DOMLib / Ethio-SPARE Manuscript Cataloguing Database

Presentation at the International Workshop "The electronic revolution? The impact of the digital on cataloguing", Copenhagen, 7-8 June 2012 (http://www1.uni-hamburg.de/COMST/meet4-3.html)

by Dr. Denis Nosnitsin, Hamburg University

When the ERC Independent Researcher Starting Grant for the Project *Ethio-SPARE: Cultural Heritage of Christian Ethiopia, Salvation, Preservation and Research* started in December 2009 it was immediately clear that some basic decisions had to be taken as far as the description of manuscripts is concerned. While during the application phase, a quite minimal data model was suggested (with very few fields – limited to Signature; Short title; Material; Measurements; Size in folios; Dating; Scribe; Author; Donor; Location original; Location current; Short contents; Incipit; Illuminations; Additiones; Further details; Bibliography), it was soon clear that such basic set would not do justice to a true catalogue description and the model would not suffice.

The question was, then, whether to use a database for cataloguing at all, or to recur to full-text cataloguing mode, either traditional, or with XML data structuring. The decision went quite soon against the traditional paper mode, as only digital manuscript cataloguing allows such necessary features as quick statistical analysis, interlinking of different data subsets, incorporation of additional layers of information, accommodation of manuscript images and, last but not least, wide accessibility of research results.

While an XML manuscript description offers much more freedom and flexibility to a scholar, with TEI currently having developed, in its <msdesc> subset, possibilities to properly encode any relevant information, this way of working requires a relatively high degree of technical preparation from all scholars involved, and, more importantly, a high degree of research independence. Without a pre-set mask that "hints" as to what features must necessarily be present in a description it may be difficult for everyone to maintain a comparable standard. This is especially important within the framework of a project where only the principal investigator has had a long-time intensive exposure to the cataloguing work and at the same time is the one who bears the responsibility for the overall quality of the project outcome – whereas other project researchers come in without much background knowledge and have first to become familiar with the process and be constantly trained and controlled.

Therefore the decision was taken in favour of the database approach, with a much extended data model. The start of the Ethio-SPARE coincided with the beginning of the COMSt network, which allowed a thorough survey of ongoing initiatives and the various approaches.

The database was then programmed on the basis of the open source MyCore (www.mycore.de) online database software. Initially programmed for UNIX machines, this database software can be now installed and developed also on MS Windows and MAC, which makes it widely useable. The fact that it is an open source software allows limitless development possibilities – and, in fact, MyCore is now being actively developed for various database applications at a series of German universities (http://www.mycore.de/list.html). Some of these applications are indeed meant for manuscript cataloguing. Already before the Ethio-SPARE, a manuscript cataloguing MyCore application was

developed in Leipzig and used for the Islamic manuscripts cataloguing project and the Refaiya library project (http://www.refaiya.uni-leipzig.de) that were presented during the COMSt launch conference in December 2009. Since these projects were sponsored by the DFG (German Research Association), the data model they used corresponded to the basic DFG requirements for manuscript cataloguing, in particular those applied to Oriental manuscripts, also within the KOHD series (Union Catalogue of Oriental Manuscripts in German Collections). When the KOHD has started to move towards online manuscript cataloguing it has thus been possible for a few initiatives to employ the same model – for example, the Turfan manuscript cataloguing (http://turfan-digital.staatsbibliothek-berlin.de/content/below/index.xml) also discussed today.

The Digital Oriental Manuscript Library (DOMLib) is an online application developed for cataloguing of and further research into the cultural heritage of the Orient (manuscripts, but also other items of scholarly and artistic interest as well as related archive sites). Its main focus is currently the manuscript tradition of Christian Ethiopia, as the catalogue was started within the framework of the European Research Council-funded project (Starting Grant 2009-14) Ethio-SPARE: Cultural Heritage of Christian Ethiopia.

The Bibliography part of the database includes, in addition to the titles used by the Ethio-SPARE project catalogue, the complete set of titles collected in the course of the European Science Foundation Research Networking Programme (2009-14) Comparative Oriental Manuscript Studies .The bibliography is expected to be extended after the completion of the COMST programme. DOMLib is a database under construction and changes are naturally expected to occur.
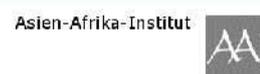
Those models, however, fail to accommodate many of the characteristics that a researcher may consider relevant in a catalogue, in particular in a digital catalogue. This was already partly noted by the members

of the Refaiya project in Leipzig, who added additional data models to provide details on "secondary entries" (additiones, colophons), bookbinding and historical documents.

The challenge of defining the fields of a digital catalogue is of course that of foreseeing from the very beginning all the details one may later want to find, in particular, all the details that one may want to filter the search accordingly. Needless to say, it makes a considerable difference for search and statistical queries if the manuscript binding is described all in one field, or sub-structured into Binding Material, Presence of Endbands, Number of Sewing stations, Number of restorations, etc.

The main data models developed by the Ethio-SPARE project include Manuscript Description, Art Item Description, Archive/Institution, and Bibliography; a data model for Persons is in development. Naturally, for cataloguing purposes the Manuscript Description model is of central importance, the other interlinked models are rather auxiliary for the moment, even if, with the help of the database, one can also get some research help when looking for the places manuscript collections were found.

The Manuscript Description cataloguing form is subdivided into General description, Physical description, Description of content, Colophon, Additiones, and Decoration, with over 100 meta data fields all in all.

The general description includes: (1) call number (necessary), (2) registered by, (3) registration date, (4) language (category selection), (5) writing system (category selection), (6) short title (necessary), (7) short content, (8) subject matter (category selection), (9) origin (link to institutions data model), (10) current location (link to institutions data model), (11) copyist/scribe, (11b) comments, (12) donor, (13) owner (fields 11-13 will become links to persons data model), (14) dating, (14b) comments, (15) state of preservation, (15b) comments, (16-26) yes/now fields: dated / palimpsest / composite / fragment / illuminated / additiones / autograph / printed part / incomplete / colophon / restoration (these fields enable quick filtering according to the set parameters, such as a chance to quickly find all dated unrestored manuscripts, or all composite manuscripts, etc.), (27) number of texts, (28) number of production units, (29) literature on the manuscript. Another category, on Manuscript form (codex/roll/leporello/others), is currently missing and will be soon added (all manuscripts we have been dealing with so far are codices).

The physical description includes: (1) binding (general), (2) binding material (category), (3) binding decoration, (4) binding originality (yes/no/partly), (5) spine binding (yes/no), (6) state (damaged/intact), (6b) details if damaged, (7) endband (yes/no), with possibility to specify in 7b and 7c whether headband or tailband are missing, (8) other information on binding, (9) writing material (category selection), (10) watermarks (yes/no + descriptive field), (11) ink, (12) number of folios, (13) blank folios, (14) number of quires, (15) structure of quires, (16) outer dimensions (three fields: width, height, thickness), (17) text area (2 fields), (18) margins (6 fields: top, bottom, left, right, up to 3 intercolumnal spaces, free comments), (19) number of columns, (20) number of lines, (21) ruling, (22) palaeography: type of script, (23) handwriting, (24) change of hand, (25) punctuation, (26) crux ansata, (27) rubrication, (28) abbreviations, (29) other details.

The description of content provides a possibility to include, for each text in a manuscript, (1) text order number, (2) text title, (3) author, (4) completeness, (5) folio range, (6) similar manuscripts, (7) editions and translations (with a link with the bibliography data model or free), (8) literature on the text (connected with the bibliography data model or free), (9) further information, (10) incipit, (11) translation, (12) explicit, (13) translation, (14…) other important passages, (15) details. All the fields can be repeated if there is more than one text. Separate tabs have been created for accommodating colophons, with fields for (16) passage itself and its (17) translation and further (18) details and for additiones, with fields for (19) short content, (20) additio (quote), (21) translation, (22) details. Once again, fields 16-18 and 19-22 can be repeated when there are several colophons or additional notes. Finally, there is a possibility to list miniatures, drawings, ornaments, initials and other decorative elements (fields 23-28), once again, multiplying the fields in case there is more than one relevant element, and, finally, a possibility to assign (29) keywords.

While we are aware of the discussion whether passages and fragments belong into catalogues that are accompanied by complete sets of images there are several considerations against leaving them out: first and foremost, the images are not searchable for text, and if one wants to find some specific formulae, it is impossible to look for them unless they are quoted in full. Second, various access degrees are conceived, and one can choose for a group to have a level of access when they see the description but not the images. Third, if one chooses to use the online catalogue as a basis for a paper version one does need the fragments to be present.

The PI has defined the minimal set of fields to be filled, as providing full descriptions for all the 2000 manuscripts collected in the course of the project would not be even approximately possible within its 5-years running time.

Among the particular features currently available are the possibility of locate the manuscript collections on the map and do the geographic search for them; the interlinking between the archives and manuscripts data models (connecting the persons data model and the bibliography data model is foreseen). Works are being carried out on user-friendlier external design and search options. The database developers at Hamburg University are currently also working on installing the data export options that will allow to view individual descriptions or sets of descriptions as an XML (TEI compatible) file as well as rich formatted text and/or PDF that should assist much in preparing print catalogues for selected portions of the digital catalogue.

It is hoped that the present model can serve as the basis for a future and larger Digital Oriental Manuscript Library that may incorporate also traditions outside Ethiopia. Fields currently lacking but relevant for other areas may be added (e.g. a more detailed structure for watermarks, link to a watermark database or similar, for those working with paper manuscripts) and I will be most grateful for discussion and feedback.