MUSEUM MEETS ACADEMIA: THE GOSLAR TO GRASMERE PROJECT

Richard Light
3 Midfields Walk
Burgess Hill
United Kingdom
richard@light.demon.co.uk

"Goslar to Grasmere" is a collaboration between the Wordsworth Trust and the Wordsworth Centre, Lancaster University. It demonstrates the potential for creating information resources which are based on museum collections but which go beyond traditional "object catalogue publication". The project has produced a web site which is generated dynamically from the records in the database, allowing new pages and material to be added at will. The transcribed manuscript sources in TEI format are not "locked into" the site, but are available for other uses. Thus, "Goslar to Grasmere" is a good example of how digital resources can be created in a format which allows their future re-use.

Background

The Wordsworth Trust is a museum and archive which holds a major Designated collection of primary sources relating to Wordsworth, his immediate circle, and the Romantic poets in general. This material has been catalogued as a "museum" collection over the past 15 years, using the MODES cataloguing software and more recently ModesXML. These records are an analysed summary of the object itself, its significance, and its collections management history. Thus a sheet of paper containing a manuscript poem might be recorded thus:
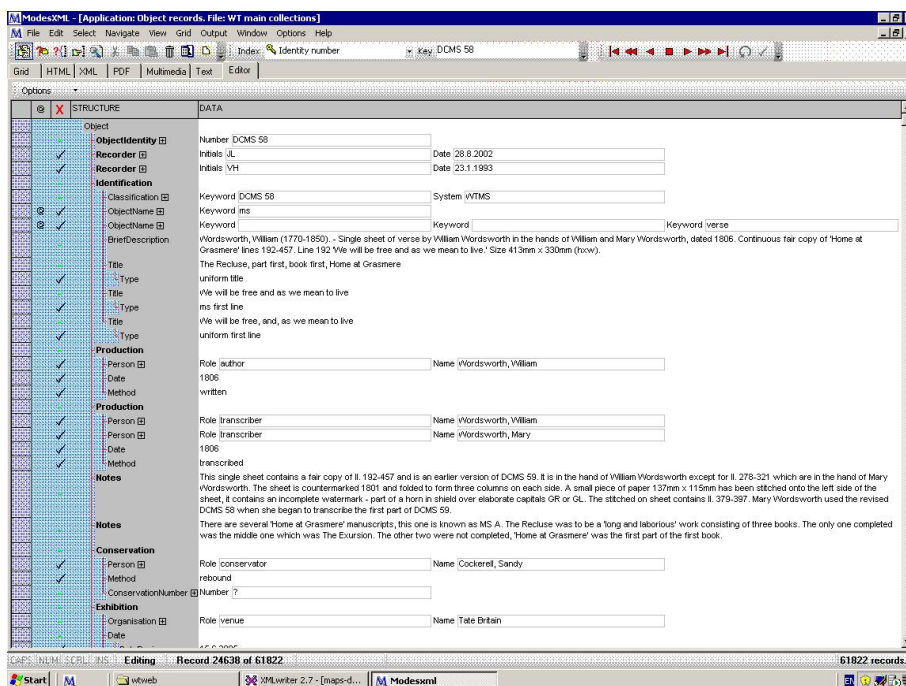
Figure 1: Object summary record

The information relating to the object's actual content is minimal: various titles and a prose description:

```
<Title>The Recluse, part first, book first, Home at Grasmere<Type>uniform title</Type></
Title>
<Title>We will be free and as we mean to live<Type>ms first line</Type></Title>
<Title>We will be free, and, as we mean to live<Type>uniform first line</Type></Title>
...
<Notes>This single sheet contains a fair copy of ll. 192-457 and is an earlier version of
DCMS 59. It is in the hand of William Wordsworth except for ll. 278-321 which are in the
hand of Mary Wordsworth. The sheet is countermarked 1801 and folded to form three
columns on each side. A small piece of paper 137mm x 115mm has been stitched onto the
left side of the sheet, it contains an incomplete watermark - part of a horn in shield over
elaborate capitals GR or GL. The stitched on sheet contains ll. 379-397. Mary Wordsworth
used the revised DCMS 58 when she began to transcribe the first part of DCMS
59.</Notes>
```

The Goslar to Grasmere project

"Goslar to Grasmere" is a collaboration between the Wordsworth Trust and the Wordsworth Centre, Lancaster University. The project aims "to explore the relationship between physical, inhabited space; represented, imagined space; and material and textual space" by an examination of two early Wordsworthian texts: the earliest surviving drafts of *The Two Part Prelude* and the manuscripts for *Home at Grasmere* in four MS notebooks. This required a re-examination of this material from a different perspective, more detailed than that required to catalogue it.

First, the sources needed to be photographed, so that layout of the original text could be presented. Standard digital photography was used to achieve this. Second, the text itself had to be transcribed, so that it could be annotated and used for searching.

Transcription of primary sources

The Text Encoding Initiative (TEI) Guidelines[1] were chosen to encode the texts. The TEI framework is widely used to encode texts of academic interest, with the intention that the results can potentially be re-used in future by other projects. It provides a rich set of tools, which allow a text to be analysed in whatever manner is relevant to the project at hand.

In this case we were interested in capturing information such as:

- physical features (crossings-out; insertions; blots)

- psychological elements (Freudian slips; inversions)

- subjects or themes in the poetry

- links to image files

- position of a line within its image file

---

1   http://www.tei-c.org/index.xml

TEI "feature structures" proved capable of encoding all of these. Here is an example of a line containing an unclear letter "i" and an example of "contamination":

```
<l n="198" id="DCMS58HAG.1.7">Far less from any momentary f<unclear>
<add type="supralinear" hand="hand.ww1">i</add></unclear>t
<fs type="line.feature"><f name="work"><str>Home at Grasmere</str></f> <f
name="hand"> <str>hand.ww1</str> </f> <f name="accidental
revision"><str>Contamination</str></f></fs><?MXMLparent DCMS58HAG.1?></l>
```

Here is a line whose offset within the image of its containing page is recorded:

```
<l n="196" id="DCMS58HAG.1.5">And of the power she has within herself
<fs type="page.offset" id="DCMS58HAG.5a"><f name="offset.left"><msr value="3"
unit="px" valueTo="308"/></f><f name="offset.top"><msr value="96" unit="px"
valueTo="130"/></f></fs><?MXMLparent DCMS58HAG.1?></l>
```

We chose to encode subjects and themes using "segments", which allow simple "type/subtype" analysis:

```
<l n="3" id="RTHAGMSB.3"><seg part="M" subtype="Repotia" type="Repetition"><seg
subtype="Parenthesis" type="poetical effects" part="M"><seg subtype="Memory /
Forgetting" type="Body and Mind">I cannot well remember</seg>, but the<seg
type="Thematic Time" subtype="Hourly"> hour</seg></seg></seg></l>
```

While the database we used – ModesXML[2] – supports XML natively, it has a simple indexing mechanism which only retrieves complete "records". We therefore analysed poems so that each line was a separate "record" in the database, and used XML processing instructions to link these records together. The examples above include "parent" links; in addition there are "child" links from the record about the whole poem to its individual lines:

---

2 http://www.modes.org.uk/index.asp?page=modes

```
<div type="page" id="DCMS58HAG.1">
<figure url="MSAcol1frrightNIK1140_jpg.jpg" rend="scalingfactor1"><fs
type="image.dimension" id=""><f name="width"><msr value="413" unit="px"
valueTo=""/></f><f name="height"><msr value="883" unit="px"
valueTo=""/></f></fs></figure>
<div type="page-body">
<lg>
<?MXMLchild DCMS58HAG.1.1?>
<?MXMLchild DCMS58HAG.1.2?>
<?MXMLchild DCMS58HAG.1.3?>...
<?MXMLchild DCMS58HAG.1.46?>
</lg></div></div>
```
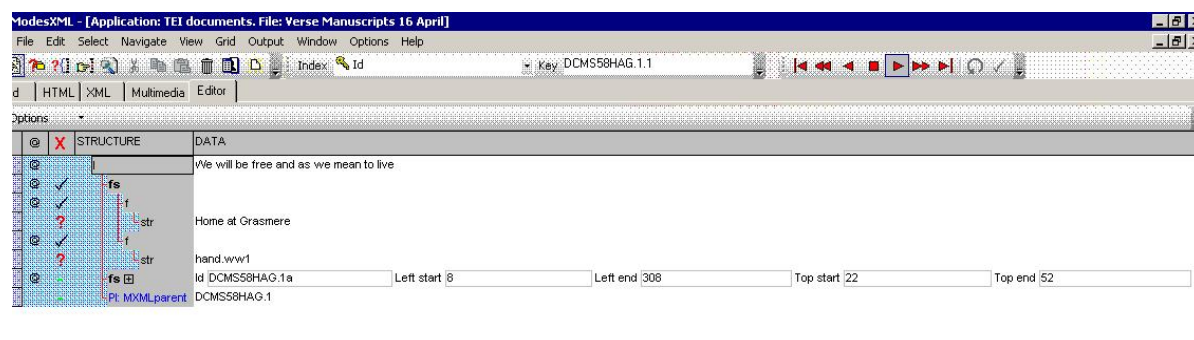
These links allow us to reconstruct all or part of any poem, as required. This is how the data looks while being edited in ModesXML:

Figure 2: Line of poem being edited

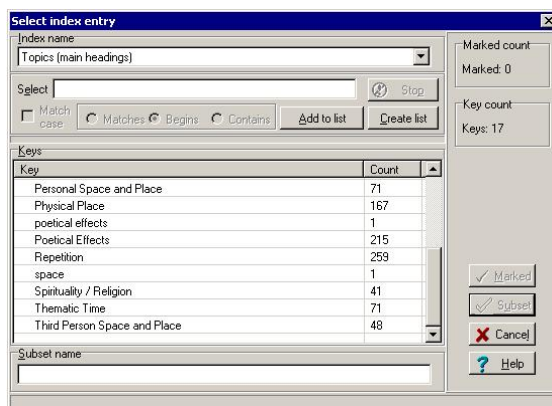The database automatically maintains indexes based on the metadata which has been added:

Figure 3: Main subjects index

The web site

We required a web site which would display and explain this material. The ModesXML database was again used: this time to hold the web site pages. However, these were authored separately using a word processor. In order to convert them to the required XML format, they were opened in Open Office and then exported as XHTML. An XSLT transform then converted this XHTML into the form of XML required for "web site" records, and the resulting documents were imported directly into the web site data file.

This is an example of a web page viewed as XML:

```
<page><pageId sortOrder="1">ChooseYourRoute</pageId>
<context><parentPageId>FrontPage</parentPageId><childPageId/></context>
<metadata><topic type="generator" value="StarOffice/OpenOffice.org XSLT
(http://xml.openoffice.org/sx2ml)"/> <topic type="author" value="Sally Bushell"/> <topic
type="created" value="2007-07-08T14:34:00"/> <topic type="changedby" value="Sally
Bushell"/> <topic type="changed" value="2007-07-08T14:44:00"/> <topic
type="navigator" value="Choose Your Route"/></metadata>
<css> ... </css>
<content> <h1 class="GGHeading1">Choose Your Route </h1> <p
class="GGnormal">This website is designed with filters to allow different users to access
different elements of the interpretative metadata. You can choose to access the entire site,
or to see only parts of it which are relevant to your interests. Select a filter by clicking
below:</p><p class="GGPageLink"><mwsPageLink pageId="FrontPage">Access Whole
```

```
Site (no filter) </mwsPageLink></p> ... </p></content></page>
```

The "web site" XML application consists of a simple "wrapper" around standard XHTML 1.0, with elements such as Cascading Style Sheet (CSS) instructions being kept separately from the page content. The XML record also contains information about this page's "parent" page, which allows tables of contents and navigation panels to be constructed dynamically.

Linking it all together

The "web site" XML is dynamically converted into HTML using an XSLT transform. Markup within the "web site" XML is used to indicate linkages between data in the different files. For example, the element given above:

```
<mwsPageLink pageId="FrontPage">Access Whole Site (no filter) </mwsPageLink>
```

indicates a cross-reference from one page on the site to another. This will be rendered within the final page as a standard HTML <a href="XXX"> hyperlink. Links to the other data files have their own coding, e.g.:

```
<mwsRecord indexkey="DCMS58HAG" indexname="Id" mode="summary1"
mwsLink="HAG"/>
```

links to the record with key "DCMS58HAG" in the file specified by the link "HAG", which in turn is defined as a link to a specific index on a specific file:

```
<mwsLink name="HAG"><application>TEI</application><filename>Verse Manuscripts 16
April</filename><indexname>Id</indexname></mwsLink>
```

The end result of all this linking is that the primary sources (the TEI-based transcription and analysis) can be used within web pages without the need to make copies of their content. Thus the following page:
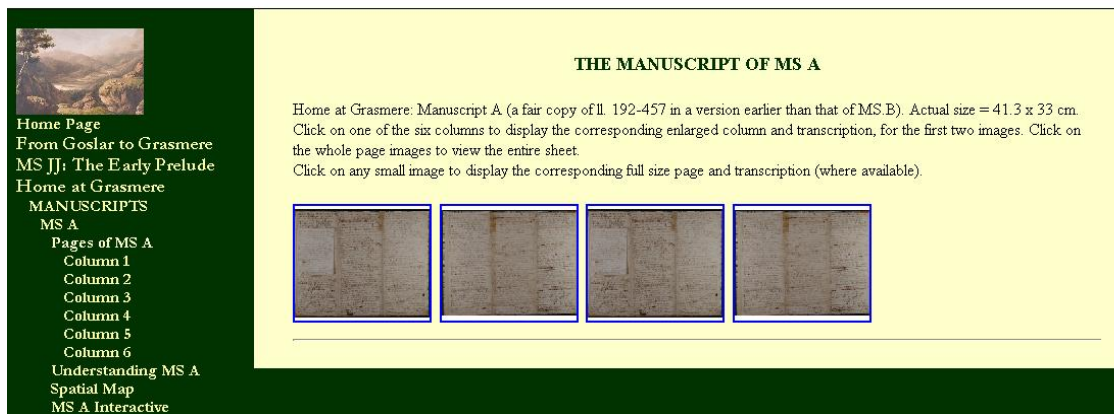

Figure 4: Web page with dynamically included text

was generated by looking up the TEI record for DCMS58HAG, then picking out a "thumbnails" child record which contains the names of the four thumbnails of images of DCMS 58. Clicking on any of these thumbnails will bring up a larger image of the selected column, together with a formatted transcription:
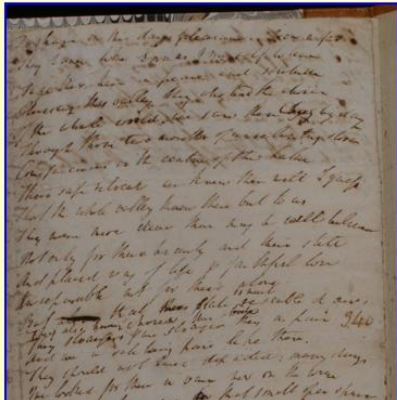
Figure 5: Web page with TEI transcription included

This site nicely demonstrates the possibility, which XML offers, of separating content from presentation. Combining that general ability with a database gives even greater flexibility. The same source material could be displayed in many different ways, using different selections of data, etc. In this site, "filters" are used to present only those pages which match a particular type of interest (general public; specialist; educationalist). Searches will pick out the lines (or groups of contiguous lines) which demonstrate a particular stylistic feature. Here we see two lines from the above extract featuring in the results of a search:

MS A: search for "({Line features}={hand: hand.ww1}) AND ({Line features}=*{accidental revision:})"

This is the result of your search for "({Line features}={hand: hand.ww1}) AND ({Line features}=*{accidental revision:})":

[198] Far less from any momentary f[ˡ]t

[208] Ambition not aspiring to the pr_{aise}ize

[209] Of outoward things but for the prize within

[213] And for the Soule I do not say too much

[218] Bleak season was it turbulent & bleank

[243] Had yieldeded up his breath the awful trance

[260] That [?faced] with a passionate welcoming 260

[326] They came like Emmas & myself to live

[329] Of the whole world: we saw them by ^{day}by day

[352] The Sheeph[?h]erd may have seized the deadly tube

[369] Of Strangers, blessd already, none would would give

Figure 6: Search results including specific transcribed lines

Sets of lines are also combined dynamically to form a "speculative text": a possible form that the complete poem might have taken at a particular point in the process of its creation.

Beyond "Goslar to Grasmere"

The Wordsworth Trust has now taken a similar approach to its main collection. Summary records like the one in Figure 1 are, where appropriate, being complemented by full TEI transcriptions, for textual material like letters. In addition, new ModesXML files are being set up to record people and places, and links to these files are being made from the object records. This allows displays like this to be produced:
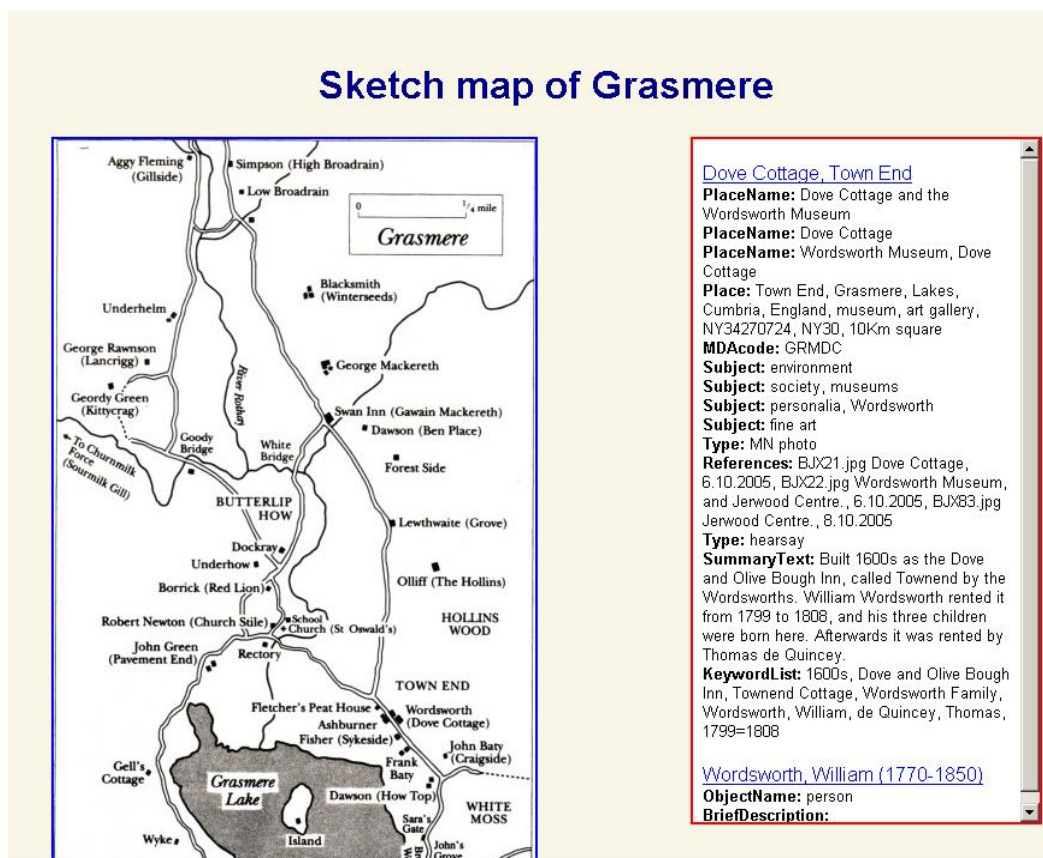
Figure 7: Web page linking map to gazetteer and biographical records

where clicking on "Dove Cottage" on the map brings up both the gazetteer entry and the biographical data associated with that place. (This link makes use of a Topic Map, which contains fine-grained links between people, places, events, etc. The Topic Map is generated automatically from the other data sources.)

Overall, this approach holds out the promise that the Wordsworth Trust will be able to record its data once, without unnecessary duplication, and then "publish" it in as many different contexts as it requires. In principle, the same information could also be made available to external bodies in machine-processible form. For example, the main collections file is now available for harvesting via the Open Archives Initiative (OAI) protocol.