



Arts & Humanities
Research Council



christine de pizan

The Making of the Queen's Manuscript (British Library, Harley 4431)

Designing a Digital Version of British Library, Harley MS 4431

Peer-Reviewed Conference Paper first presented at:
L'écrit et le manuscrit à la fin du Moyen Âge
2e colloque international du GRMF
Université catholique de Louvain 12-14 mai 2005

JAMES LAIDLAW &
CHARLIE MANSFIELD
UNIVERSITY OF EDINBURGH

The Making of the Queen's Manuscript

This research programme centres on British Library, Harley MS 4431, the collection of works by Christine de Pizan which she presented to Queen Isabeau of France early in 1414. Funding for the project, which runs from October 2004 to September 2008, has been provided by a grant of £199,716 (ca. €300,000) from the UK Arts and Humanities Research Council. Important material help and support is being given by the British Library. The project is based in the University of Edinburgh (see <http://www.pizan.lib.ed.ac.uk>), and is being carried out by an international team of scholars.

The project team:

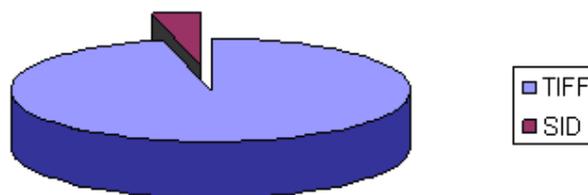
François Avril (Bibliothèque nationale de France - Consultant)
 Justin Clegg (British Library - Liaison Officer)
 Marie-Thérèse Gousset (Bibliothèque nationale de France)
 Andrew Grout (University of Edinburgh - Digital Library Officer)
 James Laidlaw (University of Edinburgh - Project Director)
 Anthony Lodge (University of St Andrews)
 Charlie Mansfield (University of Edinburgh)
 Gilbert Ouy (CNRS - Consultant)
 Gabriella Parussa (Université François Rabelais, Tours)
 Christine Reno (Vassar College)
 René Stuip (University of Utrecht)
 Tania Van Hemelryck (Université catholique de Louvain)

Page and Image Management During the Transcription Phase

The British Library's main contribution to the project is a complete set of digitised images of Harley MS 4431. The scanned photographic images were delivered in December 2004 and January 2005. These 800 TIFF (Tagged-Image File Format) files have an average size of 170 megabytes; only four files fit on one CD-ROM. Images of this size and format were too big to deliver across Europe and the USA to the team of six charged with the transcription of the manuscript. We chose to convert them to a manageable format called MrSID (MultiResolution Seamless Image Database), developed in the USA by LizardTech.

First we correct each page image through Adobe PhotoShop using an Unsharp-Mask setting of Amount = 90%, Radius = 1.4 pixels and Threshold = 1 level. About half the images have to be rotated very slightly (between 0.2 and 0.8) so that the text lies square on the canvas.

We run a batch process to convert the TIFFs to SIDs. The SID files display in the Insight package and in a SID File Viewer package which was available from LizardTech in 2004, but has now been superseded. The SID files have a size of about 7 megabytes each, which means that over 80 files fit on a CD-ROM. More importantly, the SID files are of a suitable size for serving over the Internet. Please see the pie-chart to compare file-sizes:



Size Relationship between compressed SID files and TIFF files.

The beginning of February 2005 saw the creation of the “Pizan Safe” on one of the Edinburgh University servers, into which the SID files are being loaded. The Safe holds almost 600 files at present; we expect that all 800 files will have been processed by the end of May 2005. Entry to the Pizan Safe is gained by a User Name and Password which must be supplied each time a member of the project team uses a web browser to access the sub-directory. A menu page, using JavaScript, has been designed to allow users to select folios from the Safe by simply typing the folio number followed by “r” or “v”, for recto or verso. The JavaScript code is as follows:

```
<head>
<script language="javascript" type="text/javascript">
function formproc(theform)
{
  {location.href=theform.text1.value+".sid";
  return true;}
}
</script>
</head>
<body>
<form>
  <p> <font face="Arial" size="2">Click first in the
answer box,<br />
  type in folio number without the letter f and <br />
  then click the "submit" button.</font>
  Please remember the r for recto or v for verso in
  lower-case after the folio number.
  <br /> <font face="Arial" size="2">
  <input type="text" name="text1" value="" size=10>
  <input type="button"
value="submit" onClick="formproc(this.form)" name="button">
  </font> </p>
  </form>
</body>
```

Experience of the e-commerce industry has taught that opening a range of channels for users to access resources gives greater flexibility. It also provides solutions to problems that have not yet been posed, after Pierre Levy’s notion of the virtual (Mansfield 2000). The team transcribing Harley 4431 are able to use either CD-ROMs or the Pizan Safe for convenient and rapid access to the folios on which they are working.

Approaching Digital Corpus Design

The Christine de Pizan project is fortunate in being able to rely on the pioneering work done by the Text Encoding Initiative (<http://www.tei-c.org.uk>) and by *The Digital Scriptorium* (<http://sunsite.berkeley.edu/Scriptorium/>). With their help we are able to set the project on firm foundations.

In creating a digital corpus it is important to define the projected user group before mark-up of the transcribed text commences. The definition proposed by *The Digital Scriptorium* offers a useful starting point.

A transcription is adequate for paleographic or codicological purposes when it records paleographic or codicological detail completely enough to allow paleographers or codicologists working with the manuscript to search in the transcription to find phenomena of interest to them.

(<http://sunsite.berkeley.edu/Scriptorium/technical/dsguide1.html#text>)

Our design for the electronic transcription of Harley 4431 aims first of all to create a diplomatic corpus made up of the thirty works which Christine included in the Queen's MS. As a second stage we will develop scholarly editions of all these works, based on the diplomatic corpus. Both these electronic versions will be searchable with the help of concordancing software. We also intend that these two design corpora should be useable by other computer systems. The generally accepted mark-up language which will enable us to achieve these aims is XML (eXtensible Markup Language), supported by a transformation language called XSLT (eXtensible Stylesheet Language for Transformation).

Once careful and detailed mark-up using XML has been completed, other scholars will be able to adapt the corpora for their own purposes. For example, the corpora will be made available for dictionary compilation by the Research Laboratory at the University of Nancy-2 - ATILF (Analyse et Traitement Informatique de la Langue Française), with which we have established contact.

One of the exciting features of XML is that the designer of a corpus provides windows into the data, which can be used by future researchers for purposes as yet unforeseen. In order to understand how a text marked-up with XML tags can work, a helpful metaphor is the provision of window embrasures during the building of a tower. Here, in a miniature from Harley 4431, we see how spaces for windows are being left as part of the building plan.



Detail from Harley 4431, folio 290 recto

© British Library

Transcription DTD (Document Type Definition)

The Digital Scriptorium has developed a Document Type Definition, “ds3.dtd” in its most recent version, which complies with the guidelines set by the Text-Encoding Initiative (TEI) for the transcription of medieval and renaissance manuscripts.

The purpose of ds3.dtd is to help individual scholars and students produce machine-readable transcriptions of medieval and renaissance manuscripts in a format compatible with the TEI Guidelines. Relevant issues are therefore a combination of textual questions (e.g., prose, verse, corrections, omissions) and physical questions (e.g., change of hand, change of ink, illegibility due to loss of support or due to deterioration of ink).

(http://sunsite.berkeley.edu/Scriptorium/technical/transcription_dtd.html)

The DTD.exe is free to download from

http://sunsite.berkeley.edu/Scriptorium/technical/transcription_tei.html

Using an inexpensive editing package called NoteTabPro we are able to hold the ds3 DTD alongside our corpus as we mark-up our own XML file. An example line of tagged text, taken from the *Epistre Othea*, shows more clearly what is involved:

```
Ressembler te couient <name type="person">Minos</name>
```

The tag for proper names is defined in the TEI guidelines. The tag has attributes which enable names of persons and places to be distinguished. The attribute added in our example above indicates that Minos is a person, not a place. All XML tags have to be switched-off with a closing tag, here </name>, so that the tagged element, in this case *Minos*, is clearly delimited. Each tag can be seen as a window, to recall the building metaphor.

Transforming the Tagged Corpus – A First Peek Through a Window

The second stage is to prepare an example transformation of the corpus. This equates to taking a first look through one of our metaphorical windows. To do this we need to code an XSL file which will accompany the XML file. XSLT is more like a programming language than XML, since it uses conditions (IF statements) to handle the manuscript text and the tags.

A typical XSL instruction reads like this:

```
<xsl:template match="name">
<xsl:apply-templates/>
<font face="verdana" size="1">
  <xsl:value-of select="@type"/>
</font>
</xsl:template>
```

The instruction is to be understood as follows: IF or WHENEVER the tag, “name”, is encountered in the XML file, then change the font size to a small Verdana face and display (through the window of the computer screen) whatever is stored in “type”; in our example which involves Minos, the attribute word is “person”.

The transformation made by the XSL program produces an HTML file, i.e. a traditional web-page that any browser can display. The above example displays like this, with the type attribute in Verdana and in smaller font-size:

Ressembler te couient **Minos**_{person}

In that way we take an initial step towards providing a scholarly transcription of the manuscript. The attribute “person” is of course too gross a category, but we will refine this by the addition of a sub-attribute, for example, `status="mythical hero"`.

Conclusion

At the outset of the paper we suggested that digitising Middle French manuscript texts usually requires the definition of the projected user group of the electronic corpus before mark-up of the text commences. By using XML, we hope that our design can so far as possible be future-proofed, so that later scholars will write their own XSL programs to transform our corpus to accommodate their research interests, in ways that we could not have anticipated at the outset of the work.

The management of large manuscript images during the development phase of a digitisation project creates its own challenges. We have approached them with good practice from e-Commerce, opening multiple channels for the distribution of the images. Image delivery times have been cut dramatically. Furthermore, we have made it possible to position the image of each folio and the transcribed text side by side on the computer screen.

Bibliography

Brown, Cynthia J. (1995) *Poets, Patrons and Printers – Crisis of Authority in Late Medieval France* Ithaca & London, Cornell University Press

Laidlaw, James (1987) 'Christine de Pizan – A Publisher's Progress' in *The Modern Language Review*, January 1987, Volume 82, p. 35-71.

Mansfield, Charlie (2000) 'What is the virtual - the writings of Pierre Lévy' in *The Web Journal of French Media Studies* Volume 3, Number 1, October 2000
ISSN 1460-6550 [En ligne] <http://wjfms.ncl.ac.uk/>

Poupeau, Gautier « Réflexions sur l'utilisation de la TEI pour coder les sources diplomatiques à partir de l'exemple du Cartulaire blanc de l'abbaye de Saint-Denis », *Le Médiéviste et l'ordinateur*, 43, 2004 [En ligne] <http://lemo.irht.cnrs.fr/43/43-12.htm>

Web Sites Referenced

The DTD.exe may be freely downloaded from
http://sunsite.berkeley.edu/Scriptorium/technical/transcription_tei.html
With full information from Berkeley University at
<http://sunsite.berkeley.edu/Scriptorium/technical/dsguide1.html#text>

Christine de Pizan: The Making of the Queen's Manuscript. Project Web-site
<http://www.pizan.lib.ed.ac.uk/>

Le laboratoire ATILF / CNRS Nancy-2
Analyse et Traitement Informatique de la Langue Française
<http://www.atilf.fr/>

JAMES LAIDLAW (james.laidlaw@ed.ac.uk)
CHARLIE MANSFIELD (charlie.mansfield@ed.ac.uk)
UNIVERSITY OF EDINBURGH

12 May 2005