

# Det nya handbiblioteket <sup>1</sup>

## Forskningsprogram, version 2

April 1993

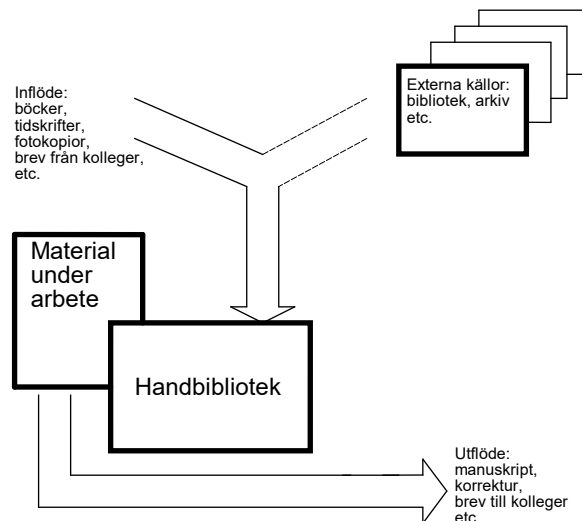
Donald Broady  
IPLab, Institutionen för numerisk analys och datalogi  
KTH, 100 44 Stockholm

### 1. Handbiblioteket

En tumregel är att datortekniken kommer till korta, eller till ingen användning alls, om den förhindrar människor att göra det de brukar göra med pappersburen information. En humanist eller samhällsvetare utvecklar med åren sofistikerade sätt att läsa och skriva och hålla ordning på sina anteckningar, pappershögar och samlingar av böcker och tidskrifter. Ett i sammanhanget brukbart dokumenthanteringssystem måste bereda utrymme för dessa mödosamt förvärvade förmågor — självfallet utan illusioner om att de maskinläsbara dokumenten skulle ersätta böcker eller tidskrifter. Vid mer intensivt umgänge med en text brukar man önska tillgång till den såväl på papper, för överblickens och de estetiska kvaliteternas skull, som i maskinläsbar form, för kunna bearbeta den, kommentera, indexera, genomsöka, extrahera, kompilera etc.

Ett bra handbibliotek är en oskattbar tillgång för den som arbetar som exempelvis lärare, forskare, redaktör, översättare eller dokumentalist. Med handbibliotek avser jag helt enkelt den samling av material som man önskar ha nära till hands i arbetsrummet därhemma eller på arbetsplatsen.

Jag har visserligen mött filosofer som påstår sig behöva föga mer än huvudet och papper och penna för att kunna arbeta, men vanligare är att man är beroende av ett rum eller åtminstone ett par bokhyllor med det material man önskar ha på armlängds avstånd: speciallitteratur, allmän referenslitteratur såsom handböcker, lexika och ordböcker, mappar med egna och kollegers skrifter, elevarbeten, egna halvfärdiga texter ...



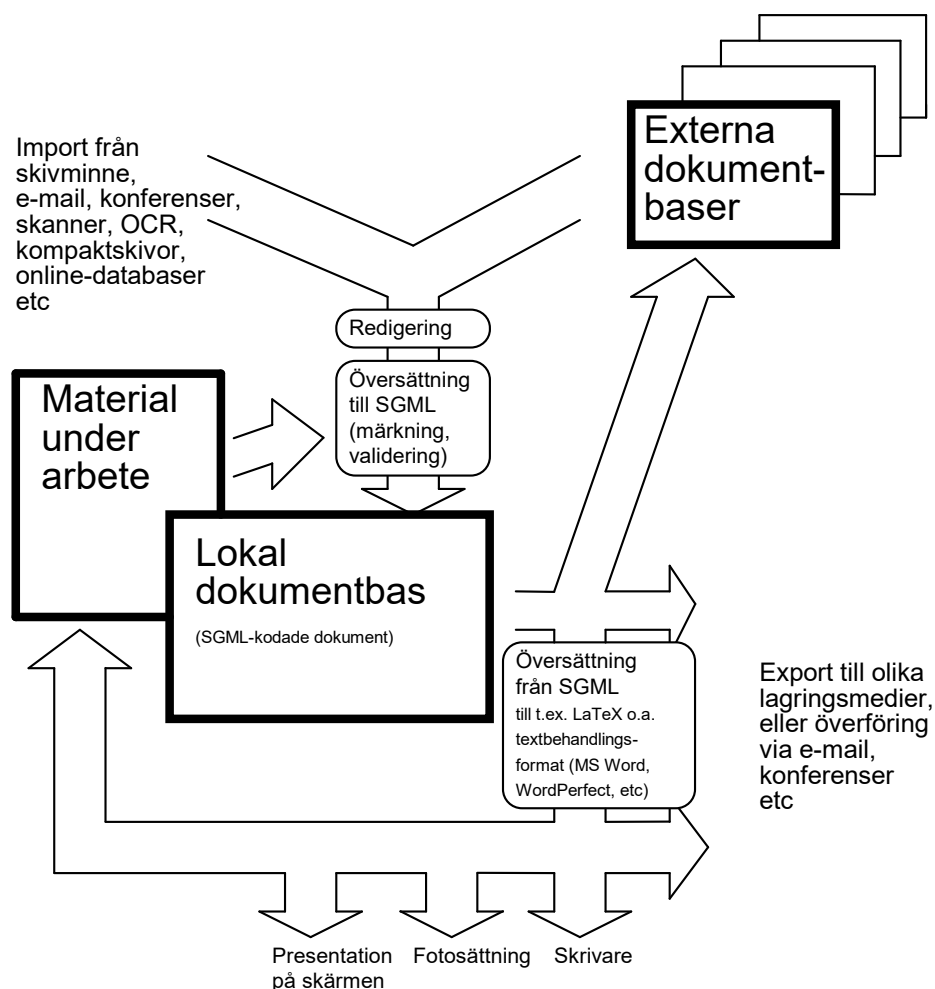
Figur 1. Flödena i ett ordinärt arbetsrum, utan särskilt mycket datorstöd

<sup>1</sup> Denna text består av reviderade utdrag ur ansökan (daterad 15 februari 1993) till ett forskningsprojekt "Det nya handbiblioteket", ställd till Forskningsrådsnämnden och Skolverket. Medsökande är Lars Dahlbäck, huvudredaktör för Nationalupplagan av August Strindbergs samlade verk, Eva Trotzig, chefsbibliotekarie vid Statens Psykologisk-Pedagogiska Bibliotek, samt Hasse Haitto, IPLab, NADA, KTH.

Arbetet bygger på erfarenheter från ett tidigare projekt "Datorstöd för kunskapsarbete", som avslutas den 30 juni 1993 och finansierats av NUTEK (ITYP 90-02737P, "Datorstöd för författande och samarbete i utbildning"), Forskningsrådsnämnden och Skolverket (delprojektet "Kunskapsverkstaden"). Mina medarbetare i detta projekt har varit Hasse Haitto, Peter Lidbaum och Magnus Tobiasson, samtliga verksamma vid IPLab, NADA, KTH.

Många lärare och forskare torde utnyttja sitt handbibliotek ungefär på det sätt som illustreras av Figur 1 härintill. In flyter böcker och tidskrifter, fotokopior, rapporter och manuskript som man ordnar på hyllor eller i travar, holkar, pärmar, kortlådor, samt datafiler lagrade på något magnetiskt eller mer sällan optiskt medium (diskett, skivminne i den egna datorn eller i en server i ett nätverk, kompaktskiva). Utöver det egentliga handbiblioteket har man på sina bord och hyllor material under arbete, dvs. ofärdiga manuskript och utkast, anteckningar och excerpter, material med anknytning till pågående forsknings- och undervisningsuppgifter, allt detta antingen på papper eller på magnetiskt medium. Man har även tillgång till externa källor såsom bibliotek, arkiv, databasvärdar, vilka inte finns till hands i varje ögonblick utan kräver att man beger sig iväg någonstans eller anlitar postverket eller telefonnätet; det viktigaste materialet kan man sedan välja att införliva med handbiblioteket genom hemlån eller fotokopiering eller nedladdning av filer. Ut flyter papper eller datafiler med manuskript till artiklar och böcker, rapporter, undervisningsmaterial etc, varav något exemplar brukar införlivas med det egna handbiblioteket.

Figur 1 visar således flödena i ett arbetsrum vars innehavare inte gör särskilt intensivt bruk av datortekniken. Figur 2 nedan är en modell för ett tänkt dokumenthanteringssystem med väl utbyggt datorstöd.



Figur 2. Flöden vid arbete med maskinläsbara, företrädesvis SGML-märkta dokument

Även i arbetsmiljön som åskådliggjorts i figur 2 är "handbiblioteket" en hjärtpunkt. Det rör sig då inte om pappersburen information utan om maskinläsbart material samlat i en lokal dokumentbas. Denna dokumentbas är lokal såtillvida att en användare (eller arbetsgrupp) i alla avseenden förfogar över materialet, ungefär som över innehållet i de egna bokhyllorna.

Genom att likna den lokala dokumentbasen vid ett handbibliotek vill jag framhäva tillgängligheten. Dokumenten skall vara tillgängliga i flera hänseenden. De skall vara lätt åtkomliga, kanske lokalt lagrade på skivminne i den egna datorn eller i ett lokalt nätverk. De skall vid behov ofördröjligen kunna bearbetas med användarens egna favoritprogram (textbehandlingsprogram, bildbehandlingsprogram, kalkylprogram etc). De skall kunna kompletteras och förses med randanmärkningar och korshänvisningar (s.k. hypertextlänkar). Samlingen skall vara överblickbar, man skall utan stort besvär erhålla automatiskt genererade innehållsförteckningar eller kartor över dokumentbasens aktuella bestånd, och man skall på enkelt sätt kunna revidera inte blott de enskilda dokumenten utan även själva dokumentbasens struktur (ungefär som när man flyttar om böckerna på hyllorna eller sorterar pappren i pärmarna).

Med metaforen handbibliotek vill jag leda tankarna till hur lärare eller forskare brukar utnyttja hyllorna och borden i arbetsrummet. Jag menar inte att samlingarna behöver vara små och få. En digital dokumentbas kan rymma mycket mer än de papper som får plats i ett arbetsrum, och man kan tänka sig att en och samma person (eller arbetsgrupp) förfogar över ett flertal dokumentbaser, var och en avpassad för särskilda arbetsuppgifter.

Materialet i den lokala dokumentbasen kan hemföras från många håll och via många medier: från kompaktskivor, disketter eller band, via näten, eller från böcker och tidskrifter via skanner i kombination med program för optisk teckenigenkänning och bildbehandling. Därtill kommer material som man själv och kollegerna framställer.

Dessutom kan material hämtas från externa dokumentbaser, varmed jag här avser material som inte är omedelbart integrerat med den lokala dokumentbasen. Sådant material skulle kunna tillhandahållas av exempelvis forskargrupper, förlag eller bibliotek. Redan i dag förekommer att forskargrupper inom vissa discipliner gör sina uppsatser och rapporter allmänt tillgängliga via de internationella universitetsnäten.<sup>2</sup> Förutsatt att de f.n. besvärliga upphovsrättsproblemen löses kommer utan tvivel allt mer maskinläsbar litteratur att saluföras — i dag billigast och enklast på kompaktskivor — och därmed fungera som externa dokumentbaser, från vilka användaren kan välja material att införliva med sin egen lokala dokumentbas.

Var och en som försöker använda sådant material i större skala vet hur svårt det är eftersom formaten är så mångskiftande. Att få ut dokumenten på papper brukar inte vålla större besvär. De anländer vanligen som ASCII- eller PostScript-filer eller i något annat bekant format, men om det rör sig om många eller komplexa dokument krävs mycket tid och datorkunskande för att stöpa om dem så att de passar min maskin och mina favoritprogram — och trots alla ansträngningar händer ofta att dokumentens innehållsliga struktur (dvs relationerna mellan enheter som kapitel, rubriknivåer, fotnoter etc), radfallet eller de nationsspecifika tecknen och vissa specialtecken deformeras. Först i och med SGML, den internationella standarden för beskrivande märkning, blir det möjligt att jämförelsevis problemfritt och felfritt låta dokument med bibehållen struktur cirkulera tvärs över nationsgränserna, oberoende av olikheter i hårdvara, operativsystem, tillämpningsprogram, tillåtna filnamn etc, och (så länge vi håller oss till västerländska språk) oberoende av skillnader mellan nationella teckenuppsättningar. Mer om SGML nedan.

Det är väsentligt att den lokala dokumentbasen är nära integrerad med material under arbete. Således skall dokumentbasens material snabbt och smärtfritt kunna överföras till exempelvis användarens ordinära ord- eller bildbehandlingsprogram, där det kan redigeras, kommenteras, sammanföras med annat material under arbete etc. Omvänt skall det finnas möjligheter att införliva material av egen tillverkning med den lokala dokumentbasen, en överföring som dock med nödvändighet blir mer komplicerad och knappast kan automatiseras fullt ut.

<sup>2</sup> En distributionskanal som snabbt vunnit popularitet är World Wide Web, som tillåter den som letar efter information att följa hypertextlänkar tvärs över Internet och att med hjälp av bläddringsverktyget Mosaic inspektera innehållet i de arkiv som forskargrupper och andra ställer till allmänt förfogande.

Från denna arbetsmiljö levereras materialet på skilda sätt: för direkt skärmpresentation (med presentationsverktyg såsom dem vi själva utvecklar vid KTH), till enkel utskrift på skrivare eller till TeX eller LaTeX för mer sofistikerad laserutskrift eller fotosättning, för distribution på magnetiskt medium (diskett, band) eller över näten meddelst elektronisk post, konferenssystem eller ftp.

Den lokala dokumentbasen kan även göra tjänst som hjärtpunkten i en skivrarverkstad, där ett lärarkollegium, en redaktionsgrupp, en grupp läromedelsförfattare eller en forskargrupp framställer mer genomarbetade och omfattande dokumentsamlingar vilka förtjänar att spridas till andra. Dessa samlingar kan därmed, för att använda terminologin i figur 2, ingå i utbudet av offentligt tillgängliga "externa dokumentbaser".

Ett naturligt användningsområde för en lokal dokumentbas vore att ur den hämta råmaterial eller halvfabrikat med vars hjälp lärare förfärdigar skraddarsydd läromedel, dvs. kompendier där utsnitt ur dokumentbasen (med automatiskt genererade "källhänvisningar") kompletteras med lärarnas egna tillskott. Förlagen eller biblioteken, författare, redaktörer och lärarkolleger skulle kunna tillhandahålla dokumentbaser som är rikhaltiga och av god kvalitet (låt säga Strindbergs samlade verk jämte kommentarer och sekundärlitteratur), varur läraren framställer undervisningsmaterial.

Ett annat användningsområde är att bibliotekarier, dokumentalister eller informatiker samlar material i en lokal dokumentbas, för att därifrån leverera maskinläsbara dokument till användarna. Detta borde vara en naturlig arbetsuppgift för bibliotek, men så är knappast fallet i dag; bland maskinläsbara dokument lämpade för humanistisk undervisning och forskning som finns offentligt tillgängliga världen över tillhandahålles enligt en aktuell inventering (MRTI, Rutgers Inventory of Machine-Readable Texts in the Humanities) blott 5% av bibliotek. Man kunde tänka sig att ett skolbibliotek bygger upp en lokal samling att användas i arbetet med lokala arbetsplaner, terminsplaneringar, läromedelsframställning etc, eller att biblioteket vid en universitetsinstitution samlar digitala versioner av uppsatser, rapporter och undervisningsmaterial som medarbetarna där framställer, eller att forskningsbiblioteken beställer och tillhandahåller mer omfattande dokumentbaser.

Närmare samarbete mellan lärare och bibliotekarier vore önskvärt. En rimlig framtidsvision är att lärarna och forskarna från sin ordinarie arbetsstation har nätförbindelse med sin institutions eller skolas bibliotek, hämtar maskinläsbart material därifrån och omvänt förser biblioteket med de uppsatser och rapporter som förtjänar att inlemmas med beståndet av offentligt tillgängliga dokument. De pliktexemplar som sänds till forskningsbiblioteken borde i vissa fall kunna levereras i digital form. De maskinläsbara dokument som biblioteken tar emot bör självfallet vara märkta på så sätt att katalogiseringen i största möjliga utsträckning kan automatiseras.

De dokument som flyter till och från biblioteken bör levereras i flyttbart format, så att mottagarna utan alltför mycket besvär kan använda dokumenten på sin egen maskin och med hjälp av sina favoritprogram. Det är dessutom väsentligt att dokumentens struktur bevaras, dvs att de olika innehållsliga element varav dokumentet består (sådant som brödtext, fotnotstext, rubriknivåer, källhänvisningar, "front matter" såsom titel, författarnamn o.a. bibliografiska upplysningar) når mottagaren i det skick som författaren avsett — det gäller att undvika situationen att överföringen resulterar i en lång teckensträng, som mottagaren får ägna mycken möda åt att redigera. Allt detta talar för en märkning enligt SGML-standarden.

## 2. System för arbete med lokala dokumentbaser, önskvärda egenskaper

- Så få antaganden som möjligt om hur dokumenten kommer att användas.
- Så få antaganden som möjligt om vilka maskiner och program som skall ta hand om dokumenten.
- Stöd för arbete med — inte enbart för presentation av — dokumenten.
- Stöd för att revidera, ordna och överblicka dokumentbasen.
- Stöd för samarbete. Behörighetskontroll avseende både enskilda dokument och dokumentbasens organisering.
- Stöd för hantering av stora och komplexa dokument.
- En och samma dokumentsamling skall kunna användas av både noviser och specialister.
- Modifierbart och utbyggbart system: fler medier och fler program skall kunna tillföras.
- Åtskilligt arbete med dokumenten skall kunna ske med enkel utrustning.
- Systemet skall stödja hantering av strukturerade dokument. Beskrivande märkning, ej procedurmärkning.

Här följer en lista över krav som kan ställas på byggstenarna i ett dokumenthanteringssystem som det nyss skisserade. (I systemet SGML Darc som utvecklas vid KTH försöker vi efter förmåga att tillfredsställa de flesta av dessa krav.)

1. Systemutvecklarna bör göra så få antaganden som möjligt om hur dokumenten kommer att användas.

Användaren kanske vill läsa dokumenten på skärmen eller på papper, redigera, komplettera, extrahera fram material till ett kompendium, eller formge materialet typografiskt för utskrift på skrivare eller för vidare befordran till ett fotosätter. Dokumentsamlingen skall också kunna användas för sökning, kompilering eller innehållslig, semantisk, syntaktisk eller lexikalisk analys etc. Systemkonstruktören skall i så liten utsträckning som möjligt försöka förutsäga och styra användarnas behov.

2. Dokumenten skall var flyttbara.

Systemutvecklarna bör göra så få antaganden som möjligt om vilka maskiner och program användaren har eller föredrar. Användaren skall till vardags kunna arbeta med t.ex. sin favoritordbehandlare eller sin favoritgrafikeditor under sitt favoritoperativsystem (i dag vanligen MS Windows ver. 3x, någon unix-variant eller Macintosh System 7), samtidigt som dokumenten skall kunna flyttas till andra miljöer. Systemet bör uppmuntra spridning av materialet, dvs källtexterna själva bör vara åtkomliga och inte levereras i kompilerade versioner som man behöver särskilda program för att komma åt.

Här kolliderar lärarnas och forskarnas behov med de kommersiella intressena. Av naturliga skäl önskar programvaruleverantörer att man använder deras produkter och helst inga andra. Därför brukar den som vill publicera material i maskinläsbar form, exempelvis på kompaktskiva, mötas av förslag i stil med: —Utmärkt, vårt företag tar hand om era dokument och stoppar in dem i vårt system som ger suveräna möjligheter till indexering, informationsåtervinning, sökning etc och kostar någon tusenlapp per CD-ROM. (Därmed förfuskas idén med flyttbara dokument. Det är som om en författare som går upp på ett förlag med sitt manuskript skulle få beskedet: — Visst, vi skall gärna ut din bok, tryckt med osynligt bläck av vår egen tillverkning, vilket förutsätter att läsekretsen införskaffar våra specialglasögon för tusen kronor styck.)

Ett annat sätt att uttrycka samma sak är att hävda att dokumentet bör sättas i centrum. Hittills har man inom persondatorvärlden på sätt och vis satt programmet i centrum. Man har definierat dokumentet i termer av det tillämpningsprogram i vilket det skapats eller senast lagrats. Man talar således om ett WordPerfect-dokument eller ett Excel-dokument. Vidare definierar man sina dokumentsamlingar med hänvisning till maskinvara och operativsystem: man säger att man arbetar

under unix eller Windows eller på Macintosh. I många sammanhang vore det bättre att kunna sätta dokumentet i centrum. Dagens teknik börjar erbjuda vissa möjligheter att bearbeta ett och samma dokument eller en och samma dokumentsamling på mångahanda sätt och med många slag av program på många slag av maskiner. För en humanist eller samhällsvetare är detta ett naturligt sätt att arbeta: jag har en källtext eller en intervju som jag önskar bearbeta med skilda redskap, jag vill inte först starta en tillämpning och därefter öppna en fil som har mening uteslutande för mitt operativsystem eller rentav för ett enda enskilt program.

3. Systemet bör stödja samarbete. Flera medarbetare skall kunna använda samma eller varandra överlappande lokala dokumentbaser.

Här är behörighetskontroll väsentlig, och denna bör gälla såväl varje enskilt dokument som hela dokumentbasens organisering. I en redaktionsgrupp eller forskargrupp kan huvudredaktören eller den vetenskaplige ledaren tilldelas vittgående befogenheter, kanske t.o.m. rätt att organisera hela dokumentbasen och att ändra allt innehåll, medan andra medarbetare med mer begränsade befogenheter ges rätt att redigera enskilda dokument, och åter andra endast rätt att läsa vissa dokument och att för eget bruk infoga korshänvisningar och skapa egna personliga kartor över basens innehåll.

4. Systemet bör stödja arbete med — inte enbart presentation av — dokumenten.

Någon läsare kanske undrar varför ordet multimedia inte förekommit i denna framställning. Det kan synas en smula otidsenligt att lägga så stor vikt vid texthantering som här sker.

Jag tror att det är olyckligt och onödigt att utvecklingen av de basala arbetsverktygen för hantering av text och stillbilder har hamnat en smula i skuggan av den mer uppseendeväckande utvecklingen av presentationsverktygen. När termen "multimedia" introducerades innebar den helt enkelt att en personator presenterar text, stillbilder, animerade bilder, ljud och videosekvenser, vilket lät sig åstadkommas genom att videospelare och högtalare kopplades till datorn. I dag har ambitionerna vuxit. Man strävar efter att bygga digitala system som mest av allt liknar en professionell videoproduktionsstudio. Man skall med andra ord samtidigt hantera ett flertal kanaler med video- och ljudinformation som kan spelas in, bearbetas, spelas in på nytt, spelas upp och sammanfogas till ett enda program. Ett sådan utrustning ger förnämliga möjligheter när det gäller att presentera en slutprodukt. Det är dock få förunnat och sällan mödan värt att hantera det egna föränderliga arbetsmaterialet på så sätt. Att skapa en fullfjädrad multimediassession kräver knappast mindre kompetens — tekniskt kunnande, självfallet, men i synnerhet kunskaper om formgivning och bildmässiga och musikaliska uttryck — än att producera en musikvideo. Bl.a. på grund av att ljudmaterial och videosekvenser är tidsberoende information och inte "maskinläsbar" i samma mening som text och vanlig grafik, är det dessutom svårt att automatisera sökning, informationsåtervinning, konvertering och liknande rutinuppgifter. Utan tvivel kommer det med tiden att bli allt lättare även för amatörer att någorlunda snabbt redigera och kombinera digitalt lagrad ljudinformation och digitaliserade rörliga bilder, men i dag utgör redan kraven på lagringsminne ett svåröverkomligt hinder. Digitalt lagrat ljud i full bandvidd kräver ca 5 MB per minut för monoåtergivning och 10 MB per minut i stereo, och en minut digitaliserad videosekvens i färg och fullskärmsformat kräver inte mindre än 1½ GB i lagringsutrymme och därtill en synnerligen snabb datorutrustning. Trots att komprimeringsalgoritmerna förfinas och lagringsutrymmet blir billigare torde det dröja ännu någon tid innan lärare och andra ordinära datoranvändare får plats med särskilt mycket digitaliserad ljud- och videoinformation på det lagringsmedium där de har sitt material under arbete.

Utveckling av system som företrädesvis arbetar med tidsberoende information, dvs rörliga bilder och ljud, är självfallet viktig men om ambitionen är att utveckla system där användarna skapar kvalificerat material av eget fabrikat, så är det det fråga om ett långsiktigt arbete. Däremot kan förbättrade verktyg för hantering av dokument innehållande strukturerad text ge stora omedelbara vinster redan i dag. Även grafik och stillbilder kan hanteras någorlunda bekymmersfritt; med de nya effektiva komprimeringsprogrammen är det möjligt att med acceptabel visuell kvalitet lagra stillbilder med komprimeringsförhållandet 1:50 eller mer, vilket drastiskt minskat minneskraven (ett exempel är

JPEG från Joint Photographic Experts Group, en komprimeringsstandard som behåller den slags information som betyder mest för det mänskliga ögat och skär bort den information som ögat knappast uppfattar, såsom smärre skiftningar i tämligen enfärgade ytor etc). Därför förefaller det tills vidare angeläget att ägna kraft åt problemen rörande texthantering och enklare bild- och grafikhantering. Om inte dessa basala problem finner tillfredsställande lösningar och om inte rikhaltiga text- och bildbaser av god kvalitet blir tillgängliga i det dagliga arbetet, kommer maskinläsbara dokument aldrig att bli naturliga inslag i exempelvis lärares och elevers vardag, trots aldrig så välgjorda animeringar och hur vackert ljudet än klingar. Här krävs avsevärda satsningar med tanke på utbildningsväsendets behov — de stora investeringarna i dokumenthanteringssystem gäller i dag sådant som underhållsmanualer hos industriföretagen eller myndigheters ärendehantering.

5. Systemet skall hjälpa användaren att revidera, ordna och överblicka materialet i den lokala dokumentbasen.

Här vill jag i synnerhet lyfta fram överblicksproblemet, som var och en som samlar sitt arbetsmaterial på ett skivminne är förtrogen med. Om det vill sig illa fungerar hårddisken som en *geniza*, det slutna rum med en liten öppning i taket där förbrukade papper och pergament med hebreisk skrift kastades ned eftersom det som skrivits på detta heliga språk enligt judisk tradition inte fick utplånas.

Även i fråga om de enskilda dokumentens innehåll gör sig överblicksproblemet snabbt påmint. Så länge jag handskas med pappersburen information utnyttjar jag mer eller mindre medvetet en rad knep för att skaffa mig överblick. Jag hittar rätt bok tack vare att jag minns dess plats i hyllan eller pärmens färg. Redan genom att ta den i handen erhåller jag en uppfattning om hur mycket den innehåller. Jag vet av erfarenhet ungefär var det lönar sig att leta efter tryckåret, innehållsförteckningen eller litteraturreferenserna. Om jag tidigare tittat i boken kanske jag bevarat ett synminne av att något stod högst upp på en högersida. Under läsningen vet jag hela tiden om jag befinner mig i början, mitten eller slutet av boken. Jag stoppar vid behov tummen i ett uppslag dit jag önskar återvända, jag lägger komihåglappar i boken, stryker under, skriver anteckningar i marginalen, viker hundöron eller spiller kaffe — ett helt spektrum av visuella och taktila orienteringsmöjligheter som saknas när informationen uppträder på skärmen. Nya slag av hjälpmedel behövs.

För att elektroniskt lagrad information verkligen skall bli *tillgänglig*, krävs nya redskap som hjälper både ”författare” och ”läsare” att orientera sig och att överblicka och organisera materialet. De i dag invanda sätten att presentera och använda pappersburen information är minst av allt naturgivna. Det dröjde sekler innan dagens typografiska konventioner för presentation av tryckt text fann sin form. De första generationerna boktryckare satte en ära i att imitera handskriften, och medeltida handskrifter ter sig för oss sena tiders barn nära nog ogenomträngliga redan på grund av textens massivitet. Konventionen att indela texten i stycken är en innovation (den medeltide skrivaren kunde föredra att fylla ut en halvfull rad med betydelslösa tecken för att erhålla en jämn högermarginal), liksom indelningen i kapitel och den systematiska användningen av rubriknivåer, olika typsnitt, sidnumrering, innehållsförteckningar, sakordsregister eller fotnoter. Allt det som vi idag förväntar oss av en läslig och läsbar tryckt text är frukten av en lång utveckling av såväl typografin som människors varseblivningsförmåga. I fråga om elektronisk information har en motsvarande utveckling av formgivningskonventioner och formkänsla nätt och jämt börjat.

Systemet bör förse användaren med orienteringshjälpmedel, sådant som kartor över hur innehållet i dokumentbasen är organiserat, eller hypertextlänkar, dvs korshänvisningar eller vägvisare mellan olika ställen i informationsmassan. Hypertextlänkar används ofta som vägvisare för att leta upp isolerade faktauppgifter o.likn., men jag tror att mer utvecklarmöda borde ägnas åt länkar som dessutom tjänar som orienteringshjälpmedel med vilkas hjälp man skaffar sig överblick hur informationen är ordnad.

Översiktskartorna bör vara ett hjälpmedel inte bara för att överblicka hur dokumentbasen är ordnad för tillfället, utan även för att stuva om i den, ungefär som när jag flyttar böckerna i mina bokhyllor. Skilda användare eller skilda grupper av användare bör kunna ordna en och samma dokumentbas på olika sätt i enlighet med sina särskilda behov. Vidare bör ett och samma dokument

vid behov kunna placeras på flera ställen i dokumentbasen (ungefär som när jag skulle behöva kopior av ett och samma papper i flera pärmar avsedda för skilda arbetsuppgifter).

Flertalet systemkonstruktörer tycks i dag favorisera en motsatt princip — man talar om ”information at your fingertips”, med Bill Gates formulering, eller om ”seamless information environment” — som innebär att användaren skall finna det han behöver utan att fundera över hur informationen är ordnad. Denna princip är rimligt då man är på jakt efter isolerade faktauppgifter eller enstaka dokument, men i mitt eget handbibliotek vill jag ha kontroll över hur samlingarna är ordnade.

6. Systemet bör stödja hanteringen av stora och komplexa dokument, samt ordnade samlingar av dokument.

Den populära kort-metaforen och annat som fragmenterar materialet och hotar att slå sönder dokumentens och dokumentsamlingarnas struktur bör bannlysas. Systemet måste tillåta arbete med stora dokument, motsvarande hundratals eller tusentals sidor, och dokumentbaser omfattande många tusen dokument. Väntetiderna får inte bli så långa att de hejdar tankens flykt.

7. En och samma dokumentsamling skall kunna användas på olika nivåer.

Såväl noviser som specialister, skolelever likaväl som universitetsstudenter, skollärare likaväl som forskare, skall med hjälp av olika slags filter och urvalshjälpmedel kunna använda en och samma dokumentbas och överföra vad de har bruk för till sitt handbibliotek. Hellre än att producera separata editioner för olika målgrupper bör leverantörerna om möjligt förse alla med samma rikhaltiga material. Skoleleven som arbetar med första kapitlet ur Röda rummet och doktoranden i litteraturvetenskap bör båda ha tillgång till hela den maskinläsbara Strindbergsutgåvan med kommentarer och konkordanser. Självfallet behöver noviser och specialister olika texter men den mest rimliga lösningen, som garanterar stora upplagor och bred spridning och som ger användaren största möjliga frihet, är att dessa texter samlas på t.ex. samma kompaktskiva. Det finns knappast några skäl (annat än möjligen kommersiella) att producera maskinläsbara motsvarigheter till de handböcker eller läromedelspaket som enbart är avsedda för studenter eller skolelever.

Jag har i andra sammanhang (se ”Kunskapsverkstaden. Om lokala dokumentbaser som arbetsverktyg för lärare”, *Datorn i utbildningen*, nr 2 1992, pp. 22-28) argumenterat för att lärarna och inte bara eleverna måste kunna använda den maskinläsbara litteratur som produceras för skolbruk. Det är min fasta övertygelse att alla, även eleverna, i det långa loppet vinner på att lärarna använder systemen och programmen i sitt eget arbete, i sin dagliga ”fortbildning”, i kursplaneringen, i lektionsplaneringen. Först då finns det chans att gedigna och rikhaltiga material matas in i systemen. Här som annars är jämförelser med biblioteksväsendet upplysande: tänk hur trist det vore om skolbiblioteken blott innehöll elementära läroböcker som lärarna inte har något att lära av. Dessutom får eleverna en bättre uppfattning om hur datorn kan användas som arbetsredskap när de möter lärare med egna erfarenheter därav.

Just nu planerar många förlag att återanvända sina maskinläsbara dokumentbaser, t.ex. underlaget för de tryckta encyklopedierna, genom därur plocka urval att saluföra till bestämda målgrupper: allt om hundar i en liten bok och allt om Afrika i en annan. Vid spridning på billigt magnetiskt eller optiskt medium finns knappast några skäl (återigen: fränsett de kommersiella) för ett sådant förfarande. Bättre vore att leverera hela encyklopedin, förutsatt att mottagaren har möjlighet att sortera fram artiklarna om hundar.

Principen bör således vara: tillhandahåll ett så rikhaltigt material som möjligt, utan tvång att allt skall användas. Med en översättning kunde gärna följa arbetet på originalspråk (i fråga om bokutgivningen kan man i Europa observera att parallellutgåvor med texten på både målspråket och originalspråket blir allt vanligare). En kortare introducerande text om ett ämnesområde kunde beledsagas av utförligare framställningar.

Här finns stora samordningsvinster om specialister och amatörer kunde använda samma dokumentsamlingar, på olika nivåer. Så sker sällan i dag, vilket illustreras av t.ex. det skåp som Apoteksbolaget tagit fram för apotekens kundutrymme. Skåpet döljer en skrivare kopplad till en persondator. Meningen är att kunden skall trycka på knappar märkta ”huvudvärk”, ”hemorrojder” etc,



varefter skrivaren skriver ut ett informationsblad med goda råd. Ett annat exempel är det ADB-system som många privatpraktiserande tandläkare använder och som bl.a. håller reda på patientjournalerna. I det ena fallet betjänas uteslutande lekmännen, i det andra uteslutande specialisterna. Det förefaller rimligare att ge fler grupper tillgång till olika utsnitt ur en och samma materialbank. På apoteket borde både personalen och kunderna kunna använda samma dokumentdatabas, och hos tandläkaren skulle även patienten kunna erhålla en utskrift av valda delar av sin journal, kompletterad med översättningar av svåra ord.

#### 8. Systemet bör vara modifierbart och utbyggbart.

Fler medier och fler program skall kunna tillföras. Systemet bör vara modulärt uppbyggt och innebära så få kommersiella bindningar som möjligt, dvs. man skall inte vara alltför beroende av enskilda hårdvaru- och programleverantörer. Internationella standarder innebär ett visst skydd mot sådana beroendeförhållanden.

#### 9. Enkel utrustning.

En användare bör utan tillgång till sofistikerade arbetsstationer eller teknisk personal kunna läsa, skriva och revidera dokument och helst också arbeta med hela dokumentbaser eller utsnitt ur sådana. En stor del av arbetet skall kunna ske med enkel utrustning som man vid behov kan ha i sommarstugan eller låna ut till en medarbetare på annan ort. Så är sällan fallet i dag, eftersom de mer avancerade dokumenthanteringssystemen är avsedda för stora företag eller förvaltningar. Över huvud taget begränsas humanister och samhällsvetare av att många datorlösningar på universitet och högskolor närmast befrämjar vad amerikanerna kallar "administrative computing". Motsatsen, "academic computing", förutsätter ett rörligare arbetssätt.

#### 10. Systemet måste kunna hantera strukturerade dokument.

Att dokumentet är strukturerat innebär att inte blott innehållet i snävare mening, texten och bilderna, är betydelsebärande. Även det sätt på vilket innehållet är ordnat, i kapitel och avsnitt, i rubriknivåer, med olika slag av texteleman för olika slag av innehåll (brödtext, fotnotstext, infogat citat, litteraturreferens, författarnamn, tryckort och tryckår etc) har betydelse för hur ett strukturerat dokument tolkas.

En konsekvens av denna kravlista är att det är ändamålsenligt att dokumenten är märkta enligt SGML-standard. Därom handlar nästa avsnitt.

### 3. Om beskrivande märkning och SGML

<p><b>SGML</b> (Standard Generalized Markup Language)</p> <p>ISO 8879, antagen 1986 Standardverk: Charles F. Goldfarb, <i>The SGML Handbook</i>. Oxford: Clarendon Press, 1990 ”Dotterstandarder” till SGML: DSSSL för typografering HyTime för hypertext samt för tidsbaserade dokument (multimedia) TEI (Text Encoding Initiative), riktlinjer för märkning av humanistisk och samhällsvetenskaplig litteratur</p>
--

SGML (Standard Generalized Markup Language) är ett språk för beskrivande märkning av maskinläsbara dokument. Jag skall inledningsvis med ett par exempel illustrera skillnaden mellan procedurmärkning och beskrivande märkning.

Procedurmärkning talar om hur den märkta informationen skall hanteras, exempelvis vad datorn och skrivaren eller fotosättningsmaskinen skall göra med en textfil. Var och en som infogat sättningsanvisningar i ett manuskript vet vad det innebär. En viss anvisning kan ange att vissa ord skall tryckas i 12 p kursiv stil. I detta fall innebär procedurmärkningen en instruktion till sätteriet.

En beskrivande märkning ger i stället besked om av vilka element texten består. Några ord med ett och samma typografiska utseende, t.ex. ord satta med 12 p kursiv, kan utgöra vitt skilda slag av element, exempelvis 1. en rubrik, 2. något författaren vill framhäva, 3. ett insprängt ord på främmande språk, 4. en boktitel. Om vi tillämpar procedurmärkning skulle i följande fiktiva text samtliga de fyra nämnda slagen av element anges på samma sätt (12 p Courier kursiv):

#### *II.3.1. Om Descartes metod*

Descartes har skapat ett mönster för vad filosofisk *metod* vill säga. Men i själva verket är Descartes berömda text ett *préface* till tre naturvetenskapliga och matematiska studier (tillgängliga i René Descartes: *Discours de la méthode*. Texte et commentaire par Étienne Gilson. Paris: Vrin, 6 uppl. 1987).

Med beskrivande märkning, som skiljer på dessa fyra slag av innehållsliga enheter, undviker man att en och samma märkning betyder olika saker. Om vi tillämpar SGML-märkning enligt version 1 av TEIs riktlinjer (mer därom strax) skulle märkningen se ut på följande sätt. Rubriknivå nr 3 markeras med ett inledande märkord `<h3>` och avslutas med märkordet `</h3>`. (Observera att ingen numrering av avsnittsrubrikerna behöver sättas ut; avsnittets plats i dokumentet gör att vi ändå vet att det rör sig om andra kapitlet, tredje avsnittet, första underavsnittet). Ett stycke inleds med `<p>` och avslutas med `</p>`. Framhävningen avgränsas av märkorden `<em>` och `</em>`. Bokstäverna *em* skall utläsas ”emfas”. (Observera att framhävning ofta representeras typografiskt av kursiv stil, men den kan även representeras på annat sätt, förr i världen med spärrad stil, i dag ibland med rak stil, exempelvis i ett förord där brödtexten är satt med kursiv stil.) Ett insprängt ord på främmande språk, här franska, kan markeras med de inledande märkorden `<gloss><foreign lang=Fr>` och avslutas med märkorden `</foreign></gloss>`. En hänvisning till en boktitel kan ske med korsreferens till ett unikt ställe; om den fullständiga referensen till Descartesutgåvan återfinns i en litteraturlista i slutet av det aktuella dokumentet och där försetts med märkordet `DES87` räcker det med att i slutet av texten ovan infoga `<xref RID=DES87>`. (`xref` uttydes korsreferens, `RID`

uttydes referensidentifikation). Så här ser ser texten ut då den på detta sätt försetts med beskrivande märkord:

<h3>Om Descartes metod</h3>

<p>Descartes har skapat ett mönster för vad filosofisk <em>metod</em> vill säga. Men i själva verket är Descartes berömda text ett <gloss><foreign lang=Fr>préface</foreign></gloss> till tre naturvetenskapliga och matematiska studier (tillgängliga i <xref RID=DES87>).

Detta var ett exempel på att man med hjälp av beskrivande märkning undviker att ett och samma typografiska utseende (kursiv stil) betecknar olika innehållsliga element. Omvänt undviker man också att ett och samma slag av innehållslig enhet märks på olika sätt. Tag citat som exempel. Citat kan utmärkas typografiskt på en rad sätt. Redan citattecknens utseende och placering varierar:

Tyska	»XXXXX« >XXXXX< „XXXXX“	Franska	« XXXXX »
		Svenska	”XXXXX” »XXXXX»
Engelska	“XXXXX” 'XXXXX'		

Även svenska typografiska konventioner kan variera. Ett citat som utgör ett eget stycke markeras ibland med indragen vänstermarginal, ibland med minskad grad och ibland med citattecken. Tag följande citat ur Viktor Rydbergs *Bibelns lära om Kristus*:

Det andra af de båda föregifna intygen lemnas af Cypriani skrift “Om kyrkans enhet“ och har följande lydelse:

“Herren säger: jag och fadren äro ett. Och återigen är det skrifvet om fadren och sonen och den helige andre: och tre äro ett.”

Onekligen har detta intyg ett viss sken för sig. Tvenne olika bibelställen äro här åsyftade. Det ena är, liksom hos Tertullianus, Joh. 10, 30; det andra är utan tvivel 1 Joh. 5, 8.

som med moderna typografiska konventioner skulle kunna se ut på flera sätt, exempelvis:

Det andra af de båda föregivna intygen lämnas af Cypriani skrift ”Om kyrkans enhet” och har följande lydelse:

”Herren säger: jag och fadren är ett. Och återigen är det skrivet om fadern och sonen och den helige andre: och tre är ett.”

Onekligen har detta intyg ett viss sken för sig. Tvenne olika bibelställen är här åsyftade. Det ena är, liksom hos Tertullianus, Joh. 10, 30; det andra är utan tvivel 1 Joh. 5, 8.

eller:

Det andra af de båda föregivna intygen lämnas af Cypriani skrift ”Om kyrkans enhet” och har följande lydelse:

Herren säger: jag och fadren är ett. Och återigen är det skrivet om fadern och sonen och den helige andre: och tre är ett.

Onekligen har detta intyg ett viss sken för sig. Tvenne olika bibelställen är här åsyftade. Det ena är, liksom hos Tertullianus, Joh. 10, 30; det andra är utan tvivel 1 Joh. 5, 8.

I franskt tryck är sägesatserna inte sällan infogade innanför citattecknen, som här på ett ställe i Claude Lévi Strauss *La pensée Sauvage*:

« Ce procédé, dit Boas, y est plus fréquent que dans tout autre langage connu de moi. »

och i engelskt tryck förekommer ofta att ett kommotecken omedelbart efter citatet placeras före citattecknet, som här i Paul Feyerabend's *Against method*:

‘If any metaphysics goes,’ writes Dr Hesse in her review of an earlier essay of mine, ‘then the question arises [...]’

Vid användning av beskrivande märkning bryr man sig i princip inte om sådana närmast estetiska spörsmål, man nöjer sig med att sätta märkordet <q> före citatet och </q> efter citatet. Sedan ankommer det på hur man styr det typografiska formateringsprogrammet och tillgängliga utskriftsmöjligheter hur den tryckta texten gestaltas.

De exempel på märkning jag här givit ansluter till den internationella standarden för rent beskrivande märkning, SGML (Standard Generalized Markup Language, ISO 8879), som antogs i oktober 1986. Standardverket är Charles F. Goldfarb, *The SGML Handbook*, Clarendon Press, Oxford 1990, som bl.a. innehåller den fullständiga ISO-texten. Arbetet går längre tillbaka i tiden. Ett förstadium till SGML var märkspråket GML, som Charles Goldfarb m.fl. från och med slutet av 1960-talet utvecklade vid IBM. Tanken var då att dokument (det rörde sig om juridiska dokument) skulle kunna märkas på ett enhetligt sätt, så att ett och samma dokument kunde matas in i skilda system för textbehandling, formatering och informationsåtervinning.

Dagens teknik öppnar vissa möjligheter att bearbeta ett och samma dokument eller en och samma dokumentssamling på mångahanda sätt och med många slag av program på många slag av maskiner. En sådan utveckling förutsätter ett generaliserat dokumentbeskrivningsspråk, ett slags esperanto om man så vill, som gör dokumenten oberoende av maskinvara och programvara och nationell teckenuppsättning. Ett och samma dokument kan med andra ord stoppas in i en SUN-maskin, en Macintosh eller en DOS/Windows-maskin och bearbetas med olika program. På detta område pågår ett livaktigt internationellt arbete med utvecklingen av SGML och med en hel svit av ”dotterstandarder” såsom DSSSL och HyTime.

DSSSL (Document Style Semantics and Specification Language, DIS 10179; akronymen uttalas vanligen ”dissel”) styr den typografiska formgivningen, dvs. sörjer för att layouten blir någorlunda densamma vid utskrift på olika medier och med olika utrustning. DSSSL förutsätter att dokumenten är SGML-märkta. DSSSL antogs som Draft International Standard i augusti 1991.

HyTime (ISO 10744), som antogs i april 1992 efter en rekordkort förberedelsestid eftersom behovet ansågs akut, är den första internationella standarden för överföring av hypertextdokument och multimediadokument (eller för att vara mer exakt: tidsberoende dokument, innehållande sådant som ljud och rörliga bilder). Även HyTime bygger på SGML-standarderna.

Detta standardiseringsarbete öppnar nya möjligheter att överskrida de gränser — nationsgränser, skillnader mellan olika slag av datorer och program, gränser mellan utbildningsväsendets nivåer och ämnen och mellan forskning och undervisning, avstånden mellan datorentusiasterna och de övriga — som i dag förhindrar samlade ansträngningar att skapa och sprida rikhaltiga dokumentbaser av god kvalitet.

Redan möjligheten att komma förbi de till synes triviala men enerverande problemen med olika teckenuppsättningar är ett stort framsteg. I SGML-standarderna (dvs ISO 8879) ingår en *reference concrete syntax*, som är utgångspunkten för översättningar mellan olika datormiljöer, nationalspråk, teckenuppsättningar m.m. I denna syntax ingår en teckenuppsättning (kallad *the base character set*) som överensstämmer med standarden ISO 646 (dvs en 7-bit-standard känd som IRV = International Reference Version). Detta är det som i egentlig mening är ASCII-teckenuppsättningen, dvs tecknen 0-127, som ligger till grund för en mängd andra internationella teckenstandarder. Med de 128 tecken som däri ingår och som nästan alla maskiner förstår kan snart sagt varje tecken i de europeiska skriftspråken representeras. Den mest flyttbara flyttbara representationen av å,ä,ö är enligt regelboken (Charles F. Goldfarb, *The SGML Handbook*, 1990, p. 506f):

Å	översätts med	&Aring;
å	översätts med	&aring;
Ä	översätts med	&Auml;
ä	översätts med	&auml;
Ö	översätts med	&Ouml;
ö	översätts med	&ouml;

SGML anger egentligen inte regler för hur man skall koda dokument. I stället är SGML en internationell överenskommelse om en uppsättning regler för hur den som märkt ett dokument skall berätta för andra hur denna märkning gått till, en berättelse som är omedelbart läsbar för mottagaren (antingen denne är en människa eller en maskin).

Att SGML är ett beskrivande språk innebär att SGML-märkningen inte säger något om den typografiska formgivningen. Vid utskrift eller fotosättning eller visning på skärmen vidarebefordras det SGML-märkta dokumentet till ett formateringsprogram som sørjer för att utseendet blir det önskade.

Det är inte meningen att den ordinäre användaren skall behöva se eller bekymra sig om SGML-märkningen. Han eller hon ser den version som passerat genom ett formateringsprogram (som ersatt &Aring; med bokstaven Å etc). Ett alternativ är att SGML-märkta dokument konverteras till användarens favorittextbehandlingsprogram och bearbetas där. Flera av de mest spridda textbehandlingsprogrammen (Microsoft Word, WordPerfect) kommer enligt tillverkarnas löften inom kort att förses med stöd för SGML.

Ett SGML-dokument består av tre delar. För det första en SGML-deklaration som talar om på vilket sätt man byggt ut eller modifierat ISO 8879. (Exempelvis anger TEIs riktlinjer att märkorden får vara upp till 128 tecken långa, vilket innebär en avvikelse från ISO 8879 som föreskriver högst 8 tecken). För det andra en DTD (Document Type Definition) som anger hur föreliggande typ av dokument märkes. En DTD kan avse affärsbrev, ytterligare en annan romaner, ytterligare en annan vetenskapliga monografier etc. För det tredje själva innehållet i dokumentet. Det är inte nödvändigt att SGML-deklaration och DTD medföljer varje enskilt dokument, men de måste (som regel som egna filer dit pekare i dokumentet hänvisar) vara tillgängliga någonstans i det system där man handskas med dokumentet.

Några ord bör sägas om det internationella projekt, TEI, Text Encoding Initiative, som tagit sig an uppgiften att utarbeta rutiner för kodning av humanisk och samhällsvetenskaplig litteratur, och som för det syftet har samlat humanister och samhällsvetare från många länder i skilda arbetsgrupper. TEIs syfte är att "formulate and disseminate guidelines for the encoding and interchange of machine-readable texts intended for literary, linguistic, historical, or other textual research."<sup>3</sup> Det första planeringsmötet skedde i november 1987. Projektet stöds av Association for Computers and the Humanities (ACH), Association for Computational Linguistics (ACL), Association for Literary and Linguistic Computing (ALLC), U.S. National Endowment for the Humanities (NEH), Andrew W. Mellon Foundation, samt Europakommissionen (XIIIe direktoratet). En första preliminär version (kallad "P1") av TEIs riktlinjer publicerades i juli 1990. En andra version (kallad "P2") av riktlinjerna är under arbete, men när detta skrives har de för märkningsarbetet viktigaste kapitlen redan gjorts tillgängliga via universitetsnäten. Det finns ingen anledning att vänta med att pröva riktlinjerna på svenska material.<sup>4</sup>

<sup>3</sup> Citerat efter Susan Hockeys och Donald A. Walkers presentation under konferensen SGML'92 i Danvers, Mass., USA, 26-29 oktober 1992.

<sup>4</sup> Kapitel 22, som blev tillgängligt redan i augusti 1992, innehöll de inte minst för bibliografiska syften väsentliga riktlinjerna för skapande av "TEI headers", dvs det avsnitt som skall inleda varje TEI-dokument och som innehåller uppgifter om titel, författare, källa, version, redaktionella principer och allt annat som kan behövas, exempelvis för automatisk katalogisering, för upprättande av index över dokumenten i ett bestånd etc. Kapitel 7 blev tillgängligt i oktober 1992 och innehöll "the bas tag set" för prosa. Kapitel 6, tillgängligt i december 1992, beskrev element som kan finnas i alla slags TEI-dokument.

På nästa sida återfinns som en avslutande illustration ett hemsnickrat exempel på SGML-märkning som följer en version av TEIs riktlinjer daterad oktober 1990<sup>5</sup>. Det är början av dokumentet vi ser. SGML-deklaration och DTD behöver som sagt inte alltid finnas med och de är här utelämnade (DTD'n representeras av en pekare `TEI . 1`, som är en av TEI föreslagen normal-DTD). Däremot är varje TEI-dokument, även detta, försett med inledande information som fungerar som ”elektronisk titelsida och elektroniskt förord”, bestående av tre delar [1] file description, [2] encoding declarations, [3] revision history — ett värdefullt underlag för bibliografiskt arbete eller för automatisk eller halvautomatisk katalogisering i ett bibliotekssystem. Sedan följer själva texten, märkt `<text>`, som inleds med en rubrik på den översta nivån, märkt `<h1>`, samt brödtextens stycken, märkta `<p>`. Avslutande märkord behövs inte eftersom element som rubriker och stycken förutsättes avslutade när ett nytt element introduceras. En rubrik är slut där vissa andra element — t.ex. en underrubrik eller första stycket i brödtexten — inleds, och ett stycke förutsättes avslutat när nytt stycke påbörjas. SGML-märkningen på följande sida är inte fullt genomförd. För läsbarhetens skull har i detta exempel (liksom i de tidigare exemplen ovan) de svenska tecknen behållits.

---

<sup>5</sup> C.M. Sperberg-McQueen and Lou Burnard (Eds): *Guidelines For the Encoding and Interchange Of Machine-Readable Texts*. Draft, Version 1.1, October 1990. Oxford, GB/Chicago, USA: The Association for Computers and the Humanities/The Association for Computational Linguistics/The Association for Literary and Linguistic Computing.

```

<TEI.1>
<TEI.header>
  <file.description>
    <title.statement>
      <title>Alva Myrdal & Gunnar Myrdal: Uppfostran i samhällets
mitt. En maskinläsbar transkription</title>
      <statement.of. responsibility>
        <role>etablerad av</role>
        <name>Donald Broady</name>
      </statement.of. responsibility>
    </title.statement>
    <extent.statement>116.300 Kb</extent.statement>
    <publication.statement>
      <creation.date>nov 1991</creation.date>
      <publication>Forum för pedagogisk historia</publication>
      <distribution>Ej för spridning</distribution>
      <release>ver. 1.0</release>
    </publication.statement>
    <source.description>
      <author>Alva Myrdal & Gunnar Myrdal</author>
      <title>Kontakt med Amerika</title>
      <imprint>
        <publ.city>Stockholm</publ.city>
        <publisher>Bonniers</publisher>
        <publ.date>1941</publ.date>
        <citn.detail>kap IV, pp. 88-132</citn.detail>
        <comment>Inflytelserik text inför reformeringen av den
svenska skolan</comment>
      </imprint>
    </source.description>
  </file.description>
  <encoding.declarations>Standardförfarande</encoding.declarations>
  <revision.history>Denna version från nov 1991 är den som är rejält
korrekturläst</revision.history>
</TEI.header>
<text>
  <h1>Uppfostran i samhällets mitt
<p>
Direkt kräver den amerikanska trosbekännelsen frihet, jämlikhet och broderskap,
indirekt kräver den uppfostran. Det ena ger formen och målet, det andra ger
innehållet och medlen. Det ena är det evigt konservativa i de amerikanska
idealen, det andra det evigt radikala.</p>
<p>
Statsgrunden, samhällsidealen har stått som oföränderliga genom en längre period
än i något annat land. Men medlen till att fördjupa och befästa idealen är
ständigt växlande. Det är i sökandet efter dem amerikanens experimentalistiska
livsinställning röjer sig. Denna paradox mellan vad som är gammalt och nytt,
stabilt och föränderligt, odiskutabelt och diskutabelt, vad som är konservativt
och radikalt i den världsdelen, som kallar sig den Nya Världen men som samtidigt är
den äldsta demokratin, har sällan blottlagts. Det ivrigt nya gäller aldrig
bottenvärderna. Men heller ingenting annat än bottenvärderna får i Amerika fortleva
utan ständiga frågetecken.

```

## 4. Några fördelar med SGML-märkta dokument

- Regelstyrd beskrivande märkning ger entydig tolkning.
- Få antaganden om hur dokumenten kommer att användas.
- Dokumenten oberoende av plattformar, tillämpningsprogram och teckenuppsättningar.
- Dokumenten kan i princip redigeras med den enklaste ordbehandlare och skrivs ut på vilken skrivare eller skärm som helst.
- Märkningen är någorlunda begriplig för människor.
- Dokumenten är flata textfiler som kan användas som databaser.
- Dokumenten kan återanvändas.
- Det finns hopp om en ny Aldus Manutius.

### 1. Regelstyrd beskrivande märkning ger entydig tolkning.

Program — eller människor — kan på ett entydigt sätt tolka ett dokument som är korrekt SGML-märkt. Man undviker med andra ord att en och samma märkning har olika betydelse, eller omvänt att en och samma betydelse representerar av olika märkningar (jfr exemplen kursivstil resp. citat ovan).

### 2. Få antaganden om hur dokumenten kommer att användas.

Några exempel på tänkbara användningar är följande. Dokumenten kanske skall överföras till textbehandlingsprogram och där redigeras; formges typografiskt för utskrift på skrivare eller för vidare befordran till fotosätter; införlivas med en fulltextdatabas eller en databas över abstracts; användas vid strängsökning (för att letar upp enstaka ord eller fraser) och/eller struktursökning (varvid sökspråket utnyttjar relationerna mellan de element varav dokumenten eller dokumentbasen är uppbyggda); användas för lingvistisk analys (lexikalisk, syntaktisk eller semantisk); utgöra underlag för indexering, för upprättande av konkordanser etc; tappas på viss specifik information med hjälp av sök- och informationsåtervinningsprogram; utgöra underlag för automatisk eller halvautomatisk katalogisering eller framställning av bibliografier; förses med hypertextlänkar. Listan kunde göras hur lång som helst. Nu senast har jag stött på flera stora projekt i olika länder som utvecklar metodik för att ur SGML-märkningen generera utskrift på punktskriftsplatta för synskadade eller för utmatning till talsyntes.

### 3. Dokumenten är flyttbara.

Kompleta SGML-märkta dokument är i princip oberoende av plattformar och tillämpningsprogram och kan (förutsatt att systemen innehåller programvara som ger SGML-stöd) flyttas mellan dessa. Spridningen är ej heller begränsade av nationalspråk, av teckenuppsättningar (så länge de är västerländska) eller av vad som låter sig skrivas in från tangentbordet.

### 4. Måttliga hårdvarukrav.

Dokumentet kan i princip redigeras med den enklaste ordbehandlare och skrivs ut på vilken skrivare eller skärm som helst. (Från början var tanken att SGML-märkning skulle kunna åstadkommas med skrivmaskin.)

### 5. SGML-märkningen är begriplig för människor.

Märkorden är mnemotekniska och återfinns omedelbart intill det textparti de avser. Detta är en viktig skillnad i jämförelse med exempelvis flertalet ordinära textbehandlingsfiler, vilka är bemängda med koder och pekare som maskinen förstår men som är obegripliga för nästan alla människor.

Det är självfallet inte meningen att man skall behöva sitta och genomföra SGML-märkningen manuellt. Ett naturligt arbetssätt är att använda sitt textbehandlingsprogram, varefter en konverteringsrutin ombesörjer omvandlingen till SGML-märkning och ett annat program validerar



märkningen, dvs undersöker om den uppfyller SGML-kraven. Men den märkning som finns är sådan att människor med en smula träning lär sig begripa det viktigaste.

6. SGML-märkta dokument är flata textfiler som kan användas som databaser.

Man kan exempelvis i ett dokument eller en dokumentssamling söka upp alla element av ett visst slag (exempelvis alla litteraturreferenser, eller alla huvudrubriker, alla abstracts, alla citat på grekiska) som innehåller vissa ord eller uppfyller andra kriterier. Man kan extrahera fram alla element som innehåller författarnamn och sedan göra vad man vill med dem. Och så vidare.

En för bibliotekarier och andra välkommen möjlighet är att TEIs riktlinjer föreskriver att all information som behövs för katalogisering och bibliografering enligt internationell standard (såsom ISBD, International Standard Book Description, etc) skall finnas omedelbart tillgänglig som väl åtskilda och väl definierade element i början av SGML-dokumentet (jfr den TEI-märkta Myrdaltexten ovan). Sådana element är författarnamn, redaktör, titel, utgivningsort, publiceringsdatum, förlag, etc, och motsvarande uppgifter avseende maskinläsbara dokument. Därmed beledsagas varje text av information som utgör underlag för en i bästa fall alltigenom automatisk katalogisering, och som även är till stor hjälp i de ordinära bibliografiska sysslor som tillhör forskares eller lärares arbete.

7. Dokumenten kan återanvändas.

Att samma dokument kan användas för de mest skilda syften innebär att vi kan räkna med att dokumenten överlever de operativsystem och tillämpningsprogram i vilka de skapats.

8. Efterlyses: en Aldus Manutius i vår tid.

Aldus Manutius var den boktryckare i Venedig i slutet av 1400-talet och början av 1500-talet som utgav en mängd antika författare i den tidens "flyttbara format": behändiga kvartovolymer som kunde stoppas i fickan, standardiserade och lättlästa typsnitt, texten kvalitetskontrollerad av lärda korrekturläsare, en med tidens mått extraordinärt stor upplaga (tusen exemplar). Under sekler betraktades dessa vackra volymer i hela den lärda världen som de pålitligaste utgåvorna av de latinska auktorena. Jag har en vision av en framtida maskinläsbar litteratur representerande samma lyckliga kombination av flyttbarhet, tillgänglighet, bred spridning och hög kvalitet. För detta krävs det internationellt accepterade dokumentbeskrivningsspråk som SGML är på väg att bli.

[[[