Towards a European Standard for Manuscript Description: the MASTER project

Lou Burnard, Muriel Gougerot, Elizabeth Lalou, and Peter Robinson

1. Introduction

This paper describes the MASTER project, an EU-funded project which in consort with several other related initiatives is attempting to define and implement a general purpose XML-based standard for the description of manuscript materials. We explain briefly the background and history of the project and discuss its links with other groups pursuing the same aim. At the time of writing, a preliminary draft of the MASTER standard is nearing completion; some indication of its underlying design principles and likely content is also included.

MASTER (Manuscript Access through Standards for Electronic Records) is a thirty-month project funded by the Telematics for Libraries section of the European Union Fourth Framework research program. The partners in the project are: De Montfort University (leader); the Royal Library, the Hague; the Arnamagnaean Institute, Copenhagen; L'Institut de recherche et d'histoire des textes, Paris; the National Library of the Czech Republic, Prague; the University of Oxford. Associated partners in the project include several major European libraries, notably the British Library, the Vatican Library, the Biblioteca Ambrosiana, and the Bodleian Library, Oxford. The project began on 1 January 1999.

The project is closely linked with the North American manuscript projects EAMMS (Electronic Access to Medieval Manuscripts) and Digital Scriptorium, and also with the international Text Encoding Initiative (TEI). While the focus of this paper is on MASTER itself, we stress that the ideas described here result from much previous work by scholars in these and many other projects. The formal mechanism for this collaboration is a Work Group on Manuscript Description, chartered by the international Text Encoding Initative (TEI) and co-chaired by Consuelo Dutschke (Columbia University) and Ambrugio Piazzoni (Vatican Library). This Work Group was charged with developing a document type definition (DTD) for the description of manuscript material which would be compatible with existing TEI recommendations for machine-readable texts. The TEI Work Group's proposals will be submitted to the TEI's Technical Review Committee for ratification and eventual inclusion in a subsequent edition of the TEI Guidelines. Like other TEI workgroups, the group has a responsibility to ensure that its recommendations are acceptable to the broadest consensus within the domain; it is particularly useful therefore that so many different initiatives are represented on it.

In practical terms, cooperation amongst these projects is very close. For example, Dutschke is leader of both the EAMMS and Digital Scriptorium projects, co-chair of the TEI workgroup, and has attended all MASTER plenary meetings as an observer; as European editor of the TEI, Burnard is a member of the TEI Work Group as well as being primarily responsible for the standards development workpackage in MASTER; Robinson is leader of MASTER, a member of the EAMMS and Digital Scriptorium advisory boards, and a long-standing associate of the TEI having chaired one of its key technical committees.

2. Background to the project

Many scholars have made suggestions concerning the making of computer-readable manuscript descriptions over the last twenty years or more, and many different computer systems have been been designed to support them. Notable pioneering initiatives include those of the IRHT in the 70s (resulting in the formation of the MEDIUM database) and of the Italian Censimento in the early 90s, and a convenient summary of several such systems can be found in the volume Bibliographic Access to Medieval and Renaissance Manuscripts: A Survey of Computerized Data Bases and Information Services, ed. Wesley Stevens, Primary Sources and Original Works, Volume 1, Numbers 3/4, 1991. Such systems have tended to reflect a particular research context, and thus to be rather narrow in scope of application; they have also almost all chosen to use database technology of some kind. By contrast, the goal of MASTER is to create such a generic system, sufficiently flexible and robust for application in many different domains of manuscript description, and the technology chosen to support this more ambitious goal is based on the international standards SGML and XML.

The advent of the world wide web and advances in digital imaging of manuscripts gave a new impetus to the work of computer based manuscript description during the mid-nineties. Almost at the same time that the libraries of Europe realised how digital imaging could offer an enormous expansion in the accessibility of manuscripts in ways never before envisaged, so they also realised the pressing need for online cataloguing of such materials. Only with this in place, would the provision of unified access to the new digital repositories springing up across

Europe make the difficult transition from the technically feasible to the practically realized. And such unified access could only be achieved by prior standardization work – by agreement about how manuscripts should be described. In the absence of any such agreement, a number of projects began attempting to define standards in this area.

In North America, a key event was the funding by the Mellon foundation, of two collaborative projects: the Electronic Access to Medieval Manuscripts project (EAMMS) and Digital Scriptorium. Both these projects began in 1996, and brought together American and European experts on computer systems and manuscript cataloguing. The inception of these American projects provoked a group of European scholars to create a European counterpart to these American projects. The stimulus for this was a talk given by Hope Mayo, then leader of the EAMMS project, at the Kalamazoo conference on Medieval Studies in May 1996, attended by Robinson. Robinson and Mayo then met later that summer, and from this came a seminal meeting at Studley Priory, near Oxford, in November 1996 which brought together the partners who later formed the core of the MASTER project. After a long process of negotiation with the European Union, MASTER began (with funding from the Fourth Framework program) on 1st January 1999.

The present proposals are based on a different approach to that of the database systems implementations characteristic of the methods developed in the 1980s and described in the Stevens volume referred to above. Firstly, the various database systems developed for describing and distributing manuscript descriptions were just that: various indeed, with multiple problems relating to compatibility of data across systems and the binding of particular databases to specific (and perishable) combinations of software and hardware. Secondly, there is real doubt as to the efficacy of any database system for description of materials as elaborately and multifariously structured as manuscript descriptions. By the time of the inception of the EAMMS and other projects (from 1996 on) an alternative to a database implementation of manuscript descriptions had emerged: the application of Standard Generalized Markup Language (SGML) encoding to the task. The work of the Text Encoding Initiative (TEI) since 1988 had created a set of guidelines for the encoding of a very wide range of humanities materials. These guidelines have become widely used in the scholarly community, and were instrumental in the explosion of electronic texts which is reshaping many areas of scholarly work.

Something like a consensus developed, among the scholars, archivists and computer experts involved in the EAMMS, Digital Scriptorium and MASTER projects, that SGML encoding might offer a more promising basis for the evolution of a general purpose computer-readable manuscript description system than database technology alone. Such an approach would add the new digital virtues of searchability and access to the strengths of the traditional, deeply-considered scholarly description. This would not mean discarding of database systems entirely: there are too many records invested in such systems, and some of the systems themselves have such virtue, for this to be desirable or possible. Accordingly, all these projects have a database component, and EAMMS (which aimed from the beginning at both database and SGML approaches) has been particularly successful in devising a MARC-based format for manuscript descriptions. However, at each end of the scale the possibilities of SGML encoding recommend themselves. At one end of the scale: SGML encoding of very simply structured records accompanied by manuscript images would lend itself to immediate web mounting. At the other end: the most complex manuscript descriptions could be housed in SGML encoding, without distortion of the description and with no compromise of scholarly detail. Similarly, the development of a universal interchange format in SGML might permit decanting of the many manuscript records from the many databases into a single, and therefore more malleable, form.

Several further factors contributed to the choice of the TEI implementation of SGML, as the base for the proposals here outlined. Firstly, there was the perceived success of the TEI guidelines in addressing many of the issues which would have to be resolved in any scheme of computer readable manuscript descriptions. Such a scheme would need to establish encoding for transcriptions of text from the manuscripts: the TEI had already set forward guidelines for transcription of primary textual material, and these guidelines had been well tested. A vocabulary would be needed for encoding of various kinds of textual division, for encoding of metadata about the description, and more: the TEI had already done much of this work. Secondly, the TEI workgroup system offered a well-proven means of bringing together domain experts and encoding experts, and harnessing their various skills towards a consensus which could form the basis of a widely-accepted standard. Thirdly, several of the people active in framing the TEI guidelines had strong interests in medieval manuscripts, and so were involved in the projects named above from their beginning.

The Studley Priory meeting which effectively inaugurated MASTER was followed by meetings of the EAMMS group at Hill Monastic Manuscript Library in December, 1996, and a year later, in November 1997, by a meeting at Columbia University which brought together many of the participants in the EAMMS, Digital Scriptorium and MASTER projects with other manuscript experts. Following this meeting, the TEI workgroup was established, with Consuelo Dutschke and Ambrogio Piazzoni as co-chairs. This workgroup first met in July 1998. The guidelines here described are the result of this meeting and of a series of further meetings involving various members of EAMMS, MASTER and the TEI workgroups: held most recently in New York in January 1999; in Paris in February 1999; in Rome in March 1999; and in Oxford in June 1999. The first draft of the dtd describved here was prepared following the Rome meeting and was thus available for consideration in May 1999 by the MASTER expert group.

The MASTER Experts Group brings together a small group of distinguished manuscript scholars to provide an independent evaluation of the project's proposals. Its members are: Dr Ian Doyle, Durham University Library (chair); Professor Peter Gumbert, Leiden; and Dr Gilbert Ouy, Paris. Following this expert group meeting, a green paper describing the draft and summarizing the discussions concerning key points within it, was posted to the web and semi-public comment invited through the members of the EAMMS discussion group (which includes all members of the TEI and MASTER projects). Following this dicussion, in June 1999 a technical subcommittee of the MASTER group met in Oxford and recommended further modification of this draft. It is this modified draft which is described in this paper. We confidently expect that publication of this draft and its associated documentation will provoke further discussion and redrafting. It should be emphasized that what is here presented has not yet been formally adopted by any of the working groups concerned, and should be understood only as the current recommendations of the authors.

3. Design principles

Standards are intended to make it easier to combine sets of data produced by different scholars in different places using different tools in different environments. Combination of such data sets is desirable when (as is the case with manuscript descriptions) they describe resources which are widely scattered and yet of global interest. Standards make such unification feasible at many levels: at one level, a standard such as that defining the electrical systems of Europe make it feasible for a computer to be transported from one place to another; at another level, a standard such as that defining the operating systems used by different computers makes it possible to share software. The level of standardization which concerns us, however, is higher still and concerns the definition of an abstract model underlying different data representations held within a computer. Provided that we have a clear agreement about the components of this abstract data model, we can expect different software systems to behave identically on different computer systems in so far as they process data in terms of that model. This is surely a better approach to standardization than the alternative, which is to insist that everyone should use identical tools and identical systems.

But how generic a model should we attempt to define? The TEI approach has always tended to adopt a broad church approach, facilitating the creation of data sets in which almost any combination of components is legal, from the most simple to the most complex. This seems particularly appropriate for manuscript descriptions, where each country, each intellectual domain, even each cataloguer for each different cataloguing project, seems to have developed a distinct cataloguing style. On the other hand, without a more narrowly defined set of rules, the task of integrating records produced to differing aspects of a very permissive standard may prove difficult or impossible.

For flexibility: this draft permits, on the one hand, the making of simple manuscript inventories which contain no more than a list of manuscript identifiers, to which may be added just a few words of description of each manuscript, or an image of the manuscript. On the other hand, the draft also aim to enable highly formalized descriptions, with elaborate structural mark-up distinguishing the various components of the description, possibly also containing complete manuscript transcriptions or digital facsimiles. The distinction is not just one of length. In this model, a lengthy description might be just a manuscript identifier accompanied by a series of unstructured paragraphs (that is: plain paragraph elements with no encoding within them discriminating their content); or a short description might be highly structured.

This flexibility is also a practical response to a recognition of the fact that, in the real world, requirements constantly change. Perhaps, in an ideal world, one might begin by making a simple inventory, with minimal details, and elaborate this by progressive addition of information. But more realistically, financial constraints may make it impossible ever to progress beyond that initial minimal inventory. Or one might be in the contrary position of having already made a substantial investment of expertise in the creation of legacy data which cannot immediately be mapped to the structures defined by the model. Here the clear answer will be to begin by importing information from the existing set of descriptions simply as unstructured prose, whether it is held in printed or digital form, or in some relational database format, so that subsequent work can take place to distinguish statements of date, place, provenance, or description.

Accordingly, this draft does not offer rigid definitions of what might constitute a short or first-level record, as against a long or full record. What is short for one purpose, may be too long for another; what is long for one purpose may be the bare minimum for another. Moreover, although we have defined a minimal level summary element, we have done so recognizing fully that its use may be entirely inappropriate in some situations. Some discussion of the thinking that lead to this apparently contradictory design decision may perhaps be helpful.

At an early stage of the design process, it was felt that a key advantage of using SGML was the ability it gave to integrate full text and database systems, by tagging arbitrary stretches of text as if they were database records. As with other seductive notions, it soon became apparent that there is a danger in this over simplification. Should we, following this principle, define a structure for our manuscript description in which almost any component could appear almost anywhere? Or should we aim for a more formal prescription, which declares: if you want to say something about the binding of the manuscript (for example) you can only say it in a `<binding>` element

within a <physDesc> element. The first approach offers the cataloguer a heady sense of near complete freedom; its danger is that the resulting descriptions are likely to be so heterogeneous as to be practically useless for purposes of integration and efficient retrieval — which was the prime justification for making the computer-readable records in the first place. The disadvantage of the second approach is that an over-rigid formalism would lead to frustration among cataloguers, and (rather quickly) refusal to adopt the standard.

Our solution was to attempt to have the best of both worlds by adopting what might be regarded as a truly British compromise. Our DTD permits the cataloguer to use either simple paragraphs of prose, which may contain anything humanly comprehensible, or to group such descriptions under more formal and precisely-defined elements for distinct and identifiable manuscript phenomena. Thus: a cataloguer is free to speak of any aspect of the manuscript binding, as it bears upon the history or intellectual content of the manuscript, within the <p> elements provided within the <history> and <msContents> elements. But if a formally-structured statement about the binding itself is required it should be located within the <binding> element provided within the <physDesc> element.

The alternative, to permit the encoder to use the <binding> element at any point, would likely lead to cataloguers feeling obliged to surround every reference to binding with the <binding> tag, regardless of context and content. This would encourage a superfluity of effort which would lead to so many different kinds of information being contained in <binding> as to render the element semantically ill-defined and therefore useless.

The danger of an open standard, such as this is designed to be, is that it may be misused. We do not encourage, for example, manuscript descriptions which consist of no more than an identifier and a lengthy prose description with no formal distinctions through markup of statements of date, origin, provenance, and the like. However, the standard itself cannot be used to prescribe that descriptions must conform to this or that model: it can only be used to enable the various models. These proposals seek to create a framework which can accommodate what we know of existing standards, and to enable (in time) greater precision and more efficient retrieval of detail in cataloguing. It is the task of various domain experts to recommend how the standard should be applied in their area. The proposers will begin this process, in the documentation which here follows and which will accompany the completed standard, but it will be the responsibility of you, the users of the standard, to further best practice.

Finally, because the background and interests of all the participants in the primary groups involved in preparation of these proposals (EAMMS, Digital Scriptorium, MASTER, the TEI workgroup) are in western European medieval manuscripts, these proposals have been prepared only with those manuscripts in mind. However, it is intended that these should in time be generalizable, at some level or other, to other manuscripts from different cultures, and we welcome reaction from experts in other manuscript areas.

4. A brief summary of the MASTER proposals

One of the consequences of the decision to define the MASTER encoding scheme as an application of the TEI document type definition is the availability of a number of standard TEI elements; another, perhaps less convenient, is that an awareness of the scope and idioms of the TEI scheme is necessary to fully understand the functionality of the proposed scheme. We do not attempt to provide such background here: the interested reader should consult the TEI web site for pointers to such information; perhaps of particular interest to francophone readers is the special edition of Cahiers Gutenberg published in 1986, which provides an excellent introduction.

An <msDescription> element may reasonably appear either within the body or within the header of a TEI conformant document. In the former case, where the document being encoded is essentially a collection of manuscript descriptions, the <msDescription> element may be placed anywhere that a paragraph might appear. In the latter case, where the description forms part of the metadata to be associated with a digital representation of some manuscript original, whether as a transcription, as a collection of digital images, or as some combination of the two, then the <msDescription> should appear within the <sourceDesc> element of the header.

The <msDescription> element may contain up to eight of the following different components, of which only the first is mandatory:

- a <msIdentifier> element, which groups all the information necessary to uniquely identify the manuscript being described within the repository where it is held;
- an optional <msHeading>element, which provides a supplied title or brief heading for the manuscript being described;
- an optional <msSummary> element, which provides minimal cataloguing information about the manuscript in a specific structured format; may be used as an alternative to more detailed cataloguing;
- an optional <msContents> element, providing an itemized list of the intellectual contents of a manuscript or manuscript part, with bibliographic descriptions etc. as appropriate;

- an optional <physDesc> element, which groups information about all physical aspects of a manuscript or manuscript part (e.g. its format and material, scripts used, foliation and paratext, binding etc.);
- an optional <history> element, which summarizes the known history of a manuscript or manuscript part, within which its origin and provenance are distinguished;
- an optional <additional> element, which groups other related information about a manuscript, in particular, administrative information relating to its current location, additional materials associated with it, details of any surrogates held etc.;
- one or more optional <msPart> elements, each of which contains a similarly structured description for a manuscript fragment which is now regarded as forming part of the manuscript being described, although originally physically distinct.

Status and Identification

In addition to the global attributes available to all TEI elements, the <msDescription> element carries a special purpose STATUS attribute which defines the compositional status of the manuscript being described , i.e. whether it is a complete unitary object, a composite of fragments, or an isolated group of fragments. This attribute is intended solely as an indication of the overall status of the manuscript being described; details of its composition are documented in various other ways, more fully described elsewhere. For example, as noted abobve, a composite manuscript, in which several originally distinct complete or fragmentary manuscripts have been physically combined, may be described by a manuscript description whose internal structure mirrors that of the manuscript, with distinct <msPart> elements for each of the identifiable fragments.

The <msIdentifier> element is used, as the name suggests, to supply the identifier used for a manuscript within the holding institution. As a minimum, the identifier should supply the name of the repository and an identifying number such as a shelfmark, as in this example:

```
<msIdentifier>
<repository>Huntington Library</repository>
<idNo>El.  26 C. 9</idNo>
</msIdentifier>
```

Further elements may additionally be supplied to specify the country, region etc. of the repository, and also to supply any alternative names or former shelfmarks associated with the manuscript in this repository, as in the following (fictitious) example:

```
<msIdentifier
<country>UK</country>
<region>Oxfordshire</region>
<settlement>Oxford</settlement>
<institution>The University of Oxford</institution>
<repository>Bodleian Library</repository>
<collection>Rawlinson</collection>
<idNo type="BOD">Rawl.  MS poet.  176</idNo>
<altName type="nick">The Tungebryht manuscript</altName>
</msIdentifier>
```

As both the above examples demonstrate, we have not yet agreed on a way of normalizing such features as the names of institutions or places by the use of authority lists or similar features. We anticipate that this will be an important further stage during the development of the standard, for which, fortunately, a range of appropriate mechanisms already exist within the TEI scheme.

The <msHeading> element is provided to make life easier for the cataloguer who wishes to use a standardised or summary name for a manuscript, which might contain some elements derived from both its self number and its intellectual content (for example) but not all of them, perhaps including additional information not strictly present in either. It is an optional element, which does not strictly form part of the <msIdentifier>, since it is a form of supplied title for the whole description.

The Manuscript Summary

Manuscript cataloguing practices vary widely, not only in the scope of the information included, but also in the amount of detail which it is feasible or desirable to record for each item. Particularly when handling legacy data, there is a frequent demand for some kind of minimal level record, whether as an end in itself or as an initial stage in the creation of a more complete record.

The MASTER dtd makes it possible for cataloguers to select fairly freely from a very wide range of detailed cataloguing possibilities, which can be tailored to specific project needs, and moreover later be expanded as appropriate. The DTD described here is designed precisely to facilitate such incremental enrichment. However, there is also a frequently voiced desire for some specific recommendation concerning the minimum practical level of detail to be recorded. The <msSummary> element is provided to meet this need, and also that for rapid conversion of legacy data. It should not be used simply to hold short summary title, 'docket', or 'tombstone' specifying a supplied title or heading applicable to the whole of a manuscript: the <msHeading> element is provided for this purpose. Moreover, it should not be used where more detail is required than can be accomodated by the <msSummary> element; indeed, good practice may suggest the removal of this element from the record when more detailed information has been included in the remainder of the manuscript description.

The <msSummary> element is unlike other components of the MASTER manuscript description in that its contents are relatively constrained. It is rather more like a traditional database record than a piece of free prose in which certain words happen to be tagged. It allows the cataloguer to specify and to mark unambigiuously information about authorship, origin (place and date), titles, and languages used within a manuscript, but little else. These components are signalled by using a mixture of pre-defined TEI elements (such as <author>, <langUsage>, and <title>) and newly-defined specialist elements (such as <origPlace> and <origDate>). The constituents of a <msSummary> must be supplied in a prescribed order, and only the <author>, <title> and <note> elements may be repeated.

Here is a simple example of the use of this element:

```
<msSummary>
<author>Domenico Cavalca</author>
<title>Vite dei santi padri</title>
<origPlace>trans.  Naples</origPlace>
<origDate>1474</origDate>
</msSummary>
```

As previously noted, we recognize the need for additional mechanisms such as authority lists and controlled thesauri to facilitate standardization of the content of these elements; these will probably involve the definition of additional attributes to hold normalized forms, or links to other components in a knowledge base. At this stage in the project, however, our focus is on defining an appropriate structure, adequate to the needs we have so far identified as common to all projects. It may well be that the elements described so far will be all that some cataloguing projects ever find necessary or feasible to supply for an initial cataloguing exercise. However, we also anticipate a need for far more detailed cataloguing, using one or more of the remaining components of the <msDescription> element.

Intellectual Content

A <msSummary> will typically supply only a single main author and possibly a supplied title. In the common case that a manuscript contains many distinct works, or parts of works, a more complex description is likely to be desirable. Rather than try to complicate the structure of the <msSummary> element, we propose the use of a specialized <msContents> element in which the structure and organmization of the intellectual content of a manuscript can be faithfully and acciurately recorded to whatever degree of complexity is appropriate.

A <msContents> element consists of one or more <msItem> elements, each of which describes a single distinct "item" of intellectuyal content, as determined by the cataloguer. It is a matter for individual cataloguing practice to decide whether, for example, each poem in a miscellany of poems, each life in a collection of saint's lives, or each charter in a cartulary, should be regarded as discrete items. The purpose of the tagging scheme we describe here is to provide a method by which the results of that decision can be communicated unambiguously, not to provide guidance on how it should be made, except insofar as that is implied by the definition we propose for distinct <msItem> elements.

A <msItem> element may simply contain running text with no further tagging; more usually, however, it will contain additional identifiable components. Many of these are standard TEI elements: these are <author>, <respStmt>, <title>, <langUsage>, <q> (for quotations), and <bibl> (for conventional bibliographic description, for example of a modern edition). These may however be combined with any appropriate combination of the following more specialized and manuscript-specific elements: <colophon>, <incipit>, <explicit>, <rubric>, and <finalRubric>.

Finally, a special purpose <locus> element is provided to specify the location in the manuscript occupied by the element within which it appears. Locations are conventionally specified as a sequence of folio numbers, but may also be a discontinuous list, or a combination of the two. This specification should be given as the content of the <locus> element. It may also be specified in a normalised form using special purpose attributes on the <locus> element. To avoid ambiguity, a <locus> element should be the first component element of the item whose location is being specified, and it should not normally be repeated within that element.

Each element within <msItem> has the same substructure, in which (following an optional <locus> element) any combination of low-level TEI phrase elements and plain text may appear.

Here is a simple example, for a manuscript containing a single item which occupies folios 1 to 223 of a manuscript, with an incipit containing the words Forte Hervei monachi on folio 1, and an explicit on the verso of f 223 which reads Benedictio salis et aquae:

```
<msContents>
<msItem><locus>f.  1-223</locus>
<author>Radulphus Flaviacensis</author>
<title>Expositio super Leviticum </title>
cf.  <bibl>Stegm&uuml;ller, RB 7093</bibl>
<incipit><locus>f.  1</locus>
Forte Hervei monachi</incipit>
<explicit><locus>f.  223v</locus>
Benedictio salis et aquae</explicit>
</msItem>
</msContents>
```

Here is a complex and perhaps more typical example:

```
<msContents>
<msItem n="1"><title><locus>fols.  5r -7v</locus>An ABC</title>:
(<ref>IMEV 239</ref>)</msItem>
<msItem n="2"><title lang="FRA"><locus>fols.  7v -8v</locus>
Lenvoy de Chaucer a Scogan</title>:  (<ref>IMEV 3747</ref>)</msItem>
<msItem n="3"><title><locus>fol.  8v</locus>Truth</title>:
(<ref>IMEV 809</ref>)</msItem>
<msItem n="4"><title><locus>fols.  8v-10v</locus>
Birds Praise of Love</title>:  (<ref>IMEV 1506</ref>)</msItem>
<msItem n="5">
<title type="supp"><locus>fols.  10v -11v</locus>Two Latin poems</title>
<msItem><title lang="LAT">De amico ad amica</title><ref>IMEV 16</ref></msItem>
<msItem><title lang="LAT">Responcio</title><ref>IMEV 19</ref></msItem>
</msItem>
<msItem n="6"><title><locus>fols.  14r-126v</locus>
Troilus and Criseyde</title>
(Bk.  1:71-Bk.  5:1701, with additional losses due to
mutilation throughout)</msItem>
</msContents>
```

As shown in the fifth item above, a manuscript item may itself contain further nested manuscript items, for example where a title is supplied for a group of works each of which is also titled. More complex situations where groups of items are nested arbitrarily deep are also imaginable.

Physical Description

Of the three remaining chief groups of the standard manuscript description, that concerned with its physical description is probably the most complex, and also that in which cataloguing practices tends to be most divergent. Recognising this, the MASTER dtd does not currently impose a strict model on the information to be collected under this heading; instead, it allows for cataloguers simply to describe what information they wish as regular prose, optionally grouping one or more paragraphs under specific categories, for which specific elements may be used.

The categories for which elements have so far been identified and which are therefore to be distinguished within a full description include the format (e.g. codex, scroll, etc.), characteristics of the support (e.g. its material) and of the layout (e.g. number of columns), and scripts used. Element definitions are also provided for information relating solely to the binding of a manuscript, its collation and other paratextual features. Finally, and perhaps more controversially, discussion of decorative features of the manuscript is currently classed under physical description, as is discussion of any music contained within it.

The majority of the elements making up a physical description will consist of plain prose descriptions, expressed as one or more paragraphs. Indeed, a <physDesc> element may contain nothing but a series of paragraphs: this would be essential in the case where the source description intermingles discussion of (say) collation, decorative features, and binding indiscriminately. More usually, discussion of each of the topics listed above will

be confined to one or more distinct paragraphs, in which case the encoder has the option to use the more specific elements (<binding>, <music>, <collation> etc.) proposed by the MASTER dtd to distinguish them. In a few cases (notably collation) a similar principle applies at a further level; that is, a <collation> element may contain either simply a series of paragraphs, or it may contain one or more of the more specialized elements <catchwords> and <signatures>, each of which contains paragraphs relating specifically to catchwords and to signatures respectively.

A particular strength of this approach is that the TEI <p> element permits of the identification of a wide range of useful objects other than simple running prose within it. These include tags for dates, numbers, titles, editorial corrections, typographic highlighting, foreign language phrases, etc. and mathematical (or other) formulae. Thus, even a <collation> element containing <p> elements only can also contain a <formula> element in which a detailed collation formula may be supplied using an appropriate notation.

This richness of the underlying "soup" from which the components of a description are made up is one important characteristic that the MASTER dtd inherits from the TEI. Another is its tendency to apply Occam's razor, In the case of decoration, for example, where other systems have tended to propose very specific (but widely differing) sets of distinctions for different classes of decorative feature, the MASTER dtd instead proposes a single category of decorative note, (<decoNote>) optionally further categorized by means of type and subclass attributes.

Historical and Curatorial Information

The remaining parts of the MASTER manuscript description are concerned with the history and curation of the manuscript as an artefact. A <history> element is used to document firstly the origin, next the provenance, and finally the acquisition of the object concerned. Attributes may be used to associate the prose descriptions within these component elements with more directly searchable normalized information such as exact dates, whether or not the evidence for a date is internal to the object, and its reliability.

The history of a manuscript should normally be presented in the order implied above. Information about the origins of the element (including any discussion of its sources) should be given as one or more paragraphs contained by a single <origin> element; any available information or discussion of distinct stages in the history of the manuscript before its arrival in its current location should be included as paragraphs within one or more <provenance> elements following this. Finally, any information specific to the means by which the manuscript was acquired by its present owners should be given as paragraphs within the <acquisition> element.

A variety of information relating to the curation and management of a manuscript may be recorded as simple prose narrative tagged using the standard <p> element, optionally grouped within one or more of the specific elements <recordHist> (record history) <custHist> (custodial history) and <availability> (a standard TEI element). Together these make up a group of elements referred to in the current DTD as <adminInfo>, as a gesture to the comparable element within the Extended Archival Description (EAD) dtd.

The <recordHist> element is provided as a means of documenting the history of the cataloguing record itself. If supplied, it contains a <source> element, followed by an optional series of <change> elements. The latter are standard TEI elements, which may also appear within the <revisionDesc> element of the standard TEI Header; their use here is intended to signal the similarity of function between the two container elements. Where the TEI Header should be used to document the revision history of the whole electronic file to which it is prefixed, the <recordHist> element may be used to document changes at a lower level, relating to the individual record.

The <source> element is used to document the primary source of information for the catalogue record containing it, in a similar way to the standard TEI <sourceDesc> element within a TEI Header. If the record is a new one, it may simply contain a <p> element as in the following example:

```
<source><p>No source:  this is an original record</p></source>
```

More usually however the record will be derived from some previously existing catalogue, which may be specified using the standard TEI <bibl> element, as in the following example:

```
<source><p>Information transcribed from
<bibl><title>IMEV</title><biblScope>1234</biblScope></bibl>
</p></source>
```

If, as is likely, a full bibliographic description of the source from which cataloguing information was taken has already been given elsewhere in the manuscript description (for example in a <listBibl> element), then it need not be repeated here. Instead, it should be referenced using the standard TEI <ref> element, as in the following example:

```
<additional>
```

```
<listBibl>
<bibl id="IMEV123">
<title>Index of Medieval Verse</title>
<!- other bibliographic details for IMEV here ->
<biblScope>123</biblScope>
</bibl>
<!- other bibliographic records relating to this ms here ->
</listBibl>
<recordHist>
<source><p>Information transcribed from
<ref target="IMEV123">IMEV 123</ref>
</p></source>
<!- ...  ->
</recordHist>
```

The <custHist> record is used to describe the custodial history of a manuscript, recording any significant events noted during the period that it has been located within the cataloguing institution. It may contain either a series of paragraphs tagged with the standard TEI <p> element, or a series of paired <date> and <custEvent> elements, each describing a distinct incident or event, further specified by a TYPE attribute.

In the following example, the cataloguer has chosen to record the key events in a manuscript's custodial history simply as a series of paragraphs:

```
<custHist>
<p>Conserved between March 1961 and February 1963 at Birgitte Dalls
Konserveringsv&aelig;rksted.</p>
<p>Photographed in May 1988 by AMI/FA.</p>
<p>Dispatched to Iceland on 13 Nov 1989.</p>
</custHist>
```

The same history might alternatively be represented in a slightly more structured and searchable form by using typed <custEvent> elements, as follows:

```
<custHist>
<date notBefore="1961-03" notAfter="1963-02"></date>
<custEvent type="conservation">
<p>Conserved between March 1961 and February 1963 at Birgitte Dalls
Konserveringsv&aelig;rksted.</p></custEvent>
<custEvent type="photography">
<date notBefore="1988-05-01" notAfter="1988-05-30">May 1988</date>
<p>Photographed in May 1988 by AMI/FA.</p></custEvent>
<custEvent type="transfer/dispatch">
<date value="1989-11-13">13 November 1989</date>
<p>Dispatched to Iceland.</p></custEvent>
</custHist>
```

Other materials

A need for information under three further headings has so far been identified to complete the manuscript description: firstly, a traditional bibliography of other works describing the manuscript in hand; secondly, information about any surrogates (copies or photographs etc.) of the manuscript, and finally information about any materials now accompanying the manuscript but not forming part of it or its binding.

The first need is simply met by the standard TEI <listBibl> element, which being fully described in the TEI reference given above we do not describe further here.

The second need is met by a special purpose <surrogates> element which enables cataloguers to provide information about any digital or photographic representations of the manuscript which may exist within the holding institution or elsewhere. Where such representations exist within published works, they will normally be documented within the <listBibl> element already mentioned. However, it is often also convenient to record information such as negative numbers, digital identifiers etc. for unpublished collections of manuscript images maintained within the holding institution, as well as to provide more detailed descriptive information about the surrogate itself.

At a later version, the content of the <surrogates> element is likely to be expanded to include elements

more specifically intended to provide detailed information such as technical details of the process by which a digital or photographic image was made. At present, this and other information may be recorded simply as prose paragraphs.

Turning to the third need identified above, where a manuscript has additional material, not originally part of the manuscript, which is bound with it or otherwise accompanying the manuscript. In cases where this additional material is clearly a distinct manuscript or manuscript fragment, the whole manuscript should be treated as a composite manuscript and the additional matter described in a separate `<msPart>`. However, there are cases where the additional matter is not self-evidently a distinct manuscript: it might be an important set of notes by a later scholar or owner, or it might be a file of correspondence relating to the manuscript. The `<accMat>` element is provided as a holder for this kind of information, as in the following example, describing a note by the Icelandic manuscript collector Árni Magnússon which has been bound with the manuscript:

```
<accMat>
<p>A slip in &Aacute;rni Magn&uacutee;sson's hand has been stuck to the
pastedown on the inside front cover; the text reads:<q>&THORN;idreks
S&oslash;gu &thorn;essa hefi eg feiged af Sekreterer Wielandt Anno 171
5 i Kaupmanna h&oslash;fn.  Hun er, sem eg sie, Copia af Austfirda
b&oacute;kinni (Eidag&aacute;s) en<expan>n</expan> ecki progenies
Br&aelig;dratungu bokarinnar.  Og er &thorn;ar fyrer eigi i
allan<expan>n</expan> m&aacute;ta samhlioda
&thorn;<expan>eir</expan>re er Sr Jon Erlendz son hefer ritad fyrer
Mag.  Bryniolf.  &THORN;esse &THORN;idreks Saga mun vera komin fra Sr
Vigfuse &aacute; Helgafelle.</q></p>
</accMat>
```

5. The next stage

Standards are defined by their users. While we have done our best to obtain input from as wide a community of domain experts as possible, we do not expect this standard to provoke more than passing interest until it has been put to the test of actual cataloguing practice, For this reason a key concept of the MASTER project plan was the requirement that the standard developed during its initial pghases should be properly validated. To that end, each of the MASTER partner libraries is funded to test the MASTER standard by making at least three hundred shorter manuscript descriptions, and fifty longer descriptions, during the life of the project. Informally, all the library partners have indicated that they intend to make many more records than this. To facilitate this process,m the project will also develop specialized software for input of the descriptions, and provide training for the cataloguers who will use it. As descriptions are created, they will be mounted on the web in a prototype on-line union catalogue to test their consistency and the extent to which the goal of an integrated set of descriptions has been achieved.

The first phase of this implementation is to commence in October 1999, following on the preparation of the draft here described and the development of software for record input. This first phase will finish in February 2000, and will be followed by revision of the standard on the basis of the partners' experiences, revision of the associated software, and a second manuscript entry phase. During the final months of the project, we plan a series of dissemination workshops where each partner will introduce the standard and the systems developed in the project to other libraries and archives.

Further information about the project is available from its website at http://www.cta.dmu.ac.uk/projects/master/ or from the authors.