COMSt - Team 4: III workshop, in cooperation with Team 3

Copenhagen, June, 7th-8th, 2012

Nordisk Forskningsinstitut, Building 27,
Københavns Universitet Amager, Njalsgade 136, 2300 København

The electronic revolution?
The impact of the digital on cataloguing

Scientific Coordination:
Team 4 : Paola Buzi, Witold Witakowski
Team 3 : Matthew Driscoll, Jost Gippert

Local Organizer:
Matthew Driscoll

Scientific Report

a) Scientific summary

On June 7th-8th 2012, the Nordisk Forskningsinstitut of Copenhagen University hosted the third workshop of the Cataloguing Team of the COMSt network (Team 4), which was realized with the determinant collaboration of the Digital approaches to manuscripts studies Team (Team 3).

The workshop, which was entitled: "The electronic revolution? The impact of the digital on cataloguing", had the following main purposes:

(1) to take into account various formats of electronic catalogues, evaluating the pros and cons of the different choices (*Session I*);

(2) to analyse the potentialities but also the limits or the problematic aspects of electronic catalogues – in all their formats –, such as the long time availability, the authorship, the interpretation and the potential chaos, the terminology etc., without forgetting the delicate matter of the relationship between printed catalogues and electronic catalogues (*Session II*);

(3) to show the basic ideas of hypercatalogues and portals that render access to distributed manuscript repositories and the like; a demonstration of existing hypercatalogues and a basic outline of the technical background was offered (*Session III*);

(4) to make an introduction to the TEI's manuscript description module and to the XML software (*Session IV*).

Following the fruitful practice of other COMSt meetings, the workshop was conceived mainly in the form of a round table. The first two sessions were coordinated by Team 4 (Paola Buzi), while the last two by Team 3 (Matthew Driscoll).

A visit to an exhibition of a selected group of manuscripts preserved in the Royal Library completed the two-days workshop.

b) Description of scientific content

As already mentioned in the "Scientific summary", the workshop was subdivided into four sessions dedicated to the analysis and evaluation of different kinds of electronic catalogues and to the technical, practical and philosophical choices which are behind them.

The final purpose of the workshop, as is obvious, was contributing to the development of the chapter of COMSt handbook which Team 4 will be responsible for (Chapter 3), but with the active collaboration of Team 3 as far as the paragraph dedicated to digital catalogues is concerned (and not only).

In order to stimulate an exchange of opinions among the speakers, the team members and the external collaborators, much time was dedicated to the discussion (above all in *Sessions II* and *IV*).

The following are the main results of the workshop:

Session I

The first session of the workshop was dedicated mainly to a survey of several ongoing projects of electronic catalogues. The invited speakers were all active collaborators of the projects taken into account, if not even founders and coordinators of them.

Each speaker was asked to expose briefly (15 minutes) the structure and the philosophy of his/her digital cataloguing project.

The communications have pointed out several crucial aspects that the construction of an electronic catalogue implies.

Eva Nyström and Patrik Granholm ("Greek Manuscripts in Sweden, a digitization and cataloguing project") explained the structure of their cataloguing project – hosted by Uppsala University Library and involving two full-time researchers plus the Library's own expertise in digital photography and electronic publishing – aiming to the cataloguing of the about 120 Greek manuscripts preserved in Swedish libraries, museums, and archives.

Before starting the project all the possible options have been evaluated: conventional printed catalogue, print-on-demand catalogue, electronic on-line catalogue, on DVD catalogue, etc. The final choice has tried to combine the diffent options as far as possible, because while an electronic catalogue can be improved and modified continuously, on the other hand a printed catalogue is indispensable when computers are not available.

The philosophy of Upssala cataloguing project takes into account above all the following aspects: (a) access; (b) time and cost efficiency; (c) updating and upgrading possibilities; (d) collaboration; (e) future endurance.

It was decided that the catalogue would be internet-published and hosted at the Uppsala University Library's digital platform. A printed (or a print-on-demand) version will also be provided. From a technical point of view, the data are tagged in XML according to the manuscript description module in TEI. Collaboration is considered a basical aspect: researchers are invited to comment on problematic parts and to suggest improvements.

Evyn Kropf ("Within the online library catalogue and beyond: Islamic Manuscripts at Michigan") described the grant-funded project of the University of Michigan Library aiming to complete the cataloguing of its Islamic Manuscripts Collection. It is interesting to stress that all the collaborators of the project have been trained before starting their work, in order to better accomplish their task; in particular a week-long workshop addressing Arabic manuscript studies (palaeography and codicology) and a multi-session workshop addressing Islamic bindings, book structures, materials and condition were organized for this purpose. The chief goal of the project is to compile a database – searchable and accessible to all via the Library's online catalogue and the HathiTrust Digital Library – of bibliographic and codicological data that may serve to further research and scholarship in Near Eastern studies and Islamicate manuscript studies in particular.

In the initial phase of the project, existing inventory data were converted to preliminary online catalogue records and digital surrogates were created for each manuscript. Next, descriptive elements are harvested from the preliminary records and combined to form a representative description posted on the project website along with a link to each manuscript's digital surrogate in the HathiTrust Digital Library.

Collaboration is again a crucial matter: colleagues from elsewhere around the world may also examine the digitized manuscripts, review existing descriptive information and submit their contributions to the descriptions as comments via the project website.

All contributions are then reviewed by the project cataloguer and refashioned for incorporation into the cataloguing records for those manuscripts, which serve as the final "published" descriptions.

Christophe Flüeler ("E-codices. Virtual Manuscript Library of Switzerland") presented the e-codices cataloguing project which aims to provide access to all medieval and selected early modern manuscripts of Switzerland via a virtual library, which will be continuously updated and extended. At the moment, the virtual library contains 894 manuscripts from 37 different libraries.

On the e-codices site, complete digital reproductions of the manuscripts are linked with corresponding scholarly descriptions. Scholarly manuscript descriptions encoded in a specialized XML format are employed as metadata for the digitized manuscripts presented on the website. The XML format used complies with the TEI guidelines for the electronic presentation of manuscript descriptions. Compliance with this international standard enables the exchange of metadata with other digitization projects, meta-search engines, and portal sites.

Each digitized manuscript is accompanied by the most relevant available scholarly description. These are selected by the participating libraries for use as "standard descriptions" on the site. Where multiple descriptions are available, the owner library chooses which should be displayed as the "standard description"; others are displayed as "additional descriptions". The idea is to make all published descriptions of each manuscript available on e-codices wherever possible.

The aim of the project is explicitly to serve not only manuscript researchers, but also the general public.

Anne-Beate Riecke ("Describing Manuscripts of the City Library in Treves with MXML for Manuscripta Mediaevalia") presented a cataloguing project regarding about 75.000 medieval western manuscripts held in German libraries, archives or museums.

Manuscripta Mediaevalia database is run by the State Libraries in Munich and Berlin and the Bildarchiv Foto. It includes a very heterogeneous variety of cataloguing descriptions. This is due to the fact, that its contents grew over decades since the 1970s and were captured with different software. Therefore, manuscript documents can show (a) collected index-entries extracted from the printed catalogues sponsored by the DFG from the 1960s to the early 2000s, (b) inventory-like catalogue-descriptions from the 1990s, that were never meant to be printed, but to collect and present online what is currently known about a manuscript, (c) inventory lists from the 2000s with a defined set of basic information on a manuscript, or (d) fully fledged descriptions, captured with a software introduced in 2007, called Manuscriptum XML (MXML) and developed especially to record the detailed specifications demanded by the DFG Guidelines for cataloguing manuscripts.

Due to this long period of data origin, there may be even more than one document for a single manuscript – covering older and more modern descriptions, descriptions focused on texts or on illuminations or – finally – on texts in different languages or parts of a manuscript. Additionally, most of the printed catalogues with relevant manuscript descriptions are digitized, presented in a special section of Manuscripta Mediaevalia and linked with the referring documents.

Currently descriptions generated in MXML form a minority among the documents in Manuscripta Mediaevalia, but they are gaining weight as the DFG will make the usage of MXML compulsory for all manuscript cataloguing projects sponsored by them.

Denis Nosnitsin's presentation ("Ethio-SPARE Manuscript Cataloguing Database") exposed very clearly that many basic and crucial decisions must be taken necessarily before starting a new (digital) catalogue, because they strongly influence the following actions.

One of the main matters was to determine the number (and the quality) of data fields for the manuscript description. Another crucial point was that of deciding 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.

The database was then programmed on the basis of the open source MyCore online database software, which is an open source software allowing limitless development possibilities.

The model however was enriched in order to accommodate many of the characteristics that a researcher may consider relevant in a catalogue, in particular in a digital catalogue.

Particular attention has been reserved to the arrangement of descriptive fields (over 100 all in all), because clearly sub-structured fields are much more useful for search and statistical queries

Simone-Christiane Raschmann ("The Union Catalogue of Oriental Manuscripts in German Collections. The manuscripts and block prints preserved in the Berlin Turfan collection") stressed the problems related to the cataloguing of fragments. Most of the 40.000 mss. from the Turfan oasis and neighbouring sites in East Turkestan are remains of large Buddhist manuscripts or block prints, so the "rules" for the compilation of the catalogues – elaborated by the KOHD project, but with a certain tolerance because of the great diversity of the Oriental manuscripts – had to be adjusted.

The speaker explained that the pattern for the description comprises 26 sub-items, like shelf number, cover, material, state of preservation, page number, format, number of lines, writing style, illumination, scribe, date, origin, author, title of the work, quotation of the first and the last line of the text as well as of the colophon, further remarks and a number of registers.

The presentation focused also on another important matter related to cataloguing, that of time. Although at the beginning it was decided not to restrict the registration of the Oriental manuscripts to inventory sheets or 'handlists' in order to provide detailed descriptions concerning the content of the single manuscript as well as full information concerning the style and the subsequent manuscript design, the still growing number of Oriental manuscripts in German collections made it necessary that in the course of the project for parts of the selected stock the catalogue entries had to be made in a much abbreviated form, similiar to handlists.

The registered manuscripts are accessible also through printed catalogues.

Marc Geoffoy ("Arabic and Syriac Manuscripts in the Mediterranean, Philosophy in Context") explained the philosophy of the French project which aims to solve two major difficulties in research in the history of philosophy: 1) understanding the conditions of the reception and the transmission in the Mediterranean World of philosophy of Greek tradition in the Arabic and Syriac languages; 2) discovering the numerous works "buried" in libraries in a systematic way, instead of by chance. Thanks to a methodical pooling of the information given in the catalogues of manuscripts and, much more, to the systematic examination of the collections, the project will identify and catalogue new texts, giving attention to the places and the institutions where the texts circulated, their copyists, sponsors, readers, etc.

The speaker stressed how this work requires a strongly interdisciplinary approach, combining a thorough knowledge in the history of philosophy, philology, codicology and computer science.

The project is based on the cross-information allowed by the ABJAD database.

Maria Teresa Ortega-Monasteiro ("Digitalization of the MSS of the Consejo Superior des Investigaciones Cientificas – CSIC") exposed the characteristics of this new project finalized to the cataloguing and digitalization of Oriental manuscripts conserved in CSIC libraries (*Tomás Navarro Tomás* in Madrid and *Escuela de Estudios Árabes* in Granada) in order to create a portal providing access to codicological and catalographic descriptions, as well as the complete text of each document, including a search system. The project has been developed by a multidisciplinary team composed of staff from the CSIC specialized in the fields of Arab, Hebrew and Hispanic studies, as well as librarians and technical staff. Some of the manuscripts have been previously restored.

The working team has published previous cataloguing projects, following new codicological and paleographical guidelines, by means of common methodologies and research protocols: Hebrew Manuscripts in Madrid libraries (2001-2006), Montserrat Abbey Manuscripts (2004-2010).

The main innovating results of the project consist in the fact that (1) it is the first internet portal in Spain that provides simultaneous access to codicological descriptions and full text for Hebrew, Arabic and Aljamiado manuscripts collections; (2) due to its multidisciplinary team, the project has a broad scope and character, being attractive for a wide public; (3) it follows the newest methodological trends in cataloguing, digitizing and metadata treatment; (4) the portal is now integrated in the CSIC libraries network and in web pages of participating institutes, with access to both collections simultaneously.
The portal is accessible in http://manuscripta.bibliotecas.csic.es

Maria Berggren ("ArkA-D, a platform for registering, cataloguing and digitizing manuscript material)" explained how Swedish libraries are now cooperating to develop
a database for the digitization of manuscripts, called ArkA-D. ArkA-D is thought to facilitate the combination of digital facsimiles with advanced metadata and to offer tools
for creating digital editions and linking commentaries to digital images.

The platform describes and digitizes manuscript collections in four steps. The first involves description at collection level in accordance with the so-called EAD (Encoded
Archival Description) standard. At a second level the produced metadata are connected
to digital images that are stored in a form that mirrors the hierarchal structure
created at step one. At the third level it is possible to add more document-related data (in formats specific for bound manuscripts, letters etc.). The fourth level allows for crowdsourcing transcriptions and commentaries. Levels three and four make use of TEI.

The *Session I* was also enriched by a more methodological communication of Elena Pierazzo ("XML schema for cataloguing Arabic MSS, Wellcome Institute and Oxford/Cambridge consortium") who anticipated in part the contents of *Session IV*.

Session II

The second session of the workshop – *Benefits and disadvantages of different "models" of electronic catalogues: printed catalogues transformed into electronic catalogues; printed catalogues transformed into online databases;* ad hoc *designed electronic catalogues (checklists, analytical catalogues, etc.)* – was introduced by a short theoretical introduction (Paola Buzi) which preceded an open discussion among all participants, who exposed their opinions about some important matters such as the terminology, the way of distribution of data in fields, etc.

The first point to be taken into account was the "distribution of data in fields" and the related "search problems". Passages of an unpublished text that Patrik Andrist delivered at the workshop *Methods and means for digital analysis of ancient and medieval texts and manuscripts* (organized in April 2012 by Tara Andrews and Caroline Macé) were read in order to guide the discussion concerning the opportunity to make use of the so called "bi-layer model" of description (opposed to a "mono-layer model" of description, which has been in use since the middle of the XX cent. in printed catalogues and is the most frequently-used model on the internet today for the cataloguing of composite manuscripts). Although not all the projects have taken into account this possibility so far, all the participants seemed to agree about the importance of such an articulated approach.

Another matter which was discussed diffusely was that of the terminology. It turned out that not all the persons responsible of cataloguing projects had perceived it as one of the priorities to be evaluated at the moment of creating a new (electronic) catalogue; moreover it has to be stressed the lack of an appropriate codicological terminology in some western languages. It is undeniable that this is still a weak point in the realization of a homogeneous standard of cataloguing.

Other delicate aspects discussed in *Session II* were, among others:

(a) the problem of dating an electronic catalogue: a printed catalogue has a certain date, while an online catalogue, which is continously updated, has not;

(b) the enormous difference between a cataloguing model elaborated for an online catalogue and that used to adapt a traditional catalogue to the electronic support;

(c) the problem of control: printed catalogues, because of the elevated costs of realization – are more controlled than the electronic ones;

(d) the delicate matter of the relationship between "traditional catalogues" and "electronic catalogues": despite all the undeniable potentialities of electronic catalogues, it was the opinion of the majority that as for the moment it is absolutely impossible to give up completely the printed ones. And it is not just a matter of tradition.

Session III

In session III, Jost Gippert presented a number of hypercatalogues and portals for manuscript descriptions, as well providing an overview of the technologies underlying them (e.g. protocols such as OAI, OAI/PMH or Z39.50, and metadata standards such as Dublin Core, METS and MODS). Ivan Boserup, head of the manuscript department at the Royal Library, then demonstrated the portal developed and maintained by CERL (the Consortium of European Research Libraries), which gives access to descriptions of manuscripts and early printed materials (up to 1830) in collections in Europe, the United States and Australia.

Session IV

In session IV, Matthew Driscoll presented the TEI P-5 manuscript description module, looking in particular at its implementation in the "Handrit" project (http://handrit.org), a union catalogue of early Scandinavian manuscripts in the Arnamagnæan Collection, divided between Copenhagen and Reykjavík, and the National and University Library of Iceland (altogether about 20,000 items). The principal XML elements for manuscript description were introduced as well as the mechanisms for linking such descriptions with multi-layered transcriptions of the text, digital facsimiles of the text-bearing objects and information about the people, places and organisations involved in their production, dissemination and reception ("authority files").

c) Assessment of the results and impact on the future direction

It can be asserted that the main goals of the workshop have been reached both because the scheduled programme was almost totally respected and because the role of all participants was very active and productive.

The main result of the workshop was definitely represented by the long and animated exchange of opinions that let compare different *praxeis* and habits in the realization of electronic catalogues.

During the "internal meeting" the development of Chapter 3 has been discussed and the speakers who attended Uppsala and Frankfurt workshops but have not delivered their texts yet have been strongly encouraged to do it by the end of August 2012.

As was explained to the participants, the state of the art of Chapter 3 is as follows:

I. *Introduction*: to be realized by Paola Buzi and Witold Witakowski

II. *Summary history of cataloguing* : about 50% of the written material has been already delived; the remaing texts are expected to be delivered by the end of August.

III. *Types and kinds of catalogues*: to be dealt with in the next workshop.

IV. *The codicological revolution? "multi-structured descriptions" vs. "simple-structured descriptions"*: a contribution written by Patrick Andrist is expected for the end of this Summer.

V. *The electronic revolution?*: to be handled and developed by Team 3; the speakers of Copenhagen workshop have been asked to deliver a text by the end of August; most of them, however, have already sent a short version of it before the workshop.

Potentially, the next workshop of Team 4 will take place at the beginning of June 2013. Possible locations are at the moment Greece (proposed by Zisis Melissakis) or Istanbul (proposed by Ewa Balicka Witakowka).

Another option might be that of a (partially) common workshop with Team 1, but this possibility, which has been already exposed to the Team leader, Marilena Maniaci, will be explored better after Arles workshop.

A final decision will be taken before the next editorial meeting in Nicosia.

d) Programme and participants

Thursday, June 7th

14.00-14.20
Opening address
Introduction to the workshop: "Printed catalogues" vs. "electronic catalogues": two different items?

14.20-16.30
I session (coordinated by Team 4)
Electronic catalogues (flat; in a database; on-line; on DVD, etc.): a survey
Invited speakers, all involved in important electronic cataloguing projects, will report briefly (10-15 min. each) on characteristics and "philosophy" of their own catalogues:

> Eva Nyström, Patrik Granholm: *Greek Manuscripts in Sweden - a digitization and cataloguing project*
> Maria Berggren: *ArkA-D, a platform for registering, cataloguing and digitizing manuscript material*
> Christophe Flüeler: *E-corpus*
> Marc Geoffoy: *Arabic and Syriac Manuscripts in the Mediterranean, PhiC*
> Evyn Kropf: *Michigan Islamic mss. project*
> Maria Teresa Ortega-Monasteiro: *Digitalization of the MSS of the Consejo Superior des Investigaciones Cientificas (CSIC)*
> Denis Nosnitsin: *DOMLib / Ethio-SPARE Manuscript Cataloguing Database*
> Simone-Christiane Raschmann: *The Union Catalogue of Oriental Manuscripts in German Collections. The manuscripts and block prints preserved in the Berlin Turfan collection.*
> Anne-Beate Riecke: *Describing Manuscripts of the City Library in Treves with MXML for Manuscripta Mediaevalia*
> Elena Pierazzo: *XML schema for cataloguing Arabic MSS, Wellcome Institute and Oxford/Cambridge consortium*

16.30-17.00
Coffee break

17.00-18.00
II session (coordinated by Team 4)
Benefits and disadvantages of different "models" of electronic catalogues: printed catalogues transformed into electronic catalogues; printed catalogues transformed into online databases; ad hoc *designed electronic catalogues (checklists, analytical catalogues, etc.). The potential problems of electronic catalogues.*

Among the other critical points, the following aspects will be discussed:

1. terminology
2. way of distribution of data in fields (in case of catalogues in a database)
3. description of manuscripts with or without their photos
4. catalogue interactivity
5. search problems
6. durability
7. respect of the authorship

18.00-19.00
<u>III session</u> (coordinated by Team 3, with the participation of Ivan Boserup, CERL portal)
Hypercatalogues and portals

1. demonstration of existing hypercatalogues
2. technical background
3. discussion of requirements in the field of Oriental manuscript cataloguing

19.00-20.00
Internal Meeting of Team 3 and Team 4

20.30
Dinner

Friday, June 8th

9.30-11.30
<u>IV session</u> (coordinated by Team 4)
Practical session

11.30-12.00
Coffee break

12.00-13.00
Final discussion and conclusions

13.30-15.00
Lunch

15.00-17.00
Visit to the Kongelige Bibliotek

***

LIST OF PARTICIPANTS

Convenors

1. Dr. Paola Buzi   Rome (IT)
2. Prof. Matthew Driscoll   Copenhagen (DK)
3. Prof. Jost Gippert   Frankfurt (DE)
4. Dr. Witold Witakowski   Uppsala (SE)

Speakers

5. Dr. Maria Berggren — Uppsala (SE)
6. Prof. Christophe Flüeler — Freiburg (CH)
7. Dr. Marc Geoffoy — Paris (FR)
8. Mr. Patrik Granholm — Uppsala (SE)
9. Dr. Evyn Kropf — Ann Arbor, Michigan (USA)
10. Dr. Denis Nosnitsin — Hamburg (DE)
11. Dr. Eva Nyström — Uppsala (SE)
12. Dr. Maria Teresa Ortega-Monasteiro — Madrid (ES)
13. Dr. Elena Pierazzo — London (UK)
14. Dr. Simone-Christiane Raschmann — Göttingen (DE)
15. Dr. Anne-Beate Riecke — Berlin (DE

Other participants

16. Dr. Tara Andrews — Leuven (BE)
17. Prof. Ewa Balicka-Witakowska — Uppsala (SE)
18. Prof. Alessandro Bausi — Hamburg (DE)
19. Dr. André Binggeli — Paris (FR)
20. Prof. Caroline Macé — Leuven (BE)
21. Prof. Andrea Schmidt — Louvain-la-Neuve (BE)
22. Ms. Evgenia Sokolinskaia — Hamburg (DE)
23. Dr. Alessandro Gori — Florence (IT)
24. Prof. Zisis Melissakis — Athens (GR)
25. Prof. Françoise Briquel-Chatonnet — Paris (FR)
26. Dr. Verena Lepper — Berlin (DE)
27. Dr. Joseph Moukarzel — Jounieh (RL)
28. Prof. Susana Torres Prieto — Madrid (ES)
29. Dr. Elzbieta Swiecicka — Stockholm (SE)
30. Dr. Natalie Tchernetska — Paris (FR)
31. Ms. Maria Widmer — Freiburg (DE)
32. Mr. Gerd Winkelhane — Berlin (DE)