

**Digital Approaches to Manuscript Studies**

**Hamburg, 23-24 July 2010**

**Report**

1. Summary (1 page)
2. Description of the Scientific Content (3-4 pages)
3. Assessment of the results and impact on the future direction of the field (2 pages)

**1. Summary**

The aim of this first meeting of Team 3 was to give an overview of technical possibilities issues in the field of manuscript studies, mostly focusing on codicology and palaeography, although some other aspects, such as cataloguing and editing, were touched upon: (1) digitization of manuscripts, (2) processing images for research purposes (codicology and palaeography), (3) use of "chemometrics" for codicological description, (4) statistical uses of codicological data, (5) linkage between images of manuscripts and texts in an XML environment (TEI). It seems suitable that the the first meeting of Team 3 was primarily occupied by questions related to the digitisation of manuscripts, since this must be the first step for any project in the field of textual scholarship in the digital age. The aspect of preservation of the manuscripts was naturally always present in the discussion about digitization and chemical analysis techniques. Since the techniques and knowledge involved in this field are more advanced for manuscripts written in Western languages (Latin and vernaculars), leading figures, both scholars and scientists, were asked from outside the COMSt community to provide their insights and experience in order to explore the applicability of those techniques and knowledge within the area covered by the COMSt project. Amongst the lectures, there was a good balance between general overviews of existing tools and techniques and presentations of concrete case studies applying those tools and techniques more specifically on Oriental manuscripts.

**2. Description of the Scientific Content**

(1) As one of the most pioneers and most important scholars active in the field of e-Science in the Humanities, Manfred Thaller showed, through the example of a digitisation project of German charters, how the digital approach is actually consisting in several layers of interpretation and how those layers are built upon one another. M. Thaller stressed the fact that most systematic digitization projects remain underused and that it is therefore important that a project does not confine itself to digitization alone but immediately builds knowledge upon it. M. Thaller describes the framework as follows:

1. digitization: importance of a persistent identification (canonical identification)
2. symbol manipulation
3. transcription

4. editing

5. research / publication / teaching: importance of a collaborative effort

Teaching (possibly e-learning) is a crucial issue: if the images and texts we are producing are not used by others and are not used in the teaching room, then they are useless.

(2) Andreas Lammer reported on the D.A.R.E. (Digital *Averroes* Research Environment) project run at the University of Cologne (Thomas-Institut), which aims at providing a full digital account of the manuscript basis containing the works of the Medieval Spanish philosopher, Averroes. His main points concerned the interrelationship of the three versions of Commentaries to Aristotle provided by the author (Short, Middle, and Long), composed in Arabic, but often preserved only in a Latin translation. Lammer largely dwelt upon the structure of the XML databases established by the project, especially upon the "semantic ontology" behind the structure. The final goal of the "ontology" is the global reconstruction of the interrelationship and interaction between manuscript and textual basis (printed editions included), bibliographical primary and secondary data, and their distribution along the historical and geographical axes.

(3) Johannes den Heijer, as the leader of Team 2 of ComSt, gave an overview of the MANUMED digitising project which is a collaborative project concerning manuscripts from around the Mediterranean Sea. MANUMED provides technical and financial support for the digitization of manuscripts (originals and microfilms), inscriptions, incunabula, but also oral traditions (folksongs, music, etc.). MANUMED hosts those images free of charge and give access to them in so far the copyright-owners allow that. Far from meeting the ambition of covering entire libraries or museums collections, the project has proved so far particularly effective in dealing with smaller, limited collections and manuscript funds.

The following two presentations were related to Multispectral Imaging of Palimpsests:

(4) Michael Phelps described a setup for high-resolution multispectral imaging of manuscripts based on a manuscript cradle, consisting of a camera with a 48 megapixel monochrome sensor, a lens with transmission from UV to IR, a filter wheel and LED-based monochromatic light sources. This permits the use of narrow-band illumination at 13 different wavelengths (including the UV and IR range), with additional filtering to capture either reflectance or fluorescence (or creating an RGB image). The lamp setup includes an additional row of LEDs to provide strafing illumination to capture surface structure of the page surface.

Some highlights from the discussion: Using backlighting has also proven useful, however integrating this has proven challenging due to physical limitations of the cradle setup. A software that automates evaluation of images (e.g. PCA based on several images from different spectra) would have to be created to make this setup, which is currently limited to Phelps' team including scientists handling the image evaluation useable by other researchers on their own.

(5) Jost Gippert reported on experiences gained in the application of the MuSiS camera system with palimpsests and other handwritten sources that require special imaging methods such as,

e.g., manuscripts with overstriking or blackened text (the examples used were letters of the early 19<sup>th</sup> century journalist Ludwig Börne that are currently being digitised and analysed at Frankfurt), faded colours (illustrated with the famous Behaim Globe of 1491 preserved in the German National Museum at Nuremberg), or other damages (e.g. in Tocharian manuscript fragments of the Berlin Turfan collection). As a preliminary result, Gippert demonstrated that both the material of the artefact to be analysed and the type of inks used determine the area of the spectre in which the best results can be obtained in applying multispectral imaging. E.g., while parchment palimpsests are still visualised best in the ultraviolet to violet range, the application of infrared may be recommended for overstriking on paper manuscripts or faded colours. One part of the presentation was devoted to recent software developments for the MuSiS system.

In the discussion, the idea of applying several different imaging systems on the same set of manuscripts (e.g., palimpsests) was put forth, with a view to a further enhancement of the techniques.

#### (6) Steven Delamarter, Ethiopian Manuscript Imaging Project's (EMIP) Digitization Stations for Manuscripts and Microfilms

The presentation gave an account of the project started in the spring of 2005 "with the mission to help preserve images of Ethiopia's manuscript heritage and make them available for scholarly study. The manuscripts concerned were in the possession of university libraries, dealers and private owners. Most have come from North America, though a few are from England, Israel and Kenya. All of the owners have graciously consented to have their manuscripts digitized and made available for scholarly study. At this time, complete image sets are available for nearly 350 codices and 340 magic scrolls" (cfr. <http://www.hmml.org/Vivarium/sgd.htm>).

Unlike most other current projects, the digitisation setup used had to be geared towards speedy digitisation of a large number of manuscripts, as materials are only accessible for short periods of time (before possibly becoming completely unavailable) and digitisation has to be carried out in the field using staff that can receive only extremely basic training. Thus, taking into account limited funds, the setup used consists of mid-level DSLRs, a simple tripod setup and fluorescent light bulbs (with fallback to daylight), allowing rapid operation even under conditions where power failures etc. would impede the use of more specialised equipment.

At the moment, the project is making available on the net as many Ethiopic mss. as possible. A number of paper catalogues has also been produced and more are planned for the future. The mss. considered till now number about 2.017. No special cataloguing samples were provided during the presentation. Yet, it is evident from the website that most of the effort has been provided in the attempt to preserve and make available for future study as many Ethiopic manuscripts as possible, and still little has been accomplished as far as the catalogation and analysis process is concerned.

Some highlights from the discussion: One of the most important technical challenges faced is in fact the further processing to archive the images and make them available. As an incentive for parties making manuscripts available, using some form of micro-payment for web access to the images is being considered.

(7) Ira Rabin, Writing materials database as a future dating tool

The aim of the database is to gather the results of chemical analyses of writing materials (inks, leather, parchment, scrolls...). The analysis (chemometrics) should be non-destructive and its scope must be as wide as possible. The purpose is to work first on dated and located manuscripts and then to use the results in order to help dating and locating other manuscripts. A special point of interest and one of the next goals of the research team led by Dr. Rabin is the development of a transportable, hopefully also a portable, analysis laboratory.

Some highlights from the discussion: Any project using methods of scientific analysis on writing supports, inks etc. should aim to provide quantified data, not merely results of relative comparison among the materials examined, so that this data can be reutilised in larger contexts.

(8) Steven Delamarter, X-Ray spectroscopy and a Fourier Transform-infrared analysis of Ethiopian Inks

The second presentation by Delamarter aimed at providing first hints on inks used in Ethiopic manuscripts on the basis of a chemometric analysis carried out in USA Labs. The result of the analysis provided evidence for the use of iron gall ink in one 18th cent. Ethiopic manuscript. Even though at its initial stage and to be carefully evaluated in accordance with specialists in chemical analysis - as Dr. Ira Rabin keenly pointed out - this first result is not at all deprived of significance because it contradicts the established opinion (based upon linguistic and philological evidence as well as present-day ethnographical analysis) that only carbon inks were used in Ethiopia. The presenter also announced the intention of mapping scribal practices correlating them to inks of chemical composition, which was considered by Dr. Rabin as probably still too ambitious on the basis of our present scientific knowledge and analysis capacity.

(9) Bernd Neumann (computer sciences) listed four areas of manuscript studies where computer tools developed for other applications can provide some help:

1. image restoration and segmentation
2. writer identification
3. content-based image retrieval (without manual annotation)
4. manuscript analysis

Those techniques are not yet fully developed to be applied successfully to manuscript studies, but the first results are encouraging.

(10) Torsten Schaßan presented some "trends" in codicology and palaeography in the digital age.

The communication focused on the presentation and critical evaluation of digitisation projects distinguished as to local (e.g., Bayerische Akademie, Bibliothèque nationale de France), regional, national (e.g., Manuscripta Medievalia, Gallica) and European (Europeana, Manuscriptorium) scales.

(11) Matthew Driscoll - Eric Haswell, Linking text and images

This presentation discussed the TEI P5 model for marking up rectangular areas of image files referenced in a TEI document's facsimile section that can then be pointed to in e.g. a transcription within the same TEI document. The UVic Image Markup Tool was presented as a simple means of creating such markup.

Some highlights from the discussion: Support of more complex area definitions (i.e. polygonal instead of rectangular) in TEI is presently being considered. No suitable and easy to use tool currently exists for complex documents referencing large numbers of images. Many projects otherwise using TEI prefer to use a METS/MODS approach for encoding structural information pertaining to facsimile images.

### **3. Assessment of the results and impact on the future direction of the field**

One concrete immediate result of the meeting is the creation of two databases which should be filled and updated by COMSt members (especially, but not only from Team 3):

- digitization projects: [http://creator.zoho.com/comstudies/comstprojects/#View:COMSt\\_Projects](http://creator.zoho.com/comstudies/comstprojects/#View:COMSt_Projects)
- digital tools: [http://creator.zoho.com/comstudies/comstprojects/#View:Digital\\_tools](http://creator.zoho.com/comstudies/comstprojects/#View:Digital_tools)

The speakers were asked to provide their presentations, which would be made available on COMSt website.

Some points of special importance:

1) It is important, as M. Thaller pointed out, to see the digital approach in the field of textual studies as a whole. There is no point in digitizing manuscripts no one is going to use: as T. Schaßan said, the manuscript as an object is being destroyed as it is being used, on the contrary the life of a digital image is guaranteed by the very fact that it is used.

2) When assessing the value and relevance of techniques and tools (best-practice), the cost-efficiency criterion should not be the only one. Other criteria are more crucial: the preservation of the object and the scholarly / scientific value of the data. A clear identification of the digital data is also a crucial point.

3) I. Rabin stressed the importance of dialogue and respect between scientists, who are providing technical help in the analysis of data, and scholars, who are asking the questions and interpreting the answers from the analysis. There is a great danger in blindly using black-box methods and the only remedy is to make efforts in order to understand each other and speaking the same language.

4) It is obvious that scholars working on oriental texts and traditions are not in a position to create standards, in most (if not all cases) we will need to use standards which have been created by other communities of scholars (like TEI or UNICODE for example). It is important though to

be aware of what exists and to follow the development of these standards, so that the specificities of oriental languages and manuscript traditions are taken into account.

5) Under the label “Digital Support of Manuscript Analysis”, some of the presentations clearly consisted on the application of “hard” scientific tools to the study of manuscripts: in this sense, the “digital support” consisted in the obvious use of PC-assisted tools. One of the future task of Team 3 will be that of a more precise and rigorous application of the “digital approach” red thread as digital encoding and treatment of manuscript-relevant information (concerning text, materials, structure, layout, intertextual and bibliographical aspects, etc.).