# Manuscriptoium – basics and compatibility

Draft version 1.2

23.2.2006

Table of contents

Manuscriptorium is a system for collecting and making accessible on the internet information on historical book resources, linked to a virtual library of digitised documents. The Manuscriptorium service is financed by the National Library of the Czech Republic and managed by AiP Beroun s.r.o. .

# The history of Manuscriptorium

Manuscriptorium is the outcome of activities begun under the auspices of the UNESCO „Memory of the World" initiative in 1992. Subsequently, objectives were set for a Czech follow-up programme, Memoriae Mundi Series Bohemica (MMSB), and means of satisfying relevant current demands were identified and implemented. At an early stage the requirements for long term viability of the project were identified and objectives set for the development of an electronic research environment for the sphere of historical book resources. Manuscriptorium brings together information provided by collaborating partners, contributors from libraries, archives and museums.

Practical experience and the co-operation of experts in the technical field and in librarianship led to the formulation of the fundamental tenets underpinning Manuscriptorium.

# The fundamentals of Manuscriptorium

### Respect for the work of our predecessors

We take the view that recent information about historical documents has been created in good faith and with the best of intentions, just as that of older provenance.  Such information came into being at different times and under varying circumstances. The views and the extent of knowledge from which it was drawn were variable.  However, all available information is of value, unless it is distorted and as long as it is documented. The purpose of our work is to ensure the accessibility of authentic existing information.

### Separation of the permanent and the variable

We have recognised the need for the strict separation of information per se from any form of instrument involved in its creation, its management and the form and manner in which it is rendered accessible. The existence and the usability of the information must be secured whenever tools are amended or replaced in any way.

The fundamental form of information preservation is XML files with a structure meeting the Library of Congress METS standard. For the preservation of descriptive metadata in the METS framework the MASTER standard is applied, specially developed by an international project team in co-operation with Oxford University Computing Services for the description of manuscripts, old printed books and other historical documents.

### Identification of common properties of information and the retention of variety

Information about historical resources is found in a great variety of systems; its content and structure is always very similar in parts.  The detail of form and content naturally exert a considerable influence on the approach of researchers, archivists and librarians and the technology they use.

Whereas we do not wish to restrict or influence  freedom to use and create information in any way, as a matter of principle we wish to obviate the need to be professionally bound by the demands of technology.

## *Openness to other projects*

We consider our activities useful, but by no means unique. They offer a service ensuring that the data entrusted to us will be preserved for the future and also for their current accessibility. We therefore consider it essential to develop tools ensuring access to the collected data, both for our own end users and via services offered to other systems

## *Accessibility and usefulness for partners*

We wish to simplify the process of contribution of data to Manuscriptorium to such an extent that co-operation with Manuscriptorium becomes an interesting and useful alternative for major institutions, complementing their own activities. For smaller institutions and individuals, on the other hand, Manuscriptorium is a means to ensure secure preservation of their data and provide modern methods of access to it. The current outcome of attempts to achieve these objectives is the Manuscriptorium Database presentation, which is accessible either via the page representing the project as a whole www.memoria.cz  or directly via www.manuscriptorium.com . We also anticipate the possibility of *individual forms* of data presentation by partners.

## *Secure and reliable preservation of information for the future*

We are aware of the great responsibility for secure preservation of the data collected, not only in terms of its physical existence but also in the sense of the preservation of its general usability, its reliability and also information about the circumstances of its creation.

We are aware of the extreme vulnerability of electronic data and its dependence on moral values, technological standards and the continuity of our society.

Manuscriptorium stores and makes accessible data and information whose origin and quality are guaranteed.

## *Openness of data storage*

Manuscriptorium was established primarily for data created in the Czech Republic under the **VISK6** project; the preservation and accessibility of this data is primarily the responsibility of the National Library of the Czech Republic. However, Manuscriptorium also respects other forms of data storage and enables access to data deposited in independent, *remote storage facilities* by means of compatible tools.

## *Flexibility of free access options*

Manuscriptorium makes metadata available to the general public without restriction.

As a matter of principle, Manuscriptorium makes all data (images, sound, full text) freely accessible to its partners, and to others via a licence.

How far the general public is given free access to data in its fullest extent is a matter for the partner to decide.

### Licensing policy

**Licensing policy** is determined by the National Library of the Czech Republic. Its objective is to manage the use of its own data, and that entrusted to it, in such a way as to ensure easy access to all available information for professional users and to ensure that at the same time the interests of indviduals are respected and that the rules regarding the use of data supplied by different partners are observed.

### Manuscriptorium compatibility certification

Partners whose information is selected for publication in Manuscriptorium and who satisfy the technical requirements for its importation receive a **Manuscriptorium Technical Compatibility certificate**. The award of this certificate is the pre-condition for the granting of a licence for unrestricted access to Manuscriptorium free of charge, i.e. access to the data provided by the other partners in its fullest extent.

Partners whose data satisfy the quality criteria receive a **Manuscriptorium Image Quality certificate**. Contributors' documents in Manuscriptorium are marked with this certification, representing information for other partners regarding the guaranteed quality of existing primary data.

# The status quo

A digital library of manuscripts, old printed books and other scarce documents is in existence. Like any other library, it contains a catalogue – in this case the OCHBR (Open Catalogue of Historical Book Resources) and its own digital documents deposited in a data storage system. The OCHBR brings together information about physical documents (manuscripts etc.) in the form of catalogue records in XML format. The Repository contains digital copies of a subset of these catalogued documents, known as complex digital documents (CDD) As a matter of principle, the Manuscriptorium system stores the metadata centrally in the OCHBR and provides access to digitised documents held in the operator's data storage facilities and in the remote storage facilities of its partners.

# Objectives of Manuscriptorium

The basis has been laid, facilitating the collection of existing data and the provision of access to it. A further objective is the development of a virtual resaerch environment for the sphere of historical resources. Facilities for the digital presentation of these resources (images and text) and for rendering them accessible are being expanded. A virtual digital library of copies of images and full texts is being established. Links to further data of relevance to these resources are being provided (analyses, further descriptive metadata, further electronic documents/objects) and the provision of information to other systems is also being secured by means of standardised procedures (communication interfaces) (**Z39.50**, **OAI-MPH**)

# Document selection in Manuscriptorium

The following is a declaration of the criteria which digital documents must satisfy in order to be accepted. Critical factors are:

1. the significance of the document and the value of its incorporation in Manuscriptorium,
2. the technical usability of the information about the document (the metadata) and the information introducing the document, i.e. images, sound files, full texts (the data).
3. the quality and reliability of the data

The National Library of the Czech Republic makes its decision regarding the incorporation of a document in the Manuscriptorium system as follows.The focus of Manuscriptorium is on rendering accessible the written cultural heritage prior to 1800, with particular reference to book materials, but it is also open to documents of an archival nature (especially, books, papers and maps) as a secondary priority. All digital documents in Manuscriptorium represent, in principle, an original physical document, whether still in existence or no longer extant.

The method of selection is described in greater detail in the document **[12]**.

# Technical compatibility

Our premise is that electronic information about historical resources is with increasing frequency usable directly in Manuscriptorium or is convertible to the required XML format, either routinely or with only a limited amount of effort. Compatibility can be readily achieved, especially in new projects and in projects whose approach has so far mainly involved digitisation and the creation of data.

Metadata created by any tools whose output consists of XML records using **DTD MASTER [5]** are considered inherently compatible. Minor variations are admissible, but they are subject to the Director's decision.

In cases where the conversion of existing information is problematical or the contributor considers that its form should not be altered, yet nevertheless has an interest in publshing it in Manuscriptorium, it is possible to create tools for routine conversion for purposes of achieving compatibility.  The contributor attaches a so-called **connector** to his system and continues to manage his data independently. The information is offered to Manuscriptorium in a compatible form via this **connector**.

For common types of records, the **connectors** are attached to the Manuscriptorium server or to the contributor's server. The use of existing **connectors** or the building of new ones is a matter for individual negotiation between the future partner, the National Library of the Czech Republic and the Manuscriptorium manager.

As a matter of principle, the National Library of the Czech Republic supports and finances the direct import of data to Manuscriptorium.

Information for which a **connector** exists or for which one is to be built is considered compatible.

The conventions followed in Manuscriptorium are described in the document **[14]**.

Acceptance of these conventions is a guarantee of total technical compatibility.

## *Storage of data and associated metadata*

The attachment of images to metadata requires the application and observance of conventions describing the path to the storage of images.

Manuscriptorium approves a broad range of data storage methods.

### Non-public data storage

If data is stored within a system which prevents direct access to it, the attachment of the data must be carried out ad hoc by means of the so-called connector, ensuring the conversion of data and metadata to a form which can be attached to Manuscriptorium.

### Public display of data

A good deal of data is or can be rendered freely accessible on the internet – each image can be reached at a unique URL address via the http protocol.
Data which is accessible in this way can usually be attached to Manuscriptorium without difficulty.

### *M-tool*

For ease of metadata creation, **M-tool** is available for download and free use – it enables the creation of basic metadata with links to image data stored at freely accessible http addresses.
**M-tool** is available at http://www.memoria.cz/mtool/.

# Compatibility of data quality and reliability

Data rendered accessible in Manuscriptorium via the internet is usually adapted to the technical requirements of the internet and is usually derived from primary data. Manuscriptorium's mission is to guarantee *secure and reliable preservation of information for the future*. For data created under the Czech Republic's VISK6 project, particular attention was devoted to this requirement. Data created with appropriate care and guarantees is marked accordingly in Manuscriptorium. The vital criteria to be met before data is awarded this mark of recognition in Manuscriptorium are described in the document **[13]**

In practice, however, it is not always possible to guarantee optimal conditions for the creation of data. For this reason data not meeting these quality criteria may also be incorporated in Manuscriptorium. The National Library of the Czech Republic reserves the right to make decisions about the incorporation of data not meeting the quality criteria.

# Connectivity of Manuscriptorium

For the use of data external to Manuscriptorium, tools are available for accessing and sharing metadata in standard formats (UNIMARC, MARC21, MASTER, Dublin Core…) via standard communication interfaces, **Z39.50**, **OAI-PMH**.

The range of such services and their extension is determined by the National Library of the Czech Republic.

# References

Reference is made in the document to the following publications:

[1]  Manuscriptorium v.2.0 – analysis of the system, Deecember 2004

[2]  Manuscriptorium v. 2.0 – the complex digital document, October 2005

[3]  Memoriae mundi series Bohemica, available at URL: http://digit.nkp.cz

[4]  Manuscript Access through Standards for Electronic Records (MASTER), available at URL:
      http://xml.coverpages.org/master.html

[5]  Reference Manual for the MASTER Document Type Definition, available at URL:
      http://www.tei-c.org.uk/Master/Reference/oldindex.html

[6]  MEdit, available at URL: http://www.memoria.cz/download/medit_cz.asp

[7]  TorXmlValid, available at URL: http://www.memoria.cz/site_cz/download.asp

[8]  jEdit, available at URL: http://www.jedit.org/

[9]  Emacs, available at URL: http://www.gnu.org/software/emacs/emacs.html

[10] NoteTabLight, available at URL: http://www.webmasterfree.com/notetablight.html

[11] Manuscriptoium – basics and compatibility:
      http://www.memoria.cz/docs/manuscriptorium_basics_and_compatibility_ENG.pdf

[12] Document selection and preparation of descriptions:
      http://www.memoria.cz/docs/manuscriptorium_document_description_ ENG.pdf

[13] Manuscriptorium Image Quality:
      http://www.memoria.cz/docs/manuscriptorium_image_quality_ ENG.pdf

[14] Manuscriptorium Technical compatible:
      http://www.memoria.cz/docs/manuscriptorium_compatibility_technical_ ENG.pdf