Digital Critical Editions.
The case of the Swedish National Edition of August Strindberg's collected works

Donald Broady

Royal Institute of Technology
Dept. of Numerical Analysis and Computing Science
S-100 44 Stockholm, Sweden

E-Mail: broady@nada.kth.se
URL: http://www.nada.kth.se/~broady

ABSTRACT

[Submitted in Spring 1996, before the conference.]

The international standards family SGML (Standard Generalized Markup Language, ISO 8879), HyTime (Hypermedia/Time-based Structuring Language, ISO 10744), and DSSSL (Document Style Semantics and Specification Language, ISO 10179) opens new opportunities in the humanities. SGML offers the opportunity to create common principles for structuring, encoding and exchanging material. HyTime offers the opportunity to create common principles for addressing and hypertext linking, as well as principles for structuring multimedia information (more precisely: time-dependent information as audio or video). DSSSL offers the opportunity to create common principles for the processing of SGML-encoded material (i.a., principles for portable style sheets).

The elaboration of these principles is, though, not given by SGML, HyTime or DSSSL, which are languages – or rather meta-languages. They merely are internationally agreed ways to define principles for describing and organising material, and for exchanging it across the boundaries between computer platforms, applications, and national languages. In every single domain the concerned specialists have to work out the principles in accordance with their own needs. In the humanities the most significant attempt in this direction is the guidelines developed by the international project TEI (Text Encoding Initiative).

The paper will present experiences from the application of the TEI guidelines to the Swedish National Edition of August Strindberg's collected works. In this context the encoding has been used as the source for the creation of both a digital version and some volumes in the ordinary printed book edition (via LaTeX to Postscript delivered to the printing house). It is a relatively large edition (70 volumes + commentary volumes), half of which already published by use of traditional typesetting. There were of course strong demands that the printed volumes to come should mimic the design of the previous ones in every detail. Thus, the encoding scheme had to be underpinned by a set of rules reflecting on the one hand the particular typographic design principles (created by Sweden's most celebrated book designer Karl-Erik Forsberg), and on the other hand the philological and editorial principles governing this particular critical edition.

The optimal combination of scholarly editorial aspects and visual presentation aspects was rather difficult to achieve. You run the risk of having to choose between two bad alternatives: either a rather unorthodox (partly presentation-oriented and therefore less useful in other respects) SGML-encoding, or a pure descriptive encoding which means time-consuming and arduous work within the layout programme.

This problem and possible solutions will be discussed in the paper, as well as other lessons learned from the Strindberg project.

The paper will also touch upon the program for the consortium NOLA (Networking of Literary Archives), the aim of which is to make material in European archives – mainly unpublished primary sources from major novelists, philosophers, musicians, painters etcetera – available in portable digital versions. The TEI recommendations for encoding and meta-description of manuscript materials will be the particular focus of NOLA, which will assess their suitability and recommend their extension and modification as necessary. A major emphasis will be placed on development of procedures for the integration and use of TEI-aware tools for the creation, management, documentation, analysis and dissemination of archival resources.

1. Introduction

Nationalupplagan av August Strindbergs samlade verk
(the National Edition of August Strindberg's collected works)
70 volumes + 20 supplement volumes

Most recent volumes. prepared at the Royal Institute of Technology:
Vol 57. *Svarta Fanor*, Norstedts, Stockholm 1995, 428 p. (prose)
Vol. 15. *Dikter på vers och prosa*, Norstedts, Stockholm 1995, 582 p. (verse)

The National Edition of August Strindberg's collected works is not – yet – a digital edition. It is a scholarly book edition, with extensive commentaries, explanations, and variants etc. There will be seventy volumes in all, whereof 37 volumes are published. The pace is modest, one, two or three volumes each year. According to the plans also 20 supplementary volumes are to be expected, mainly presenting variants and none of them published so far. This is Sweden's largest publication project ever in the humanities.

The two most recent volumes, both published in late 1995, represent, if you wish, a step into the digital age.

The technical production has been undertaken by us at the Royal Institute of Technology, all the way from the raw text files delivered by the editors to the final formatted print-ready PostScript files (illustrations and cover excluded). The most important collaborator in this work was Hasse Haitto.

Our aim was to create a digital resource adapted to the state of art of portable markup in the humanities, that is TEI compliant. This resource should at the same time as much as possible support the layout work needed to produce a printed book edition.

The aspect upon which I will focus today is, thus, the use of digital versions as the basis for hardcopy output. This may not be the most advanced or intriguing use of digital techniques. It is, though, important for several reasons, not the least the actual demand. It is also, so far, the only task that we have implemented in a real production project.

2. The demand

Besides the Strindberg project we are engaged in several other projects.

- the publication of the Swedish land registrers (Sw. *jordeböcker*) of the medieval ages and the 16th century, in co-operation with the Central Board of National Antiquities;
- the publication of Swedish court protocols from the 17th century, in co-operation with virtually all concerned Swedish archives, museums and libraries;
- publication of educational material, in co-operation with the Swedish National Agency for Education.

The projects mentioned have one thing in common. The scholarly community and the funding agencies first and foremost need to get the printed book editions out. That is their prime concern. If the digital processing helps to make the editorial work and the book publication more effective and reliable, and cheaper, in that case they are seriously interested. In other case they are not. For the time being it seems impossible to obtain funding in Sweden for stand-alone digital critical editions. At the

Royal Institute of Technology we have to accept that our own main ambition, to pave the way for future digital editions that offer more than the printed ones, has to remain a more or less hidden agenda when we co-operate with the archives, libraries, museums and funding agencies.

3. Implementation, tools

The coding of the two Strindberg volumes at the Royal Institute of Technology was undertaken in close collaboration, often daily contacts, with the general editor (Docent Lars Dahlbäck at Stockholm University), the editorial committee, and the volume editors, as well as with book designers and typography experts. At the Royal Institute of Technology we used the following methods and tools.

- SGML-encoding: emacs (Unix); MS Word (PC, Mac); Author/Editor (SGML-editor for PC)
- Validation: OmniMark (Mac)
- Scripts for input to formatter: OmniMark (Mac)
- Formatting: LaTeX (Unix)

From the editors we received scanned and OCR-interpreted material or word processing files or raw text files, which we converted and encoded in SGML. We used the general purpose and SGML-aware translation software OmniMark from Exoterica in order to validate the encoding and to write the scripts that produced the input to the formatter (layout-program) LaTeX, which is a powerful public domain software primarily used for the layout of high quality scholarly publications in mathematics and hard sciences. The output from LaTeX was print-ready PostScript files that we delivered to the printing house.

Our long-term aim is to refine and reuse the SGML-files in a future digital edition. Therefore we have as well already accomplished a similar crude encoding of all the previous published 35 volumes, as a prelude to a future digital edition. The publishing house has granted us the rights to distribute it for free on the internet for purely scholarly purposes.

There were absolute demands that the result should mimic the design of the previous volumes in every microscopic detail. Thus, the encoding scheme and the scripts controlling the formatter had to be underpinned by a set of rules reflecting on the one hand the particular typographic design principles, created by Sweden's most celebrated book designer Karl-Erik Forsberg, and on the other hand the philological and editorial principles governing this particular critical edition.

The challenge was to put as much as possible of this knowledge into the computer system, and not the least to procure a bridge between the two knowledge domains, which meant offering the editors a guarantee that if let's say a certain type of dialogue in the manuscript for a forthcoming drama volume was correctly encoded according to the editorial principles, then the printed output would as far as possible meet with the typographic standards set by the book designers. Which was, of course, only possible to a certain degree – our aim was to investigate to what degree.

Form the editors' point of view it gives them better control the more information there is already in the markup, of which they knew the typographic consequences, instead of the tiresome trial and error routines of sending blueprints sent back and forth between editors, book designers and the printing house, with corrections causing new errors to pop up all the time.

4. TEI

TEI (Text Encoding Initiative)

URL http://www-tei.uic.edu/orgs/tei

Burnard, Lou/Sperberg-McQueen, C.M (Eds.): *Guidelines for Electronic Text Encoding and Interchange* (TEI P3). Volume I-II. Text Encoding Initiative, Chicago University/Oxford University, 1994.

Our encoding followed the guidelines proposed by Text Encoding Initiative. As most of you know, these guidelines are the result of seven years of work by many committees engaging scholars in the humanities and computer scientists in many countries. Thanks to the Text Encoding Initiative, the humanities community now is in the possession of a common basis for platform-independent encoding and exchange of a broad range of transcriptions of ancient manuscripts or contemporary speech, prose, poems, and drama, historical sources, linguistic language corpora etc.

The TEI guidelines derive their strength from being an implementation of SGML.

5. SGML, introduction

SGML (Standard Generalized Markup Language)

SGML is a language used to describe how documents are built. The structural and content elements, their attributes and their mutual relations are described, but (typically) not the typographic look.

ISO 8879, accepted 1986

Most authoritative book, including the standard itself:
Charles F. Goldfarb, *The SGML Handbook*, Clarendon Press, Oxford 1990, ISBN 0-19-853737-9.

Best introduction:
Eric van Herwijnen, *Practical SGML*, Kluwer, Boston/Dordrecht/London, 2 ed. 1994,
ISBN 0-7923-9434-8

At the risk of boring the eminent SGML- and TEI-specialist in this room, among others both the European editor, Lou Burnard, and the American editor, Michael Sperberg-McQueen, of the TEI guidelines, I will spend a few minutes to give those of you who might not be familiar with SGML a hint on what this standard is about. After all, this is not a dedicated technical conference.

SGML is not a markup scheme that prescribes what markup (or "tags") to put in your information. It is a general language offering an international agreement on how to tell someone else (a human being or a computer system) how you have chosen to markup your information.

Take for example the following piece of text.

1.3.1. On Descartes' method

Descartes created a model for *the* philosophic method. Descartes' celebrated text on the method was, however, published as a *préface* to three studies on natural sciences and mathematics (available in René Descartes: *Discours de la méthode & Essais*. Éd. Ch. Adam & P. Tannery. OEuvres VI. Vrin, Paris, nouvelle édition 1982).

Here you find four occurrences of italics. They share the same typographical appearance but represent four very different types of elements: heading level 3, emphasis, a gloss from a foreign language, reference to a source.

If you wish to achieve nothing more than a certain appearance on paper or on screen you could use a format instruction such as Courier 12 p italics, as you normally do within your word processor. However, the main idea behind SGML and the TEI guidelines is to describe not what things look like but what they are.

Here is the same text provided with SGML encoding (using an early version of the tag sets proposed by TEI):

\<h3\>On Descartes' method**\</h3\>**
\<p\>Descartes created a model for **\<em\>**the**\</em\>** philosophic method.
Descartes' celebrated text on the method was, however, published as a **\<gloss\>**
\<foreign lang=Fr\>préface**\</foreign\>\</gloss**\> to three studies on
natural sciences and mathematics (available in **\<xref RID=DESCARTES1982\>**).**\</p\>**

The heading level 3 is enclosed by the start tag **\<h3\>** and the end tag **\</h3\>**. There is no need for a number if the system is able to count the chapters, sections and sub-sections and produce the correct number in the output. A paragraph is opened by **\<p\>** and closed by **\</p\>**. The emphasis is enclosed by the tags **\<em\>** and **\</em\>**, and the French word by the **\<gloss\>** and **\</gloss\>**, within which we find another pair of tags informing us (or the system, for example for a spell checker) that this is a foreign word, and that the language is French. Finally, the reference to another book is represented by a cross reference element with an ID attribute which makes it possible to retrieve the correct reference somewhere else in the document, perhaps in a reference list at the end.

In order to get the document portable you should also translate the diacritics into international standardised entities, as **\é** for é.

At the same time you have represented the hierarchical structure: the start of a section, including a third level header, a paragraph (later on followed by other paragraphs), and within each paragraph emphasis, references etc.

In this case descriptive markup helps to avoid that the same typographic look (italics) designates four different types of elements. The other way around descriptive markup might help to avoid that one and the same element type is represented by several different looks. Quotations is one example. In different languages (and within each language) there are various typographical conventions when it comes to present inline quotations on paper or on screen. For example:

German	»XXXXX«	French	« XXXXX »
	›XXXXX‹		
	„XXXXX"	Swedish	"XXXXX"
			»XXXXX»
English	"XXXXX"		
	'XXXXX'		

When applying descriptive markup you might not need to bother about these differences when editing a text. It might be sufficient to encapsulate the quotation within a pair of tags, **`<InlineQuote>`** and **`</InlineQuote>`**. The output system can use the appropriate typographic conventions in the output on paper or screen.

In the same manner a tag like **`<blockquote>`** could be used to encapsulate parts of the text that are to be treated differently in accordance with different national and typographic conventions. In French, different from contemporary Swedish, you most often wish intended block quotes.

A few words on standards related to SGML:

Additional standards to SGML

HyTime (Hypermedia/Time-based Structuring Language, ISO/IEC 10744:1992),
a standard for hypertext (i.e., linking mechanisms) and multimedia (i.e., time-dependent information such as audio and video)

DSSSL (Document Style Semantics and Specification Language, ISO/IEC 10179:1996),
for the processing of SGML-encoded material, i.a., for exchange of typographical information

---

The international standards family SGML (Standard Generalized Markup Language, ISO 8879), HyTime (Hypermedia/Time-based Structuring Language, ISO 10744), and DSSSL (Document Style Semantics and Specification Language, ISO 10179) opens new opportunities in the humanities. SGML offers the opportunity to create common principles for structuring, encoding and exchanging material. HyTime offers the opportunity to create common principles for addressing and hypertext linking, as well as principles for structuring multimedia information (more precisely: time-dependent information as audio or video). DSSSL offers the opportunity to create common principles for the processing of SGML-encoded material (i.a., principles for portable style sheets).

You could also, which is what TEI does, enrich the encoding scheme with more content oriented tags signifying name (as Descartes, in our example), date and time, a gap in a transcription due to a damaged source, or information on the conjunctures made by a certain editor interpreting the source.

Now you have a platform-independent SGML source from which you might generate a nicely formatted output on screen or paper or a speech synthesis or Braille, or as input to a database, or something else.

End of tutorial, back to Strindberg.

---

6. Encoding problems

The orthodox view is, thus, that SGML has nothing to do with the processing of the information. You should encode what things *are*, not how they are intended to *look* on for example a printed page or a screen.

Our dilemma was that we wanted both: both a TEI-compliant SGML-encoding, as a digital resource for multi-purpose usage, *and* at the same time there was high demands for an accurate result in print.

Furthermore, we wished to combine those two ambitions, which meant that as much as possible of the information of relevance to the typography should be present already in the SGML source, before it was filtered to the formatter, a filtering process that we wanted to automate as much as possible. Thereby we hoped to maximise the usability and the *re*usability of the SGML-file and to minimise the manual work required within the formatter. The technique meant that we tried to identify every type of constituent in the source that was to receive a unique typographic look and to furnish every one of those types of constituents (for which I coined the term typographemes) with a unique tagging.

6This was realised by scripts that downtranslated the SGML tagging into formatting instructions. The constituents could of course be nested into each other, and since our scripting tool OmniMark understand how to handle nested elements, we could profit from the SGML encoding. Take for example in the middle of the reproduced page below the ending of a letter from the novel character Reverend Nathanael Skåre who uses the salutation phrase "Din Återlöste" (Yours, by God's mercy saved).

SJUNDE KAPITLET

Jesu Efterföljelse.

Följande morgon väcktes han av städerskan som framlämnade
ett brev vilket befanns vara av följande lydelse.

Timot. Kap. X, v. 27, 28, 29. Första Korint.
Kap. VI, v. 3, 4, 5.

Dyre Br!
Vårs H:s J. Kr. Nåd och Frid, Fadrens kärlek och D. H. A:s
delaktighet etc. Amen!

Jag såg av Gråkappan i går afton att Du ämnar utgiva
Försoningsfacklan. Sök mig i min verksamhet i morgon bittida
före 9.

Din Återlöste
Nathanael Skåre.

Nu förstod han Lundells gåtor, till en del! Han kände visserli-
gen icke den store gudsmannen Skåre personligen och visste
intet om Försoningsfacklan, men han var nyfiken och beslöt att
hörsamma den närgångna kallelsen.

Kl. 9 stod han på Regeringsgatan framför det väldiga fyra-
våningshuset, vars fasad var klädd med skyltar ifrån källarvå-
ningen ända upp till taklisten. *Kristliga Boktryckeri Aktiebolaget
Friden* 2 tr. upp. *Redaktionen av Guds barns arvedel* ½ tr. upp.
Expeditionen av Yttersta Domen, 1 tr. upp. *Expeditionen av Fridsba-
sunen* 2 tr. upp. *Redaktionen av Barntidningen: Föd mina Lamm* 1 tr.
upp. *Direktionen för Kristliga Bönhusaktiebolaget Nådastolen verk-
ställer utbetalningar och beviljar lån mot första inteckning i fastighet* 3
tr. upp. *Kom till Jesus* 3 tr. upp. Obs. ☞ Ordentliga utsäljare
som kunna ställa borgen erhålla sysselsättning därstädes. *Mis-*

Those two right-aligned lines could be tagged as follows.

```
<close>
<salute>din återlöste</salute>
<signed>Nathanael Skåre</signed>
</close>
```

Since SGML allows nested tagging one and the same encoding could be interpreted differently depending of the surrounding context. In this case the element `close` together with its sub-elements `salute` and `signed` are contained within a larger element representing, as a suggestion, a division, in this case a chapter, within a suitable even more encompassing element representing the novel. When encoding instead let's say Strindberg correspondence corpus, published chronologically in a dozen volumes, the element `close` will probably be encapsulated within an element representing a letter, which in turn in the hierarchical structure could be subordinated to an element representing year. SGML-aware software did allow us to choose to render of the element `signed` in italics (which mirrors the typography of the correspondence volumes) if and only if it is included into an element representing a letter to or from Strindberg. More important than the typographical concerns are of course the flexible opportunities for precise searches, filtering and so on. Using the context-dependence of the element `signed` allow us to sort out senders of letters to Strindberg, but not literary characters such as the Reverend Nathanael Skåre.

By the way you might observe one of the inline symbols or glyphs that are rather frequent in some of Strindberg's texts. Here there is a hand at the bottom of the page.

There are, of course, alternative approaches. You could use the SGML file as input into a traditional typesetting process, just as you normally use an ASCII file or a word processing file, in which case SGML is not very useful since the main bulk of the layout work would have to be done in the formatting system anyway.

Or you could use an unorthodox more presentation-oriented approach to SGML and include lots of processing instructions intended for the typesetting system, which would have made the files less useful for other purposes and even for exchange with other typesetting systems. Certainly also SGML-aware systems might handle some processing instructions, but it is of course preferable to have the crucial information contained within the proper SGML markup. It is for example reasonable to instruct the system to apply a certain typographical look to each element `signed` in a certain context. This context is in our case defined by the inclusion in an element `close`, in turn encapsulated in some higher order element representing a chapter in a novel or some other division. The alternative, to send the processing instruction "italics" to the typesetting machine, means that you get trapped by the specific typesetting system from which the material can be extracted only at the cost of losing crucial information. Problems usually arise when a revised edition is to be printed and the previous typesetting system is found to be outdated. In the entire Strindberg project we only inserted a handful of processing instructions in the SGML files, in the rare cases when attempts to find other solutions had failed.

Or you could instead of TEI use some other SGML application, more directly suited for the management of printed output. In fact, initially we tried to use the international SGML-based standard in this domain, ISO 12083, which is designed exactly for this purpose, to support the publication industry's production of books and journals. But ISO 12083 did not fill our needs. It is obviously more useful when you start from scratch with an electronic manuscript and wish to automate the formatting as much as possible without too specific demands on the final output and when a more uncomplicated conventional layout will do. The tag set is consciously kept very limited, there is no such thing as let's say the `<closer>` element that I mentioned, or the `<milestone>` element which we used a lot as a placeholder where the formatter should insert for example page division symbols like the hairline bar. Using ISO 12083 would have forced us to define a multitude of such elements which are already defined in the TEI guidelines.

7. Two types of demands

What made our endeavour so demanding was that we had to represent the given physical appearance, in two respects:

First the look of the original source, Strindberg's own hand-written manuscripts or letters or some first printed edition. This is in fact a task that the TEI guidelines to a large extent is devoted to.

Second we had to slavishly mimic and duplicate the typographic conventions used in the 35 previous volumes published by means of traditional typesetting. The layout was originally created by Sweden's most celebrated book designer Karl-Erik Forsberg, and the demands on the accuracy of the layout work was extremely high. This turned out to be the most difficult, time-consuming and arduous task – and not at all what the TEI guidelines are meant for.

One reason for stretching the TEI guidelines in this direction is that the hardcopy version will probably for many years – for a century, in the opinion of the editors – to come be the standard Strindberg edition. A reader using a digital edition should therefore have available all relevant information in the hardcopy edition, including layout properties as pagination, page breaks, line breaks etc. Thus a future digital edition would anyway have to include information on not only the physical appearance of the primary sources but also on the printed National Edition.

If you wish you should be able to use a digital version in order to make extracts that in print or on screen in essential aspects look like the book edition, and automatically generate page number and maybe line number of the printed edition.

8. Already in the SGML files

Many things of significance to the typography could be prepared already in the SGML file to be interpreted by the scripts that filtered this file to the formatter.

We used for example emacs' capacity of regular expressions to automate routine tasks like inserting non-breaking space between a digit and the name of a month (the digit one, fixed space, July) or between an initial and a last name, or the requested amount of space before or after different types of dashes etcetera, or fixed space between an n-dash and the next character in a dialogue.

9. Typefaces and unusual symbols

One of the first problems to solve was to find the appropriate typefaces and to edit the kerning tables, the position of the dots and circles in the Swedish diacritic type characters, and to choose the right font size and weight for different types of text. Here our interactive mode of work allowed us to produce pages on the laser printer and let the book designers scrutinise alternative typefaces and their size – 9 point one, 9 point 15, 9 point 20 etcetera – and weight, text area, spacing between letters, words and lines etc.

We had some problems in finding the equivalent of a rather specific Greek (Strindberg often wrote living persons' names in Greek), modifying it, and then insert Greek text in the file. Since we did not succeed to create a portable solution we simply inserted an entity in the SGML file which later on was resolved directly to a string in the system specific Greek font to which we switched for the duration of the quote.

In order to produce the peculiar small symbols and glyphs that Strindberg was found of, we created PostScript files and used them to render those signs inline. This was done for example to get rune glyphs, and the hand that you saw a minute ago:

> tr. upp. *Kom till Jesus* 3 tr. upp. Obs. ☛ Ordentliga utsäljare

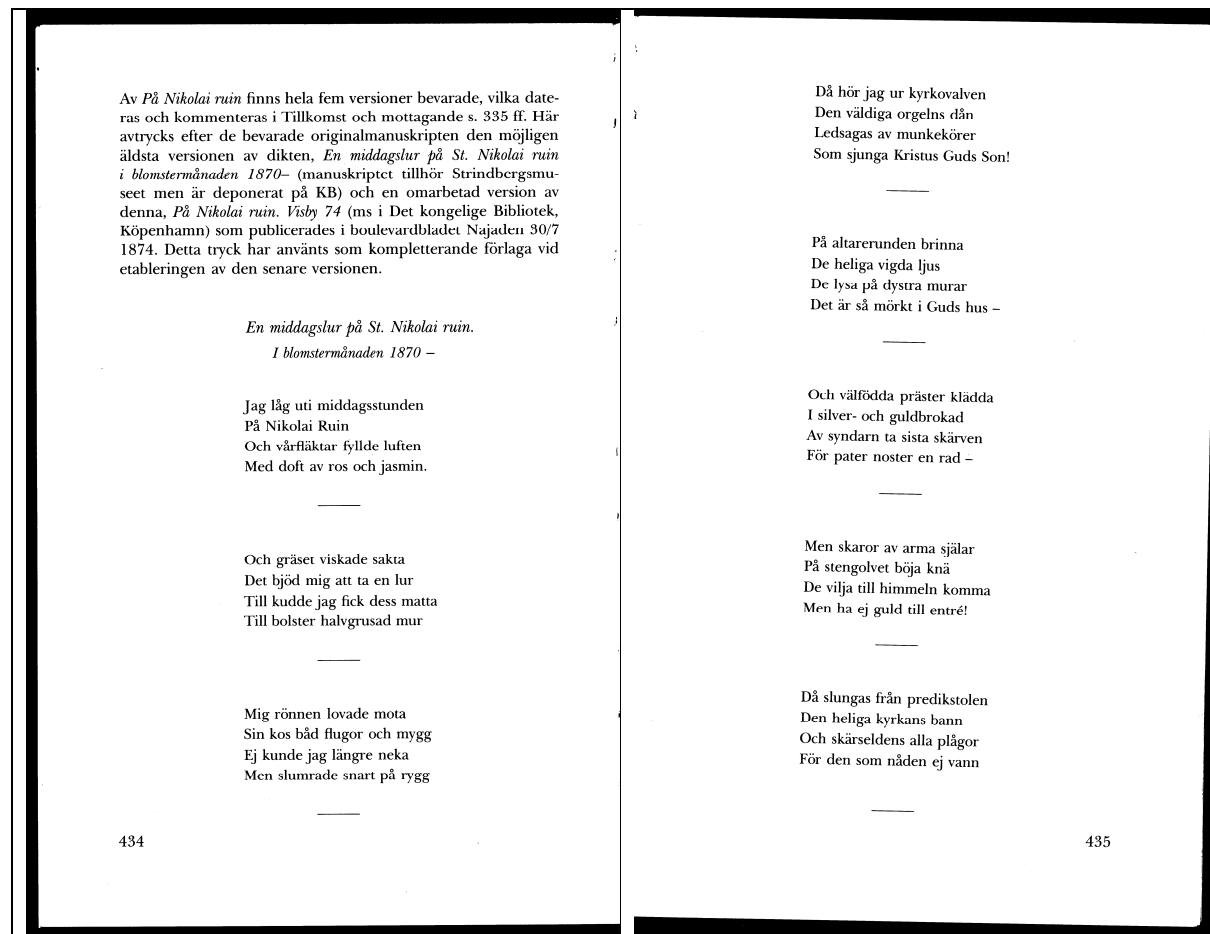(August Strindberg, *Röda rummet*, Samlade Verk 6, Norstedts, Stockholm 1981, p. 81.)

And this symbol of a seagull:

> rade den. Ett exempel! Där en vanlig människa blott sett en fågel, ser du en hök. Hur vet du att det är en hök, då denna fågel och fåglar i allmänhet på målningar ha detta utseende ∩.

(August Strindberg, *Tjänstekvinnans son I–II*, Samlade Verk 20, Norstedts, Stockholm 1989, p. 310.)

10. Left to be accomplished in the formatter

Many things had, though, to be accomplished in the formatter (i.e. LaTeX), as illustrated by the two pages below from the poetry volume. This volume was extremely cumbersome, since almost every one of the 582 pages demanded a unique solution. You need to know the page breaks and line breaks in order to judge how to distribute the text over the page. There were lots of peculiar hanging indentations, centring concerns etcetera.

Av *På Nikolai ruin* finns hela fem versioner bevarade, vilka dateras och kommenteras i Tillkomst och mottagande s. 335 ff. Här avtrycks efter de bevarade originalmanuskripten den möjligen äldsta versionen av dikten, *En middagslur på St. Nikolai ruin i blomstermånaden 1870–* (manuskriptet tillhör Strindbergsmuseet men är deponerat på KB) och en omarbetad version av denna, *På Nikolai ruin. Visby 74* (ms i Det kongelige Bibliotek, Köpenhamn) som publicerades i boulevardbladet Najaden 30/7 1874. Detta tryck har använts som kompletterande förlaga vid etableringen av den senare versionen.

En middagslur på St. Nikolai ruin.
I blomstermånaden 1870 –

Jag låg uti middagsstunden
På Nikolai Ruin
Och vårfläktar fyllde luften
Med doft av ros och jasmin.

———

Och gräset viskade sakta
Det bjöd mig att ta en lur
Till kudde jag fick dess matta
Till bolster halvgrusad mur

———

Mig rönnen lovade mota
Sin kos båd flugor och mygg
Ej kunde jag längre neka
Men slumrade snart på rygg

———

434

Då hör jag ur kyrkovalven
Den väldiga orgelns dån
Ledsagas av munkekörer
Som sjunga Kristus Guds Son!

———

På altarerunden brinna
De heliga vigda ljus
De lysa på dystra murar
Det är så mörkt i Guds hus –

———

Och välfödda präster klädda
I silver- och guldbrokad
Av syndarn ta sista skärven
För pater noster en rad –

———

Men skaror av arma själar
På stengolvet böja knä
De vilja till himmelen komma
Men ha ej guld till entré!

———

Då slungas från predikstolen
Den heliga kyrkans bann
Och skärseldens alla plågor
För den som nåden ej vann

———

435

(August Strindberg, *Dikter på vers och prosa*, Samlade Verk 15, Norstedts, Stockholm 1995, pp. 434–435)

A reoccurring problem was the need for optical centring not only of some poems but also of some other text sections. Optical centring means mathematically that some lines centre around the weighted axes for all relevant lines, with exceptions for extra-long lines or other irregularities. In the end the optical centring had to be adjusted manually by the help of a designer's trained eye.

I the following paragraphs, on top of page 244 in *Svarta fanor*, the book designers had opted for a hanging indent for the three lines in the midst of a paragraph, and centred around the axes of the two words "Hvita Dukar," without the three letters "o.s.v." (signifying etcetera) interfering with the centring. It is not obvious how to describe such a thing in the tagging scheme, and even less so in the formatter. The problem is, again, that you cannot judge on this until you are able to experiment with the outcome of the page and line breaks.

terna) som finnas der, och rädda dem från förgängelsen och domen?
De kunde ju kallas: *Hvita Dukar*
Ur
Svarta Fanor.
1, 2, 3, o.s.v.
Det om kemin kan uteslutas, liksom Crex, men om Hemmet och Vexterna, Uret o.s.v. om spökerier etc. Äfven strötankar utryckta och samlade? Hvad synes Eder.

Vad Strindberg syftar på är ganska tydligt: »kemin» avser revisorns föreläsningar (klosterskrifterna nr 7 och 8), »Crex» syftar på revisorns dialog om »Kornknarren och Göken» (nr 6) och »Hemmet och Vexterna, Uret o.s.v.» på greve Max' monolog om »Materien som levande väsen» (nr 4); »spökerier» kan anspela både på revisorns utredning av naturliga förklaringar till spökerier (nr 9) och greve Max' dialog »Om järtecken» (nr 2).

(August Strindberg, *Svarta fanor*, Samlade Verk 57, Norstedts, Stockholm 1995, p. 244)

We also had to calculate the space for the illustrations, which was complicated since the designers in a case like this preferred to have the accompanying text on the bottom of a verso and the illustration itself on the following recto, as below.

i Stormen, Paolo Malatesta och ett till, 1/4 ark (=15) hvilka alla falla under Högsommar och sålunda hafva en 6 veckor på sig». »Förordet skall», tillägger han, »ofördröjligen afgå!» (Av dessa texter utfördes veterligen endast Förordet och »Afsked», dvs. *Första Natten* i *Sömngångarnätter*.) Den 1/9 aviseras flera »Svenska Afsked», försåvitt »Barndopet» inte skulle bli klar (Brev 3, s. 290).

Avresan till Grez skedde den 12/9. Strindberg insjuknade på resan (bl.a. var han sjösjuk) och fick lägga sig till sängs i Stralsund; uppehåll gjordes därefter i Lübeck, Hamburg, Bremen, Köln och Paris (barnsköterskan Eva Carlssons dagbok, KB; Brev 3, ss. 298, 302 f.; jfr M. Brundin, Eva Carlssons dagbok och Strindbergs resor i Europa 1883–1889, seminarieuppsats, Stockholms univ., 1977, s. 9 ff.). Strindberg korrekturläste inte själv diktsamlingen utan överlät detta på Karl Otto Bonnier och Staaff (Brev 3, ss. 286, 321; jfr brev från Strindberg till Staaff 27/11, återgivet i K. Dahlbäck, Ändå tycks allt vara osagt, 1994, s. 143, samt brev från Staaff till Strindberg 27/9 och 24/10, KB). Från tåget mellan Bremen och Osnabrück 18/9 uppmanar han Staaff att ändra »verkliga idiotier» om sådana påträffas, samt att inte låta någonting »uteslutas» (kanske fruktade Strindberg för censuringrepp). Staaff svarar 27/9 oroligt: »Idiotier skall jag ändra, säger du! Ja, men det är ett krångligt göra! Det är väl bäst att ändra minsta möjliga, egentligen blott omeningar eller obegripligheter. Jag skall handla efter bästa förstånd, men du måste komma ihog att det är rätt kinkigt och skälla ner mig måttligt, om du tycker att någonting blifvit åt helfvete genom min försorg» (KB).

Efter ankomsten till Frankrike meddelar Strindberg brevledes förlaget ytterligare ändringar. Värt att notera är att många av dessa sena ändringar gäller *Landsflykt:* 26/9 begär han från Grez avskrifter av »Kristiania» *(Landsflykt II)* och »Paris» *(VII),* 8/10

På nästa sida återges framsidan av ett handskrivet blad med rättelser till Landsflykt, *vilket medföljde ett brev till Albert Bonnier omkring 8/10 1883 (Bonniers förlagsarkiv). Som framgår av Strindbergs anmärkning till den näst sista rättelsen i* Landsflykt II, *»Ni hade skrifvit hosarna», hade han fått avskrifter från förlaget av de ännu inte (av sättaren) uppsatta dikterna.*

354
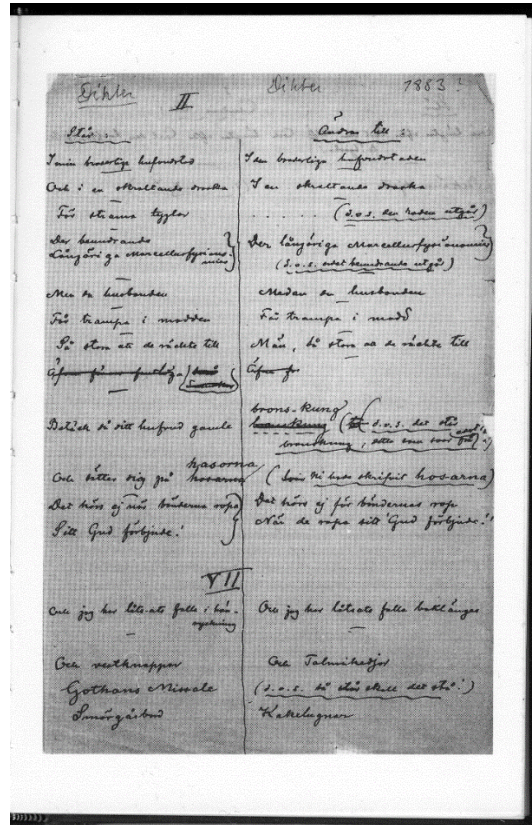
(August Strindberg, *Dikter på vers och prosa*, Samlade Verk 15, Norstedts, Stockholm 1995, pp. 354–355)

The following is an example of indentions within indentions.

Bliv som Du varit alltid glad och ung
Och kvinnorna helt ärbart dyrka
Men när din själ blir kvald och tung
Köp »Väktarn» och gå in i Beskows kyrka.

Ja, lycka till Du tjocka Bouff,
Förlåt mitt djävla sätt att gratulera.
Och måtte lyckan ge dig snart en knuff
Att Du blir miljonär och Riddare m.m.

Nu – Calle går det rent för hin
Med rim reson och sångmöns fläkt
Och därför stoppar jag helt visligt in
I säck min pipa – »Och nu är lampan släckt».

P.S.

Schål på Dig gamle Ravaillac!!!!
Och må din skugga ej bli större –
När Du är ute och »slår klack»
Så akta att Du ej blir »Djurre»

Men kors för tusan jag glömt bort
Att på din namnsdag gratulera
Jag vill då göra saken kort
»Jag har den äran, hum! m.m.»

Ja – lycka till Ditt gamla as
[---]

446

(August Strindberg, *Dikter på vers och prosa*, Samlade Verk 15, Norstedts, Stockholm 1995, p. 446)

The structure might cause problems. In the Strindberg edition either a hairline bar or one, two or three stars are used as division marker, that is for example subdivisions within a larger division within a chapter, which we were happy to represent by the `milestone` element. Until we met with both a hairline bar and a constellation of stars.

In some cases we used the formatters reference mechanism to get the correct page references.

In the TEI guidelines there is primarily one attribute, REND, meaning rendition, that you are supposed to use for typographic information of the kind that we were interested in. We often longed for more alternatives. Even if the attribute string might in principle be indefinitely long, you need to have a processing program that understands it. This was cumbersome, for example when we wanted to give multiple rendition attributes – such as subscript and italics, or hanging indent within hanging indent – to one and the same element,. A cleaner solution, that we tried to use as much as possible and which is more in accordance with the TEI approach, is to use elements instead of attributes.

The typographic end result met with high quality standards, thanks to that LaTeX is a most powerful and flexible formatter. In several respects – good widow control and line break control, low number of hyphenations, long distances between lines ending with hyphens – the result was superior to that of previous volumes produced by traditional more or less manual typesetting. LaTeX uses so called dynamic programming. Every character is placed in a box and then composed to words, lines, pages, sections. The interword spacing is flexible; it can be stretched or contracted to a user-defined degree. The same applies for spacing generally, both vertical and horizontal, of which we made frequent use, for example to make the space before and after sections elastic in order to let LaTeX calculate the optimal page and line breaks; here we relied on the judgement of the book designer to find the most suitable parameters. LaTeX has powerful widow control and avoids ugly hyphenation by calculating on larger portions of the text than most layout programs. The principle is this: LaTeX puts according to the kerning table letters together to form words, noting where there are permissible hyphenation locations. It proceeds to assemble the words for a paragraph at a time, while keeping the line widths optimally spaced; if this is not possible, it will attempt hyphenation. This process continues until all options have been investigated and the "best" found. During this process LaTeX continuously throws away the impossible attempts (for example hyphenation in the beginning or midst of a line) and only calculates on the possible ones, which makes it fast,

Further, LaTeX is very helpful in pointing out all occurrences where it is not possible to meet all the combined user defined demands, thereby leaving the decision to the user. You might also force LaTeX to override the best solution and go for a second best.

In order to achieve this, LaTeX lets you define or defines per default punishment scores for ugliness ("badness" is the ordinary technical term). Hyphenation at the end of two lines after one another gives a high ugliness score, widows of different length receive different scores, higher the shorter the line etcetera. There are several hundreds of these parameters, used in cubic equations, and it was an interesting exercise to interview the book designers in order to have them define such parameters; their knowledge is normally tacit, not explicit and quantifiable as is needed in order to build it into a computer program.

If, given the values of those parameters, LaTeX is unable to solve a problem, for example a hyphenation or a widow, it lets the line go out in the right margin, and produces a list of the location of such too long lines and their exceeding width (which might be only some thousandths of a millimetre). We could then make printouts and discuss possible solutions with the book designers. Often we tried to find some character like a dash that might, or according to some schools in typography should, go out in the right margin.

Demands on layout are often contradictory. Sometimes there exist evidently superior solution. For example, the number of lines on a page with running text is often fixed, but to solve a difficult widow problem you might go back to add or remove one line from the first page of the chapter, as this in our case has about half a page of vertical space. You could also make a paragraph more or less dense in

order to force LaTeX to procure a line break other than allowed by the current values of the parameters.

The back side of this is that LaTeX runs in batch mode and is not very user friendly. There exist filters from SGML to more user friendly wysiwyg formatters, like SoftQuad WEnabler to Quark Xpress, but as far as I know you are unable to reach the same rule-based accuracy as with a TeX or LaTeX solution.

12. Note on NOLA

Finally, I wish to mention the existence of the consortium NOLA (Networking of Literary Archives), which is an initiative with the aim to make material in European archives – especially unpublished primary sources from major novelists, philosophers, musicians, painters etcetera – available in portable digital versions. The TEI recommendations for encoding and meta-description of manuscript materials will be the particular focus of NOLA, which will assess their suitability and recommend their extension and modification as necessary. A major emphasis will be placed on development of procedures for the integration and use of TEI-aware tools for the creation, management, documentation, analysis and dissemination of archival resources. The activities are co-ordinated from University of Bergen and its Wittgenstein archives, and also Lou Burnard and his collaborators here at Oxford are engaged.[1]

11. Conclusions

A small well organised team of experts is sufficient to accomplish a production like this. You need competence in five areas: scholarly editing, book design, SGML, the management of the formatter, and the editing of typefaces.

The learning curve is steep in all five areas. In order to spread the competence a thorough documentation is necessary. Else you are in trouble if one of the experts disappear.

The procedure is rather cost effective given that the publications are sufficiently homogeneous within each publication and between publications. The initial cost is, though, substantial. Our two volumes took approximately 12 person months, all included. There is of course more to gain the more material of the same type you are to publish. We could for example reuse the design and scripts used for the explanations in the commentary part, which meant that the one-hundred-page long section with explanations in the second volume was very straight foreword to produce.

Much might be automated, but with heterogeneous material and high demand on the typographic output there is very much that has to be accomplished within the formatting system. The main workload in our case was the typography. By comparison the SGML knowhow needed seemed almost trivial. This was especially the case with the poetry volume, where almost every page demanded unique design solutions. This poetry volume took about ten times as much work as the prose volume.

You do not need expensive software. Commercial off the shelf software or public domain programs is sufficient for most of the work.

---

[1] The NOLA consortium is co-ordinated by Norwegian Computing Centre for the Humanities (NCCH), University of Bergen, Norway (Lars G. Johnsen, Director) in collaboration with the Wittgenstein Archives at the University of Bergen, Norway (Claus Huitfeldt). Other participants are Oxford University Computing Services, Great Britain (Lou Burnard); University of Innsbruck Library, Austria (Heinz Hauffe); Forschungsinstitut Brenner-Archiv (FIBA), University of Innsbruck, Austria (Allan Janik); the Dorset County Museum, Dorchester, Great Britain (Richard M. de Peyer); Royal Institute of Technology, Stockholm, Sweden (Department of Numerical Analysis and Computing Science, Donald Broady); The Royal Library, Stockholm, Sweden (Anders Burius).

Compared to traditional typesetting, when manuscripts, dummies and proofs are sent back and forth, this more exploratory style of work gave the editorial board a better control. We were in contact with the editors nearly every day. They knew that general revisions in the SGML file or the scripts had to be done only once to take effect everywhere and ever after. They were given laser printouts for control of singular pages or passages. In order to minimise the risk for unintended faults during the editing we used file comparison programs to automatically trace changes between the final SGML file (the markup excluded) and earlier versions. All this would have been impossible if we prematurely had poured it all into a proprietary layout program. Instead we inserted the corrections in the SGML file for as long as possible, constantly downtranslating it into the formatter to produce PostScript output on screen or on paper to show to the editors and book designers.

The editors grasped of course the opportunity to perform a great number of changes during the process, many hundreds, possibly several thousands, many more than in any of the previous published volumes, which meant delays. We are not sure that we ourselves wish to do this again. Right now for the volumes to come a commercial printing house is trying to accomplish the same with the aid of our experiences.

Some of Strindberg's manuscripts constitute a serious challenge for the encoder, as might be seen from my last image, below, from one of the most difficult manuscripts, the famous *Ockulta Dagboken* (the Occult Diary), never before presented in a scholarly edition but currently in preparation, with some of the writing up-side down, all kind of curious symbols, strike-through, underlining, parenthesises, and drawings, and newspaper cuttings, calendar sheets, train tickets or even plants glued to the pages.

Thank you for our attention.
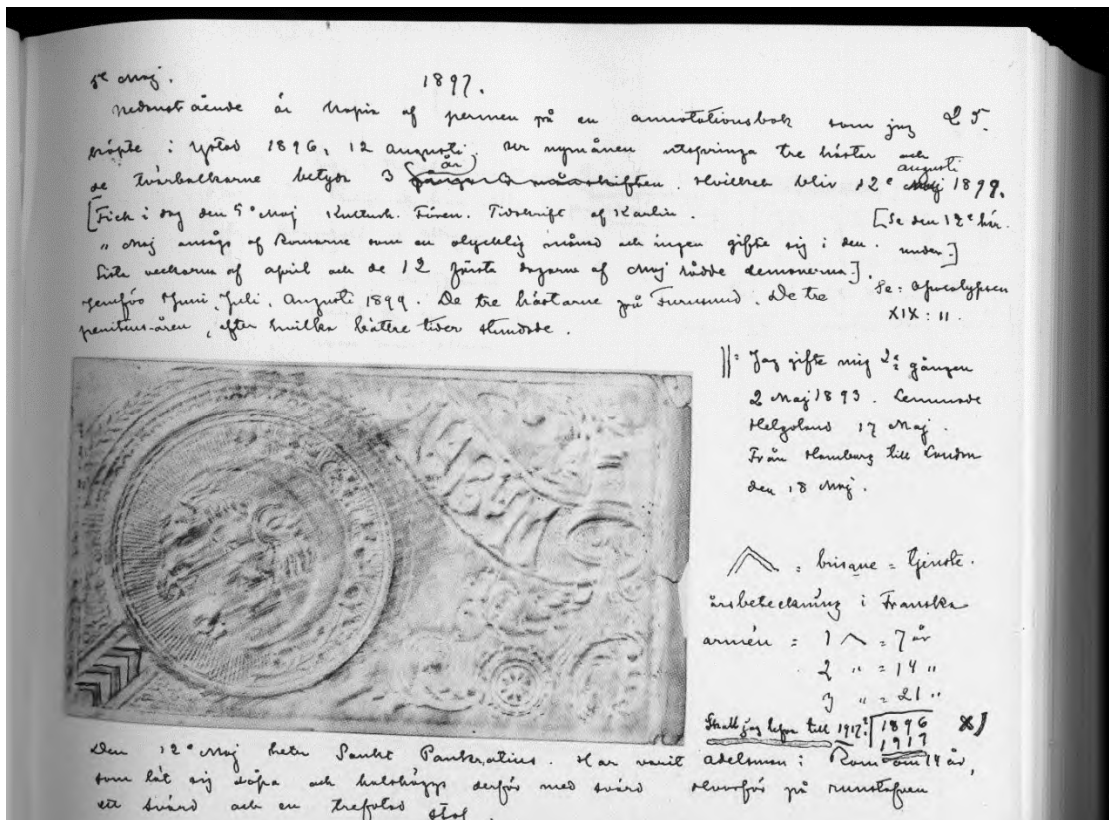
(August Strindberg. *Ockulta Dagboken*, written 1867–1908, facsimile edition, Gidlunds, 1977, p. 25)