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* Kapitel 1.2 "Underlying Principles and Intended Use". Burnard & Sperberg-McQueen,

medför att andra viktiga aspekter såsom lättanvända verktyg, begripliga och användbara exempel och utbildningsinsatser fått stå tillbaka, något som också avspeglas i kritiken mot TEI.³⁴⁸

Nancy Ide och Michael Sperberg-McQueen anser, utan att ge några exempel, att det TEI åstadkommit genom riktlinjerna ganska snart skulle visa sig vara värdefullt inte bara för humanistiska och samhällsvetenskapliga forskningsområden utan även för textindustrin generellt. Författarna klargör redan 1995 tydligt vikten av noggranna förberedelser för möjligheter till framtida användning av digitalt innehåll och de poängterar att:

It has become crucial for both research and industry to ensure that any text that is created can be used and, more importantly, reused for any number of applications, including applications which have not yet been imagined or developed.³⁴⁹

Vid TEI:s konstituerande möte för 20 år sedan – vilket i dessa sammanhang är en ansevärd tidsperiod – identifierade alltså ett antal representanter för den datorstödda humanistiska forskningen vikten av genomtänkta strategier för informationsutbyte och datalagring, en fråga som numer finns med på vetenskapliga institutioners och myndigheters agenda.³⁵⁰ För att inte ge intrycket av att ingen verksamhet inom området eller att inga digitala texter eller samlingar existerade före 1987, vill jag påminna om att *Oxford Text Archive* startade sin verksamhet 1976 och att 1965 var tidskriften *Computers and the Humanities* första årgång. I det allra första numret publicerades en kartläggning av litterära verk i maskinläsbar form. Kartläggningen omfattar 25 sidor med information om författare, verk, språk, tillgängligt format, storlek på samlingen (uttryckt i ord, sidor, verser, rader), status på arbetet (slutfört/pågående), om analys utförts etcetera.³⁵¹

Redan 1964 togs frågan om ett standardformat för utbyte och lagring av texter upp till diskussion under en konferens om databearbetning av litterära texter. Vid mötet förutspådde man att inom en nära framtid skulle antalet texter som transkriberats från tryckta till digitala medier öka dramatiskt. Förhoppningarna om att biblioteken skulle kunna tillhandahålla texter på hålkort eller magnetband formuleras också.³⁵² Önskan om att kunna beskriva och ordna

Text Encoding Initiative. Guidelines for electronic text encoding and interchange s. 6 f. Se: <www.tei-c.org/P4X/AB.html#ABDPIU>.

³⁴⁸ Denna kritik tas bl.a. upp på s. 113 f.

³⁴⁹ Ide & Sperberg-McQueen, "The TEI: History, Goals, and Future" s. 6.

³⁵⁰ Till exempel inriktningen i EU:s ramprogram och statskontorets utredningar för svenska myndigheter. *European Interoperability Framework for pan-European eGovernment Services*, Statskontoret, *Öppen programvara*, Statskontoret, *Öppen programvara. Erfarenheter av produkter som bygger på öppen källkod inom förvaltningen*.

³⁵¹ G. Carlson, "Literary Works in Machine-Readable Form", *Computers and the Humanities* 1 (1967).

³⁵² Martin Kay, "Report on an Informal Meeting on Standard Formats for Machine-readable Text" *Literary Data Processing Conference 1964 Proceedings*, red. Jess B. Bessinger, (N.Y.: White Plains, 1965).

digitalt innehåll på ett sätt som tillåter utbyte och flexibel användning har funnits med på agendan inom humaniora i över 40 år.

Intresset för TEI och digitala texter har även lett till att ett nytt forskningsområde kring teorier om märkning och dess praktik växt fram, som Mylonas och Renear beskriver inför TEI:s 10-års jubileum 1997.³⁵³ Frågor om innehållets ordning har framför allt intresserat filosofer och som tidigare nämnts är teorin om innehållets hierarkiska ordning och metateorier om textualitet³⁵⁴ exempel på viktiga bidrag till området.³⁵⁵ Innan jag fortsätter med en beskrivning av innehållet i riktlinjerna ger jag nedan en kort bakgrund till TEI-konsortiets uppbyggnad.”

4.3 Från projekt till konsortium

Fram till år 2000, då TEI-konsortiet bildades, finansierades TEI bland annat av amerikanska donationer och stiftelser för humaniora, EU och kanadensiska forskningsrådet för samhällsvetenskap. Mellan 1987 och 1999 drevs TEI som ett projekt. Till projektet knöts den rådgivande gruppen (*Advisory Board*) med företrädare för en rad olika vetenskapliga sammanslutningar.³⁵⁶ Den rådgivande gruppen hade som uppgift att se till att forskningens behov tillgodosågs i riktlinjerna (snarare än administrativa eller industriella behov). Den rådgivande gruppen hade även i uppgift att tillgodose att disciplinernas vägledande teorier och analysmetoder för texthantering avspeglades i riktlinjerna samtidigt som de skulle beakta de speciella behov och önskemål användarna kunde komma att ha. Dessa uppgifter är förstås inte helt oproblematiska att kombinera. TEI:s riktlinjer har därför fått utstå en hel del kritik för att vara för omfattande, kom-

³⁵³ Mylonas & Renear, ”The Text Encoding Initiative at 10: Not Just an Interchange Format Anymore – But a New Research Community”.

³⁵⁴ Renear m.fl. använder här termen *textuality*. Enligt Mats Dahlström innefattar begreppet textualitet: ”[e]tt sätt att via text och bärare materialisera eller ’mediera’ det immateriella innehållet, som också sätter ramar för hur texterna produceras, sprids, framträder, upplevs och bedöms.” Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 72.

³⁵⁵ Frågor om innehållets hierarkiska ordning och dokumentstruktur behandlas bl.a. i James Coombs, Allen H. Renear & Steven DeRose, ”Markup systems and the future of scholarly text processing”, *Communications of the ACM* 30.11 (1987); Renear, ”Out of Praxis: Three (Meta)Theories of Textuality”; Renear, Durand & Mylonas, ”Refining our notion of what text really is”. Frågor om textualitet behandlas i Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion*.

³⁵⁶ Dessa 15 organisationer företrädades i TEI:s *Advisory Board* fram till TEI-konsortiet bildades år 2000: American Anthropological Association, American Historical Association, American Philological Association, American Philosophical Association, American Society for Information Science, Association for Computing Machinery, Special Interest Group for Information Retrieval, Association for Documentary Editing, Association for History and Computing, Association Internationale Bible et Informatique, Canadian Linguistic Association, Dictionary Society of North America, AAP Electronic Publishing Special Interest Group, International Federation of Library Associations and Institutions, Linguistic Society of America samt Modern Language Association. Se <www.tei-c.org.uk/Vault/ABlist.html>.

plicerade och ordrika samt innehålla för få eller dåliga exempel som inte speglar disciplinernas olika, och inte sällan motstridiga, teoretiska perspektiv.³⁵⁷

TEI har, fram till förra året, publicerat fyra versioner av riktlinjerna och till hösten 2007, lagom till TEI:s 20-års jubileum släpptes den omarbetade femte upplagan.³⁵⁸ Redaktörskapet för TEI:s utgivning av riktlinjer delas, för närvarande finns en redaktör i Europa och en i USA.³⁵⁹

Löpande arbetsinsatser och utvecklingsarbetet i TEI utförs på liknande sätt som den distribuerade arbetsmodell (basarmodellen) som beskrivits ovan. Arbetsfördelningen sköts genom värdskap som löper i fyraårsperioder där ansvaret för organisationen fördelas mellan fyra universitet (eller andra forskningsinstitutioner).³⁶⁰ Arbetsgrupperna är en viktig del i TEI:s organisation, likaså de användare som sedan använder riktlinjerna i verkligheten så att säga. Problem uppenbaras oftast inte förrän märkningen pågått ett tag. Arbetsgrupperna utför det tekniska "grovjobbet" vilket innebär att de utarbetar nya principer, moduler och element eller reviderar existerande moduler. Speciella intressegrupper finns också för de användare som inte vill eller kan medverka direkt i den tekniska utvecklingen. Den allmänna diskussionslistan är ett viktigt forum framför allt för lärdomar av andras problem och lösningar, inte så mycket för att påverka. Till det behövs ett mer initierat engagemang.

Hur ser då medlemsfördelningen ut? Vilka projekt och institutioner har valt att bli medlemmar? I tabellen nedan tar vi oss en närmare titt på TEI:s medlemmar som för närvarande är 93 till antalet.³⁶¹ Vad gäller den geografiska fördelningen dominerar USA, med nästan hälften av alla medlemmar. Därefter är det flest medlemmar från Storbritannien. Den resterande dryga tredjedelen medlemmar är spridda över framför allt Europa, undantaget några medlemmar från Kanada, Nya Zeeland och Taiwan.

Genom att kategorisera verksamheter och institutioner i sju typer av medlemmar i TEI illustreras att satsningar på specifika digitala samlingar är vanligast i medlemsfördelningen.³⁶² Dessa digitala samlingar har vanligen ett universitet

³⁵⁷ Riktlinjerna omfattar cirka 600 olika element och nästan 1300 sidor i den tryckta versionen, (TEI P3 (P som i Proposal) som publicerades 1994 omfattar 1292 sidor och TEI P4 som publicerades 2002 omfattar 1067 sidor. Kritiken mot riktlinjerna och principerna har levererats både inom och utom TEI projektet, se vidare s. 113 f.

³⁵⁸ Första utkastet TEI P1 (P för proposal) gavs ut 1990 för revision och kritisk granskning, P2 gavs ut i 1993 och 1994 publicerades P3, som var den första formella publikationen. P4 publicerades 2002 och innebar att TEI:s riktlinjer nu fanns i XML. P1-P3 finns tillgängliga vid <www.tei-c.org.uk/Vault/Vault-GL.html>. De publicerade utgåvorna P3 och P4 finns tillgängliga vid <www.tei-c.org.uk/Guidelines2/>.

³⁵⁹ Vilka dessa redaktörer är, se not 371 s. 113.

³⁶⁰ Brown University, Oxford University, University of Virginia samt tre forskningsinstitutioner (ATILF, INIST, Loria) i Nancy i Frankrike, se vidare <www.tei-c.org/About/board.xml>.

³⁶¹ Uppgifterna avser endast institutionellt medlemskap.

³⁶² Klassificering är gjord av mig och avser särskilja olika typer av institutioner och organisationer men även nivåer (institution, fakultet, universitet).

eller annan statlig som huvudman och sysslar oftast inte med undervisning.³⁶³ Syftet med dessa samlingar är vara ett stöd för forskning och att bevara kulturella objekt. Vissa av universiteten, vanligen i USA, som ansvarar för digitala samlingar, men även framstående humanistiska centra för forskning och utveckling, har valt att bli medlemmar på universitetsbasis. De amerikanska universitetsbiblioteken är också relativt sett väl representerade i samlingen av medlemmar, med 12 medlemmar av totalt 93. Universiteten som sådana är även företrädare av de fakulteter och institutioner som finns med i medlemsregistret, där vissa institutioner är specifikt inriktade på utbildning snarare än forskning. Några av de medlemmar som i praktiken har stort inflytande i TEI är de centra som är specifikt inriktade på forskning kring och utveckling av digitala samlingar och datorstödd humanistisk forskning och undervisning.³⁶⁴ Vissa av medlemmarna kategoriseras som intresseorganisationer som här främst företräder arkiven, biblioteken samt några verksamheter inom elektronisk publicering. Samtliga intresseorganisationer är hemmahörande i USA.

³⁶³ Digital samling används i betydelsen textarkiv eller digitalt arkiv, det vill säga en samling texter som gjorts digitalt tillgängliga.

³⁶⁴ Dessa center benämns F/U center i tabellen och avser antingen avdelningar eller centrum-bildningar som bedriver forsknings- och utvecklingsverksamhet och som oftast är baserade vid olika universitet. Några forskargrupper ingår även i denna kategori. Till de främsta centren inom området räknas framför allt: Oxford Text Archive; Centre for Computing in the Humanities vid King's College i London, Scholarly Technology Group for Humanities, Brown University; Centre for Textual Criticism and Document Studies i Belgien samt de humanistiska centren för textanalys som finns i Kanada (McMaster University, University of Victoria, University of Alberta, University of Toronto, Université de Montréal, University of New Brunswick).

Tabell 1. Jämförelse TEI:s medlemmar per land och typ av medlemsinstitution. Absoluta tal.

	Biblio- tek	Digital samling	FoU center	Fakultet	Intres- seorga- nisation	Institu- tion	Univer- sitet	Totalt
Australien			1					1
Belgien		1	1					2
Bulgarien		1		1		1		3
Danmark		1						1
Frankrike			1	1		3		5
Holland		1		1		2		4
Indien		1						1
Italien						1	1	2
Kanada	1	1		1				3
Norge		2					1	3
Nya Zeeland			1					1
Polen	1							1
Serbien		1						1
Slovenien				2		1		3
Spanien	1							1
Storbritannien	1	5	4				2	12
Sverige		1						1
Taiwan		1						1
Tjeckien		1					1	2
Tyskland			1					1
Ukraina				1				1
Ungern							1	1
USA	12	5	5	1	8		11	42
Totalt	16	22	14	8	8	8	17	93

Anm: Uppgifterna i tabellen är från TEI:s medlemsregister på webben den 10 juli 2007, se <www.tei-c.org/Membership/current.xml>.³⁶⁵

Med utgångspunkt i medlemsregistret kan man även konstatera att huvuddelen av medlemmarna återfinns bland stödjande verksamheter och forskningscentra snarare än inom utbildningsområdet.

Vad man inte kan utläsa av uppgifterna ovan är i vilken omfattning de digitala samlingarna används i undervisningen. Ett mindre antal av medlemmarna i TEI tillhandahåller utbildning inom området såsom digital informationshantering, textmärkning och digital publicering. Detta anser jag avspeglar TEI:s fokus på forskningsverktyg och stöd, snarare än utbildningsfrågor.³⁶⁶ Vad som kan

³⁶⁵ Den enda svenska medlemmen är Forskningsprogrammet Digital Litteratur, Uppsala universitet, se <www.skeptron.uu.se/broady/dl/index.htm>.

³⁶⁶ Sedan några år tillbaka finns en speciell intressegrupp (*Special Interest Group, SIG*) för utbildning i och om TEI. Se <www.tei-c.org/Activities/SIG/Education/>.

konstateras är att de allra flesta medlemmar har utvecklat, och utvecklar, digitala samlingar där TEI:s riktlinjer används för beskrivning, lagring och utbyte av innehållet. Såväl kompetens som innehåll är ju dessutom en förutsättning för att senare bedriva någon form av utbildning om digital informationshantering, textmärkning och digital publicering. Det är ingen tillfällighet att *Centre for Computing in the Humanities* vid *King's College* i London erbjuder utbildningsprogram från grundläggande upp till forskarutbildningsnivå. Där finns en mångårig erfarenhet av forskning och forskningsstödjande verksamhet samt tillgång till flertal riktigt stora och komplexa digitala samlingar av olika slag.³⁶⁷ Andra verksamheter med mångårig erfarenhet har istället fortsatt sin i huvudsak stödjande verksamhet och erbjuder endast kortare utbildningsinsatser utan examina. Ett exempel på det är *Oxford Text Archive (OTA)* vid Oxforduniversitetet. OTA saknar varken kompetens eller digitalt material i sitt omfattande arkiv, men OTA sorterar organisatorisk under universitets centrala data- och serviceavdelning³⁶⁸ och har som uppgift att stödja forskning och utbildning. Därför erbjuder inte OTA någon formell utbildning eller examina, däremot ges kortare utbildningar i digital informationshantering, textmärkning och arkivkunskap.

Genom att jämföra den geografiska fördelningen mellan de projekt som anmält sig till TEI:s projektdatabas och kan antas använda sig av riktlinjerna och mellan TEI:s betalande medlemmar i tabellen ovan, ser vi att spridningen är något annorlunda än för medlemsstatistiken. De projekt som anmält sig till TEI:s register är digitala samlingar i en rad olika genrer, såsom prosa, poesi, klassiska kanoniska texter, manuskriptsamlingar, språkkorpora eller lexikon. Vissa av projekten är stora, väl finansierade och sedan länge etablerade digitala samlingar, andra projekt är mindre och mer avgränsade.

USA:s och Storbritanniens dominans bland projekten är inte lika markant som i medlemsregistret även om dessa länder är värdar för över hälften av de 137 projekten i registret. Norge är ett av de länder där TEI-aktiviteten är hög, sju av projekten är norska. Projektregistret visar även att några nya länder tillkommit, bland annat Australien, Finland, Japan, Rumänien, Schweiz, Spanien och Österrike.

TEI:s projektregister innehåller uppgifter från 1996 och i tabellen nedan kan vi se att närapå hälften av samtliga projekt registrerades under den första treårsperioden. Efter 1998 har tillväxten av nyanmälda projekt istället mattats av. Ett något förvånande resultat kan man tycka då märkverktygen och metoderna blivit alltmer lättillgängliga under de senaste åren. Dessutom förs diskussioner om nödvändigheten av ett hållbart filformat och möjligheter till plattformsober-

³⁶⁷ Se <www.kcl.ac.uk/schools/humanities/cch> för ytterligare information.

³⁶⁸ OTA är en del av *Research Technologies Service (RTS)* som i sin tur sorterar under *Oxford University Computing Service (OUCS)*. *Research Technologies Service* bistår forskningen med expertis inom informationshantering, öppen programvara, avancerad gridteknik för naturvetenskaperna, den engelska nationella språkkorpusen (BNC), virtuella forskningsmiljöer, etc., se <www.oucs.ox.ac.uk/rts/>.

roende lösningar numer inte bara inom gruppen initierade forskare utan även på en allmänpolitisk nivå, till exempel i IT-standardiseringsutredningen.³⁶⁹

Tabell 2. *Fördelning av världsländer för digitala samlingar som använder TEI:s riktlinjer under tidsperioden 1996-2007. Absoluta tal.*

Världsländer	1996-1998	1999-2001	2002-2004	2005-2007	Totalt
Australien	2	1			3
Belgien			1	1	2
Danmark	1				1
Finland	1				1
Frankrike	2		2	1	5
Holland	3		2		5
Irland	2				2
Italien	2		2		4
Japan	2	3			5
Kanada	4	1	3		8
Kroatien				1	1
Norge	3	2	2		7
Nya Zeeland				1	1
Rumänien			1		1
Schweiz	1				1
Serbien	1				1
Slovenien	1	1		1	3
Spanien		1			1
Storbritannien	7	5	5	1	18
Sverige	1		1		2
Taiwan				1	1
Tyskland		3		2	5
Ungern			1		1
USA	28	12	11	6	57
Österrike				1	1
Totalt	61	29	31	16	137

Anm: Uppgifterna i tabellen är från TEI:s projektregister på webben i juni 2007, se <www.tei-c.org/Applications/>.

TEI-gemenskapen kan genom den ovan presenterade aktivitetsfördelningen möjligen framstå som en ganska begränsad skara entusiaster som genom omständligt arbete iordningställer digitala samlingar som används av ett fåtal initierade forskare. Det beror förstås vilka verksamheter man jämför med men det faktum att så många arbetat så länge, inom så många discipliner, mot samma mål torde vara något ganska unikt, åtminstone inom humaniora:

In the end well over 100 scholars from many disciplines, professions, and countries, were active in the core group that was developing the Guidelines. This

³⁶⁹ Se avsnitt 1.1.3 s. 30 f.

alone would make the Guidelines an exemplary achievement in collaboration, one on a scale fairly rare in the history of the humanities.³⁷⁰

Som tidigare nämnts har TEI:s riktlinjer utvecklats av en samling engagerade forskare inom många olika discipliner.³⁷¹ Man kanske kan tycka att denna skara är liten men man bör då betänka att bakom de institutionella medlemmar, som uppgifterna ovan beskriver, finns förstås många individer per medlemsinstitution. Hur många är svårt att säga då dessa uppgifter inte finns tillgängliga. Inte desto mindre är TEI en inflytelserik intresseorganisation som medverkat och påverkat utvecklingen inom ett flertal områden, och då inte enbart för litterära utgåvor eller språkkorporusar, utan även informationshantering på webben. Inte minst inom World Wide Web Consortium (W3C) där bland annat en av TEI:s förgrundspersoner, Michael Sperberg-McQueen, har en ledande position sedan mitten av 1990-talet och har sedan dess lett arbetet med utvecklingen av XML.³⁷²

Som ovan nämnts avser riktlinjerna spegla den mångfald och den komplexitet som utmärker de forskningsområden som medverkat i utformningen av dem. Resultatet av det mångåriga samarbetet beskrivs övergripande nedan. Riktlinjerna innehåller många hundra aspekter på hur innehåll och delar av innehåll kan beskrivas inom en rad genrer och för flera discipliner. Vidare innehåller riktlinjerna verktyg för att dokumentera hur beskrivningen och tolkningen gått till. Därför koncentrerar jag beskrivningen nedan genom att redogöra för riktlinjernas modulära uppbyggnad i relation till den här studien.³⁷³

4.4 Riktlinjernas utgångspunkter och principer

TEI:s riktlinjer bygger på antagandet om att "alla" texter delar ett antal gemensamma kännetecken eller innehållsliga delar, som logiskt relaterar till varandra.

³⁷⁰ Mylonas & Renear, "The Text Encoding Initiative at 10: Not Just an Interchange Format Anymore – But a New Research Community" s. 4.

³⁷¹ Lou Burnard och Michael Sperberg-McQueen anses vara bland de som betytt allra mest för TEI. Mylonas & Renear uppmärksammar detta: "The large and collaborative nature of the TEI also makes it hard to identify and distribute credit – without sliding down a long slippery slope. But as this is an introduction to a celebratory volume, we will not entirely relegate credit to a footnote: so we salute here the two editors, Michael Sperberg-McQueen (University of Illinois at Chicago), and Lou Burnard (Oxford University) who were – and still are – the uncontested intellectual leaders of this enterprise." Ibid. Lou Burnard (Oxford University) är kvar på posten som europeisk redaktör. Sperberg-McQueen har ersatts av Syd Bauman (Brown University) som USA:s redaktör för utgivningen av TEI:s riktlinjer.

³⁷² För mer information om extern märkning och länkning i XML, se: *TEI Workgroup on Stand-Off Markup, XLink and XPointer*, <www.tei-c.org/Activities/Workgroups/SO/>.

³⁷³ Som grundliga presentationer av hur TEI:s riktlinjer kan användas är följande publikationer en bra utgångspunkt: Burnard, O'Brien O'Keefe & Unsworth, red, *Electronic textual editing*, Hockey, *Electronic texts in the humanities: principles and practice*, Nancy Ide & Jean Vâronis, red, *Text encoding initiative: background and context* (Dordrecht; London: Kluwer Academic, 1995); Schreibman, Siemens & Unsworth, red, *A companion to digital humanities*.

En annan utgångspunkt är det pluralistiska perspektiv som ovan presenterats samt forskningens krav på explicita och rigorösa beskrivningar av de innehållsliga objekten.

Sperberg-McQueen och Burnard framhåller att endast ett rigoröst utarbetat schema kan uppnå forskningens krav på generalitet och samtidigt ge möjligheter till automatiserad texthantering.³⁷⁴ Därtill finns förstås behovet av beskrivning av innehållet utifrån olika discipliners teorier och konceptuella modeller. Ett och samma innehåll tolkas och ges olika vikt beroende på perspektiv. TEI-schemat innehåller därför metoder och verktyg för beskrivning av flera vyer, eller dimensioner, av texten, något som man avser uppnå genom dokumentmodellens olika moduler.³⁷⁵

Mylonas och Renear framhåller, angående svårigheterna att nå konsensus mellan olika discipliner och forskningsområden, att det är lätt att tala om interdisciplinärt samarbete och kommunikation mellan intellektuella skiljelinjer men en helt annan sak att verkligen genomföra det. Hindren och problemen som skall övervinnas är många innan konsensus kan nås:

Not only do different disciplines have quite different interests and perspectives, but also, it seems, different conceptual schemes: fundamentally different ways of dividing up the world. What is an object of critical contest and debate for one discipline is theory-neutral data for another, and then completely invisible to a third. What is structured and composite for one field is atomic for another and an irrelevant conglomeration to a third. Sometimes these differences of perspective occurred within a discipline, sometimes across historical periods of interest, sometimes across national or professional communities. Practices that would seem to have much in common could vary radically – and yet have enough in common for differences to be a problem! And even where agreement in substance was obtained, disagreements over nuances, of terminology for instance, could derail a tenuous agreement.³⁷⁶

Dessa svårigheter har också varit föremål för en del av den kritik som riktats mot TEI:s riktlinjer. Kritiken mot TEI har framför allt rört två stora frågeområden:

- Den strukturorienterade modellen för hierarkiskt ordnade innehållsliga objekt (OHCO) med vars hjälp innehållets struktur identifieras och beskrivs.
- Den deskriptiva märkningens problematik där det ifrågasätts om de strukturer i texten som beskrivs existerar *in situ* eller om dessa strukturer konstitueras med hjälp av intersubjektiva förevändningar?

³⁷⁴ Michael Sperberg-McQueen & L. Burnard, "The Design of the TEI Encoding Scheme" *Text encoding initiative: background and context*, red. Nancy Ide & Jean Vâeronis, (Dordrecht; London: Kluwer Academic, 1995) s. 18.

³⁷⁵ Benämningen TEI-schema används för den formella specifikationen av TEI:s riktlinjer. Det kan dock förekomma att dessa två begrepp används synonymt.

³⁷⁶ Mylonas & Renear, "The Text Encoding Initiative at 10: Not Just an Interchange Format Anymore – But a New Research Community" s. 5.

Textsociologen Jerome McGann³⁷⁷ kritiserar TEI för att pragmatiska målsättningar – som att enkelt kunna söka, transformera och bearbeta texterna i stor utsträckning – styrt utformningen av schemat. McGann hävdar att genom utformningen av en modell som fokuserar på beskrivning av objekt och deras relationer har man tagit fram verktyg som ligger utanför många forskares intresseområden. Schemat är bara så objektivt som systemets *à priori* kategorier medger. Sådana verktyg kan därför endast användas för mycket specifika ändamål.³⁷⁸ I detta sammanhang skulle det betyda att en sådan modell endast fungerar enligt en lokal diskursiv logik men att den inte kan vara särskilt användbart på materialet i den här studien.

Claus Huitfeldt³⁷⁹ framhåller att de diskussioner som följt TEI:s strävan efter att inkorporera så många vetenskapliga teorier varit stimulerande. Men Huitfeldt kritiserar, liksom Paul Fortier, även den deskriptiva märkningen som är en grundläggande utgångspunkt och ett arv från SGML.³⁸⁰ Huitfeldt påpekar att vi inte kan beskriva något utan att tala om vad vi beskriver, därtill är det är omöjligt att fånga en texts samtliga möjliga strukturer och dimensioner. Därför är han också kritisk till den hierarkiska dokumentmodell som ger upphov till en rad problem för beskrivning av till exempel strukturer. Vad som är struktur och vad som är innehåll beror på *hur* texten analyseras. En text innehåller många olika typer av strukturer: fysisk, komposition, narrativ och grammatisk struktur. Och eftersom TEI baseras på SGML, och därmed den hierarkiska dokumentmodellen, tvingas textens många dimensioner in i *en* överordnad struktur.³⁸¹ Vilken struktur skall då anses vara överordnad de övriga och därmed beskrivas som ”den” logiska? Enligt TEI:s riktlinjer är frågan om vilken struktur som skall vara överordnade upp till användarna att besluta om då tolkningen styrs av

³⁷⁷ Enligt textsociologer såsom McGann, McKenzie och Chartier är verkets sociala och materiella genealogi en förutsättning bland andra för verkets tillblivelse och tolkning. Chartier, *Böckernas ordning: läsare, författare och bibliotek i Europa mellan 1300-tal och 1700-tal*; Chartier, *Forms and meanings: texts, performances, and audiences from codex to computer*; Jerome J. McGann, *Radiant textuality: literary studies after the World Wide Web* (New York; Basingstoke: Palgrave, 2001); Jerome J. McGann, ”Making texts of many dimensions” *A companion to digital humanities*, red. Susan Schreibman, et al., (Malden, MA; Oxford: Blackwell Pub., 2004); McKenzie, *Bibliography and the sociology of texts*. Mats Dahlström identifierar textsociologins kritiska förhållningssätt: ”[...] söker textsociologin sålunda verket i den text eller de texter som gemensamt upprättades av såväl författare som förläggare, tryckare, illustratörer, kommentatorer, kritiker och publik under ett bestämt historiskt avsnitt.” Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 157 f.

³⁷⁸ McGann, ”Making texts of many dimensions” s. 202.

³⁷⁹ Claus Huitfeldt har varit vetenskaplig ledare för Wittgensteinarkivet vid universitetet i Bergen, ett filologiskt forskningscenter och digitalt arkiv med rigoröst märkta (både normaliserade och diplomatiska versioner) av Wittgensteins komplexa, mångdimensionella och opublicerade manuskript och anteckningar. Se <<http://wab.aksis.uib.no/>> (Jag noterar att flertalet av de personer som refereras till i avhandlingen återfinns i listan över Wittgensteinarkivets gästforskare).

³⁸⁰ Claus Huitfeldt, ”Multi-Dimensional Texts in a One-Dimensional Medium”, *Computers and the Humanities* 28 (1995) s. 237. Paul Fortiers och TEI Literature Working Group kritik från 1991 finns tillgängligt i TEI:s arkiv som ”The TEI Guidelines (version 1.1 10/90): A Critique”)dokument ai3w05), se <www.tei-c.org/Support/Learn/tei_bibliography.xml> eller <<ftp://ftp.ifi.uio.no/pub/SGML/TEI/AI3W5.DOC>>.

³⁸¹ Ibid.

den TEI-modul som väljs för beskrivning av strukturen. Dock kvarstår det faktum att TEI:s grundläggande moduler anses vara generella nog för att beskriva alla slags texter inom alla slags genrer.³⁸²

En metod att överkomma den hierarkiska ordningen är att tillåta överlappande hierarkier, något man kan uppnå med olika metoder. En sådan metod går ut på att alla element skall kunna överlappa alla andra element vilket medför att den hierarkiska ordningen sätts ur spel. En annan metod är extern märkning, som ovan nämnts, och som tillåter att multipla beskrivningsdimensioner lagras utanför måldokumentet.

Med utgångspunkt i kritik mot SGML hierarkiska struktur konstruerade Huitfeldt och andra vid Wittgenstienarkivet ett eget kodsysteem som tillät kodning av överlappande hierarkier på ett annat, mindre ordrikt och ”enklare” sätt än vad SGML då (1990-tal) kunde erbjuda.³⁸³

Ett problem med användning av överlappande element är att de flesta verktyg inte kan validera kodningen då det inte finns någon förutbestämd struktur, hierarkisk eller annan, att kontrollera mot. Detta medför därmed att inget ”kvitto” kan ges på att innehållet är beskrivet enligt den modell man tänkt sig. Därtill gör det valideringen av märkningen mindre förutsägbar, något som begränsar utbyte och återanvändning i andra miljöer än där de utformats.

Kritiken angående problemen med OHCO-modellen har gett upphov till flera olika projekt och teoretiska modeller, dock med begränsad spridning.³⁸⁴ För TEI:s räkning finns sedan 2003 en speciell intressegrupp som utreder frågan.³⁸⁵ Huitfeldt intar dock en pragmatisk hållning när han menar att ibland måste vad som är praktiskt genomförbart få råda över principerna och att viss intersubjektivitet är möjlig:

There are no facts about a text which are objective in the sense of not being interpretational. However being interpretational does not mean being entirely sub-

³⁸² Jfr. inledningen till avsnitt 4.4 s. 113.

³⁸³ Multi-Element Code System (MECS), <<http://helmer.aksis.uib.no/claus/mecs/mecs.htm>>, MECS är enklare i teknisk bemärkelse då schemat tillåter ett godtyckligt antal innehållsliga objekt beskrivas utan krav på nästling. Schemat är dock inte enkelt i någon annan betydelse, utan kräver att den som kodar innehållet kan hålla reda flera beskrivningsdimensioner av texten samtidigt. En reviderad och uppdaterad version av MECS, TexMecs, (Trivially extended MECS). För ytterligare information om detta schema, se <<http://decentius.aksis.uib.no/mlcd/>>.

³⁸⁴ Vid konferensen Extreme Markup Languages har sedan 2001 flera papers behandlat frågor om överlappande hierarkier. Se t.ex. Michael Sperberg-McQueen. ”Rabbit/duck grammars: a validation method for overlapping structures”. *Extreme Markup Languages 2006*, Montréal, Québec, 2006; Michael Sperberg-McQueen. ”Representation of overlapping structures”. *Extreme Markup Languages 2007*, Montréal, Québec, 2007; Jeni Tennison & Wendell Piez. ”The Layered Markup and Annotation Language (LMNL)”. *Extreme Markup Languages 2002*, Montréal, Québec, 2002. Sperberg-McQueen presenterar i 2006 års paper (se ovan) en formell modell för hur överlappande hierarkier kan valideras genom att tilldela markörerna fyra olika syntaktiska typer. Varje element/markör tillhör minst en typ av syntax, som anger hur elementet kan användas.

³⁸⁵ TEI Overlapping Markup SIG, <www.tei-c.org/Activities/SIG/Overlap/>.

jective – there are some things about which all competent readers agree, at least for practical purposes.³⁸⁶

En annan ståndpunkt intar litteraturvetaren Ian Lancashire³⁸⁷ som kritiserar TEI för att man tar för givet att den som kodar texten har tillräckliga kunskaper om författaren eller verket för att kunna tolka innehållet. Lancashire är, tvärt emot Huitfeldt, av den åsikten att märkning och tolkning av text är två olika företeelser. Märkningen i sig behöver inte innefatta någon tolkning, utan tolkningen bör helt överlåtas åt den forskare som använder och analyserar texterna.³⁸⁸ Enligt denna logik skulle den uppsättning element som används för att beskriva texternas mest grundläggande struktur (styckesindelning, rubriker, meningar, ord etcetera) vara tillräcklig. Men arbetet med att utforma TEI-schemat har visat att detta inte är önskvärt eller ens möjligt utan att förlora för mycket konkretion. Man har, som tidigare nämnts, inom olika genrer skilda uppfattningar om den primära ordningen av de objekt som utgör innehållet. Däremot är det möjligt att – genom benämning av elementen, konstruktionen av schemat och genom att använda TEI-headerns element – tillföra information om innehållet som explicit visar vilket teoretiskt perspektiv som väglett märkningen.³⁸⁹

Ett exempel på genre där beskrivning av innehållets multipla ordningar är vanliga är poesin. Från början (i TEI P1) fanns inte möjlighet att infoga beskrivning av metriken samtidigt med en typografisk märkning. Men efter kritik från arbets- och referensgrupper lanserades efter ett par år (i TEI P2, 1993) en ny modul med möjligheter till metrisk märkning oberoende av versradernas struktur, vilket medförde att beskrivningar (och analyser) av versernas relationer till typografins (och vise versa) blev möjliga.³⁹⁰

McGann finner det lingvistiska och informationsbaserade perspektivet på text, som han anser TEI är exempel på, oanvändbart på genrer som till exempel poesi. Att i beskrivningen av texterna överordna en logisk struktur de materiella aspekterna urvattnar den process som intresserar de flesta litteraturvetare. Vad

³⁸⁶ Huitfeldt, "Multi-Dimensional Texts in a One-Dimensional Medium" s. 237.

³⁸⁷ Ian Lancashire, litteraturprofessor vid University of Toronto (renässans- och äldre engelskt drama samt poesi och bibliografiska studier). Specialist på, och utvecklare av, verktyg för textanalys, t.ex. TACT (*Text-Analysis Computing Tools*).

³⁸⁸ Ian Lancashire. "Early Books, RET Encoding Guidelines, and the Trouble with SGML". *Scriptorium conference*, November, Calgary, 1995.

³⁸⁹ För exempel på hur dessa uppgifter registreras se, Bilaga 6. Registrering av metadata i TEI-header. Tillgänglig vid CD:/bilagor/bilaga6.pdf se även CD:/xml/khls-post.xml för hela märkning av hela posten i XML-format.

³⁹⁰ I P2 infördes elementet <seg> med tillhörande attribut som kan användas för att markera grupper av ord i relation till versraderna samt elementet <caesura> för markering av "pauser" i versraderna. Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 221 f. & 745. Se även <www.tei-c.org/P4X/VE.html#VESE> för exempel. Se även Michael Sperberg-McQueens inlägg i TEI:s diskussionslista (TEI-L) daterat 1993-11-26 (subject: new chapter: verse) vid <<http://listserv.brown.edu/archives/cgi-bin/wa?A0=TEI-L>>.

McGann är ute efter är konflikten mellan TEI:s strävan efter tydlighet och de poetiska uttryckens strävan efter att maximera dubbeltydigheten genom rika och dubbelbottnade betydelser i de språkliga uttrycken.³⁹¹

Sperberg-McQueen och Burnard försvarar TEI:s pluralistiska perspektiv och intersubjektiva ansats och när de framhåller att fullständig omfattning och objektivitet är omöjlig att uppnå:

It is not possible to define a tag set – at least not a useful one – in a purely atheoretical way, but it is possible, given a specific set of sufficiently explicit theoretical approaches to a field, to define a tag set which allows adherents of competing theories to encode textual features they find of interest.³⁹²

Man kan ana att lärdomar från det initiala intensiva arbetet med utformningen av riktlinjerna och samarbetet mellan disciplinerna kom att vägleda förutsättningar för riktlinjernas fortsatta interdisciplinära användning, utveckling och vidare spridning. Om man inte tar hänsyn till vad forskare och lärare inom de discipliner som riktlinjerna vänder sig till anser vara användbart kommer de heller inte att använda verktygen. TEI-schemat, menar jag, skiljer sig därför från vissa andra mer dataorienterade metoder och verktyg där genrespecifika frågor eller teoretiska perspektiv inte prioriteras eller ens uppmärksammas.³⁹³

4.5 Modulerna

Nedan ger jag en översiktlig presentation av TEI-schemats design och innehåll. Jag gör inga anspråk på att här beskriva samtliga aspekter av schemat eller de forskningspraktiker som det representerar, för en sådan uppgift är schemat alldeles för komplext och omfattande. Förutom en grundläggande översikt presenterar jag framför allt sådant som är direkt eller indirekt relaterat till den här studien.³⁹⁴

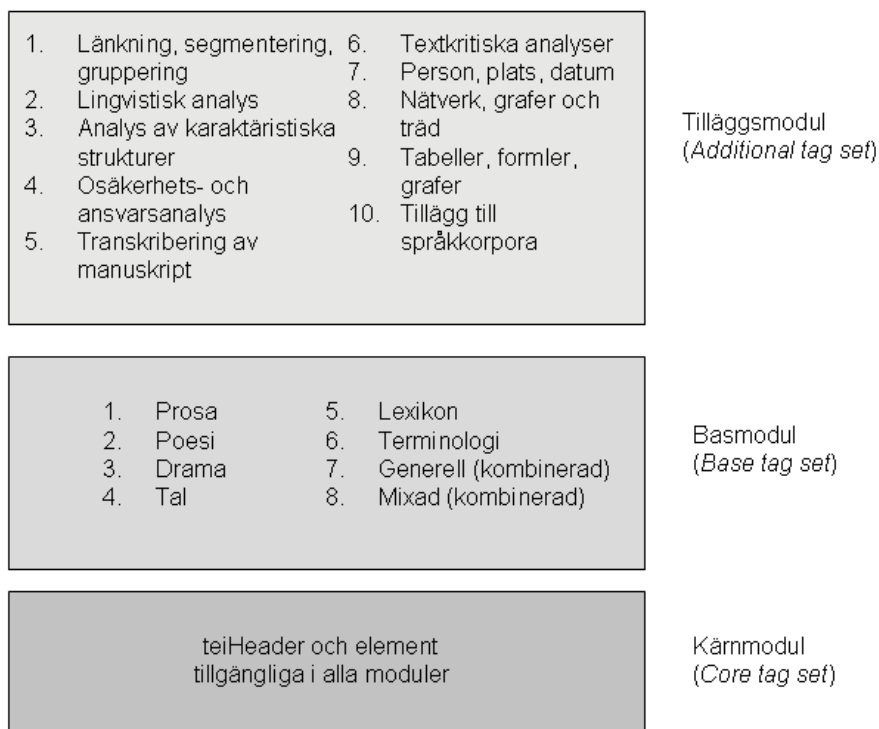
I *Figur 13* kan vi se att schemat består av tre övergripande moduler, var och en med en specifik funktion. Syftet med denna design är att tillhandahålla valfrihet för användare gällande genre eller typ av text, i relation och med hänsyn till dokumentformatets krav på stringens och konformitet. Därför finns möjligheter att kombinera olika typer av element lämpliga för olika genrer och teoretiska perspektiv samt möjligheter att definiera egna element vid behov.

³⁹¹ McGann, "Making texts of many dimensions" s. 202.

³⁹² Sperberg-McQueen & Burnard, "The Design of the TEI Encoding Scheme" s. 18.

³⁹³ Vissa verktyg (t.ex. RDF) som tagits fram för märkning av metadata på webben har en generell konstruktion som medför att varje användare definierar hur det semantiska innehållet beskrivs, vilket enligt min mening väsentligen försvårar utbyte av innehåll.

³⁹⁴ För den läsare som önskar en preciserad beskrivning av TEI-schemat hänvisas till TEI:s riktlinjer. Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange*. Även tillgänglig vid <www.tei-c.org/release/doc/tei-p4-doc/html/>.



Figur 13. Översikt TEI:s överordnade moduler (*Pizza Chef Model*) i TEI P4

TEI-schemat (i detta fall TEI P4³⁹⁵) byggs alltså upp av tre delar som brukar beskrivas som pizzamodellen.³⁹⁶ Kärnmodulen motsvarar tomatsåsen och osten, basmodulen själva degen och tilläggsmodulen av fyllningen som kan kombineras efter smak. Genom webbgränssnittet *Pizza Chef* beställs sedan vilka moduler som skall ingå i dokumenttypsdefinitionen (DTD:n).³⁹⁷ I DTD:n, som är en formell specifikation av hur dokumentet skall struktureras, deklarerats följande:

de element som skall ingå i schemat och hur de får användas
benämning av elementen
hur elementen får nästlas

³⁹⁵ I studien har TEI P4 använts av den anledningen att P5 inte var tillgängligt när märkningen av texterna utfördes. En prosopografisk modul introducerades i P5 under 2007. P5 publiceras i början av november 2007, utkast och betaversioner har dock funnits tillgängliga tidigare.

³⁹⁶ Se <www.tei-c.org.uk/pizza.html> (pizza enligt den amerikanska modellen, i Italien är det inte alls säkert att pizzan består av precis dessa ingredienser).

³⁹⁷ *Pizza Chef* är numera ersatt av verktyget Roma för konstruktion av önskat TEI-schema i P5 och där moduler, element och attribut kan väljas eller tas bort, se vidare s. 126 f.

elementens globala och specifika attribut³⁹⁸
entiteter (enheter/data som kan inkluderas i dokumentet och som inte omfattas av den teckenkodning som används, eller data som återanvänds många gånger, samt hänvisning till externt lagrad data som till exempel en annan DTD).

Kärnmodulen³⁹⁹ består av element som är tillgängliga i alla moduler. En sådan uppsättning element är **<teiHeader>** som används för metadata, Andra uppsättningar av element är sådana som används för beskrivning av rubriker, indelning av dokumentets delar, styckesindelning, punktering, citering, bibliografiska referenser, citering och förkortningar samt enklare beskrivning av namn, plats och datum med mera. Det vill säga sådant som arbetats fram inom TEI för att dessa egenskaper anses vara generella karaktäristika för texter inom olika generer.⁴⁰⁰ Från nästa modul, basmodulen, väljs sedan en, och endast en, uppsättning element.⁴⁰¹ Prosamodulen kan användas för beskrivning av de flesta typer av löpande text och innehåller därför samma element som kärnmodulen. Inga nya element introduceras här om inte någon annan av uppsättningarna i tilläggsmodulen aktiveras. Uppsättningen av prosaelement räknas i detta sammanhang alltså som den mest elementära modulen för beskrivning av text. I Tabell 3 nedan ges en översikt av samtliga elementuppsättningar i basmodulen.⁴⁰²

³⁹⁸ Vissa attribut är gemensamma för samtliga element, det vill säga de är *globala attribut*. Exempel är attributen **type** och **id** medan andra attribut är specifika för vissa elementgrupper. Se s. 77 f. för exempel.

³⁹⁹ Jag använder moduler för *tag sets* för att tydliggöra schemats design och särskilja dessa objekt från de klasser (*classes*) som alla enskilda element och attribut tillhör. En modul består av en mängd element som enligt TEI-schemat innehar gemensamma egenskaper eller funktioner.

⁴⁰⁰ Kap. 6 "Elements Available in All TEI Documents", se Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 125 -190. Även <www.tei-c.org/P4X/CO.html>

⁴⁰¹ Undantaget de kombinerade elementuppsättningarna.

⁴⁰² Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 217-340.

Tabell 3. *Basmodulens (base tag set) elementuppsättningar i TEI P4.*

Namn	Elementuppsättning	Exempelhänvisning
Prosa (<i>Prose</i>)	Element användbara för flertalet texter	<i>Figur 15</i> , s. 124 samt < http://www.tei-c.org/P4X/DS.html#DSDIV
Poesi (<i>Verse</i>)	Element för metrisk märkning, rytm, pauser, etc.	< http://www.tei-c.org/P4X/VE.html#VEME
Drama (<i>Drama</i>)	Element för rollistor, akt, scenanvisningar etc.	<i>Figur 9</i> , s. 86 samt < http://www.tei-c.org/P4X/DR.html#DRBOD
Tal (<i>Speech</i>)	Element för transkribering av tal för lingvistiska analyser	< http://www.tei-c.org/P4X/TS.html#TSBA
Lexikon (<i>Dictionaries</i>)	Element för detaljerade lexikografiska analyser	< http://www.tei-c.org/P4X/DI.html#DIEN
Terminologi (<i>Terminology</i>)	Element för terminologiska databaser	< http://www.tei-c.org/P4X/TE.html#TEBA
Generell (kombinerad) (<i>General base</i>)	Kombinerade element från basmodulens uppsättningar. Endast element från en uppsättning får användas för varje textdel.	< http://www.tei-c.org/P4X/ST.html#STMA
Mixad (kombinerad) (<i>Mixed base</i>)	Kombinerade element från basmodulens uppsättningar. Element från samtliga uppsättningar för användning på valfri plats	< http://www.tei-c.org/P4X/ST.html#STMA

Elementuppsättningarna i tabellen ovan motsvarar de discipliner och verksamheter som deltagit, och deltar, i utvecklingen av TEI. Uppställningen ovan motsvarar även den pluralism som genomsyrar TEI och de antaganden om vikten av beskrivningar som speglar ett visst teoretiskt perspektiv, även ett eklektiskt sådant.

För en rigorös märkning kan dock förfinade verktyg behövas. Då kan ett valfritt antal elementuppsättningar väljas från tilläggsmodulen. Tilläggsmodulens elementuppsättningar innehåller element som är relevanta för olika slag av bearbetning eller en speciell typ av forskning.⁴⁰³ Elementen kan även vara sådana som förekommer i flera genrer men som inte är tillräckligt vanliga för inkludering i basmodulen.

⁴⁰³ Ibid. s. 341-572.

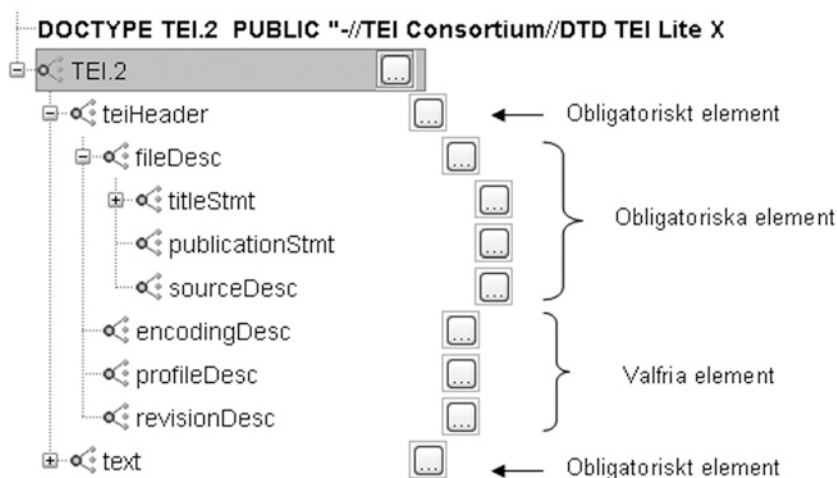
Tabell 4. Tilläggsmodulens elementuppsättningar i TEI P4.

Namn	Elementuppsättning	Exempelhänvisning
Länkning, segmentering, gruppering (<i>Linking, Segmentation, Alignment</i>)	Element för länkning till intern eller extern text, gruppering, korrespondens och aggregering av element, associering av element till specifika tolkningsscheman etc.	< http://www.tei-c.org/P4X/SA.html
Lingvistisk analys (<i>Simple Analytic Mechanisms</i>)	Beskrivning och klassificering av textsegment (meningar, ord, morfem, fraser, tecken etc.) Attribut för lingvistisk analys, tolkning och associering. Lingvistisk annotering	< http://www.tei-c.org/P4X/AI.html
Karaktäristiska strukturer (<i>Feature Structures</i>)	Element för benämning, identifiering och gruppering av data/objekt som kan inordnas i valfria karaktäristiska strukturer. fs-modulen kan beskrivas som ett metaspråk	< http://www.tei-c.org/P4X/FS.html
Osäkerhets- och ansvarsanalyser (<i>Certainty and Responsibility</i>)	Element för beskrivning och gradering av osäkerhet inför textens äkthet, datering, tolkning eller liknande. Element för märkning av vem som ansvarar för tolkningen.	< http://www.tei-c.org/P4X/CE.html
Transkribering av primärkällor (<i>Transcription of Primary Sources</i>)	Element för beskrivning av primärkällors (manuskript eller andra ej tryckta källor) typografiska, materiella och annan specifik information (t.ex. sidhuvud, rubrikinformation).	< http://www.tei-c.org/P4X/PH.html#PHSK
Textkritiska analyser (<i>Critical Apparatus</i>)	Element för märkning av textkritiska utgåvor för märkning av verkets varianter, manuskript, utgåvor, citeringar etc.	< http://www.tei-c.org/P4X/TC.html
Namn och datum (<i>Names and Dates</i>)	Element för märkning av egennamn (person, plats, organisationer) samt specificerad märkning av datum eller andra typer av tidsperioder	< http://www.tei-c.org/P4X/ND.html
Grafer, nätverk och träd (<i>Graphs, Networks, and Trees</i>)	Element för märkning av grafer, noder och trädstrukturer såsom definierade inom matematiken	< http://www.tei-c.org/P4X/GD.html
Tabeller, formler och grafik (<i>Tables, Formulae, and Graphics</i>)	Element för märkning av tabeller, diagram, formler och grafisk information. Särskild fokus på layout och presentation samt beskrivning av olika bildformat	< http://www.tei-c.org/P4X/FT.html
Språkkorpora (<i>Language Corpora</i>)	Elementuppsättningar för märkning av komposition och beskrivning av textsamlingar bestående av språkliga uttryck (tryckta eller talspråk)	< http://www.tei-c.org/P4X/CC.html

Tilläggsmodulens elementuppsättningar är exempel på TEI-schemats komplexitet och omfattning. I den här studien har elementuppsättningarna för prosa samt den för namn och datum använts för beskrivning och strukturering av texterna i databasen. Initialt planerade jag att använda den mycket komplexa uppsättning element som kan användas för att beskriva och ordna valfria egenskaper och strukturer (*Feature Structures*). Ansatsen övergavs dock ganska snart då modulen visade sig vara för komplex för de egenskaper som behövde beskrivas och ordnas. För märkning av de egenskaper, framför allt olika typer av egennamn och tidsperioder, valdes att istället använda kombinationen av elementuppsättningen för prosa (*Prose*) samt tilläggsmodulen för namn och datum (*Names and Dates*).

4.5.1 Obligatoriska och valfria element

En av de viktigare utgångspunkterna för TEI är att man genom riktlinjerna vill erbjuda användarna valmöjligheter snarare än att påbjuda en uppsättning regler. En större andel av de närapå 500 elementen som inryms i de uppsättningar som ovan beskrivits är därför valfria.⁴⁰⁴ De element som är obligatoriska återfinns framför allt i `<teiHeader>` där den bibliografiska informationen dokumenteras och där innehållet beskrivs samt hur märkningen utförts.⁴⁰⁵



Figur 14. TEI-dokumentets grundläggande struktur och de obligatoriska elementen i `<teiHeader>`

Den obligatoriska `<teiHeader>` består av fyra delar:

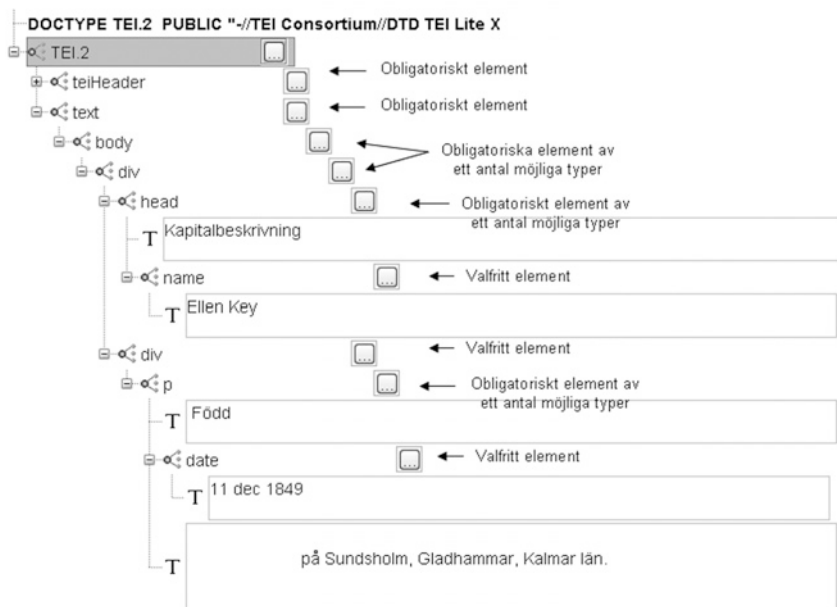
1. `<fileDesc>` för bibliografisk beskrivning av den digitala versionen och dess källor,
2. `<encodingDesc>` för beskrivning av hur märkningen utförts,
3. `<profileDesc>` för beskrivning av innehållet utöver den bibliografiska informationen,
4. `<revisionDesc>` för information om versioner, uppdateringar och ändringar.

Endast elementuppsättning, `<fileDesc>`, är obligatorisk men om övriga tre används måste de användas i den ordning som de framställs i Figur 14 ovan. Ele-

⁴⁰⁴ TEI P4 innehåller 441 element och antalet element i TEI P5 uppgår till 493.

⁴⁰⁵ Den modifierade bilden i figuren nedan samt i Figur 15 på s. 124 baseras på ett urklipp ur programmet Oxygen XML Editor (gränssnittet Tree Editor). Oxygen finns tillgängligt för nedladdning vid: www.oxygenxml.com.

menten som ingår i <fileDesc> innehåller bibliografisk information som identifierar och beskriver filens innehåll. I TEI-schemat har man eftersträvat att den bibliografiska information skall motsvara de beskrivningar och de benämningar som används av biblioteken.⁴⁰⁶ De tre obligatoriska elementen i <fileDesc> innefattar information om titel, publicering och distribuering samt information om den tryckta källan.⁴⁰⁷ För de dokument som inte har någon tryckt källa finns inget specifikt element, här används <sourceDesc> även för dokument som skapats digitalt.⁴⁰⁸



Figur 15. Exempel på grundläggande struktur och element i TEI-dokumentets kropp

Om vi lämnar <teiHeader> och istället övergår till textens kropp visar Figur 15 den mest grundläggande strukturen av en löpande text och exempel på några av de få obligatoriska elementen. Elementet <text> är obligatoriskt och detta element innesluter i sin tur ett antal valbara element, där något av dem måste väljas. I exemplet ovan visas elementet <div> (för *division*) som kan användas för att dela in och avgränsa innehåll på flera olika sätt relaterat till genre och typ av text. I TEI:s riktlinjer har man, där det är möjligt, valt att använda generella

⁴⁰⁶ Richard Giordano, "The TEI header and the Documentation of Electronic Texts" *Text encoding initiative: background and context*, red. Nancy Ide & Jean Vâronis, (Dordrecht; London: Kluwer Academic, 1995) s. 76 f.

⁴⁰⁷ <sourceDesc> supplies a bibliographic description of the copy text(s) from which an electronic text was derived or generated. <www.tei-c.org/P4X/HD.html#HD2>.

⁴⁰⁸ En lösning är att i <sourceDesc> helt enkelt beskriva att dokumentet är ett digitalt originaldokument.

benämningar på elementen. Med utgångspunkt i den innehållshierarkiska modellen (OHCO) är många texter uppbyggda av ett godtyckligt antal innehållsliga objekt ordnade i en viss följd.⁴⁰⁹ För att ange vilken typ innehållsobjekt är används det globala attributet *type* som anger om det skall beskrivas som ett kapitel, en vers eller en indelning utifrån vilken sorts egenskaper innehållsobjektet innehåller, det vill säga som elementet `<div>` har använts i den här studien.⁴¹⁰

Efter en grundläggande indelning av texten går det sedan att infoga valfri beskrivning antingen på en mer rudimentär nivå eller mycket detaljerat. Beroende på vilka aspekter som skall märkas finns, som jag ovan försökt visa, ett stort antal valmöjligheter. Svårigheterna är möjligen att välja den uppsättning element som bäst motsvarar de teoretiska utgångspunkterna och/eller de pragmatiska målen som datamodellen beskriver.

4.5.1.1 *TEI Lite* – en lagom lättversion

Ofta är det så att inte alla element som finns verkligen behövs vid märkning av olika slags innehåll. Då finns möjligheter att använda en förminskad version av TEI-schemat. *TEI Lite* innehåller ungefär en femtedel av elementuppsättningen består av de mest använda elementen.

TEI Lite is a specific customization of the TEI tagset, designed to meet "90% of the needs of 90% of the TEI user community".⁴¹¹

Fördelarna med *TEI Lite* flera: Det är lättare för en nybörjare att komma igång med innehållsmärkningen när det inte är så många olika element och hierarkiska strukturer att hålla reda på. För en inte så detaljerad märkning är den också självskriven. Varför använda de större elementuppsättningarna om det inte behövs? Nackdelen med lättversionen är att för detaljerade och "djupa" beskrivningarna räcker elementen inte alltid till för en adekvat beskrivning. Men då finns alltid möjligheter att använda de ämnesspecifika modulerna.

TEI Lite finns inte i riktlinjerna för P5 av den anledningen att det inte finns behov av den längre. Å ena sidan, i och med att användare fritt kan välja elementuppsättningar, där det dessutom finns möjlighet att välja bort de element som inte behövs eller lägga till sådana som saknas, relativt det innehåll som skall beskrivas, kanske lättversionen blir överflödiga. Å andra sidan är den en bra utgångspunkt för nybörjare då det inte är fullt lika många element att hålla reda på, välja bland och ta ställning till.

⁴⁰⁹ Burnard & Sperberg-McQueen, *Text Encoding Initiative. Guidelines for electronic text encoding and interchange* s. 191-202.

⁴¹⁰ Exempel i Bilaga 2 `<CD:/bilagor/bilaga.2pdf>` samt i Bilaga 8 `<CD:/bilagor/bilaga.8pdf>`. Dessa dokument är även tillgängliga i XML-format, se `<CD:/xml/ffo-kapdesc.xml>` och `<CD:/xml/klhs-post.xml>`.

⁴¹¹ Se `<www.tei-c.org/Guidelines/Customization/Lite/>`

4.6 TEI P5 – från pizza till smörgåsbord

Sedan den fjärde utgåvan av TEI:s riktlinjer 2002 (TEI P4) har aktiviteten varit hög med att uppdatera, utöka och förenkla användningen av schemat.⁴¹² Vid TEI:s årsmöte i november 2003 identifierades tre målsättningar för TEI P5:⁴¹³

- Interoperabilitet genom att nyttja arbete utfört av andra.
- Intern revision genom rensning av ett decenniums tillväxt.
- Expansion genom att adressera outforskade områden.

Prioriteringen ovan är enligt Burnard en följd av att TEI tidigare framför allt riktat sig till de egna leden men att det nu blivit viktigt att anpassa sig till utvecklingen av XML och webben.⁴¹⁴

För interoperabilitet och bättre portabilitet rekommenderar TEI att Unicode används för teckenkodning av sådana tecken som inte går att representera i ASCII-formatet istället för att som tidigare definiera tecknen ad hoc.⁴¹⁵ Av samma anledning (det vill säga interoperabilitet) har TEI:s attribut för att ange språk ersatts av det av W3C definierade attributet `xml:lang` där språken definieras enligt en ISO standard.⁴¹⁶ Genom denna, till synes, måttliga förändring har man övergått till användning av två olika standarder istället för en internt definierad modell.⁴¹⁷ Därtill satsas det nu även på internationalisering genom att riktlinjerna i sin helhet översatt till några av de stora språken.⁴¹⁸ Vidare tillåts användning av så kallade namnrymder som tillåter att andra scheman kan integreras.

Den interna revisionen av TEI-schemat har inriktats på att utveckla specifikation och kontroll av tillåtna attributvärden. Det fanns endast begränsade möjligheter att styra attributvärdenas datatyper i SGML. I TEI löste man detta genom att för vissa attribut ange vilka värden som överensstämde med dokumenttypsdefinitionen (DTD:n), vilka värden som rekommenderades eller gavs det exempel på möjliga värden. Ett sådant exempel är elementet `<date>` där man tidigare rekommenderade normaliserade attributvärden för tidsangivelser. Ge-

⁴¹² Burnard. "New tricks from an old dog: An overview of TEI P5".

⁴¹³ Information och dokumentation om TEI P5 tillgängligt vid: www.tei-c.org/Guidelines/P5/.

⁴¹⁴ Burnard. "New tricks from an old dog: An overview of TEI P5".

⁴¹⁵ ASCII utformades ursprungligen för amerikanska syften och har därför inte haft stöd för alla språk även om standarden byggts ut under årens lopp. Unicode stöder teckenkodningen i alla språk och blir allt vanligare.

⁴¹⁶ Se www.w3.org/TR/REC-xml/#sec-lang-tag.

⁴¹⁷ Att TEI:s definitioner skulle vara interna är möjligen något överdrivet, men som jämförelse med W3C:s rekommendationer och ISO standarder anser jag att det är relevant.

⁴¹⁸ Översättningar till kinesiska, franska, tyska, italienska, japanska och spanska är nästan helt färdiga i november 2007. Även gränssnitt och hjälptexter i den applikation som används för att generera scheman och dokumentation har översatts till flera olika språk. Där har jag gjort den svenska översättningen. <http://tei.oucs.ox.ac.uk/Roma/>

nom att schemat numera kan beskrivas med XML-notation finns även möjligheter att styra dataformat för inmatning av data (i detta fall attributvärden), precis som i en databas.⁴¹⁹ I TEI P5 används W3C:s rekommendationer för datatyper i XML istället för de tidigare interna rekommendationerna.⁴²⁰ I TEI P4 rekommenderas att använda följande datumnotation:

```
<date value="2007-01-01"> 1:a januari 2007 </date>
```

I P5 har man ett specificerat datumformat som anger vilken datatyp som är tillåten att lagra i attributet. Exemplet ovan annoteras med TEI P5 enligt figuren nedan:

```
<date when="2007-01-01"> 1:a januari 2007 </date>
```

I P5 har man infört attributet **when** för normaliserade värden av tidsangivelser. Inte några stora synliga förändringar i notationen, dock är det ett steg mot ytterligare möjligheter till utbyte genom användning av internationella standarder.

Den interna dokumentationen av själva TEI-schemat kallas för ODD (*One Document Does it all*). Denna specifikation och dokumentation är en nu en modul som alla andra. ODD karaktäriseras av några specifika funktioner:

- Element- och attributklasserna formella specifikationer beskrivs med ett speciellt XML-vokabulär.
- Specifikationsspråket innehåller stöd för: macron, ett hierarkiskt system för element- och attributklasser samt design av modulerna.
- Dokumentmodellerna för element och attribut formuleras med RELAXNG XML notation. Det finns dock verktyg för att generera scheman i flera olika metaspråk: andra RELAXNG varianter, DTD notation eller W3C schema.
- Dokumentation för beskrivning av element och attribut etcetera tillhandahålls tillsammans med dessa specifikationer, användarfall, exempel och andra typer av hjälpmaterial.⁴²¹

⁴¹⁹ Se t.ex. Lou Burnard & Sebastian Rahtz. "RelaxNG with Son of ODD". *Extreme Markup Languages 2004*, Montréal, Québec, 2004.

⁴²⁰ W3C XML Schema Part 2: Datatyper används för tidsangivelser enligt den gregorianska kalendern, {TEI Consortium, 2007 #264@Appendix D Datatypes and Other Macros, se <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-MACRO,S.html>

⁴²¹ Burnard & Rahtz. "RelaxNG with Son of ODD".

Denna genomgripande förändring innebär att man i TEI P5 tagit bort skillnaderna mellan olika elementuppsättningar och därmed redundant beskrivning av vissa element. Varje element definieras endast en gång och vilken kombination av element som helst är tillåten för uppbyggnaden av ett schema. Detta är förstås ett steg mot ytterligare modularitet. Även om kärnmodulen och *teiHeader* fortfarande finns kvar har användare möjlighet att bygga sitt schema helt efter egna önskemål. De restriktioner som fanns i pizza-modellen har tagits bort. Nu är det tillåtet att plocka de ”godbitar” man vill.⁴²²

En annan förändring är att TEI i P5 övergår till en standard för pekare och länkning genom att använda *Uniform Resource Identifier* (URI) som är avsedd för att lokalisera, identifiera eller namnge en resurs.⁴²³

4.6.1 Modulen namn, datum, personer och platser

Under hösten 2007 publicerade TEI en ny modul i schemat (P5) av särskilt intresse för den här studien. Det är en utökad och förändrad variant av modulen som hanterar beskrivning av namn, plats och datum.⁴²⁴

This module also provides elements for the representation of information about the person, place, or organization to which a given name is understood to refer and to represent the name itself, independently of its application. In simple terms, where the core module allows one simply to represent that a given piece of text is a *name*, this module allows one further to represent a *personal name*, to represent the *person* being named, and to represent the *canonical name* being used. A similar range is provided for names of places and organizations. The intended application for this module is in biographical, historical, or geographical data systems such as gazetteers and biographical databases, where these are to be integrated with encoded texts.⁴²⁵

I modulen finns alltså specifika element för beskrivning av historiska, geografiska, biografiska och mer specifikt prosopografisk information.

Elementen i modulen utgår från tre olika typer:

- Personliga egenskaper och karaktäristiska (kulturella, fysiska, politiska, religiösa etcetera)

⁴²² Utformning av TEI-schema i P5 tillgängligt vid: <<http://tei.oucs.ox.ac.uk/Roma/>>.

⁴²³ Se <www.w3.org/Addressing/> samt <www.w3.org/Addressing/Activity/>.

⁴²⁴ När den här studien inleddes fanns ingen specifik modul för beskrivning prosopografiska material. Den reviderade modulen för beskrivning av namn, platser och annat i prosopografiska material publicerades i text Encoding Guidelines P5, november 2007 och speglar på ett bättre sätt än tidigare hur prosopografiska innehållsobjekt kan beskrivas. {TEI Consortium, 2007 #264@”13 Names, Dates, People, and Places”}

Se även <www.tei-c.org/release/doc/tei-p5-doc/html/ND.html>.

⁴²⁵ P5: Guidelines for Electronic Text Encoding and Interchange, första sidan, andra stycket vid TEI Consortium, red, *TEI Consortium. Guidelines for Electronic Text Encoding and Interchange. P5* (2007) ”13 Names, Dates, People, and Places”. <www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>.

- Tidsbestämda tillstånd (giftermål, boendeort, yrke etcetera)
- Händelser (förändringar av olika slag t.ex. födelse, skilsmässa, död)⁴²⁶

Dessa tre beskrivningstyper överensstämmer mycket väl med de egenskaper och karaktäristika som märkts upp i FFO-databasen. Istället för de projektspecifika lösningar som hittills använts finns nu möjligheter att beskriva prosopografiska data utifrån en gemensam datamodell.⁴²⁷ I kapitlet om märkning av arkivposter diskuteras denna reviderade modul och de enskilda elementen närmare.⁴²⁸

Ett problem med verktyg som TEI-schemat, och andra liknande scheman och datamodeller, är att ju mer specificerade och detaljerade de blir desto svårare blir det att verkligen förstå principerna och upptäcka all funktionalitet. TEI, och de deskriptiva märkspråkens inneboende hierarkiska logik kritiserar som vi ovan sett för att premiera *en* logisk ordning och utelämna andra parallella eller alternativa ordningar. Då en modell samtidigt döljer som den visar fram är avvägningen mellan en trivial modell som avskärmar vyerna, eller en komplex som vidgar dem, inte alltid självklar.⁴²⁹ En modell som avser fånga in alla möjliga aspekter och vyer av ett visst innehåll riskerar att förlora skärpan. En modell som förstärker skärpan riskerar dölja mer än den visar. Så vilken väg skall man välja? Matthew Driscoll förespråkar den pluralistiska vägen:

Indeed, to all intents and purposes there is no limit to the information one can add to a text – apart, that is, from the limits of the imagination. One thing is clear: the more we put into a text, the more we can get out of it. If we are lucky we wind up even finding things we didn't know were there; if we are very lucky, we find things one didn't even know it was possible to look for.⁴³⁰

TEI-schemat gör det möjligt att beskriva innehållet utifrån flera vyer, eller modeller, vilket öppnar för fler möjliga användningsområden snarare än färre.

⁴²⁶ En preliminär version presenterades av Matthew Driscoll i maj 2007 under en workshop vid Uppsala universitet för prosopografiska metoder. Matthew James Driscoll. "XML markup of prosopographical data". *Exploring New Methods for Prosopography in the Humanities and the Social Sciences*, Uppsala, Sweden, 2007.

⁴²⁷ Märkningen med den nyare modulen presenteras i delkapitel 6.2 Beskrivande märkning av arkivmaterial s. 184 f.

⁴²⁸ Se kapitel 6.2 s. 184 f.

⁴²⁹ Jfr. McCarty, *Humanities computing* s. 50-53.

⁴³⁰ Matthew James Driscoll, "Levels of Transcription" *Electronic textual editing*, red. L. Burnard, et al., (New York, N.Y.: Modern Language Association of America, 2006) s. 261.

4.7 Avslutning

I detta kapitel har utgångspunkter, principer och användningsområden för verktyget TEI presenterats och diskuterats. En del utrymme har ägnats åt en organisatorisk och historisk översikt. Avsikten med det var att rama in och belysa de insatser som ligger bakom TEI:s riktlinjer. Forskare, lärare, bibliotekarier och andra från en rad olika kunskapsområden har under 20 år lyckats inte bara komma överens om principerna för dessa riktlinjer utan även fortsatt att ständigt utveckla dem. TEI:s femte version av riktlinjerna publicerades lagom till tjugoaårsjubileet hösten 2007. Trots min lovsång här till TEI finns förstås fläckar även på denna sol. En sådan är tillgängligheten, det vill säga svårigheter för nybörjare att börja använda TEI-schemat som med sin komplexitet har en hög lärandetröskel och borde kunna skrämman den lagom intresserade men troligen inte den mycket intresserade. Detta problem är också uppmärksammat inom TEI-gemenskapen, där exempel på *good practice* och andra försök att förbättra tillgängligheten förbättrats.⁴³¹ Problemet med flera av de instruktiva guiderna är att innehållet sällan finns tillgängligt. Mitt bidrag till det är den märkning, de exempel och de tillämpningar jag lägger fram i den här avhandlingen. Det material som presenteras i avhandlingen finns tillgängligt dels som exempel på den bifogade CD:n, dels finns allt material tillgängligt i webbarkivet.⁴³²

Riktlinjernas modulära design har också presenterats och åskådliggjorts. Skillnaderna mellan TEI:s förra version P4 och den nya, P5, är framför på det sätt modulerna kan väljas och kombineras. I den nya versionen är det möjligt att plocka de godbitar som är av intresse. En annan nyhet i TEI P5 är den reviderade modulen för beskrivning av namn, datum, person och plats som utökats och förbättrats.

Detta var sista kapitlet i delen II om innehållsdesign. I nästa del av avhandlingen redovisas de tillämpningar som bygger på det som presenterats i den här delen av avhandlingen.

⁴³¹ Se t.ex. <www.tei-c.org/Support/Learn/>.

⁴³² Aktuella versioner av webbarkiv och databas tillgängliga på vid <www.skeptron.uu.se/monicalz/ffo/>.

Del III Tillämpad innehållsdesign

5 Prosopografisk metod, verktyg och material

I detta kapitel kommer jag att illustrera och exemplifiera resultatet av modelleringen av det empiriska material som insamlats inom projektet Formering för offentlighet. Inledningsvis presenterar jag den prosopografiska metoden, modellering och klassificering samt ger exempel på prosopografiska tillämpningar. Därefter ges en kort introduktion om projektet Formering för offentlighet samt en översikt över materialet som samlats in och sammanställts. Även de verktyg som valts, det vill säga TEI-schemat, diskuteras i relation till beskrivning av det prosopografiska materialet. Kapitlet ägnas även åt utgångspunkter för modelleringen där särskild vikt läggs vid de olika arbetsmomenten i utförandet.

5.1 Om metod

Tiden runt sekelskiftet 1900 erbjuder många möjligheter för studier av kvinnors väg till offentlighet, kamp för utbildning, deras verksamheter, inträde på arbetsmarknaden eller kamp för rösträtt.⁴³³ Avhandlingen handlar dock inte primärt om dåtidens borgerliga kvinnors väg till offentlighet, utan om att skapa förutsättningar för sådana studier.

Den prosopografiska metoden som närmare diskuteras nedan består i att systematiskt samla in ett stort antal uppgifter om en avgränsad grupp människor som har något gemensamt, till exempel kön eller en viss socialgrupps tillhörighet. I det här sammanhanget är det dels borgerliga kvinnor kring sekelskiftet 1900 och dels lärarinnestuderande under perioden 1861-1943 som avgränsar gruppen. Materialet i studien är ganska disparat och metoderna likaså. Detta tillvägagångssätt kan förstås diskuteras men en anledning är att för socialhisto-

⁴³³ Exempel på sådana studier är bl.a.: Christina Florin, *Kampen om katedern: feminiserings- och professionaliseringsprocessen inom den svenska folkskolans lärarkår 1860-1906*, Umeå studies in the humanities, 82 (Umeå & Stockholm: Almqvist & Wiksell, 1987); Ebba Heckscher, *Några drag ur den svenska flickskolans historia: under fleres medverkan samlade* (Stockholm: Norstedt & Söner, 1914); Josefin Rönnbäck, *Politikens genusgränser: den kvinnliga rösträttsrörelsen och kampen för kvinnors politiska medborgarskap 1902-1921*, Atlas akademi (Stockholm: Atlas, 2004); Tord Rönnholm, *Kunskapens kvinnor: sekelskiftets studentskor i mötet med den manliga universitetsvärlden*, Forskningsrapporter från Institutionen för historiska studier vid Umeå universitet, 12 (Umeå: Institutionen för historiska studier Univ. distributör, 1999); Annika Ullman, *Stiftarinnegenerationen: Sofi Almqvist, Anna Sandström, Anna Ahlström* (Stockholm: Stockholmia, 2004); Eva Österberg & Christina Carlsson Wetterberg, red., *Rummet vidgas: kvinnor på väg ut i offentligheten 1880-1940* (Stockholm: Atlantis, 2002).

riska studier är det ofta nödvändigt att vara eklektiker. John Tosh belyser varför det behövs en eklektisk strategi för historiska studier av sociala grupper:

Sociala grupper efterlämnar inga kollektiva arkiv. Deras sammansättning och plats i den sociala strukturen måste rekonstrueras ur ett brett spektrum av källor som har satts samman av högst banalare skäl.⁴³⁴

Avhandlingen baseras just på ett sådant tillvägagångssätt där källmaterialet hämtats från olika typer av källor och metoderna för insamling, sammansättning och analyser varierar. Avhandlingen handlar även om klassificeringsproblematiken och andra ställningstaganden som är nödvändiga att göra när vi konstruerar historiskt material om sociala grupper.

5.1.1 Modellering som metod

Min forskningsmetod kan definieras som tillämpad modellering med en explorativ och kritisk ansats. Studien har en pragmatisk utgångspunkt där principer och metoder för beskrivning av innehåll samt utformning av prosopografiskt material utforskas genom att sätta verktygen och metoderna i arbete.

Modellering kan grovt delas in i två distinkta faser: konstruktion och manipulering.⁴³⁵ Gränsen mellan dessa två faser är suddig och de överlappar ofta varandra, speciellt initialt när konstruktionen kanske är behäftad med många fel och brister. Distinktionen är ändå viktig då de är varandras förutsättningar. Det är genom att manipulera konstruktionen som konstruktionen kan förbättras eller förändras. Som exempel kan nämnas att de första försöken att konstruera något användbart av det historiska material som utformades för projektet, resulterade i ändringar i konstruktionen efter att jag haft möjlighet, tack vare konstruktionen, att manipulera innehållet. Genom dessa upprepade förflyttningar mellan konstruktion och manipulering kunde modellen sedan förfinas och förbättras.

Willard McCarty anser att modellen fungerar som en brygga mellan teori och empiri, samt mellan subjekt och objekt genom dess olika slags relationer:

[...] the manipulatory essence of modeling, with its connotation of embodied action, physically or metaphorically; the mediating role and ternary relationship modelling establish between knower and known; the directed vector-like engagement of the inquirer's attention, *through* the model he or she has made *to* the object of study; and the model's consequent function as an artificial agent of perception and instrument of thought.⁴³⁶

⁴³⁴ John Tosh, *Historisk teori och metod*, 2., rev. uppl. ed. (Lund: Studentlitteratur, 2000) s. 114.

⁴³⁵ McCarty, *Humanities computing* s. 27. McCarty använder även distinktionen "model *of*" (konstruktion) respektive "model *for*" (manipulation). Se t.ex. Willard McCarty, "Knowing true things by what their mockeries be: modelling in the humanities", *Text Technology* 12.1 (2003); McCarty, *Humanities computing*.

⁴³⁶ McCarty, *Humanities computing* s. 38.

En modells fokuserade och samtidigt förenklande funktionalitet ger oss möjligheter att fokusera det vi letar efter men gör det också möjligt att finna det vi inte söker efter. Problem uppstår givetvis om modellen inte fungerar på det material som skall modelleras. Det ”överskott” av materialen som modellen förkastat kan hanteras på olika sätt. Antingen modifieras modellen för att bättre passa det material som modelleras eller så används ”överskottet” som ett resultat i sig samt för att analysera och problematisera modellens förtjänster och brister. I praktiken används ofta en kombination av dessa två angreppssätt. Distinktionen mellan dem är ändå viktig inte minst för att skilja ett kritiskt angreppssätt från ett som enbart är praktiskt.

Willard McCarty framhåller att modellering har växt fram ur praktiken och inte ur teorierna, vilket gjort att den är starkt rotad i den tysta kunskapen, det vill säga sådan kunskap som inte bara är oartikulerad utan som kanske inte kan artikuleras.⁴³⁷ Med det vill jag säga att även om jag försöker vara tydlig i alla steg i modelleringen av detta material kan det tänkas att vissa aspekter förblir oartikulerade helt enkelt för att de är *för* självklara delar av modelleringen och att de därför inte blir någon aspekt att beakta eller förhålla sig kritisk till.

5.2 Prosopografins utgångspunkter och karaktäristika

Den prosopografiska metoden har använts som utgångspunkt för insamling av det material som tillämpningarna bygger på. I avsnittet ges därför en översikt av den prosopografiska metodens utgångspunkter och användningsområden samt exempel på prosopografiska tillämpningsområden och inriktningar.

Prosopografi är en metod som länge använts för socialhistoriska studier av samhällets eliter och innehåller inslag av genealogi, onomastik⁴³⁸ och demografi. I Europa var metoden redan under 1500-talet förknippad med idén om att samla fakta om berömda män i de sociala eliterna i kollektiva biografier. Prosopografin föddes ur dåtidens strävan om att samla in, förteckna och organisera all världens kunskap.⁴³⁹

Den biografiska och den prosopografiska metoden är relaterade, och i vissa avseenden överlappar de varandra. Den främsta skillnaden är att den prosopografiska analysen syftar till att studera en viss grupp, dess historia och struktur, (och inte i första hand den enskilde individen) genom analyser av aggregerade kollektiva biografiska detaljer och uppgifter.⁴⁴⁰

Praktiskt består metoden i att systematiskt samla in omfattande information om gemensamma karaktäristika och tillgångar för var och en av individerna i

⁴³⁷ McCarty, ”Knowing true things by what their mockeries be: modelling in the humanities” s. 3.

⁴³⁸ Namnforskning, vanligen person- och ortnamn.

⁴³⁹ Karl Ferdinand Werner, ”L'apport de la prosopographie à l'histoire sociale des élites” *Family trees and the roots of politics: the prosopography of Britain and France from the tenth to the twelfth century*, red. K. S. B. Keats-Rohan, (Woodbridge: Boydell Press, 1997) ssk. s. 3-5.

⁴⁴⁰ Donald Broady, ”French prosopography: definition and suggested readings”, *Poetics* 30 (2002).

urvalet. Denna information används sedan för att undersöka grupper av individer som har något gemensamt, antingen verksamma inom ett specifikt fält eller verksamma under en viss tidsperiod.⁴⁴¹ Metoden ger därmed möjligheter att å ena sidan använda individers egenskaper för att belysa och analysera skeenden och strategier. Å andra sidan kan kollektivets samlade tillgångar användas för att illustrera individers positioner och fördelning av tillgångar i de sociala rum som kan konstrueras med utgångspunkt i individernas egenskaper och tillgångar. Genom uppgifter om individernas tillgångar, som har sitt ursprung i olika tidsperioder i individernas liv, ges även möjligheter att studera både kollektiva gemensamheter och förändringar över tid.⁴⁴² Kombinationen av individ- och kollektiv-perspektiv ger även möjligheter att analysera sociala nätverk i de fall uppgifter om sociala kontakter finns.

På grund av komplexiteten i såväl datainsamling som bearbetning lämpar sig prosopografi särskilt väl för digitalisering. Genom att digitalisera prosopografiskt material är det möjligt att hantera omfångsrika prosopografier med hundratusentals individer och som dessutom är mycket informationstät.⁴⁴³

Den prosopografiska metoden används idag inom samhällsvetenskap såväl som humaniora. Skillnaden i tillämpningsområden består framför allt i syftet och på vilka grunder individernas egenskaper registreras samt hur den prosopografiska produkten sedan används. I samhällsvetenskaperna introducerades metoden inom den franska sociologiska traditionen.⁴⁴⁴

I Donald Broadys definition av prosopografi inom den franska sociologiska traditionen betonas särskilt att det är fältets historia och struktur som studeras, inte individen *per se*.⁴⁴⁵ En annan, men snarlik, definition svarar historikern Lawrence Stone för.⁴⁴⁶ Såväl Broady som Stone skriver om nödvändigheten av att isolera ett socialt sammanhang som skall studeras ("field" respektive "universe") samt att de uppgifter som samlas in skall vara enhetliga och spegla individernas livsbanor, egenskaper och tillgångar. Till skillnad från Stone betonar Broady det specifika symboliska kapitalet och ställningstaganden i frågor av betydelse inom ett visst fält.

⁴⁴¹ För metoddiskussioner i anknytning till projektet, se Agneta Linné & Kerstin Skog-Östlin, red, *Biografi och kollektivbiografi som historisk metod: NFPF:s kongress 15-18 mars 2001 i Stockholm* (Uppsala: Forum för pedagogisk historia SEC, Uppsala Universitet, 2001).

⁴⁴² Se Larsson, *Från adlig uppföstran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866* s. 54 f.

⁴⁴³ *Prosopography of Anglo-Saxon England*, är både omfångsrik och informationstät, se <www.pase.ac.uk/>.

⁴⁴⁴ Ett välkänt exempel är studien om det akademiska fältet i Paris, se Pierre Bourdieu, *Homo academicus* (Stockholm/Stehag: Brutus Östlings bokförl. Symposion, 1996). Ett annat exempel är Gisele Sapiros prosopografiska studie om det franska litterära fältet under andra världskriget, Gisele Sapiro, "The structure of the French literary field during the German Occupation (1940–1944): a multiple correspondence analysis", *Poetics* 30 (2002).

⁴⁴⁵ Broady, "French prosopography: definition and suggested readings" s. 381–382.

⁴⁴⁶ Lawrence Stone, "Prosopography" *Historical Studies Today*, red. Felix Gilbert & Stephen R. Graubard, (New York: Norton, 1972) s. 107.

I inledningskapitlet berörde jag det faktum att prosopografiska studier ofta kräver tillgång till många typer av uppgifter och ofta krävs en insamling av relevanta uppgifter från många olika källor. Dessa uppgifter behöver sedan beskrivas, ordnas, bearbetas, analyseras och presenteras. Mångfalden av källor och uppgifter som samlas in om alla individer ger inte bara möjligheter till att utföra kvantitativa studier utan dessutom möjligheter till fallstudier om enskilda individers livsbanor och gärningar.⁴⁴⁷ För detta arbete finns olika metoder för registrering och lagring av uppgifter att tillgå. Det sammanlagda materialet, det vill säga prosopografin, resulterar ofta i en komplex samling, dels för att uppgifter baseras på olika typer av källor, dels för att dessa källor var för sig och tillsammans skall värderas och modelleras. Därför anser jag att den prosopografiska metoden är lämplig att använda som studieobjekt för innehållsdesign.

Jag har som nämnts valt att använda TEI-schemat, och därmed en hierarkisk dokumentmodell, som verktyg för beskrivning och organisering av det prosopografiska materialet. Ett annat, fullt möjligt och mer vedertaget alternativ, hade varit att välja en relationsdatabas.⁴⁴⁸ Anledningen till valet av TEI-schemat istället för en relationsdatabas är att jag ville utforska hur humanistiska metoder och verktyg (märkning av innehåll med TEI-schemat) kunde kombineras med analysmetoder och statistiska verktyg som framför allt används inom samhällsvetenskaperna. En annan anledning till valet av TEI-schemat var en målsättning att lagra materialet i *ett* och samma format – ett format som dessutom ger förutsättningar för fler användningar baserade på ett och samma ursprungsmaterial. TEI/XML som filformat ger även möjligheter att flytta materialet mellan olika plattformar. Därtill tillåter formatet en annan typ av modulär design av relativt ostrukturerade data som inte på samma sätt är möjligt med en relationsdatabas.

Den modulära designen ger även förutsättningar för presentationer av individuella eller grupprelaterade levnadsbeskrivningar samt olika tematiska ”ingångar” i materialet. Tillika ger formatet möjligheter att utföra olika slags analyser, att kontrastera variabler, analysera relationer mellan individerna eller teckna kartor över de tillgångar och egenskaper som individerna i materialet besitter.

De prosopografiska tillämpningarna som beskrivs nedan använder sig av relationsdatabaser. Skillnaden mellan dem och den studiens prosopografin är arbetssättet, källmaterialets karaktär och inte minst arbetsresurser. Dessa projekt är stora långsiktiga satsningar som har pågått i årtionden med många inblandade forskare och väl utvecklat tekniskt stöd. I mitt forskningsarbete har

⁴⁴⁷ Larsson, *Från adlig uppfostran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866* s. 63–66.

⁴⁴⁸ I en relationsdatabas organiseras data i relationer, dvs. i tabeller, som relateras till varandra med eller utan restriktioner för dessa relationer. Metoden skiljer sig därmed väsentligen från hierarkiska system. För analys av design av prosopografier med relationsdatabaser se: John Bradley & Harold Short, ”Using Formal Structures to Create Complex Relationships” *Resourcing Sources* (Oxford: Unit for Prosopographical Research, 2002); John Bradley & Harold Short, ”Texts into Databases: The Evolving Field of New-style Prosopography”, *Literary and Linguistic Computing* 20.1 (2005).

jag ensam bistått ett väl avgränsat forskningsprojekt under några år med sammanställning, utformning och presentation av en begränsad mängd material. Det säger sig självt att kvalitet och omfattning av prosopografierna i den här studien inte kan mäta sig med de projekt jag presenterar nedan.

5.3 Prosopografiska tillämpningar – en översikt

Prosopografiska metoder har traditionellt handlat om studier av samhällets manliga eliter och används särskilt av senantik- och medeltidshistoriker.⁴⁴⁹ För dessa tidsperioder består källmaterial oftast av information, vanligen inskriptioner, om religiösa och politiska eliter.⁴⁵⁰ Dessa prosopografiska är beroende av källornas detaljrikedom, ursprung och pålitlighet, något som också påverkar analyserna av materialet.⁴⁵¹

I projektet *Formering för offentlighet* har det skrivits texter om kvinnor, och för vissa av dem är dessa texter baserade på ett tämligen magert material. Ofta kan det vara problematiskt, även i modernare källor, att finna uppgifter om kvinnor eller om personer i samhällets lägre sociala klasser. Boken *Obemärkta* innehåller beskrivningar om elva kvinnors och mäns liv och verksamhet. Bristen på källor och material är vanligare för kvinnor än för männen framför allt på grund av den (då) rådande könsordningen.⁴⁵²

Detta problem uppmärksammas bland annat i en prosopografi om Ryska kvinnliga revolutionärer under perioden 1870-1917. I den studien hade man stora svårigheter att hitta uppgifter om vissa individer, speciellt om de kvinnliga arbetare som deltog i revolutionen. Trots att flera av dem var berömda pionjärer, ansågs de möjligen som mindre viktiga för revolutionen. En annan orsak till bristen på källmaterial kan vara att kvinnor från de lägre samhällsklasserna inte efterlämnade några skrifter. För somliga kvinnor har forskarna skrivit en slags minibiografier och de har lagt ner mycket arbete på att rekonstruera uppgifterna om de kvinnliga arbetarna. Lättast var det att finna information om de kvinnor vars män också var aktiva med revolutionära aktiviteter.⁴⁵³

⁴⁴⁹ Se Averil Cameron, red, *Fifty years of prosopography: the later Roman Empire, Byzantium and beyond* (Oxford: Published for the British Academy by Oxford University Press, 2003); Katharine Keats-Rohan, *Resourcing sources* (Oxford: Linacre College (University of Oxford), Unit for Prosopographical Research, 2002).

⁴⁵⁰ Cameron, red, *Fifty years of prosopography: the later Roman Empire, Byzantium and beyond*.

⁴⁵¹ För en belysande diskussion om relationerna mellan att använda inskriptioner som källor och svårigheter att analysera skeenden i romarriket, se Werner Eck, "The Prosopographia Imperii Romani and Prosopographical Method" *Fifty years of prosopography: the later Roman Empire, Byzantium and beyond*, red. Averil Cameron, Proceedings of the British Academy, 118 (Oxford: Published for the British Academy by Oxford University Press, 2003) s. 17 f.

⁴⁵² Ronny Ambjörnsson & Sverker Sörlin, *Obemärkta: det dagliga livets idéer* (Stockholm: Carlsson, 1995) ssk. s. 10.

⁴⁵³ Anna Hillyar & Jane McDermid, *Revolutionary women in Russia, 1870-1917: a study in collective biography* (Manchester: Manchester University Press, 2000). Prosopografin innehåller uppgifter om ca 1200 kvinnor som på ett eller annat vis var involverade i ryska revolutionen.

Skillnaderna mellan omfattning, informationsrikedom och användningar av prosopografier varierar därför också. Exempel på svenska och väl avgränsade prosopografiska studier är Ingrid Heymans om omvårdnadsforskningens utveckling i Sverige⁴⁵⁴ samt Esbjörn Larssons historiska studie om Krigsakademien och kadettutbildningen på Karlberg.⁴⁵⁵ Larsson har i sin avhandling upprättat en prosopografi med uppgifter om över hälften av samtliga inskrivna kadetter vid Kungl. Krigsakademien under perioden 1792-1866.⁴⁵⁶

För vissa klassiska prosopografier har arbetet pågått under mycket lång tid. En av de äldsta prosopografierna är *Prosopographia Imperii Romani* (PIR), den första utgåvan publicerades 1874, och innehåller uppgifter om varje individ nämnd i texter från romarriket.⁴⁵⁷ Den andra utgåvan, som påbörjades 1933 har nu hunnit till bokstaven R. Ett annat exempel är den otroligt omfångsrika och avancerade databasen om det bysantinska riket. Projektet påbörjades 1950 och de arbetar fortfarande med att registrera all information om varenda individ någonsin nämnd i de bysantinska texterna.⁴⁵⁸

Två prosopografiska databaser handhas vid *Centre for Computing in the Humanities* vid *Kings College* i London, den ovan nämnda databasen om det bysantinska riket och en databas om det anglosaxiska England.⁴⁵⁹ De är exempel på den ”nya stilens prosopografi” enligt John Bradley och Harold Short.⁴⁶⁰ Skillnaden mellan traditionell prosopografi och den nya stilens prosopografi är framför allt att traditionella prosopografier sammanställdes som ett hjälpmedel för forskare att analysera databasliknande sammanställningar av stora historiska material, ofta bestående av artiklar, för varje individ i den grupp av personer som studerades.⁴⁶¹ En artikel i detta sammanhang kan liknas vid en registerpost där samma typ av uppgifter om levnadsbanor registreras för var och en av individerna. Strukturen är alltså förutbestämd vilket gör att artikeln skiljer sig från biografien som genre, som inte utmärks av samma stringens. Traditionella prosopografier publicerades ofta som böcker som bestod av en sammanställning av dessa artiklar och som var sökbara via olika index. Det säger sig självt att utan datorstöd blir sammanställningar av uppgifter om hundratal, kanske tusentals, individer ett gigantiskt arbete.

För traditionella prosopografier har datorstödet vanligen bestått i avskrifter av redan väl strukturerade källor, såsom uppgifter från mantalslängder och folk-

⁴⁵⁴ Ingrid Heyman, *Gånge hatt till: omvårdnadsforskningens framväxt i Sverige: sjuksköterskors avhandlingar 1974-1991*, Skeptronserien (Göteborg: Daidalos, 1995).

⁴⁵⁵ Larsson, *Från adlig uppfostran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866*.

⁴⁵⁶ Urvalet är gjort med utgångspunkt i två perioder, den första och den sista perioden under Krigsakademiens tid. Ibid. s. 57.

⁴⁵⁷ Eck, ”The *Prosopographia Imperii Romani* and Prosopographical Method”.

⁴⁵⁸ Projektets syfte presenteras vid <www.pbw.kcl.ac.uk/content/aboutpbw/aboutpbw.html>.

⁴⁵⁹ Se *Prosopography of the Byzantine Empire*, tillgänglig vid <www.pbw.kcl.ac.uk/> samt *Prosopography of Anglo-Saxon England*, <www.pase.ac.uk/>.

⁴⁶⁰ Bradley & Short, ”Using Formal Structures to Create Complex Relationships”; Bradley & Short, ”Texts into Databases: The Evolving Field of New-style Prosopography”.

⁴⁶¹ Bradley & Short, ”Texts into Databases: The Evolving Field of New-style Prosopography”.

räkningar.⁴⁶² För den nya stilens prosopografi är däremot produkten, det vill säga den prosopografiska databasen, ett slutmål i sig självt. Databaskonstruktionen har inte bara varit ett hjälpmedel för modellering av innehållet utan har även medverkat till en förändring av arbetssättet:

We have found that a highly structured model works well for our prosopographies when we use the database not to model the relevant content of the source materials (which, because many of the texts are historical narrative sources, are of course not well served by relational structures), but to model some aspects of how the prosopographer thinks about their materials, and their task.⁴⁶³

I den nya stilens prosopografi har författandet av artikeln, som tidigare var startpunkten för prosopografin, förflyttats till något som sker efter det att uppgifterna om individerna samlats in, ordnats och granskats, vilket presenteras i avsnittet om *faktoider* nedan. Arbetet med de prosopografiska databaserna ger alltså inte bara möjligheter att sammanställa mer komplexa material och utföra mer avancerade analyser än tidigare, utan kan även ge upphov till en förändring av forskningens praktik.

Detta förändrade arbetssätt har resulterat i att man i dessa prosopografier använder en specifik princip för att organisera uppgifter om en viss person. Denna organiserande princip har man valt att benämna *faktoid*:

The ironic flavour of the name 'Factoid' is not accidental. It reflects the historian's worry when a tiny extract is taken out of the context of a larger text and the historical period in which it was written and presented as a 'fact'.⁴⁶⁴

Principen att använda faktoider organiserar de data som registreras enligt följande principer: Denna *källa* vid *denna tidpunkt* hävdar *detta* om *denna person*.⁴⁶⁵

[...] historical persons are formed out of their references in various surviving sources.⁴⁶⁶

Dessa principer syftar alltså till att relatera och samla de källor som refererar till en specifik person. Med hjälp av denna metod kan man inte bara strukturera information om personerna utan även om deras tillgångar och egenskaper såsom boende- och verksamhetsort, yrke etcetera.

⁴⁶² Ibid., s. 2.

⁴⁶³ Ibid. s. 1.

⁴⁶⁴ Ibid. s. 6.

⁴⁶⁵ För en utförlig beskrivning av databasens principer, uppbyggnad och användningar, se Bradley & Short, "Using Formal Structures to Create Complex Relationships" s. 3-21.

⁴⁶⁶ John Bradley, "Documents and Data: Modelling Materials for Humanities Research in XML and Relational Databases", *Literary and Linguistic Computing* 20.1 (2005) s. 138.

I dessa prosopografier är inte analyser av aggregerade data det mest intressanta utan det är *källorna* som är av primärt intresse. Dessa sammanlänkade faktoider fungerar därmed som verktyg för att med olika grad av säkerhet kunna fastställa, eller utesluta, historiska individers identitet, levnadsbanor, verksamhet, sociala nätverk och öden.

Bradley och Short framhåller också att det väl strukturerade innehållet ger flera möjligheter till att utföra avancerade sökningar och sammanställningar samt tillhandahåller flera ingångar i materialet baserat på till exempel person, plats, tid eller typ av källa.⁴⁶⁷ Med utgångspunkt i litteraturen om prosopografiska metoder inom historieämnet konstaterar jag att kvantitativa analyser hittills inte varit det primära intresset inom vare sig traditionella eller den ”nya stilens prosopografi”.

För den mer sociologiskt intresserade är aggregaten av karaktäristika och analyser därav det primära och inte i första hand källkritiska analyser. Detta är också något som var en utgångspunkt i projektet *Formering för offentlighet*. Som stöd för en stringent insamling av uppgifter och inför författandet av artiklar till prosopografen, som i projektet benämndes ”kapitalbeskrivningar”, användes en mall med riktlinjer för vilka typer av uppgifter som behövde registreras. Själva registreringen av uppgifter innebär därmed egentligen inte någon större skillnad jämfört med en traditionell historisk prosopografi. Den stora skillnaden torde vara med vilka utgångspunkter denna mall hade konstruerats. Vid konstruktionen av mallen använde man i projektet bland annat Bourdieus verktygslåda för studier av sociala fält.⁴⁶⁸ Ett av projektets syften var dock att utvidga gränserna för traditionella studier av fält genom att studera kvinnornas nätverk, något som påverkade konstruktionen av mallen och insamlingen av uppgifter.⁴⁶⁹

Vad som karaktäriserar den typ av prosopografiska studier som utförts inom den franska sociologiska traditionen och inom projektet *Formering för offentlighet* är det primära intresset att samla in data utifrån på förhand uppställda analysvariabler – som i sin tur grundar sig i studier av sociala fält – till vilka källorna förhåller sig sekundärt. Denna frågedrivna metod skiljer sig åt jämfört med de rigoröst källkritiska prosopografier har presenterats ovan. Det andra som utmärker dessa prosopografiska studier är användningen av korrespondensanalyser för analyser av sociala fält. Sådana analyser illustrerar individers positioner och fördelning av tillgångar i de sociala rum som konstrueras med utgångspunkt i individernas egenskaper och tillgångar.⁴⁷⁰ Databasen är inte

⁴⁶⁷ John Bradley & Harold Short, ”Texts into Databases: The Evolving Field of New-style Prosopography”, *Ibid.*

⁴⁶⁸ Ett socialt fält kan i korthet beskrivas som: *[e]tt system av relationer mellan positioner besatta av människor och institutioner som strider om något för dem gemensamt*. Broady, et al., *Formering för offentlighet. En kollektivbiografi över Stockholmskvinnor 1880-1920*, s. 4. Se även t.ex. Bourdieu, *Homo academicus*.

⁴⁶⁹ Se s. 155 f. för vidare diskussion om mallen, mallen i sin helhet redovisas i Bilaga 1, se <CD:/bilagor/bilaga1.pdf>.

⁴⁷⁰ En annan och mer korrekt benämning på dessa analyser är ”geometrisk dataanalys” (Geometric Data Analysis). Standardverket inom detta område är: Brigitte Le Roux & Henry Rouanet,

slutmålet utan har framför allt varit ett medel för geometriska dataanalyser för studier av fördelning av individernas tillgångar och egenskaper som ovan beskrivits. Broady föreslår därför benämningen *geometrisk prosopografi* för dessa mer sociologiskt inriktade prosopografiska studier.⁴⁷¹

Det som särskiljer geometrisk prosopografi från den nya stilens prosopografi är att den senare utmärks av det primära intresset för källorna, modelleringen av materialet samt själva produkten, det vill säga databasen. Av sekundärt intresse är statistiska analyser – som är primärt för geometrisk prosopografi – eller författande av artiklar. Medan det för traditionella historiska prosopografiska studier är källor och analysvariabler *per se* samt författandet av artiklar som är av större intresse än modelleringen av innehållet och statistiska analyser.

Målsättning med utformningen av det prosopografiska materialet i den här studien har, som ovan nämnts, varit att förbereda det för flera olika användningsområden. För detta är kunskaper och metoder som utvecklats inom den nya stilens prosopografi applicerbara, speciellt deras fokus på att modellera innehållet. Om ”enbart” korrespondensanalyser och nätverksanalyser varit målet med prosopografien hade det varit tillräckligt att mata in uppgifterna i ett register.⁴⁷²

Avslutningsvis kan konstateras att utmärkande för samtliga prosopografiska tillämpningar är den systematiska användningen av strukturerande scheman (mallar) och variabler vid insamlingen av data. Det som skiljer de tre typerna av prosopografiska tillämpningar från varandra är framför allt med vilka utgångspunkter dessa strukturerande scheman och variabler konstrueras. Annat som är särskiljande är i vilken utsträckning den prosopografiska databasen eller sammanställningen används för vidare analyser samt vilken betydelse källgranskningen ges.

Geometric data analysis: from correspondence analysis to structured data analysis (Dordrecht: Kluwer Academic, 2004).

⁴⁷¹ Anförande vid konferensen ”Exploring New Methods for Prosopography in the Humanities and the Social Sciences” finansierad av European Science Foundation och Riksbankens Jubileumsfond, 10-11 maj 2007.

⁴⁷² Uppgifterna hade då kunnat matas in i t.ex. Excel, för analyser och för vidare överföring till andra format i statistiska program som t.ex. SPSS och SPAD.

5.4 Introduktion till projektet Formering för offentlighet

Forskningsprojektet ”Formering för offentlighet. En kollektivbiografi över Stockholmskvinnor 1880-1920” pågick under åren 2000 till 2006.⁴⁷³ Projektets syfte var att studera kvinnors vägar till offentligheten – med särskilt intresse för Stockholm som mötesplats för dessa kvinnor. I projektet ville man kartlägga kvinnornas sociala nätverk och mötesplatser samt studera deras betydelse för olika sociala fält; särskilt har deras insatser inom pedagogik, filantropi, sjukvård och kulturliv studerats.

De centrala frågorna i projektet har varit att studera kvinnornas tillgångar, såväl nedärvda som förvärvade samt vilka strategier kvinnorna utvecklade för att vinna tillträde till det offentliga samhällslivet. Vilken funktion hade sällskap, föreningar och sammanslutningar som mötesplatser? På vilket sätt användes de sociala nätverken för inträdet i offentligheten? Dessa och liknande frågor har varit vägledande.

Som forskningsmetod har huvudsakligen Bourdieus verktygslåda används. Det innebär att kapital och fält har använts. Här kunde dessa begrepp prövas på ett material som troligen skulle bjuda en del motstånd.⁴⁷⁴

En av svårigheterna med att studera dessa offentliga kvinnors insatser och verksamheter är att de var verksamma inom flera olika fält och verksamhetsområden samtidigt. En viktig del i projektet Formering för offentlighet var därför att studera kvinnors insatser i skilda miljöer och sociala kontaktnät:

[---]som ofta torde ha varit tämligen osynliga från en utkikspunkt inom de etablerade fälten, nämligen de *nätverk* som kvinnor upprättade och utnyttjade som en bas varifrån de kunde göra sin tids stora frågor till föremål för offentlig diskussion.⁴⁷⁵

Om man – som Broady framhåller i projektplanen – studerar kvinnors insatser inom olika områden från *ett* perspektiv inom *ett* fält eller verksamhetsområde, till exempel politiken, blir det svårt att urskilja de enskilda insatserna eller att upptäcka eventuella skuggfält.⁴⁷⁶ Genom att kombinera studier av fält med studier av sociala nätverk kan man möjligen fånga in aspekter som inte upptäcks med studier av fält. Skillnaden mellan nätverk och fält i detta sammanhang är att i ett *nätverk* avses relationer mellan personer medan ett *fält* avser relationer

⁴⁷³ Projektinformation tillgänglig vid <www.skeptron.uu.se/broady/sec/ffo.htm>.

⁴⁷⁴ Se Broady, et al., *Formering för offentlighet. En kollektivbiografi över Stockholmskvinnor 1880-1920*.

⁴⁷⁵ Citat från den reviderade projektplanen, daterad september 2000, tillgänglig:

<www.skeptron.uu.se/broady/sec/p-ffo-00.htm>.

⁴⁷⁶ I projektplanen tas flickskolorna upp som exempel på skuggfält, som fungerade och fanns samtidigt, men i skuggan av, som läroverken. Begreppet skuggfält diskuteras bl.a. i: J.P. Roos & Anna Rotkirch, ”Fält i skuggan av fält: sensocialismens dubbelliv” *Kulturens fält: Cultural fields: en antologi*, Skeptronserien, (Göteborg: Daidalos, 1998).

mellan positioner.⁴⁷⁷ I sociala nätverk har institutionerna ingen plats eller funktion, utan det handlar om mänskliga relationer. Det sociala nätverket kan i sin tur användas på minst två metodologiskt olika sätt; dels som ett kvantifierbart mått på mellanmänskliga relationer, dels som ett sätt att analysera dessa relationers kvalitet, det vill säga deras innehåll och funktion.⁴⁷⁸ Ambitionen att studera kvinnornas sociala nätverk innebar att prosopografin kompletterades med märkning som möjliggjort analyser av kvinnornas sociala nätverk.⁴⁷⁹ I avhandlingens nätverksanalyser har dock analyser av relationernas innehåll och kvaliteter lämnats åt sidan, framför allt på grund av att denna typ av information inte finns tillgänglig i tillräckligt många kapitalbeskrivningar.

Lite drygt ett dussintal medarbetare har varit knutna till projektet. En rad studier och en mängd publikationer har producerats, bland annat en avhandling om kvinnokonferenser och Svenska Kvinnors Nationalförbund, en bok om "stiftarinnegenerationen", som handlar om tre kvinnliga skolgrundare och deras livsverk, samt ett flertal skrifter om kollektivbiografisk metod och artiklar om lärarinneutbildning och lärarinnor.⁴⁸⁰

Min roll i projektet har varit den tekniske redaktörens. Efter att projektet pågått ett par år blev jag involverad i sammanställning och organisering av insamlat material. Jag har inte medverkat i konstruktion av mallen eller författandet av levnadsbeskrivningarna (kapitalbeskrivningar), ej heller har jag gjort de avskrifter som ligger till grund för den största delmodulen i FFO-databasen, FFO-LUNDSTROM. För just den modulen av FFO-databasen har jag så att säga tagit hand om material som andra konstruerat, skrivit och författat. Detta innebar att jag från början inte hade så god kännedom om innehållet i materialet och att ett nära samarbete med projektdeltagarna var nödvändigt. Det innebar även att jag, speciellt i början, gjorde en hel del misstag framför allt när det

⁴⁷⁷ Termen nätverk används inom en rad olika områden och kräver därför en precisering. Socialt nätverk betecknar här som "personliga relationer mellan människor som är nära förbundna med varandra." se Donald Broady, "Nätverk och fält" *Sociala nätverk och fält*, red. Håkan Gunneriusson, Opuscula Historica Upsaliensia, 28. (Uppsala: Historiska institutionen, Uppsala Universitet, 2002) s. 51. Termen fält definieras vanligen som ett "system av relationer mellan positioner besatta av människor och institutioner". Broady, "Nätverk och fält" s. 49.

⁴⁷⁸ Jfr. Niklas Stenlås, "Det sociala nätverkets logik eller vad händer i sociala nätverk?" *Nätverk i historisk forskning: metafor, metod eller teori?* red. Peter Aronsson, et al., Rapporter från Växjö universitet. Humaniora, 1 (Växjö: Växjö universitet, 1999).

⁴⁷⁹ Hur detta genomförts beskriver jag i avsnittet om märkning på s. 172 ff.

⁴⁸⁰ Projektets medarbetare har under årens lopp publicerat en rad olika skrifter om de kvinnliga pionjärerna: om stiftarinnegenerationen, om pedagogerna, om de socialt verksamma, om litteraturkritiker och kulturlivet samt om metodiska aspekter. Se t.ex.: Agneta Linné. "Lärarinnor, pedagogiskt arbete och offentlighet. Kvinnliga pionjärs liv och strategier kring år 1900 - ett jämförande perspektiv". *Svenska Historikermötet*, 22-24 april, Uppsala, 2005; Linné & Skog-Östlin, red, *Biografi och kollektivbiografi som historisk metod: NFPF:s kongress 15-18 mars 2001 i Stockholm*, Lovisa af Petersens, *Formering för offentlighet: kvinnokonferenser och Svenska kvinnornas nationalförbund kring sekelskiftet 1900*, Stockholm studies in history, 87 (Stockholm: Acta Universitatis Stockholmiensis, 2006); Ullman, *Stiftarinnegenerationen: Sofi Almqvist, Anna Sandström, Anna Ablström*. En förteckning över dessa skrifter finns tillgänglig vid <www.skeptron.uu.se/broadly/sec/ffo-bib.htm>.

gällde att beskriva och klassificera innehållet.⁴⁸¹ Däremot har jag medverkat till beslut om vilka indikatorer som kunde användas och som skulle klassificeras. Ett flertal revideringar har genomförts av vilka uppgifter som skulle eller kunde användas för statistiska analyser. Klassificering av föreningar, utbildning, yrkesverksamhet etcetera har också genomgått ett antal revideringar och ändringar.

Målsättningen har redan från början varit att iordningställa materialet för flera olika användningar. Det skulle vara möjligt att utföra olika slags kollektiva och statistiska analyser samt att publicera kapitalbeskrivningar och annat material på webben.⁴⁸² Vilka metoder och verktyg som valdes för att iscensätta denna målsättning är bland annat det avhandlingen handlar om.

5.5 Modelleringens utgångspunkter

I följande avsnitt presenteras modelleringens två utgångspunkter; den ena är klassificering och den andra är datamodellen. De principer som ligger till grund för hur materialet klassificeras och hur det beskrivs och ordnas påverkar, tillsammans med själva innehållet, dels de arbetsmoment som står till buds, dels den färdiga produkten och dess användningsområden.⁴⁸³

5.5.1 Klassificering

Modellering av innehållet inbegriper även klassificering som tjänar två huvudsakliga syften där båda två är lika viktiga: att separera och att gruppera. Resultatet av klassificeringen synliggör därmed, i bästa fall, de ställningstaganden och de utgångspunkter som utgör de kategoriserade objektens karaktäristika.⁴⁸⁴ En grundläggande uppgift för klassificeringar är alltså att på ett systematiskt sätt indela och organisera fenomen eller objekt⁴⁸⁵, vilket samtidigt innebär vissa prioriteringar och urvalsförfaranden. Detta kan ske utifrån teorier eller utifrån mer pragmatiska överväganden.

Ett exempel på en välkänd klassificeringspraktik är biblioteksvärldens klassifikationssystem. Under slutet av 1800-talet skapades bibliotekens universella klassifikationssystem som grundades på föreställningar om att klassificeringen och dess struktur i en objektiv mening korresponderade mot kunskapsområdets

⁴⁸¹ Dessa aspekter diskuteras närmare i relation till relevant avsnitt samt i avhandlingens sammanfattande diskussion.

⁴⁸² Material som iordningsställts för projektets räkning finns tillgängligt eller hänvisat till vid <www.skeptron.uu.se/monicalz/ffo/>.

⁴⁸³ En färdig produkt är svårt att åstadkomma i en modulariserad design, den prosopografiska databasen är inte färdig i ordets rätta bemärkelse då den hela tiden har utökats och förändras. Genom att publicera avhandlingen är databasen som den ser nu en "färdig" produkt.

⁴⁸⁴ Michael Sperberg-McQueen, "Classification and its Structures" *A companion to digital humanities*, red. Susan Schreiban, et al., Blackwell companions to literature and culture; 26 (Blackwell Pub., 2004).

⁴⁸⁵ Objekt är den tekniska term som används i klassificeringsteori. Ibid.

egen struktur.⁴⁸⁶ Detta synsätt ifrågasätts från många håll inte minst inom vetenskapssociologin.⁴⁸⁷ I den här studien rör det sig om att klassificera och konstruera objekt med utgångspunkt i relativt konkreta sociala företeelser som utbildning, socialt ursprung och individers verksamhetsområden. Ett bra exempel på svårbedömda jämförelser är klassificeringen av kvinnornas yrkestitlar under den studerade perioden kring sekelskiftet 1900. Den klassificering som nu används baseras på en specifik grupp och inte i jämförelse med befolkningen i stort. Därtill bör tolkningarna utgå från andra faktorer än exempelvis enbart yrkesverksamhet och för detta ändamål används indikatorerna för socialt och olika slags symboliskt kapital.

Att beskriva och att ordna genom att klassificera är i praktiken ofta föremål för återkommande förhandlingar, omförhandlingar och förändringar, något som ingår som en del av modelleringen. Begreppet klassificering syftar alltså till *resultat* såväl som själva verksamheten att *klassificera*. En av de grundläggande principerna för klassificering är just nödvändigheten av att somligt innehåll inkluderas och accentueras samtidigt som annat dämpas eller exkluderas. Ett objekt betecknas av vad det "är" i samma utsträckning som vad det inte "är".

Klassificeringen kan, enligt Geoffrey Bowker och Leigh Star beteckna:

Verksamheter och infrastrukturer, som exempelvis utbildning, sjukdomar eller yrkesklassificeringar. Många av dessa klassificeringar är inbäddade i föreställningar om att det handlar om vedertagen praxis. De historiska utgångspunkterna är ofta bortglömda och klassificeringarna anses vara både triviala och tråkiga och därmed inte är värda att undersökas närmare.

System och specifika regler, systematiska mätningar och behandlingsmetoder inom medicinen, tekniker och förklaringsmodeller för indelning av exempelvis människor är exempel på detta.

De tillämpade klassifikationerna, det vill säga hur verksamheter och system organiserar och organiseras av praktiken. I boken används sjuksköterskornas systematiserade aktiviteter, rutiner och normer som studieobjekt och exempel. Andra exempel på tillämpade klassificeringar är på det sätt undervisning organiseras i föreläsningar, seminarier och lektioner; det vill säga de aktiviteter som både formar och formas av undervisningens praktik.⁴⁸⁸

⁴⁸⁶ Se Karin Pettersson, *Den manliga heterosexuella normen: en analys av SAB-systemet utifrån ett feministiskt och queerteoretiskt perspektiv* (Borås: Högsk. i Borås Bibliotekshögskolan/Biblioteks- och informationsvetenskap, 2001).

⁴⁸⁷ Se t.ex. Bourdieu, *Homo academicus*; Karin D. Knorr-Cetina, *Epistemic cultures: how the sciences make knowledge* (Cambridge, Mass.: Harvard University Press, 1999); Andrew Pickering, *Science as practice and culture* (Chicago: Univ. of Chicago Press, 1992).

⁴⁸⁸ Geoffrey C. Bowker & Susan Leigh Star, *Sorting things out: classification and its consequences*, Inside technology, (Cambridge, Mass.: MIT Press, 1999) s. 17-28.

Indelningen ovan är användbar för att klargöra skillnaden mellan de metoder och verktyg som används i studien och i tillämpningarna. Hur man klassificerar, beskriver och ordnar forskningsmaterial styrs ju av en rad olika faktorer, såsom vedertagen praxis, normer eller forskningsområden. Klassificeringen i sig har därmed alltid en bias som konstitueras genom dess sammanhang, historiska förändringar och rumsliga förflyttningar. Detta är något som formella teorier för informationshantering försöker överkomma genom att:

[---] regularize the movement of information from one context to another; to provide a means of access to information across time and space.⁴⁸⁹

Ett exempel på ett försök att tillämpa en sådan formell logik är den hierarkiska dokumentmodellen. Denna formella modell påverkar sedan hur ett visst innehåll konkret beskrivs och ordnas relativt en abstrakt modell. Modellen tjänar då som redskap för beskrivning och ordning av materialet men påverkas i sin tur av teoretiska utgångspunkter eller av frågor av mer praktisk natur såsom källmaterialets kvalitet, forskningsgruppens sammansättning och kunskapsområden. Därmed ges även en formell modell en visst mått av regional eller lokal förankring genom de logiker som är rådande i det sammanhang där den brukas. Jag menar att detta inte går att undkomma, frågan är snarare i vilken utsträckning man bör ta hänsyn till lokala eller regionala logiker för att modifiera en formell och allmängiltig modell? Denna problematik har jag brottats med flera gånger under arbetet med utformningen av materialet. När kan, eller bör, materialets ”egna” eller ”inneboende” egenskaper ta över den formella modellens logik?⁴⁹⁰ Om den formella modellens logik strikt tillåts råda riskerar vissa innehållsobjekt att hamna utanför – vilket resulterar i att vi får residualkategorier. Dessa residualkategorier, ofta hopslagna som ”övriga”, påverkar med sin tvetydighet ändå strukturen och ordningen i ett material (eller i ett socialt sammanhang för den delen) och kan ge upphov till en kumulativ röra som senare kan bli svår att reda ut.⁴⁹¹ Frågeställningar liknande den ovan, om det specifika för ett visst projekt eller kunskapsområde visavi en mer generell design, kommer jag att diskutera i den senare delen av avhandlingen.⁴⁹²

Vilka tillvägagångssätt kan användas för att försöka undvika en alltför lokal klassificeringslogik? För att bryta vår vardagliga förståelse behöver vi rekonstruera alla indikatorer som förutsätter det studerade fenomenet. Dessutom behöver vi vaska fram de relevanta egenskaper inom det specifika fält där dessa

⁴⁸⁹ Ibid. s. 290.

⁴⁹⁰ För exempel och diskussion se s. 196 f. Se även Ibid. s. 290-293; McCarty, *Humanities computing* s. 168-170.

⁴⁹¹ Bowker & Star, *Sorting things out: classification and its consequences* s. 39.

⁴⁹² Detta är ett problem som har uppstått i utformningen av det prosopografiska material som sammanställdes allra först. Se diskussion på s. 196 f.

egenskaper har någon betydelse – i det här fallet betyder det de kapitalformer som har något värde inom det fält som studeras.⁴⁹³

Ett relationellt synsätt på konstruktioner av fakta är framträdande hos Ludwik Fleck som menar att vårt sätt att uppfatta fakta och att tolka fakta är en produkt av våra vardagsföreställningar.⁴⁹⁴ Vardagsföreställningarna påverkas av tankestilens tvång som i sin tur är grundad i vår medverkan i olika tankekollektiv. En kunskapsprocess består i att skapa användbara resultat – faktum – ur givna förutsättningar, det vill säga den kunskap vi redan har. Resultaten påverkas av ett visst tvång – tankestilen – som har en stiliserande effekt och som uppfattas av oss som den objektiva verkligheten. Alltså; de kategorier vi skapar, är produkter av rum såväl som tid, något som har betydelse för hur vi tolkar de egenskaper och fenomen som vi sedan försöker beskriva. Att ha bostad på Östermalm vid tiden runt förra sekelskiftet innebar inte alls samma status som idag. Att som kvinna studera på universitetet för 100 år sedan var en kamp och förbehållet några få, att kvinnor dessutom inte kunde få statlig anställning efter utbildningen påverkade förstås också familjens benägenhet att bekosta utbildning för sina döttrar.⁴⁹⁵

I den här studien handlar klassificeringen konkret om att beskriva, ordna och särskilja uppgifter om gemensamma karaktäristika och tillgångar för individerna i det prosopografiska materialet. Intresse riktas mot skillnader i konstruktioner av vetenskapliga analytiska kategorier av sociala företeelser genom användning av antingen (externt) etablerade klassificeringar eller internt konstruerade klassificeringar som i sin tur är avsedda att spegla det specifika innehållets interna logik och ordning. Av särskilt intresse är dock den ovan beskrivna, och ofrånkomliga, diskrepansen mellan formella modeller och regionala logiker.⁴⁹⁶ För att förankra denna diskussion om formella modeller och regionala logiker kan en närmare beskrivning av det material som används som studieobjekt presenteras närmare, innan dess en diskussion om den här prosopografins utgångspunkter.

5.5.2 Datamodellen

För modellering av ett prosopografiskt material behövs som ovan beskrivits någon sorts modell eller princip för insamling av data – förutom metoder och verktyg för beskrivning och registrering av data.

I den här studien används, som tidigare nämnts, TEI-schemat som metod och som verktyg för beskrivning och lagring av texterna. I TEI:s nya prosopo-

⁴⁹³ Bourdieu, *Homo academicus* s. 44.

⁴⁹⁴ Fleck, *Uppkomsten och utvecklingen av ett vetenskapligt faktum: inledning till läran om tankestil och tankekollektiv*.

⁴⁹⁵ Se t.ex. Rönnholm, *Kunskapens kvinnor: sekelskiftets studentskor i mötet med den manliga universitetsvärlden*. En studie om det kvinnliga studenterna i en manlig miljö, deras sociala kapital, bakgrund och framtid kring förra sekelskiftet.

⁴⁹⁶ Se diskussioner i avsnitt 6.1 s. 172 ff.

grafiska modul som färdigställdes under 2007 finns element för beskrivning av just den sortens uppgifter som samlades in av projektets medlemmar.⁴⁹⁷ Vid tiden för författandet av kapitalbeskrivningarna fanns inte TEI:s prosopografiska modul utan då användes de element – som inte var optimala för beskrivningen av innehållet – som fanns att tillgå för ändamålet.⁴⁹⁸ TEI:s prosopografiska modul i P5 har dock använts för märkning av delar av materialet om Kungl. Högre Lärarinneseminariet.⁴⁹⁹

Modelleringen av de två materialen, FFO-KAPITAL⁵⁰⁰ och FFO-HLS,⁵⁰¹ har haft olika karaktär dels på grund av skillnader i typen av texter, dels på grund av att insamlingen skett med olika utgångspunkter. I projektet *Formering för offentlighet* användes en mall vid författande av kapitalbeskrivningar. Detta medförde att innehållet modellerades relativt de verktyg, och teoretiska utgångspunkter, om studier av sociala fält och nätverk redan vid insamlingen.⁵⁰²

Till databasen FFO-HLS användes arkivhandlingar som källmaterial. Dessa arkivhandlingar var redan strukturerade enligt den praktiska modell som användes vid seminariet för registrering av uppgifter om eleverna. Den största skillnaden har således med materialets karaktär och texternas genre att göra: i FFO-KAPITAL, används bearbetat källmaterial och i FFO-LUNDSTROM sekundärmaterial i form av avskrifter ur en biografisk uppslagsbok. I FFO-HLS används faksimil och avskrifter av arkivmaterial.⁵⁰³

Att använda den hierarkiska dokumentmodellen på ett redan modellerat material har inte varit problemfritt.⁵⁰⁴ Bradley och Short hävdar att den hierarkiska dokumentmodellen och de XML-tillämpningar som finns att tillgå i dag inte är tillräckligt komplexa för modellering av prosopografiskt material.⁵⁰⁵ De förespråkar istället relationsdatabaser som är adekvat utformade och innehåller många relationer. Registrering av data i en enda tabell är inte att betrakta som en relationsdatabas. De menar att många historiska projekt strävar efter att konstruera digitala representationer av en förlaga istället för att skapa nya digita-

⁴⁹⁷ TEI P5 publicerades i november 2007, betaversioner av schemat har dock funnits tillgängliga tidigare.

⁴⁹⁸ Här användes modulen för prosa med tilläggsmodulen för namn och datum. En översikt av modulerna ges i avsnitt 4.5 s. 118 f.

⁴⁹⁹ En introduktion till den prosopografiska modulen i TEI P5 ges i 4.6 s. 126 f.

⁵⁰⁰ Benämning på det material som har sammanställts för projektet ”*Formering för offentlighet*”, se s. 152 för en kortare översikt.

⁵⁰¹ Benämning på det material som har samlats in och sammanställts utifrån arkivet för Kungl. Högre Lärarinneseminariet, se s. 152 för en kortare översikt.

⁵⁰² Jag deltog inte i den initiala registreringen av uppgifter och har alltså inte skrivit någon kapitalbeskrivning eller transkriberat text från de biografiska uppslagsböckerna. Däremot har jag reviderat själva databasen löpande genom att samla in och infoga nya uppgifter, som t.ex. revidering av föreningsaktiviteter och styrelseposter.

⁵⁰³ Digitala fotografier av arkivposterna samt avskrifter av textinnehållet i dessa foton.

⁵⁰⁴ Detta diskuteras och konkretiseras i de två nästföljande kapitlen om märkning och utformning av de två prosopografiska materialen.

⁵⁰⁵ Bradley & Short, ”*Texts into Databases: The Evolving Field of New-style Prosopography*”.

la versioner där analys och utgångspunkter är synliga och som därmed kan tillföra något nytt, annat än själva digitaliseringen.⁵⁰⁶

Som tidigare nämnts förutsätter den hierarkiska dokumentmodellen att en övergripande struktur ges företräde medan andra strukturer får stå tillbaka.⁵⁰⁷ Detta har förstås sina nackdelar och har lett till att jag använt en del innovativa och oortodoxa lösningar för beskrivning av materialet, speciellt materialet i FFO-KAPITAL som är relativt ostrukturerat till sin *form* men som har en relativt tydlig innehållslig struktur.⁵⁰⁸ Materialet i FFO-HLS har däremot en tydlig struktur både till form och innehåll och där har TEI-schemat använts utan några anpassningar till någon projektspecifik datamodell som kapitalbeskrivningsmallen var för det andra materialet. Detta har betydelse för hur beskrivningen och själva märkningen av materialen gått till förutom det faktum att TEI:s schema uppdaterats mellan första och andra beskrivningstillfället.

De metoder och verktyg som används för modellering av prosopografiskt material i denna studie kan beskrivas i fyra punkter:

- *Projektspecifika datamodeller*: till exempel de mallar för kapitalbeskrivningar som använts för insamling och registrering av sekundärmaterial i databasen FFO-KAPITAL.
- *Generella hierarkiska dokumentmodeller*: i detta sammanhang avses de TEI-moduler som använts för märkning av det prosopografiska materialet
- *Generella principer* för beskrivning av semantiska innehållsobjekt: sådana generella principer avser beskrivning av till exempel egennamn, platser, datum och dylikt, det vill säga sådana egenskaper som kan beskrivas med relativt distinkta regler för tolkning. Till detta hör även sådana egenskaper som kan kontrolleras mot använda källor eller valideras mot till exempel auktoriserade namnformer.⁵⁰⁹
- *Projektspecifika klassificeringar av element*, som till exempel klassificering av typer av egennamn, olika typer av organisationer och föreningar för att beteckna ett visst innehållsobjektet på ett distinkt sätt och därmed kunna särskilja elementets specifika egenskaper relativt andra liknande element.

⁵⁰⁶ Ibid. s. 5.

⁵⁰⁷ Denna problematik har diskuterats i avsnitt 4.4 s. 113 f.

⁵⁰⁸ Dessa oortodoxa lösningar har bl.a. inneburit att jag använt element på ett ibland icke avsett sätt. Varför beskrivs närmare i avsnitt 6.1 s. 172 f.

⁵⁰⁹ T.ex. kontrollerade och normaliserade stavningar av personnamn och geografiska platsnamn. Särskilt viktigt för inmatning av poster vid arkivinstitutioner och för bibliotekskatalogisering. För svenska auktoriserade namnformer tillhandhålls sådana listor bl.a. av Kungl. Biblioteket, se vidare KB:s auktoritetspraxis vid: <www.kb.se/auktoritetspraxis/>.

Punkterna ovan sammanfattar de metoder och huvudsakliga principer med vilka beskrivning och ordning av materialet utförts. Dessa principer har dels sin utgångspunkt i beslut som togs innan märkningsarbetet påbörjades, dels har de reviderats eller vuxit fram under själva arbetsprocessen som pragmatiska lösningar på problem som uppstått. Användningen av hur TEI-schemat skulle användas bestämdes innan märkningen påbörjades, men de projektspecifika datamodellerna och kategoriseringarna har förändrats allt eftersom märknings- och analysarbetet fortskridit. Det visade sig så småningom, efter några revideringar, vilka lösningar som var utförbara och vilka som kunde överges. Ett exempel på detta är att märkningen från början tillät ett större mått av tolkning än vad som visade sig vara praktiskt genomförbart. Datamodellen modifierades därför så att den modell som sedan använts tillåter märkning av uppgifter för en teknisk redaktör, med små sakkunskaper om den enskilda kvinnan, ändå kan besluta om lämpliga element för beskrivning av innehållet. Beskrivningen avser företrädesvis sådant som kan kontrolleras internt inom projektet eller med hjälp av andra källor. Exempel på det är att jag inte märkt upp utsagor i kapitalbeskrivningarna om kvinnornas kapitalvolym.⁵¹⁰ Istället beskrivs och registreras indikatorer på sådant som senare kan användas som mått på kapitalvolym i senare analyser, som exempelvis utbildning och studieresor utomlands.⁵¹¹

Med utgångspunkt i datamodellen har jag vinnlagt mig om att beskriva och märka innehållsobjekten (till exempel en viss egenskap såsom kvinnans eller faderns yrke) så detaljerat och så nära den faktiska utsagan som möjligt. Tolkningen av materialets projektspecifika resultat eller kvinnornas fältspecifika egenskaper bör helst göras i analysen snarare än vid märkningen av materialet för att ge företräde åt en empirinära beskrivningspraktik framför en projektspecifik märkning. Självklart innehåller även den här sortens märkning tolkning, men det som eftersträvas är att inte frestas att föregripa kommande analyser redan under modellering och beskrivningen av innehållet.

I följande kapitel, som handlar om de två prosopografiska modulerna i FFO-databasen, beskrivs de principer som använts för beskrivning och ordnande av de olika slags källor som använts för tillämpningarna i den här studien. Nedan ges först en kort översikt av utformningen av den prosopografiska databasens moduler.

5.6 Introduktion till det prosopografiska materialet

I detta avsnitt ges en översiktlig beskrivning av de moduler som ingår i FFO-databasen och i detta avsnitt presenteras deras utformning. Därefter följer en presentation av modulen FFO-LUNDSTROM som i sin tur innehåller en del-

⁵¹⁰ T.ex. utsagor om att en viss kvinna hade stort mått av kapital inom det litterära fältet.

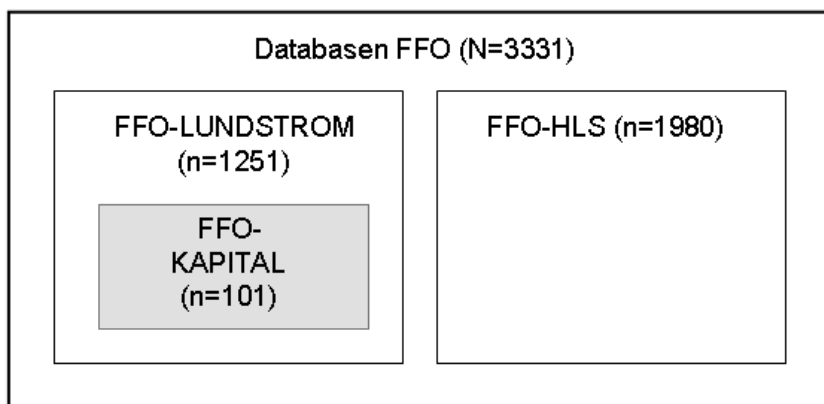
⁵¹¹ Fler indikatorer presenteras i avsnitt 5.7 s. 157.

modul benämnd FFO-KAPITAL. Dessa två moduler innehåller information av olika karaktär. I FFO-LUNDSTROM finns relativt sett få uppgifter registrerade för ett större antal kvinnor medan FFO-KAPITAL innehåller många uppgifter om ett mindre antal kvinnor. Slutligen presenteras den andra modulen i FFO-databasen, nämligen FFO-HLS.

5.6.1 Den prosopografiska databasens moduler

Det prosopografiska materialet utgörs av databasen FFO som i sin tur består av två huvudsakliga moduler FFO-LUNDSTROM och FFO-HLS. FFO-LUNDSTROM innehåller dessutom en delmodul, FFO-KAPITAL som består av de nyskrivna kapitalbeskrivningarna.⁵¹²

Målsättningen med databasen är att på ett flexibelt sätt organisera och skapa förutsättningar för analyser och presentationer av det prosopografiska material som samlats in. I figuren nedan illustreras databasens modulariserade utformning.



Figur 16. Illustration av FFO-databasens utformning

I databasen finns material i form av såväl kortare som längre texter samt digitala bilder (fotografier) av arkivmaterial. Databasen innehåller en mängd olika uppgifter om offentliga kvinnor samt lärarinnestuderande. Dessa texter, avskrifter, noteringar och arkivposter, har sedan märkts och organiserats med TEI:s riktlinjer och lagrats i TEI/XML-format. XML-filerna har fungerat som förutsättningar för olika slags transformeringar, bearbetningar och analyser, vilket presenteras i avsnitten 6.3, 6.4, 6.5 och kapitel 7.

⁵¹² Termerna modul och register används ibland synonymt med varandra. Ordet modul avser beskriva databasens logiska indelning, medan register nyttjas för det material som iordningställts för de statistiska analyserna. Databas används för att beteckna hela den prosopografiska samlingen, inklusive registerposter, digitala texter och digitala bilder.

Uppgifterna i prosopografin baseras som ovan nämnts på flera typer av källor. Till kapitalbeskrivningarna har till exempel biografier, biografiska uppslagsverk samt brev, matriklar och annat arkivmaterial använts.⁵¹³ Modulen FFO-LUNDSTROM, baseras på uppgifter hämtade ur biografiska uppslagsverk.⁵¹⁴

Den andra huvudsakliga modulen, FFO-HLS, i FFO-databasen baseras på digitaliserat arkivmaterial om 1980 kvinnor som utbildade sig vid Kungl. Högre Lärarinneseminariet från starten 1861 till skolan stängdes 1943.⁵¹⁵

Databasen har konstruerats så att modulerna kan användas var för sig eller tillsammans. Detta medför att nya moduler kan infogas, under förutsättning att de innehåller liknande typ av uppgifter som redan använts.⁵¹⁶ Likaså tillåter modulariteten tillämpning av valfri modul eller delar av en modul. Det går alltså bra att till exempel extrahera ett urval bestående av kvinnor födda i Stockholm och som tagit examen vid Kungl. Högre Lärarinneseminariet. I de två följande kapitlen illustreras hur dessa möjligheter nyttjats.

5.6.2 Offentliga kvinnor kring sekelskiftet 1900

I projektet *Formering för offentlighet* studeras de kvinnliga sociala nätverkens betydelse, framför allt i Stockholm, med betoning på ett antal föregångskvinnors insatser inom pedagogik, socialt arbete, sjukvård och kulturliv. Av särskilt intresse är kvinnornas strategier och formationer samt kvinnornas ackumulering och förvaltning av olika sorters symboliskt kapital.

Vilka kvinnor är det då som ingår i urvalet? I Lundströms biografiska uppslagsbok definieras en offentligt verksam kvinna som till exempel en politiskt engagerad person, läkare, advokat, lektor, föreståndare, författare, konstnär, journalist eller någon som på annat sätt "gjort sitt namn känt och uppmärksammat."⁵¹⁷ Det fanns förstås inte möjlighet för alla kvinnor att lämna det privata livet och träda ut i offentligheten utan det var framför allt de borgerliga kvinnorna. En och annan kvinna från de lägre samhällsklasserna finns också med i urvalet.⁵¹⁸

Uppgifterna som samlats in och registrerats är av olika slag men framför allt handlar det om kvinnornas sociala bakgrund, utbildningsnivå, verksamhetsom-

⁵¹³ För metoddiskussioner i anknytning till projektet *Formering för offentlighet*: Linné & Skog-Östlin, red, *Biografi och kollektivbiografi som historisk metod: NFPF:s kongress 15-18 mars 2001 i Stockholm*.

⁵¹⁴ I delkapitel 5.8 s. 162 ges exempel på vilket sorts uppgifter som använts.

⁵¹⁵ Se bilaga 7 för uppgifter om digitaliserade arkivhandlingar, <CD:/bilagor/bilaga7.pdf>.

⁵¹⁶ För en grundlig genomgång av utformning av prosopografiska databaser, se: Bradley & Short, "Using Formal Structures to Create Complex Relationships".

⁵¹⁷ Lundström, *Svenska kvinnor i offentlig verksamhet: porträtt och biografier*.

⁵¹⁸ I kapitalbeskrivningar är det en kvinna, Agda Östlund, som härrör från lägre samhällsklasserna. Agda var en av de första kvinnorna i riksdagen 1921. I FFO-LUNDSTROM (dvs. bland samtliga 1251 offentliga kvinnor) är det endast 8 kvinnor som kommer från det vi klassificerat arbetarklass. De allra flesta kvinnor härrör från samhällets övre sociala skikt. Se även Zetterman, *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet*.

råden, föreningsdeltagande och socialt umgänge. Dessa uppgifter har sedan märks med TEI/XML, bearbetats och organiserats i en prosopografisk databas. Den modul i FFO-databasen som detta kapitel handlar om består i sin tur av två delmoduler. Dels den största FFO-LUNDSTROM som innehåller uppgifter om totalt 1251 kvinnor, dels de 101 kapitalbeskrivningarna i delmodulen FFO-KAPITAL.⁵¹⁹

Vilka kvinnor som inkluderats i urvalet och vilka som exkluderats kan förstås diskuteras. Det är inte min uppgift att bedöma urvalet av kvinnor, men som jag förstår det har man i projektet haft ambitionen att i första hand studera de kvinnor som var verksamma under perioden 1880-1920 i det geografiska, sociala och kulturella rum som var Stockholm. Detta har också gjorts och de kvinnor som ingår i FFO-KAPITAL var i första hand pedagoger, socialt verksamma eller publicister verksamma i Stockholm. Därtill tillkommer dessutom de data som registrerats utifrån kortare lexikaliska biografiter om ytterligare kvinnor från Stockholm såväl som övriga delar av landet. Dessutom hör det till saken att tack vare FFO-dabasens modulära design är det möjligt att utvidga materialet genom att lägga till ytterligare personposter eller kapitalbeskrivningar. I delmodulen FFO-LUNDSTROM finns fler verksamhetsområden företrädade än i FFO-KAPITAL såsom kulturellt verksamma i vid mening (teater, musik, dans etcetera) samt kvinnor verksamma inom juridiska, tekniska och ekonomiska områden. Syftet med detta tilläggsmaterial var att skapa möjligheter till analyser på kollektivnivå samt att använda dem som en fond för det material som utgörs av kapitalbeskrivningarna.

5.6.3 Kungl. Högre Lärarinneseminiariets arkiv

I den här modulen av den prosopografiska databasen ingår samtliga kvinnor som skrevs in vid Högre Lärarinneseminarier och inte enbart offentligt verksamma kvinnor.⁵²⁰ Inledningsvis ges här en kort presentation om seminariets historia.

5.6.3.1 Bakgrund

Kvinnors möjligheter till högre utbildning var under 1800-talet begränsade. Detta förändrades dock under senare delen av 1800-talet och genom statliga reformer fick kvinnor även tillträde till offentliga arbetsområden där lönerna var relativt låga.⁵²¹ Folkskolorna var det första offentliga arbetsområde som

⁵¹⁹ Antalet kvinnor FFO-LUNDSTROM inkluderar de 101 kvinnorna i FFO-KAPITAL. Uppgifter i FFO-LUNDSTROM är baserad på avskrifter ur Lundström, *Svenska kvinnor i offentlig verksamhet: porträtt och biografier*. Uppgifterna i Lundströms biografiska uppslagsbok har transkriberats av Rakel Höglund.

⁵²⁰ Modulen kan antingen användas tillsammans med FFO-KAPITAL eller FFO-LUNDSTROM eller som en fristående modul.

⁵²¹ Mellan 1856-58 motionerades i riksdagen om att kvinnliga lärare skulle få undervisa i folkskolorna. Greta Wieselgren, *Den höga tröskeln: kampen för kvinnas rätt till ämbete*, Kvinnohistoriskt arkiv, 7 (Lund: Gleerup, 1969) s. 9-10.

kvinnor fick tillträde till och 1859 startade det första folkskollärarinneseminariet. Flickor och pojkar från de lägre sociala skikten undervisades tillsammans i folkskolan. I mitten av 1800-talet ansåg många att det borde räcka med att flickor fick gå i folkskolan, någon ytterligare utbildning behövde inte staten anordna.⁵²² Kvinnor i de högre sociala skikten hade förstås helt andra förutsättningar till utbildning än kvinnor hemmahörande i de lägre sociala skikten. I borgerskapets och överklassens hem sköttes ofta undervisningen av guvernanter och en del dem grundade sedermera så kallade flickpensioner som var föregångare till flickskolorna.⁵²³

Kritiken mot den undermåliga undervisningen vid många av flickpensionerna växte och 1861 inrättades ett statligt seminarium för utbildning av lärarinnor: Kungliga Högre Lärarinneseminariet. Föregångaren till seminariet var *Lärokursen för Fruntimmer* som pågick 1859-1861. Flera av lärarna och de studerande på *Lärokursen* blev sedan anställda på Kungl. Högre Lärarinneseminariet.⁵²⁴ I början av 1870 fick kvinnorna tillträde till universiteten⁵²⁵ men det skulle dröja ganska länge innan de fick behörighet till högre tjänster såsom tjänstgöring som adjunkt eller lektor vid de statliga läroverken.⁵²⁶

5.6.3.2 Arkivmaterialet

Högre Lärarinneseminiariets arkiv (SSA 2132) är omfattande och många handlingar från seminariets verksamhetsperiod finns kvar. För skolans samtliga verksamhetsår under perioden 1861–1943 finns uppgifter om elevernas födelse- och bostadsort, deras ansökningshandlingar, inträdesansökningar och skrivningar, valda kurser, betyg, uppförande, anmärkningar och frånvaroliggare med mera.⁵²⁷ Förutom detta finns även uppgifter om kvinnornas sociala ursprung, tidigare utbildningar, språkkunskaper, utlandsvistelser samt anställningar efter lärarinnexamen för perioden 1914–1943.

Därutöver tillkommer ett omfattande material om elevföreningar, protokoll från direktioner och lärarkollegium, journaler över gåvor och inköp, biblioteksarkiv (utlåning och bestånd), skrivelser, lektionsböcker, journaler, fotografier, kassaböcker, huvudböcker med mera.

⁵²² Rudolf Fåhraeus, *Högre lärarinneseminiariets historia* (Stockholm: Norstedt, 1943).

⁵²³ För en studie om dessa flickors undervisning under 1800-talet, se Angela Rundquist, *Blått blod och lifevita händer: en etnologisk studie av aristokratiska kvinnor 1850-1900* (Stockholm: Carlsson, 1989) s. 69 ff.

⁵²⁴ Fåhraeus, *Högre lärarinneseminiariets historia*.

⁵²⁵ Sekelskiftets kvinnliga studenter studeras bland annat i: Rönnholm, *Kunskapens kvinnor: sekelskiftets studentkor i mötet med den manliga universitetsvärlden*. Flera av kvinnorna i den studien finns med i FFO-databasen.

⁵²⁶ Om strider kring förslaget om kvinnors behörighet till tjänster på statliga läroverk samt om det något spända förhållandet mellan de seminarieutbildade lärarinnorna och de kvinnliga lärarna med akademisk utbildning, se Ullman, *Stiftarinnegenerationen: Sofi Almqvist, Anna Sandström, Anna Abström* s. 177-191.

⁵²⁷ Till och med de tio utlåningsliggarna från skolans bibliotek finns kvar, totalt 10 liggare under perioden 1881–1941.

Den första liggaren för inskrivningsuppgifter användes under hela verksamhetsperioden 1861–1943 och innehåller uppgifter om namn, inskrivningsdatum, diplom, examensdatum, studieår, födelseort, samt anmärkningar och faderns yrke.⁵²⁸ Samtliga 1980 poster i denna matrikel är digitaliserad, det vill säga fotograferad och avskriven.⁵²⁹

Därtill finns ytterligare tre inskrivningsmatriklar för perioden 1914–1943 som innehåller ytterligare uppgifter om elevernas sociala ursprung, tidigare utbildning, arbetsplats och anställning efter examen vid seminariet. Samtliga poster för åren 1914–1943 är fotograferade, tillika ett urval av de detaljerade betygskataloger som finns för varje läsår, ett antal frånvaroliggare, ansökningshandlingar samt vissa årsböcker.⁵³⁰ Ett urval av posterna i detta material är avskrivna och har försetts med beskrivande märkning med TEI:s schema för prosopografiska material.⁵³¹

Därtill tillkommer ytterligare en källa som i vissa fall kan komplettera arkivposterna genom att utvidga informationen om lärarinnelevernas liv efter examen. I en utgåva från 1943, samma år som skolans lades ned, har Valter Fevrell samlat minnesrunor om några av lärarna, föreståndarinnorna och eleverna på skolan.⁵³² Totalt 47 före detta elever har fått en minnesruna i boken där 36 av dem tjänstgjorde som lärarinnor eller föreståndarinnor vid andra läroanstalter. Resterande 11 före detta lärarinnelever var verksamma inom andra områden än det pedagogiska.⁵³³ I dessa minnesrunor finns, i varierande omfattning, en del användbara uppgifter för studier av livet efter examen. Några av kvinnorna i den boken finns även med i FFO-KAPITAL.⁵³⁴

Syftet med att infoga ytterligare en modul (FFO-HLS) till FFO-databasen var flerfaldigt. En anledning var att jag inte funnit någon studie om lärarinnelevernas rekrytering, sociala bakgrund, undervisningen och tillvaron vid seminariet eller om denna utbildnings betydelse för deras senare yrkesverksamhet, trots att seminariet spelat en stor roll och att arkivmaterialet är så omfattande.⁵³⁵ En

⁵²⁸ Serie D1. Inskrivningsmatriklar, Vol. 1: 1861–1941

⁵²⁹ Förteckning över samtliga 1980 inskrivna elever i Bilaga 11, se <CD:/bilagor/bilaga11.pdf>. Avskrifter och registrering av uppgifter i denna inskrivningsmatrikel har utförts av Agnes Langerth, även Alice Langerth har medverkat i detta arbete.

⁵³⁰ Förteckning över de digitaliserade volymerna i Kungl. Högre Lärarinneseminiariets arkiv, i bilaga 7, se <CD:/bilagor/bilaga7.pdf>.

⁵³¹ Se avsnitt 4.6.1 s. 128.

⁵³² Valter Fevrell var lektor vid seminariet. Valter Fevrell, *Kungl. Högre lärarinneseminiariet in memoriam: minnesrunor* (Stockholm: Geber, 1943).

⁵³³ De 47 f.d. lärarinneleverna i Fevrells bok har markerats i förteckningen över samtliga lärarinnelever i bilaga 11, se <CD:/bilagor/bilaga11.pdf>.

⁵³⁴ T.ex. Anna Maria Roos, Anna Sandström och Anna Whitlock.

⁵³⁵ En modernare studie om seminariet finns som fokuserar seminariernas tillkomst och utveckling snarare än sociologiska och pedagogiska frågor, se Silbano de Lopez, "Kungliga Högre lärarinneseminiariet och flickskolans framväxt" *Sekelskiftets utmaningar: essäer om välfärd, utbildning och nationell identitet vid sekelskiftet 1900*, red. Ann-Katrin Hatje, (Stockholm: Carlsson, 2002). Därtill har andra typer av utgåvor såsom seminariets historieskrivning och minnesrunor författade av: Fevrell, *Kungl. Högre lärarinneseminiariet in memoriam: minnesrunor*; Fähræus, *Högre lärarinneseminiariets historia*.

annan anledning var att pröva användningen av material där foton och text kombineras. Slutligen var det ett utmärkt tillfälle att skapa en modul med ett annat tillvägagångssätt än för det tidigare materialet genom att skapa en diplomatisk digital samling av primärkällor.⁵³⁶ Målsättningen har varit att utforma ett prosopografiskt material som kan användas för utbildningshistoriska studier av olika slag, exempelvis studier av utbildningstidens betydelse för kvinnornas sociala nätverk, föreningsliv och yrkesverksamhet.⁵³⁷ I och med att arkivmaterialet är så omfattande finns flera möjligheter till olika typer av utbildningshistoriska studier, till exempel studier av lärarinneutbildningens innehåll och uppläggning vid den här tiden.

5.7 Kapitalbeskrivningar som verktyg för insamling av material

I detta kapitel diskuteras det verktyg som använts för insamling av material till delmodulen FFO-KAPITAL, som är den modul som innehåller de mest detaljerade uppgifterna. Informationen har samlats in och skrivits ned i levnadsteckningar eller ”minibiografier” (här benämnt *kapitalbeskrivningar*). Benämningen har valts i enlighet med projektets metodologiska och teoretiska utgångspunkter:

Kapitalbeskrivningarnas utformning är ett uttryck för projektets teoretiska och metodologiska hemhörighet i en Bourdieutradition; begreppen kapital och fält liksom positioner och dispositioner brukas som forskningsredskap och kapitalbeskrivningarna är utarbetade enligt prosopografisk metod.⁵³⁸

Kapitalbeskrivningarna är producerade av medarbetarna i projektet och innehåller detaljerade uppgifter om kvinnornas tillgångar, såväl nedärvda som förvärvade.⁵³⁹ Informationen i det här fallet representeras av indikatorer för olika sorters tillgångar och egenskaper. Det handlar med andra ord om vissa specifika uppgifter, som samlats in och nedtecknats i kapitalbeskrivningarna för att:

⁵³⁶ FFO-KAPITAL är också till viss del en diplomatisk utgåva i den bemärkelsen att jag inte redigerat författarnas texter i någon större utsträckning. Jfr. s. 180.

⁵³⁷ Denna modul skall kompletteras med ytterligare uppgifter om kvinnornas föreningsverksamheter under skoltiden.

⁵³⁸ Linné. ”Lärarinnor, pedagogiskt arbete och offentlighet. Kvinnliga pionjärs liv och strategier kring år 1900 - ett jämförande perspektiv”.

⁵³⁹ Kapitalbeskrivningarna är författade av: Donald Broady, Boel Englund, Jonas Gustafsson, Lena Hammarlund, Ingrid Heyman, Lena Kåreland, Agneta Linné, Eva Trotzig, Kerstin Skog-Östlin.

[i] kondenserad form teckna varje kvinnas resurser i termer av sociala, kulturella, pedagogiska, organisatoriska, politiska, filantropiska, religiösa och ekonomiska tillgångar.⁵⁴⁰

Dessa indikatorer används sedan som mått på olika former av kapital och för olika slags deskriptiva statistiska analyser. Följande typer av indikatorer för kvinnornas tillgångar och egenskaper har varit vägledande vid författandet av kapitalbeskrivningarna (exempel på specifik typ av indikator inom parentes):

- socialt ursprung (far/mor, yrke, utbildning, positioner),
- utbildning (examina, avslutad/oavslutad, utlandsvistelser),
- verksamheter (avlönad/oavlönad),
- socialt kapital (civilstånd, umgänge),
- ekonomiskt kapital (eget, rika släktingar, bouppteckningar),
- specifikt symboliskt kapital (intellektuellt kapital, organisationskapital, politiskt kapital, kapital knutet till filantropi, utbildningspolitiskt och pedagogiskt kapital),
- publikationer (facklitteratur, prosa, artiklar),
- utmärkelser (Illis Quorum⁵⁴¹),
- eftermälen.

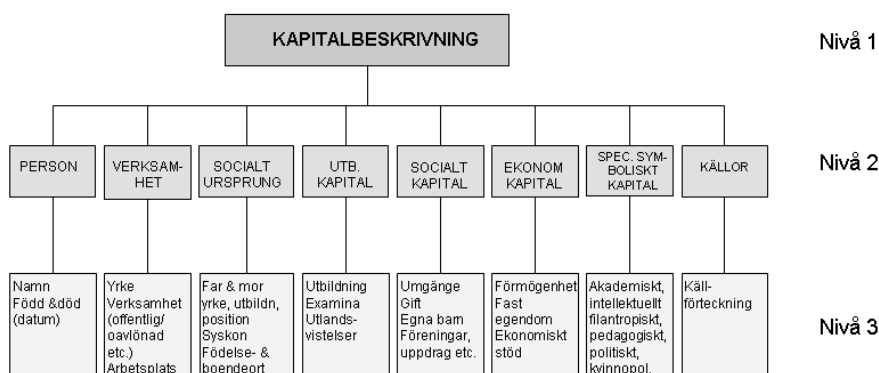
Av dessa indikatorer har sedan, bland annat med hjälp av märkningen, en mängd variabler konstruerats för användning i olika slags analyser och tillämpningar.⁵⁴² En projektspecifik mall användes för att vägleda författarna vid insamling och skrivande i syfte att harmonisera indikatorerna och säkerställa att vissa obligatoriska uppgifter registrerades. Mallen var därför uppbyggd kring olika rubriker för att identifiera och gruppera de egenskaper och tillgångar som var relevanta. Huvudrubrikerna i mallen är hierarkiskt arrangerade med olika logiska nivåer förutom att den exemplifierar i vilken ordning texten bör skrivas. I figuren nedan visas den hierarkiska struktur och logiska ordning som mallen föreskriver.⁵⁴³

⁵⁴⁰ Linné. "Lärarinnor, pedagogiskt arbete och offentlighet. Kvinnliga pionjärs liv och strategier kring år 1900 - ett jämförande perspektiv".

⁵⁴¹ Medalj som kan utdelas av den svenska regeringen för betydande insatser inom kulturellt, socialt eller konstnärligt arbete.

⁵⁴² För ytterligare diskussion om metod och utgångspunkter för material och kapitalbeskrivningarna, se t. ex. Ingrid Heyman, Eva Trotzig & Donald Broady, *Formering för offentlighet: rapport från konferens i Uppsala den 22 november 2000*, Meddelanden från Forum för pedagogisk historia, Nr 4 (Uppsala: Forum för pedagogisk historia SEC ILU Univ., 2001); Linné. "Lärarinnor, pedagogiskt arbete och offentlighet. Kvinnliga pionjärs liv och strategier kring år 1900 - ett jämförande perspektiv".

⁵⁴³ Se mall för kapitalbeskrivning i bilaga 1 vid <CD:/bilagor/bilaga1.pdf>.



Figur 17. Översikt av kapitalbeskrivningsmallens logiska ordning och hierarkiska struktur

Kapitalbeskrivning påbörjas alltid med information om kvinnans namn, vem som författat den samt grundläggande persondata om tidpunkt för födelse och död. Varje huvudrubrik (nivå 2) innehåller ett antal underrubriker (nivå 3). Det är meningen att data för alla underrubriker skall nedtecknas under tillhörande huvudrubrik. Detta är inte alltid möjligt då vissa sorters information logiskt kan tillhöra flera huvudrubriker. Till exempel kan mer allmän information som egentligen bör sorteras under rubriken socialt kapital även sägas tillhöra en annan huvudrubrik.

I vissa kapitalbeskrivningar är dock texten inte alls strukturerad utifrån mallens principer. Det är dock möjligt att applicera den modellen, som mallen är ett uttryck för, även om det inte alltid blir rätt eller bra. Ibland har uppgifter om läroanstalter, som hör hemma i nivå 3 under Utbildningskapital (Nivå 2), förtecknats i stycket om socialt kapital eller i ett eget stycke motsvarande nivå 2. Detta har jag löst genom typning av alla element som inte direkt motsvarar mallens logiska ordning, men som gör det möjligt att identifiera rätt innehållsobjekt.⁵⁴⁴

Mallen föreskriver inte något om det stilistiska och texterna varierar både vad gäller stil och omfång. Vissa texter består till stor del av uppräkningsdata, medan andra kapitalbeskrivningar mer har karaktären av självständiga artiklar.⁵⁴⁵ Mallen som teoretiskt verktyg får därmed en konkret och praktisk funktion i forskningsarbetet genom att den både hjälper och begränsar skrivandet. Genom att i förväg försöka modellera innehåll och ordning av det förväntade innehållet blir det också lättare att upptäcka skevheter och misstag i datainsamlingen men även felaktigheter i själva mallen. Om man dessutom kombinerar flera verktyg kan det medföra en ytterligare skärpning men även en breddning

⁵⁴⁴ Se exempel på detta i Figur 7 s. 79.

⁵⁴⁵ Samtliga kapitalbeskrivningar finns tillgängliga i olika format vid www.skeptron.uu.se/monicalz/ffo/kapital/.

genom att olika verktyg fokuserar på olika aspekter. Mallen hjälper ju till att belysa vissa aspekter av materialet medan andra får stå tillbaka.

5.7.1 Kapitalbeskrivningarnas källmaterial

Som underlag till författandet av kapitalbeskrivningarna har källor av olika karaktär använts, arkivmaterial såväl som sekundärkällor.⁵⁴⁶ Till att börja med finns en mängd databaser och översiktsverk som kan vara lämpliga att börja med.⁵⁴⁷ I Riksarkivets nationella arkivdatabas (NAD) på webben kan man söka efter vad som finns var i olika arkiv.⁵⁴⁸ De tryckta källorna är framför allt lexikala biografiska verk såsom *Svenskt biografiskt lexikon* (SBL) där några av kvinnorna i FFO-databasen är biograferade.⁵⁴⁹ Artiklarna i SBL författas av specialister och ackompanjeras av fullständiga bibliografier. Därför är de också en bra utgångspunkt för fortsatt forskning.⁵⁵⁰ Numer skrivs artiklar endast om avlidna personer som gjort insatser inom samhälls- och kulturliv. Första bandet utgavs år 1918 och hittills har 31 band givits ut och nu har man kommit fram till bokstaven S.⁵⁵¹ För de kvinnor, i det här materialet, vars efternamn börjar på bokstäver efter S saknas det alltså biografiska artiklar i SBL.

Andra lexikala verk som använts för kapitalbeskrivningarna är *Svenska män och kvinnor* som gavs ut åren 1942-45 och som innehåller kortfattade artiklar.⁵⁵² Även *Svenska kvinnor i offentlig verksamhet* samt *Svenska kvinnor från skilda verksamhetsområden* har varit till nytta för insamlingen av uppgifter om kvinnorna – förutom att dessa biografiska uppslagsböcker använts som källor till de större registren i projektet.⁵⁵³ Andra exempel på tryckta källor är *Vem är det?*⁵⁵⁴

⁵⁴⁶ En förteckning över de källor som använts som underlag till kapitalbeskrivningarna redovisas i bilaga 14 vid CD:/bilagor/bilaga14.pdf. Dessa källor har inte inkluderats i den här bokens källförteckning då de förhåller sig sekundärt till kapitalbeskrivningarna som ju är de källor jag använder mig av.

⁵⁴⁷ En grundlig genomgång av källor lämpliga för biografiskrivande ger Ingrid Heyman, "Vilket underlag kan bidra till en trovärdig biografi? Källor, arkiv och övrigt material" *Biografi och kollektivbiografi som historisk metod*, red. Agneta Linné & Kerstin Skog-Östlin, Meddelanden från Forum för pedagogisk historia, 5 (Uppsala: Forum för pedagogisk historia SEC, Uppsala Universitet, 2001) s. 4-9.

⁵⁴⁸ Sökning kan ske genom olika sökingångar: person, ort, arkiv och arkivinstitutioner, se <www.nad.ra.se/>.

⁵⁴⁹ T.ex. Sophie Adlersparre, Elin Cederblom, Astrid Cleve von Euler, Klara Johansson och Ellen Key.

⁵⁵⁰ Förutom det markeras SBL:s position av att redaktionen är en statlig myndighet.

⁵⁵¹ Artiklarna finns inte publicerade på webben, där finns dock ett index, utan man måste beställa en CD eller läsa artiklar på plats i biblioteket. Se <www.sblexikon.se/>.

⁵⁵² Torsten Dahl & Nils Bohman, *Svenska män och kvinnor: biografisk uppslagsbok* (Stockholm: Bonnier, 1942).

⁵⁵³ Hedberg & Arosenius, *Svenska kvinnor från skilda verksamhetsområden: biografisk uppslagsbok*; Lundström, *Svenska kvinnor i offentlig verksamhet: porträtt och biografier*.

⁵⁵⁴ Svensk biografisk handbok, *Vem är det?*, utgavs av Norstedts förlag under perioden 1912 till 2001. Uppgifterna i baseras på material som personerna själva skickade in till redaktionen. Fr.o.m. 2007 tog Nationalencyklopedin AB över utgivningen.

Eftermälen har också använts i ganska stor utsträckning, flera av kvinnorna blev flitigt omskrivna i media efter sin död. Exempel på annan lexikal litteratur är förstås de vanliga uppslagsverken som till exempel *Nordisk familjebok* och *Svensk Uppslagsbok*.⁵⁵⁵ *Adresskalendrar för Stockholm* har vissa författare använt sig flitigt av, tillika olika slag av jubileums- och minnesskrifter.

De otryckta källorna utgörs av offentliga arkiv såsom mantalslängder och bouppteckningar. Bouppteckningarna har varit en värdefull källa i flera av kapitalbeskrivningarna. Tyvärr saknas det bouppteckningsuppgifter för många de aktuella kvinnorna vilket försvårar användningen av dessa uppgifter för kollektiva analyser. Kollektiva analyser förutsätter att uppgifterna finns för flertalet av individerna i urvalet. Detta är särskilt viktigt eftersom materialet inte tillåter några vidlyftiga tolkningar utanför själva kohorten. Därför är det viktigt att det interna bortfallet är så litet som möjligt, för att man skall kunna uttala sig om strukturer eller förhållanden giltiga för kohorten *per se*. Däremot är det förstås fullt möjligt att utföra analyser på en delmängd av FFO-KAPITAL – säg för de kvinnor som det finns bouppteckningsuppgifter för.

Även källor av mer privat karaktär såsom privata arkiv och brevsamlingar har använts. Flera av brevsamlingarna finns tillgängliga på de större forskningsbiblioteken samt vid Kungl. Biblioteket.⁵⁵⁶ Till viss del har även muntliga källor, som till exempel släktingar, använts för insamling av data.

Medarbetarna har skickat kapitalbeskrivningarna i Wordformat till mig som sammanställt dem och upprättat ett register med information om författare, kvinnans namn, datum och version. Alla kapitalbeskrivningar har därefter migrerats från wordformatet till vanligt textformat och sedan till XML-formatet. Varje kapitalbeskrivning har först genomgått en rudimentär TEI-märkning innan den lagts till i masterfilen, det vill säga den fil som innehåller all text som skrivits om samtliga 101 kvinnor i FFO-KAPITAL.⁵⁵⁷

Designen av detta prosopografiska material har framför allt påverkats av kapitalbeskrivningsmallen men även av de verktyg som valdes för märkningsarbetet. Förutom det är de kvinnor som ingår i materialet förstås en förutsättning för prosopografien. Eller snarare kan man betrakta de som konstruerade individer, som konstrueras genom de egenskaper och kategorier som projektet valt för att karaktärisera dem.⁵⁵⁸ Bourdieus verktygslåda har även kombinerats med studier av kvinnornas nätverksformationer och där har uppgifterna om kvinnornas ytliga kontakter, vänner och ovänner kommit till användning.

⁵⁵⁵ *Nordisk familjebok* utgavs i 38 band 1904-26, i 23 band 1923-27 och 22 band 1951-55. *Svensk uppslagsbok*, utkom med 30 band 1929-37 och 32 band 1947-55.

⁵⁵⁶ De kvinnohistoriska samlingarna vid Göteborgs universitet har god tillgång till källmaterial som t.ex. brev, bilder, dagböcker, manuskript etc. med tyngdpunkt på 1900-talets första hälft. Förutom dessa enskilda arkiv finns även arkiv från flera svenska kvinnoföreningar. <www.ub.gu.se/kvinn/handskrifter/arkiv.xml>.

⁵⁵⁷ Märkningen av kapitalbeskrivningar har skett i olika steg något som beskrivs närmare nedan s. 172 f.

⁵⁵⁸ Epistemiska individer, se Bourdieu, *Homo academicus* ssk. s. 53-56.

Kapitalbeskrivningarnas konstruktion har även påverkat det sätt som TEI-schemat använts. Kapitalbeskrivningarna är författade med utgångspunkt i projektets syften, metodval och forskningsfrågor. Därmed har de också varit vägledande för modelleringen av det digitala materialet tillsammans med verktyget TEI.⁵⁵⁹

Sist men inte minst påverkar förstås även källorna själva och författarna innehållet i kapitalbeskrivningarna. För författarna handlar det om att söka efter ledtrådar och lämningar efter kvinnorna som sedan skall sättas in i ett sammanhang.⁵⁶⁰ För mig handlar det om att sammanställa, beskriva och ordna det material som författats. Däremot skulle det vara mycket svårt för mig som teknisk redaktör, inte minst tidsmässigt, att avgöra om tolkningar och slutsatser i en kapitalbeskrivning är för vidlyftiga. För detta arbete har mallen varit betydelsefull. Där har jag haft det verktyg som dirigerar vad som skall beskrivas och hur det skall göras. Jag behöver inte lägga mig i eventuella tolkningsproblem, så länge det inte påverkar de uppgifter som märks upp. Genom att använda indikatorer som så långt det är möjligt håller sig på en ”objektiv” eller snarare materialistisk nivå, ges möjligheter att senare kontrollera mot källorna. Detta har också gjorts för vissa indikatorer exempelvis för några av de vanligaste föreningarna.⁵⁶¹

5.8 Biografiska uppslagsböcker som källmaterial i prosopografin

Uppgifterna i FFO-LUNDSTROM bygger på avskrifter ur samtida biografiska uppslagsböcker.⁵⁶² Syftet med att registrera uppgifter för en större mängd kvinnor än vad som skulle kunna ingå i kapitalbeskrivningarna var att konstruera ett större jämförelsematerial. Detta material ansågs vara användbart som fond för jämförelser, deskriptiva statistiska analyser och korrespondensanalyser.

Det första materialet som skrevs av och registrerades var data hämtat ur *Svenska kvinnor från skilda verksamhetsområden*.⁵⁶³ Materialet till den boken sammanställdes inför baltiska utställningen i Malmö, där kvinnorna hade en egen paviljong.⁵⁶⁴ Uppgifterna i boken om de 843 kvinnorna skrevs av med utgångs-

⁵⁵⁹ Principerna och tillvägagångssättet för denna märkning beskrivs på s. 172 ff.

⁵⁶⁰ Ingrid Heyman menar att verktyget, dvs. kapitalbeskrivningsmallen, fungerar som ett ”enande kitt bland torra fakta” och därmed underlättar konstruktionen av kvinnornas sammanhang. Heyman, ”Vilket underlag kan bidra till en trovärdig biografi? Källor, arkiv och övrigt material” s. 9.

⁵⁶¹ Kontroll och revidering av kvinnornas föreningsaktivitet har skett för de föreningar som flest kvinnor var aktiva inom såsom Fredrika Bremerförbundet, Sällskapet Nya Idun samt Föreningen för Gift Kvinnas Eganderätt genom att granska medlemsmatriklar för de aktuella åren.

⁵⁶² Hedberg & Arosenius, *Svenska kvinnor från skilda verksamhetsområden: biografisk uppslagsbok*; Lundström, *Svenska kvinnor i offentlig verksamhet: porträtt och biografier*.

⁵⁶³ Hedberg & Arosenius, *Svenska kvinnor från skilda verksamhetsområden: biografisk uppslagsbok*.

⁵⁶⁴ Herman Fredrik Ahlström, *Officiell berättelse öfver Baltiska utställningen i Malmö 1914* (Malmö: Baltiska utställningen, 1915).

punkt i en mall för vilka indikatorer som skulle registreras (se Tabell 5 nedan). Det framgår inte i förord eller någon annanstans i boken hur många kvinnor som fick förfrågan att delta i boken.

Att föreliggande biografiska uppslagsbok över nu levande svenska kvinnor förtär åtskilliga luckor, äro undertecknade de första att erkänna. Men dels har den knappt tillmätta tiden icke medgivit efterforskningar och förfrågningar i den utsträckning som önskligt hade varit, dels ha flera, vilkas namn man skulle väntat finna här, särskilt anhängit att icke bli medtagna, och denna deras önskan har naturligtvis måst respekteras. Slutligen har också åtskilliga, trots upprepade påminnelser, icke insänt de begärda uppgifterna, som icke heller på annat sätt kunnat anskaffas. För allt detta ävensom för övriga möjligen förekommande misstag och brister anhålles om benäget överseende.⁵⁶⁵

Med utgångspunkt i förordet i Hedberg och Arosenius bok kan man dra slutsatsen att bortfallet blev relativt stort. Både vad gäller de önskade uppgifterna och både vad gäller antalet kvinnor som skickade in dem. Uppgifterna i boken visade sig också vara av skiftande kvalitet. Bland annat saknades det i många fall uppgifter om kvinnans utbildning. Ytterligare en anledning var att denna bok publicerades 1914 och räckte därför inte in hela perioden 1880-1920.

På grund av dessa problem men även för att fler kvinnor fanns förtecknade här beslutades det att ytterligare en biografisk uppslagsbok skulle skrivas av, *Svenska kvinnor i offentlig verksamhet*.⁵⁶⁶ Inbjudan till medverkan i boken skickades ut till cirka 3000 kvinnor under år 1922. Lite drygt 1200 offentligt verksamma kvinnor lämnade in egna biografier. Det handlar alltså om ett verk där de biograferade fortfarande var i livet vid utgivningen. Det står ingenting i för- eller efterord om varför en del kvinnor saknas medan andra finns med.

Uppgifterna som används i FFO-LUNDSTROM är avskrifter av dessa korta biografiska texter. För att få en stringent registrering av så många uppgifter som möjligt användes samma mall för indikatorer som för avskrifterna ur *Svenska kvinnor från skilda verksamhetsområden*. De uppgifter som i detta fall var av intresse är följaktligen av samma typ av indikatorer på kvinnornas tillgångar och egenskaper som i kapitalbeskrivningarna då dessa avskrifter skulle användas som jämförelsematerial.

Intressant är att båda dessa böcker är ett slags styrd form av självbiografi då det som finns med i biografierna är vad kvinnorna själva valt att lämna ut, vilket bör beaktas vid analys av materialet. Noteras bör även att flera av kvinnorna som finns förtecknade i Hedberg & Arosenius även finns med i Lundströms bok. I och med att samma typ av indikatorer samlats in för båda verken är det

⁵⁶⁵ Hedberg & Arosenius, *Svenska kvinnor från skilda verksamhetsområden: biografisk uppslagsbok*, förord på onummerad sida.

⁵⁶⁶ Lundström, *Svenska kvinnor i offentlig verksamhet: porträtt och biografier*. Det är framför allt uppgifter hämtade ur Lundströms biografiska uppslagsbok som används i FFO-databasen. Uppgifter om 1206 kvinnor som är biograferade i boken har skrivits av och registrerats av Rakel Höglund. Förklaring till avskrifterna och kodnyckel, se <www.skeptron.uu.se/monicalz/ffo/lundstr-anv.pdf>.

möjligt att integrera kvinnor från Hedberg & Arosenius till FFO-LUNDSTROM, eller att göra jämförande analyser om uppgifternas innehåll för de kvinnor som finns med i båda uppslagsböckerna för att undersöka hur informationen skiljer sig åt mellan de två verken.

Tabell 5. *Jämförelse av indikatorer och registrerade värden för kvinnornas egenskaper och tillgångar från två biografiska uppslagsböcker.*⁵⁶⁷

Indikator	Lundström (N=1205)		Hedberg & Arosenius (N=842)	
	Registrerade värden	Andel bortfall (procent)	Registrerade värden	Andel bortfall (procent)
Nr	unik id	0,0	unik id	0,0
Födelseår	årtal	0,2	årtal	0,8
Namn	angivet namn (efternamn, förnamn)	0,0	angivet namn (efternamn, förnamn)	0,0
Titel	angiven titel/titlar	0,1	angiven titel	0,1
Född i ⁵⁶⁸	S/L/U	3,0	S/L/U	0,4
Bodde i	S/L/U	0,6	S/L/U	0,4
Faderns yrke	angiven titel	4,8	angiven titel	8,1
Utbildning	angiven utbildning(ar)	13,5	S/läroanstalt	60,8
Utland/typ	N/angiven anledning(ar)	0,5	N/studier/arbete/bosatt	0,5
Anställd	J/N	1,2	J/N	0,6
Pionjär	J/N	0,2	J/N	0,2
Uppdrag	J/N	0,5	J/N samt ev. föreningsnamn	0,0
Sfär	tolkat (t.ex. Kultur:litteratur)	0,9	tolkat (t.ex. Kultur:litteratur)	1,3
Publicerad	J/N	0,7	J/N ev. även kod för typ av publikation	1,2
Utmärkelse	J/N	2,6	J/N	0,2
Gift	J/N (även J2 som avser omgift)	0,0	J/N	0,4
Makes yrke ⁵⁶⁹	angiven titel	52,6	angiven titel	60,0
Anm.	Ytterligare upplysningar, t.ex. om släktskap, betydande insatser, utmärkelser, föreningar etc.	20,0	Ytterligare upplysningar, t.ex. om släktskap, eller om betydande insatser	63,5

⁵⁶⁷ Kolumnen "Andel bortfall" i tabellen avser internt bortfall per typ av indikator och anges i procent för respektive register. Detta bortfall kan bero på tomma poster, att "uppgift saknas" har angetts eller felskrivningar. Bortfall i kategorin "Makes yrke" avser alla ogifta kvinnor.

⁵⁶⁸ Avser om kvinnan är född i Stockholm, på landsorten eller i utlandet.

⁵⁶⁹ Bortfallet för makes titel överensstämmer närapå exakt med andelen ogifta kvinnor. I Lundströms bok är andelen ogifta 52 procent och i Hedberg & Arosenius är andelen ogifta kvinnor 60 procent.

I tabellen ovan presenteras de indikatorer som förtecknats med utgångspunkt i de biografiska uppslagsböckerna. I tabellen ingår även en jämförelse mellan omfattningen av det interna bortfallet för respektive indikator.⁵⁷⁰

I tabellen kan vi se att den information som förtecknats för varje typ av indikator till stora delar samstämmer mellan de två registren. Av de kvinnor som finns med i böckerna och alltså skickade in den begärda informationen har de flesta angivit det redaktörerna vill ha. Ett stort bortfall, om bortser från bortfallet i ytterligare upplysningar, är det för uppgifterna om kvinnornas utbildning samt för "ytterligare upplysningar", i Hedberg & Arosenius. Bortfallet på 60 procent beror på grund av att uppgifter om utbildning ofta saknas i den boken. Vidare saknas det uppgifter om faderns yrke för var tolfte kvinna i Hedberg & Arosenius och för var tjugonde kvinna i Lundströms bok. Faderns yrke är en särskilt viktig indikator då den används som grund till för variabler om social klassificering i analyserna. Bortfallet för makens yrke kan förklaras med att det överensstämmer med antalet ogifta kvinnor i varje bok.

Ytterligare information som kan användas för andra indikatorer återfinns under "anmärkningar". Även där har Hedberg & Arosenius störst bortfall framför allt på grund av den väldigt kortfattade information i boken. I Lundströms bok är de biografiska notiserna något fylligare – dessutom ackompanjerade av foton som kvinnorna skickat in. Detta sammantaget med skälen som ovan angivits om fördelar med Lundströms bok gjorde att jag valde den som jämförelsematerial till kapitalbeskrivningarna.

Vissa uppgifter har kompletterats i efterhand, till exempel information om publicering för kvinnor boende i Stockholm.⁵⁷¹ Alla typer av indikatorer i tabellen ovan finns förtecknade i varje kapitalbeskrivning, under förutsättning att uppgifter fanns att samla in. Därtill registrerades en stor mängd annan information som beskrivits i kapitel 5.7., s. 157.

FFO-LUNDSTROM har i efterhand utökats med ytterligare information baserat på aggregerat av den avskrivna information samt kombinationer med information som hämtats från anmärkningar eller andra källor. Exempel på det är att jag skapat aggregerade variabler för sociala klasser, inflyttning från landsort till Stockholm och publiceringsmönster. Däremot har det inte varit möjligt att skapa regionala kategorier med utgångspunkt i födelse- och bostadsort eftersom registreringen av orter skett på en redan aggregerad nivå med endast tre möjliga värden (Stockholm, landsort och utland).

Uppgifterna som baseras på avskrifter ur Lundströms biografiska uppslagsbok är inte lika användbara på grund av att de redan var kategoriserade vid

⁵⁷⁰ Originalmaterialet för avskrifterna från Lundström finns tillgängligt vid <www.skeptron.uu.se/monicalz/ffo/totreg-matrix.htm> och originalmaterialet för avskrifter ur Hedberg & Arosenius finns tillgängligt vid: <www.skeptron.uu.se/monicalz/ffo/ha-orig.htm>.

⁵⁷¹ För många av de beskrivande analyserna har använts en delmängd av FFO-LUNDSTROM som enbart består av de 600 kvinnor som bodde i Stockholm, dvs. fler kvinnor än de 101 som utgör FFO-KAPITAL.

registreringen.⁵⁷² Den exakta födelseorten angavs inte utan kategoriserades redan från början som en variabel med tre värden (Stockholm, Landsort, Utland). På liknande sätt hade uppgifterna publicering, utlandsvistelser, anställning, uppdrag och utmärkelser registrerats. Endast binära värden (J/N) finns registrerade – istället för uppgifter som faktiskt finns om antal, titel och genrer av utgivningen eller var kvinnan var anställd vid tiden för uppgifternas inlämnande. Uppgifterna om kvinnans och faderns utbildning och yrkestitel/verksamhet har däremot registrerats såsom de anges i utgåvan.

5.8.1 Källvärde

Möjligheten att använda uppgifter från en sådan speciell genre som självbiografen för beskrivande statistiska analyser av en grupp individer kan förstås diskuteras. Även om, eller kanske snarare på grund av att, vi använder de korta biografiska texterna till att förteckna faktiska företeelser och egennamn är det inte oviktigt att vara medveten om självbiografins begränsningar. Jag menar dock att den självbiografiska informationen har ett värde på grund av dess subjektivitet. Vad väljer individen att berätta, och på vilka grunder? Ett exempel är att många av kvinnorna titulerar sig som fröken eller utifrån börd och giftermål i Lundströms bok trots att de är yrkesverksamma.⁵⁷³ Den rådande könsordningen och samhällseliga normer påverkade säkert detta faktum.⁵⁷⁴ En annan orsak kan vara att titeln "fröken" vid sekelskiftet användes som social markör. Tidigare var "fröken" avsedd för ogifta kvinnor ur medel- och överklass och ännu längre tillbaka för grevedöttrar och adelns döttrar.⁵⁷⁵

Hur subjektiva de biografiska noteringarna är kan dock i detta fall diskuteras eftersom kvinnorna bidrog med texter som de på förhand hade fått anvisningar om vilken typ av information som önskades. Man kan snabbt konstatera att i uppslagsböckerna är biografierna snarare uppräknningar än retrospektiva levnadsbeskrivningar. Inte desto mindre kan man ju tänka att de uppgifter som lämnades ut i vissa fall mer speglar det individen önskade vara än hur det verkligen var. För detta vidkommande har vissa uppgifter ur böckerna till viss del kunnat kompletteras med uppgifter från andra källor vilket gör att eventuella skevheter och ofullständiga kan korrigeras.⁵⁷⁶ Genom att kombinera olika källmaterial kan den praktik som studeras skisseras – eller åtminstone konturerna

⁵⁷² Dessa avskrifter producerades för projektet Formering för offentlighet strax innan jag blev involverad som teknisk redaktör. Dock är det inte säkert att jag vid den tidpunkten hade uppmärksammat nackdelarna med att kategorisera informationen vid själva transkriberingen.

⁵⁷³ Se avsnitt 7.1.1 s. 203.

⁵⁷⁴ Jfr., Ambjörnsson & Sörlin, *Obemärkta: det dagliga livets idéer* s. 10.

⁵⁷⁵ Svenska Akademiens ordbok, < www.saol.se/ >.

⁵⁷⁶ Genom att "samla bevisbitar från ett brett spektrum av källor relevanta för det aktuella problemet" är det förstås lättare att avslöja felaktigheter i en självbiografi, eller en annan källa. Citatet (något omarbetat) ur: Tosh, *Historisk teori och metod* s. 75.

av den.⁵⁷⁷ Det som talar för användningen av självbiografiska texter är att många av kvinnorna som annars kanske skulle ha passerat mer eller mindre obemärkta genom historien finns noterade i dessa böcker – åtminstone några aspekter av deras liv.

5.9 Arkivmaterial som källor i prosopografin

Utöver kapitalbeskrivningar och självbiografiska notiser har inskrivningsmatriklar och annat arkivmaterial fotograferats och transkriberats. Innehållet i FFO-HLS bygger som tidigare nämnts på foton och avskrifter av arkivmaterial. Hitills har en bråkdel av det totala arkivmaterialet i Kungl. Högre Lärarinnesemariets arkiv digitaliserats.⁵⁷⁸

Utgångspunkterna här är att företeckna samtliga de uppgifter som källmaterialet erbjuder. Metodansatsen kan därmed betecknas som motsatsen till kapitalbeskrivningarna, då det är källorna som avgör vad som registreras och skrivs av till skillnad från kapitalbeskrivningarna, där insamlingen styrs av på förhand bestämda frågor. Påståendet är dock en sanning med modifikation. Arkivmaterialet är intressant för att det erbjuder den typ av data som en prosopografi brukar innehålla – och mer därtill. Hade målsättningen varit att studera utbildningens innehållsliga dimension hade fotograferingen också startat med andra volymer än inskrivningsmatriklarna.

Avsikten har varit att förteckna och sammanställa källmaterialet med så lite påverkan på innehållet som möjligt. Den digitala transkriberade versionen har utformats med en diplomatarisk ansats då det som står i arkivet ordagrant skrivits av, även förkortningar och felskrivningar.⁵⁷⁹ Detta får konsekvenser utöver märkningen av innehåll då jag till skillnad från kapitalbeskrivningar även beskriver innehålllets struktur i dessa matriklar.⁵⁸⁰

5.9.1 Bildfångst och textfångst

Av praktiska skäl snarare än några bevarandeskäl beslöt jag mig att fotografera arkivmaterialet för att reproducera *start*dokumentet i ett annat *måldokument*. En sideeffekt bli att måldokumentet även representerar en tvådimensionell grafisk yta för ett startdokument. I detta fall var det representationer av inskrivningsliggare och annat otryckt arkivmaterial. Valet stod mellan två strategier:

⁵⁷⁷ Esbjörn Larsson skriver även att självbiografins värde snarare ligger i att studera bilden av en företeelse snarare än företeelsen *per se*. Larsson, *Från adlig uppfostran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866* s. 63-66.

⁵⁷⁸ För en förteckning över vilka volymer och poster som digitaliserats, se bilaga 7.

⁵⁷⁹ Jfr. s. 180.

⁵⁸⁰ Hur märkningen gått till illustreras i avsnitt 6.2 s. 184 f.

Vetenskaplig utgivning arbetar med två strategier för transponering: en textorienterad och en bildorienterad. Den förra yttrar sig i *transkription*, den andra i *faksimilering*.⁵⁸¹

Transkription var den metod som användes för de biografiska posterna i FFO-databasen. Vanliga metoder för bildfångst är digitalfotografering och skanning.⁵⁸² Att valet av metod för bildfångst föll på digitalfotografering och inte skanning var av den enkla anledningen att jag hade tillgång till en digitalkamera men inte en handburen skanner. Att betala för att arkivpersonalen skulle skanna de fysiska förlagorna kom ju inte på fråga för ett avhandlingsprojekt som detta.

Praktiskt utfördes fotograferingen så att en sida i taget fotograferades, och inte hela uppslaget, beroende på liggarnas format.⁵⁸³ Man bör ha i åtanke att inte fotografera för stora områden. Även om det går att förstora bilden i det enklaste bildredigeringsprogram är det en tidsödande manöver att utföra för varenda post som skall granskas eller transkriberas. Exakt sju kvinnor har plats per sida i den liggare som innehåller inskrivningsuppgifter för samtliga verksamhetsår.⁵⁸⁴

De övriga inskrivningsböckerna som fotograferats innehåller detaljerade uppgifter för varje kvinna och där används en sida per kvinna för förteckning av uppgifter om hennes födelseår och ort, social bakgrund, tidigare utbildning och anställning, studiegång vid seminariet och tiden efter seminariet. Även här har en sida i taget fotograferats, dessutom en användbar indelning med tanke på tillämpningar.⁵⁸⁵ Dessa böcker innehåller även ett alfabetiskt index med kvinnornas namn och hänvisning till sidnummer i liggaren. Det finns tre liggare med dessa mer detaljerade uppgifter för åren 1914-1943.⁵⁸⁶ Övriga volymer som fotograferats är av skiftande karaktär, betygskatalogerna innehåller ”kunskapsbetyg” i olika ämnen samt ett tämligen omfattande bruk av marginalanteckningar.⁵⁸⁷

För bildformat rekommenderas oftast att i första hand spara digitala bilder i formatet TIFF och i andra hand i bildformatet JPEG.⁵⁸⁸ Anledningen till detta är att i TIFF-formatet lagras information för varje enskilt bildelement (pixel) om färg enligt ett schema. Detta filformat lagrar även metadata om bildens

⁵⁸¹ Kursivering i original. Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 202.

⁵⁸² Numera är även de lite enklare digitalkamerorna fullt användbara och har tillräckligt hög upplösning för fotofaksimilering. Kevin Kiernan, ”Digital facsimiles in editing” *Electronic textual editing*, red. Lou Burnard, et al., (New York, N.Y.: Modern Language Association of America, 2006).

⁵⁸³ Behjälplig vid fotograferingen var Agnes och Alice Langerth. Fotograferingen utfördes under juni månad 2006.

⁵⁸⁴ D1.1 (1861-1941), för exempel på liggarens första uppslag där de allra första eleverna registrerades 1861, se <CD:/bilder/hls-matrikel1.jpg>.

⁵⁸⁵ Se exempel i *Figur 38* s. 218.

⁵⁸⁶ Varför det saknas liggare för perioden innan 1914 vet jag inte, jag har sökt i arkivet i alla tänkbara volymer och en hel del protokoll kring den tiden utan att hitta de saknadet materialet.

⁵⁸⁷ För exempel se, <CD:/bilder/hls-matrikel1.jpg>.

⁵⁸⁸ Tagged Image File Format (TIFF) samt Joint Photographic Experts Group (JPEG).

storlek och färgschema.⁵⁸⁹ I JPEG-formatet används däremot en lagringsteknik som kan medföra att kvaliteten på bilden försämras vid komprimering och då går det inte att återskapa den ursprungliga bildkvaliteten.

I detta fall har JPEG-formatet används för att lagra bilderna och kvaliteten var fullt tillräckligt för det mesta av arkivmaterialet.⁵⁹⁰ JPEG-formatet höll dock inte tillräcklig bildkvalitet för betygskatalogerna, som var i folioformat och väldigt stora, trots att en hög bildkvalitet användes. Kvaliteten på bilderna medförde att för vissa av dem var det inte möjligt att se de allra minsta marginalanteckningarna vid förstoring, då de blev för oskarpa. Däremot gick det bra att läsa uppgifterna om elevernas namn, tabellrubrikerna och betygen vid förstoringen.⁵⁹¹

När bilderna är tagna infinner nästa arbetsmoment, nämligen att ordna filerna på ett koncist och självförklarande sätt för god hanterbarhet. En viss tveksamhet infann sig över vilken ordnande princip som är mest användbar. Valet stod mellan att ordna bilderna tematiskt eller hierarkiskt. Om en tematisk ordning skulle användas var det logiskt att nyttja den arkivförteckning där hierarkier och kronologier redan var upprättade. För en kronologisk ordning kunde materialet fördelas enligt perioder först och sedan tematiskt. Valet föll på att använda den arkivförteckning som fanns som ordnande princip men att dela in kronologin i kortare perioder än i arkivförteckningen. Detta för att minska antalet fotografier i varje katalog. Därtill är det fullt möjligt att lägga till ett kronologiskt index med pekare till perioder i filsystemet.

Det kanske kan tyckas trivialt att ägna så mycket utrymme till filernas organisering. Men eftersom syftet var att koppla ihop bilder med text är det viktigt att strukturen är genomtänkt från början så att sökning av och länkning till bilderna blir hanterbart vid utformningen och användningen av samlingen.

För att fånga in själva texten i ett bildbaserat startdokument måste någon sorts överföring till måldokumentet ske. Det gör man antingen genom att transkribera textsekvenserna i bilden eller genom att använda tekniska hjälpmedel som kan fånga text i bilder. Det finns möjligheter att med OCR-teknik läsa av en bild och sedan låta en programvara identifiera texten och migrera den till maskinläsbart format för vidare överföring till ytterligare program för formatering och märkning. Detta kräver till att börja med en noggrann korrekturläsning av resultatet. Därtill krävs det att texten kan fångas med hjälp av OCR-teknik. Textsekvenserna i detta arkivmaterial är handskrivna och av skiftande läsbarhet, det är många förkortningar och i vissa dokument har bläcket bleknat. OCR-tekniken är dessutom optimerad för tryckt text snarare än handskriven snirklig text.

⁵⁸⁹ Morrison, Popham & Wikander, *Creating and Documenting Electronic Texts: A Guide to Good Practice* s. 10-20.

⁵⁹⁰ JPEG anses vara det bästa alternativet för webbpublicering genom möjligheterna att komprimera filerna och därmed minska storlek och tid för nedladdning.

⁵⁹¹ För exempel på ett uppslag i betygskatalogen från 1866 som illustrerar marginalanteckningar och betygssättning, se <CD:/bilder/hls-betyg.jpg>.

If the characters in the text are not fully formed or there are instances of broken type and damaged plates, the software will have a difficult time reading the material.⁵⁹²

Med utgångspunkt i dessa råd och med tanke på att texten är handskrivna på-börjades transkriberingen av innehållet.

Uppgifterna i den första inskrivningsmatrikeln registrerades i ett kalkylblad. Innehållet var redan ordnat i tabellform i liggaren och här prioriterades först möjligheten att relativt snabbt mata in och skapa ett heltäckande register. Dessutom var det enkelt att migrera detta innehåll till XML-format efteråt eftersom innehållet var ordnat i en enkel tabellstruktur. I figuren nedan illustreras vilka uppgifter som den första inskrivningsmatrikeln innehåller.

Namn	Födelse		Utgått	Anmärkningar
	år och dag	ort	Betyg: 1-4 <i>Läsåret 1861-62</i>	
Ungqvist, Carolina.	1825 7e	Jönköping	1863 3e	Med diploma
Lusth, Emma.	1825 7e	Stockholm	1864 7e	Med Diploma
Bagge, Eriq.	1822 7e	Stockholm	1863 3e	Med diploma
Bair, Alice.	1841 3e	Norrmalm	1864 7e	Med Diploma

Figur 18. Utdrag ur inskrivningsmatrikeln för läsåret 1861-1862 (Serie D1, vol. 1).

Uppgifterna i matrikeln är: namn, födelseort och datum, examensdatum, hänvisning till betygskatalog och anmärkningar.

I likhet med FFO-KAPITAL är det individen som är den ordnande principen i grundregistret. Till detta grundregister kan sedan ytterligare information infogas allt eftersom arkivmaterialet digitaliseras. Redan nu har vi kopplat ytterligare poster med mer detaljerad information till detta register och dessa inskrivningsuppgifter återfinns som tidigare nämnts i tre andra liggare. Ett urval av dessa har hittills transkriberats; de har skrivits av enligt en diplomatarisk princip och samtidigt märkts upp med TEI:s nya modul för prosopografiska material.⁵⁹³ Transkriberingen kräver tillgång till två datorskrämar, en som visar

⁵⁹² Morrison, Popham & Wikander, *Creating and Documenting Electronic Texts: A Guide to Good Practice* s. 19.

⁵⁹³ Förutom de bilder och filer som bifogas i CD:n är exempel på de digitaliserade posterna samt grundregistret tillgängligt vid <www.skeptron.uu.se/monicalz/ffo/>.

bilden och en som visar den avskrivna texten (i en texteditor). Den beskrivande märkningen av arkivposterna presenteras i avsnitt 6.2, s. 184.

Intentionen med digitaliseringen av arkivmaterialet var att skapa en reproduktion av originaldokumenten med så lite påverkan på innehållet som möjligt och objektivitet relativt materialet. Pragmatiska överväganden har dock gjort att dessa principer emellanåt behövde åsidosättas. Intentionen att vara så objektiv som möjligt vid produktionen kan behöva nyanseras. Även i den här typen av "otolkade" material där källorna görs tillgängliga och där användarna kommer närmare originaldokumenten innehåller flera moment där tolkningen påverkat utformningen. Inte minst vid transkriberingen av texterna men även vid produktionen av de digitala bilderna ingår det manipulerande moment.⁵⁹⁴ Genom det sätt som materialet organiseras, beskrivs och presenteras medför att själva innehållet blir mer eller mindre manipulerbart, något som exemplifieras bland annat i avsnitt 7.4, s. 216. Presentationerna, inte minst, både möjliggör och begränsar de frågor andra kan ställa till den digitaliserade samlingen.⁵⁹⁵

5.10 Avslutning

Detta kapitel har ägnats åt prosopografiska metoder och tillämpningar. Min avsikt har varit att illustrera prosopografiska tillämpningar som kan tillföra kunskap om hur utformning av digitala prosopografiska samlingar kan göras. Jag har även presenterat och problematiserat de källor som används som underlag för de uppgifter som finns registrerade i FFO-databasen. Även verktygen *per se* har diskuterats i relation till modellering och utformning av prosopografiskt material.

Därtill har jag försökt belysa praktiska och konkreta moment i arbetsprocessen. Detta är ingen lätt uppgift då risken är att beskrivningen av praktiken, som balanserar mellan det triviala och det avancerade, blir platt och intetsägande. Genom att förse läsaren med exempel på de olika momenten jämte bilder, filer och grafer i texten såväl som bilagor på CD:n ges läsaren en inblick i de möjligheter som innehållsdesign kan medföra för såväl forskning som undervisning.

⁵⁹⁴ Digital bildfångst är, menar Dahlström, det område där skillnaden mellan subjektivitet och objektivitet är som störst för att producenten "antas träda i bakgrunden" samtidigt som digitala bilder erbjuder många tillfällen till manipulation. Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 246.

⁵⁹⁵ Exempelvis på det sätt som bild kopplats ihop med innehåll, se tillämpningsexempel på s. 218.

6 Innehållsdesign av prosopografiskt material

Detta kapitel ägnas åt beskrivning av och diskussion om de verktyg som används för att beskriva materialet i prosopografin. Inledningsvis presenteras märkningen av kapitalbeskrivningarna, därefter märkningen av arkivmaterialet. Avslutningsvis ger jag exempel på hur märkningsarbetet kan underlättas och därefter presenteras metoder för migrering av innehåll mellan filformat. Slutligen, och som en övergång till nästföljande kapitel om tillämpningar, beskrivs tillvägagångssätt och överväganden för klassificering av innehåll.

6.1 Beskrivande märkning av kapitalbeskrivningar

Ett syfte med beskrivningen av innehållet och bearbetningen av det aktuella materialet var att det skall vara modifierbart, utbyggbart och plattformsoberoende. I första hand inte syftet varit att samla in uppgifter för att studera enskilda individer utan avsikten med sammanställningen av FFO-KAPITAL var att kunna utföra studier av sociala strukturer och nätverk.⁵⁹⁶

Som tidigare nämnts har jag använt mig av TEI:s modul för prosaliknande texter med tillägget för namn och datum för märkning av kapitalbeskrivningarna.⁵⁹⁷ Prosamodulen är den mest generella av TEI:s moduler och innehåller sådana element som passar för de flesta typer av texter, vilket jag nedan ger exempel på.⁵⁹⁸

6.1.1 Utgångspunkter och principer

För att senare kunna klassificera information för analyser och variabler behövdes principer för vägledning av beskrivningen och kategoriseringen. Innan själva märkningsarbetet påbörjades formulerades därför fyra vägledande principer för hur innehållet skulle beskrivas och ordnas.

⁵⁹⁶ Kapitalbeskrivningarna har dock ett värde var och en för sig, t.ex. för studier av enskilda individer levnadsbanor och livsöden.

⁵⁹⁷ I början användes en lättversion av TEI-schemat, kallad TEI-Lite, se även avsnitt 6.1.2 s. 176 f.

⁵⁹⁸ Se s. 121 för en sammanställning av TEI-modulerna i P4.

Dessa principer kan sammanfattas i fyra huvudklasser: *person, plats, tid* och *företeelse*.⁵⁹⁹ Syftet med principerna var dels att underlätta för den som skulle märka materialet, dels att säkerställa en stringent beskrivning av innehållet. Principerna användes för att enskilt eller tillsammans beteckna de indikatorer som beskrev de egenskaper och tillgångar som var relevanta för datamodellen. En mycket viktig aspekt var att märkningen av kvinnornas egenskaper och tillgångar skulle ske med så lite tolkning som möjligt. Detta innebär att den märkning som tillförts avser beteckna konkreta företeelser – däremot inte antaganden om omfattning eller värderingar av indikatorer på kvinnornas egenskaper och tillgångar. Som exempel kan nämnas märkning av uppgifter om kvinnornas medlemskap i organisationer och föreningar samt i förekommande fall märkning av ytterligare information om ordförandeskap eller dylikt; det vill säga sådana uppgifter som kan kontrolleras i efterhand mot primärkällor och som inte är behäftade med alltför mycket värdering och tolkning.

För analysernas del var det således inte särskilt viktigt att beskriva textens eller dokumentens interna struktur. Men märkning av dokumentets strukturer såsom rubriker och styckesindelningar är nödvändiga för att skapa en grundläggande ordning innan den semantiska märkningen påbörjades. I vissa fall har även raderna tillfogats märkning. Däremot har inte sidbrytningar märkts upp, dels för att det inte var nödvändigt för själva prosopografin, dels för att det inte behövs för kommande bearbetningar. Nämnas kan också att dokumenten inte var avskrivna från en och samma fysiska förlaga utan redan var elektroniska lagrade i Word-format. Informationen om sidbrytningar förloras ändå vid överföringen, eller migreringen, från Word via textformat till XML-formatet.⁶⁰⁰ Migreringen till XML var förstås nödvändig för att märkning med TEI skulle kunna påbörjas.

Hänvisningar och noteringar om de primärkällor som använts som underlag vid författandet av kapitalbeskrivningarna var oftast rudimentära. Ibland fanns endast hänvisningar i form av förkortningar av källorna som till exempel ”SBL” som betecknade att uppgifterna hade hämtats ur artikeln underförstått om personen i fråga i Sveriges Biografiska Lexikon. Dessa uppgifter har märkts som de är nedskrivna i dokumentet för att de senare skulle kunna sammanställas till en gemensam bibliografi för prosopografin.

Sist men inte minst har utgångspunkterna för märkning av detta material även varit att använda verktygen såsom de är konstruerade. Detta innebär att jag inte modifierat TEI-schemat för att beskriva de innehållsobjekt och de strukturer som behövde beskrivas. Anledningarna till detta är tvåfaldigt. För det första hade jag inte mycket erfarenhet av märkningsarbete i början av studien,

⁵⁹⁹ *Företeelse* inbegriper framför allt egenskaper eller tillgångar såsom socialt umgänge, utbildningskapital eller organisationskapital, men även enskilda händelser som ägt rum såsom t.ex. utlandsvistelser.

⁶⁰⁰ Om dokument som är producerade digitalt skall bevara information för paginering måste information om detta infogas, t.ex. genom beskrivande märkning. En annan lösning är att lagra dokumentet i PDF-format.

en modifiering av TEI-schemat ansåg jag vara för tidsödande inte minst med tanke på bristande resurser. För det andra gjorde jag den bedömningen att de elementuppsättningar som fanns verkade räcka till och att jag i studien skulle inrikta mig på att utforska vad TEI:s moduler kunde erbjuda för beskrivning och utformning av den här sortens innehåll.

Beslutet att använda TEI utan modifieringar har resulterat i en del okonventionella lösningar och några fall av *tag abuse*. Användningen av divisioner (elementet <div>) för kategorisering av semantiska innehållsobjekt snarare än beskrivning av dokumentets struktur kan nog uppfattas som en oortodox användning av dessa riktlinjer.⁶⁰¹

[...] elements will be grouped together hierarchically into textual divisions and subdivisions, such as chapters or sections. [...] all such textual divisions be regarded as occurrences of the same neutrally named elements, with an attribute type used to categorize elements independently of their hierarchic level.⁶⁰²

I riktlinjerna framhålls å ena sidan att divisionselementet bör användas för beskrivning av hierarkiska delar av dokumentet och att benämningen *division* används för att å ena sidan på ett neutralt sätt beskriva innehållsobjekt som på något sätt relaterar till varandra. Å andra sidan kan *type*-attributet användas till att karaktärisera en division på ett eller annat sätt oberoende hierarkisk nivå.⁶⁰³ Detta tolkar jag som att det är fullt godtagbart, om än okonventionellt, att använda divisionerna för beskrivning av semantiskt innehåll snarare än strukturella aspekter. Dessutom var inte TEI:s modul för namn och datum utformad för prosopografiska beskrivningar vid tiden för märkningen.⁶⁰⁴

I likhet med ett annat snarlikt projekt, *The Orlando Project*, har det primära här varit att beskriva det semantiska innehållet och inte dokumentens hierarkiska struktur. I Orlandoprojektet har man byggt upp en stor historisk och litterär textdatabas om och med kvinnliga författare från brittiska öarna.⁶⁰⁵

Orlando: Women's Writing in the British Isles from the Beginnings to the Present is a highly dynamic textbase. It is a rich resource for researchers, for students, and for readers with an interest in literature, women's writing, or cultural history more generally. With about five and a half million words of text, it is full of interpretive information on women, writing, and culture. It includes documents on the lives

⁶⁰¹ Detta element introducerades i avsnitt 4.5.1 s. 123, och i avsnitt 6.1.2, s 176 f. presenteras hur märkningen av kapitalbeskrivningarna utförts.

⁶⁰² Citat från avsnitt 4.1 *Divisions of the Body*, i TEI P5, se <www.tei-c.org/release/doc/tei-p5-doc/en/html/D,S.html#DSDIV> (första och andra stycket).

⁶⁰³ ”@type characterizes the element in some sense, using any convenient classification scheme or typology.” Ibid.

⁶⁰⁴ TEI P5 utgavs som tidigare nämnts i november 2007.

⁶⁰⁵ För information om projektet, se <www.ualberta.ca/ORLANDO/> . Större delen av materialet är inte fritt tillgängligt utan kräver prenumeration men exempel på innehåll och metoder finns på projektets webbplats. Däremot kan jag inte hitta deras DTD:er på projektets webbplats, endast beskrivningar om dem.

and writing careers of about a thousand writers, together with a great deal of contextual historical material on relevant subjects, such as the law, economics, science, writing by men, education, medicine, politics.⁶⁰⁶

Till skillnad från många andra digitaliseringsprojekt har man i Orlandoprojektet inriktat sig på författarnas historia, kulturella och samhälleliga förutsättningar och sammanhang. Det man också gjort är att utveckla speciella verktyg eller snarare datamodeller (DTD:er) för att beskriva individers egenskaper och tillgångar och sociala livsvillkor, men även författarnas litterära stilar. Datamodellerna avser spegla Orlandoprojektets teoretiska utgångspunkter i likhet med kapitalbeskrivningsmallen även om de griper över ett större område än vad vi gör.

Förutom de element som finns i TEI har fem egna datamodeller konstruerats, bland annat en mer allmän för märkning av kronologisk information. Två av datamodellerna är utformade specifikt för projektet och används för beskrivningar om författarnas liv, deras texter och litterära karriärer.⁶⁰⁷ En av dessa benämns *Life DTD* och används för beskrivning av kvinnornas egenskaper och tillgångar och har en tämligen hierarkisk struktur. Den innehåller ungefär samma sorts beskrivningskategorier som används i det här materialet. Det vill säga element för beskrivning av namn, födelse, familj, utbildning, vänner etcetera.

Den andra datamodellen, *Writing DTD*, används för att beskriva kvinnornas produktion (utgivning, refusering, datum första publikation etcetera), litterära egenskaper (genre, stil etcetera) och reception (erkännanden, mest kända verk, influerade etc.) och har en plattare och mer flexibel struktur än *Life DTD*. Bland annat använder de sig av en extensiv kronologisk märkning med tid som den överordnade och strukturerande principen.⁶⁰⁸ Denna märkning används sedan för att bearbeta och ordna om det ursprungliga innehållet. Anledningen till att konstruera egna projektspecifika DTD:er var att man i projektet redan från början hade en vision om att beskriva händelser men att huvudsakligen innesluta dem i en kronologisk ordning.⁶⁰⁹ Den beskrivande märkningen används även för nätverksanalyser och formationer samt för analyser av historiska förändringar i likhet med den här studien. Vid Orlandoprojektets webbplats finns inga masterfiler eller DTD:er tillgängliga för granskning eller nedladdning, det är därför svårt att närmare utvärdera deras angreppssätt.

Valet av en relativt platt och därmed flexibel struktur för den beskrivande märkningen – ett val som jag delar med Orlandoprojektet – går egentligen emot

⁶⁰⁶ <www.ualberta.ca/ORLANDO/index.htm> (första stycket).

⁶⁰⁷ DTD och märkningsprinciper tillgängliga <www.ualberta.ca/ORLANDO/tagsets.htm>.

⁶⁰⁸ En kronologisk märkning används till alla typer av händelser och företeelser, se: Isobel Grundy, et al., "Date and ChronStructs: Dynamic Chronology in the Orlando Project", *Literary and Linguistic Computing* 15.3 (2000). Se <www.ualberta.ca/ORLANDO/Chronolpaper.htm>.

⁶⁰⁹ Susan Brown, et al., "Facing the Deep: The Orlando Project Delivery System 1.0", *Text Technology* 14.2 (2005) s. 23.

de grundläggande principerna för den strikt hierarkiska datamodellen som utvecklats med utgångspunkt i strukturerade dokument av olika genrer. Enligt dessa principer har textens semantiska och intellektuella innehåll också en direkt koppling till dokumentets struktur och uttrycks bäst genom en hierarkisk dokumentmodell.⁶¹⁰ Identifiering och beskrivning av semantiskt innehåll kräver möjligen en mer flexibel struktur där det finns möjligheter att placera och flytta om innehållsobjekten (eller ”lådorna”, se *Figur 19* nedan) menar jag. Om det är möjligt att åstadkomma detta utan en eget projektspecifik DTD eller XML-schema kommer jag att diskutera i följande avsnitt.

6.1.2 TEI i arbete

Som inledningsvis nämndes har jag använt mig av TEI:s prosamodul inklusive modulen för beskrivning av namn och datum. När jag påbörjade märkningen använde jag dock en minimerad variant av TEI, *TEI Lite*, som innehåller ungefär en femtedel av elementuppsättningen. Det är alltså ett komprimerat schema som omfattar de element som oftast används.⁶¹¹ Fördelen med denna lättversion är att det är lättare att komma igång med innehållsmärkningen när det inte är så många olika element och hierarkiska strukturer att hålla reda på – kanske flera hundra olika som i de fullständiga modulerna.⁶¹² Nackdelen med lättversionen är att för de mer detaljerade och djupa beskrivningarna räcker elementen inte alltid till för en adekvat beskrivning. Men om man, som jag, påbörjat märkningen i *TEI Lite* går det oftast att ändra elementens namn senare med sök och ersättningsfunktioner – förutsatt att elementen försetts med information om innehållsobjektets typ, går det att identifiera de rätta elementen likväl som med de nya och mer precisa elementnamnen.⁶¹³ Olika TEI-moduler har alltså används för märkningen men principerna är ändå i stort sett desamma oavsett elementuppsättning.

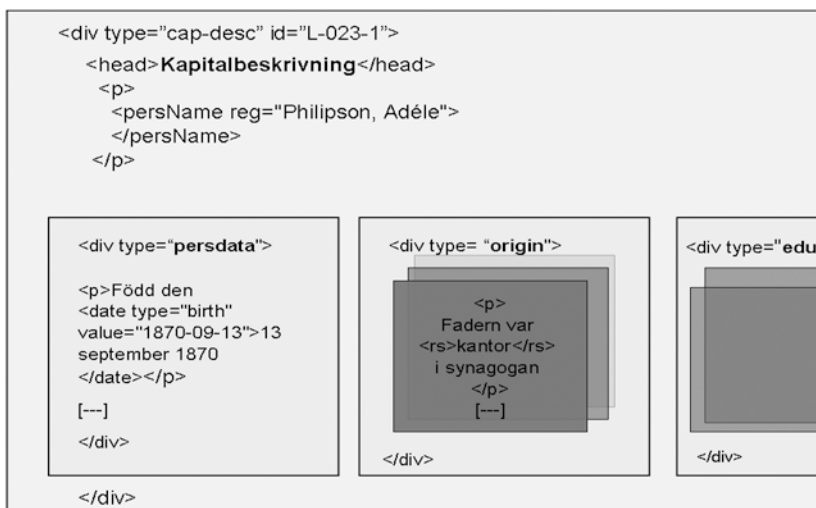
De principer som användes för att återspegla kapitalbeskrivningarnas innehåll med hjälp av TEI:s dokumentmodell illustreras i *Figur 19*.

⁶¹⁰ Jfr. avsnittet om innehållets hierarkiska ordning (OHCO) i kapitel 3.3 s. 77 f.

⁶¹¹ ”TEI Lite is a specific customization of the TEI tagset, designed to meet 90% of the needs of 90% of the TEI user community”. Se <www.tei-c.org/Guidelines/Customization/Lite/>.

⁶¹² I TEI P5 är detta inte längre ett ”problem” eftersom man vid konstruktion av sin egen version av TEI-schemat kan välja bort vissa element som inte kommer användas.

⁶¹³ Vid användning av sök- och ersättningsfunktioner gäller det att vara mycket noggrann och spara kopior av filerna före varje moment. För den ovane är det lätt hänt att det blir fel och att fel ord ändras.



Figur 19. Principer för beskrivning av innehåll i kapitalbeskrivningarna

Genom märkningen kan man säga att texten placeras i "lådor", eller *innehållsobjekt*, som var och en motsvarar ett specifikt innehåll.⁶¹⁴ Här handlar det om att beskriva indikatorer av en viss sort. I figuren ovan visualiseras detta som "lådor" som i sin tur kan innehålla en godtycklig mängd mindre lådor. Dessa innehållsobjekt benämns här "divisioner" <div> i enlighet med benämningen i TEI-schemat. För att beskriva och för att hitta rätt innehållsobjekt kategoriseras elementet <div> genom attributet <type>. Detta ger möjlighet att till exempel extrahera endast de förekomster av Paris som varit föremål för kvinnornas utlandsvistelser och inte alla andra förekomster av Paris, exempelvis information om någons möster som bodde i Paris.

Jag har strävat efter att benämna divisioner koncist såväl som tydligt i syfte att olika typer av elementet <div> i princip skall vara självförklarande.⁶¹⁵ Dessa element kan förekomma var som helst i kapitalbeskrivningens löpande text. Varje <div> kan även innesluta en godtycklig mängd "mindre" divisioner – hur många beror på vad texten innehåller och vilka innehållsobjekt som skall beskrivas. Den överordnade divisionen är dock alltid den avgränsade kapitalbeskrivningen, det vill säga den kvinna texten handlar om.

Elementet <div> används vanligen inte till beskrivning av semantiskt definierade innehållsobjekt utan är framför allt avsedd att beskriva interna dokumentstrukturer såsom kapitel eller avsnitt och dylikt enligt TEI:s riktlinjer.⁶¹⁶ Denna pragmatiska lösning användes för att infoga semistrukturerad och icke-

⁶¹⁴ Principer för dessa innehållsobjekt, här i bemärkelsen kapitaldivisioner, har redovisats i avsnitt 2.2 s. 55 f. samt i avsnitt 3.3 s. 77 f.

⁶¹⁵ Hur divisionerna har benämnts redovisas, se bilaga 3 <CD:/bilagor/bilaga3.pdf>.

⁶¹⁶ Jfr. diskussionen på s. 175.

hierarkiskt innehåll i en hierarkisk dokumentmodell som TEI-schemat representerar.

Viss beskrivning om dokumentets interna struktur förekommer också såsom rubriker <head>, stycken <p> och rader <l>.

6.1.2.1 Märkning av personer och föreningsaktiviteter

Inuti divisionerna har sedan innehåll som kan sägas tillhöra någon av typerna *person*, *plats*, *tid* och *företeelse* försetts med märkning. För beskrivning av egennamn har i stor utsträckning elementet <name> använts och för märkning av tid har elementet <date> använts. Senare byttes flera av <name> elementen ut till de mer specialiserade elementen för egennamn som modulen för namn och datum tillhandahåller. Men principen är i stort densamma även om typningen av de specifika namnelementen blir överflödiga och med att detta redan är gjort i elementets benämning. De specifika namn element som hittills använts är:

- <placeName> för beskrivning av platser till exempel födelse- och boendort i förekommande fall tillsammans med elementet <address>,
- <orgName> för beskrivning av kvinnornas föreningsaktiviteter, sammanslutningar och sällskap,
- <persName> för beskrivning av personnamn.

Jag har inte märkt upp för och efternamn var för sig då detta inte skulle tillföra tillräckligt många fördelar för vare sig bearbetning, sökning eller presentation. I och för sig skulle detta förfinas sökfunktionen. Separeringen av för- och efternamn har dock gjorts i databasen vilket jag anser räcka.

För beskrivning av en viss sorts *företeelse* har antingen elementet <name>, med varianter, eller <rs> använts. Skälet till det är att jag inte kunde finna något lämpligt element i TEI Lite (P4) för beskrivning av företeelser. En företeelse kan beteckna exempelvis medlemskap i en kvinnopolitisk förening, som tidigare illustrerades i *Figur 7* s. 79 som exemplifierar märkning av Klara Johanssons medlemskap i två olika föreningar. Jag hade i och för sig kunnat skapa ett element för *events* kombinerat med <date> för detta ändamål men det gjordes som sagt aldrig.

Inuti ”organisationsdivisionen” <div type=“orgcap”> finns en godtycklig mängd innehållsobjekt som beskrivits och kategoriserats, till exempel föreningsaktiviteter. Här användes till en början attributet <name type=“org”> för att särskilja vilken sorts egennamn som avses. Föreningens eller sammanslutningens namn har alltså använts som indikator på föreningsaktivitet. Genom användning av *type*-attributet klassificeras vilken sorts egennamn som avses och med hjälp av *key*-attributet relateras den typen med en specifik underkategori. Kombinationen <name type=“org” key=“kvinnopol”> användes till en början för en struktur som beskrev medlemskap i organisationer och föreningar. Genom att använda attributet *key=“kvinnopol”* kopplades attributvärdet ”kvinnopol” ihop med en post som beskriver vilka föreningar och sammanslutningar som vi i projektet klassi-

ficerat som kvinnopolitiska.⁶¹⁷ Tillvägagångssättet modifierades något senare när de specifika elementen för egennamn började användas, *key*-attributet används dock fortfarande på samma sätt. I *Figur 20* nedan kan vi se vad posten för typen ”kvinnopolitiska sammanslutningar” innehåller.⁶¹⁸

```
<div type="org_kvinnopol" id="kvinnopol">
  <head>Kvinnopolitiskt inriktade sammanslutningar</head>
  <p>
    <orgName> Akademiskt Bildade Kvinnors Förening</orgName>
    <orgName> Federationen</orgName>
    <orgName> Fredrika Bremerförbundet <abbr>FBF</abbr></orgName>
    <orgName> Föreningen för gift kvinnas eganderätt</orgName>
    <orgName> Föreningen för kvinnans politiska rösträtt <abbr>FKPR</abbr></orgName>
    <orgName> Kvinnors demokratiska Världsförbund</orgName>
    <orgName> Radikala föreningen</orgName>
    <orgName> Svenska kvinnors medborgarförbund</orgName>
    <orgName> Svenska Kvinnors Nationalförbund</orgName>
    <orgName> Tidevarvet</orgName>
    <orgName> Dräktreformföreningen</orgName>
    <orgName> Diskussionsklubben</orgName>
  </p>
</div>
```

Figur 20. Exempel på poster som värdet av ett *key*-attribut relaterar till

Genom att relatera *key*-attributets värde ”kvinnopol” till divisionselementets unika id får vi reda på att i kategorin ”kvinnopol” ingår dessa föreningar och sammanslutningar. Det går även att infoga ytterligare information om de enskilda föreningarna till exempel länkar till information om föreningen. I posten som *key*-attributet relateras till skulle man även kunna definiera rekommenderad stavning av de egennamn som posten innehåller. Jag har dock valt att använda denna möjlighet för var och en av de taggade förekomsterna av föreningsnamn. Detta ger upphov till redundant märkning, eftersom samma stavningsform upprepas flera gånger i dokumentet. Fördelarna är dock att masterfilen med samtliga texter inte är beroende av en länkning till ett externt dokument vid extrahering eller migrering av data.

Metoden för normalisering av stavningsvarianter är i princip samma som för attributet <key>. För presentation av materialet och för att säkerställa att det beskrivna innehållet senare kan återfinnas har alla egennamn försetts med normaliserade värden i attributet <reg>.⁶¹⁹ Denna procedur har inte varit helt oprob-

⁶¹⁷ Dessa poster kan antingen förvaras internt i samma dokument som kapitalbeskrivningarna eller i ett särskilt externt XML-dokument avsett för beskrivning av föreningar. Skillnaden består i på vilket sätt man länkar till denna information.

⁶¹⁸ Dokument med samtliga föreningsposter, se <CD:/xml/orgs.xml>.

⁶¹⁹ Ett exempel är ”Föreningen för gift kvinnas äganderätt” där följande varianter förekommer ”äganderätt” och ”eganderätt”. Se *Figur 25* s. 188 för exempel på användningen av *reg*-attributet.

lematisk, då olika stavningsformer förekommer i texterna beroende på tidsperiod, källa och författare. Därför är det desto viktigare att normalisera stavningskonventionen. Genom att använda ett attribut där den normaliserade stavningen lagras slipper man ändra i den ursprungliga texten.

Detta är förstås särskilt viktigt att ta ställning till vid avskrifter av fysiska förlagor. Beroende på syftet med märkningen kan man antingen producera en diplomatarisk eller en normaliserad text eller varianter där emellan. I den diplomatariska utgåvan skrivs texten av såsom den är, inklusive felskrivningar och annat. I den normaliserade utgåvan ändras texten utifrån bestämda principer.⁶²⁰ Fördelarna med den beskrivande märkningen är att det blir möjligt att göra flera varianter av en och samma bastext eller förlaga. Även om jag i det här fallet har transkriberat en fysisk förlaga kan jag genom att normalisera stavningen skapa olika lager av textens beskrivning: en diplomatarisk och en normaliserad. Matthew Driscoll framhåller ytterligare en fördel med digitaliseringen och det är möjligheterna att utan alltför stora åthävor återvända till texten och ändra, ta bort eller lägga till märkningen.⁶²¹ I detta material skulle jag kunna infoga en normaliserad stavning baserat på en annan norm än den jag använt till exempel för att visa hur stavningen av egennamn varierat under 100 år.

6.1.2.2 Märkning av sociala kontakter

En inte helt trivial uppgift var beskrivning av kvinnornas sociala kontakter. För att begränsa nätverksanalyserna till en hanterbar analysmängd valde jag att märka de kontakter som kvinnor hade med andra kvinnor i gruppen. En annan anledning var att information om externa kontakter saknas i flera av kapitalbeskrivningarna, underlaget var helt enkelt för tunt för att motivera en sådan märkning. Som tidigare påpekats är det fullt möjligt att infoga ytterligare märkning om de "externa" kontakter kvinnorna hade.⁶²²

Även för beskrivning av kvinnornas sociala nätverk användes elementet <name>. Denna gång kategoriserades namnelementet med ett annat attributvärde för att beskriva att en viss kvinna haft kontakt med en annan kvinna i materialet.

⁶²⁰ En variant av den normaliserade utgåvan är den moderniserade som anpassats till modernt språkbruk. Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion* s. 205.

⁶²¹ Driscoll, "Levels of Transcription" s. 258.

⁶²² Extern i förhållande till den undersökta gruppen.

```

<div type="soccap">      <head>UMGÅNGE SOM VUXEN </head>
<p>
  <name type="friend" key="ffo" reg="Whitlock, Anna">
    Anna Whitlock</name> och Julia Kjellberg
    Ca vintern <date>68-69 </date> i beröring med
    <name type="friend" key="ffo" reg="Adlersparre, Sophie">
      Sofie Adlersparre </name>
    [---] Tidigt /ca <date>1868</date>??/ bekant med
    <name type="friend" key="ffo" reg="Hierta Retzius, Anna">
      Anna Hierta </name>
    [---]
  </p>
  [---]
</div>

```

Figur 21. Märkning av en kvinnas kontakt med andra kvinnor i materialet

Exemplet ovan, som är ett utdrag ur kapitalbeskrivning om Ellen Key, visar hur märkningen av kvinnornas kontaktnät utförts i den löpande texten. Här fungerar namnelementets attribut som en pekare, eller länk, till andra personnamn. Attributvärdet "friend" beskriver att det är personer som kvinnan i någon bemärkelse haft kontakt med. Benämningen "friend" är egentligen missvisande då kvinnorna inte alls behövde vara vänner för att kvalificeras som kontakt. Tolkningar av vilken sorts kontakt de hade med varandra (vän, ovän eller tillfällig bekantskap) har inte gjorts med hjälp av märkningen. Sådana tolkningar har överlämnats till författarna av kapitalbeskrivningarna. I nätverksanalyserna använder jag däremot ordet *kontakt* i och med att det kan röra sig om ytliga bekanta, som i *Figur 21* ovan, likväl som vänner eller professionella kontakter samt för att undvika tolkningar med utgångspunkt i vad som står eller inte står om kvinnornas umgänge i kapitalbeskrivningarna.⁶²³

Personnamnen har även de försetts med normaliserad stavning, i likhet med märkningen av föreningar och organisationer, i syfte att bättre kontrollera extrahering av önskade data. De personer som inte ingår i FFO-KAPITAL finns ännu inte med i märkningen av personnamn. I exemplet i *Figur 21* ovan är det Julia Kjellberg som inte märkts upp med namnelementet för att hon inte ingår i FFO-KAPITAL.

En nackdel med att märka kvinnornas kontakter på det sätt som beskrivits ovan är att relationen blir enkelriktad istället för ömsesidig. Det är endast i de kapitalbeskrivningar som kontakter med andra som explicit nämns som kan registreras genom märkningen. För vissa kvinnor kan det då hända att det är fler pekare *från andra* till henne eller tvärtom. För andra kvinnor är det fler pekare *till andra* kvinnor än från andra i kapitalbeskrivningarna. Information om dessa kvinnors sociala nätverk blir därmed beroende av vad författaren valt att ta upp eller haft kännedom om. Dessa enkelriktade relationer kan dock användas för analyser av strukturen på kvinnornas kontaktnät som exemplifieras i *Figur 33*, s. 211.

⁶²³ Exempel på nätverksanalyserna presenteras i kapitel 7.2 s. 208 f.

6.1.2.3 Märkning på olika logiska beskrivningsnivåer

Tillvägagångssättet som ovan presenteras ger därmed upphov till att flera olika hierarkier kan beskrivas. En hierarki beskriver textens struktur genom märkning av stycken och rubriker. Den överordnade semantiska hierarkin är de divisioner som beskriver textens semantiska innehållsobjekt, det vill säga kvinnornas olika typer av egenskaper eller kapitalformer. Även kapitalbeskrivningsmallen egen struktur har påverkat märkningen. Genom märkningen ordnas innehållet såväl hierarkiskt som logiskt, dels på grund av mallen, dels på grund av XML:s hierarkiska dokumentmodell.⁶²⁴ Därför är det viktigt att divisioner av samma typ placeras på rätt logiska nivå relativt mallen.

Ett exempel på det är beskrivningen av utbildning. I kapitalbeskrivningsmallen har information om faderns yrke, utbildning och position samlats under en egen rubrik. För kvinnornas del har yrke, verksamhet och utbildning särskiljts vilket förstås förenklar märkningen då informationen, i de flesta fall, redan är logiskt strukturerad. Jag har märkt upp indikatorer på utbildning både för kvinnorna och för deras fäder med divisionen `<div type="edcap">`. Noggrannhet är i detta fall en dygd. Typen "utbildning" betecknar samma logiska slag men för olika individer. I mallen, där ju individen är den överordnande principen, hamnar beskrivningen för faderns utbildning i en hierarkisk nivå under nivå 3, se *Figur 17*, s. 159. Det måste den också göra annars hamnar faderns utbildning på samma nivå som kvinnans, det vill säga nivå 3, vilket inte bara blir logiskt fel utan även fel vid extraheringen av data. Om inte de två divisionerna som betecknar utbildning hamnar på rätt nivå i dokumentets hierarki, för kvinnan respektive hennes far, kommer fel värde att extraheras.⁶²⁵ Ett alternativ till att säkerställa att rätt indikator betecknar rätt person i kapitalbeskrivningen är att tillföra ytterligare en typning av elementen. Genom att använda ett attributvärde som särskiljer fadern från kvinnan `<div type="edcap" n="f">` kan man lätt identifiera och filtrera ut endast de värden som avser till exempel faderns utbildning.

Jag har haft för avsikt att beskriva de strukturer som beskriver medlemskap i föreningar och kvinnornas kontaktnät och som inte är beroende av indelningen i divisioner. Bradley beskriver hur märkning av multipla hierarkier kan utföras med hjälp av TEI-schemat.⁶²⁶ Den hierarkiska dokumentmodellen tillåter endast beskrivning av ett mindre antal hierarkiska strukturer. Om beskrivningen och länkningen blir för komplex, till exempel om man behöver beskriva fler än tre relationer mellan personer, duger inte TEI-schemat eller snarare den hierarkiska

⁶²⁴ Jfr. *Figur 14* s. 123 samt *Figur 17* s. 159.

⁶²⁵ I filen `div-edu.xml` exemplifieras märkningen av utbildning Sophie Adlersparre och hennes far. Se `<CD:/xml/div-edu.xml>`. De skript som har använts för extrahering av data ur xml-filerna skulle, om faderns utbildningsindikator låg på samma hierarkiska nivå som kvinnans, hitta den som kommer först i dokumentet, i detta fall faderns. Detta beror på att skriptet läser in (*parse*) informationen i xml-filen sekventiellt och därför läser in den data som först påträffas – förutsatt att det är den som eftersöks.

⁶²⁶ Bradley, "Documents and Data: Modelling Materials for Humanities Research in XML and Relational Databases".

dokumentmodellen särskilt bra längre.⁶²⁷ Det är mycket möjligt men för den märkning som beskrivits här har TEI-schemat varit tillräckligt.

Dock genererar de tillvägagångssätt som ovan beskrivits en hel del redundant märkning, till exempel att behöva typa de element som redan befinner sig skall vara inuti rätt division ("låda") som i fallet med kvinnans och faderns utbildning.

6.1.2.4 Avslutning

Genom det tillvägagångssätt som ovan presenterats är att texterna lätt kan för-ses med ytterligare märkning som beskriver andra typer av innehållsobjekt eller senare helt rensas från all märkning som tidigare har infogats. Den främsta fördelen är emellertid att tack vare den märkning som jag ovan gett exempel på kan ett och samma textmaterial – som för närvarande innehåller 101 relativt omfattande kapitalbeskrivningar och drygt 1200 kortare uppgiftssammanställningar samt de 1980 arkivposter som registrerats – kan användas i flera tillämpningar.⁶²⁸ Detta är dessutom genomförbart med fria programvaror och öppna standarder.

Genom exemplen har jag även visat hur beskrivning och kategorisering av innehållsobjekten gått till. I filen <gagner-marielouise.xml> ges exempel på märkning av en ganska omfattande kapitalbeskrivning, i filen <hermelin-honorine.xml> ges exempel på en mer rudimentär märkning och i bilaga 2 presenteras en sammanställning över den kategorisering av de innehållsobjekt som använts för att beskriva det semantiska innehållet och den logiska ordningen.⁶²⁹

6.1.3 Sammanfattning av märkningsarbetet

Beskrivningen av det prosopografiska materialet har skett utifrån följande verktyg och principer:

- 1) Projektets datamodell (*kapitalbeskrivningarna*)
- 2) De grundläggande principerna för beskrivning av innehållsobjekten (*person, plats, tid & företeelse*)
- 3) TEI:s hierarkiska dokumentmodell (prosamodulen inkl. *namn & datum* i TEI P4)
- 4) Kategorisering av innehållsobjekten med hjälp av *type-attributet* (överordnad kategori) och i förekommande fall även med *key-attributet* (underordnad kategori till *type-attributet* som används för

⁶²⁷ Ibid. s. 140-141.

⁶²⁸ Hela FFO-databasen innehåller vid en räkning i juni 2007 cirka 270 000 ord och flera än 40 000 element. Kapitalbeskrivningarna omfattar drygt 200 000 ord och drygt 16 000 element.

⁶²⁹ Samtliga dokument tillgängliga i katalogen <CD:/xml/> (för filerna) samt i katalogen <CD:/bilagor/> för bilagan med kategoriseringar.

att lagra ytterligare information om typen av objektet eller objektet i sig)⁶³⁰

- 5) Normaliserad stavning av egennamn med hjälp av *reg-attributet*

Märkningen av innefattar såväl interna struktur (avsnitt, stycken och rubriker) som märkning av de semantiska innehållsobjekten (kategoriserade divisioner och egennamn). För att beskriva den semantiska strukturen har en oortodox användning av TEI:s hierarkiska dokumentmodell och prosamodulen använts. De element som används för beskrivning av textens interna struktur har använts för beskrivning av semantiskt definierade innehållsobjekt. De innehållsobjekt som betecknar egennamn har grundligt kategoriserats för att kunna återfinnas oavsett placering i dokumentet.

Detta tillvägagångssätt medför att olika hierarkier och strukturer i materialet kan beskrivas. En pragmatisk och projektspecifik lösning har alltså använts för att infoga semistrukturerad och icke-hierarkisk information (kapitalbeskrivningar) i en väl strukturerad hierarkisk dokumentmodell (TEI:s prosamodul). Projektets specifika intressen har därmed kunnat kombineras med den generella modellens förutsättningar. En sådan strategi kan dock försvåra framtida utbyte av det beskrivna innehållet och förutsätter en noggrann dokumentation om tillvägagångssätt.

6.2 Beskrivande märkning av arkivmaterial

Märkningen har för innehållet i FFO-HLS skett med TEI:s schema P5, som publiceras hösten 2007.⁶³¹ Den största skillnaden jämfört med märkningen av FFO-KAPITAL är att de prosopografiska element som används för beskrivning av personer nu är mer precisa än i tidigare versioner av TEI P4.

En annan skillnad är förstås karaktären på det källmaterial som har försetts med beskrivande märkning. Detta källmaterial är i sig självt väl strukturerat och följer samma mönster på alla sidor i böckerna, något som underlättar märkning såväl som val av lämpliga element. Som tidigare nämnts innehåller alla poster rubriker för information om namn, socialt ursprung, tidigare utbildning, studiegång vid seminariet, anställningar efter examen etcetera⁶³²

Jag har tidigare beskrivit vilka rubriker som anger vilken information som förtecknats i matrikeln. I bilden nedan illustreras hur en sida i matrikeln är strukturerad. Därefter ges exempel på märkning av en sida ur inskrivningsmatrikeln för de elever som skrevs in under perioden 1914–1928.

⁶³⁰ Användningen av key-attributet i den här studien beskrivs på s. 181.

⁶³¹ Schemat har varit tillgängligt i olika versioner sedan 2005.

⁶³² Se s. 155 för exempel webbapplikation av just denna post.

42

Namn: Ingeborg Anna Margaretha Tehning

Födelsår och födelseort: den 2 september 1908 i Kristinehamns förs., Kärnelän

Faderns yrke: Folksamman

Utbildning före inskr. vid seminariet: Förgångsbetyg från Kristinehamns elementarskolor för flickor år 1925

Föregående anst. som lärarinna (älv. privatundervisn.): Guvernant 1925-26

Sommarplats 1930 hos grevinnan Wetterberg

Studieår vid seminariet: 1927 - 1930 (dec.)

Lägre kursens valfria ämnen: Svenska, engelska, franska, historia, geografi.

Figur 22. Exempel på post i inskrivningsmatrikel för perioden 1914–1928 (Serie D1, vol. 2)

Ovan illustreras de första raderna för en av de mer detaljerade posterna och nedan illustreras märkningen av dessa uppgifter.⁶³³

```
<pb rend="top-right" n="42"/>
<div type="name">
  <fw type="name" place="margin-left" n="1">
    <hi rend="bold">Namn:</hi>
  </fw>
  <p>
    <persName>
      <forename>Ingeborg</forename>
      <forename type="main">Anna</forename>
      <forename>Margaretha</forename>
      <surname>Tehning</surname>
    </persName>
  </p>
</div>
<div type="birth">
  <fw type="birth" place="margin-left" n="2">Födelsår och födelsort:</fw>
  <p>Den <date when="1908-09-02">2:a september 1908</date> i <placeName
    type="parish">Kristinehamn</placeName>s
    <choice><abbr>förs.</abbr><expan>församling</expan>
    </choice>.
```

Figur 23. Märkning av arkivpost (urklipp)

Överst används det tomma elementet `<pb/>` för att beskriva sidbrytning, där anges även sidans nr och sidnumreringens placering.

Elementet `<fw>` används för begrepp som upprepas i sidhuvuden, i marginaler etcetera. Därför ansåg jag det lämpligt att använda här då varje sida ser lika-

⁶³³ XML-filen som innehåller hela den märkta posten bifogas vid: `<CD:/xml/khls-example.xml>`.

dan ut.⁶³⁴ I exemplet ovan anges i attributen för <fw> även information om typ av innehåll samt placering av texten i dokumentet.

Elementet <persName> är mer kraftfullt än det mer allmänt hållna namn-elementet som tidigare beskrivits. Detta med anledning av att varje typ av personnamn kan märkas med avsett element som gjorts i exemplet ovan. Det hade gått lika bra att använda <name> även för detta och att använda type-attributet för att särskilja för- och efternamn. Type-attributet i <persName> kan då istället användas för ytterligare klassificering av namntyper, exempelvis "pseudo" eller "namn_som_ogift".

En annan användbar nyhet i TEI P5 är införandet av elementet <choice> som tillåter beskrivning av flera varianter av ett ord. I detta fall har man i inskrivningsmatriklarna använt förkortningar i väldigt stor utsträckning. Dessa har märkts upp såsom de står. Men genom att använda elementet <choice> anges även det utskrivna ordet. Vid bearbetningar och presentationer kan man sedan välja vilken av varianterna som skall visas. I de fall då osäkerhet föreligger över vad förkortningen avser beteckna finns möjligheter att ange även detta. Elementen <abbr> och <expan> har även en rad olika attribut som tillåter ytterligare beskrivning, kategorisering eller länkning till andra förekomster i samma dokument.⁶³⁵

Eftersom den transkriberade texten skall sammankopplas med den bild som avbildar originaldokumentet är det särskilt viktigt att använda ett gångbart system för detta. Här används den redan tidigare fastslagna strukturen med serier och volymer beroende på typ av arkivmaterial. Här följer exempel på hur länken mellan text och bild beskrivs i XML-filen.

```
<div xml:id="D1.2-057" n="057">
  <linkGrp xml:base="#D1.2-057" type="picture">
    <ptr target="/pict/hls/D1/2/057.jpg"/>
  </linkGrp>
  <linkGrp xml:base="#D1.2-057" type="reg">
    <ptr target="D1.1-1649"/>
  </linkGrp>
```

Figur 24. Märkning av unikt id samt länkning till externa dokument

I figuren ovan illustreras hur dokumentets id hänger ihop med övriga attributvärden. Den första raden beskriver en unik identifikation av just den här posten, samt dess löpnummer, som är samma som bildens. På nästa rad anger en

⁶³⁴ "<fw> (forme work) contains a running head (e.g. a header, footer), catchword, or similar material appearing on the current page." TEI Consortium, red, *TEI Consortium. Guidelines for Electronic Text Encoding and Interchange. P5*, <www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-fw.html>.

⁶³⁵ <choice> groups a number of alternative encodings for the same point in a text. Ibid. <www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-choice.html>.

grupp länkar som på något sätt hör ihop och attributvärdet `xml:base="#D1.2-057"` avser en relativ referens som senare kan omvandlas till en absolut referens. Inuti detta element anges en pekare till den bild som texten är en avskrivning av. Denna information kan nu användas dels för att sammanföra text och bild, dels för att säkerställa att texten i fråga relaterar till rätt fotografi.

I nästa grupp länkar används attributvärdet för att sammankoppla denna post med motsvarande post i det första registret, där samtliga kvinnor är registrerade. På detta sätt finns möjlighet att koppla ihop hela eller delar av många olika dokument. Något som blir användbart när, eller ifall, övriga dokument i arkivet blir transkriberade.

Märkningen av resterande poster i dokumentet går till på samma sätt, med beskrivning av rubrikerna och registrering av de uppgifter som där har förtecknats.

I TEI:s modul för märkning av datum, namn, person och plats finns ett specifikt element `<person>` för identifiering och beskrivning av en individ och dennes egenskaper. Till detta element finns underelement som beskriver yrke, födelsedatum, födelseplats etcetera. Det vill säga ett sådant element som hade varit mycket användbart för detta material. Tyvärr är det endast tillåtet att använda detta element i *TEI-header*, det vill säga inte i den löpande texten utan i den del av dokumentet där olika sorters metadata antecknas. En möjlighet hade då varit att förteckna samma uppgifter som i *TEI-header*, något som jag egentligen anser inte är ett alternativ. Att på detta sätt lägga ner tid på tidsödande registrering av redundant information känns onödigt. Samma sorts uppgifter om namn, yrke etcetera registreras ju i själva filen.

På samma sätt som för den beskrivande märkningen av kapitalbeskrivningarna har här innehållets struktur påverkat märkningen. Men till skillnad från kapitalbeskrivningarna har jag här kunnat använda TEI:s modell såsom den är tänkt att användas och inte på det något okonventionella sätt som gjordes i tidigare.

6.3 Verktyg för semi-automatiserad märkning

I detta avsnitt ger jag exempel på hur tidsödande märkning av samma typ av ord delvis kan automatiseras. Eftersom kapitalbeskrivningarna är skrivna i löpande text förekommer det att uppgifter egentligen hör hemma i en annan division beroende på vilken typ av egenskap eller tillgång som beskrivs. Data-modellen jag använt med kategoriserade element för beskrivning av till exempel egennamn gör det möjligt att identifiera dessa element oavsett var i dokumentet uppgifterna förekommer. Detta under förutsättning att elementet är rätt kategoriserat.

Verktyg för automatisk taggning av ordklasser och egennamn utvecklas framför allt inom datorlingvistik. Jag prövade ett webbaserat verktyg för svensk text som märkte olika slags egennamn och datum av en given text-

mängd.⁶³⁶ Metoden fungerade inte särskilt bra på kapitalbeskrivningarna och en anledning till det var att taggningsverktyget var avsett för tidningstexter snarare än prosaliknande texter som emellanåt innehöll namn med äldre stavning. Min bedömning av denna utprovning var att det skulle kräva en så pass stor arbetsinsats att korrigera misstagen att det var mer tidsbesparande att själv utföra märkningen med hjälp av verktyget DL-ANTE, som utvecklades för just den här sortens repetitiva märkning och typning av innehållsobjekt.⁶³⁷

För att automatisera märkningsarbetet av innehållsobjekt som återkommer upprepade gånger i en text utvecklades ett verktyg som automatiskt infogar märkning och samtidigt typar och numrerar de märkta elementen. Verktöget DL-ANTE (Digital Litteratur: Automatisk numrering och typning av element) är en mer avancerad sök- och ersättningsfunktion och fungerar enligt följande principer.

För att märka alla förekomster av till exempel "Sällskapet Nya Idun" i en xml-fil anges i programmet:

- först och främst det namn man önskar söka efter (*Nya Idun*),
- därefter vilken typ av egennamn det är (*org*),⁶³⁸
- den normaliserade stavningen av detta egennamn (*Sällskapet Nya Idun*),
- samt en gemensam identifierare som blir en del av den första delen i ID-numret och där den andra delen består av ett löpnummer.

Detta resulterar i att applikationen söker efter förekomster av textsträngen *Nya Idun* i den fil som lästs in (det vill säga som är öppnad). Varje textsträng omges samtidigt av elementet <name> inklusive attribut och attributvärden som angetts. Nedan ett exempel på resultatet, det vill säga en förekomst av föreningsnamnet.

```
<name type="org" id="SNI.078" reg="Sällskapet Nya Idun" key="sallsk">Nya Idun</name>
```

Figur 25. Semiautomatisk märkning av föreningsnamn

⁶³⁶ I Nomen Nescio-projektet har man utvecklat metoder för igenkänning, klassificering och annotering av egennamn i löpande text. Se Janne Bondi Johannessen, et al. "The Nomen Nescio Project - Scandinavian Named Entity Recognition". *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (ALLC/ACH)*, June 11 - 16, Göteborg University, 2004. För ytterligare information och testverktyg se projektets webbplats <<http://g3.spraakdata.gu.se/nn/>>.

⁶³⁷ Tack till Albert Rigo som våren 2003 var behjälplig i en tidig version av DL-ANTE.

⁶³⁸ Detta moment behövs inte om man använder elementet <orgName> då det redan är typat. Men då kan man istället använda attributet *type* till att beskriva en underklass på samma sätt som jag använder attributet *key* i exemplet ovan.

För att utföra en komplett märkning på just *Sällskapet Nya Idun* kan det behövas flera sökningar då applikationen inte kommer att hitta förekomster som använder förkortningar såsom ”SNI”. Sök- och ersättfunktionen är programmerad så att den exakta strängen som skrivs in också är den som eftersöks. Detta för att förhindra att fel ord eller benämningar blir märkta och klassificerade av misstag. I detta fall hade annars alla förekomster av ”Idun” eller ”Nya” blivit märkta. Idun är även namnet på en tidskrift och en manlig motsvarighet till det kvinnliga sällskapet, ”Nya” kan finnas med i en rad andra ord. Det krävs att man först söker igenom filen för att finna möjliga varianter av den textsträng man önskar märka. För detta kan man ta reguljära uttryck till hjälp som tillåter sökning på delar av ett ord, olika stavningar etcetera.⁶³⁹ För märkning av stora textfiler där samma typ av ord skall märkas är det mycket användbart att lära sig använda reguljära uttryck eller andra metoder som underlättar den ofta tidsödande märkningen. Verktaget DL-ANTE⁶⁴⁰ som här presenterats är användbart för sådana ord, till exempel ofta förekommande egennamn, som är av samma slag. Men trots dessa hjälpmedel krävs förutom kontroll ofta att ändringar genomförs manuellt eftersom det hela tiden handlar om att tolka vad orden betecknar.

6.4 Metoder för migrering av filformat

I följande avsnitt kommer jag att beskriva de moment som kan behövas vid migrering av datafiler och dessa innehåll mellan olika program och filformat. Inledningsvis ges en översiktlig beskrivning över tillvägagångssätt för olika material. Avslutningsvis ges ett mer detaljerat exempel på hur man ur text som försetts med beskrivande märkning kan extrahera exakt de ord som är av intresse.

En förutsättning för en mångsidig användning av ett digitalt material för forskning och undervisning är att innehållet går att använda i flera olika applikationer. Målsättningen med bearbetning och presentation av de prosopografiska databaserna är tillgänglighet och modularitet. Grundmaterialet, det vill säga kapitalbeskrivningar, matriser, avskrifter av källor, bilder etcetera finns tillgängligt på webben för forskare och andra. Vidare är det möjligt att ladda ner större delar av materialet i FFO-KAPITAL samt FFO-LUNDSTROM.⁶⁴¹

⁶³⁹ Den normaliserade stavningen av Sällskapet Nya Idun som använts utgår från den stavning som anges i *Nordisk Familjebok*, Vol. 28, 1919 (Uggleupplagan 38 band, 1904-1926). Sällskapet Nya Idun bildade 1885 i syfte att sammanföra verkssamma kvinnor inom det vetenskapliga, litterära och konstnärliga området. Ellen Key var en av grundarna och Ellen Fries dess första ordförande, båda kvinnorna finns med i FFO-KAPITAL. Idun fanns också som manligt sällskap sedan tidigare, därav benämningen Nya för kvinnornas sällskap. Idun är också namnet på en tidning som utgavs under perioden 1887-1963.

⁶⁴⁰ DL-ANTE bifogas i avhandlingens CD:/verktyg/DL-ANTE/ tillsammans med en enkel användarhandledning. Se även <www.skeptron.uu.se/monicalz/ffo/tools>.

⁶⁴¹ Se <www.skeptron.uu.se/monicalz/ffo/>.

Användare skall helst inte behöva installera nya program eller vara tvungna att investera i dyr programvara. Ibland går det dock inte målsättningen att uppnå. De interaktiva kartorna är exempel på det, för att använda dem behövs att just den programvaran installeras i användarens dator och även om den är gratis kan det vara en tröskel för många. Min strävan har varit att så långt som möjligt använda öppna program och standarder för bearbetning och presentation av detta material. Detta har inte alltid varit möjligt att uppnå. För flera av de statistiska bearbetningarna, som till exempel geometriska dataanalyser, krävs specifika programvaror som det inte finns motsvarigheter till med fria program.⁶⁴²

6.4.1 Migrering av innehåll och lagringsformat

Migrering av filformaten har skett i olika steg, och ibland behövs innehållet i filerna modifieras något medan det vid andra tillfällen är enklare. Nedan kommer jag att beskriva migreringens olika moment.

Först och främst behövdes det Wordformat som kapitalbeskrivningarna levererades i omvandlas till vanlig textformat. Detta görs enkelt genom att öppna Wordfilerna i Word eller Open Office och spara filen i enkelt textformat. Därtill gäller det att ge textformatet rätt teckenkodning. Vidare gäller det att XML använder Unicode för teckenkodning vilket som deklarerar i början av filen.⁶⁴³ Att ändra teckenuppsättning är inte särskilt svårt det görs utan större svårigheter i ett enklare editeringsprogram.⁶⁴⁴ Man öppnar filen och väljer att spara den valfri teckenkodning, för XML ska det vara Unicode (UTF).⁶⁴⁵ Om man ändå väljer att använda ASCII-kodning av textfiler bör man ersätta de tecken som inte ingår i Latin-1, till exempel våra tecken för å, ä, ö, med avsedda entiteter om man vill vara säker på att texten visas korrekt i en webbläsare.⁶⁴⁶

Figur 26 nedan illustrerar arbetsmomenten för den första delen av filformatsöverföringen.

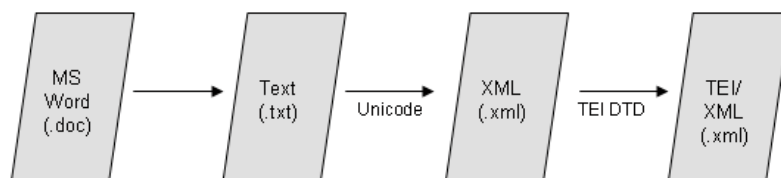
⁶⁴² Här avses statistikplattformen SPSS samt programmet SPAD som använts för geometriska dataanalyser (korrespondensanalyser), se exempel på korrespondensanalys i Zetterman, *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet..* Många av de enklare deskriptiva analyserna går det att utföra i Calc (eller Excel som inte är fri programvara), flera av analyserna kräver användning av program som SPSS som inte heller är fritt. Information om SPSS vid <www.spss.com/se/> och om SPAD vid <www.spad.eu/>.

⁶⁴³ I ett XML-dokument deklarerar formatet och teckenkodningen allra först. För exempel se någon av de bifogade xml-filerna i katalogen <CD:/xml/>.

⁶⁴⁴ T.ex. i Notepad (Anteckningar).

⁶⁴⁵ UTF, *Unicode Transformation Format*, teckenkodningsformat som utökar datorernas möjligheter att representera världens skriftsystem och samtidigt vara kompatibelt med äldre teckenrepresentationer, som huvudsakligen var ASCII.

⁶⁴⁶ ASCII (*American Standard Code for Information Interchange*), teckenkodning som används för att koda bokstäver och andra tecken i datorer, ersätts nu allt oftare av Unicode. Entiteter är de objekt som används för att representera andra tecken: t.ex. bokstaven å motsvaras av "å". Numera klarar ofta webbläsarna av att automatiskt läsa av dokumentens teckenkodning så att innehållet visas med rätt uppsättning tecken. Problemet med olika teckenkodningar är inte på något sätt trivialt och kan ställa till mycket problem vid presentation och inläsning av text.



Figur 26. Översikt för migrering av filformat

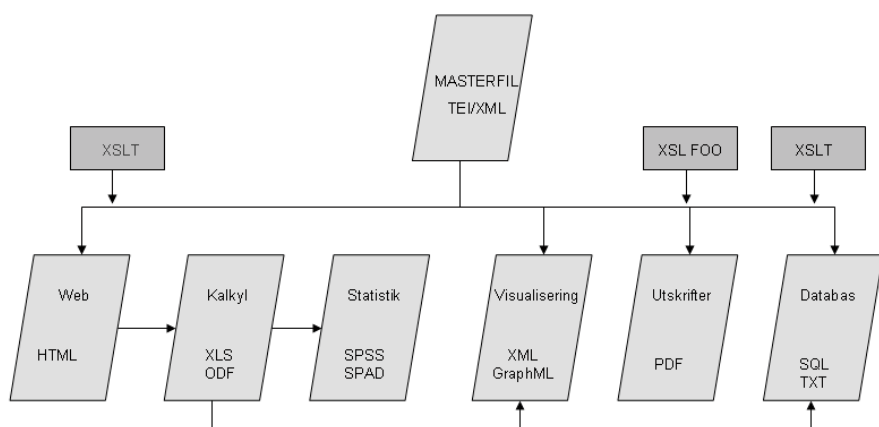
Nästa steg efter att filen konverterats till en textfil är den deskriptiva märkning-
en i TEI/XML-formatet. Detta utgör det viktigaste momentet och här skapas
den masterfil som sedan är utgångspunkt för bearbetning, tillämpningar och
presentationer av det prosopografiska materialet.

Migreringen från textformat till XML-format utförs helt enkelt genom att
läsa in textfilen i det editeringsprogram som skall användas för märkningen. Ett
program avsett för märkning av XML-filer bör innehålla ett minimum av hjälp-
funktioner, såsom funktioner för validering (relativt DTD:n eller schemat),
grammatisk validering (relativt XML:s syntax), automatisk utfyllnad av element,
avancerade sök- och ersättningsfunktioner, koppling till XSLT-parser samt möjlighe-
ter att visualisera och analysera dokumentstrukturen på olika sätt.⁶⁴⁷ Sist men
inte minst är det viktigt att editeringsprogrammet kan hantera den aktuella
DTD eller det schema som används i märkningsarbetet.⁶⁴⁸

Under arbetet med masterfilen har flera revideringar har ägt rum. Även in-
nehållet i den har ändrats några gånger i och med att nya kapitalbeskrivningar
tillkommit under hand. Med masterfilen som utgångspunkt har valda delar av
innehållet förflyttats till andra filformat och olika program och plattformar som
illustreras i Figur 27 nedan.

⁶⁴⁷ Det finns en rad olika editeringsprogram varav flera fritt tillgängliga, andra som har billiga
licenser och övriga som är mycket dyra. En sammanställning över verktyg för editering och
transformering av XML-filer TEI:s Wiki <www.tei-c.org/wiki/index.php/Category:Tools> se
även den gamla adressen: <www.tei-c.org.uk/Software/>. En annan webbplats för allt om
SGML/XML är Cover Pages (OASIS). Som är :”The original all inclusive overview of every-
thing” (citrat från TEI:s webb) <<http://xml.coverpages.org/sgml-xml.html>>.

⁶⁴⁸ Svensk handledning för nybörjare i XML-märkning, se <www.uu.se/monicalz/ffo/tools/>. Vid TEI:s webbplats finns guider och handledningar för att komma igång med XML-märkning,
se <www.tei-c.org/Support/Learn/index.xml>.



Figur 27. Migrering från TEI/XML masterfil till andra filformat och användningsområden

Migreringen till html-formatet har skett med hjälp av skriptspråket XSLT. Inläsning av data till kalkylbladet har därefter skett via de tabeller i html-format som skapades med hjälp av XSLT. Detta moment presenteras närmare i nästföljande stycke. Även för migrering till databasen (MySQL) har XSLT använts och då för extrahering av utvalda data från masterfil till textformat. Databashanteraren läser in sekventiellt ordnade data i textformat till tabeller och poster. De data som redan var lagrade i kalkylbladet konverterades där till textformat för inläsning i databasen.

Därtill har masterfilen varit utgångspunkt för utskrift av kapitalbeskrivningar i PDF-formatet. Detta utförs med hjälp av instruktioner i märkspråket XSL-FO, som används för beskrivning av instruktioner för formatering av dokumentens layout. För konstruktion av PDF-filerna har jag använt de XSL-FO filer som var inkorporerade med det editeringsprogram jag oftast använt.⁶⁴⁹ Förutsättningarna för att dessa prefabricerade stilmallar skall fungera är att dokumentmodellen är välformulerad och validerad. Instruktionerna i stilmallen kräver att innehållet i XML-filerna är välformulerade enligt TEI-schemat, om så inte är fallet kommer inte transformationen att fungera.⁶⁵⁰

En fördel är att jag genom användningen av ett vedertaget schema som TEI även får tillgång till verktyg, i detta fall en färdig stilmall i XSL-FO. Det är förstås fullt möjligt att definiera sitt eget schema, men då fungerar inte de prefabri-

⁶⁴⁹ Jag har framför allt använt oXygen en editings- och utvecklingsmiljö för XML och en rad andra webbt tekniker och skriptspråk. oXygen installeras med en rad olika verktyg och märknings-scheman, såsom TEI, DocBook mfl. Det finns gott om användarhandledningar och exempelsamlingar på webbplatsen och programmets mest grundläggande funktioner är relativt intuitiva även för den ovane. oXygen går att ladda ner och använda gratis under en månad, därefter kostar licensen en mindre summa, forskare, lärare och studenter får rabatt. <www.oxygenxml.com/>.

⁶⁵⁰ Vid andra tillfällen har jag skapat PDF-filer av ett annat XML-format, DocBook, men principen är densamma.

cerade verktyg som finns tillgängliga, utan det blir istället nödvändigt att skapa egna.

I de vanligaste kalkylprogrammen (*MS Excel*, *Open Office Calc*) går det alldeles utmärkt att importera html-tabeller.⁶⁵¹ Programmet läser in den önskade filen och sedan är det bara att spara den i det kalkylformat som programmet använder. *Open Office Calc* använder en öppen standard för filerna, *Open Document Format* (ODF) som baseras på XML, något som förenklar vidare överföring.⁶⁵² Calc hanterar dock flera proprietära filformat. Kalkylprogrammen är användbara för lagring och hantering av väl strukturerade data som kan ordnas i matriser och tabeller. I kalkylprogrammen finns också möjlighet att exportera i flera olika format såsom textfiler, xml, databashanterare, html etcetera.

Sammanfattningsvis konstaterar jag att i de flesta fall finns många möjligheter att överföra innehåll från ett filformat till ett annat. Programvarornas egna funktioner för import och export av olika filformat brukar räcka långt. Dessutom får fler och fler program inte bara stöd för XML utan allt oftare används XML-formatet som grund för programmets filformat.

6.4.2 Transformerings av innehåll med skriptspråket XSLT

Som nämnts kan prosopografin användas för statistiska analyser samt för presentationer av olika slag. För att göra detta behövs emellertid en ganska omfattande bearbetning. Verket XSLT som använts för att extrahera information är ett deklarativt skriptspråk avsett för transformering av XML.⁶⁵³ I själva transformeringen används märkningen och strukturen i XML-filen för att med stor precision kunna utföra bearbetningar av innehållsobjekten.

Den transformerade och bearbetade informationen migreras sedan till i en ny fil för import i de program som används för statistiska analyser.⁶⁵⁴ Den ursprungliga informationen kan oftast inte användas för analyserna i obearbetat skick. Exempelvis indikatorer för kvinnornas sociala ursprung. För detta ändamål används faderns yrke som ursprungligen har 200 unika förekomster i detta material. Denna information måste klassificeras för att bli användbar i analyserna.

⁶⁵¹ MS Excel är inte fritt tillgängligt, men mycket vanligt förekommande.

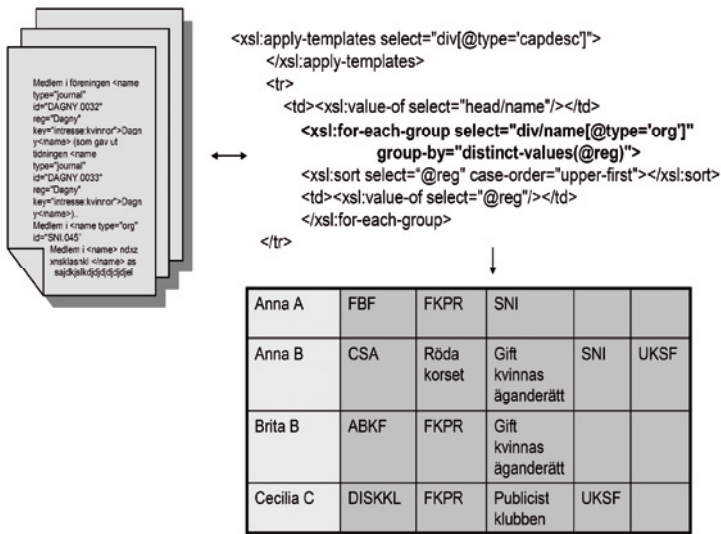
⁶⁵² MS Office använder också XML, dock inte *Open Document Format*, ODF, utan ett eget schema som kallas Open XML. ODF och Open XML är två konkurrerande scheman. Open XML ansökte hösten 2007 om att få bli en ISO-standard men fick avslag.

⁶⁵³ XSLT introduceras i avsnitt 3.5 s. 82 och behandlas inte närmare i avhandlingen pga dess komplexitet. För exempel på hur skriptspråket syntax och funktion se bifogade exempel i katalogen <CD:/verktyg/xml/>. För fullständig referens till skriptspråket XSLT, se Kay, *XSLT: programmer's reference*, Kay, *XSLT 2.0 programmer's reference*. För en bra introduktion till XSLt för nybörjare, se Jeni Tension, *Beginning XSLT 2.0: from novice to professional* (Berkeley, CA: Apress, 2005).

⁶⁵⁴ Den nyskapade filen kan vara av olika format, t.ex. text, XML eller HTML. Här används vanligtvis HTML eller XML för import till kalkylblad, samt till statistikprogrammen SPSS och SPAD för vidare bearbetning.

Korrespondensanalyser och deskriptiva statistiska analyser kan sedan utföras. I den aktuella studien utforskas olika sätt att göra nätverksanalyser och här experimenteras med olika sätt att illustrera kvinnornas kontaktnät i grafer från korrespondensanalyser. Det finns även möjligheter att kontrastera grupper av individer för att undersöka hur kvinnornas kontaktnät strukturerar sig. Exempel på dessa analyser presenteras i *Figur 33*, s. 211.⁶⁵⁵

Kvinnornas föreningsaktiviteter har registrerats och används sedan för analyser av det sociala rum där kvinnorna var verksamma. Tidigare presenterades märkning av kvinnornas föreningsaktiviteter.⁶⁵⁶ Nedan presenteras det tillvägagångssätt som använts för att extrahera och sortera information om kvinnornas föreningsaktiviteter. Som jag beskrivit förut har elementet <name> försetts med några attribut. Förutom att attributen hjälper till att beskriva innehållet är de absolut nödvändiga för transformering och bearbetning av information i XML-filen. I exemplet nedan används attributet type="org" som beskriver att innehållet betecknar ett namn på en organisation. Id-attributet (id="SNI.055") informerar om att detta är den 55:e förekomsten av Sällskapet Nya Idun i FFO-KAPITAL.⁶⁵⁷ Attributet reg (reg="Sällskapet Nya idun") anger den normaliserade stavningen eller benämningen av begreppet och i key (key="kvinnopol") lagras uppgifter om projektets klassificering av organisationen.



Figur 28. Extrahering av önskad information med XSLT, från masterfil till tabell

⁶⁵⁵ Här används olika kategorier såsom boendeort, huvudsakligt verksamhetsfält (kultur, pedagogik etc.) och de huvudsakliga grupperingarna (sfärer av kvinnorna i kapitalbeskrivningarna), pedagoger, publicister eller socialt verksamma, som samtliga var boende i Stockholm.

⁶⁵⁶ Se avsnitt 6.1.2.1 s. 178.

⁶⁵⁷ ID-attributen har numrerats med hjälp av DL- ANTE, se avsnitt 6.3 s. 187.

I figuren ovan visas hur verktyget XSLT används för att extrahera exakt den information som används för analyserna.⁶⁵⁸ I detta exempel används de element som beskriver kvinnornas föreningsaktiviteter. Informationen extraheras genom attributet `type` och dess attributvärde `"org"` (`xsl:for-each-group select="div/name[@type='org']"`).

Informationen ordnas sedan genom att använda den normaliserade stavningen i attributet `reg` (`group-by="distinct-values(@reg)"`) samt en funktion för alfabetisk ordning (`xsl:sort select="." case-order="upper-first"`).

Därefter placeras innehållet exakt där vi vill ha den, i detta fall i avsedd cell i en HTML-tabell. Denna tabell har sedan importerats till ett kalkylprogram där html-tabellen sparats i ett annat filformat. Därefter har jag i kalkylprogrammet bearbetat filen, såsom skapa tabellrubriker etcetera innan överföringen till statistikprogrammet SPSS.⁶⁵⁹ I SPSS vidtar sedan arbetet med att skapa omvandla textsträngar till numeriska variabler och konstruera nya, kombinerade och aggregerade, variabler.⁶⁶⁰ Efter importen till SPSS, och innan konstruktionen av nya variabler, görs ytterligare en kontroll av att rätt innehåll extraherats ur masterfilen. En bra och enkel metod för detta är skapa korstabeller för det som skall kontrolleras, då är det lätt att se om ett förväntat värde inte finns i den cell där det borde vara. I detta fall kontrolleras att rätt organisationsnamn har placerats tillsammans med rätt kvinna. I SPSS finns sofistikerade funktioner för omvandling av format för variabler. Exempelvis går det att automatiskt omvandla numeriska variabler från en variabel med värdet lagrat i textformat. Varje textsträng tilldelas då ett numeriskt värde och själva textsträngen används som en beskrivande etikett för det numeriska värdet. Likaså finns möjligheter att omvandla numeriska värden till textsträngar.

På liknande sätt kan valfritt innehåll i masterfilen bearbetas eller flyttas mellan olika tillämpningar och filformat, under förutsättning att de begrepp som eftersöks är märkta.

Dokumentstrukturen eller de data som skall märkas kan ibland vara så pass regelbunden att det går att automatisera den märkningen som jag gett exempel på. Möjligheter finns även att nyttja den textens struktur i form av stycken och mer implicita märkning, såsom radbrytningar, interpunktion etcetera för att automatisera konvertering av textfiler till XML-format. Ett exempel på detta tillvägagångssätt är en studie som utfördes på äldre avskrifter av medeltida drama på vers. Där användes dels den litterära genrens struktur, dels den ringa icke-hierarkiska COCOA-märkning som utgångspunkt för att skapa hierarkiska

⁶⁵⁸ Det XSLT-skript som använts här bifogas i `<CD:/xml/xsl/orgs.xml>`.

⁶⁵⁹ Innan dess utförs upprepade kontroller av att den extraherade informationen stämmer med intentionerna. För att detta moment inte skall ta alltför lång tid i anspråk har stickkontroller använts.

⁶⁶⁰ Kombinerade variabler skapas av värden ur två eller flera olika variabler, t.ex. en variabel för samtliga kvinnor födda på landsorten som utbildat sig på Kungl. Högre Lärarinneseminariet.

TEI/XML filer med nästlade element.⁶⁶¹ Drama på vers använder sig av ett enkelt men väl strukturerat format och passar bra som studieobjekt för den här typen av tester.⁶⁶² I likhet med denna studie utfördes migreringen av innehåll och format i olika steg med hjälp av skript skrivna i Perl och XSLT.⁶⁶³ Först gjordes textfilen om till en XML-fil försedd med några milstolpar baserat på den COCOA-märkning som redan fanns. Resultatet blev en XML-fil med platt struktur, eftersom elementen inte var nästlade utan sekventiellt ordnade i filen och markerade med tomma element.⁶⁶⁴ Därefter grupperades de element som var av samma sort och inordnades i en nästlad hierarkisk struktur.⁶⁶⁵ I och med att varje moment i migreringen består av fristående moduler och skript bör metoden kunna användas på texter av andra genrer och med andra typer av markörer än de som används i COCOA – förutsatt att de inte är för komplexa och oregelbundna i sin struktur. Därutöver krävs det en viss modifiering av de skript som använts för transformering.

Ett problem vid migrering är att resultatet påverkas av kvaliteten i den ursprungliga källfilen, eftersom förbättringar kan vara svåra att genomföra. I den här studien var till exempel texten skriven enbart med versaler, såsom brukligt med gamla elektroniska texter. För att arbetet med migrering av innehåll och format skall vara meningsfull krävs en noggrann analys av innehållet med avsikt att hitta de mönster och strukturer som kan användas för transformering. Detta är mycket tidsödande men helt nödvändigt. Automatiserad märkning är därför mer intressant för projekt med behov av att migrera och transformera större mängder texter och där det finns besparingar att göra.

6.5 Ordning och nyordning – klassificering av indikatorer

I detta avsnitt presenteras tillvägagångssätt för klassificering av egenskaper och tillgångar. Som indikatorer på socialt ursprung har fädernas yrkestitlar använts.⁶⁶⁶ Klassificering av sociala förhållanden kan vara problematiskt då en och

⁶⁶¹ James Cummings, "Liturgy, Drama, and the Archive: Three conversions from legacy formats to TEI XML", *Digital Medievalist* 2.1 (2006); Cummings & Langerth Zetterman, "Legacy Data Migration: A pilot study on the methodological feasibility of conversion and enhancement of electronic resources."

⁶⁶² I avsnittet om märkning med COCOA ges exempel på en av de texter som migrerades i den här studien. Se *Figur 9* s. 86 för COCOA-syntax.

⁶⁶³ Perl är ett skriptspråk utvecklat och ofta använt av lingvister. Perl är gratis, distribueras under GNU *General Public Licence* och kan användas på de flesta operativsystem. Se <www.perl.org/>.

⁶⁶⁴ Ett tomt element används som platshållare eller markör för beskrivning av sådant som inte är innehåll, t.ex. sidbrytningar eller andra milstolpar.

⁶⁶⁵ Cummings, "Liturgy, Drama, and the Archive: Three conversions from legacy formats to TEI XML". Se <www.digitalmedievalist.org/journal/2.1/cummings/#N1055A>. I artikeln hänvisas till de figurer och de skript som använts för migreringen.

⁶⁶⁶ I FFO-KAPITAL har fädernas högsta position registrerats, i FFO-LUNDSTROM har den information som finns i Lundströms biografiska uppslagsbok använts.

samma individ oftast tillhör ett flertal olika sociala sammanhang, sammanhang som dessutom förändras över tid.⁶⁶⁷

Det förekommer till exempel över 200 olika yrkestitlar för kvinnornas fäder. Denna information måste klassificeras och aggregeras för att bli brukbar i analyserna.⁶⁶⁸ Alltför många unika värden, *modaliteter*, i en och samma variabel är inte önskvärt då det försvårar såväl analys som tolkning. Alltför få värden är ej heller alltid önskvärt då analyserna istället blir för grova och därmed riskerar bli intetsägande.⁶⁶⁹

Kravet att samla in så många indikatorer som möjligt för samtliga individer i urvalet medför att de ursprungliga indikatorerna ibland behövt modifierats eller förenklats för att få en heltäckande insamling av de uppgifter som fanns tillgängliga för alla och envar. Som exempel kan nämnas att som mått på kvinnornas sociala ursprung – och eventuella klassförflyttning – användes initialt fädernas yrkeskarriär enligt vissa ”mätpunkter” i en kvinnas liv: När kvinnan föddes, när hon var 10 år, när hon flyttade hemifrån samt höjdpunkten i faderns yrkeskarriär. Men det visade sig att enhetliga data inte gick att uppbringa för alla dessa mätpunkter för många av kvinnorna i urvalet. Det fanns helt enkelt inte tillräckligt med uppgifter om fädernas yrkeskarriärer i källorna. Därför användes istället endast faderns höjdpunkt i yrkeskarriären som indikator för kvinnornas sociala ursprung.

Ett annat konkret problem är klassificering av kvinnornas yrkesverksamheter. Eftersom samtida klassificeringar utgår från männens fält och verksamhetsområden speglar de därmed sociala hierarkier med en manlig verksamhet som norm snarare än de verksamhetsområden kvinnorna på väg ut i offentligheten hade att välja på. Eftersom materialet handlar om förra sekelskiftet är det inte lämpligt att kategorisera yrkestitlar och verksamhet utifrån nutida sociala klassificeringar.⁶⁷⁰ En social klassificering bör spegla den historiska period som studeras då de sociala strukturerna och hierarkierna hela tiden förändras. I det här fallet handlar det rent konkret om att så många av kvinnorna tituleras enligt börd eller civilstånd trots att de var yrkesverksamma. Många av kvinnorna som var verksamma som oavlönade inom olika verksamhetsområden, främst med sociala- och bildningsfrågor, tituleras som filantroper eller socialt verksamma i kapitalbeskrivningarna. Med en modern klassificering skulle många i urvalet placeras i kategorin övriga, då de inte har någon egentlig yrkestitel, vilket i sig skulle göra en sådan kategorisering tämligen verkningslös i analyserna av kvin-

⁶⁶⁷ Stone, ”Prosopography” s. 121.

⁶⁶⁸ För fädernas sociala klassificering, se Lennart Thorsell, ”Den svenska parnassens 'demokratisering' och de folkliga bildningsvägarna” *Samlaren* (Uppsala, 1958).

⁶⁶⁹ Undantaget är booleska variabler, det vill säga sådana variabler som registrerar en egenskap som man antingen har eller inte har. Exempelvis variabler för medlemskap i föreningar: kvinnan var antingen medlem eller inte.

⁶⁷⁰ För en historisk genomgång av social klassificering i Sverige, se Mikael Börjesson, ”Om att klassificera sociala grupper” *Fältanteckningar: utbildnings- och kultursociologiska texter tillägnade Donald Broady*, red. Mikael Börjesson, et al., Skeptronserien (Uppsala universitet: Forskningsgruppen för utbildnings- och kultursociologi ILU Uppsala universitet, 2006).

nornas egenskaper. För kvinnornas den aktuella studien är det därmed lämpligare att tala om huvudsaklig verksamhet snarare än yrkestitel.

Kvinnornas position var vid den här tiden nära relaterad till makens position alternativt till faderns sociala position.⁶⁷¹ Kvinnornas sociala positioner i detta material, med utgångspunkt i titlar och verksamhetsområden, skulle möjligen kunna ses som ett skuggfält av sociala positioner med egna hierarkier och logiker. Eller så är det så att man i projektet blandar äpplen med päron genom att jämföra titlar och yrkeskategorier med sekundära sociala positioner baserat på nedärvda tillgångar och giftermålsstrategier.

En meningsfull, och tillika funktionell, klassificering – för såväl det aktuella materialet som för den tidsperiod som studeras – är essentiellt för konstruktionen av de kvinnokollektiv som studeras. Därför har klassificeringen av kvinnornas titlar, som oftast motsvarar deras yrkesverksamhet, använt en modifierad version av den sociala klassificering Thorsell använder i sin studie om författare bildningsvägar och sociala härkomst.⁶⁷² Därutöver har en projektspecifik klassificering utformats med egna lokala kategorier som avser spegla den aktuella tidsperioden såväl som det aktuella urvalet.

I den lokala klassificeringen differentieras till exempel publicister från övriga kvinnor verksamma inom kultursfären, skolgrundare från andra chefsbefattningar, samt socialt verksamma kvinnor från kvinnor som tituleras enligt börd eller giftermål. För exemplet med fädernas titlar har en samtida klassificering med åtta kategorier för differentiering av sociala skikt använts.⁶⁷³

Intellektuell överklass: ämbets- och tjänstemän, fria yrken med akademisk examen som utbildning samt officerare, läkare, advokater, litteratörer

Tjänstemannaklass: studentexamen eller jämförligt, men ej akademisk utbildning. Tillhörande social grupp I samt några kategorier ur socialgrupp II, bl.a. folkskolelärare

Godsägarklass: godsägare och förvaltare, bruksägare och disponenter.

Högborgerlig över- och medelklass: affärsmän, större företagare och andra näringsidkare inom borgerliga stadsnärningar

Bondeklass: självägande bönder och arrendatorer

Lägre, småborgerlig/hantverkare/lägre tjänstemannaklass: hantverksmästare, småhandlare, verkstämmästare, kontorister, underofficerare, underbefäl, vaktmästare o.a. tjänstemän av lägre grad i allmän tjänst

Arbetarklass: arbetare av alla kategorier från typografer till grovarbetare och sjömän. Dock ej jordbruksarbetare

Jordbrukets underklass: torpare, backstugesittare, statare och drängar.

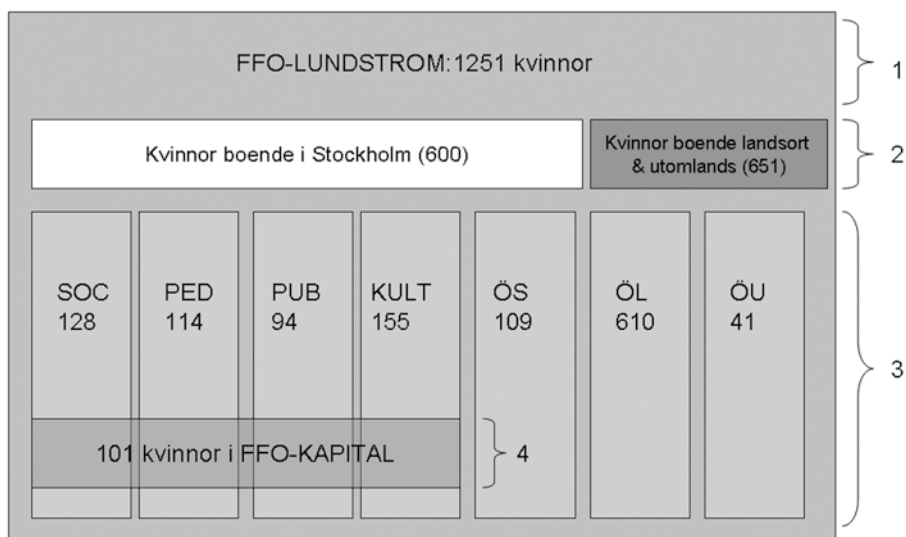
⁶⁷¹ För en diskussion om sociala positioner, parallella måttstockar, ståndstillhörighet och yrkestitlar, se Esbjörn Larsson, *Att studera titlar som kapital: en teoretisk diskussion kring militära titlar i 1700-talets svenska samhälle* (2003); Larsson, *Från adlig uppfostran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866* s. 57-62.

⁶⁷² Thorsell, "Den svenska parnassens 'demokratisering' och de folkliga bildningsvägarna".

⁶⁷³ Ibid. s. 85-102.

Valet av Thorsells principer för social klassificering kan motiveras av att den tidsperiod han studerar (1880-1950) överensstämmer med tidsperioden för projektet Formering för offentlighet (1880-1920).⁶⁷⁴ Thorsell kombinerar i sin klassificering socioekonomiska faktorer med "sociala bildningstyper" som kategoriserar utbildningsnivå, inspirerad av Cederblads bok.⁶⁷⁵ Vad gäller klassificeringen av utbildning för FFO-databasen valdes att göra ungefär på samma sätt som med kvinnornas utbildning. Dels användes de "sociala bildningstyper"⁶⁷⁶ som Thorsell använder, dels användes en lokal modifiering av denna kategorisering, för att särskilja studier på till exempel flickskolor från läroverksstudier.⁶⁷⁷ Återigen var det nödvändigt att modifiera en klassificering som grundade sig på männens villkor, i detta fall utbildning, för att bättre spegla kvinnornas villkor.

Figur 29 illustrerar hur registret har delats in i ett antal delregister. En indelning som har gjorts enligt två principer; kvinnans boendeort och huvudsakliga verksamhetsområde (sfär).



Figur 29. FFO-LUNDSTROM registrets 4 nivåer.

Anm: Socialt verksamma (SOC), Pedagoger (PED), Publicister (PUB), Kvinnor inom den övriga kultursfären i Stockholm (KULT), Övriga kvinnor boende i Stockholm ej tillhörande någon av kategorierna ovan (ÖS), Övriga kvinnor på landsbygden (ÖL), Övriga kvinnor boende utomlands (ÖU)

⁶⁷⁴ Thorsell studerar författares (män och kvinnor) sociala rekrytering och bildningsvägar under åren 1880-1950.

⁶⁷⁵ Carl Cederblad, *Bildningens väg: idéerna och livet* (Uppsala: Lindblad, 1932).

⁶⁷⁶ De sociala bildningstyperna presenteras i: Ibid. s. 166-188; Linné. "Lärarinnor, pedagogiskt arbete och offentlighet. Kvinnliga pionjärs liv och strategier kring år 1900 - ett jämförande perspektiv".

⁶⁷⁷ Se vidare avsnitt 6.5 s. 196.

FFO-LUNDSTROM består av 4 nivåer som var och en kan användas som en fristående modul. Den första nivån omfattar hela, det vill säga samtliga kvinnor som omnämns i Lundströms biografiska uppslagsverk samt kvinnorna i kapitalbeskrivningarna. Nivå 2 är en uppdelning mellan kvinnor boende i Stockholm och kvinnor boende på landsorten eller utomlands. Nivå 3 differentieras av olika verksamhetsområden, eller sfärer. Var och en av dessa sfärer är i sin tur uppdelad efter boendeort. Nivå 4, FFO-KAPITAL, som bildar en vertikal modul i nivå 3 är den mest detaljerade och informationsrika och innehåller de 101 kvinnor som förekommer i kapitalbeskrivningarna.

Kvinnor boende i Stockholm, oavsett sfär, utgör nästan hälften av det totala antalet kvinnor i FFO-LUNDSTROM. I Stockholm finns dels en kultursfär och dels en sfär bestående av publicister. I kategorierna ”övriga” i Stockholm och på landsort ingår kvinnor som var verksamma inom teknik, medicin, juridik, politik med mera. I FFO-LUNDSTROM finns inte födelse eller boendeort registrerat, annat än ”Landsort” eller ”Utland”, för kvinnor som inte var födda eller boende i Stockholm. Detta medför att analyser av till exempel den geografiska spridningen eller analyser av flyttströmmar inte är möjlig annat än genom dikotomin Stockholm – landsort.

Kvinnornas föreningsaktiviteter har också varit föremål för kategorisering och här används en lokal klassificering för beskrivning av de något mer än 120 föreningar, sammanslutningar och organisationer som kvinnorna i FFO-KAPITAL varit aktiva inom.⁶⁷⁸ Tanken med att registrera föreningsaktiviteter var att studera kvinnornas nätverk och positioner inom olika verksamhetsområden. Förutom själva föreningsaktiviteten har uppgifter samlats in om kvinnans position i föreningen eller organisationen. Det är en väsentlig skillnad att vara ”vanlig” medlem jämfört med att vara ordförande eller stiftarinna.

I avsnittet har jag diskuterat exempel på hur den nedtecknade informationen i den biografiska uppslagsboken och i kapitalbeskrivningarna klassificerats och ordnats.

⁶⁷⁸ Förteckning över föreningar och föreningstyper i bilaga 6, se <CD:/bilagor/bilaga6.pdf>.

7 Tillämpningar av det prosopografiska materialet

En stor del av avhandlingen har hittills behandlat förutsättningar för tillämpningar av prosopografiska digitala material. Nu är det dags att visa vad som är möjligt att göra med det digitala material som iordningställt och utformats. I detta kapitel presenteras de tillämpningar som utformats i den här studien.

Klassificeringen av de olika egenskaper och tillgångar som individerna i det här materialet besitter är användbara i flera tillämpningar, för den beskrivande statistiken såväl som för korrespondensanalyser och nätverksanalyser. I webbarkivet har bland annat sfär, socialt ursprung och utbildning används som ”ingångar” till materialet och som organiserande princip för databasens moduler.

I kapitlet används inte skärmdumpar i någon stor utsträckning för beskrivning av webbarkivet.⁶⁷⁹ Dels för att vissa av tillämpningarna bifogas i avhandlingens CD:n och bäst granskas i webbläsaren, dels för att dessa tillämpningar finns tillgängliga på, och är konstruerade för, webben snarare än för pappersutskrifter.⁶⁸⁰ Därför inriktas detta kapitel så långt som möjligt på att visa principer och tillvägagångssätt för de tillämpningar som här presenteras.

7.1 Beskrivande statistik

En stor del av användningen av det prosopografiska materialet har hittills varit deskriptiva statistiska analyser. Som ovan beskrivits har de data som registrerats och transkriberats utformats för olika slag av aggregat och kombinerade variabler. Med utgångspunkt i de data som samlats in och förtecknats har över 200 olika variabler skapats.⁶⁸¹ I och med att den ursprungliga informationen oftast

⁶⁷⁹ Skärmdump (*screenshot*) är en digital bild som avbildar hela, eller delar av bildskärmen vid ett givet tillfälle. Undantaget här är exempel på databasens webbgränssnitt då den inte kan bifogas CD:n eftersom databasen kräver programvara som är installerat på webbservern.

⁶⁸⁰ Vissa tillämpningar kräver programvara som körs på datorn/servern, t.ex. databasen och de interaktiva kartorna. För samtliga tillämpningar som presenteras i kapitlet ges hänvisningar till den bifogade CD:n samt till webben. Allt material som finns tillgängligt på webben nås via <www.skeptron.uu.se/monicalz/ffo/>

⁶⁸¹ De 239 variablerna listas i bilaga 12, se <CD:/bilagor/bilaga12.pdf>.

är på en mycket detaljerad nivå finns även möjligheter till senare aggregering och kombinerings av olika data.⁶⁸²

Variablerna kan delas in i ett antal grupper relativt karaktär, eller snarare typ av indikator. Följande variabelgrupper finns för närvarande representerade i registret FFO-LUNDSTROM:

- Personliga data (namn, födelseår, generation, födda per decennium)
- Geografiska data (födelsort, boendeort, flyttströmmar)
- Socialt ursprung (faderns angivna yrkestitel, faderns sociala klass)
- Kvinnans verksamhetsområde (namngiven titel, typ av verksamhet, sfär, sfär-boende)
- Civilstånd (gift, ev. makes yrkestitel, makes sociala klass)
- Utbildningsvariabler (utbildad typ av utbildning, utbildad lärarinneseminariet, utlandsvistelse)
- Specifika tillgångar (publicering, utgivningstyper⁶⁸³, uppdrag, utmärkelser)

Registret FFO-KAPITAL innehåller ytterligare variabler som baseras på uppgifter extraherade ur kapitalbeskrivningarna. Dessa variabelgrupperingar består av:

- Sociala nätverk (kontakter med namngiven annan person, antal kontakter, antal omnämnanden i andra kapitalbeskrivningar, antal nämnda kontakter i den egna kapitalbeskrivningen)
- Föreningsaktiviteter (aktiv i namngiven förening, aktiv i antal föreningar, aktiv i typ av förening, aktiv i antal föreningstyper, styrelseledamot i namngiven förening, årtal)
- Utmärkelser (namn, typ, antal utmärkelser, årtal)
- Uppdrag (namn, typ, antal, årtal)
- Övrigt (indexerat värde bouppteckning, årtal)

I FFO-HLS har följande variabler registrerats:

- Persondata (namn, födelsedatum och ort)
- Social bakgrund (faderns yrke)
- Tidigare utbildning och anställning
- Studiegång vid seminariet (studieår, valfria, obligatoriska, övnings- och fackämnen)

⁶⁸² Med vissa undantag för FFO-LUNDSTROM, se t.ex. s. 164 f.

⁶⁸³ Uppgifter om genrer finns endast för kvinnor som bodde i Stockholm.

- Studier utomlands
- Tiden efter seminariet (anställning och övriga levnadsförhållanden)

De variabler som visat sig vara speciellt användbara för andra slags analyser än beskrivande statistik är kontakt- och föreningsvariablerna, dels med tanke på typen av material, dels för forskningsfrågorna som formulerades i projektet. Dessa variabler har använts i flera olika typer av analyser, både för beskrivande statistik, nätverksanalyser och korrespondensanalyser.

7.1.1 Något om de offentliga kvinnorna

I syfte att ge en inblick i innehållet i den prosopografiska databasen samt belysa vad FFO-databasen kan erbjuda för den beskrivande statistiken presenteras nedan en översiktlig framställning om fördelningar och samband mellan några indikatorer för kvinnornas egenskaper och tillgångar.⁶⁸⁴ Vad gäller källornas tillförlitlighet baseras, som tidigare nämnts, de avskrivna uppgifterna i Lundströms på uppgifter som kvinnorna själva rapporterat in, med undantag för medlemskap och styrelseuppdrag i vissa föreningar, där informationen har kontrollerats utifrån medlemsmatriklar.⁶⁸⁵

Nästan hälften av kvinnorna bodde i Stockholm (48 procent) och en lika stor andel på landsorten (49 procent) medan bara en mindre andel var bosatta utomlands (3 procent).⁶⁸⁶ Vad gäller kvinnornas yrken och verksamhetsområden titulerade sig var tredje kvinna antingen via giftermål eller börd (friherinna, doktorinna etcetera) eller som ”fröken” (39 procent). Valet av titeln fröken istället för yrkestitel tolkas på olika sätt. Å ena sidan var det vid den här tidsperioden kvinnor från medel- och överklass som använde benämningen ”fröken” och markerade genom det en kvinnas sociala klasstillhörighet. Å andra sidan bör det även betänkas att uppgifterna i Lundströms biografiska uppslagsbok inlämnades av kvinnorna själva. Kanske titulerade vissa av kvinnorna sig som fröken, trots att de var yrkesverksamma, för att de inte ville framhäva sig själ-

⁶⁸⁴ Presentationen av kvinnorna avser FFO-LUNDSTROM (N=1251) om inget annat anges. I annat fall avses modulen FFO-KAPITAL (n=101). För samtliga beskrivningar som presenteras i detta avsnitt samt för fler analyser hänvisas till rapporten: Zetterman, *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet*.

⁶⁸⁵ På grund av självbiografiers begränsade källvärde är det extra viktigt att kombinera olika typer av källor. Larsson, *Från adlig uppfostran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866* s. 66. De medlemsmatriklar som kontrollerats är Sällskapet Nya Idun (publ.), Fredrika Bremerförbundet och Föreningen för kvinnans eganderätt (avskrifter av publ.), se referenser s. 265 ff.

⁶⁸⁶ Vissa kategorier är i framställningen ovan hopslagna i syfte att ge en enkel översikt. Till exempel är de kulturellt verksamma kvinnorna i databasen uppdelade i två kategorier som tillsammans utgör 21,2 procent; kulturell produktion (författare) samt kulturförmedling (teater, opera etc.).

va?⁶⁸⁷ Samtidens förväntningar på kvinnors ställning i samhället kan ha bidragit till valet av titulering.

De kulturellt verksamma kvinnorna, i vid bemärkelse, är den näst vanligaste kategorin i FFO-LUNDSTROM, där de utgör drygt 20 procent. Efter dem är de största grupperna lärare och intellektuellt fria yrken (läkare och advokater)som vardera utgör 9 procent av kvinnorna. En annan ganska stor grupp kvinnor var verksamma som föreståndare eller studierektorer (7 procent). Kvinnornas utbildningsnivå var hög, var fjärde kvinna hade med den tidens mått en högre utbildning: antingen en akademisk examen (12 procent) eller lärarinneexamen (14 procent), flertalet utbildade vid Kungl. Högre Lärarinneseminariet.⁶⁸⁸ Många hade också yrkes- och hantverksutbildningar (14 procent) eller olika former av konstnärliga utbildningar (13 procent). En mindre andel av kvinnorna hade folkskola som högsta utbildningsnivå (3 procent).

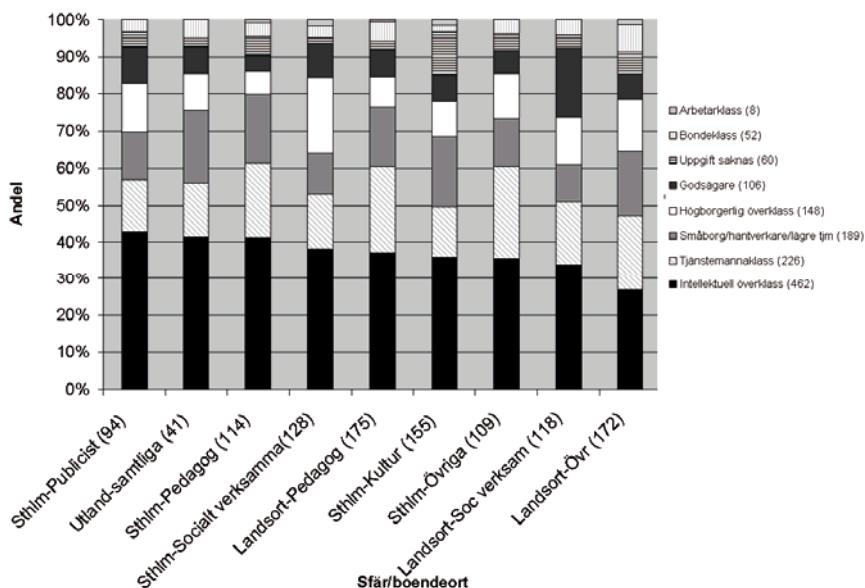
Åldersfördelning bland kvinnorna är relativt jämnt fördelad mellan 30 till 50-åringar; en fjärdedel av kvinnorna var i mellan 43-52 år, en tredjedel mellan 33-42, en femtedel i mellan 23-32 år vid tiden för bokens tillkomst 1922.⁶⁸⁹ Färre än hälften i de grupperna var eller hade varit gifta (48 procent). Detta är en betydligt mindre andel jämfört med övriga kvinnor i Sverige, i samma ålderskategori (30-60 år), där mer än 80 procent var eller hade varit gifta 1920.⁶⁹⁰ En ännu mindre andel, 38 procent, av kvinnorna i FFO-KAPITAL var gifta. Sett till fördelningen inom gruppen FFO-KAPITAL var pedagogerna de som i störst utsträckning var ogifta, medan publicisterna var de kvinnor som i störst utsträckning gifte sig.

⁶⁸⁷ ”Det har uppenbarligen inte känts naturligt för dessa [kvinnor] att tala i egen sak. De har underordnat sig en rörelse, kanske inte i första hand av osjälviska skäl utan därför att tidens könsroller gjorde det svårt för kvinnor att framträda offentligt.” Ambjörnsson & Sörlin, *Obemärkta: det dagliga livets idéer* s. 10.

⁶⁸⁸ Hit räknas även ”Lärokursen för fruntimmer” som gavs 1859-1861, det vill säga åren innan Kungl. Högre Lärarinneseminariet startade sin verksamhet.

⁶⁸⁹ För kvinnor födda 1865 var medellivslängden 50 år för kvinnor och 47,5 år för män. År 1922 var medellivslängden för kvinnor ca 60 år, för män något lägre. Sverige. Statistiska centralbyrån, *Från folkebrist till en åldrande befolkning: glimtar ur en unik befolkningsstatistik under 250 år: fakta inför 2000-talet* (Stockholm: Statistiska centralbyrån (SCB), 1999) s. 48-49.

⁶⁹⁰ Ibid. s. 68.



Figur 30. Kvinnornas sociala ursprung; jämfört med boendeort och sfär. Procent (N=1251).

I Figur 30 kan vi se att något mindre än hälften av de i Stockholm boende pedagogerna och publicisterna (drygt 40 procent) kom från intellektuell överklass.⁶⁹¹ Ungefär lika många (38 procent) av de socialt verksamma kvinnorna hade fäder hemmahörande i den intellektuella överklassen. De socialt verksamma kvinnorna som bodde i Stockholm har den största andelen med ursprung i högborgerlig över och medelklass.⁶⁹² Kvinnor inom pedagogiska och den kulturella sfären boende i Stockholm kom även i ganska stor utsträckning från småborgerliga hem.

Vad gäller sambandet giftermål och social klass är det inga direkt oväntade resultat. Kvinnorna gifte sig hög i utsträckning inom eller över sin egen sociala klass. Kvinnorna från intellektuell överklass gifte sig oftast med män från samma klass. Kvinnorna från tjänstemannaklassen gifte sig även dem med män från den intellektuella överklassen. Mer precist visar beräkningarna att kvinnor inom kultursfärerna i Stockholm och på landsorten gifte sig i mycket hög utsträckning med män från den intellektuella överklassen. De var de socialt verksamma kvinnorna från intellektuell överklass boende på landsorten, men även kvinnor i kultursfären, som oftast gifte in sig i högborgerliga miljöer. Sammanfattningsvis

⁶⁹¹ Inom parentes anges antal kvinnor per sfär samt antal kvinnor per social klass.

⁶⁹² För klassificeringsprinciper för kvinnornas sociala ursprung, se 6.5 s. 196 f.

kan vi se att, oavsett kvinnornas sociala ursprung, gifter de sig oftast med män från den intellektuella överklassen och näst oftast med tjänstemän.⁶⁹³

Kvinnorna var mycket aktiva inom ett flertal föreningar och sammanslutningar, för de 101 kvinnorna i FFO-KAPITAL finns 335 unika förekomster av 113 olika föreningar registrerade. Det betyder att varje kvinna i genomsnitt var medlem i 3,3 föreningar. Vilka föreningar var då kvinnorna med i?

De vanligaste föreningarna var Sällskapet Nya Idun, Fredrika Bremerförbundet, Landsföreningen för kvinnans politiska rösträtt, Uppsala kvinnliga studentförening samt Akademiskt bildade kvinnors förening. Spännvidden var emellertid enorm och föreningslivet präglades av frågor inom allt från rösträtt, sjukvård, socialt arbete, bildning för kvinnor, dräktreformer till yrkesorganisationer, fredsföreningar och läsesalonger. Det som särskiljer kvinnorna är vilken typ av föreningar de var aktiva inom. De som var aktiva med sociala frågor, freds- och kvinnopolitiska frågor är också de som var med i ett större antal föreningar än de andra kvinnorna. Dessa kvinnor tillhörde även de som i störst utsträckning hade många kontakter med övriga kvinnor i urvalet. De som var med i sällskap där man framför allt ägnade sig åt intellektuell samvaro engagerade sig inte alls i föreningar som hade samhällsfrågor på agendan. En grupp som skiljer sig åt från övriga är de kvinnor som engagerade sig kvinnopolitiskt men inte alls i sociala frågor eller fredsfrågor.

Sammanfattningsvis kan jag konstatera att de kvinnor som engagerade sig i samhällsfrågor även var mycket engagerade inom flera andra typer av föreningar och hade dessutom, eller som en följd därav, ett stort socialt kontaktnät. De som inte engagerade sig i samhällsfrågor var aktiva inom föreningar med intellektuell och yrkesrelaterad karaktär. För de kvinnopolitiskt intresserade kvinnorna i detta urval var vissa specifika samhällsfrågor som till exempel rösträttsfrågor vanliga.⁶⁹⁴

7.1.2 Något om lärarinneleverna

Kungl. Högre Lärarinneseminariet hade en viktig roll för utbildning av lärarinnor och skolföreståndare men även för andra yrkesområden. Flera av lärarinneleverna gjorde karriär inom andra områden, till exempel Anna-Maria Roos arbetade till exempel aldrig som lärarinna och enligt Fevrell är den troliga orsaken till det att hennes yrkesval inte var ”förenligt med familjens höga sociala ställning”.⁶⁹⁵

⁶⁹³ De deskriptiva analyser som ligger till grund för påståenden om giftermålsstrategier se bilaga 13 vid CD:/bilagor/bilaga13.pdf.

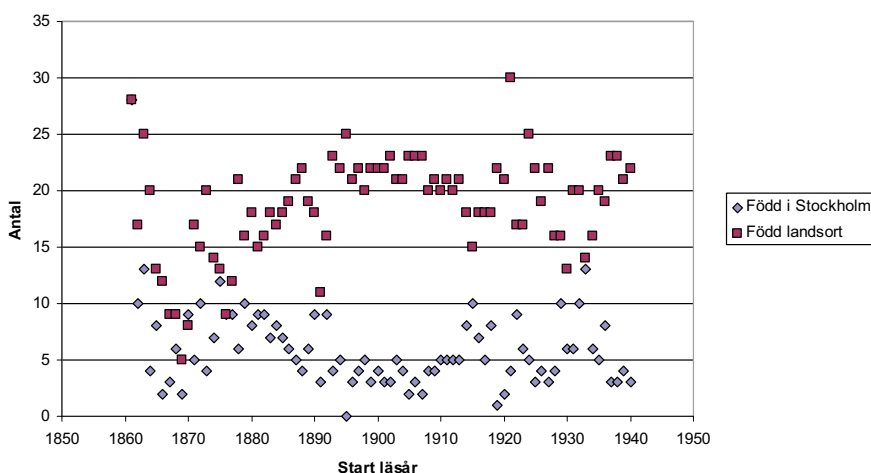
⁶⁹⁴ De tabeller, diagram och klusteranalyser som slutsatser om kvinnornas föreningsaktiviteter baserar sig på redovisas och diskuteras i Monica Langerth Zetterman. ”Människor av god vilja! - om utbildningens betydelse för kvinnor kring sekelskiftet 1900”. *Tredje pedagogikhistoriska konferensen*, 28-29 september, Lärarhögskolan i Stockholm, 2006; Zetterman, *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet*. Den förra innehåller flera analyser om föreningsaktiviteter än den senare.

⁶⁹⁵ Fevrell, *Kungl. Högre lärarinneseminariet in memoriam: minnesrunor* s. 301.

I likhet med den andra modulen i FFO-databasen (FFO-LUNDSTROM) har uppgifter samlass in om kvinnornas sociala bakgrund, tidigare utbildning, verksamhet vid seminariet, samt anställningar och verksamhetsområde efter Kungl. Högre Lärarinneseminariet.

Totalt 1980 kvinnor var inskrivna vid Högre Lärarinneseminariet. Endast 4 procent avbröt utbildningen på grund orsaker som giftermål, sjukdom, dödsfall eller annan utbildning (exempel sjuksköterskestudier). De allra flesta, 88 procent, tog examen ”Med Diplom” och endast 1 procent slutade utbildningen utan examen. Flertalet av de studerande slutförde utbildningen inom utsatt tid, endast ett fåtal var eftersläntrare.

I matrikeln finns bland annat födelseuppgifter registrerade och i figuren nedan kan vi se att rekryteringen av lärarinnestuderande framför allt skedde från övriga delar av landet än Stockholm.⁶⁹⁶ I början av seminariets verksamhetsperiod var rekryteringen från Stockholm högre än runt sekelskiftet.



Figur 31. Lärarinneelever vid Kungl. Högre Lärarinneseminariet; årsvis jämförelse av födelseorter

De blivande lärarinnorna var unga, sett ur vårt tidsperspektiv, när de påbörjade sina studier, de allra flesta var mellan 18-22 år (1493 kvinnor) men det finns även några 14-16-åringar i registret.

Vad det blev av dessa lärarinneelever efter examen går delvis att spåra med utgångspunkt i det material som för närvarande finns tillgängligt i FFO-databasen. För de närmsta åren efter deras examen finns uppgifter om anställning och yrkesverksamhet för perioden 1914–1943. Dessutom finns tillgång till

⁶⁹⁶ Orterna som ingår i ”Landsort”, i FFO-HLS är registrerade såsom orten benämns i arkivmaterialet (städer, byar, län etc.). Ortnamnen är alltså inte normaliserade eller på förhand klassificerade som de biografiska uppgifterna i FFO-LUNDSTROM.

uppgifter för 116 kvinnor i FFO-LUNDSTROM som tog sin examen vid Kungl. Högre Lärarinneseminariet. Dessa kvinnor kan därmed följas från inskrivning till deras yrkesverksamma liv efter examen.

Flertalet av dessa 116 kvinnor hade anställning av något slag (94 procent). De flesta arbetade som föreståndare på privata flickskolor och läroverk (40 kvinnor), som lärarinnor (14 kvinnor) eller som författare (10 kvinnor). Tre kvinnor blev rektorer och lika många var skolgrundare.⁶⁹⁷ Tretton kvinnor använde sig av sin ärvda eller genom giftermål förvärvade titel och sjutton kvinnor titulerar sig som fröken.⁶⁹⁸ Om dessa kvinnors verksamhet efter lärarinneexamen är det svårt att sja, men troligen var det så att åtminstone några av dem inte yrkesarbetade efter examen.

7.2 Nätverksanalys

Utöver rena frekvensbeskrivningar och korstabeller finns, inom prosopografin, med sina generaliserande ambitioner, även behov av metoder för analyser av nivån mellan individen och kollektivet. En sådan metod är nätverksanalys, en annan är korrespondensanalysen.⁶⁹⁹ Nätverksanalysen ger möjligheter till att studera individers relationer men även strukturen på nätverket:

[---]sociala nätverk som en beskrivning av relationer mellan individer eller grupper (organisationer) i termer av noder sammankopplade med länkar ett gott heuristiskt eller explorativt begrepp.⁷⁰⁰

Metodens främsta användningsområde är därmed inte att förklara, fastslå eller klassificera, utan att genom det explorativa arbetssättet ge möjlighet till att se strukturer, skeenden och förhållanden ur nya perspektiv.

Nätverksanalys av sociala system har utvecklats med bakgrund i tre olika forskningsfält. Det första utvecklades i USA under 1930-talet inom socialpsykologi, där man med sociometriska metoder och grafteori analyserade sociala gruppers dynamik. Det andra området var socialantropologiska studier av informella relationer. Den tredje grenen var socialantropologer vid universitetet i Manchester som studerade konflikter i sociala system.

Nätverksanalysens formaliserande och kvantitativa framställningar är ett kännetecken och används ofta för studier av olika sociala skikt, exempelvis

⁶⁹⁷ Här ingår Sofi Almqvist och Anna Sandström, se Ullman, *Stiftarinnegenerationen: Sofi Almqvist, Anna Sandström, Anna Ahlström*.

⁶⁹⁸ Se diskussion om fröknar och andra titlar på s. 153.

⁶⁹⁹ För exempel på korrespondensanalys av materialet i FFO-databasen, se Zetterman, *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet*.

⁷⁰⁰ Peter Aronsson, Solveig Fagerlund & Jan Samuelson, *Nätverk i historisk forskning: metafor, metod eller teori?* Rapporter från Växjö universitet. Humaniora, 1 (Växjö: Univ., 1999) s. 2.

elitors nätverk, används även för demografiskt inriktade studier av exempelvis nätverk i lokalsamhällen.⁷⁰¹ Exempel på historiska studier där de sociala nätverkens betydelse har studerats är en studie om kvinnors nätverk och deras betydelse för att driva gemensamma frågor som till exempel fredsfrågan.⁷⁰²

De metoder som används för att studera nätverksanalys kan i korthet beskrivas i följande steg. För det första studeras nätverket som helhet, snarare än de individer som ingår. Hur ser nätverket ut? Hur formerar sig noderna? Sedan används metoder för att beskriva en individs relation till övriga i nätverket som i exemplet med hur Lydia Wahlström utmärker sig i förhållande till de övriga pedagogerna. Slutligen finns det metoder för att studera nätverkets:

- *täthet*, det vill säga vilken densitet nätverket har beror på antalet noder;
- *likhet* hur enskilda noder korrelerar med varandra;
- *centralitet* för undersökning av nodernas placering centralt eller lokalt relativt nätverk;
- *positioner* undersöks genom att leta individer som har liknande egenskaper som sina grannar;
- *undergrupper* där olika metoder används för att identifiera kluster eller andra typer av gemenskaper.⁷⁰³

Syftet med avsnittet är att visa hur ett prosopografiskt material kan användas för nätverksanalyser. Det finns en rad olika programvaror att välja mellan varav långt ifrån alla är fritt tillgängliga. Det program jag har använt valdes för att det hanterade flera filformat såväl kalkylblad som XML-format, dessutom var det lätt att använda.⁷⁰⁴ Utgångspunkten för de första två exemplen i detta avsnitt är de kontaktnätverk som finns registrerade för kvinnorna i FFO-KAPITAL. Dessa kontaktnät beskrev jag i XML med hjälp av XML-schemat GraphML för att programmet kunde använda detta format.⁷⁰⁵ Programmet använder även kalkylblad för värdena som är ordnade i matriser.⁷⁰⁶

⁷⁰¹ Ibid. s. 2-11. För introduktion till nätverksanalys se även Einar Hreinsson, Tomas Nilson & Lars Hermanson, *Nätverk som social resurs: historiska exempel* (Lund: Studentlitteratur, 2003).

⁷⁰² Irene Andersson, *Kvinnor mot krig: aktioner och nätverk för fred 1914-1940*, Studia historica Lundensia, 1 (Lund: 2001).

⁷⁰³ För en mer ingående beskrivning av metoder och tillvägagångssätt för nätverksanalyser se Aronsson, Fagerlund & Samuelson, *Nätverk i historisk forskning: metafor, metod eller teori?* ssk. s. 45-59.

⁷⁰⁴ Graferna i *Figur 32* och *Figur 33* har skapats och presenteras med applikation VisuaLyzer, se <www.mdlogix.com/> för information och nedladdning av demoversion av programmet. Bilder av nätverksgrafer bifogas i <CD:/figurer/figur25.jpg> samt CD:/figurer/ figur26.jpg>.

⁷⁰⁵ Programmet använder formatet graphML, en tillämpning av XML med egen DTD, i likhet med TEI. XML-filen <ffo-net.graphml> innehåller alla de relationer och attribut, som behövs för att göra samma typ av nätverksanalyser som i detta avsnitt, se <CD:/xml/ffo-net.graphml>.

⁷⁰⁶ Kalkylbladet inklusive alla data som behövs för att göra samma sorts analyser som här bifogas i medföljande CD vid <CD:/docs/ffo-net.xls>.

Konstruktionen av graferna, noderna och de riktade länkarna baseras på binnära relationer. Varje nod representerar en individ och har också ett antal attribut som anger t ex social bakgrund, utbildning, föreningsaktiviteter etcetera.

Informationen som används i nätverksanalyserna har i likhet med de register som skapats för de statistiska analyserna utvunnits från masterfilen i XML-formatet. På liknande sätt och med samma sorts verktyg har kvinnornas kontakter extraherats ur startdokumentet för att infogas i måldokumentet. De attributvärden som överförts till måldokumentet kan användas för filtrering av presentationerna, för analyser och för att formulera sökfrågor.

I båda graferna nedan visas samtliga 101 kvinnors 450 kontakter med kvinnor inom gruppen, men på olika sätt.



Figur 32. Nätverksanalys av 101 kvinnors interna kontakter i FFO-KAPITAL

Grafen ovan är ett exempel på hur kvinnornas sociala kontakter kan visualiseras genom nätverksanalys.⁷⁰⁷ För att tillföra ytterligare information, utöver nätverksstrukturen, har jag använt symboler för att markera social bakgrund (fa-

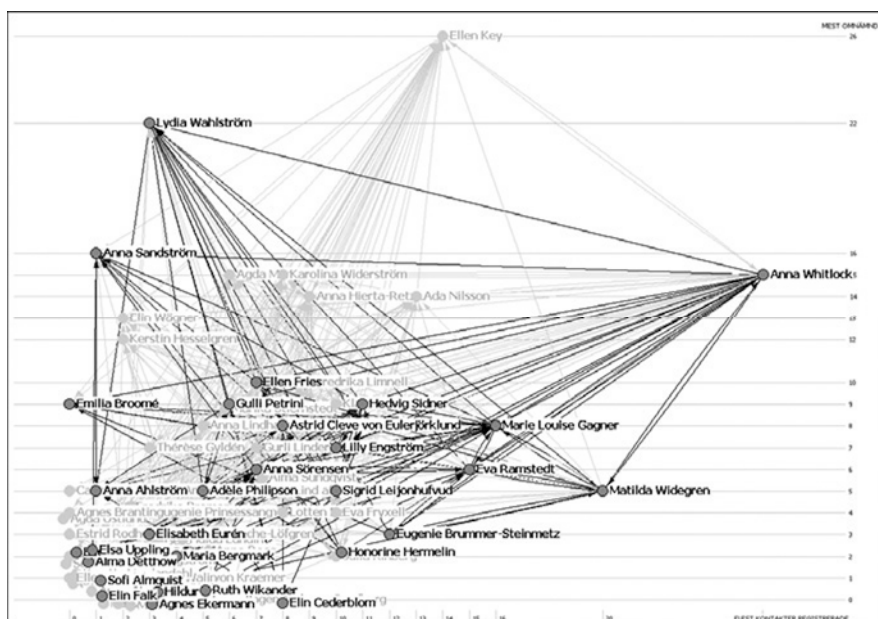
⁷⁰⁷ PED: pedagoger, PUB: publicister, SOC: socialt verkssamma. För beskrivning av den sociala klassificeringen se s. 199.

derns yrke) och färger för att markera sfärtillhörighet i noderna som avser representerar kvinnorna.⁷⁰⁸

De kvinnor som hade många kontakter, med övriga kvinnor i databasen FFO-KAPITAL, befinner sig i centrum av nätverksgrafen och kvinnorna med få kontakter återfinns således i periferin. I färgversionen av grafen kan urskiljas att pedagogerna är en relativt samlad grupp i ost-sydlig riktning medan publicisterna grupperar sig i syd-sydväst. De socialt verksamma är den grupp som är mest utspridd och de som har flest kontakter med kvinnor utanför den egna sfären. De formerar sig i mindre kluster med några få kvinnor i varje, bland annat återfinns en grupp socialt verksamma bland pedagogerna nere i söder, till skillnad från pedagoger och publicister.

Grafen visar också kvinnornas sociala bakgrund, baserat på uppgifter om faderns yrke. Vi kan här se att de socialt verksamma kvinnorna oftare kom från godsägarhem och bondehem än vad pedagogerna gjorde som istället oftare kom från intellektuell överklass. Publicisterna växte oftast upp antingen i intellektuell överklass eller i högre tjänstemannahem.

En nätverksanalys kan manipuleras och analyseras på flera olika sätt. I *Figur 33* är noderna (som representerar kvinnorna) sorterade i enligt flest kontakter med andra kvinnor (X-axel) och mest omnämnd i andras kapitalbeskrivningar (Y-axel).

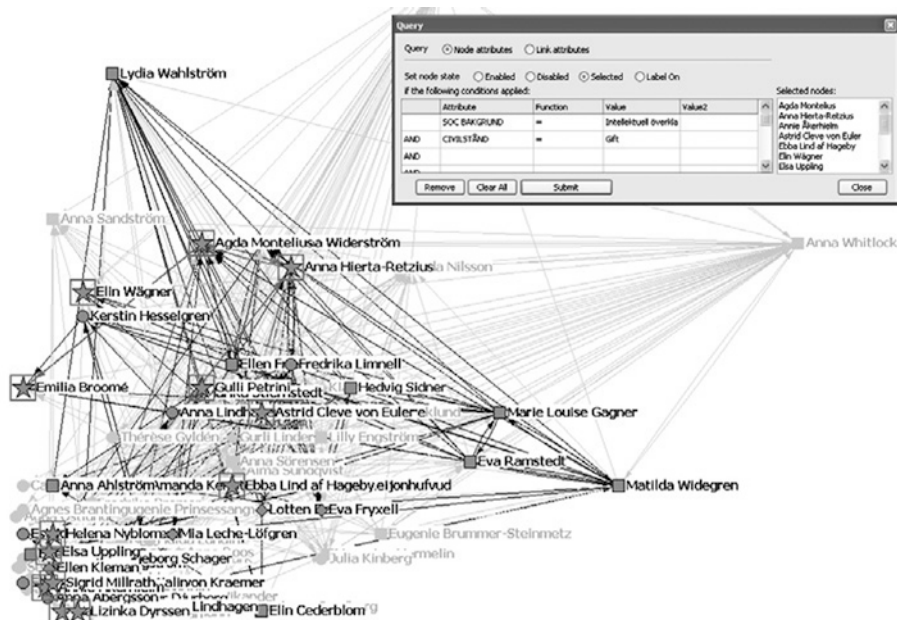


Figur 33. Nätverksanalys av FFO-KAPITAL: flest kontakter och mest omnämnd

⁷⁰⁸ För samma graf i färg, se bifogad digital bild på CD:n <cd:/figurer/figur32.jpg>.

De kvinnor som både är mest omnämnda och både har flest kontakter registrerade i sin egen kapitalbeskrivning, och därmed flest antal kontakter totalt, placeras i nordöstra hörnet av grafen.⁷⁰⁹ Lydia Wahlström utmärker sig genom att vara omnämnd i 22 andra kapitalbeskrivningar men har för egen del inte så många nämnda kontakter med andra (endast tre registrerade kontakter). Anna Whitlock har istället både många andra kontakter (15) och är både mycket omnämnd (26).⁷¹⁰ Grafen illustrerar även att förutom de kvinnor som utmärker sig och de som har medelstort kontaktnät, så är en grupp pedagoger samlade i det sydvästra hörnet. Man kan även se att de kvinnor som har ett stort kontaktnät också har fler kontakter utanför sitt eget verksamhetsområde, i detta fall pedagogik.

Andra användbara funktioner i denna analysmetod är möjligheten att göra olika filtreringar av den nätverksformation som skapats. Ett exempel är att kunna accentuera en viss grupp individer i analysen genom att filtrera utifrån attributvärden, frekvens, typ av relationer eller med frågor.⁷¹¹ Nedan följer ett exempel på hur en fråga används för att hitta och markera en viss grupp kvinnor.



Figur 34. Exempel på sökning i en nätverksanalys

⁷⁰⁹ I nätverksgrafen ovan accentueras pedagogerna, medan publicister och socialt verksamma får i exemplet stå tillbaka och är därför gråtonade.

⁷¹⁰ Ellen Key, som i exemplet är gråtonad, utmärker sig både genom flest antal kontakter totalt (omnämnd och egna registrerade), inte bara till andra publicister utan även till övriga sfärer, det vill säga jämfört med pedagoger och filantroper.

⁷¹¹ En sökning i en databas kallas för fråga (query), så det är egentligen en strukturerad sökning som utförs.

I *Figur 34* är det alla kvinnor som härrör från intellektuell överklass som accentueras och bland dem har de som var gifta skiljts från de ogifta genom att noderna som betecknar de gifta kvinnorna antagit formen av en stjärna. Det går till så att man väljer den variabel (attribut) och sedan vad operatorerna skall utföra, i detta fall ”lika-med”, och därefter väljs det värde som skall användas för sökningen. Resultatet av denna fråga visar att flertalet av de gifta kvinnorna hade kontakt med varandra (även om det inte syns så bra i bilden ovan) och att de gifta kvinnorna har en relativt god spridning mellan de med få och de med ganska många kontakter.⁷¹²

Förutom frågefunktionen finns en rad andra funktioner, till exempel för presentation och analys, till exempel att kunna identifiera kluster och gemenskaper, att analysera individers roller och positioner, att studera nodernas avstånd eller vägar. Genom det kan nätverket utforskas på flera olika sätt, inte minst tack vare möjligheter att ”interagera” med nätverket och att vissa aspekter kan accentueras medan andra får stå tillbaka.

I detta avsnitt har jag visat exempel på hur forskare, lärare och studerandes skulle kunna använda nätverksanalys för utforskningar av prosopografiska material. Att kunna visualisera komplexa material kan nyansera beskrivande statistiska analyser. Ibland kan det vara mer klagörande att visualisera komplexa samband där inte bara de binära relationerna studeras utan även andra aspekter som exempelvis social bakgrund och verksamhetsområde. I nästa avsnitt ges exempel på hur användare kan utforska prosopografiska material med hjälp av interaktiva semantiska ”kartor”.

7.3 Interaktiv visualisering

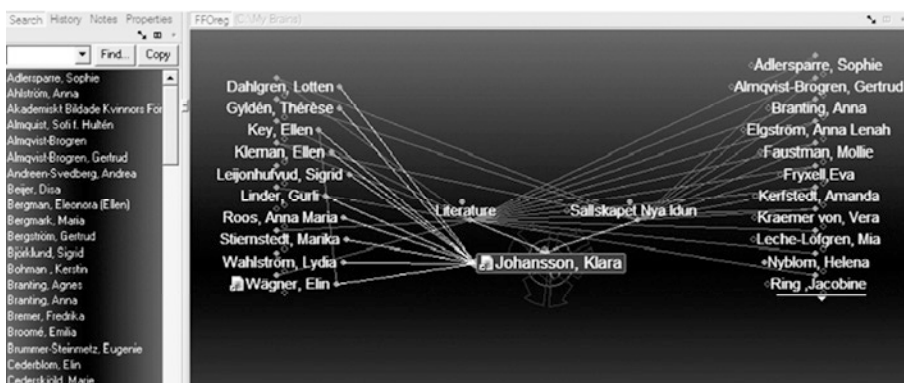
Utöver nätverksanalyserna kan detta material även användas till att skapa ett slags interaktiva kartor som visar både sammanhang och relationer.⁷¹³ I det aktuella exemplet är det kvinnorna själva som utgör en typ av nod och där organisationer och sfärer (grupperingar) är andra typer av noder. Till detta har bland annat standarden Topic Maps⁷¹⁴ använts. Topic Maps är avsedd att beskriva och kategorisera innehåll och dess relationer som tidigare nämnts.⁷¹⁵ Genom att applicera den här sortens märkning kan man skapa index, placerade utanför själva dokumentet till skillnad från den märkning som ovan beskrivits som appliceras inuti filen.

⁷¹² Se därför figur 34 vid <CD:/figurer/figur34.jpg>

⁷¹³ Detta är gjort på ett urval i prosopografen, nämligen de 101 kvinnor i kapitalbeskrivningarna som det finns utförlig information om.

⁷¹⁴ Se <www.topicmaps.org/> samt <www.ontopia.net/>.

⁷¹⁵ Se s. 99 f.



Figur 35. Exempel på en kvinno-ny

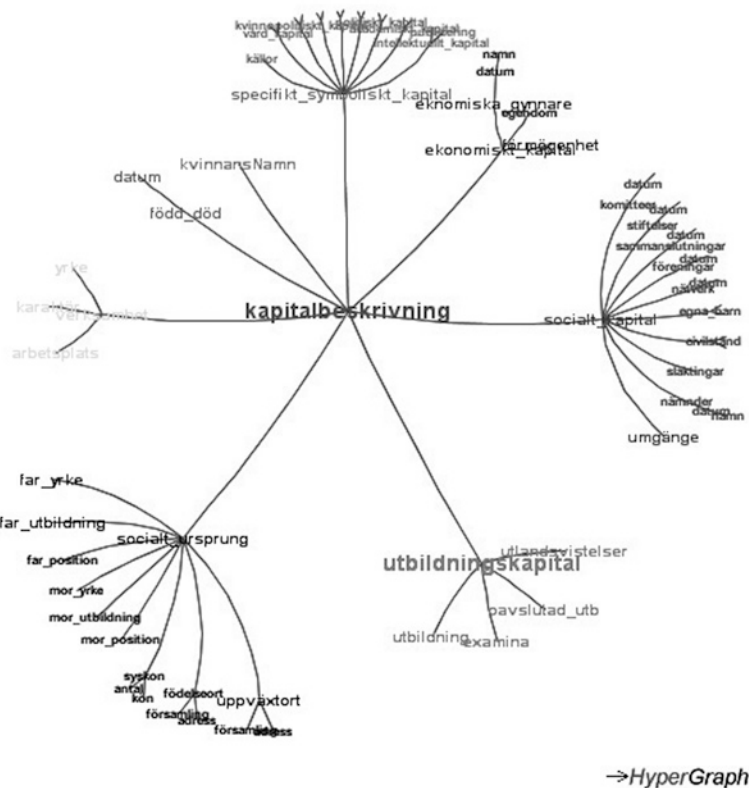
I Figur 35 presenteras ett exempel på en sådan karta där en kvinno-ny utkikspost är i centrum.⁷¹⁶ Här är det Klara Johanssons kontakter som accentueras, vilken sfär hon tillhörde och en av organisationerna hon var medlem i (sällskapet Nya Idun). Om användaren klickar på en annan nod, till exempel ett annat namn, kommer denna i centrum i stället. Det finns även index- och sökfunktioner samt länkar till mer information om Klara Johansson (länk till kapitalbeskrivningen vänster om namnet). Dessa interaktiva kartor ger användaren möjlighet att studera materialet genom att växla mellan individnivå och andra mer övergripande nivåer. Nackdelen med kartorna, speciellt de som är informationsrika, är att det kan vara ganska lätt att "gå vilse" bland all information.

XML-filerna kan även användas med hjälp av den typ av generella verktyg som *Text Analysis Portal for Research (TAPoR)*⁷¹⁷ tillhandahåller. Där finns en rad, fritt tillgängliga verktyg att välja bland vid TAPoR:s webb som kan användas för att studera och analysera XML-filer. Nedan följer ett exempel på hur en kapitalbeskrivning i XML-format transformerats till en interaktiv graf.⁷¹⁸

⁷¹⁶ I gränssnittet används färgsättning för att illustrera periferi och centrum, det blir ljusare längre ut i periferin och mörkare närmare centrum för den aktuella noden. Även linjerna som illustrerar relationer ändrar färg beroende på var muspekaren befinner sig.

⁷¹⁷ "[...] tools for sophisticated analysis and retrieval, along with representative texts for experimentation." Se vidare <<http://tapor1.mcmaster.ca/>>.

⁷¹⁸ Grafen gjord med visualiseringsverktyget HyperGraph som är baserat på XML. Tillämpning tillgänglig vid: <<http://taporware.mcmaster.ca/~taporware/xmlTools/hypergraph.shtml?>>>. Den dåliga återgivningen för några av förgreningarna beror på att verktyget använder för ljusa färger. För färgbild, se <CD:/bilder/hypergraph1.jpg>.

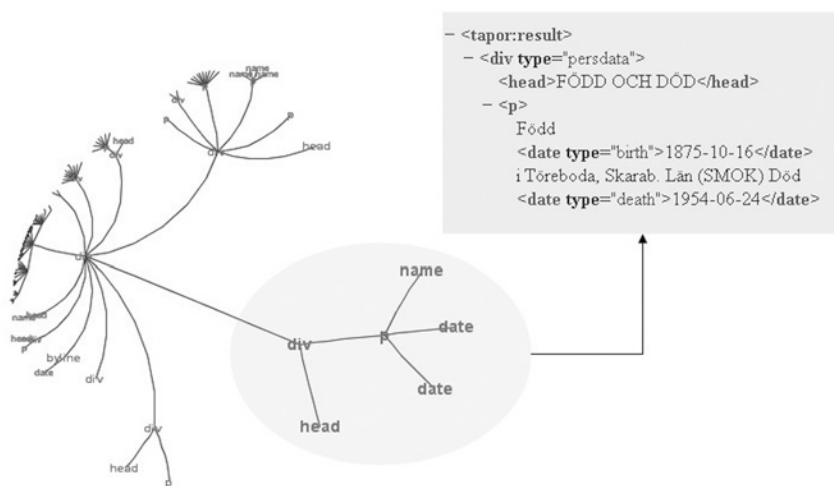


Figur 36. Visualisering av XML-dokumentets struktur

Här kan vi se att den överordnade principen, kapitalbeskrivningen (nivå 1), är i centrum varifrån de underliggande hierarkierna förgrenar sig.⁷¹⁹ Längst ut på de grenar som utgår från mitten återfinns de minsta grenarna som visualiserar de divisioner (nivå 3) som har något innehåll (det vill säga där själva texten finns).

I bilden nedan illustreras hur en ”trädgren” i en kapitalbeskrivning, det vill säga en XML-fil, accentuerats och hur innehållet i denna gren presenteras via den dynamiska grafen.

⁷¹⁹ Jfr. denna visualisering med den hierarkiska beskrivningen av kapitalbeskrivningsmallen i Figur 17 s. 159.



Figur 37. Illustrering av trädstrukturens yttersta gren och dess innehåll

Verktyget kan användas för att utforska innehållets struktur tillsammans med själva innehållet, något som gör det möjligt att studera hur trädstrukturen förgrenas ned till minsta innehållsobjekt.⁷²⁰ Detta kan vara användbart för att analysera uppbyggnaden och designen av stora filer eller för att lära sig hur innehåll kan beskrivas med hjälp av XML. Verktyget HyperGraph kan även användas för indexering av innehåll på webben där en användare kan manipulera grafen och klickar på en nod för att sedan länkas vidare till något annat, exempelvis en annan webbsida.

7.4 Att förena bilder med text

I FFO-HLS har den beskrivande märkningen haft en annan inriktning än för FFO-KAPITAL, dels för att en målsättning varit att sammanföra fotografierna med den transkriberade texten, dels på grund av källmaterialets beskaffenhet. Det märkta arkivmaterialet i FFO-HLS ger dock ungefär samma typ av möjligheter till bearbetning och analyser som FFO-KAPITAL, förutom att databasen FFO-HLS ger dessutom ytterligare möjligheter för presentation och analys. Tack vare XML-formatet kan de fotograferade arkivposterna och de transkriberade bildtexterna kopplas ihop och publiceras på webben.⁷²¹

⁷²⁰ Bilden är en modifierad version av två presentationer i olika fönster i webbläsaren, den dynamiska grafen visas i ett fönster och innehållet i ett annat. Grafen är konstruerad med HyperGraph, som refereras i not 718, s. 214. För färgbilder, se <CD:/bilder/hypergraph2.jpg> som illustrerar en kvinnas kapitalbeskrivningar samt <CD:/bilder/hypergraph3.jpg> som illustrerar innehållsobjektet i figuren ovan.

⁷²¹ Exempel tillgängliga vid <www.skeptron.uu.se/monicalz/khls/bilder.htm>.

Att förena transkriberad text med avbildningen av originalet kan vara användbart på flera sätt. Inte minst för källmaterial som är svåråtkomligt eller skört. Men även för att användaren kan studera källmaterialets grafiska utseende samt kontrollera avskrifterna mot originalets avbildning. Vid Kungl. Biblioteket har flera tillämpningar av det här slaget tagits fram. Ett exempel är praktverket *Suecia antiqua et hodierna*, ett planschverk över stormaktstidens Sverige som påbörjades år 1661 och där tre band, index och bokbandet digitaliserats.⁷²² I den digitala utgåvan kan användare bläddra bland de grafiska avbildningarna, titel- och bildinformation, katalogfakta (information om den aktuella arkivposten), länk till den samling bilden tillhör, zooma in bilden, sökning i verkets databas etcetera. Sammantaget en mycket informationsrik webbplats och värdefull källa för forskning och undervisning. I anslutning till bildvisningen finns även bildens titel avskriven. Denna text relateras dock inte till var den är placerad i själva bilden. Flera av originalbilderna innehåller text, exempelvis placerad i nedre marginalen. Denna bildtext visas inte i den transkriberade versionen för användaren, utan här får zoomverktyget användas för eventuell läsning.⁷²³ Om den transkriberade bildtexten (inklusive en översättning från latinet) hade placerats i anslutning till avbildningen hade detta kunnat förhöja användbarheten av den digitaliserade samlingen ytterligare.

Detta är tekniskt möjligt att utföra med hjälp av XML-applikationer, exempelvis med hjälp av *Scalable Vector Graphics*, SVG.⁷²⁴ I denna XML-applikation används tvådimensionell vektorgrafik för bildrepresentation, vilket i korthet innebär att bildens form, storlek och placering anges med koordinater. Då SVG är en tillämpning av XML uttrycks också koordinaterna i XML-formatet.

Nedan presenteras exempel på hur fotografiet och den transkriberade texten kan förenas med hjälp av beskrivande märkning.⁷²⁵

⁷²² "Det forna och nuvarande Sverige". Se den digitaliserade samlingen och läs bakgrund till verket som sådant, om källorna, KB:s bestånd samt om upphovsrätten för *Suecia antiqua et hodierna* vid <www.kb.se/samlingarna/digitala/suecia/>.

⁷²³ Se t.ex. <www.kb.se/samlingarna/digitala/suecia/andra-bandet/Arsta/>.

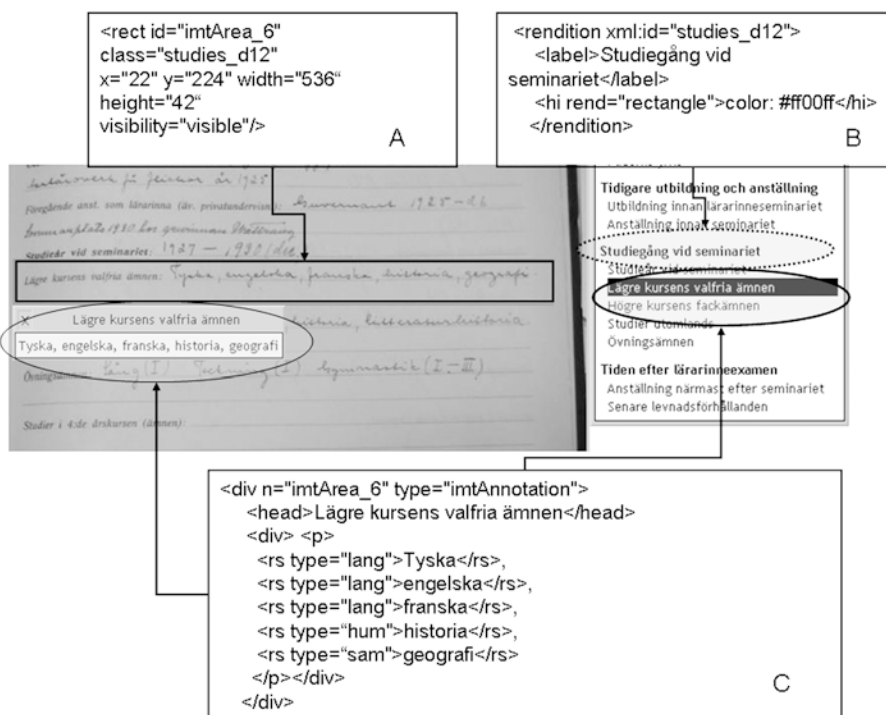
⁷²⁴ <www.w3.org/Graphics/SVG/>.

⁷²⁵ Tillämpningen som presenteras nedan baseras bland annat på *Image Markup Tool* (IMT 1.6), se <www.tapor.uvic.ca/~mholmes/image_markup/>. Andra liknande verktyg på marknaden använder antingen proprietära format är eller är aningen för komplicerade för icke-expert, jämfört med IMT (det verktyg som jag använt) som använder XML-och TEI-formaten.



Det verktyg jag använt för märkning av fotografiet är en applikation utvecklad i syfte att tillgodose behovet av lättanvända verktyg för beskrivning och märkning av bilder. I *Figur 39* nedan presenteras hur verktyget fungerar samt de principer och metoder som möjliggjort denna tillämpning där fotografierna och den avskrivna texten förenas.

⁷²⁷ Förutom XML-filerna behövs en del andra verktyg för att få detta att fungera i en webbläsare, bl.a. används Java-skript för interaktivitet samt objektens placering och CSS stilmallar används för formatering av typsnitt, färg etc.



Figur 39. Exempel på relaterande märkning som förenar bild och transkriberad text

I programmet Image Markup Tool utgår man från fotografiet där, man markerar de områden som skall rama in den aspekt som skall accentueras. I figuren ovan är det den "lägre kursens valfria ämnen" som markerats. I programmet genereras sedan koordinater som motsvarar den grafiska markeringen placering i bilden. I ruta A i figuren ovan visas den XML-märkning som beskriver markeringens placering och form. Den transkriberade texten (ruta C) relateras till markeringen genom ett löpnummer (imtArea_6). I ruta B, bredvid bilden, finns en innehållsförteckning bestående av de transkriberade huvudrubrikerna med tillhörande underrubriker som är kopplade till bildens markeringar. Genom att klicka på en underrubrik i ruta C markeras den samtidigt som motsvarande områdesmarkering visas i bilden.⁷²⁸

Att på detta sätt enkelt kunna koppla ihop bilder med den transkriberade texten kan vara värdefullt, inte minst som verktyg för att utforska ett digitaliserat arkivmaterial samt möjligheter att kontrollera den transkriberade texten gentemot de avbildade källorna. Vidare kan den märkning som infogas och relateras till varje bild och varje markering användas till sökfunktioner och in-

⁷²⁸ Den XML-fil som innehåller all märkning för just denna bild/post bifogas vid <CD:/xml/imt-example.xml>.

dexering av de typer av innehåll som beskrivits, till exempel att skapa index utifrån rubrikernas typning.

I och med att detta arkivmaterial redan är så väl strukturerat har det varit möjligt att förenkla och skala upp den manuella processen för markering och märkning av bilderna. Samtliga poster (sidor) i inskrivningsmatriklarna har samma form: rubrikerna är placerade på samma ställe, innehållet skrivs på samma rader och i samma ordning. Genom att nyttja de avbildade dokumentens struktur har en mall skapats där markeringarnas placering och annan metadata används som mall och där endast innehållet byts ut. Detta kräver dock att bilderna sparas med exakt samma storlek för att områdesmarkeringarnas koordinater skall placeras på samma plats i samtliga bilder. Mallen används sedan som en behållare som fylls med den transkriberade och redan märkta texten för varje enskild arkivpost. Proceduren utförs med hjälp av skriptspråket XSLT som innehåller funktioner för transformationer av innehåll från ett startdokument (avskrifter av inskrivningsposter) till ett måldokument (dokument för där bild och text kopplas ihop). Detta tillvägagångssätt har snabbat på märkningsprocessen betydligt, men kräver en del arbetsinsatser för kontroll av att områdesmarkeringarna är rätt placerade och inte minst att rätt uppgifter kopplats till rätt bild.⁷²⁹

Exemplen ovan är gjorda med en tidigare version av applikationen. Från och med version IMT 1.7 har applikationen kopplats till TEI:s riktlinjer. Nu används elementen <facsimile>, <surface> och <zone> i modulen *Representation of Primary Sources* i TEI P5 istället för SVG-formatet som tidigare användes.⁷³⁰ Funktionaliteten är ungefär densamma i den nyare versionen har dock möjligheter till utbyte och överförning mellan plattformar och projekt väsentligen förbättras i och med övergången från SVG till TEI-formatet.

Den tillämpningen som här presenterats är en del av den webbplats som utformats med utgångspunkt i det prosopografiska materialet och som presenteras i nästföljande avsnitt.

7.5 Webbarkivet

Den sista tillämpningen som avhandlingen tar upp är det webbarkiv som byggts upp med hjälp av det prosopografiska materialet. Webbarkivet innehåller inte enbart FFO-databasen utan även kompletterande material i form av register, statistiska analyser, skript och stilmallar för transformering och migrering, förteckningar och listor etcetera, det vill säga den dokumentation som beskriver

⁷²⁹ Start- och måldokument bifogas, se <CD:/xml/khls-example.xml> samt <CD:/xml/imt-example.xml>, därtill bifogas det skript som använts för migrering av innehåll från start- till måldokumentet, se <CD:/xml/verktyg/imt-example.xml>.

⁷³⁰ Se ”11 Representation of Primary Sources” TEI Consortium, red, *TEI Consortium. Guidelines for Electronic Text Encoding and Interchange. P5*. <www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html> för närmare information om modulen för märkning av källmaterial.

arbetsmomenten i innehållsdesign tillsammans med det konkreta resultatet i form av den prosopografiska databasen.

Webbarkivet är tillgängligt vid <www.skeptron.uu.se/monicalz/ffo/> och innehåller länkar och hänvisningar till de teman som presenteras nedan.

Den CD som bifogas avhandlingen innehåller en del av allt material i webbarkivet. En anledning till att inte allt finns med på CD:n är att vissa program som behövs för funktionaliteten i webbarkivet måste vara installerade för att dessa tillämpningar skall fungera. Ett exempel på det är relationsdatabasen för FFO-KAPITAL med webbgränssnitt som presenteras närmare nedan. Detta delkapitel kan därför betraktas som en summering av de tillämpningar som tidigare presenterats dock med tillägg av FFO-databasen. Webbarkivet är för närvarande strukturerat enligt olika teman som var och en innehåller fristående moduler med skilda funktioner.

Tabell 6. *Översikt över webbarkivets moduler och funktioner.*

Modul	Beskrivning	Funktion
Fulltextarkiv	Samtliga transkriptioner av kapitalbeskrivningar och arkivposter i XML- och HTML-format	Fritt tillgängliga namnförteckningar och masterfiler för nedladdning och återanvändning
Dokumentationsarkiv	Mallar för kapitalbeskrivningar, Klassificeringar och förteckningar av olika slag	Tillgång till dokument där arbetet beskrivs och förtecknas
Källmaterial	Förteckningar och beskrivning av det källmaterial som har använts samt samlad bibliografi för FFO-KAPITAL	Dokumentation om innehåll och insamling av det prosopografiska material
Märkningsarkiv	Dokumentation om märkningens strategi och genomförande	Fritt tillgänglig dokumentation med exempel om märkning av historiskt och prosopografiskt material
Databas med webbgränssnitt	Databas som för närvarande innehåller allt material i FFO-KAPITAL samt delar av FFO-HLS	Webbgränssnitt med sökfunktioner samt tillgång till fulltext (FFO-KAPITAL) samt bilder (FFO-HLS) via ett webbaserade gränssnitt
Registerarkiv	Register som iordningställts för olika slags analyser. Även originalregistren för avskrifterna ur Lundström och Hedberg & Arosenius biografiska uppslagsböcker	Register för påseende eller nedladdning.
Analys	Beskrivande statistik, korrespondensanalys och nätverksanalyser	Exempel på olika slags analyser inklusive beskrivning av utförande och tolkning

Modul	Beskrivning	Funktion
Verktygslåda	Förteckningar och hänvisningar till de metoder och verktyg som använts för studien. Inkluderar skript, stillmallar och databasmodell som används för transformering och migrering av det digitaliserade materialet	En samlad fritt tillgänglig dokumentation för märkning och bearbetning av samhällsvetenskapligt och humanistiskt material (särskilt historiskt och prosopografiskt material)
Länkar	Sparsam hänvisning till digitala arkiv av särskilt intresse, t.ex. prosopografiska samlingar, kvinnohistoriska samlingar	Hänvisningar till andra digitala samlingar med goda exempel eller annat särskilt intressant material

Den tematiska indelningen i moduler är utformad i syfte att de var och en skall vara relativt fristående. En modul skall kunna färdigställas eller revideras utan att alla andra blir väsentligen påverkade.

Nedan presenteras en databasmodul och dess webbgränssnitt lite närmare för att ge ytterligare exempel på hur masterfilen har kommit till användning.

7.5.1 Databasmodulen FFO-KAPITAL

Den databasmodul som för närvarande är utvecklad är den som innehåller materialet i FFO-KAPITAL. Med utgångspunkt i kapitalbeskrivningarna och i den klassificering som gjorts av bland annat utbildning, föreningsaktiviteter och yrkestitlar har en relationsdatabas iordningsställt.⁷³¹

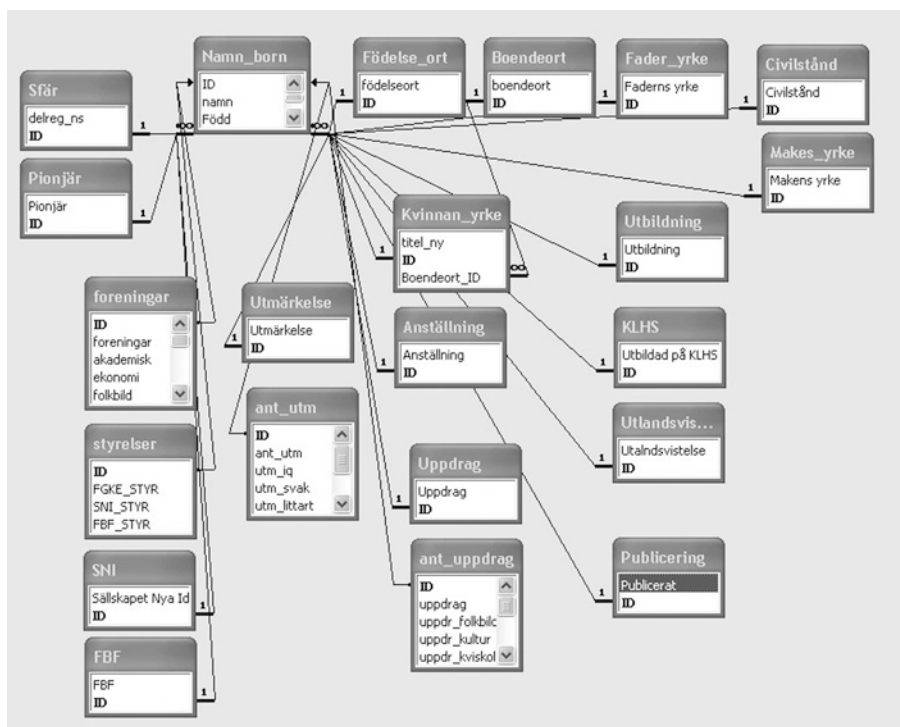
En relationsdatabas kan i korthet beskrivas som en mängd av data som ordnas och lagras i tabeller, eller snarare relationer, som försetts med vissa villkor antingen inom sig eller mellan varandra. I tabellerna lagras sedan poster med samma attribut i rader och kolumner. Tabellerna relateras med varandra genom olika sorters restriktioner. I en relationsdatabas eftersträvas minimal lagring av redundant information, till skillnad från den hierarkiska dokumentmodellen som har använts för masterfilerna.⁷³²

I *Figur 40* nedan visas en grafisk vy av relationsdatabasens schema som anger vilka data som skall lagras i tabellerna samt hur dessa tabeller är kopplade till varandra.⁷³³

⁷³¹ En databashanterare är ett program som lagrar och hanterar databaser, kan därmed hårbärgera flera databaser samtidigt. Den databashanterare som används för denna modul är MySQL som använder frågespråket SQL. MySQL är utvecklad i Sverige och den enklare databashanteraren är fri programvara, licensierad under GNU General Public License. MySQL såldes i januari 2008 till IT-företaget Sun. Andra kända relationsdatabashanterare är Oracle och Microsoft SQL Server.

⁷³² För en överskådlig och bra introduktion till relationsdatabaser, se Thomas Padron-McCarthy & Tore Risch, *Databasteknik* (Lund: Studentlitteratur, 2005).

⁷³³ Skillnaden mellan schema och datamodell är att i datamodellen bestäms hur schemat skall utformas och i schemat bestäms vilka data som får lagras i databasen och var de skall lagras.



Figur 40. Databasschema för FFO-KAPITAL

Detta schema är egentligen inte särskilt optimerat med tanke på att en hel del redundant information lagras i databasen. För en absolut frånvaro av icke-redundant data skulle det behövas en bättre design med många fler tabeller som lagrade informationen så att varje typ av data eller information separerades från varandra. I denna databas skulle man kunna tänka sig att en tabell enbart innehöll namn på föreningstyper. Denna tabell skulle relateras till en tabell för föreningsaktivitet 1, en annan tabell för föreningsaktivitet 2 etcetera som var och en lagrade data om föreningstyper. På detta sätt skulle de kvinnor som inte hade någon föreningsaktivitet alls inte behöva registreras i någon av tabellerna där data om föreningsaktiviteter lagras. Nu är informationen om föreningsaktiviteter istället kombinerat i en och samma tabell där data lagras i attributen i form av binära värden, antingen var kvinnan med i en viss föreningstyp eller inte.

Denna databasmodul har ett gränssnitt så att användare kan söka och utforska materialet på webben. I *Figur 41* nedan ges exempel på poster som kan användas för sökning i databasen.

Sök i FFO databasen

Sök på en eller med flera kriterier

Sök i FFO databasen	
Namn	
Sfär	
Titel	
Född	
Födelseort	
Utbildning	
Kungl.högreläroinneseminarier	☒ ☐ Ej utbildning på HLS ☐ Utbildning på HLS
Faderns yrke	
Civilstand	Bondeklass Gödsägarklass Högborgerlig över- och medelklass Intellektuell överklass Småborgerlig/hantverkare/lägre tim Tjänstemannaklass Uppgift saknas
Makens yrke	
Sni	☒ ☐ Ej medlem i Sällskapet Nya Idun ☐ Medlem i Sällskapet Nya Idun
Fbf	☒ ☐ Ej medlem i Fredrika Bremer förbundet ☐ Medlem i Fredrika Bremer förbundet
Back Search Clear	

Figur 41. Webbaserat gränssnitt för sökning i FFO-KAPITAL

I den webbaserade sökfunktionen kan flera en eller flera av posterna användas för enkla eller sammansatta strukturerade sökningar.⁷³⁴ Vid vissa sökfält anges fördefinierade värden, dels för att undvika att sökningen sker med värden som inte finns registrerade i databasen, dels som en hjälp till användaren vid konstruktion av sökfrågan. I detta exempel kan flera olika parametrar användas var och en för sig eller genom att kombineras. Ju fler parametrar som används, desto mer snävas sökningen in och desto färre träffar kommer den att generera.

Nästa exempel visar hur resultatet av en sökning om publicister i FFO-KAPITAL presenteras.

⁷³⁴ Det webbaserade sökgränssnittet är gjort med MySQL och PHP. PHP är ett fritt tillgängligt skriptspråk som ofta används för webbapplikationer som hanterar dynamiskt innehåll och passar tillsammans med MySQL. Databasen tillgänglig via <www.skeptron.uu.se/monicalz/ffo/>.

Bläddra eller sök poster i databasen
eller visa kapitalbeskrivningar i fulltext (html- och xmlformat)

Hittade 21 post(er)

Exportera poster. [Välj format]

Namn	Star	Titel	Född	Födelseort	Bondeort	Utbildning	Kapitalbeskrivning i XML-format	Kapitalbeskrivning i HTML-format
Adlerparne, Sophie	Publicister	Skriftställarinna	1823	Född på landsort	Boende Stockholm	Lärarinneseminarium	adlerparne-sophie.xml	adlerparne-sophie.html
Almqvist-Brogren, Gertrud f. Almqvist	Publicister	Författarinna	1875	Född på landsort	Boende Stockholm	Akademisk utbildning (UG)	almqvist-brogren-gertrud.xml	almqvist-brogren-gertrud.html
Bendixson, Nanna	Publicister	Barnboksöförfattare	1860	Född Stockholm	Boende Stockholm	Konsthärlig utb.	-	-
Brandting, Anna	Publicister	Författarinna	1855	Född Stockholm	Boende Stockholm	Flickskola/läroverk flickor	brandting-anna.xml	brandting-anna.html
Dahlgren, Lotten	Publicister	Skriftställare	1851	Född Stockholm	Boende Stockholm	Övr utbild	dahlgren-lotten.xml	dahlgren-lotten.html
Eigström, Anna Lenah	Publicister	Författarinna	1884	Född på landsort	Boende Stockholm	Konsthärlig utb.	eigstrom-anna-lenah.xml	eigstrom-anna-lenah.html
Faustman, Mollie	Publicister	Konstnär	1883	Född Stockholm	Boende Stockholm	Konsthärlig utb.	faustman-mollie.xml	faustman-mollie.html
Frynell, Eva	Publicister	skriftställarinna	1828	Född Stockholm	Boende Stockholm	Uppg. Saknas	frynell-eva.xml	frynell-eva.html
Johansson, Klara	Publicister	Skriftställare	1875	Född på landsort	Boende Stockholm	Akademisk examen	johansson-klara.xml	johansson-klara.html
Kerfstedt, Amanda	Publicister	Författarinna	1835	Född på landsort	Boende Stockholm	Flickskola/läroverk flickor	kerfstedt-amanda.xml	kerfstedt-amanda.html
Key, Ellen	Publicister	skriftställarinna	1849	Född på landsort	Boende Stockholm	Lärarinneseminarium	key-ellen.xml	key-ellen.html
Kleman, Ellen	Publicister	Redaktör	1867	Född på landsort	Boende Stockholm	Flickskola/läroverk flickor	kleman-ellen.xml	kleman-ellen.html

Figur 42. Webbgränssnitt för visning av resultat efter sökning i FFO-KAPITAL

Resultatet visar de poster som stämde överens med den ställda sökfrågan och som hämtats från tabellerna i databasen. Hur resultaten presenteras på webben kan variera, i det här exemplet visas resultaten, det vill säga de kvinnor som överensstämde med sökkriterierna i tabellform. I framställningen finns länkar till kapitalbeskrivningarna i fulltext (XML- och HTML-format). Resultattabellen kan dessutom sorteras utifrån de enskilda kolumnerna genom att klicka på kolumnrubriken. Därtill kan samtliga poster exporteras till ett nyskapat dokument i olika filformat.⁷³⁵ Genom den webbaserade databasen ges möjlighet för användare att söka strukturerat efter vissa poster, eller välja alla visning av samtliga poster, därtill möjlighet att läsa hela kapitalbeskrivningen i fulltext.

Jag har i detta avsnitt gett exempel på hur det prosopografiska materialet gjorts tillgängligt och förhoppningsvis även brukbart för olika grupper av användare.

7.6 Sammanfattande kommentarer om tillämpningar

Detta kapitel har ägnats åt de tillämpningar som utvecklats med hjälp av det digitala materialet. Min ambition har varit att påvisa den mångfald av möjligheter med det digitala material som iordningställts och utformats. I inledningen betonade jag behovet av en bättre beställarkompetens och goda exempel för forskare, lärare och studenter. Dessa tillämpningar är ett bidrag till en ökad kunskapsfond i det avseendet. Genom tillämpningarna, som inte ger exempel baserade på enkla påhittade exempel, utan är baserade på forskningsmaterial är

⁷³⁵ För närvarande XML-, text- (TXT), Excel- (XLS) och HTML-formaten.

min förhoppning att det kan tjäna som användbara exempel på möjliga tillvägagångssätt.

Jag har några gånger berört att innehållsdesign av större material är relativt tidsödande. Men en estimerad, och verklig, tidsåtgång bör även jämföras med den mångfald av möjligheter ett sådant material kan ge, för forskning såväl som för undervisning.

I tabellen nedan illustreras en uppskattad tidsåtgång för de olika arbetsmomenten.

Tabell 7. *Uppskattad tidsåtgång för olika moment i arbetet med den prosopografiska databasen*

Typ av moment	Beskrivning	Uppskattad tidsåtgång
Insamling/arkivering av kapitalbeskrivningar	Mottagande av fil, namngivning, föra in i register och spara i rent textformat.	Löpande arbete som svårt att uppskatta, ca 10 min per kapitalbeskrivning.
Design av datamodell och märkningsschema	Utforma och ta beslut om t.ex. TEI-moduler, datamodell	2-3 dagar.
Märkning kapitalbeskrivningar	Märkning av grundläggande struktur, samt namn, datum o platser, bibliografi etc. Tidsåtgången för märkning beror i hög grad på hur väl strukturerad texten är från början samt i vilken utsträckning det går att automatisera märkningen av vissa begrepp.	En omfattande kapitalbeskrivning (8000 ord) > 3 tim (nybörjare), ca 1 tim för erfaren person. Mindre omfattande kapitalbeskrivning ca 1-2 tim nybörjare, ca 30 min av mer erfaren person.
Utformning av XSLT-skript	Hitta rätt funktioner i XSLT:s bibliotek, undersöka om det redan finns andra skript, testa och ”provköra”, revidera	Från 1 tim till 2 dagar beroende på hur avancerade skript det rör sig om.
Bearbetning av register	Klassificering, variabler, namnge etiketter, etc. göra och spara skript som kan återanvändas.	Svårt att uppskatta, tar längre tid än man tror.
Fotografering	Fotografering på Stockholms stadsarkiv, inkl. bedömning av de frambeställda källorna och annat kringarbete, drygt 1500 bilder. inkl. väntetid för framställning av nya kapslar.	1,5 – 2 dagars arbete för 1-2 personer.
”Arkivering” av foton	Namngivning och arkivering av fotografierna i kataloger, kontroll.	Ca 1- 2 dagars efterarbete för att arkivera drygt 1500 fotografier.
Registrering av arkivposter	Transkribering av 1980 fotograferade personposter med namn, födelesår, födelseort, div. läsårsdata, anmärkningar, examen, examensdatum.	Strax under 2 veckors arbete för 1 person.
Märkning av detaljerade arkivposter samt länkning	Märkning arkivposter som är väl strukturerade, går delvis att automatisera.	ca 1 tim per personpost, beroende bl.a. på hur läslig handstilen är samt antalet korrigeringar och anteckningar för den aktuella posten i liggaren

Typ av moment	Beskrivning	Uppskattad tidsåtgång
Design av databas med sökgränssnitt	Design av databas, databas-schema och webbgränssnitt, installation på server, inläsning av data etc.	Cirka 2 veckors arbete för en fungerande version, längre tid för arbete med fler funktioner och bättre gränssnitt.
Transformerering/migrering	Överföring av data mellan plattformar	Svårt att uppskatta, allt från 1 min till 8-10 timmar om något krånglar.
Nätverksanalyser	Iordningställa materialet för nätverksanalyser, dvs. transformera XML-filer till annat XML-format och/eller till matriser.	ca 2 dagar för FFO-KAPITAL
Sammanställning och dokumentation	Skriva handledningar, upprätta register och allmänna noter, ingar om sådant som annars glöms bort	Ett antal dagar/veckor.

Som tabellen illustrerar varierar tidsåtgången mellan väsentlig mellan, och inom de olika arbetsmomenten och det är med andra ord mycket svårt att göra en saklig uppskattning. Det som framför allt bidrar till att minska tidsåtgången under hand är, förutom större erfarenhet av momenten, hur många av delmomenten som går att automatisera. Alla skript och alla rutiner som kan generaliseras så pass att de är värda att återanvändas spar mycket tid. Sedan bidrar förstås ökad rutin till en snabbare process, därtill och inte minst verktygen användbarhet.

Med utgångspunkt i de tillämpningar som här presenterats har min avsikt varit att belysa hur till exempel de statistiska analyserna kan tjäna som ett bra komplement för att förstå individuella kvinnors strategier och insatser för samhällslivet. I undervisningen skulle de här exemplen kunna användas för att visa hur statistiska analyser kan användas för att klargöra vissa tendenser och strukturer, men att tolkningen av resultaten måste förstås utifrån överlagda historiska bedömningar likväl som bedömningar om källornas tillförlitlighet och uppgifternas jämförbarhet.⁷³⁶

Underlag som erbjuder exempel på svårbedömda jämförelser kan också vara användbar i undervisningssituationer.⁷³⁷ Exemplet här är problemen med klassificering av kvinnornas yrkestitlar under den här studerade perioden. Den klassificering som nu används baseras på en specifik grupp och inte i jämförelse med befolkningen i stort. Därtill bör tolkningarna utgå från andra faktorer än exempelvis enbart yrkesverksamhet och för detta ändamål används indikatorerna för egenskaper av socialt- och olika slags symboliskt kapital.

Andra aspekter på sådant som skulle kunna användas som underlag för vidare analyser är nätverksanalyserna. Exempelvis att analyser av kvinnornas verksamhet, föreningsaktiviteter eller sociala nätverk kanske enbart bekräftar redan

⁷³⁶ Se t.ex. Tosh, *Historisk teori och metod* s. 192-212.

⁷³⁷ Jfr. Johansson, *Ådalen 1931 - minuterna som blev historia: metodbok till en digital dokumentsamling*.

kända förhållanden, samband eller tendenser, eller tolkningen att de mest aktiva och kända kvinnorna i urvalet hade stora kontaktnät inom många olika samhällsfrågor var en effekt av deras. Jag hävdar dock att det som framkommer i analyserna av till exempel kvinnornas sociala nätverk och föreningsaktiviteter kan hjälpa till att precisera, eller förkasta, den typen av antaganden.

Webbarkivet har utformats i syfte att dokumentera arbetet och göra det fritt tillgängligt för forskning såväl som för undervisning som andra kan ta del av materialet för att utforska innehållet, göra egna analyser, använda i andra tillämpningar och som exempel på innehållsdesign. Detta webbarkiv innehåller inte i första hand en mängd tillrättalagda exempel utan konkret forskningsmaterial som även är tänkt att kunna användas till undervisning, antingen som exempel på form, det vill säga innehållsdesign eller antingen för att utforska ett visst innehåll, det vill säga ett utbildningshistoriskt material. Genom att tillgängliggöra och strukturera innehållet tematiskt är målsättningen att webbarkivet skall kunna användas av forskare, lärare och studerande, som utgångspunkt för nya analyser och nya frågor, som underlag för kritik av forskningsmetod eller kategorisering av innehållet.

De tillämpningar som presenterats är inte utvecklade med någon särskild pedagogik eller undervisningsmetod i åtanke utan de utgår från material tillhands. Min förhoppning är att dessa tillämpningar skall kunna inspirera andra att använda öppna standarder och fria program vid design av forskningsmaterial. För den som anser att den hierarkiska dokumentmodellen är för strukturerad och strikt finns flera alternativ, några av dem har helt kort berörts i avhandlingen. Oavsett vilka metoder och verktyg som väljs är modellering av innehåll en viktig faktor inom forskning såväl som undervisning. Om inte materialet modelleras och struktureras med utgångspunkt i några principer eller någon modell blir det svårt, om än inte omöjligt, att sedan bearbeta och manipulera på något användbart sätt.

Del IV Avslutning

8 Sammanfattande diskussion

Den studie som här presenterats har närmast karaktär av metodutveckling. Ambitionen har varit att finna och analysera ändamålsenliga principer, metoder och verktyg för att beskriva, ordna och bearbeta digitalt material. För konktionens skull har jag lyft fram särskilda tillämpningar som har med forskning och undervisning inom utbildningshistoria och prosopografi (ung. kollektivbiografi) att göra, men analyserna är tillämpliga på många fler arbetsuppgifter som går ut på att bygga upp digitala arkiv för forsknings- och undervisningsändamål.

I början av avhandlingen identifieras området innehållsdesign. Begreppet introducerades av Donald Broady vid KTH 1997 för att identifiera ett allt viktigare verksamhetsområde när det gäller framställning av digitalt material. Termen ”innehållsdesign” valdes för att framhäva områdets egenart i förhållande till systemdesign och grafisk design, det vill säga i förhållande till utveckling och användning av datorprogram respektive formgivning av det som visas på skärmen, skrivs ut på papper eller presenteras i form av talsyntes eller på annat sätt. Däremellan, mellan de tekniska förutsättningarna å ena sidan och presentationsformerna å den andra, finns ett område, innehållsdesign, inte att förväxlas med det skapande av innehåll som författare, fotografer och andra ägnar sig åt. Innehållsdesign innebär inte att skapa innehåll utan att utforma innehållet på sådant vis att det blir brukbart. En innehållsdesigners uppgifter liknar närmast en teknisk redaktörs insatser när digitala utgåvor färdigställs. Med den skillnaden, som Broady uppmärksammade, att medan digital publicering ditintills varit inriktad på att leverera innehållet i en särskild på förhand given form, som en pappersbok, på en CD eller som filer på en webbplats, var teknikens utveckling på väg att öppna möjligheter för att flyttbart och mångsidigt användbart innehåll, vars framtida användning innehållsdesignern inte kan förutse och inte bör föregripa.⁷³⁸

I avhandlingen har jag strävat efter att dels urskilja och känneteckna – med hjälp av kategorier som beskrivande märkning, modellering, klassificering, databashantering, digital infrastruktur, statistisk och annan bearbetning, digital utgivning, webbpublicering – olika moment som kan ingå i innehållsdesign, dels identifiera, beskriva och analysera ändamålsenliga principer, metoder och verktyg för innehållsdesign.

⁷³⁸ Broady, *Content design. Methods and tools for the creation of portable hypermedia archives. Notes for a proposed CID project. Version 2*; Broady, ”Digitala arkiv och portföljer”.

Därtill har jag försökt visa – teoretiskt och praktiskt – hur de olika metoderna och verktygen kan användas i ett konkret sammanhang, nämligen i utformningen av digitalt material för forskning och undervisning, med tillämpningar inom utbildningshistoria och prosopografi.

8.1 Varför innehållsdesign?

En fördel med termen design är att den fungerar både i både verb- och substantivkonstruktioner. Verbet designa avser utveckling och planering av utformningen av ett system, en produkt eller en modell, medan substantivet design avser själva resultatet, det vill säga modellen, systemet eller produkten. Därtill är den engelska översättningen, *content design*, enkel och precis.

Vad skiljer innehållsdesign från exempelvis webbdesign? Innehållsdesign omfattar inte bara om sådana tillämpningar som jag uppehåller mig vid, såsom att skapa underlag för digitala utgåvor eller att konstruera register för statistiska analyser av prosopografiskt material. Innehållsdesign är ett bredare område än så, då målsättningen är att skapa förutsättningar för en digital infrastruktur som tillåter att innehållet används till webbpresentationer såväl som i databaser och för statistiska analyser. Därtill kan innehållet bearbetas och presenteras på olika sätt tack vare att det semantiska innehållet och dess form har särskiljts vid beskrivningen. Att förbereda ett material för ett på förhand bestämt ändamål, till exempel för publicering i XML eller HTML-format på webben, är en mycket snävare uppgift.

I slutet av avhandlingens kapitel 2 resonerar jag om öppen källkod och öppna standarder. *Open Source* och *copyleft*-rörelsen är besläktad med min ambition att renodla innehållsdesignaspekter, frikopplade från frågor om på vilka plattformar och med hjälp av vilka programvaror innehållet skall hanteras och presenteras. Jag argumenterar också för att fritt tillgängliga metoder och verktyg skall användas i mesta möjliga mån, och hävdar att öppna standarder och öppna program är bra eftersom de kan medverka till funktionella lösningar och en beständig infrastruktur av digitala material för forskning och undervisning. Där är avhandlingen av ganska normativ ton, eftersom jag förespråkar vissa slag av metoder, verktyg och angreppssätt, som exempelvis den hierarkiska dokumentmodellen.

Med en mindre normativ uppläggning skulle avhandlingsarbetet ha kunnat ägnas åt exempelvis en metodologisk studie av ett urval prosopografiska tillämpningar där tillvägagångssätt och metoder analyserats ur olika perspektiv. Eller en för en mer allmänt hållen metodologisk analys av exempelvis TEL:s riktlinjer. Istället valde jag ett angreppssätt där vägen till det ”färdiga” empiriska materialet i hög grad är själva resultatet. En fördel med min uppläggning är att jag kunnat behålla fokus på hur den hierarkiska dokumentmodellen kan tillämpas. En annan fördel är att avhandlingen resulterat i tillämpningar som faktiskt har använts för forskning och i någon mån undervisning, exempelvis de register

som ligger till grund för den beskrivande statistiken, korrespondensanalyserna och nätverksanalyserna. Dessutom har avhandlingen resulterat i materialiserade webbtillämpningar och databaser, som kan användas av andra forskare eller i undervisningssammanhang.

En nackdel med uppläggnigen är att alternativa tillvägagångssätt blivit rudimentärt och översiktligt behandlade. Här tänker jag framför allt på extern annotering, Topic Maps och olika användningar av relationsdatabaser eller objektsdatabaser.

Dock menar jag att, med sina begränsningar, min studie om hur innehållsdesign kan användas för att beskriva, ordna, lagra och bearbeta digitala material avkastat kunskap som är brukbar redan i dag. I en framtid kan man tänka sig att det slags innehållsdesign som jag utforskat kan bidra till realiseringen av det som Ivan Illich beskrivit som *nätverk av information* som skapar nya möjligheter till lärande som inte kanaliseras via skolor och lärare.⁷³⁹ Detta lärandenätverk (*learning web*) – som Illich formulerade sig åratals innan webbens tillkomst – skulle innehålla en rad olika verktyg för utbyte, samarbete, interaktion av innehåll såväl som färdigheter och där den viktigaste faktorn var tillgänglighet till innehåll (*learning objects*) och till andras kunskap.⁷⁴⁰

8.2 Återanvändningens aspekter

Möjligheter till återanvändning av det digitala materialet är ett annat argument för, men även ett resultat av, innehållsmärkning med XML. På flera ställen i avhandlingen har jag illustrerat hur materialet kan göras utbyggbart och modifierbart och dessutom flyttbart mellan plattformar. Filformatet XML leder också till att andra kan bruka materialet utifrån egna önskemål under förutsättning att det görs tillgängligt i brukbara format. Ett steg i den riktningen är att sätta själva innehållet i centrum för beskrivningen, snarare än den applikation eller de funktioner man för tillfället anser sig behöva. Detta har i möjligaste mån gjorts i det arbete som redovisas i avhandlingen i och med att den databas (FFO-databasen) som upprättats visserligen primärt styrts av det berörda forskningsprojektets (Formering för offentlighet) frågor och behov, men utformats så att innehållet kan migreras till andra plattformar och återanvändas i sammanhang forskningsperspektiven och frågorna är annorlunda. Ett exempel är verktygen för olika sorters textanalyser av digitala material.⁷⁴¹ Ett annat exempel är möjlig-

⁷³⁹ Ivan Illich, *Deschooling society*, World perspectives, 44 (New York: Harper & Row, 1971); Ivan Illich, "Education without school: How can it be done", *The New York Review of Books* 15.12 (1971).

⁷⁴⁰ Ett exempel på sådana lärande nätverk är de *wikis* som finns för alla möjliga ämne och inom olika gemenskaper. I en *wiki* finns tillgång till innehåll och vägledning tillika möjligheter att kommunicera med varandra direkt i *wikin*. Det finns inga "garantier" om kvalitet på de kunskaper som presenteras. Wikipedia är troligen den mest kända *wikin*, se <<http://wikipedia.org/>>.

⁷⁴¹ T.ex. verktyg tillgängliga vid TaPOR textanalysportal <<http://portal.tapor.ca/tools/>>.

heterna att lägga till nya data, till exempel uppgifter om de data som ingår i databasen (pionjärkvinnornas publikationer, tidningsartiklar och inlägg i den offentliga debatten). Det är därtill fullt möjligt att filtrera bort all märkning som jag infogat och på detta sätt rensa innehållet från den beskrivning, eller snarare de metadata, som tillförts, samtidigt som detta innehåll fortfarande är läsbart för människor eller datorer.

Dessutom erbjuder en djup eller rik beskrivande märkning, framtida möjligheter att använda samma märkning för andra syften än de ursprungligen förutsedda. Det är enklare att ta bort märkning och information som visar sig överflödigt än att i efterhand tillföra mer märkning. Vidare öppnar märkningen möjligheter till olika ingångar i materialet eller att skapa olika lager. Väsentligt är också att märkningen tillåter att man extraherar delar ur en digitaliserad samling utan att det sammanhang vari delarna ingår i går förlorat.

En annan aspekt av återanvändbarhet är lagringsformatet. Digitala samlingar behöver underhållas och uppdateras precis som materiella samlingar. Därför bör det format som används för lagringen vara så universellt och plattformsoberoende som möjligt. XML-formatet lagras som vanligt textformat vilket medför att filerna är läsbara på alla datorplattformar oavsett i vilken editor filerna ursprungligen skapades i.⁷⁴² En framtida användare behöver inte tillgång till den ursprungliga dokumentmodellen eller schemat för att läsa filen, däremot eventuellt för att förstå hur märkningen genomförts och vad den beskriver.⁷⁴³

Det valda tillvägagångssättet har flera fördelar än de som ovan framhållits. I och med att innehållet beskrivits och ordnats enligt TEI/XML finns alla möjligheter att migrera det till en relationsdatabas, eller andra tillämpningar, något jag också påvisat i den tillämpade delen av avhandlingen. Om jag däremot istället för XML-formatet hade använt en relationsdatabas som utgångspunkt, hade materialet inte kunnat användas i sin "ursprungliga" form, exempelvis som avskrifter av arkivposter eller som läsbara biografiska notiser – som i projektet FFO kallats "kapitalbeskrivningar", eftersom tonvikten ligger på de biografierade personernas tillgångar, det vill säga "kapital" i Pierre Bourdieus mening". För att åstadkomma detsamma i en relationsdatabas hade jag behövt lagra en mängd redundant information, eller göra varje kapitalbeskrivning till en egen post, vilket inte är förenligt med en god konstruktion av en relationsdatabas. Jag har också strävat efter att inte skapa verktyg som är begränsade till att tjäna som instrument enbart för inmatning och registrering av uppgifter, vilket annars ofta är fallet i prosopografiska tillämpningar.⁷⁴⁴

Slutligen vill jag betona att den faktor som påverkar återanvändningen i stor utsträckning är den beskrivande märkningen. Den medför framför allt att den tillfogade informationen blir explicit och därmed även möjlig för andra att kriti-

⁷⁴² Jämför med html-formatet där det heller inte har någon betydelse för presentationen med vilken dator eller med vilken editor filen är skapad.

⁷⁴³ Som tidigare nämnts försöker man i ett schema namnsätta element och attribut samt dessa värden på ett sätt som förstås av människa såväl som maskin.

⁷⁴⁴ Exempelvis de stora prosopografierna vid King's College som presenteras på s. 138 f.

sera, förändra, utvidga eller minska. Alla texter är märkta dels på en maskinnära nivå för att det behövs teckenkodning så att innehållet kan lagras och presenteras med hjälp av datorer, dels är de oftast implicit ”märkta” till exempel genom vårt sätt att arrangera dem i kapitel, stycken och ord. Detta sätt att skriva och läsa hierarkiskt och sekventiellt ordnad text är resultat av gamla praktiker och normer och något vi lär oss i unga år. Det finns dock många andra sätt att representera text, där den sekventiella presentationen av text är en vanlig variant, något jag diskuterar i nästa avsnitt.

Genom den beskrivande märkningen har alltså ytterligare en beskrivningsnivå tillförts innehållet, något som i sig är av värde inte minst av bevarandeskäl och för att garantera möjligheter till återanvändning. Problemet vid återanvändning är att finna verktyg som inte enbart beskriver dokumentets interna strukturer och relationer, utan som samtidigt kan hantera det semantiska innehållet och dokumentmodellen.⁷⁴⁵

Den beskrivande märkningen blottlägger, eller åtminstone synliggör, de underliggande, teoretiska eller pragmatiska, antaganden som använts för beskrivning och klassificering av innehållet. Texten tillsammans med den infogade informationen om texten, manifesterar därmed kunskapsorganisationens utgångspunkter som i sin tur har en manipulerande inverkan på presentationen och användningen av materialet. Det material som beskrivits och utformats med hjälp av väl avgränsade frågeställningar och verktyg, som exempelvis de nämnda kapitalbeskrivningarna, påverkas oundvikligen av detta faktum. Vad som kan vara värdefullt att ha i åtanke är att många produkter, må det vara digitala material eller andra sorters framställningar, inte ackompanjeras av explicita beskrivningar av varför en viss klassificering använts eller vad som styr insamlingen av material. Detta ger konsekvenser för återanvändning inom andra områden än de från början avsedda. Om generella och explicita beskrivningar saknas förblir det digitala materialet bundet till den lokala logiken. I exempelvis undervisningssituationer är det förstås till en viss gräns genomförbart att använda ett material utan tillfogad explicit beskrivning men utan karta blir det svårt att söka sig utanför de redan iordninggjorda stigarna.

Konkreta exempel på den sistnämnda svårigheten finner den som försöker utnyttja flertalet av de ämnesportaler – av skilda slag men som har det gemensamt att de tillhandahåller material för forskning såväl som undervisning – som erbjuder flera, men förutbestämda, ingångar i materialet. De kvinnohistoriska samlingarna vid Göteborgs universitet som nämns i början av avhandlingen tillhandahåller på webben flera innehållsrika och i sitt slag utmärkta portaler avsedda för forskare, lärare och studerande. De där presenterade ämnena är tematiskt indelade och det rika digitaliserade källmaterialet, exempelvis tidskrifter och offentliga utredningar, är som i viss mån sökbart, men framför allt är

⁷⁴⁵ RDF framförs som ett verktyg som utvecklats för webben som ett bra alternativ. Dino Buzzetti & Jerome J. McGann, "Editing in a digital horizon" *Electronic textual editing*, red. Lou Burnard, et al., (New York, N.Y.: Modern Language Association of America) s. 63.

det tillgängligt. Dock är det mycket svårt att återanvända detta material på annat sätt än så som det presenteras eftersom det är ”hårdkodat” i form av HTML-märkning eller i PDF-format som inte medger egna bearbetningar eller andra ordningar.⁷⁴⁶ Om lärare och studerande vill fördjupa sig eller utforska materialet på annat sätt blir det stopp. Dessutom kan det förhålla sig så att det underliggande materialet är rikt och fullt av intressanta och användbara uppgifter, som inte syns på presentationsnivån eller också har funktionerna begränsats i enlighet med några i förväg valda angreppssätt.

Ett exempel på det motsatta, det vill säga en mångsidigt användbar webbplats, är den demografiska databasen, också den omnämnd i avhandlingen. Där erbjuds användare tillgång till ett rikt material, därtill med tydligt angivna utgångspunkter. Det finns bearbetade och iordningställda introduktioner, förslag på möjliga studier och lärarhandledningar och användare kan söka i de delar av databasen som är fritt tillgängliga. Det går att spara resultatet av en sökning och ladda ner det till sin egen dator. Därtill finns möjlighet för forskare, lärare och studerande att få tillgång till ytterligare register för vetenskapliga studier.⁷⁴⁷ I detta fall är alltså materialets återanvändbarhet mycket god tack vare den modulära designen men även tack vare att de vetenskapliga utgångspunkterna är klart formulerade.

Dessa två exempel, de kvinnohistoriska samlingarna och den demografiska databasen, illustrerar två motsatta sätt att ta sig an de pedagogiska utmaningar som följer med utformningen av forskningsmaterial – utmaningar som ställer större krav med ökande ambitioner att göra material tillgängligt både på webben och via andra media, och för olika användargrupper. Slutsatsen är att största möjliga klarhet i utgångspunkterna och minsta möjliga restriktioner vad gäller framtida användningsområden bör eftersträvas.

8.3 Innehållets ordning

En av de klassiska ambitionerna i innehållsdesign av digitalt material är att genom märkning särskilja innehåll från form. Är det ens möjligt att särskilja innehållet från dess form? I avhandlingen visar jag genom beskrivning av märkspråkens grundläggande antaganden och teoretiska utgångspunkter samt med exempel på märkning och tillämpningar att det är möjligt.⁷⁴⁸ En användbar modell för beskrivning av skillnaden mellan innehåll och form är Aarseths definition av textens funktion såsom bärare av information.⁷⁴⁹ Han menar att text byggs upp av teckensträngar med två modaliteter, dels såsom de framträder för en använ-

⁷⁴⁶ Se <www.ub.gu.se/kvinn/portaler/>.

⁷⁴⁷ Se <www.ddb.umu.se/>.

⁷⁴⁸ En av dessa utgångspunkter är den hierarkiska dokumentmodellen som presenteras på s. 77 f. och problematiseras på s. 113 f. En annan är Aarseths modell för cybertexter, som diskuteras nedan.

⁷⁴⁹ Se s. 78 för en presentation av Aarseths modell.

dare (*scriptons*), dels såsom de är arrangerade i texten (*textons*).⁷⁵⁰ Med hjälp av beskrivande märkning kan vi både hantera arrangerandet av *textons* och förbereda för användningen av *scriptons*.

En digital samling texter av det slag som färdigställts i samband med detta avhandlingsarbete kan beskrivas på en skala från statisk till dynamisk beroende av funktionaliteten och möjligheter till manipulation. Om funktioner för ändring av *textons* eller olika transversala funktioner möjliggörs blir texten dynamisk, något som Aarseth benämner *textonic*.⁷⁵¹ Ett exempel på en text som enligt Aarseths texttypologi kan anses som explorativ är det webbgränssnitt jag skapat för sökning och hämtning av innehållet i FFO-databasen. En användare kan själv påverka vilket innehåll som presenteras beroende på utformning av sökfrågan och därmed själva välja vägar genom innehållet. Det finns dock vissa begränsningar inbyggda i systemet genom att sökfunktionen innehåller vissa förvalda värden, vilket både begränsar och stödjer användarens möjligheter att arbeta explorativt. Förvalda kategorier är ofta till stor hjälp när sökfrågor formuleras, speciellt om användaren är obekant med materialet. En användare, säg en studerande, som första gången möter ett historiskt material kan inte förväntas formulera sökfrågor som speglar innehållet. Genom de förvalda kategorierna synliggörs utformningen av materialet, såväl som innehållet, och därmed ges användaren en chans att välja mellan olika innehållsobjekt och vägar in i materialet. I FFO-databasens sökgränssnitt används bland annat utbildning, socialt ursprung och yrkestitlar som parametrar för sökning (istället för andra tänkbara alternativ såsom kvinnornas eventuella publicerade skrifter, fritidsintressen eller hänvisningar till källmaterial).⁷⁵² Detta speglar tydligt vilka utgångspunkter som styrt utformningen, och därmed klassificeringen, av innehållsobjekten, något som också påverkar möjliga användningar av och vägar in i materialet.

Ett annat exempel från utvecklingsarbetet i anslutning till avhandlingen är tillämpningen för de digitaliserade arkivposterna. Där kan användaren föra muspekaren över fotografiet så att den avbildade boksidan tillsammans med den avskrivna texten syns i webbläsaren, medan arkivpostens rubriker presenteras bredvid boksidans avbildning. Denna applikation är med sin relativt ringa interaktivitet en mer statisk text än den förra, men samtidigt försedd med en annan sorts explorativ funktion.⁷⁵³ Tack vare den beskrivande märkningen har således texter försetts med funktioner som gör det möjligt att skapa mer eller mindre dynamiska varianter av källorna. I det sist nämnda fallet kan en användare inte välja väg genom materialet, men får i gengäld tillgång till funktioner för kontroll av överensstämmelsen mellan transkribering och avbildad källa, samt hjälpmedel för läsning av den handskrivna texten. Här är utgångspunkterna möjligen inte lika genomskinliga, ej heller den avsedda användningen av det

⁷⁵⁰ Aarseth, *Cybertext: perspectives on ergodic literature* s. 70.

⁷⁵¹ Ibid. s. 72.

⁷⁵² Se exempel på webbarkivet s. 222 f. Det är fullt möjligt att använda andra och fler parametrar i sökgränssnittet.

⁷⁵³ Se exempel på arkivposternas bilder på s. 216 f.

material som presenteras. En tillämpning av det här slaget kan ändå dock vara ett gott komplement till en databas (vi bortser här från fallet då bildpresentationerna av faksimil är huvudsaken).

På vilket sätt befrämjar eller motverkar den hierarkiska dokumentmodellen tillämpningarnas funktionalitet och dynamik? TEI- schemat och andra scheman som baseras på XML kritiserar ibland för att *en* sorts hierarkisk ordning får råda på bekostnad av andra alternativ. Bland dem som kritiserar den hierarkiska dokumentmodellen finns Claus Huitfeldt som menar att den ger upphov till en rad problem vid beskrivning av olika strukturer. Vad som är struktur och vad som är innehåll beror på *hur* texten analyseras. Då TEI:s riktlinjer började utformas i anslutning till den då nya SGML-standarden (och därmed den hierarkiska dokumentmodellen), tvingades textens många dimensioner i *en* överordnad struktur.⁷⁵⁴ Huitfeldt har en poäng i att vad som utgör struktur och vad som utgör innehåll beror på vilken aspekt som tillerkänns tolkningsföreträde.

Jag vill dock framhålla att trots att en hierarkisk dokumentmodell förvisso tenderar att ge företräde åt en enda överordnad princip, finns det ändå möjligheter att manipulera och omordna samma innehåll. Det beror bland annat på vilken betydelse och funktion den beskrivande märkningen får vid utformning av tillämpningar. Genom att, vilket skett i mitt arbete med FFO-materialet, märkningen använts som bas för extrahering av exakt de uppgifter som behövs för statistiska analyser av det samlade materialet, kan faktiskt ett icke-hierarkiskt måldokument genereras ur ett hierarkiskt startdokument. Återigen är det tack vare den beskrivande märkningen som masterfilen kan användas för flera olika tillämpningar, låt vara att andra verktyg än märkningen behövs för att manipulera och migrera innehållet. Vilken funktion en och samma märkning får är beroende av hur kraftfulla verktyg som används för manipulation av innehållet. När masterfilen används för presentation av den löpande texten i webbläsaren krävs inte lika avancerade verktyg som när ett visst innehåll skall extraheras och dessutom placeras på rätt ställe i måldokumentet. Detta reciproka förhållande beror på att verktygens inneboende egenskaper påverkar tillämpningarnas funktionalitet i lika stor utsträckning som omfattning och inriktning av märkningen. I det här fallet påverkar även källornas karaktär tillämpningarnas funktionalitet, på grund av det heterogena källmaterialet. Hur mycket beskrivande märkning som än har infogats i materialet påverkas inte en källas kvalitet, ej heller hjälper en extensiv märkning till att analysera ett material om inga andra verktyg appliceras. Omvänt är det också så att en sparsam eller ofullständig märkning påverkar utfallet av de möjliga bearbetningarna. En slutsats blir därmed att på samma sätt som man kan särskilja innehåll och form är det möjligt, och nödvändigt, att särskilja själva beskrivningen (vilken åstadkommes med märkningen) från tillämpningarnas funktionalitet.

Märkning av det semantiska innehållet kräver förstås tolkning – precis som märkning av strukturen gör. Det innehåll som beskrivs och märks kan sedan

⁷⁵⁴ Huitfeldt, "Multi-Dimensional Texts in a One-Dimensional Medium".

arrangeras om, bearbetas och manipuleras på en rad olika sätt beroende på tolkning, typ av tillämpning och vilka aspekter av innehållet som är beskrivet, märkt och kategoriserat. Detta har jag påvisat, i den tillämpade delen av avhandlingen, bland annat genom att utveckla och tillämpa XSL-skript som användes för att extrahera exakt den information (rörande märkta innehållsobjekt i kapitalbeskrivningarna) som behövdes för statistiska analyser. Den struktur som märktes upp utgick från målsättningen att beskriva möjliga indikatorer på kvinnornas egenskaper och tillgångar. Men denna märkning kan även användas till att hitta och skapa andra sorters strukturer. Till exempel beskrivning av hur stora kontaktnät kvinnan hade, något som jag också visat i tillämpningen för nätverksanalyser där extrahering av namn ur startdokumentet (masterfilen) använts för att skapa ett måldokument med information om kvinnornas kontakter med varandra inom en avgränsad grupp. Ett annat exempel ur arbetet med FFO-data är att märkningen av yrkestitlar kan användas för strukturering av innehållet utifrån typ av yrkesverksamhet istället för utifrån personernas namn.⁷⁵⁵

Dock är det förstås så att den dokumentstruktur som märks och ordnas i innehållsobjekt såsom kapitel, stycken och rader, gör att just denna struktur får företräde framför andra och därmed blir överordnad andra tänkbara strukturer. Om man använder den semantiska beskrivningen av egenskaper och tillgångar på det sätt som jag illustrerat i avhandlingen ges möjligheter för användare att utforska innehållet på flera olika sätt och med flera olika ingångar i materialet.

Ett bra exempel på en digital samling med avancerade funktioner för forskning såväl som undervisning är Wittgensteinarkivet i Bergen.⁷⁵⁶ Samlingarna innehåller diplomatariska och normaliserade utgåvor av manuskript, anteckningar och annat digitaliserat material om Wittgenstein, som föreläsningar, introducerande material, publikationer och tillämpningar baserade på dessa samlingar.

I arbetet med ganska många av de publikt tillgängliga digitala samlingarna har man ägnat årtal åt märkning varefter materialet ändå presenteras som texter med relativt få funktioner och där användaren som vill hitta och använda texterna erbjuds ingångar via några få färdiga index avseende exempelvis författare, genre, årtal och titel.⁷⁵⁷ I många fall är detta naturligtvis tillräckligt, som exempelvis för studier av äldre tryck där användare behöver studera och skriva

⁷⁵⁵ Emellanåt upplever jag att kritiken mot XML och den hierarkiska dokumentmodellen grundar sig på att man förbisett de möjligheter som finns att skapa nya måldokument med helt annan struktur än i startdokumentet. Skriptspråket, XSLT, som kan användas för denna manipulering är utformat just för *transformering* av XML-filer.

⁷⁵⁶ Se <<http://wab.aksis.uib.no/>> Innehållet i samlingarna baseras den listning Georg Henrik von Wright gjorde om Ludwig Wittgensteins manuskript som han lämnade efter sig vid sin död 1951.

⁷⁵⁷ Ett exempel på en sådan informativ digital samling inom vetenskapshistoria är *The Chymistry of Isaac Newton*, där olika slags material samlats och där man kan beskåda diplomatariska och normaliserade transkriptioner av laborationsanteckningar, index över alkemiska substanser samt Newtons avskrifter av andra källor, se <<http://webapp1.dlib.indiana.edu/newton/>>.

ut den avbildade källan tillsammans med den transkriberade texten, kanske en diplomatisk version som är ackompanjerad av en normaliserad transkription. Exempel på en sådan digital samling är Arkivet för Nordiska medeltidstexter.⁷⁵⁸ I andra sammanhang är inte presentationen av texterna det primära utan att texterna finns arkiverade och tillgängliga, som exempelvis vid *Oxford Text Archive*, vars främsta funktionalitet ligger i möjligheterna för användare att söka bland de tusentals texter som arkivet innehåller och ladda ned det man behöver.⁷⁵⁹

Den digitala samling som utformats och använts som exempel på tillämpningar i den här studien kan förstås inte mäta sig med dessa omfattande och mångåriga projekt. Men det jag kan tillföra med den här studien är exempel på märkning av prosopografiskt material, fri tillgång till masterfiler samt exempel på tillämpningar och deras utförande. Min ambition har inte enbart varit att påvisa *vad* som går att göra utan även ge exempel på *hur* det kan utföras, och vad innehållsdesign innebär i sammanhanget.

8.4 Verktygens styrkor och svagheter

Det slags prosopografiska material, det vill säga uppgifter om ett större antal personer, som utgjort underlaget för tillämpningarna i mitt avhandlingsarbete, lämpar sig väl för märkning enligt TEI:s riktlinjer. Även om materialet baseras på olika sorters källor, som skall modelleras var för sig och tillsammans, förekommer ändå tydliga strukturer, och såväl datainsamlingen som bearbetningen av innehållet är så pass komplexa företag att de svårligen låter sig genomföras helt manuellt. Men att genomföra beskrivande märkning är tidsödande. Därför måste man fråga sig om det är värt besväret.

Det bör även poängteras att arbetet med de olika momenten – beskriva, ordna och bearbeta – inte på något sätt varit så okomplicerat som avhandlingen möjligen kan ge sken av. Det är många ställningstaganden och val som måste göras under arbetet med vart och ett av dessa moment. Märkningsarbetet har inneburit att en rad olika problem uppstått med exempelvis klassificeringsprinciper, lämpliga beskrivningar, möjliga bearbetningar och tänkbara användningar. Hur kan kvinnornas föreningsaktiviteter beskrivas och märkas på ett lämpligt sätt? Vilken struktur bör vara den överordnade; datum eller föreningsnamn? Det är inte heller alltid lätt, eller självklart, att i förväg inse hur den kommande bearbetning påverkas av hur innehållet är beskrivet och ordnat. Vilka konsekvenser får en viss beskrivning av innehållsobjekten, för hur materialet senare kan användas för exempelvis statistiska analyser? Dessa och många andra frågor och problem har jag brottats med under arbetets gång. Genom flera förändringar och revideringar av verktyg och tillvägagångssätt har det inneburit ständiga analyser och överväganden, vilket även medfört god kännedom om mate-

⁷⁵⁸ Se <www.menota.org/>.

⁷⁵⁹ Se <<http://ota.ahds.ac.uk/>>.

rialets beskaffenhet såväl som verktygens funktionalitet. Den kännedom jag skaffat mig om material och verktyg under arbetets gång medverkade förstås även till att tidsåtgången för varje arbetsmoment minskade allt eftersom arbete fortskred. En annan effekt är att en liknande arbetsinsats på något annat, men likartat material, skulle ge bättre utgångspunkter vad gäller kunskaper om modellering av innehållet, troliga ställningstaganden, lämpliga verktyg och bearbetningar.

Som jag tidigare noterat vore en relationsdatabas det alternativ som skulle ligga närmast till hands, särskilt med tanke på målsättningen att göra statistiska analyser och skapa fristående moduler. Men tack vare att XML-formatet valdes kan masterfilerna användas i olika tillämpningar mer flexibelt än en relationsdatabas skulle tillåta. Ett bra exempel är att masterfilen i XML-format kan användas för att utforska olika metoder för textanalys utan ingående tekniska kunskaper om själva verktygen. De många på webben tillgängliga lättanvända verktygen avsedda för olika slags textanalyser inbjuder till ett explorativt arbetssätt. Dock behövs förstås historiska eller liknande specialistkunskaper för att tolka resultaten.⁷⁶⁰ Ett annat exempel på fördelarna med XML-formatet är hur masterfilen använts för presentation av kapitalbeskrivningarna i fulltext.

En annan anledning till valet av TEI/XML istället för en relationsdatabas är att pröva hur märkningen skulle kunna utnyttjas för att generera underlaget för bestämda statistiska analyser såsom korrespondensanalys och andra metoder tillhörande familjen geometrisk dataanalys vilka utnyttjas inom den kultursociologiska tradition där projektet hörde hemma. Ytterligare en anledning var önskemålet om att materialet skulle lagras i ett format som möjliggjorde migrering mellan olika plattformar.

XML-formatet är trots sin hierarkiska struktur och strikta form relativt flexibelt såtillvida att det tillåter många artskilda sätt att beskriva och lagra innehåll. I masterfilen är det enkelt att lägga till eller ta bort delar av innehållet utan att innehållsobjekten behöver mista förbindelsen med sin kontext. Med andra ord är det möjligt att extrahera valfria delar, eller moduler, ur en masterfil men ändå identifiera det sammanhang vari den ingår, dock under förutsättning att det finns en unik identifierare. XML-formatet är således flexibelt vad gäller hur innehållet kan beskrivas och ordnas, men även såtillvida att det är jämförelsevis enkelt att ändra i innehåll eller utformning. Emellertid hade en databas varit mer flexibel i andra avseenden, bland annat hade det varit enklare att genomföra större förändringar i datatyper eller mer genomgripande förändringar i klassificering av innehållsobjekten.

Vid en jämförelse mellan relationsdatabasens mer dataorienterade angreppssätt och märkningens mer textorienterade, kan även andra skillnader noteras. En dataorienterad innehållsdesign förutsätter att ett tänkt innehåll skall passa in i en modell som utformas *à priori*. Revideringar av själva datamodellen är inte

⁷⁶⁰ I *Figur 36* s. 215 ges exempel på hur ett webbaserat verktyg används för att visualisera innehållslets struktur och inbördes relationer.

lika lätta att genomföra som revideringar av de data som lagras. Den textorienterade modellen är mer flexibel relativt det innehåll som skall beskrivas och ordnas. Här utgår man från materialets struktur och använder den modell som, i bästa fall, överensstämmer med de innehållsobjekt man avser att beskriva. Den textorienterade modellen för dock med sig mer redundans i lagringen av data och är inte särskilt lämplig för en stringent och precis inmatning av data.

En annan nackdel med den hierarkiska dokumentmodellen är att *en* övergripande struktur ges företräde. För att hantera det problemet har jag använt en okonventionell lösning för märkning av kapitalbeskrivningarna (varmed som nämnts avses biografiska notiser med uppgifter om personers egenskaper och tillgångar). Dessa är relativa ostrukturerade till sin form (författade av forskare vilka som regel saknar kännedom om vad den kommande XML-märkningen innebär) men med en tydlig semantisk struktur (eftersom forskarna följt en mall som föreskriver att uppgifter om födelseår, faderns yrke när den biograferad personen lämnade föräldrahemmet, och så vidare skall inkluderas). Här har en empirinära märkning använts, vilket innebär att så långt som möjligt undvikande tolkningar vad gäller innehållsobjektens omfattning och karaktär. Självfallet kan inget innehåll märkas utan ett visst mått av tolkning, men det som bör eftersträvas är att det tydliggörs vilka kriterier som styr märkningen och därmed tolkningen. Exempel på en sådan empirinära tolkning är att märkningen inte använts för att beskriva de biograferades kapitalvolym, det vill säga deras samlade tillgångar i form av socialt ursprung, kulturellt kapital, socialt kapital etcetera.⁷⁶¹ Istället beskrivs och registreras *indikatorer* på sådant som senare i analyserna kan användas som mått på kapitalvolym. Med andra ord har tolkningen så långt som möjligt skjutits på framtiden, fram till tidpunkten för de mer samlade historiska och kultursociologiska analyserna. En extra bonus är att chanserna ökar att märkningen i framtiden skall bli brukbar även för analyser av annat slag än de som varit av primärt intresse inom FFO-projektet. De väl förankrade utgångspunkterna, koncentrerade i kapitalbeskrivningsmallen, har varit ett nödvändigt verktyg för en någorlunda stringent tolkning och avgränsning av vad som kan anses vara indikatorer på en viss sorts kapital och vad som inte är det.

De huvudsakliga principerna för modelleringen av det prosopografiska materialet har bidragit till det slutliga resultatet. Dessa principer kan, med tanke på att de förenar det specifika och empirinära med det generella och abstrakta, även användas som vägledning för modellering av andra slags material. Kombinationen av en projektspecifik datamodell med generella beskrivningsprinciper resulterar i digitala material som både speglar projektets inriktning men som även kan utbytas med och användas av andra. Detta tillvägagångssätt innebär också möjligheter att senare kunna lägga till andra projektspecifika datamodeller för beskrivning och klassificering av innehållsobjekt.

Här kan jag hämta stöd i proveniensprincipen, ett begrepp från arkivbildningsområdet. Principen innebär bland annat att ett arkiv ska ordnas och för-

⁷⁶¹ T.ex. utsagor om att en viss person innehade ett stort mått av kapital inom det litterära fältet.

tecknas så att det återspeglar det sätt på vilket det har uppkommit. För att åstadkomma detta måste man i processen pendla mellan å ena sidan överordnade principer och å andra sidan innehållet i de dokument och den information som ska bilda arkivet. I mitt fall hämtar jag principerna från TEI-schemat och från kapitalbeskrivningsmallen och ytterligare information från det material som projektets forskare sammanställt.

Annat som måste övervägas vid val av verktyg för att beskriva och ordna är vilken metod som skall användas för insamling av, i detta fall, prosopografiska data. För det material som rubricerats FFO-KAPITAL användes en frågedriven metod för insamling av data till skillnad från det källorienterade tillvägagångssätt som användes för insamling av FFO-HLS. Om det är viktigt att inmatning av data sker på ett stringent och precist sätt är inte samlande av uppgifter i texter tillhörande den biografiska genren, såsom kapitalbeskrivningarna i FFO-projektet, att föredra. Den frågedrivna metoden är inte heller alltid att föredra, då risken är att viss information förbises, exempelvis sådant som från början inte ansågs intressant eller viktigt att samla in. Fördelar med ett frågedrivet tillvägagångssätt för insamling är att det underlättar utforskandet av relevanta källor, under förutsättning att dessa är tillgängliga, och underlättar en pragmatisk och "eklektisk" användning av källor vilket ofta är nödvändig för att ett prosopografiskt material över huvud taget skall kunna sammanställas. Här består utmaningen i att använda och sammanställa material av disparat karaktär, men även i att tillämpa verktyg som inte i första hand är utformade för det slags material man använder.⁷⁶²

I den här studien har den uppsättning verktyg och den modell som från början användes för att beskriva det prosopografiska materialet modifierats. Anledningen var som ovan nämnts behovet att bättre kunna tillämpa den relativt det material som skulle utformas. I slutfasen av studien utvecklades dock verktyget och fungerade bättre för att beskriva materialets beskaffenhet.⁷⁶³

Svagheter med XML-formatet och den hierarkiska dokumentmodellen är att märkningen är relativt tidsödande, speciellt om märkningen bör vara riklig eller när tolkningen av en viss sorts innehållsobjekt blir problematisk. Men för en mer rättfram märkning av, låt säga, egennamn finns flera möjligheter till automatisering av arbetet. En möjlighet är att använda avancerade sök- och ersättningsfunktioner som i den här studien framför allt utförts med det verktyg som utvecklades för ändamålet. Ett annat tillvägagångssätt är att utföra märkningen i flera skilda steg, med början i märkning av divisioner, rubriker, stycken, rader etcetera som då strukturerar och gör innehållet till en sammanhängande helhet med identifierbara moduler som underlättar den djupa och semantiska märkningen av andra innehållsobjekt.

⁷⁶² Här avses framför allt märkningen av prosopografiska data i kapitalbeskrivningarna, se s. 172 ff.

⁷⁶³ Här avses utvecklingen av TEI: riktlinjer. När studien inleddes fanns ingen modul för beskrivning av prosopografiskt material. En sådan färdigställdes under 2007.

I den här studien har XML-märkningen fungerat väl för det aktuella prosopografiska materialet. En annan lösning, som exempelvis en relationsdatabas hade varit möjlig men absolut inte nödvändig för att uppfylla kraven på beskrivning, organisering och bearbetning av innehållet i prosopografin. Tvärtom, XML-formatet har visat sig vara användbart för samtliga moment som jag räknat till innehållsdesign, till att börja med för att *beskriva* och *ordna*. För *bearbetning* och *migrering* av innehåll har applikationer av XML-formatet använts, och likaså för presentationsmomentet. Därtill har ett flertal andra verktyg, som databashanterare, statistiska programvaror och webbaserade verktyg för textanalyser använts för ytterligare bearbetningar, analyser och presentationer.

8.5 Innehållsdesign för forskning och undervisning

Det övergripande syftet med föreliggande studie har varit att ge ett bidrag till metodutveckling och att föreslå tillvägagångssätt för innehållsdesign och utformning av digitala material för forskning och undervisning. Ett delsyfte har varit att avgränsa och inventera moment som ingår i innehållsdesign, och ett annat delsyfte att i tillämpningar undersöka de inventerade metodernas och verktygens begränsningar och möjligheter. Jag har strävat efter att synliggöra förutsättningar för arbetsmomenten i innehållsdesign, det vill säga beskrivning, ordnande och bearbetning av digitala material, eftersom dessa moment ibland behandlas lättvindigt vid utformning av material för forsknings- och undervisningsändamål.

Beskrivning av innehållet är ett av de grundläggande momenten i innehållsdesign. Det är, som tidigare påpekats, tack vare beskrivningen av innehållsobjekten som materialet kan bearbetas, användas och presenteras. Beskrivningen medverkar dessutom till att göra underliggande antaganden om utformningen av innehållet explicita. I praktiken innebär det att generella eller specifika beskrivningsmodeller, eller kombinationer av dessa, används, vilket i sin tur innebär att berörda forskare, lärare och studerande i förväg behöver tänka igenom vilket slags beskrivning av innehållet som krävs. En beskrivningsmodell förutsätter dessutom underliggande antaganden om vad själva innehållet är, en ontologi om man så vill – med andra ord antaganden om vad som finns eller inte finns i det system av innehållsobjekt och dessas relationer som skall beskrivas. Vilka aspekter skall lyftas fram? Vilka skall stå tillbaka? Vilka strukturer kan urskiljas? Genom beskrivningen kan innehållsobjekt grupperas och samtidigt särskiljas från varandra, precis som i Aaresths texttypologi, där *textons* särskiljs från *scriptons* varigenom dynamiska texter särskiljs från statiska. Andra exempel är den kapitalbeskrivningsmodell som här har använts för att identifiera indikatorer för sådana egenskaper och tillgångar som kan relateras till Bourdieus kapitalbegrepp och avgränsa dem från andra uppgifter som inte är av intresse i sammanhanget.

Beskrivningsmomenten menar jag kan användas för att förfinas och förbättra material för forskning och undervisning, genom att göra de underliggande antagandena om utgångspunkter, teoretiska eller pragmatiska, synliga. I undervisningen skulle ett väl beskrivet forskningsmaterial kunna användas som underlag för kritiska analyser av just underliggande antaganden om materialets beskaffenhet och struktur. Vad har man valt att spegla? Varför har vissa indikatorer försetts med beskrivande märkning medan andra lämnats därhän? Vilka innehållsobjekt har använts för att underbygga de resultat som framställs? Kan några tydliga tendenser i urval och värderingar urskiljas? Vilka uppgifter används för att beskriva en person och hur avspeglas detta i märkningen? Därtill kan ett väl beskrivet material tjäna som goda exempel på tillvägagångssätt eller som utgångspunkt för vidare analyser och märkning av helt andra aspekter än de som är synliggjorda.

En effekt av beskrivningen är att man kan överblicka hur såväl innehållet i dess helhet som de specifika innehållsobjekten har organiserats i ett givet material. Genom klassificeringar avgörs vilka innehållsobjekt som grupperas ihop och vilka som skiljer sig från varandra. Även här kan olika metoder användas och valet står mellan etablerade klassificeringssystem, revideringar av klassificeringar eller lokala och innehållsspecifika klassificeringar. I studien har en kombination av dessa tre alternativ använts, något som var nödvändigt på grund av att enbart de etablerade klassificeringssystemen inte räckte till för att identifiera och avgränsa de aspekter som skulle analyseras. Det gällde i synnerhet skillnader i de biograferade personernas yrkesverksamheter och utbildningsstrategier. Då materialet handlar om en tidsperiod, 1880-1920, då rådande klassificeringssystem inte avspeglade kvinnors offentliga verksamheter var det nödvändigt att modifiera etablerade klassificeringar eller uppfinna egna. Detta ger förstås metodiska begränsningar framför allt med tanke på resultatens jämförbarhet med andra liknande studier eller överförbarhet till andra sammanhang. Men eftersom beskrivningen och ordningen av innehållsobjekten utgår från det som faktiskt står i texten är det fullt möjligt att applicera andra ordningar och klassificeringar än dem som tillämpats i den här studien.

I undervisningssammanhang skulle ett på angivet sätt märkt prosopografiskt material kunna användas till att förse studerande med uppgifter som att utforma andra klassificeringar och ordningar eller granska hur presentationer och analyser skiljer sig beroende på hur innehållet ordnas och klassificeras. Vilka ingångar i materialet kan konstrueras med utgångspunkt i hur det är organiserat och klassificerat? Vilka nya variabler kan skapas med utgångspunkt i de innehållsobjekt som är beskrivna? Vilka parametrar är mest användbara för att särskilja olika grupper av kvinnor vid sökning i databasen eller i nätverksanalyserna? Vilka utbildningsbanor eller yrkesbanor, för kvinnor och för män, finns inte kvar idag? Hur kan utbildningsanstalter och yrken som gått ur tiden beskrivas och klassificeras? På vilka sätt påverkas analyserna av de klassificeringar som används?

Nästa moment som innefattar bearbetning och manipulering – och de därtill nära knuta förberedelserna för presentation – är det som kräver mest uppfinningsrikedom. Här skall nu verktygen sättas i arbete med det innehåll och material man har för handen. Ofta handlar bearbetnings- och presentationsmomenten om att lösa specifika uppgifter eller skapa specifika funktioner. Det är register som skall iordningställas som underlag för statistiska analyser, presentationer eller söksidor för webben som ska designas eller databaser som skall sättas ihop. Ett material som försetts med genomtänkt beskrivande märkning ger fler möjligheter till att lösa uppgifter eller att skapa nya funktioner.

Här kan man tänka sig att i undervisningssammanhang problematisera variablernas utformning och innehållet i registren, de statistiska analyser som utförts eller kan utföras med variablerna tillhands. På vilka sätt bidrar framställningar som tabeller, diagram eller nätverksgrafer till att ge överblick av innehållet? Vilka andra tillvägagångssätt kan användas? Hur hanteras kategorier som ”övriga” vid bearbetning och i analyser? Eller frågor om hur teoretiska utgångspunkterna speglas i materialet presenteras på webben. Vilka tilläggsmaterial, som exempelvis kopplingar till besläktade projekt och material, kan tillföras presentationerna och hur kan det genomföras?

Dessa och många andra frågor skulle även kunna tjäna som utgångspunkter för vidare utforskning om kunskapsorganisation i allmänhet och betydelsen av innehållsdesign i synnerhet. Inom utbildningsforskningen skulle sådana studier av innehållsliga aspekter kunna utgöra komplement till de betydligt vanligare studierna av kommunikation, interaktion och lärprocesser.

8.6 Avslutande kommentarer

Erfarenheterna från den här redovisade studien visar att det är möjligt att urskilja och avgränsa olika centrala moment inom innehållsdesign. Jag har påvisat hur dessa moment kan kopplas till principer, metoder och verktyg som är lämpade för forskning och undervisning inom samhällsvetenskap och humaniora. Jag har också funnit belägg för att den hierarkiska dokumentmodellen är användbar – och i många avseenden till och med överlägsen andra alternativ – för beskrivning, ordnande och bearbetning av uppgifter från det slag av historiska källmaterial som används inom projekt i stil med *Formering* för offentlighet.

De tillämpningar jag presenterat i tidigare kapitel utgör exempel på metoder och verktyg ägnade för uppbyggnaden och studier av utbildningshistoriska och prosopografiska material. I detta diskussionskapitel har jag diskuterat begränsningar och möjligheter som sådana metoder och verktyg erbjuder. Min slutsats är att möjligheterna väger tyngre än begränsningarna. Inte minst eftersom dessa metoder och verktyg tvingar fram ett tillvägagångssätt som blir mer stringent och som gör underliggande antaganden explicita, vilket gäller hela arbetsgånge från modellering av innehållet till presentationen av materialet.

Samtidigt illustrerar tillämpningarna att det gäller att vara pragmatisk. Jag har utforskat principerna, metoderna och verktygen genom att sätta dem i arbetet och funnit att en hel del revideringar behövs. Min ambition har inte varit att bara beskriva och utforska ett antal principer, metoder och verktyg utan även, om möjligt, generera nya kunskaper om hur vi kan utforma och använda digitalt material för forskning och undervisning. Jag har inte gett mig in i mer omfattande diskussioner eller analyser av lärande eller undervisningens inriktning utan hållit mig till att utforska metoderna och verktygen *per se*.

En viktig aspekt som nästan inte alls behandlats i avhandlingen är frågor om upphovsrätt och vad digitaliseringen kan innebära för tillgång till och publicering av liknande digitala material. Denna diskussion ligger dock utanför arbetet i den här studien men kommer förstås att kräva framtida insatser i form av studier av lagstiftning och praxis samt tillämpning av riktlinjer.

Att jämföra denna avhandling med andra metodstudier inom ämnen som pedagogik eller utbildningshistoria är inte helt enkelt. Där har jag knappast funnit några liknande studier, fränsett några undantag.⁷⁶⁴ Däremot har jag kunnat positionerat mitt arbete i förhållande till forskningen inom området *Humanities Computing*.⁷⁶⁵ Samtidigt som min studie är ganska teknisk till sin natur är den en deskriptiv undersökning av tillämpningar med analytiska och kritiska inslag.

Vilken användning kan resultaten och tillämpningarna få? Min förhoppning är att avhandlingen kan vara till någon nytta för forskare, lärare och studerande som vill utforma och återanvända och arbeta med befintliga digitala material eller skapa egna. De principer, metoder och verktyg som jag behandlat är, oavsett vad man tycker om dem, centrala i den nu pågående utvecklingen, och som ytterligare åskådningsmaterial görs tillämpningar och verktyg tillgängliga på den CD som medföljer avhandlingen och på webben.

Avhandlingen är bitvis ganska teknisk till sin karaktär. Till den läsare som tycker att det hela verkar svårt vill jag avslutningsvis säga att det inte är nödvändigt att behärska alla de moment som ingår i innehållsdesign för att kunna dra fördel av beskrivande märkning:

⁷⁶⁴ Eftersom avhandlingen fokuserar på metodutveckling kan det vara svårt att finna relevanta liknande avhandlingar då metodutveckling kan "gömma" sig i andra typer av studier. Några exempel på relaterade svenska studier är: Dahlström, *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion*; Johansson, *Ådalen 1931 - minuterna som blev historia: metodbok till en digital dokumentsamling*; Larsson, *Från adlig uppföstran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866*. Andra undantag är exempel på pedagogiska studier om undervisning med digitaliserade arkivmaterial, se t.ex. Grensjö & Sutter, "Forskande lärande i skolans och arkivens värld"; Jonas Liliequist, "Arkivet i undervisningen eller hur man kan få studenter att intressera sig för äldre handskrivet källmaterial" *Arkiven i forskningens tjänst: 20 år med Forskningsarkivet vid Umeå universitet*, red. Göran Larsson, (Umeå: Forskningsarkivet Umeå universitet, 2003).

⁷⁶⁵ Här avses det område som identifieras och av framför allt Willard McCarty professor i *Humanities Computing*, se Willard McCarty, "Humanities Computing: Essential Problems, Experimental Practice", *Literary and Linguistic Computing* 17.1 (2002); McCarty, "Knowing true things by what their mockeries be: modelling in the humanities"; McCarty, *Humanities computing*.

Creating an electronic edition is not a one-person operation; it requires skills rarely if ever found in any one person.⁷⁶⁶

Om det hade varit nödvändigt att ensam behärska alla moment i förväg, skulle framställningen av de flesta nu existerande digitala samlingar aldrig ha kommit igång. Med denna studie hoppas jag att ha demonstrerat att det är fullt möjligt för intresserade forskare, lärare och studerande att utforma digitala material för fler ändamål än bara enkel webbpublicering eller utskrift på papper. Innehållsdesign ger möjligheter att utforma digitala material för såväl forskning som undervisning och som tack vare den modulära och väldefinierade uppbyggnaden kan användas utanför det sammanhang där det från början producerats. Och som därmed kan bidra inte bara till utveckling av mer mångsidigt användbara digitala arkiv utan även till nya forsknings- och undervisningsmetoder.

⁷⁶⁶ Shillingsburg, *From Gutenberg to google: electronic representations of literary texts* s. 94.

Summary

Content design. Principles, methods, tools and applications in history of education

This thesis is about content design – an area which encompasses the practices and the conceptions of description, organisation, manipulation and presentation of digital content. The point of departure in this study was to identify and investigate principles, methods and tools applicable for content design of digital materials for research and teaching in history of education.

Background

During the last decades, information technology has changed the research practices for many scholars. One important aspect is the augmented opportunities for cooperation and exchange within research communities where the web has transformed the way in which scholarly knowledge may be produced and disseminated. The transformation has been most notably in the sciences and humanities since the adoption of digital technologies for research, analysis, communication, and teaching, has been rather slowly introduced in the social sciences, even though there has been some progress in these fields lately. This development is of immense interest and importance for anyone concerned with how to acquire, preserve, and make accessible content descending from different research practices.

One problem with the growing amount of digital archives is that content often is moulded into proprietary and application dependant formats and therefore difficult to retrieve, access and reuse in any other context than that in the one in which they are produced.

Fifteen years ago, individual researchers, research groups and archivists typically used proprietary non-portable formats for their depositories of digitalized sources, for example in the humanities electronic versions of printed texts, transcribed manuscripts or language corpora. Today the *de facto* encoding standards, such as the *Text Encoding Initiative*, TEI, within the humanities, permits scholars to create, use and share collections of well structured high quality digital sources. TEI was originally developed as an application of SGML – an interchange language for textual data – but is now fully implemented in the XML-format – a subset of SGML designed for describing and sharing structured data. This interdisciplinary effort has been very successful and is now widely

recognised as the standard format for scholarly text encoding and textual interchange. The TEI community, have had an impact on the emerging mark-up initiatives far beyond its own scholarly domain by the contribution to the development of XML, the dominant standard for information interchange format for web services.

The same kind of development has taken place within most scholarly domains, as well as in many industries but not in the same extent in the teaching practices in higher education. This means that the educational milieus are lagging behind the research communities. However, most teachers and students do not take part in this development. They are still referred to printed material, poorly structured HTML pages or non-reusable PDF files. The qualified digital sources they encounter are more often than not only available on proprietary systems and formats that may function well as long as the material stays on a particular website. Students and teachers will meet problems as soon as they try to incorporate material into their learning environment, for example in order to accomplish projects or to create tailor-built courseware. Course content is, thus, not sufficiently integrated into the learning environment. Content issues are often neglected and there are fewer studies that investigate and explore how technology might help researchers and teachers to use and reuse their materials. In other words to work with the *actual content* rather than proposing a special teaching method or instructional technique.

An important task then, would be to explore how we can make information applicable for different purposes and uses in research and educational practices. This might be achieved by a systematic exploration of content design for research with the auxiliary practical goal that such understanding could be valuable for educational practices in higher education. Namely in terms of how teachers and students could be given opportunities to separate what is learnt from how it is taught.

If the content is separated from its presentational forms one can easily imagine how the same information could be used in many different occasions and for several different purposes in research and education. Similarly, if the content is digitally produced and stored with no predestined use or predefined teaching method the same content could be used and reused in different teaching and learning situations. Consequently, this also means that the same original content might be personalized, accessed and manipulated by teachers and students at different educational levels or different disciplines and subjects.

Above I have argued that there is a need for archives of content modules on the web - easy to share, to navigate, and to combine and reuse in new contexts as opposed to proprietary file formats. If such archives are to be beneficial in various research and educational settings, and if they are to stimulate freedom of choice in higher education, it is crucial that the content is adapted to emerging international agreements and standards on mark-up languages and metadata. Otherwise, the content will be locked into proprietary platforms and applications.

In situations when digital archives, educational systems and resources proliferate, the need for standardisation becomes apparent. As in any standard driven initiatives, standardisation applied to research, scholarly communities and educational settings should be focused towards enabling, reusing and interoperability among heterogeneous computer platforms and software systems.

To achieve this, consensus is needed on a number of issues e.g. architectures, services, protocols, data models, classification schemes and interoperability between different standards – *de facto* standards or established by official standardisation organisations. In this study I will investigate and explore the most important de-facto standards for preparation, preservation and interchange of digital content for research and teaching, in the humanities and the social sciences in general. More specifically it will be an exploration of the preparation and application of prosopographical method and material for studies in history of education.

The prosopographical method could briefly be described as a kind of collective biography used both in historically oriented studies of ancient, medieval and modern societies and both in sociologically oriented studies of modern and contemporary societies. It is a kind of collective study of common characteristics of groups of individuals, most often historical actors but in sociological studies also contemporary individuals. By means of comprehensive data collection and thorough investigations on common background characteristics such as social origin, occupation, inherited and acquired positions in society, education, place of residence, religion, economical status etcetera, scholars explore and analyse the collected material. The result, that is the prosopography itself, will be a rather complex collection of different kinds of data which have to be organised in order to be applicable. This complexity is one good reason for using prosopographical material for an explorative study on content design. Another reason was the challenge to compile and model the heterogeneous source material.

Object of study and research questions

Given the present possibilities in tools and methods for content markup aimed to separate the structure and the organisation from the actual content, my overall aim was to study how to make digital content applicable for different purposes and uses in research and educational practices. In trying to reach that aim I explored, through different kinds of source materials and by prosopographical applications for studies in history of education, the possibilities of content design in such an area. The focus of my explorations was principles, methods and tools developed mainly within the Digital Humanities communities and to some extent also in the social sciences.

The overall aim was to contribute to methods development by proposing modes of procedure in content design for research and teaching. Often the prerequisites for adequate content design – to describe, organise, manipulate

and present – are not accounted for or trivially treated during working processes of preparation of digital content. As a result, digital content of many kinds are at risk to be tied into proprietary platforms and specific modes of presentational options. Thus, a secondary aim was to delimitate and identify the fundamental components – principles, methods and tools – of content design for digital materials in humanities and the social sciences. Another aim then, was to investigate the limitations and opportunities of the identified methods and tools with applications for studies in history of education.

These interrelated but yet discrete aims were guided by the following questions:

- In what way can the components of content design be delimited and how can they be characterised?
- What kind of principles, methods and tools are appropriate for content design within the humanities and the social sciences?
- In what way could these principles, methods and tools be used in applications for studies in history of education?
- What kind of possibilities and limitations could be discerned for these methods, tools and applications?

The first two questions were mainly used for the inventory and analyses of principles, methods and tools for content design. The last two questions mainly guided the experimental exploration of applications based on source materials collected for prosopographical studies.

The prosopographical applications explored in this study were based on two different types of source materials. Firstly, I used source materials collected and created within the research project *Formation for the public sphere. A Collective Biography of Stockholm Women 1880–1920*. The main purpose of this project was to investigate the social strategies of the first generations of women entering the public sphere in Sweden.

Secondly, I collected materials from the comprehensive archive on *Royal Advanced Seminary for Women Teachers*, the first official seminar for female teachers in Sweden which operated during the period 1861–1943.

My primary focus was uses and applications of tools and methods for preparation of text-based resources in general – since text is the dominant form of communicating knowledge in research and education – although the main foci were on prosopographical materials.

Methods and tools for content design

The study's inventory of methods and tools resulted in an historical exposé of the development and an analysis of descriptive markup which have been used for managing electronic texts since the late 1960's. In the early days the markup of texts was based on the procedural method, where processing instructions for were inserted in the texts. Nowadays the dominant method for text encoding is

descriptive markup which is built on the notion of separating the description and the processing of the content.

The most common method for content markup of textual content emanate from the textual theory of the hierarchical order of content objects in texts. This principle implies that one superior structure precedes all other possible structures when describing a text's content. Usually documents are divided into sections, paragraphs, lines, words etcetera down to the smallest division of semantic meaningful parts. Besides such a structure there are of course other meaningful structures in any give text. In this study the main focus has not been to markup the documents inherent structure or ordering of paragraphs. Here the task has been to markup indicators for what could be counted for as different kinds of capital both in an objective sense, both relative the investigated group of women.

The first descriptive markup language which gained any spread uses was the *Standard Generalized Markup Language*, SGML, a comprehensive meta language for descriptive markup. At first it was mainly used in industry but was soon picked up by scholars who needed a tool for textual analyses. Next step in the development and dissemination came almost thirty years after the first development when a, simplified but yet more strict, subset of SGML was launched in 1998. It was the *Extensible Markup Language*, XML, which is a general purpose markup language for storage and interchange of digital content with the primary purpose to facilitate the sharing of structured data across different information systems, particularly via the internet. The classification as an extensible language is due to the fact that users can define their own elements. Locally defined element can of course cause problems if the content ever is to be shared with others. That is why standardisation efforts are so important. The dominant *de facto* standard today in the humanities and the social sciences is the *Text Encoding Initiative*, TEI, found at <www.tei-c.org>.

The TEI community's original overall aim was to develop interchange guidelines allowing projects to share data and theories about data and to promote the development of common tools. More than 100 persons, from many different specialities and disciplines have been working hard since 1987, when the TEI project was launched, to produce and rewrite the guidelines. This effort have been very successful and the TEI guidelines are now widely accepted as the standard interchange format for textual data and there are quite a few full text archives and projects currently using the TEI guidelines.

In this study the TEI guidelines were used to describe and organise the propographical materials collected. A range of other tools, some of related to the XML family, were used to manipulate, analyse and present the marked-up content.

Prosopography – method and materials

Originally prosopography was mainly concerned with studies of the social elites. The concept of prosopography has been traced to the 16th century in Europe – then closely related to the contemporary endeavour to collect, register and organise all forms of knowledge. Development and contributions in prosopography applies to some different schools, either concerned with in-depth and detailed studies of smaller groups of for example social elites, or less detailed but larger numerical sets of individuals belonging to the same field.

Regardless the prosopographical school the main interest for a prosopographer is the investigation of a chosen group, the structures and transformations, rather than individuals or institutions. It is a kind of collective biography, particularly applicable for heterogenous datasets in historical research and also a method of necessity due to the fact that the objects of study are dead and therefore not available for interviews. Prosopographical research is a time-consuming endeavour when investigating a number of primary and sometimes secondary sources for detailed data collection.

There is also an important distinction in handling and analysing data between prosopography in the humanities and the French sociologically oriented prosopographical tradition. In traditional prosopography the sources play an important role in terms of certainty of the facts and their claims about individuals in the material. In sociologically oriented prosopographies the main concern is to collect the same kind of data for all persons in order to analyse the material with either multivariate quantitative techniques or with qualitative methods depending on the quality of the data. Thus, the quality of the data is most important independently of prosopographical orientation. The main methods for analysing the prosopographical data in the sociologically oriented studies are either geometrical data analysis especially correspondence analysis, or other more qualitative oriented approaches.

Traditionally the outcome of a prosopography was to publish results in a printed book. Along with the advancement and development of modern computer technology come possibilities for researchers to explore, use and handle extensive prosopographical materials in a variety of novel ways. Prominent examples are the *Prosopography of the Byzantine Empire* and the *Prosopography of Anglo-Saxon England*, both very rich and valuable resources in ancient and medieval history. Included in the *Prosopography of Anglo-Saxon England* are all individuals who lived in or were closely connected with Anglo-Saxon England during the period from 597 to 1042.

The main purpose in the project *Formation for the public sphere* was to investigate the social strategies of the first generations of women entering the public sphere in Sweden around the turn of the century 1900. This period was of crucial significance for women engaged in philanthropy, reform pedagogy, modern health care, literature and music. These women's strategies, investments and careers differed from their male contemporaries and their contributions are not

easily recognisable. In order to discern and interpret their contributions to the establishment of the modern welfare state institutions, a modern educational system and the modern cultural fields, methods from the French sociological tradition founded by Pierre Bourdieu have been used.

A central endeavour when preparing the prosopography was to collect information on the women's social origin, social intercourse, their networks, educational trajectories and matrimonial status. Such information is here called assets or capital. In Bourdieu's sense certain types of capital are acknowledged within certain social groups but not by everyone. Each field that is sufficiently autonomous has its own rules for inclusion, exclusion and rewards, and specific species of capital. By analysing the distribution of certain types of capital among the pioneer women the project aim was to try to map the structure, the hierarchies and the polarities of domains like female culture, education and philanthropy.

Since the project favoured the collection of sociologically interpretable data it was important to collect comprehensive information on names, dates and places, e.g. where and when and with whom she lived, where she worked, in order to trace the "meeting places" and networks. Hence a mandatory core set of data was, whenever possible, harvested to depict some of the most crucial assets:

Social origin: father's and mother's occupation, education, positions.

Number of brothers and sisters: Woman's and parents place of birth and place of upbringing.

Educational capital: kind of basic and further education. Sojourns abroad.

Social capital: influential relatives, matrimonial status, number of children, housing, member of state commissions, foundations.

Economic capital: wealth, earthly goods and relations to patrons.

Political and religious capital: positions in political/religious organisations, standpoints in such matters.

Specific symbolic capital: assets being valued either within certain fields or domains or within women's networks.

Of course many of the biographical texts cover much more, but the main aim was trying to make the collection of these core data as comprehensive as possible for each and every woman included.

Empirical materials

The prosopographical collection consists of rather dispersed kinds of resource material. There are three kinds of datasets of biographical accounts which has been collected by rather different means. One of the datasets is called capital descriptions and consists of 101 rather extensive texts written in running prose by the project members. The scholars have explored archival material such as

letters, diaries or estate reports, as well as printed newspapers, journals or books – and of course existing biographies and autobiographies.

The second dataset are 1251 condensed texts based on excerpted information from a biographical reference book published by Nils Styrbjörn Lundström in 1924. The excerpts are based on a collection of autobiographical short articles on Swedish women who had entered the public sphere around the turn of the century 1900.

The third kind of dataset consists of photographed and transcribed archival sources from the *Royal Advanced Seminary for Women Teachers* archive in Stockholm such as enrolment registers, registers for marks and alike. A total of 1980 women are registered in this dataset.

My role in the project was the technical editor. I compiled and encoded the written materials which others had collected for the module with capital descriptions. The archival material I collected myself. Then I compiled and prepared the three different collections for further analyses and applications.

In the collection and the writing of the capital descriptions, the scholars were guided by a structured schema corresponding to the species and subspecies of capital: date and place of birth, address of parental home, father's and mother's occupation, etcetera. This was the mandatory information and the "core set". The scheme was not intended to restrict the writing too much. Its main purpose was to ensure that the core data are as comprehensive as possible and as congruent as needed. There is no need to confine oneself to merely a line-up of facts. Several of the biographies are not only extensive and rich in information but also good essays, quite enjoyable to read and publishable as they are. Other capital descriptions are meagre, often because the information needed was not to be found in the archives.

The capital descriptions and the archival records were then XML encoded with the TEI scheme instead of constructing a project and content specific encoding scheme as some other similar projects has done. Instead, we use for example the additional TEI module for "Names and Dates" which provides tools for encoding prosopographical data such as information on proper names, names of organisations and names of geographical places, ambiguous date periods and precise dates in a detailed manner.

In the prosopographical content model each woman corresponds to a main division that contains subdivisions and further subdivisions. In principle each subdivision or sub-subdivision corresponds to one type of capital such as <div type="soc-orig">. The basic features marked in a subdivision are:

- name(s)
- a date(range)
- an event or an occurrence

For each of these elements there is a reciprocal relationship to a biography and an arbitrary category respectively. This means that a division is only instantiated when there is a relationship or an occurrence of a certain kind. Obviously, some instances are mandatory: I need for example to record a name and a birth date in order to instantiate a record for a woman.

The structural and logical hierarchies of content objects were used as a starting point when extracting exactly the data needed for the analyses. Thus, we can extract, the content "Paris" out of the element "placeName" when and only when it occurs within "div type= 'travel'". I might thus extract information on that the woman in question did travel to Paris, and not have to bother with all other Parises that might appear: books published in Paris, dresses made in Paris, father born in Paris. Obviously a consistent use of the content model providing for the accuracy of the markup become very important since so much depends on it.

The approach to apply the content model and do the encoding in running prose is not unproblematic. I have faced a few rather difficult decisions on the ways to encode and how to denote the aspects that needs to be captured. The consistent use of divisions corresponding to the content model is important, since it guarantees that the internal hierarchy of elements is no hindrance for finding the specific data. The collected material, images and transcriptions, from the archival records for the teaching seminar were very structured themselves compared to the material described above. Thus the markup in this case was an easier task altogether partly depending on my improved skills I assume.

The prosopographical collection, stored in a master file in XML format, and provided with TEI markup, was then used for further transformations and manipulations with other tools and on other platforms.

Applications

The applications developed in this study were based on the master XML file. Through different steps I migrated and extracted the content needed for different analyses. One such application is a subset of content objects stored in the master file which was migrated into statistical software.

The corresponding types of data from the two of the datasets in the prosopographical collection, comprising capital descriptions and autobiographical notes, was merged into one dataset for further processing and statistical analyses. The data needed was extracted into tables using the script language *Extensible Stylesheet Language Transformations*, XSLT, in order to import the tables into the statistical software for analyses. The statistical software is used for converting string values to numerical values and for organising, combining, classifying and aggregating variables.

Other applications based of the master file where migration of content objects into tools for social network analysis and interactive topic maps where the structure of the content, as well as the content itself, could be explored. The

applications also comprise a web archive with a database backend and a web interface for searching the prosopographical database with access to the full text capital descriptions (the biographical sketches) which can be downloaded in a choice of file formats. Another example was the use of an application designed for annotating images, using markup with the TEI/XML format. In this application the images of the archival record from the teaching seminar and the corresponding transcribed texts could be joined and as an output the image and the text were displayed simultaneously. When the user clicks on the image this particular areas text will be accentuated and vice versa if the user clicks on the index adjacent the image, the corresponding area on the image will be shown.

All applications, all the texts provided with markup, all scripts used for manipulation and migration of content objects, are freely available on the web site and on the CD provided with this thesis.

Conclusion

In this thesis the overall aim was to identify and describe principles, methods and tools for content design. Another aim was to as to investigate the limitations and opportunities of the identified methods and tools by means of applications in history of education thus contributing to methods development by proposing modes of procedure in content design for research and teaching

I have shown that a complex set of heterogeneous material calls for some measures to guarantee the consistency of data. If we had chosen a database solution, such as for example the relational database it would be a matter of course to opt for a controlled vocabulary. Here, the categorisation was undertaken prior to data input into order to ensure the congruency of the recorded aspects. This is an important aspect of the study. The major decisions on the categorisation have been taken prior to the collection of information and prior to the writing of the biographies in terms of indicators denoting the different types of capital.

The TEI Guidelines have been used as far as possible for describing but also for the principles of classification of content objects. These content objects, i.e. elements and attributes, are represented by XML markup inserted into the running prose text.

What might separate this undertaking from similar prosopographical collections is the aim to maximise portability. The material is available through an ordinary web browser, and downloadable and reusable with intact markup. In other words, the users should be able to import the material to their own systems, and not forced to stay on the website.

Another difference is that the encoded data are used as the input to multivariate quantitative analyses, as illustrated at the end of this paper. The most important difference is, however, the aim to establish a sociological prosopography where information is used to indicate the individuals' symbolic and ma-

terial assets and strategies in order to understand the relations between their positions and standpoints.

My conclusion is that the TEI encoding scheme serves well as a valuable tool for encoding complex historical materials. Especially useful are the possibilities to use a TEI/XML encoded corpus for multiple purposes, for archiving, distribution, exchange, transformations, and for qualitative analyses, and as input into multivariate statistical analyses. Thanks to the encoding the web archive provides enhanced navigation besides presentation of different views of the material and filtering possibilities.

Besides, a digital collection provided with markup allows texts to be treated as research tools in themselves. That is, digital texts lend themselves to much more than access, retrieval and reading. Encoded texts and other materials can help scholars do other kinds of research work and posing new questions by the modularised design, and by making the underlying theories of the content design explicit.

Potentially these kinds of international standardisation efforts such as the TEI will be useful not just for the scholarly community but also teachers and students. Provided that digital collections with research material, courseware and other educational resources are made available in modular shape and adequately marked-up, one and the same content might be combined and reused for different audiences with different needs in different educational settings. Teachers and students might be encouraged to find personal ways of learning and less inclined to use pre-constructed syllabi and selections of course content.

Innehållsförteckning för digitala bilagor

I denna förteckning listas innehållet i den bifogade CD:n, som innehåller de digitala bilagorna samt de figurer som framställs i boken som bifogas som digitala bilder. Därutöver innehåller CD:n ytterligare bilder med fler exempel och ett antal xml-filer med exempel på olika sorters märkning samt ett utdrag ur FFO-KAPITAL i XML-format, avsedd att användas som exempelsamling tillsammans med de XSLT-skript som bifogats som exempel hur transformeringen utförts. På CD:n finns ytterligare några dokument: en sammanställning över de verktyg och metoder som använts i avhandling (förslag till verktygslåda). Därtill bifogas de matriser som använts som underlag för nätverksanalyser och för beskrivande statistik samt två användarhandledningar; en för XSLT-transformeringen och för verktyget DL-ANTE (användas för automatisera märkning och numrering av innehållsobjekt).

Denna CD är avsedd för att ge exempel på den märkning och de tillämpningar som presenteras i boken samt ge tillgång till de bilagor och annat som jag hänvisar till men även att ge förslag till en verktygslåda för innehållsdesign. Innehållet på CD:n är främst avsett att nås via de html-baserade index-sidorna. CD:n har en katalogstruktur som motsvarar uppdelningen i förteckningen nedan.

De inkluderade filerna, med undantag för DL-ANTE.exe, fungerar på alla vanliga plattformar och är kontrollerade på PC (Windows XP), Ubuntu (Linux), Debian (Linux) och Mac.⁷⁶⁷ Det som inte inkluderats på CD:n är SQL-databasen och webbarkivet. Webbplatsen innehåller samtliga aktuella kapitalbeskrivningar, register och analyser samt en del annat material i anslutning till projektet Formering för offentlighet och databasen innehåller ett webbaserat gränssnitt för sökningar av poster och hämtning av fulltextversioner av kapitalbeskrivningarna. Allt detta är tillgängligt via webbplatsen <www.skeptron.uu.se/monicalz/ffo/>.

⁷⁶⁷ Filformaten på CD:n är i html-format för webbsidor och matriser, pdf-format för textdokument, jpg-format för bilder, xml-format (tei & graphml) för märkning samt xsl-format för xslt-skripten. DL-ANTE.exe har endast använts på PC (Windows XP).

Bilageförteckning

Förteckning över digitala bilagor och sökväg.

Bilaga nr	Beskrivning	Sökväg
Bilaga 1	Mall för kapitalbeskrivningar	CD:/bilagor/bilaga1.pdf
Bilaga 2	Beskrivande märkning av indikatorer för egenskaper och tillgångar	CD:/bilagor/bilaga2.pdf
Bilaga 3	Exempel på beskrivande märkning av kapitalbeskrivningar	CD:/bilagor/bilaga3.pdf
Bilaga 4	Personförteckning över samtliga kvinnor i FFO-LUNDSTROM (inkl. FFO-KAPITAL)	CD:/bilagor/bilaga4.pdf
Bilaga 5	Exempelposter ur Lundströms biografiska uppslagsbok	CD:/bilagor/bilaga5.pdf
Bilaga 6	Förteckning över föreningar, sammanslutningar och organisationer	CD:/bilagor/bilaga6.pdf
Bilaga 7	Förteckning över digitaliserade handlingar i Kungl. Högre Lärarinneseminiariets arkiv	CD:/bilagor/bilaga7.pdf
Bilaga 8	Exempel ur inskrivningsmatriklar i Kungl. Högre Lärarinneseminiariets arkiv	CD:/bilagor/bilaga8.pdf
Bilaga 9	Registrering av metadata i TEI header	CD:/bilagor/bilaga9.pdf
Bilaga 10	Exempel på beskrivande märkning av inskrivningsmatriklar i Kungl. Högre Lärarinneseminiariets elevmatrikel	CD:/bilagor/bilaga10.pdf
Bilaga 11	Förteckning över inskrivna elever vid Kungl. Högre Lärarinneseminiariet i Stockholm under perioden 1861 -1943	CD:/bilagor/bilaga11.pdf
Bilaga 12	Förteckning över variabler i registret FFO-LUNDSTROM	CD:/bilagor/bilaga12.pdf
Bilaga 13	Exempel på Litteraturbankens annoteringsfunktion för e-texter	CD:/bilagor/bilaga13.pdf

Figurförteckning

Förteckning över figurer i avhandlingen, bifogade som digitala bilder, och sökväg

Figur nr	Beskrivning	Sökväg
Figur 1	Exempel på den deskriptiva märkningens grundläggande principer: elementet	CD:/figurer/figur1.jpg
Figur 2	Exempel på den deskriptiva märkningens hierarkiska struktur med hjälp av nästlade element	CD:/figurer/figur2.jpg
Figur 3	XML: hierarkisk dokumentmodell	CD:/figurer/figur3.jpg
Figur 4	Exempel på märkning av en viss kvinnas egenskaper och tillgångar	CD:/figurer/figur4.jpg
Figur 5	Urklipp ur nätverksgraf med Ellen Keys egenskaper (attribut) accentuerade	CD:/figurer/figur5.jpg
Figur 6	Exempel på procedurrell märkning med LaTeX	CD:/figurer/figur6.jpg
Figur 7	Exempel på märkning av föreningsaktiviteter i FFO-KAPITAL	CD:/figurer/figur7.jpg

Figur nr	Beskrivning	Sökväg
Figur 8	Märkspråket COCOA, exempel på syntax	CD:/figurer/figur8.jpg
Figur 9	Exempel på COCOA märkning av 1600-talspjäs	CD:/figurer/figur9.jpg
Figur 10	Jämförelse av bruket av sluttagg för nästlade element mellan SGML och XML.	CD:/figurer/figur10.jpg
Figur 11	Periodisk sökning av ”semantic web”, antal träffar per angiven tidsperiod.	CD:/figurer/figur11.jpg
Figur 12	Exempel på Topic Maps datamodell och grundläggande principer	CD:/figurer/figur12.jpg
Figur 13	Översikt över TEI:s moduler i TEI P4 (<i>Pizza Chef Model</i>)	CD:/figurer/figur13.jpg
Figur 14	TEI-dokumentets grundläggande struktur och de obligatoriska elementen i <teiHeader>	CD:/figurer/figur14.jpg
Figur 15	Exempel på grundläggande struktur och element i TEI-dokumentets kropp	CD:/figurer/figur15.jpg
Figur 16	Illustration av FFO-databasens utformning	CD:/figurer/figur16.jpg
Figur 17	Översikt av kapitalbeskrivningsmallens logiska ordning och hierarkiska struktur	CD:/figurer/figur17.jpg
Figur 18	Utdrag ur inskrivningsmatrikeln för läsåret 1861-1862.	CD:/figurer/figur18.jpg
Figur 19	Principer för beskrivning av innehåll i kapitalbeskrivningarna.	CD:/figurer/figur19.jpg
Figur 20	Exempel på poster som värden i "key"-attributet relaterar till	CD:/figurer/figur20.jpg
Figur 21	Märkning av en kvinnas kontakt med andra kvinnor i materialet	CD:/figurer/figur21.jpg
Figur 22	Exempel på post i inskrivningsmatrikel för perioden 1914–1928.	CD:/figurer/figur22.jpg
Figur 23	Märkning av arkivpost (urklipp).	CD:/figurer/figur23.jpg
Figur 24	Märkning av unikt id samt länkning till externa dokument	CD:/figurer/figur24.jpg
Figur 25	Semiautomatisk märkning av föreningsnamn	CD:/figurer/figur25.jpg
Figur 26	Översikt för migrering av filformat.	CD:/figurer/figur26.jpg
Figur 27	Migrering från TEI/XML masterfil till andra filformat och användningsområden	CD:/figurer/figur27.jpg
Figur 28	Extrahering av önskad information med XSLT, från masterfil till tabell	CD:/figurer/figur28.jpg
Figur 29	FFO-LUNDSTROM registrets 4 nivåer	CD:/figurer/figur29.jpg
Figur 30	Kvinnornas sociala ursprung (boendeort och sfär). Procent (N=1251)	CD:/figurer/figur30.jpg
Figur 31	Jämförelse födelseorter läsårsvis	CD:/figurer/figur31.jpg
Figur 32	Nätverksanalys av 101 kvinnors interna kontakter i FFO-KAPITAL.	CD:/figurer/figur32.jpg
Figur 33	Nätverksanalys av FFO-KAPITAL: flest kontakter och mest omnämnd	CD:/figurer/figur33.jpg
Figur 34	Exempel på sökning i en nätverksanalys	CD:/figurer/figur34.jpg
Figur 35	Exempel på en kvinnas vy	CD:/figurer/figur35.jpg
Figur 36	Visualisering av XML-dokumentets struktur	CD:/figurer/figur36.jpg
Figur 37	Illustrering av trädstrukturens yttersta gren och dess innehåll	CD:/figurer/figur37.jpg
Figur 38	Exempel på hur bild och text kan användas tillsammans.	CD:/figurer/figur38.jpg

Figur nr	Beskrivning	Sökväg
Figur 39	Exempel på relaterande märkning som förenar bild och transkriberad text.	CD:/figurer/figur39.jpg
Figur 40	Databasschema för FFO-KAPITAL.	CD:/figurer/figur40.jpg
Figur 41	Webbaserat gränssnitt för sökning i FFO-KAPITAL	CD:/figurer/figur41.jpg
Figur 42	Webbgränssnitt för visning av resultat efter sökning i FFO-KAPITAL	CD:/figurer/figur42.jpg

Bildförteckning

Förteckning över bifogade bilder och sökväg

Bild nr	Beskrivning	Sökväg
Bild 1	Fotografi av uppslag i betygsliggaren i Kungl. Högre Lärarinneseminiariets arkiv (Serie 2A, vol. 1)	CD:/bilder/hls-betyg.jpg
Bild 2	Fotografi av inskrivningsmatrikel i Kungl. Högre Lärarinneseminiariets arkiv (Serie 2A, vol. 1)	CD:/bilder/hls-matrikel.jpg
Bild 3	Visualisering av kapitalbeskrivningsmallens struktur	CD:/bilder/hypergraph1.jpg
Bild 4	Visualisering av en kapitalbeskrivningsstruktur	CD:/bilder/hypergraph2.jpg
Bild 5	Illustrering av trädstrukturens yttersta gren och dess innehåll	CD:/bilder/hypergraph3.jpg

Filförteckning

Förteckning över bifogade filer och sökväg

Fil nr	Beskrivning	Sökväg
Fil 1	Exempel på märkning av en relativt omfattande kapitalbeskrivning	CD:/xml/gagner-marielouise.xml
Fil 2	Exempel på märkning av en mindre omfattande kapitalbeskrivning	CD:/xml/hermelin-honorine.xml
Fil 3	Exempel på beskrivande märkning av indikatorer på utbildningskapital i en kapitalbeskrivning	CD:/xml/div-edu.xml
Fil 4	Exempel på beskrivande märkning av arkivposter samt märkning av unikt ID och länkning till externt dokument	CD:/xml/klhs-example.xml
Fil 5	Exempel på märkning för nätverksgraf (GraphML)	CD:/xml/ffo-net.graphml
Fil 6	Exempel på registrering av metadata i TEI-header	CD:/xml/khls-header.xml
Fil 7	Utdrag ur masterfilen för FFO-KAPITAL avsedd för test av xsl-skript eller andra verktyg	CD:/xml/ffo-utdrag.xml

Dokumentförteckning

Förteckning över bifogade digitala dokument och sökväg

Dokument nr	Beskrivning	Sökväg
Dokument 1	Matris för nätverksanalyser av FFO-KAPITAL	CD:/docs/ffo-net.html
Dokument 2	Matris som kan användas för beskrivande statistik med ett urval variabler i FFO-LUNDSTROM	CD:/docs/ffo-stat.html
Dokument 3	Verktygslåda för innehållsdesign: Förteckning av metoder och verktyg för innehållsdesign (HTML)	CD:/docs/toolbox.html
Dokument 4	Verktygslåda för innehållsdesign: Förteckning av metoder och verktyg för innehållsdesign (PDF)	CD:/docs/toolbox.pdf
Dokument 5	Handledning för test av XSL-skript på utdrag ur FFO-KAPITAL (anvisningar för fil nr 7 & verktyg nr 2, 3 eller 4)	CD:/docs/ffo-xsl-test.pdf

Verktygsförteckning

Förteckning över bifogade verktyg och sökväg

Verktyg nr	Beskrivning	Sökväg
Verktyg 1	DL-ANTE: Digital Litteratur: Automatisk Numrering och Typning av Element	CD:/verktyg/dl/DL-ANTE.exe
Bilaga 1a <i>till verktyg 1</i>	Användarhandledning DL-ANTE (Digital Litteratur: Automatisk Numrering och Typning av Element)	CD:/verktyg/dl/dl-ante.pdf
Verktyg 2	XSLT-skript för extrahering/transformering av organisation & föreningsnamn	CD:/verktyg/xsl/ffo-org.xsl
Verktyg 3	XSLT-skript för extrahering/transformering av personnamn	CD:/verktyg/xsl/ffo-name.xsl
Verktyg 4	XSLT-skript för transformering och räkning av sociala kontakter	CD:/verktyg/xsl/ffo-net.xsl
Verktyg 5	XSLT-skript för räkning av element	CD:/verktyg/xsl/count.xsl

Referenser

Otryckt material

Kungl. Högre Lärarinneseminariets arkiv

Arkivbeteckning: SE/SSA/2132 Högre Lärarinneseminariet

Serie D 1 Inskrivningsmatriklar:

Vol. 1 1861–1941

Vol 2. 1914–1928

Serie D 2B Betygsliggare över inträdesprövningar:

Vol. 1. 1861 – 1893

Se även bilaga 7 för fullständig förteckning över fotograferade och transkriberade handlingar i Kungl. Högre Lärarinneseminariets arkiv.

Medlemsmatrikel

Sällskapet Nya Iduns medlemsmatrikel i (*Sällskapet Nya Idun 1885-1935: femtioårshistorik: medlemsförteckning*).

Digitalt material

Biografiskt material

Ämnesportaler vid Kvinnohistoriska samlingarna vid Göteborgs universitetsbibliotek <<http://www.ub.gu.se/kvinn/>> och där särskilt av följande:

Biografiska artiklar vid portalen ”Kvinnors kamp för kunskap” om: Sophie Adlersparre, Astrid Cleve von Euler, Ellen Fries, Honorine Hermelin, Ellen Key, Betty Pettersson, Anna Sandström, Lydia Wahlström, Anna Whitlock

Biografiska artiklar vid portalen ”Kvinnors kamp för rösträtt” om: Gulli Petrini, Lydia Wahlström, Anna Whitlock, Elin Wägner

Bearbetat digitalt material

Avskrifter och register

Avskrifter ur Lundström, Nils Styrbjörn. *Svenska kvinnor i offentlig verksamhet: porträtt och biografier*. Uppsala, 1924, avseende 1206 kvinnor, se personförteckning i bilaga 4 vid <CD:/bilagor/bilaga4.pdf>.

Kjell Östbergs register om drygt 1000 kvinnors organisationstillhörighet och styrelseuppdrag, tyngdpunkt på 1930-talet (återskapat från Dbase IV-format) (odat.).

Förteckning över styrelseledamöter i *Fredrika Bremer-förbundets styrelse* 1885-1900 enligt verksamhetsberättelserna. Av Boel Englund (odat.).

Förteckning över medlemmar 1886-1895 i *Föreningen för gift kvinnas eganderätt* (FGKE), inkl. ständiga medlemmar, enligt förteckning funnen i Fredrika Bremerförbundets arkiv, Riksarkivet, juni 2001. Av Boel Englund (2001-06-19).

Kapitalbeskrivningar

Kapitalbeskrivningar avseende följande 101 kvinnor (FFO-KAPITAL), årtalen anger födelseår.

1823, Sophie Adlersparre, Skriftställarinna, (KHLS)
1863, Anna Ahlström, Skolgrundare
1844, Sofi (f. Hultén) Almqvist, Skolgrundare, (KHLS)
1875, Gertrud (f. Almqvist) Almqvist-Brogren, Författarinna
1888, Andrea Andreen-Svedberg, Läkare
1876, Disa Beijer, Barnträdgårdslärlarinna
1860, Nanna Bendixson, Barnboks författare
1842, Eleonora (Ellen) Bergman, Sånglärlarinna
1868, Maria Bergmark, Seminarieledarinna
1853, Anna Bergström-Simonsson, Sånglärlarinna
1869, Gertrud Bergström, Skolköksinspektör
1866, Sigrd Björklund, Flickskolelärlarinna
1855, Kerstin Bohman, Socialt verksam
1862, Agnes Branting, Textilkonstnärinna
1855, Anna Branting, Författarinna
1801, Fredrika Bremer, Författarinna
1866, Emilia Broomé, Socialpolitiker
1846, Eugenie Brummer-Steinmetz, Skolgrundare
1863, Sigrd Carlheim-Gyllensköld, Pianist
1869, Elin Cederblom, Seminarieledarinna
1815, Marie Cedersköld, Föreståndarinna
1875, Astrid Cleve von Euler, Lärarinna
1850, Calla Curman, Filantrop
1851, Lotten Dahlgren, Skriftställare
1855, Alma Detthow, Skolgrundare
1851, Hildur Djurberg, Seminarieföreståndarinna, (KHLS)
1836, Drottning Sophia, Socialt verksam
1866, Lizinka Dyrssen, Socialt verksam
1849, Agnes Ekermann, Seminarieadjunkt, (KHLS)
1834, Augusta Ekman, Socialt verksam
1884, Anna Lenah Elgström, Författarinna

1843, Lilly Engström, Seminarieledarinna, (KHLS)
1864, Elisabeth Eurén, Lektor, (KHLS)
1872, Elin Falk, Gymnastikinspektör
1883, Mollie Faustman, Konstnär
1855, Ellen Fries, Studierektor
1829, Eva Fryxell, Skriftställarinna
1868, Marie Louise Gagner, Lektor, (KHLS)
1843, Selma Giöbel, Stuckatör
1842, Thérèse Gylden, Socialt verksam
1886, Honorine Hermelin, Flickskolelärlarinna
1872, Kerstin Hesselgren, Yrkesinspektör
1841, Anna Hierta-Retzus, Filantrop, (KHLS)
1875, Klara Johansson, Skriftställare
1866, Anna Kempendahl, Slöjdlärlarinna
1835, Amanda Kerfstedt, Författarinna
1849, Ellen Key, Skriftställarinna
1874, Julia Kinberg von Sneidern, Läkare
1881, Gerda Kjellberg, Läkare
1867, Ellen Kleman, Redaktör
1878, Vera von Kraemer, Publicist
1878, Mia Leche-Löfgren, Skriftställarinna
1862, Sigrd Leijonhufvud, Publicist
1816, Fredrika Linnell, Filantrop
1840, Ebba Lind af Hageby, Filantrop, (KHLS)
1881, Elisabeth Lind, Sjuksköterska
1865, Gurli Linder, Författarinna, (KHLS)
1870, Anna Lindhagen, Barnavårdsinspektör
1863, Emmy Lindhagen, Sjuksköterska
1840, Augusta Lundin, Modeskapare
1847, Hulda Lundin, Seminarieföreståndarinna, (KHLS)
1866, Agda Meyerson, Sjuksköterska
1866, Gerda Meyerson, Skriftställarinna
1872, Sigrd Millrath, sekreterare
1850, Agda Montelius, Filantrop
1872, Ada Nilsson, Läkare
1881, Kerstin Nordendahl, Sjuksköterska

- 1857, Bertha Nordenson, Socialt verksam
 1843, Helena Nyblom, Författarinna
 1873, Ebba Pauli, Publicist
 1861, Sally Petersson, Sjuksköterska
 1867, Gulli Petrini, Flickskolelärarinna
 1838, Betty Pettersson, Lärarinnapåläroverk
 1870, Adele Philipson, Flickskolelärarinna
 1830, Prinsessan, Eugenie, Socialt verksam
 1879, Eva Ramstedt, Docent
 1835, Emmy Rappe, Sjuksköterska
 1862, Jacobine Ring, Skriftställare
 1877, Estrid Rodhe, Sjuksköterska
 1862, Anna Roos, Författarinna, (KHLS)
 1862, Ellen Sandelin, Läkare
 1854, Anna Sandström, Skolgrundare, (KHLS)
 1872, Ingeborg Schager, Skolkökslärarinna
 1852, Hedvig Sidner, Lektor, (KHLS)
 1875, Marika Stiernstedt, Författarinna
 1872, Alma Sundqvist, Läkare
 1875, Anna Sörensen, Rektor
 1876, Alfhild Tamm, Läkare
 1886, Carin Ulin, Lärarinna
 1861, Elsa Uppling, Sånglärarinna
 1869, Lydia Wahlström, Publicist
 1868, Ingeborg Walin, Hushållslärare, (KHLS)
 1870, Bertha Wellin, Riksdagskvinna
 1852, Anna Whitlock, Skolgrundare, (KHLS)
 1863, Matilda Widegren, Skolföreståndarinna, (KHLS)
 1856, Karolina Widerström, Läkare
 1888, Ruth Wikander, Lektor
 1882, Elin Wägner, Skriftställare
 1871, Anna Åbergsson, Revisor
 1869, Annie Åkerhielm, Författarinna
 1870, Agda Östlund, Riksdagskvinna

Övriga tryckta referenser⁷⁶⁸

- Aarseth, Espen. *Cybertext: perspectives on ergodic literature*. Bergen: Univ. of Bergen, 1995.
 Ahlström, Herman Fredrik. *Officiell berättelse öfver Baltiska utställningen i Malmö 1914*. Malmö: Baltiska utställningen, 1915.
 Ambjörnsson, Ronny, & Sverker Sörlin. *Obemärkta: det dagliga livets idéer*. Stockholm: Carlsson, 1995.
 Andersson, Irene. *Kvinnor mot krig: aktioner och nätverk för fred 1914-1940*. Studia historica Lundensia, 1. Lund, 2001.
 André, Jacques, Vincent Quint, Richard Furuta, & Institut national de recherche en informatique et en automatique. *Structured documents*. The Cambridge series on electronic publishing. Cambridge: Cambridge University Press, 1989.
 Aronsson, Peter, Solveig Fagerlund, & Jan Samuelson. *Nätverk i historisk forskning: meta-for, metod eller teori?* Rapporter från Växjö universitet. Humaniora, 1. Växjö: Univ., 1999.
 Barnard, David, Lou Burnard, Jean-Pierre Gaspard, Lynne A Price, Michael Sperberg-McQueen, & Giovanni Batista Varile. *Hierarchical Encoding of Text. Technical Problems and SGML Solutions*. Computers and the Humanities, The Text Encoding Initiative: Background and Contents. red. Nancy Ide and Jean Vâconis 29, 1995.
 Berners-Lee, Tim, & Mark Fischetti. *Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor*. 1. uppl. San Francisco: Harper, 1999.
 Berners-Lee, Tim, James Hendler, & Ora Lassila. "The Semantic Web." *Scientific American*. May (2001).
 Bondi Johannessen, Janne, Eckhard Bick, Kristin Hagen, Dorte Haltrup, Åsne Haaland, & Andra Björk Jónsdóttir. "The Nomen Nescio Project - Scandinavian

⁷⁶⁸ *Modern Language Association* anvisningar har används för utformning av referenser i fotnoter och i förteckningen övriga tryckta referenser. MLA:s anvisningar har modifierats något; i avhandlingen används kursiv stil för refererade titlar samt "s." och "f(f)" för sidhänvisningar i fotnoterna.

- Named Entity Recognition". *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (ALLC/ACH)*, June 11 - 16, Göteborg University, 2004.
- Borin, Lars. "Vi som går köksvägen. Språkteknologer och korpuslingvister i Litteraturbanken." i *Fältanteckningar: utbildnings- och kultursociologiska texter tillägnade Donald Broady*. red. Mikael Börjesson, Ingrid Heyman, et al. Uppsala: Forskningsgruppen för utbildnings- och kultursociologi ILU Uppsala universitet, 2006 (399-404).
- Bourdieu, Pierre. *Homo academicus*. Övers. Lars Eberhard Nyman and Mats Rosengren. Stockholm/Stehag: Brutus Östlings bokförl. Symposion, 1996.
- Bowker, Geoffrey C., & Susan Leigh Star. *Sorting things out: classification and its consequences*. Inside technology. Cambridge, Mass.: MIT Press, 1999.
- Bradley, John. "Finding a Middle Ground between 'Determinism' and 'Aesthetic Indeterminacy': a Model for Text Analysis Tools." *Literary and Linguistic Computing* 18.2 (2003).
- . "Documents and Data: Modelling Materials for Humanities Research in XML and Relational Databases." *Literary and Linguistic Computing* 20.1 (2005): 133-151.
- Bradley, John, & Harold Short. "Using Formal Structures to Create Complex Relationships." i *Resourcing Sources*. Oxford: Unit for Prosopographical Research, 2002 (3-21).
- . "Texts into Databases: The Evolving Field of New-style Prosopography." *Literary and Linguistic Computing* 20.1 (2005): 3-24.
- Bradley, John, & Paul Vetch. "Supporting Annotation as a Scholarly Tool - Experiences From the Online Chopin variorum Edition." *Literary and Linguistic Computing* 22.2 (2007): 225-241.
- Broady, Donald. *Sociologi och epistemologi: om Pierre Bourdieus författarskap och den historiska epistemologin*. Skeptronserien. 2., korr. uppl. uppl. Stockholm: HLS Förlag, 1991.
- . "Det nya handbiblioteket." i *Biblioteken, Kulturen och den sociala intelligensen. Aktuell forskning inom biblioteks- och informationsvetenskap*. red. Lars Höglund. Göteborg: Forskningsrådsnämnden/Valfrid, 1995 (83-107).
- . *Content design. Methods and tools for the creation of portable hypermedia archives. Notes for a proposed CID project. Version 2*. Nada/CID, KTH, 1997.
- . "Hur kodar man Röda Rummet?" i *ABM, IT och forskningen. Rapport från en konferens på Kungliga Biblioteket den 17 november 1999*. red. Mats Rolén. Stockholm: Riksbankens Jubileumsfond, 2000.
- . "Digitala arkiv och portföljer." i *IT i skolan: mirakelmedicin eller sockerpiller?* Observatoriet för IT lärande kunskap och kompetens, Rapport 45/2001. Stockholm: IT-kommissionen, 2001 (67).
- . "French prosopography: definition and suggested readings." *Poetics* 30 (2002): 381-385.
- . "Nätverk och fält." i *Sociala nätverk och fält*. red. Håkan Gunneriusson. Opuscula Historica Upsaliensia, 28. Uppsala: Historiska institutionen, Uppsala Universitet, 2002 (49-72).
- Broady, Donald, Boel Englund, Ingrid Heyman, Agneta Linné, Kerstin Skog-Östlin, Eva Trotzig, & Annika Ullman. *Formering för offentlighet. En kollektivbiografi över Stockholmskvinnor 1880-1920*. Uppsala: SEC/ILU, Uppsala universitet, 1998.

- Broady, Donald, & Hasse Haitto. "Internet and the humanities: the promises of Integrated Open Hypermedia". *Contemporary computer and network technologies*, 17-18 Jan., Moskva, 1996.
- Broady, Donald, & Monica Langerth Zetterman. "Formation for the public sphere – a collective biography". *DRH2003 (Digital Resources in the Humanities 2003)*, 31 aug-3 sept, University of Gloucestershire, Cheltenham, England, 2003.
- Brockman, William, Laurua Neumann, Carole Palmer, & Tonya Tidline. *Scholarly Work in the Humanities and the Evolving Information Environment*. Washington, D.C.: Digital Library Federation, Council on Library and Information Resources, 2001.
- Brown, Susan, Clements Patricia, Isobel Grundy, Jeffrey Antoniuk, & Sharon Farnel. "Facing the Deep: The Orlando Project Delivery System 1.0." *Text Technology* 14.2 (2005): 21-40.
- Burnard, Lou. "Report of Workshop on Text Encoding Guidelines." *Literary and Linguistic Computing* 3 (1988): 131-133.
- . "New tricks from an old dog: An overview of TEI P5". *Dagstuhl Seminar Proceedings. Digital Historical Corpora - ARchitecture, Annotation, and Retrieval*, Dagstuhl, Germany, 2007.
- Burnard, Lou, Katherine O'Brien O'Keeffe, & John Unsworth, red. *Electronic textual editing*. Vol. Volym. New York, N.Y.: Modern Language Association of America, 2006.
- Burnard, Lou, & Sebastian Rahtz. "RelaxNG with Son of ODD". *Extreme Markup Languages 2004*, Montréal, Québec, 2004.
- Burnard, Lou, & Michael Sperberg-McQueen. *Text Encoding Initiative. Guidelines for electronic text encoding and interchange*. XML Version. Oxford, Providence, Charlottesville, Bergen: Text Encoding Initiative Consortium, 2002.
- Buzzetti, Dino, & Jerome J. McGann. "Editing in a digital horizon." i *Electronic textual editing*, red. Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York, N.Y.: Modern Language Association of America (53-73).
- Börjesson, Mikael. "Om att klassificera sociala grupper." i *Fältanteckningar: utbildnings- och kultursociologiska texter tillägnade Donald Broady*, red. Mikael Börjesson, Ingrid Heyman, et al. Skeptronserien. Uppsala universitet: Forskningsgruppen för utbildnings- och kultursociologi ILU Uppsala universitet, 2006 (249-258).
- Cameron, Averil, red. *Fifty years of prosopography: the later Roman Empire, Byzantium and beyond*. Vol. Volym. Oxford: Published for the British Academy by Oxford University Press, 2003.
- Carlson, G. "Literary Works in Machine-Readable Form." *Computers and the Humanities* 1 (1967): 75-102.
- Cederblad, Carl. *Bildningens väg: idéerna och livet*. Uppsala: Lindblad, 1932.
- Chartier, Roger. *Böckernas ordning: läsare, författare och bibliotek i Europa mellan 1300-tal och 1700-tal*. Övers. Jan Stolpe. 1. uppl. uppl. Göteborg: Anamma, 1995.
- . *Forms and meanings: texts, performances, and audiences from codex to computer*. Philadelphia: Univ. of Pennsylvania Press, 1995.
- Coombs, James, Allen H. Renear, & Steven DeRose. "Markup systems and the future of scholarly text processing." *Communications of the ACM* 30.11 (1987): 933 - 947.
- Cummings, James. "Liturgy, Drama, and the Archive: Three conversions from legacy formats to TEI XML." *Digital Medievalist* 2.1 (2006).

- Cummings, James, & Monica Langerth Zetterman. "Legacy Data Migration: A pilot study on the methodological feasibility of conversion and enhancement of electronic resources." *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities*. Gothenburg, Sweden, 2004.
- Dahl, Torsten, & Nils Bohman. *Svenska män och kvinnor: biografisk uppslagsbok*. Stockholm: Bonnier, 1942.
- Dahlström, Mats. "Nya medier, gamla verktyg." *Human IT* 6.4 (2002): 77-116.
- . *Under utgivning: den vetenskapliga utgivningens bibliografiska funktion*. Skrifter från Valfrid, 34. Borås: Valfrid, 2006.
- DeRose, Steven. "Markup Overlap: A review and a Horse". *Extreme Markup Languages*, August 2-6, 2004, Montréal, Québec, 2004.
- DeRose, Steven, David G. Durand, Ellen Mylonas, & Allen H. Renear. "What is Text, Really?" *Journal of Computing in Higher Education* 1.2 (1990): 3-26.
- DeRose, Steven J. *The SGML FAQ book: understanding the foundation of HTML and XML*. Boston: Kluwer Academic, 1997.
- Driscoll, Matthew James. "Levels of Transcription." i *Electronic textual editing*, red. L. Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York, N.Y.: Modern Language Association of America, 2006 (254-261).
- . "XML markup of prosopographical data". *Exploring New Methods for Prosopography in the Humanities and the Social Sciences*, Uppsala, Sweden, 2007.
- Eck, Werner. "The *Prosopographia Imperii Romani* and Prosopographical Method." i *Fifty years of prosopography: the later Roman Empire, Byzantium and beyond*, red. Averil Cameron. Proceedings of the British Academy, 118. Oxford: Published for the British Academy by Oxford University Press, 2003 (11-22).
- Engman, Jonas. "Handling fine-meshed filtering from archive to resource level. Combining RDF-metadata with XML content markup." Umeå University, 2004.
- Erson, Eva. "*Det är månen att nå*": en studie i några datorintresserade pojars språk och föreställningsvärld. Umeå: Institutionen för nordiska språk, Umeå Universitet, 1992.
- European Interoperability Framework for pan-European eGovernment Services*. Luxembourg: Office for Official Publications of the European Communities, 2004.
- Fevrell, Valter. *Kungl. Högre lärarinneseminariet in memoriam: minnesrunor*. Stockholm: Geber, 1943.
- Flanders, Julia. "Gender and the Electronic text." i *Electronic text: investigations in method and theory*, red. Kathryn Sutherland. Oxford: Clarendon, 1997 (127-143).
- . "Learning, Reading, and the Problem of Scale: Using Women Writers Online." *Pedagogy: Critical Approaches to Teaching Literature, Language, Composition, and Culture* 2.1 (2002): 49-59.
- Fleck, Ludwik. *Uppkomsten och utvecklingen av ett vetenskapligt faktum: inledning till läran om tankestil och tankekollektiv*. Övers. Bengt Liliequist. Eslöv: B. Östlings bokförl. Symposion, 1997.
- Florin, Christina. *Kampen om katedern: feminiserings- och professionaliseringsprocessen inom den svenska folkskolans lärarkår 1860-1906*. Umeå studies in the humanities, 82. Umeå & Stockholm: Almqvist & Wiksell, 1987.
- Fähraeus, Rudolf. *Högre lärarinneseminariets historia*. Stockholm: Norstedt, 1943.
- Garshol, Lars Marius. "Living with topic maps and RDF". *XML Europe 2003*, May 5-8, London, 2003.

- Giordano, Richard. "The TEI header and the Documentation of Electronic Texts." i *Text encoding initiative: background and context*. red. Nancy Ide and Jean Vâeronis. Dordrecht; London: Kluwer Academic, 1995 (75-98).
- Goldfarb, Charles F. "SGML: The Reason Why and the First Published Hint." *Journal of the American Society for Information Science* 48.7 (1997).
- Goldfarb, Charles F., & Yuri Rubinsky. *The SGML handbook*. Oxford: Clarendon, 1990.
- Grensjö, Bengt, & Berthel Sutter. "Forskande lärande i skolans och arkivens värld." i *Arkiven i forskningens tjänst: 20 år med Forskningsarkivet vid Umeå universitet*. red. Göran Larsson. Umeå: Forskningsarkivet Umeå universitet, 2003 (160-167).
- Grepperud, Gunnar. *Fleksibel utbildning på universitets- og høyskolenivå: forventninger, praksis og utfordringer. 3: Fleksibel utbildning som utfordring*. Tromsø: UVETT Universitetets videre- og etteruddannelse Universitetet i Tromsø, 2005.
- Grundy, Isobel, Patricia Clements, Susan Brown, Terry Butler, Rebecca Cameron, Grey Coulombe, Susan Fisher, & Jeanne Wood. "Date and ChronStructs: Dynamic Chronology in the Orlando Project." *Literary and Linguistic Computing* 15.3 (2000): 265-289.
- Gunder, Anna. "Berättelsens spel - berättarteknik och ergodicitet i Michael Joyces afternoon, a story." *Human IT*:3 (1999).
- . "Forming the Text, Performing the Work - Aspects of Media, Navigation, and Linking." *Human IT*:2-3 (2001).
- Heckscher, Ebba. *Några drag ur den svenska flickskolans historia: under fleres medverkan samlade*. Stockholm: Norstedt & Söner, 1914.
- Hedberg, Valborg, & Louise Arosenius. *Svenska kvinnor från skilda verksamhetsområden: biografisk uppslagsbok*. Stockholm: Bonnier, 1914.
- Heyman, Ingrid. *Gänge hatt till: omvårdnadsforskningens framväxt i Sverige: sjuksköterskors avhandlingar 1974-1991*. Skeptronserien. Göteborg: Daidalos, 1995.
- . "Vilket underlag kan bidra till en trovärdig biografi? Källor, arkiv och övrigt material." i *Biografi och kollektivbiografi som historisk metod*. red. Agneta Linné and Kerstin Skog-Östlin. Meddelanden från Forum för pedagogisk historia, 5. Uppsala: Forum för pedagogisk historia SEC, Uppsala Universitet, 2001 (89).
- Heyman, Ingrid, Eva Trotzig, & Donald Broady. *Formering för offentlighet: rapport från konferens i Uppsala den 22 november 2000*. Meddelanden från Forum för pedagogisk historia, Nr 4. Uppsala: Forum för pedagogisk historia SEC ILU Univ., 2001.
- Hillyar, Anna, & Jane McDermid. *Revolutionary women in Russia, 1870-1917: a study in collective biography*. Manchester: Manchester University Press, 2000.
- Hockey, Susan M. *Electronic texts in the humanities: principles and practice*. Oxford: Oxford University Press, 2000.
- Hreinsson, Einar, Tomas Nilson, & Lars Hermanson. *Nätverk som social resurs: historiska exempel*. Lund: Studentlitteratur, 2003.
- Huitfeldt, Claus. "Multi-Dimensional Texts in a One-Dimensional Medium." *Computers and the Humanities* 28 (1995): 235-241.
- Ide, Nancy, & C. M. Sperberg-McQueen. "The TEI: History, Goals, and Future." i *Text encoding initiative: background and context*. Dordrecht; London: Kluwer Academic, 1995 (5-15).
- Ide, Nancy, & Jean Vâeronis, red. *Text encoding initiative: background and context*. Vol. Volym. Dordrecht; London: Kluwer Academic, 1995.

- Illich, Ivan. *Deschooling society*. World perspectives, 44. New York: Harper & Row, 1971.
- . "Education without school: How can it be done." *The New York Review of Books* 15.12 (1971).
- Initiative, Text Encoding. *The TEI Consortium: A Membership Prospectus*, 2001.
- IT-standardiseringsutredningen. *Den osynliga infrastrukturen: om förbättrad samordning av offentlig IT-standardisering: betänkande*. Statens offentliga utredningar, 2007:47. Stockholm: Fritze, 2007.
- IT och skola 2005. En sammanfattning av en enkätstudie riktad mot skolledare, lärare och elever under våren 2005.*: Skolverket, Myndigheten för skolutveckling, Sveriges kommuner och landsting, Stiftelsen för kunskap- och kompetensutveckling, Microsoft, 2005.
- Johansson, Egil. *En studie med kvantitativa metoder av folkundervisningen i Bygdeå socken 1845-1873*. Akademiska avhandlingar 2. Umeå: Pedagogiska institutionen, Umeå universitet, 1972.
- Johansson, Roger. *Ådalen 1931 - minuterna som blev historia: metodbok till en digital dokument-samling*. Lund: Studentlitteratur, 2006.
- Johansson, Veronica. "Elektroniska dokument i informationspolitisk belysning. En fråga om makt, kontroll och förhandlingar." *Human IT* 7.2 (2004): 197-264.
- Johns, Adrian. *The nature of the book: print and knowledge in the making*. Chicago: Univ. of Chicago Press, 1998.
- Kay, Martin. "Report on an Informal Meeting on Standard Formats for Machine-readable Text." i *Literary Data Processing Conference 1964 Proceedings*. red. Jess B. Bessinger. N.Y.: White Plains, 1965 (327-328).
- Kay, Michael. *XSLT: programmer's reference*. 2. uppl. Birmingham: Wrox Press, 2001.
- . *XSLT 2.0 programmer's reference*. 3rd uppl. Indianapolis, IN: Wiley Pub., 2004.
- KB-utredningen. *KB - ett nav i kunskapssambället: Kungl. biblioteket - Sveriges nationalbibliotek: verksamhet och visioner: betänkande*. Statens offentliga utredningar, 2003:129. Stockholm: Fritzes offentliga publikationer, 2003.
- Keats-Rohan, Katharine. *Resourcing sources*. Oxford: Linacre College (University of Oxford), Unit for Prosopographical Research, 2002.
- Kiernan, Kevin. "Digital facsimiles in editing." i *Electronic textual editing*. red. Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York, N.Y.: Modern Language Association of America, 2006 (262-268).
- Knorr-Cetina, Karin D. *Epistemic cultures: how the sciences make knowledge*. Cambridge, Mass.: Harvard University Press, 1999.
- Kronman, Ulf, & John Parnefjord. "Resource Description Framework. Metadata för Internet." *Human IT*.4 (1999).
- Kulturellt innehåll - digitalisering, elektronisk tillgång och digitalt bevarande*. Europeiska unionens officiella tidning, 2006.
- Lancashire, Ian. "Early Books, RET Encoding Guidelines, and the Trouble with SGML". *Scriptorium conference*, November, Calgary, 1995.
- . *Using TACT with electronic texts: a guide to text-analysis computing tools*. New York: Modern Language Association of America, 1996.
- Langerth Zetterman, Monica. *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet*. Rapporter från Forskningsgruppen för utbildnings- och kultursocio-

- logi. 1 uppl. 41. Uppsala: Forskningsgruppen för utbildnings- och kultursociologi, Under utgivning.
- Larsson, Esbjörn. *Att studera titlar som kapital: en teoretisk diskussion kring militära titlar i 1700-talets svenska samhälle*. 2003.
- . *Från adlig uppföstran till borgerlig utbildning: Kungl. krigsakademien mellan åren 1792 och 1866*. *Studia historica Upsaliensia*, 220. Uppsala: Acta Universitatis Upsalien-sis: Uppsala University Library distributör, 2005.
- Le Roux, Brigitte, & Henry Rouanet. *Geometric data analysis: from correspondence analysis to structured data analysis*. Dordrecht: Kluwer Academic, 2004.
- Levy, David M. *Scrolling forward: making sense of documents in the digital age*. 1. uppl. New York: Arcade Pub, 2001.
- Liliequist, Jonas. "Arkivet i undervisningen eller hur man kan få studenter att intressera sig för äldre handskrivet källmaterial." i *Arkiven i forskningens tjänst: 20 år med Forskningsarkivet vid Umeå universitet*. red. Göran Larsson. Umeå: Forskningsarki-vet Umeå universitet, 2003 (168-175).
- Linné, Agneta. "Lärarinnor, pedagogiskt arbete och offentlighet. Kvinnliga pionjärs liv och strategier kring år 1900 - ett jämförande perspektiv". *Svenska Histori-kermötet*, 22-24 april, Uppsala, 2005.
- Linné, Agneta, & Kerstin Skog-Östlin, red. *Biografi och kollektivbiografi som historisk metod: NFPF:s kongress 15-18 mars 2001 i Stockholm*. Vol. Volym. Uppsala: Forum för pedagogisk historia SEC, Uppsala Universitet, 2001.
- Lopez, Silbano de. "Kungliga Högre lärarinneseminariet och flickskolans framväxt." i *Sekelskiftets utmaningar: essäer om välfärd, utbildning och nationell identitet vid sekelskif-tet 1900*. red. Ann-Katrin Hatje. Stockholm: Carlsson, 2002 (181-198).
- Lundström, Nils Styrbjörn. *Svenska kvinnor i offentlig verksamhet: porträtt och biografier*. Upp-sala, 1924.
- Marshall, Catherine, & Bernheim B. Alice J. "Exploring the Relationship between Per-sonal and Public Annotations". *ACM/IEEE Joint Conference on Digital Libraries (JCDL04)*, June 7-11, Tucson, Arizona, 2004.
- Marshall, Catherine, & Frank Shipman. "Which Semantic Web?" *Fourteenth ACM Confe-rence on Hypertext and Hypermedia*, Nottingham UK, 2003.
- McCarty, Willard. "Humanities Computing: Essential Problems, Experimental Practi-ce." *Literary and Linguistic Computing* 17.1 (2002): 103-125.
- . "Knowing true things by what their mockeries be: modelling in the humanities." *Text Technology* 12.1 (2003).
- . *Humanities computing*. Basingstoke; New York: Palgrave Macmillan, 2005.
- McDermott, Drew, & Dejing Dou. "Representing Disjunction and Quantifiers in RDF". *First International Semantic Web Conference*, June 9-12, 2002.
- McGann, Jerome J. *Radiant textuality: literary studies after the World Wide Web*. New York; Basingstoke: Palgrave, 2001.
- . "Making texts of many dimensions." i *A companion to digital humanities*. red. Susan Schreibman, Raymond George Siemens and John Unsworth. Malden, MA; Oxford: Blackwell Pub., 2004.
- McKenzie, Donald. *Bibliography and the sociology of texts*. Cambridge: Cambridge Universi-ty Press, 1999.
- Melkersson, Hans. "Innehållsdesign. Om principer för märkning och lagring av inne-håll." i *Fältanteckningar: utbildnings- och kultursociologiska texter tillägnade Donald*

- Broady. red. Mikael Börjesson, Ingrid Heyman, et al. Skeptronserien. Uppsala: Forskningsgruppen för utbildnings- och kultursociologi ILU Uppsala universitet, 2006.
- Morrison, Alan, Michael Popham, & Karen Wikander. *Creating and Documenting Electronic Texts: A Guide to Good Practice*. Oxford: Oxford Text Archive, 2000.
- Müller, Eva, Uwe Klosa, Stefan Andersson, & Peter Hansson. "The DiVA Project - Development of an Electronic Publishing System." *D-Lib Magazine* 9.11 (2003).
- Mylonas, Ellen, & Allen H. Renear. "The Text Encoding Initiative at 10: Not Just an Interchange Format Anymore – But a New Research Community." *Computers and the Humanities* 33.1-2 (1997): 1-9.
- Nissen, Jörgen. *Pojkarna vid datorn: unga entusiaster i datateknikens värld*. Linköping studies in arts and science, 89. Stockholm; Stehag: Symposion Graduale, 1993.
- Ong, Walter J., & Lars Fyhr. *Muntlig och skriftlig kultur: teknologiseringen av ordet*. 2:a uppl. uppl. Göteborg: Anthropos, 1991.
- Padron-McCarthy, Thomas, & Tore Risch. *Databasteknik*. Lund: Studentlitteratur, 2005.
- Pepper, Steve. *The TAO of topic maps*, 2002.
- Petersens, Lovisa af. *Formering för offentlighet: kvinnokonferenser och Svenska kvinnornas nationalförbund kring sekelskiftet 1900*. Stockholm studies in history, 87. Stockholm: Acta Universitatis Stockholmiensis, 2006.
- Pettersson, Karin. *Den manliga heterosexuella normen: en analys av SAB-systemet utifrån ett feministiskt och queerteoretiskt perspektiv*. Borås: Högsk. i Borås Bibliotekshögskolan/Biblioteks- och informationsvetenskap, 2001.
- Pickering, Andrew. *Science as practice and culture*. Chicago: Univ. of Chicago Press, 1992.
- Piez, Wendell. "Beyond the "descriptive vs procedural" distinction". *Extreme Markup Languages 2001*, 2001.
- Rahtz, Sebastian, Norman Walsh, & Lou Burnard. "A unified model for text markup: TEI, Docbook, and beyond". *XML Europe 2004*, 18-21 April, Amsterdam, The Netherlands, 2004.
- Raymond, Eric S. *The cathedral and the bazaar. Musings on Linux and open source by an accidental revolutionary*. Rev. uppl. Sebastopol, Calif.: O'Reilly, 2001.
- Renear, Allen H. "Out of Praxis: Three (Meta)Theories of Textuality." i *Electronic text: investigations in method and theory*. red. Kathryn Sutherland. Oxford: Clarendon, 1997 (107-126).
- . "The Descriptive/Procedural Distinction is Flawed." *Markup Languages: Theory and Practice* 2.4 (2000): 411-420.
- . "Text from Several Different Perspectives, the Role of Context in Markup Semantics". *Proceedings of the 2003 Conference on Computers, Literature, and Philology*, Florence: University of Florence, 2003.
- . "Text Encoding." i *A companion to digital humanities*. red. Susan Schreibman, Raymond George Siemens and John Unsworth. Malden, MA; Oxford: Blackwell Pub., 2004 (218-239).
- Renear, Allen H., David G. Durand, & Ellen Mylonas. "Refining our notion of what text really is." i *Research in Humanities Computing 4: Selected papers from the 1992 ALLC/ACH Conference*. red. Susan M. Hockey and Nancy Ide. Oxford: Oxford University Press, 1997 (263-280).

- Roos, J.P., & Anna Rotkirch. "Fält i skuggan av fält: sensocialismens dubbelliv." i *Kulturens fält: Cultural fields: en antologi*. Skeptronserien. Göteborg: Daidalos, 1998 (453-466).
- Rundquist, Angela. *Blått blod och lifje vita händer: en etnologisk studie av aristokratiska kvinnor 1850-1900*. Stockholm: Carlsson, 1989.
- Rönnbäck, Josefin. *Politikens genusgränser: den kvinnliga rösträttsrörelsen och kampen för kvinnors politiska medborgarskap 1902-1921*. Atlas akademi. Stockholm: Atlas, 2004.
- Rönholm, Tord. *Kunskapens kvinnor: sekelskiftets studentskor i mötet med den manliga universitetsvärlden*. Forskningsrapporter från Institutionen för historiska studier vid Umeå universitet, 12. Umeå: Institutionen för historiska studier Univ. distribuerad, 1999.
- Sapiro, Gisele. "The structure of the French literary field during the German Occupation (1940-1944): a multiple correspondence analysis." *Poetics* 30 (2002): 387-402.
- Schreibman, Susan, Raymond George Siemens, & John Unsworth, red. *A companion to digital humanities*. Vol. Volym. Malden, MA; Oxford: Blackwell Pub., 2004.
- Shadbolt, Nigel, Wendy Hall, & Tim Berners-Lee. "The Semantic Web Revisited." *IEEE Intelligent Systems* 21.3 (2006): 96-101.
- Shillingsburg, Peter. "Principles for Electronic Archives, Scholarly Editions, and Tutorials." i *The literary text in the digital age*. red. Richard J. Finneran. Editorial theory and literary criticism. Ann Arbor: University of Michigan Press, 1996.
- Shillingsburg, Peter L. *From Gutenberg to google: electronic representations of literary texts*. Cambridge: Cambridge Univ. Press, 2006.
- Sowa, John F. *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove: Brooks/Cole, 2000.
- Sperberg-McQueen, Michael. "Classification and its Structures." i *A companion to digital humanities*. red. Susan Schreibman, Raymond George Siemens and John Unsworth. Blackwell companions to literature and culture; 26: Blackwell Pub., 2004 (160-176).
- . "Rabbit/duck grammars: a validation method for overlapping structures". *Extreme Markup Languages 2006*, Montréal, Québec, 2006.
- . "Representation of overlapping structures". *Extreme Markup Languages 2007*, Montréal, Québec, 2007.
- Sperberg-McQueen, Michael, & L. Burnard. "The Design of the TEI Encoding Scheme." i *Text encoding initiative: background and context*. red. Nancy Ide and Jean Vâeronis. Dordrecht; London: Kluwer Academic, 1995 (17-40).
- Statskontoret. *Öppen programvara*. Stockholm: Statskontoret, 2003.
- . *Öppen programvara. Erfarenheter av produkter som bygger på öppen källkod inom förvaltningen*. Stockholm: Statskontoret, 2004.
- Stenlås, Niklas. "Det sociala nätverkets logik eller vad händer i sociala nätverk?" i *Nätverk i historisk forskning: metafor, metod eller teori?* red. Peter Aronsson, Solveig Fagerlund and Jan Samuelson. Rapporter från Växjö universitet. Humaniora, 1. Växjö: Växjö universitet, 1999 (17-27).
- Stone, Lawrence. "Prosopography." i *Historical Studies Today*. red. Felix Gilbert and Stephen R. Graubard. New York: Norton, 1972 (107-140).
- Svedjedal, Johan. *The literary web: literature and publishing in the age of digital production: a study in the sociology of literature*. Skrifter utgivna av Avdelningen för litteratursociologi

- vid Litteraturvetenskapliga institutionen i Uppsala, 42. Stockholm: Kungl. bibl., 2000.
- . *Den sista boken*. Stockholm: Wahlström & Widstrand, 2001.
- . *En svensk Litteraturbank? Utredning för Svenska Akademien, avlämnad i januari 2003*, 2003.
- Sverige. Statistiska centralbyrån. *Från folkebrist till en åldrande befolkning: glimtar ur en unik befolkningsstatistik under 250 år: fakta inför 2000-talet*. Stockholm: Statistiska centralbyrån (SCB), 1999.
- Säljö, Roger, & Maria Södling. *Utbildning på vetenskaplig grund: röster från fältet*. Stockholm: Höskoleverket, 2006.
- Sällskapet Nya Idun 1885-1935: femtioårshistorik: medlemsförteckning*. Stockholm: Sällskapet Nya Idun, 1935.
- TEI Consortium, red. *TEI Consortium. Guidelines for Electronic Text Encoding and Interchange*. P5. Vol. Volym, 2007.
- Tennison, Jeni. *Beginning XSLT 2.0: from novice to professional*. Berkeley, CA: Apress, 2005.
- Tennison, Jeni, & Wendell Piez. "The Layered Markup and Annotation Language (LMNL)". *Extreme Markup Languages 2002*, Montréal, Québec, 2002.
- Thompson, John B. *Books in the digital age: the transformation of academic and higher education publishing in Britain and the United States*. Cambridge: Polity, 2005.
- Thorsell, Lennart. "Den svenska parnassens 'demokratisering' och de folkliga bildningsvägarna." i *Samlaren*. Uppsala, 1958 (53-135).
- Tosh, John. *Historisk teori och metod*. 2., rev. uppl. uppl. Lund: Studentlitteratur, 2000.
- Ullman, Annika. *Stiftarinnegenerationen: Sofi Almqvist, Anna Sandström, Anna Ahlström*. Stockholm: Stockholmia, 2004.
- Vanhoutte, Edward. "An introduction to the TEI and the TEI Consortium." *Literary and Linguistic Computing* 19.1 (2004): 9-19.
- Werner, Karl Ferdinand. "L'apport de la prosopographie à l'histoire sociale des élites." i *Family trees and the roots of politics: the prosopography of Britain and France from the tenth to the twelfth century*. red. K. S. B. Keats-Rohan. Woodbridge: Boydell Press, 1997 (1-21).
- Vetenskapsrådet. *Om forskningens infrastrukturer inom humaniora och samhällsvetenskap i Sverige*. Vetenskapsrådet, 2005.
- Wieselgren, Greta. *Den höga tröskeln: kampen för kvinnas rätt till ämbete*. Kvinnohistoriskt arkiv, 7. Lund: Gleerup, 1969.
- Winograd, Terry. "Thinking machines: Can there be? Are we?" i *The Boundaries of Humanity: Humans, Animals, Machines*. red. James Sheehan and Morton Sosna. Berkeley: University of California Press, 1991 (198-223).
- Zetterman, Monica Langerth. "Using Markup for Multivariate Analyses in the Prosopographical Study "Formation for the Public Sphere"." *ACH/ALLC Conference*, June 15 - June 18, Victoria University, Canada, 2005.
- . "Människor av god vilja" - om utbildningens betydelse för kvinnor kring sekelskiftet 1900". *Tredje pedagogikhistoriska konferensen*, 28-29 september, Lärarhögskolan i Stockholm, 2006.
- . "Märkvärt. Noteringar om en prosopografi över kvinnor i offentlighet kring sekelskiftet 1900." i *Fältanteckningar: utbildnings- och kultursociologiska texter tillägnade Donald Broady*. red. Mikael Börjesson, Ingrid Heyman, et al. Skeptronserien.

- Uppsala: Forskningsgruppen för utbildnings- och kultursociologi ILU Uppsala universitet, 2006.
- . *Kvinnors mötesplatser och sociala och kulturella tillgångar kring sekelskiftet 1900. Beskrivning av prosopografiskt material inom projektet Formering för offentlighet*. Rapporter från Forskningsgruppen för utbildnings- och kultursociologi. 1 uppl. 41. Uppsala: Forskningsgruppen för utbildnings- och kultursociologi, 2008.
- Österberg, Eva, & Christina Carlsson Wetterberg, red. *Rummet vidgas: kvinnor på väg ut i offentligheten 1880-1940*. Vol. Volym. Stockholm: Atlantis, 2002.

Acta Universitatis Upsaliensis UPPSALA STUDIES IN EDUCATION

Editors: Donald Broady, Christina Gustafsson, Sverker Lindblad, Ulf P. Lundgren, Ulla Riis

1. *Inga Elgqvist-Saltzman* 1976: Vägen genom universitetet. En forskningsöversikt och en empirisk analys av några studievägar inom filosofisk fakultet i Umeå 1968–1975.
2. *Sverker Lindblad* 1976: Simulering i undervisning. En kritisk granskning av artificiell praktik i allmänutbildning.
3. *Urban Dahlöf* 1977: Reforming Higher Education and External Studies in Sweden and Australia.
4. *Leif Isberg* 1978: Arbetsuppgifter, krav och utbildning. En studie rörande domarproblematiken inom lagspelet fotboll.
5. *Mats Myrberg* 1978: Towards an Ergonomic Theory of Text Design and Composition.
6. *Henry Werlinder* 1978: Psychopathy: A History of the Concepts. Analysis of the Origin and Development of a Family of Concepts in Psychopathology.
7. *Håkan Jenner* 1979: Mål och verklighet i ett terapeutiskt samhälle. En processoch effektstudie av en behandlingsinstitution för drogmissbrukare.
8. *Christer Stensmo* 1979: Kollektivet Trollängen. En utvärdering på organisationsnivå.
9. *Lars Ryhammar* 1979: Barn och trafikundervisning. En undersökning i grundskolans årskurser 3–6.
10. *Agnes Nobel* 1979: Boken i skolan. En analys med särskild inriktning på bibliotekets funktion i grundskolan.
11. *Annika Andre* 1980: Grundskola i glesbygd. Försöksverksamhet med årskurslöst högstadium.
12. *Britta Ericson* 1980: Lässvårigheter och emotionell störning.
13. *Karl-Georg Ahlström & Maud Jonsson* 1980: Flexible Grouping of Pupils and Team-work between Teachers.
14. *Ola Westin* 1981: On Political Socialization and Education: Investigations into an Argumentation for a Good Political Belief System.
15. *Gunnar Berg* 1981: Skolan som organisation. En analys av skolans organisatoriska struktur i ett förändringsperspektiv.
16. *Birgitta Willén* 1981: Distance Education at Swedish Universities. An Evaluation of the Experimental Programme and a Follow-up Study.
17. *Magnus Söderström* 1981: Personalutveckling i arbetslivet och återkommande utbildning. Några problem och utvecklingsmöjligheter.
18. *Lars Holmstrand* 1982: English in the Elementary School: Theoretical and Empirical Aspects of the Early Teaching of English as a Foreign Language.
19. *Lame Strömberg Sölveborn* 1983: Doktoranders studiesituation. En undersökning vid samhällsvetenskapliga fakulteten, Uppsala universitet.
20. *Max Scharnberg* 1984: The Myth of Paradigm-shift, or How to Lie with Methodology.
21. *Margareta Björndahl* 1984: Klassföreståndares inställning till daglig elevvård. En analys av uppfattningar och pedagogiskt handlande i sexton högstadielklasser.
22. *Bert Ståhlhammar* 1984: Rektorsfunktionen i grundskolan. Vision – verklighet.
23. *Gunnar Richardsson, Paul Lönnström & Staffan Smedberg* 1984: Utbildningshistorisk forskning. Problem, källmaterial, metodik.
24. *Ulla Holm* 1985: Empati i läkar-patientrelationen. En teoretisk och empirisk analys.
25. *Tomas Englund* 1986: Curriculum as a Political Problem.
26. *Silva Sköld* 1987: FAL – Fortbildning i Administration för Läkare. Studie av ett personalutvecklingsprojekt.
27. *Gertrud Grabn* 1987: Educational Situations in Clinical Settings. A Process Analysis.
28. *Stellan E. Ljfdahl* 1987: Fysikämnet i svensk realskola och grundskola. Kartläggning och alternativ ur fysikdidaktisk synvinkel.
29. *Björn Wallin* 1988: Frostigmetoden. Litteraturgranskning och två empiriska studier.
30. *Rune Axelsson* 1989: Upper Secondary School in Retrospect. The View of Former Students.

31. *Agneieszka Bron-Wojciechowska* 1989: Workers and Post-Secondary Education. A Cross-Policy Perspective.
32. *Rikard Palmér* 1989: Utvecklingsbetingelser hos utvecklingsstörda. En longitudinell studie.
33. *Gösta Wennberg* 1990: Geografi och skolgeografi. Ett ämnes förändringar. En studie med exempel.
34. *L. Kathrine Käller* 1990: Fostran till andrarang. En studie av dominansprocessen vid skolstart och via vägar genom utbildningssystemet ur ett kvinnovetenskapligt perspektiv.
35. *Per Hansson* 1990: Styrning och kultur. En studie om förändringsbetingelser i kyrklig församlingsverksamhet.
36. *Anders Engquist* 1990: Utvecklingssamtal som förändringsinstrument. Teoretisk och empirisk analys.
37. *Rolf Helldin* 1990: Den unika speciallärarkompetensen? – Några speciallärares uppfattningar och relationen till emotionellt och socialt missgynnade ungdomar.
38. *Arabella Wennerström Nyland* 1991: Oriental Learning and Western Knowledge. The Encounter of Educational Traditions in Bengal 1781–1835.
39. *Bo Nestor* 1991: Studieledare i Grundskolan. En skolledningsfunktion med förhinder?
40. *Anna Löthman* 1992: Om matematikundervisning - innehåll, innebörd och tillämpning. En explorativ studie av matematikundervisning inom kommunal vuxenutbildning och på grundskolans högstadium belyst ur elev- och lärarperspektiv.
41. *Elisabet Hammar* 1992: « La Française. » Mille et une façons d'apprendre le français en Suede avant 1807.
42. *Birgitta Ekman* 1992: Livsvillkor, livsformer, utbildning. En kommunstudie i pedagogiskt perspektiv.
43. *Héctor Pérez Prieto* 1992: Skola och erfarenhet: elevernas perspektiv. En grupp elevers skolerfarenheter i ett longitudinellt perspektiv.
44. *Abelardo Castro Hidalgo* 1992: Det traditionella och det "andra" universitetet. En början till samspel?
45. *Sonia Bentling* 1992: I idéernas värld. En analys av omvårdnad som vetenskap och grund för en professionell utbildning.
46. *Christina Andersson* 1993: The Children of Maria. Adolescent Substance Abusers, their Families and Schooling.
47. *Max Scharnberg* 1993: The Non-Authentic Nature of Freud's Observations. Vol. I: The Seduction Theory.
48. *Max Scharnberg* 1993: The Non-Authentic Nature of Freud's Observations. Vol. II: Felix Gattel's Early Freudian Cases, and the Astrological Origin of the Anal Theory.
49. *Monica Wåglund* 1993: Orientering mot ledarskap. Analys av en chefsutbildning för läkare.
50. *Thomas Tydén* 1993: Knowledge Interplay. User-Oriented Research Dissemination through Synthesis Pedagogics.
51. *Leif Ribom* 1993: Föräldraperspektiv på skolan. En analys från två håll.
52. *Mehdi Sedigh Zadeh* 1994: Att studera i ett främmande land. En studie av motiv och studiesituation bland utländska studerande vid svenska högskolor.
53. *Carl Anders Sjöström* 1994: Makt och mening. Förutsättningar för en innehållsfokuserad pedagogisk forskning.
54. *Birgitta Almquist* 1994: Approaching the Culture of Toys in Swedish Child Care. A Literature Survey and a Toy Inventory.
55. *Kjell Gustavsson* 1994: Vad är idrottandets mening? En kunskapssociologisk granskning av idrottens utveckling och läromedel samt en organisationsdidaktisk kompetensanalys.
56. *Urban Dabblöf and Staffan Selander* (eds.) 1994: New Universities and Regional Context. Papers from an International Seminar held at Sundsvall, Sweden, 14–18 June, 1992.
57. *Gunilla Roos* 1994: Kommunerna och det pedagogiska utvecklingsarbetet inom barnomsorgen. Omfattning, inriktning och villkor.
58. *Ulf Nyttell* 1994: Styra eller styras? En studie av skolledares arbete och arbetsvillkor.
59. *Gunnar Finnbogason* 1994: Från utbildningsplanering till kursplaner. Den isländska grundskolereformen 1974.
60. *Håkan Sandberg* 1995: Att få andas och kunna möta. Förutsättningar för målorientering inom öppen barn- och ungdomspsykiatrisk verksamhet.

61. *Leif Östman* 1995: Socialisation och mening. No-utbildning som politiskt och miljömoraliskt problem.
62. *Gunilla Lindqvist* 1995: The Aesthetics of Play. A Didactic Study of Play and Culture in Preschools.
63. *Catharina Andersson* 1995: Läras för skolan eller skolas att lära. Tankemodeller i lärarutbildning.
64. *Max Scharnberg* 1996: Textual Analysis: A Scientific Approach for Assessing Cases of Sexual Abuse. Vol. I: The Theoretical Framework, the Psychology of Lying, and Cases of Older Children.
65. *Max Scharnberg* 1996: Textual Analysis: A Scientific Approach for Assessing Cases of Sexual Abuse. Vol. II: Cases of Younger Children, Including a Case of Alleged Necrophilia, and the Shortcomings of Judicial Logic.
66. *Urban Dabblöf* and *Staffan Selander* (Eds.) 1996: Expanding Colleges and New Universities. Selected Case Studies from Non-metropolitan Areas in Australia, Scotland and Scandinavia.
67. *Evanthia Kalpazidou Schmidt* 1996: Forskningsmiljöer i ett nordiskt perspektiv. En komparativ studie i ekologi och kunskapsproduktion.
68. *Carsten Ljunggren* 1996: Medborgarpubliken och det offentliga rummet. Om utbildning, medier och demokrati.
69. *Lars-Göran Carlsson* 1997: Perspektiv på barnuppfostran. En studie av föräldraskap i kristen miljö.
70. *Anders Garpelin* 1997: Lektionen och livet. Ett möte mellan ungdomar som tillsammans bildar en skolklass.
71. *Kajsa Falkner* 1997: Lärare och skolans omstrukturerings. Ett möte mellan utbildningspolitiska intentioner och grundskollärares perspektiv på förändringar i den svenska skolan.
72. *Ulf Olsson* 1977: Folkhälsa som pedagogiskt projekt. Bilden av hälsoupplysning i statens offentliga utredningar.
73. *Britta Wingård* 1998: Att vara rektor och kvinna.
74. *Michael A. A. Wort* 1998: Distance Education and the Training of Primary School Teachers in Tanzania.
75. *Eva Österlind* 1998: Disciplinering via frihet. Elevers planering av sitt eget arbete.
76. *Lillemor Kim* 1998: Val och urval till högre utbildning. En studie baserad på erfarenheterna av 1977 års tillträdesreform.
77. *Maud Söderlund* 1998: "En mänsklig atmosfär". Trygghet, samhörighet och gemenskap – God vård ur ett patientperspektiv.
78. *Anders Andersson* 1999: Mångkulturalism och svensk folkhögskola. Studie av en möjlig mötesarena.
79. *Else-Maj Falk* 1999: Lärare tar gestalt. En hermeneutisk studie av texter om lärarblivande på distans.
80. *Finn Calander* 1999: Från fritidens pedagog till hjälplärare: Fritidspedagogers och lärares yrkesrelation i integrerade arbetslag.
81. *Valdy Lindbe* 1999: Greening Education. Prospects and Conditions in Tanzania.
82. *Inger Eriksson* 1999: Lärares pedagogiska handlingar. En studie av lärares uppfattningar av att vara pedagogisk i klassrumsarbetet.
83. *Birgitta Lidholt* 1999: Anpassning, kamp och flykt. Hur förskolepersonal handskas med effekter av besparingar och andra förändringar i förskolan.
84. *Roland Forsberg* 1999: Musik som livsstil och vetenskap. Studerandes möte med universitetsämnet musikvetenskap.
85. *Fritjof Sahlström* 1999. Up the Hill Backwards. On Interactional Constraints and Affordances for Equity-Constitution in the Classrooms of the Swedish Comprehensive School.
86. *Jan-Olof Hellsten* 2000: Skolan som barnarbete och utvecklingsprojekt. En studie av hur grundskoleelevers arbetsmiljö skapas – förändras – förblir som den är.
87. *Inger Ahlgren* och *Annalisa Gummeson* 2000: Utbildningsplaner och reformimplementering. Sjuksköterskeutbildningen i fokus.
88. *Mia Marie F. Sternudd* 2000: Dramapedagogik som demokratisk fostran? Fyra dramapedagogiska perspektiv – dramapedagogik i fyra läroplaner.

89. *Joyce Kemuma* 2000: The Past and the Future in the Present. Kenyan Adult Immigrants' Stories on Orientation and Adult Education in Sweden.
90. *Evelyn Säll* 2000: Lärarrollens olika skepnader; estradör, regissör och illuminatör. En longitudinell studie av blivande lärares föreställningar.
91. *Jonas Gustafsson* 2000: Så ska det låta. Studier av det musikpedagogiska fältets framväxt i Sverige 1900–1965.
92. *Ulrika Tornberg* 2000: Om språkundervisning i mellanrummet – och talet om ”kommunikation” och ”kultur” i kursplaner och läromedel från 1962 till 2000.
93. *Eva Forsberg* 2000: Elevinflytandets många ansikten.
94. *Karin Hjalmskog* 2000: ”Democracy begins at home”. Utbildning om och för hemmet som medborgarfostran.
95. *Anders Westlin* 2000: Teknik och politiskt handlande. Rationalitet och kritik i den samhällsorienterande undervisningen.
96. *Elvy Westlund* 2001: Undervisnings- och utbildningsplanering – för vem? En studie av ingenjörsutbildning vid en svensk högskola.
97. *Guy Karnung* 2001: Rösterna om kvalitativ forskning. En karaktäristik utifrån vetenskapliga texter.
98. *Irène Lind Nilsson* 2001: Ledarskap i kris, kaos och omställning. En empirisk studie av chefer i företag och förvaltning.
99. *Kent Ericsson* 2002: From Institutional Life to Community Participation. Ideas and Realities Concerning Support to Persons with Intellectual Disability.
100. *Anna Henningsson-Yousif* 2003: Skolperspektiv – utveckling av redskap för analys av politikers, lärares och elevers resonemang om skolan.
101. *Pia-Maria Ivarsson* 2003: Barns gemenskap i förskolan.
102. *Elisabet Nihlfors* 2003: Skolchefen i skolans styrning och ledning.
103. *Katarina Wijle* 2003: Planning and Implementing Health Interventions. Extrapolating Theories of Health Education and Constructed Determinants of Risk-Taking.
104. *Margareta Svennbeck* 2004: Omsorg om naturen. Om NO-utbildningens selektiva traditioner med fokus på miljöfostran och genus.
105. *Annalill Ekman* 2004: Lärande organisationer i teori och praktik. Apoteket lär.
106. *Christina Hellblom-Thibblin* 2004: Kategorisering av barns ”problem” i skolans värld. En undersökning av skolhälsovårdsrapporter läsåren 1944/45–1988/89.
107. *Cathrin Martin* 2004: From *other* to *self*. Learning as Interactional Change.
108. *Niclas Månsson* 2004: Negativ socialization. Främlingskapet i Zygmunt Baumanns författarskap.
109. *Per Ekstrand* 2005: ”Tarzan och Jane”. Hur män som sjuksköterskor formar sin identitet.
110. *Mia Heikkilä* 2006: Kommunikativa resurser för lärande. Barns gester, blickar och tal i tre skolmiljöer.
111. *Katarina Gustafson* 2006: Vi och dom i skola och stadsdel. Barns identitetsarbete och sociala geografier.
112. *Marie Karlsson* 2006: Föräldraidentiteter i livsberättelser.
113. *Henrik Román* 2006: Skönheten och nyttan. Om synen på gymnasiesvenskans litteraturundervisning 1947–1985.
114. *Hans Nyttell* 2006: Från kvalitetsidé till kvalitetsregim. Om statlig styrning av skolan.
115. *Carina Fast* 2007: Sju barn lär sig läsa och skriva. Familjeliv och populärkultur i möte med förskola och skola.
116. *Carina Carlbed* 2007: Medicinens lyskraft och skuggor – om trosföreställningar och symbolisk makt i habiliteringen 1960–1980.
117. *Monica Langerth Zetterman* 2008: Innehållsdesign – principer, metoder och verktyg samt tillämpningar inom utbildningshistorisk forskning och undervisning.