

CHARTERS IN THE BERGEN UNIVERSITY LIBRARY – XML ENCODING AND DIGITISATION IN COOPERATION WITH THE FACULTY

Rune KYRKJEBØ

BERGEN UNIVERSITY LIBRARY / AKSIS / NORWEGIAN RESEARCH COUNCIL

Rune.kyrkjebo@ub.uib.no

Abstract. The paper presents an ongoing digitisation project on old charters in the Bergen University Library Special Collections. The idea for this project originated in cooperation between the faculty and the library and AKSIS, an information technology development department associated with our university. Together we handed in an application for funding to the Norwegian National Research Council, and we were so lucky as to receive the funding. In this paper I wish to underline the value of cooperation between university library and faculty. I also aim to outline the material and technical matters of our project on old charters, with emphasis on the XML text encoding standards used.

”THE USER COMMUNITY AS RESPONSIBILITY AND RESOURCE” AS PRINCIPAL GUIDELINE FOR E-PROJECTS

The quotation in this heading is taken from an interesting paper by David Seaman (1997), where he shares his experiences on how to build a sustainable digital library, as he calls it, in cooperation with the different groups of users. Seaman concludes that for a special collection department in a university library, ”the [user] community you create quickly becomes a vital part of the resources on which you draw”, this happens when users start creating electronic texts, re-purposing existing resources, and ”providing critical observation on the various aspects of the digital library in which they invest themselves” (Seaman 1997).

In my opinion Seaman here formulates what is a very sound general aim or ideal situation for special collections moving their books and manuscripts into the electronic age. Especially in the case of university libraries, the user communities have both the interest and the capability of contributing to contents, use, and design of our electronic resources. In the case of old manuscript material it is a general experience that faculty staff and higher degree students supply the library with new information on items in our collections. The student of one particular manuscript quickly becomes an expert on exactly that item. This student may produce information which might be added to the library’s catalogue, for example details on dating, origin, places and names, or historical significance of the document. Sometimes a student produces something larger, like a full transcription of a document. That transcription is today typically an electronic representation which we should be able also to take into the library’s electronic collections.

When we believe, then, that there will be something to gain from engaging staff and students in the creation of electronic special collection resources, this affects the way in which those resources should be designed by our libraries. What we do is simply invite our expert users and ask them: How should this text best be displayed? What phenomena do you want to search for in this text? What parts of the text should be marked with this or that type of text encoding? How should it be linked up with other information sources?

Manfred Thaller also gives many interesting considerations to our subject in a paper titled ”From the digitized to the digital library” (2001). His paper contains principal discussions of the concept of digital library, with a focus on manuscript material. I shall quote here only one of his several theses

which I find particularly enlightening, it regards the question of what manuscripts to digitise: "The primary topic of digitisation projects should not be the absolute top of the range treasures of a collection, but those materials that we always have wanted to promote if they were just marginally more important." (Thaller 2001). His point is that "our treasures" typically have been edited once or more during the 19th or 20th century, maybe produced in facsimile also. So what the scholarly community really has the most to gain from, is digitisation of classes of material that are not "top of the range" but nevertheless scientifically interesting.

I believe Thaller is completely right in this judgment. We face the question, then, of what are the scientifically interesting not-top-of-the-range materials. Cooperation with faculty staff is here in my opinion the quickest and safest way to decide what parts of our collections falls into this category. The question will be what part of our un-edited, unpublished manuscripts are the most promising material from a scholarly point of view, where are the historians interested in investing their research resources. In this respect university libraries are privileged when having to make priorities, because we typically have many experts close at hand, in our faculty and at the institutes that the library serves.

BERGEN UNIVERSITY LIBRARY CHARTER COLLECTION – A PROJECT ("DIPLOMPROSJEKTET")

A handful of documents in the charter collection at Bergen University Library have the status of "treasures", or they have least been old enough to be included in national editions from the 19th or 20th centuries. This goes especially for the medieval documents in the charter collection. The collection as a whole however comprises several hundreds of documents from the whole period between the years 1293 and ca. 1750. When, last year, the National Research Council issued a special invitation to apply for digitisation of humanities research material, our library went into discussion with staff members at the University of Bergen Department of History. Together we overviewed the collection and brought together the library's catalogue material and professional interests, with the knowledge and considerations and academic wishes of the history department. It quickly became clear that the part of the collection dated between 1570 and 1660, some 400 documents, for research reasons was most worthy of digitisation and editing.

We also defined as a goal for this project to produce an electronic infrastructure suitable for handling images, full texts, and catalogue information for the charter collection. Encoding of the textual information is the point at which it is most important, in my view, to choose open and well defined standards to ensure the maximum interchangeability and usability of the textual data, in order to provide users with the best possible data, and to provide ourselves with the best possibility for maintenance and use of our data in the library. A part of our vision for the future is that our special collections department should be able to receive electronic textual material from staff and students, material that they have enriched with new information or another kind of higher quality. The most obvious example of this is electronic transcriptions with some degree of text encoding added, for one purpose of study or another.

EXAMPLE ITEM FROM THE CHARTERS PROJECT

The website for our charter collection will incorporate lists where you can choose to view the shorter catalogue entry (fig. 1), or a longer entry with more details. From the catalogue entry one will be able to open the full text transcription, if this particular document has been electronically transcribed.

An electronic handling of images and data like that in figure 1, can of course be achieved in many ways using different techniques. There are several ways to produce a similar result for the end user, for example by using a database. The solution we are developing is not primarily based on use of a database, but on so called *transformation* of the XML catalogue files and transcription files.

Vis de viktigste data Vis alle data

Tittel: 1585, 19. oktober Sogn og Fjordane, Stryn, Ytre Eide
Inst. ansv. for tilgang: UBB
Fysisk plassering: UBBHF Spesialsamlingene, Diplomsamlingen

Sammendrag: Kristofer Pepler, fogd i Nordfjord, Morten Pederssøn, sogneprest i Innvik prestegjeld, Jacob Hanssøn, hjelpeprest samme sted, og 10 edsvorne menn i Innvik er tilstede ved et arvebytte etter Anders Torgilssøn på Ytre Eide - mellom enken Anna Oluffsdatter på den ene side og sønnen Anders Anderssøn samt hans halvsøsken på den annen side. De får hva han har gitt dem i gårdene Robjørgene i Innvik, Meland, Ytre Eide og Kirkekeide. Enken har bl. a. fått 1 1/2 laup i Brynnestad i Olden.

[Transkripsjon](#)

UBB-DiplSaml-1585-okt-19-01r.jpg
 UBB-DiplSaml-1585-okt-19-01v.jpg

Fig. 1.– Short catalogue entry which shows dating, place name, physical location, abstract of content for the charter, and link for the transcription. The image on the right hand side leads to a full size image of the charter. The text concerns inheritance of farmland in Nordfjord, Western Norway.

We encode the catalogue into text files where each part of the contents are marked with start tags and end tags, like in this example:

```

<date>1585, 19. oktober</date>
<geogname normal = "Ytre Eide">Sogn og Fjordane, Stryn, Ytre Eide</geogname>
<repository>UBB </repository> .....
  
```

On the internet the XML file is run through a so called XSLT stylesheet, which produces the layout, the typography with labels, boxes and borders, fonts and so on. That stylesheet is a powerful tool, not just for viewing the document as HTML. It can in fact also produce a PDF, or it can rearrange the contents and markup of one XML file into another XML file.

STANDARDS FOR TEXTUAL ENCODING – EAD, TEI /MENOTA

We use XML documents to contain all textual data. Our choice of XML scheme for the catalogue information is the "Encoded Archival Description" (EAD). This was chosen because it focuses on large collections rather than on single documents. One aim of our project is namely to produce an electronic infrastructure that is as general as possible. In the long run we wish to possess the tools for keeping and updating the whole manuscript catalogue for our library, which in turn means that we will have to handle several levels of series and subseries, and larger numbers of early modern and modern manuscripts. The EAD is suited for exactly this, it is an XML document type definition which "sees" a large, complex archive, rather than a single document, so to speak. The main reason for choosing EAD is therefore mostly a strategic one, since the charter collection in fact is a "flat" collection without subseries.

The well known XML based standard TEI is in fact best suited for encoding philological information about single manuscript items. The TEI has an element called <sourceDesc>, which has a number of sub-elements allowing for encoding of most kinds of philological information. Thus one could say that while EAD "sees" an archive, TEI "sees" the single document. The TEI standard

is currently being used by the Menota-project, also based in Bergen. Menota has modified TEI to make it better suited for encoding medieval text in a very diplomatic, manuscript close manner. Associated with the Menota project is another project called Mufi, which aims at developing unicode fonts for electronic versions of medieval texts. Both Menota and Mufi are projects which our charter project has some cooperation with, since they are also in the field of old manuscript digitisation.

CONCLUSION

XML and related standards were chosen for digitising The Bergen University Library charter collection, in order to create electronic texts that are as openly accessible as possible, as easy to use and as well documented as possible. This is a practical advantage for the user community, and this is preparing the ground for our users to enrich our electronic collection by adding information themselves to electronic texts like transcriptions. In working out the details of choice of material, and of encoding and display techniques, faculty staff and higher degree students have played an active part by specifying their needs, and by helping us see what the possibilities of such a project are.

REFERENCES

1. **SEAMAN, David.** 1997. The User Community as Responsibility and Resource. Building a Sustainable Digital Library. *D-Lib Magazine*.
<http://www.dlib.org/dlib/july97/07seaman.html>
The etext library at University of Virginia, USA, <http://etext.lib.virginia.edu/>
2. **THALLER, Manfred.** 2001. From the Digitized to the Digital Library. *D-Lib Magazine*.
<http://www.dlib.org/dlib/february01/thaller/02thaller.html>
Project “Codices Electronici Ecclesiae Coloniensis (CEEC)”: <http://www.ceec.uni-koeln.de/>

Website of the Bergen University Library charter project (in Norwegian):

<http://gandalf.aksis.uib.no/diplom/>

Website of the Bergen University Library Medieval Parchment Fragments project (in English):

<http://gandalf.aksis.uib.no/mpf/>

Website of the AKSIS centre (available in English):

<http://www.aksis.uib.no/>

Website of the Encoded Archival Description (EAD), Library of Congress, USA:

<http://www.loc.gov/ead/>

Website of the Text Encoding Initiative (TEI):

<http://www.tei-c.org/>

Website of the Medieval Nordic Text Archive (Menota) (in Scandinavian):

<http://www.menota.org/>

Website of the Medieval Unicode Font Initiative (Mufi) (in English):

<http://gandalf.aksis.uib.no/mufi/>