

# **BETEXT: Cross Institutional Cooperation in the Digitalization of Cultural Heritage Documents in XML**

## **Authors**

Michael Davis, Blekinge Institute of Technology, Karlskrona, Sweden - michael.davis@bth.se  
Peter Linde, Blekinge Institute of Technology, Karlskrona, Sweden - peter.linde@bth.se  
Hoa Tran, AerotechTelub Information & Media AB, Växjö, Sweden - hoa.tran@aerotechtelub.se

## **Abstract**

The following describes a digitalization and on-line publication project, led by members of Blekinge Institute of Technology, located in the south of Sweden. The project aimed to create the foundation for a regional network of small cultural institutions that would develop, support and utilize on-line full text archives in the service of tourism, marketing and education. Its pilot project, The Chapman Archive, involved the digitalization, encoding with the TEI DTD in XML and web-publication of two rare 18<sup>th</sup> century texts of both regional and international importance.

## **1. Background**

### **1.1 Introduction**

The project discussed here involves three distinct but interrelated projects: COVAX, BETEXT and the Chapman Archive. In 1999, inspired by the tremendous success of the Electronic Text Center at the University of Virginia (U.Va), Michael Davis and Peter Linde began considering the possibility of a similar center for their university, Blekinge Institute of Technology (BIT). As David Seaman writes, the key to the center's long-lived achievement at U.Va was that it did not live an isolated life in the basement of the library and on the web, but actively enlisted both the university's teaching staff and archive community to develop and promote its work.<sup>1</sup> Thus, the center not only had access to a wide range of locally available manuscripts to digitize, along with the expertise of the archivists, but to the energies and interests of teachers and students as well. In other words, through the center, the archival and educational communities have come together to provide on-line access to material that is actually used in the classroom. The question that Davis and Linde posed for themselves was how to knit that kind of network at BIT. The University of Virginia, the oldest publicly funded university in the United States, with 19,000 students, 12,000 employees, library archives of 10.3 million pages and an operating budget of \$816.3 million, has tremendous resources to call upon.<sup>2</sup> We, at a technical college barely 10 years old, have 3,000 students, 425 employees, an operating budget of \$32.5 million and no archives of our own to speak of.<sup>3</sup> If we were going to build a similar center, we would have to reach out to the archive and academic communities around us. Our solution is called the Blekinge Electronic Text Center, or BETEXT.

### **1.2 COVAX**

In 2000, we became a partner in the COVAX, one of the European Union Fifth Framework IST (Information Society Technology) projects. COVAX, which stands for Contemporary Culture Virtual Archives in XML, sought to develop single-web portal access to archive, library, museum and full-text collection databases from across Europe, using XML and XQL as its common "language." The goal was to increase accessibility to electronic resources, regardless of their location. A major element of the project was the conversion of existing bibliographical and full-text records to XML. We were the main providers of full-text documents: developing the Chapman Archive as well as converting, encoding and publishing chapters from museum yearbooks, an exhibition catalogue and 20 odd works of non-copyrighted Swedish fiction. Our conversion and encoding process was primarily an automated one of searching and replacing in MS Word and a simple Visual Basic program. All documents were encoded according to the TEI-lite DTD, and it was originally decided that COVAX would not only search and display their header-section metadata, but, if the user so desired, the body-sections as well. We thus hoped that COVAX would provide us with an interface for future BETEXT archives.

Unfortunately, the complications of creating a speedy retrieval system for both bibliographic records and full-text documents eventually became unwieldy, and the project decided to focus solely on metadata. Metadata searches that resulted in a BETEXT full-text reference would be linked to our own interface for full-text searching and display. The user issues and technical solutions sections that follow in the final third of this paper is thus a product of that relatively late decision.

### 1.3 BETEXT

Despite these complications, the principle goals for BETEXT have remained unchanged. We have conceived of it as a logistical center, a node between the various cultural and educational institutions in the region, functioning as a resource for electronic-text project development and for user education. The center would provide hardware and software resources for digitalization, encoding and publication, as well as expertise and training in the entire process of electronic publication. It would provide the archive producer with a portal shared by other on-line documents, thereby increasing the likelihood that his or her work would be found and used by the research and education communities. It would also, by going into those communities to provide education in the use of such work, guarantee that it got the kind of exposure to promote increased use and interest in further e-text development.

Our partners for both the BETEXT and COVAX projects consisted of the primary cultural institutions of the region. They were:

- the municipality of Karlskrona
  - represented by the city library and the Karlskrona Reading Society, which was active as one of the nation's first circulating libraries between 1794 and 1863, and now functions as an archive of the society's collections
- The Museum of Blekinge
- The Swedish Naval Museum
- The Swedish Navy
  - Represented by the Naval Officers Association and Officers College, which supervise libraries and archives in Karlskrona dating from the late 18<sup>th</sup> century.

This group formed our archive community and would serve us with digitalization project ideas, manuscripts and knowledge about the manuscripts. Our connections to the education community was formed by the university library at BTH, Peter Linde's workplace -- which has a longstanding tradition of providing education in the use of on-line databases for students and instructors at the high-school, professional and university levels -- and the department of Humanities and Social Sciences, where Michael Davis teaches literature. It would be their tasks to "market" BETEXT to readers and teachers.

### 1.4 Chapman

One look at the list of our partners reveals the dominance of the naval culture in the region. Karlskrona, the region's largest city, was the second-largest municipality of the nation at the time of its founding and became one of the most modern naval bases of the 18<sup>th</sup> century, serving as the model for a number of other cities in Europe at the time, including Helsinki. Although the city's national importance has declined dramatically since then, the surrounding archipelago remained largely closed to foreigners for reasons of naval security until the mid 1990's, and in 1998, it was added to the U.N. World Heritage list<sup>4</sup>. As we looked around for interesting manuscripts to digitize, it shouldn't be surprising that we found ourselves sifting through the work of perhaps the 18<sup>th</sup> century's most important ship builder, Fredrik af Chapman (1721 – 1808).

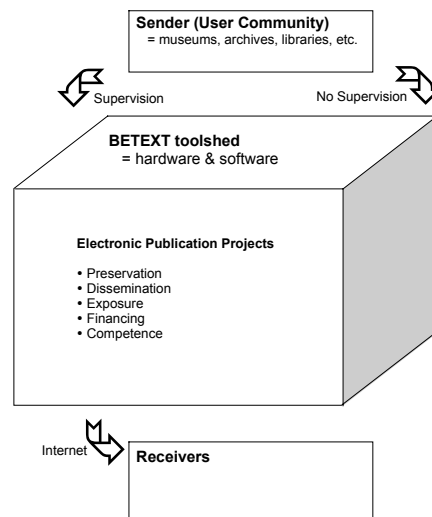
Chapman's international fame is built upon two works: *Architectura Navalis Mercatoria* (begun in the 1760s, published 1786) and *Tractat om Skepps Byggeriet* (1775). The former consists of detailed sketches of current, successful ship design from Greenland to North Africa and has become an indispensable resource to naval historians. Of greater importance at the time, however, was the latter, which was the first successful attempt – in classic enlightenment style – to transform the slow craft of ship building into a mathematical science, with the possibility for "error-free" assembly line systematization. What makes the two works even more interesting for a publishing project is that each set of charts and tables in the *Tractat* is referenced to a specific ship sketched in the *Architectura*. While a wide variety of 18<sup>th</sup> and 19<sup>th</sup> century material was available to us from our various partners, we decided that the challenges and potential benefits of electronic publication (ex. hyperlink references and high-quality image reproduction) were made for these two works.

## 2. Logistical Issues

Before we can discuss the specific technical solutions employed in the Chapman archive, however, it would be useful to review the logistics of our cooperative network.

### 2.1 Cooperation - Issues why and how

While the concept of BETEXT as a toolshed/workshop matured, a number of key components developed. They can graphically be described as below:



To get the attention and cooperation of the senders in the BETEXT project we have had to assure them that the digitization effort will result in at least five things:

- Preservation: that the digitized object is preserved for a reasonably long period of time, that constancy is assured
- Dissemination: that the digitized object is searchable, visible and linked to other in multiple ways that can't be done otherwise
- Exposure: that the whole process of creating digital copies of real life objects get the institutions greater exposure among potential visitors.
- Financing - There is a way... somehow☺
- A raise in digitisation competence - Do it yourself!

### 2.2 Preservation

There is no easy solution for the problem of preserving digital documents for the future. It's a giant problem that one must be aware of and prepared to tackle if you are in the process of producing or administrating digital documents for the future. Archivalist, librarians and museum people know this but have few possibilities to deal with the problem since they are fully occupied with preserving and disseminating their non-virtual documents and objects.

In a study from the Electronic Librarians(eLib) programme on the preservation of Electronic materials<sup>5</sup> a framework for the subject digital preservation is defined where the preservation part is just one of seven modules:

1. Data Creation
2. Data Selection & Evaluation
3. Data Management
4. Resource Disclosure
5. Data Use
6. Data Preservation
7. Rights Management

The needs of librarians, archivalists and museum curators differs of course but most of the concerns of these professionals regarding digitization could be described by the above framework. This framework could easily be transformed into a digitizing policy for a collection and for an organisation. But the average small or middle sized museum or library most likely don't have such a framework or never even thought about it. Therefore BETEXT must be able to offer competence firstly in preservation techniques but also in the other related areas such as digital rights, data management etc.

When we discuss digital preservation we always stress that BETEXT uses a file format that is nonproprietary, untied to a specific computer platform and easy to migrate and export. SGML and XML are generally unheard of, and there is no use of trying to explain XML in technical terms. The need for some sort of center of excellence in these operations is clear when you study the experiences gathered by Warwick and Carty and published in a article called "Only Connect: A study of the problems caused by platform specificity and researcher isolation in Humanities Computing Projects."<sup>6</sup> What they found agreed with our experiences: that there is a lack of technical know-how and a sense of isolation among digitalization and e-publication efforts, especially in the humanities, from where most of the staff at libraries, museums and archives are recruited. Moreover, the know-how is isolated from even the local academic communities. In fact, there seem to be greater competence divides between different humanities specialties than in the technical or medical faculties. As Warwick and Carty wrote:

"It seems that often the most useful information and advice comes from other people working in humanities computing but that contact of this kind can be limited. Conferences go some way towards solving this problem. However, there is a clear need for improved communication between different. It was telling for example that many projects were under the impression that the difficulties they experienced with DTDs were unique to themselves, and were unaware that such difficulties were relatively common. The existence of centres such as the HRI and the CCH go some way towards achieving this, but there is room for much improvement, particularly in for projects which have access to little or no institutional support or expertise. Often interviewees were not aware of the existence of some of the other projects discussed in this report."

Warwick and Carty report that all of the projects experienced problems with proprietary formats or software, and wanted to guard against such difficulties by adopting standard formats. The main reason for the interest of a platform-independent format such as SGML or XML was the need for long term preservation. Another motivating factor behind the move to and interest for SGML was that funding bodies seemed to encourage the use of electronic archives with standard platforms. Other reasons that Warwick and Carty encountered for using SGML include its flexibility and the ability to produce both hard copy and electronic publications from one master and the fact that SGML is recognized as a international standard.

## 2.3 Electronic Publication

Convincing your partner that its meaningful to digitize anything from their collection means that you must be able to offer extensive searching mechanisms, mechanisms that are not possible to attain otherwise. This is a pretty obvious a point to stress since by using XML you have every opportunity to decide what is going to be searchable and how it is supposed to function via the interface. The new digital media we promote is good for at least three things when it comes to dissemination: searching, linking and displaying pictures.

For the Chapman Archive pilot study, we developed our criteria for search capabilities together with specialists at the Naval Museum simply by discussing the two Chapman documents, considering how both the layman and the specialist would go about searching them for information. In this case we were looking for clues about what to tag in the text. Our joint decision was to limit the tags to as few as possible, mainly using attributes for the name tag such as "vesseltype", "person", "place" etc. We could have expanded these searching possibilities almost without limit by adding attributes for details in the rigging, in the armament, in the planking etc, which in turn would make searching and combining these specifics possible. However, since our target users are most likely to be layman, we decided that the extra tags would not be worth the extra amount of tagging labor they would require.

Linking was another great possibility. After looking at the two books for the pilot study and discussing with the librarian at the Naval Museum, we understood that this was a perfect project for digitization since references were made in the *Tractat* to the *Architectura*, but that these two books were very seldom found in the same collection! Hypertext linking the references would, we agreed, help the layman to understand the extent of Chapman's achievement.

The *Architectura* contains 62 drawings of different types of ships. These are all in black and white but they are huge, about 80\*40 cm. In this case photographs were the only choice since no scanner of that magnitude was available in the region. The point here is though that pictures, especially color pictures, are really cheap to reproduce in the digital medium. The resolution might not be that great, but most of the time it's sufficient for the non-specialist. When discussing the advantages of digitalization for such an image-heavy text, a good example to present is the Blake archive<sup>7</sup>. What is done here digitally could not have been realized in print because of costs and the technical limitations of the printed media.

## **2.4 Exposure**

Especially for the small, specialized and middle-sized libraries, museums and archives there seem to be a want for opportunities to tell the world that they exist and that they stock interesting objects well worth checking out. For most of them being able to digitize a unique text or collection would mean that they for the first time would have the potential to reach a national and international audience. It would mean that they would get publicity locally and that the isolation felt would loose its grip a bit. It would also mean that they enter a network, provided by BETEXT, that puts them in a working contact with other heritage institutions in the region and maybe even with new national and international connections.

On a grander scale this has been investigated in the COVAX-project where collections from small and medium-sized memory institutions were converted and made available through one common interface. In this respect our partners in the COVAX project have made similar experiences of the value for museums and archives to be visible on the web as well as in the physical world in order to increase their audience.

The people responsible for promotion in the museum world know that more and more tourists are planning their holidays in advance by searching the net for information about particular regions and collections and exhibitions are always something that can attract visitors that are passing through.

## **2.5 Financing and ownership**

The model for cooperation in BETEXT was to provide hardware, software, methods and competence to the specialists at partner museums, libraries and archives. However the Chapman pilot project has shown that this is not enough. While large institutions like the University of Virginia can provide a long term economic base for an electronic text center, institutions of our scale must look for methods of cooperative financing and joint project applications if projects are to be realized.

The project owner obviously knows what financing sources that might be available for the type of project he has in mind and what kind of organizations support special initiatives of institutions in this area. In BETEXT, there is of course competence about what resources are available for all kinds of mark-up and digitizing-projects. This means that content owners and owners of know-how have to cooperate from the conception of a project till the very end and jointly participate in the different phases of a project.

Copyright laws are not an easy thing to deal with. Therefore we have tried to avoid material that is under copyright. And it is obvious that if you are dealing with rare books and manuscript most of it does not fall under copyright restrictions. When we created the Chapman website as a pilot project for BETEXT the idea was to link it to the Naval Museum's website and to give all copyright for the site to the museum. BETEXT should not possess any copyright for materials produced. Having a clear copyright policy for the electronic text that all partners could agree upon was essential for our success.

## **3. Basic XML Issues**

### **3.1 Why TEI**

It is not unusual that small and medium-sized museums and archives keep their records in homemade MS Access-, FileMaker-etc. Such was the case when the partners at Blekinge Institute of Technology during the conversion phase of the COVAX-project had to deal with several small catalogues in different formats for conversion to XML using standard DTDs. We also converted a sample of full-text documents from local archives, libraries and museums - exhibition catalogues, yearbooks, novels, from print to XML-markup according to the TEI-lite DTD.

Choosing a DTD for the full text documents was never really any great dilemma. The collections being focused on by BETEXT, and the full-text samples needed for the COVAX project were cultural text-documents from museums, archives and libraries. The TEI (Text Encoding Initiative) is an international project aimed to develop guidelines for the encoding of exactly these types of documents. The project has developed a set of very comprehensive DTDs that today are being used as standards for marking up and interchanging cultural texts with SGML and XML. Very soon we found that one of these, the TEI-lite DTD, suited our purpose of converting cultural heritage documents to XML, very well. After having a look at the hundreds of pages of the original TEI-DTD we decided we were not interested in writing our own DTD and that the simplified TEI-lite version seemed to cover all the resources we intended to digitize.

### 3.2 Who and how: Transcription and markup

Tagging is a very time consuming, but a straight forward process as long as you are clear about what DTD, character set and elements to use in your documents.

In BETEXT we try to give the user enough knowledge to make intelligent choices about tagging and TEI-markup. The idea is that the project owner takes the responsibility for tagging the texts and creates valid XML-files. But everyone who has tagged longer stretches of texts knows that it is very easy to prioritize other human activities.

The main text of the Chapman project, the *Tractat*, is about 270 pages long and contains several tables and a lot of mathematical equations. So none of the writers of this paper volunteered for tagging. The question was: Who in the world with enough knowledge of TEI can be tempted to do such a boring job for the very low salary we can offer? The answer was a student with a examination from the distance course "Information Architecture & Electronic Publishing in Practice" that was created for BETEXT as a training and continuing education effort for professionals, but also was given as distance education course of its own in the spring term of 2001.<sup>8</sup> Much of this course material will be transformed into self-teaching modules to be accessible in the center.

We introduced our student-tagger to the text, which he started to scan page by page. These pages then were sent to a typewriting agency where the original paper manuscript was transformed into a Word-file. This file was then imported into XML-spy editor and converted to XML.

All the pictures for the *Architectura Navalis Mercatoria* was sent to us on a CD-disc, from The National Maritime Museum in Stockholm, who had used them for putting out a version of the book on their ChapmanNet web site.<sup>9</sup> These pictures were in the TIFF-format but of no great quality. Since the Chapman project was a pilot study we decided to use the TIFFs anyway, converting them to JPGs for use on the web.

HTML pages were constructed with all the 62 pictures with links to the corresponding text in the "Tractat". The *Tractat* on the other hand was being tagged in XML according to the TEI-lite but all the mathematical formulas were giving us trouble. Since the writing agency had used MS Words Equation Editor all the formulas were specially inserted and could easily be converted to pictures (a lot of pictures!). Instead of trying to convert the formulas to proper XML using TEI (which we doubt can be done) the formulas load as JPG pictures. The pictures are very small and don't slow loading significantly.

The bottom line of this is that tagging can be very time consuming and very expensive if you don't have the necessary skills. This is why all the large publishers and other organizations that are pouring out digital texts in XML-format use the services of companies such as AEL Data - Data Conversion Services.<sup>10</sup> You will find these companies in places like India, Sri Lanka, China and other countries that have extremely low costs for workers salaries. And this is why we must include these kinds of services the next time we do a major conversion. What they do is that they key the documents and also code the document with XML-tags if the customer wants to. The companies we have been in touch with guarantee 99.99% accuracy for a service with double keying, which means that two separate typists types two versions of the manuscript that are afterwards checked against each other for mistakes.

The cost for this is about 0.30 Euros per 1000 keystrokes. Your typical A4 format with double spacing contains about 4000 keystrokes, which then will cost about 1.2 Euros, around 1 dollar per page. Now, if you want the text to be XML-coded as well you just have to add about 30% of the cost per page, which means that the XML tagging would add about 30% extra keystrokes to the text, since the conversion service companies mainly charge by numbers of keystrokes. They don't charge extra for doing XML tagging per se. You just instruct them what

elements are supposed to go where in the text, what DTD is being used etc. Especially if you are planning to do extensive conversions to XML this is an option that must be considered. Prices can always be negotiated.

For everyday XML markup we use XML Spy which is a general tool useful for accessing and importing data from flat files or databases, including Microsoft Access. We found it quite easy to work with XML-Spy since it has all the basic functions we needed. However, it, like most software, is still not very intuitive, so you can't just put it into the hands of the user without quite extensive education efforts. While a limited amount of tagging can be automated (ex. paragraphs and often-repeated names), most of the conceptual tagging can't. Thus BETEXT's role will always be to deliver education to interested memory institutions or, for those who can't afford to dedicate their own personnel to this kind of project work, to promote a network among the institutions for the sharing of the financial burden.

### 3.3 Basic Tagging choices

All the documents we XML encode are tagged according to the TEI-lite standard, which includes the three "supertags": <TEI.2></TEI.2> which tells us that the document conforms to the TEI rules and comprise two tags: <teiHeader></teiHeader> which contains information about the printed source, the electronic version, revisions, and etc., and <text></text>, which contains the text itself

In our experience it's the header part of the TEI markup that is one of the most important and most difficult aspects of tagging to understand for the user. For the benefit of our user partners we have constructed a template that is supposed to make life easier for the novice tagger. Via a web form the user can easily create header tags that can be copied and pasted into any XML-document. This is done very much like how several Dublin Core metatag-creator-services work,<sup>11</sup> using HTML forms on a web page that lead the user through the steps of filling in the TEI-lite header fields. On submission, the JavaScripts generate the beginnings of a TEI XML document, with a well-formed and validatable header.

In the text part of the document we have always stressed the importance of limiting the number of elements to as few as possible, concentrating on the user aspect - What does the user want to find? This is essentially up to every partner in the BETEXT project to determine. They know their data and their users and therefore must have free hands to determine useable elements within the TEI-lite DTD. The only thing apart from making it simple, that we stress from BETEXT is that they use the header elements that we demand and that can be easily created through the header template.

### 4. Basic User issues

For a potential user of the BETEXT archives, the COVAX decision to make document display the responsibility of the BETEXT site, presented a number of challenges. COVAX header-section searches would link the user to the document, but we would have to find a way to then provide users with body-section searching. We would also have to find a method of displaying what had become very large web files, the *Architectura*'s simple XML file transforming into over 5 megabytes when all its images are included. This meant not only discovering workable and attractive XSLT methods – methods that could be routinized for the BETEXT center later -- for image embedding and linking from one document to another in XML; it also meant developing a technique for delivering the document in small sections, that is to say, hyper-linked individual pages, using the original pagination. Virtually all SGML and XML encoded document databases deliver little more than simple, formatted text, in which entire documents or perhaps chapters are presented in long, scrolling web pages. Our chief technological challenge would be to deliver the kind of user-friendly document we had been forced to design.

### 5. Technical Solutions

Based on the user issues at stake and the needs and interests of our institutional community, we wrote a technical proposal and sent it around to a number of consulting firms. The proposal was very precise about its requirements for searching, browsing and displaying. In the end we contracted Aerotechtelub Information & Media to develop the application, built of course, on our TEXTML server from IXIASOFT.<sup>12</sup>

Searching and Browsing functions were relatively easily achieved, using Xpath to trigger index-creation in the TEXTML server. In fact, perhaps the only serious problem we've faced, technically speaking, as explained, we wanted to extend the search application functionality to display a single page of the whole document at a time.

XML files stored on the server follow markup rules according to the TEI LITE dtd. The XML instance contains empty page-break elements that reflect the layout of the real document. That is to say, it is similar to the <br> tag in HTML; it tells us when the page breaks but does not contain the page. This creates difficulties if you want to extract page by page from the XML-file; there's no page element to extract. Another thing to take into consideration is transformation with the aid of XSLT and style sheets. The XML-file is a digitized form of a "real world" document, and every type of "real world" document has a specific layout. We try to reflect that layout in the digitized version with the aid of style sheets. However, when you extract the document page by page, tags that are started on the previous page are 'lost,' creating problems for displaying with correct layout. A possible solution of this could be the use of a SAX parser and applying a default layout when viewing document by single page.

## 6. General Conclusions

For most small- and medium-sized museums, libraries and archives, being on the Web means no more than information about opening hours and collections and not the on-line display of objects, texts and manuscripts. Moreover, even this limited amount of information remains largely in the hands of commercial web-agencies and beyond the control and competence of the information owner. Although one would think that these memory institutions have had the opportunity to learn the processes of digitization and electronic publication during the 15 years since computers made their every day lives, they have not.

This is why we are convinced that the BETEXT idea of a regional tool-shed for digitizing important and unique material from local memory institutions is sound. As the axis of a logistical network, we can offer or coordinate technical expertise, knowledge and labor for projects that our partners could not develop on their own. In fact, we currently work on connecting to national organizations, hoping to form a larger network (which in turn would have contact with international bodies like TEI and OCLC) that can potentially coordinate support, standards and training.

Nevertheless, it's neither been the structure of markup nor the technical solutions that have been most difficult to develop in the BETEXT project; it's been to inspire engagement and project-work out of the partners. At the start we were forced to present, in concrete terms, the possibilities and benefits of electronic publication, at a long series of meetings, as well as to being the driving force behind the writing of our grant applications. Now, with the pilot project completed, it appears that it will take the same process all over again to build upon it with new projects. Ultimately, our greatest skills at the center may have to be rhetorical.

---

<sup>1</sup> <http://etext.lib.virginia.edu/mission.html>

<sup>2</sup> <http://www.virginia.edu/stats&facts/>

<sup>3</sup> <http://www.bth.se/valkommen.nsf> (see the "Snabbfakta" link)

<sup>4</sup> Despite what you may have heard, the policy also applied to the Soviet submarine corps.

<sup>5</sup> Hendley, Tony. "Comparison of Methods of Digital Preservation": A JISC/NPO study within the Electronic Libraries (eLib) Programme on the preservation of electronic Materials. May 1998.

<sup>6</sup> Warwick, Claire and Celine Carty. "Only Connect: A study of the problems caused by platform specificity and researcher isolation in humanities computing projects." In: *2001 in the digital publishing odyssey: proceedings of an ICC/IFIP conference held at the University of Kent at Canterbury, Kent, United Kingdom, 5-7 July 2001*. Eds. Arved Hüler, Peter Linde and John W.T. Smith.

<sup>7</sup> <http://www.blakearchive.org/>

<sup>8</sup> <http://www.bth.se/ihu/elpubkur.nsf>

<sup>9</sup> <http://130.242.29.96/sjohistoriska/Avdelningar/ritarkiv/chapmanintro.html>

<sup>10</sup> <http://www.aeldata.com/index.htm>

<sup>11</sup> <http://www.lub.lu.se/cgi-bin/nmdc.pl>

<sup>12</sup> Detailed information on TEXTML is available at <http://www.ixiasoft.com>