ECP 2006 DILI 510049

ENRICH

Survey results and their interpretation

Deliverable number	*D-2.1*
Dissemination level	*Public*
Delivery date	*20 March 2008*
Status	*Draft*
Author(s)	*Zdeněk Uhlíř (NKP), Tomáš Psohlavec (AIP)*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

Document Version Control

Version	Date	Change Made (and if appropriate reason for change)	Initials of Commentator(s) or Author(s)
draft 0.0.1	2008-02-15	AIP/TP	Founding of the document, summarizing available information
draft 0.0.2	2008-02-19	NKP/ZU	Abstract Summarization for ENRICH, quantifying
draft 0.0.3	2008-02-20	AIP/TP	Removing internal notes, translation
draft 0.0.4	2008-03-05	NKP/TK	Translation

Document Review

Reviewer	Institution	Date and result of the review
JH	CCP	2008-03-17, formal check and content review, amendments of Chapter 1 and 5

Approved By (signature)	**Date**

Accepted by at European Commission (signature)	**Date**

1 Executive Summary

The goal of the ENRICH project is to integrate existing scattered electronic content through the way of the metadata enrichment and coordination between heterogeneous metadata and data standards.

The first steps were undertaken within the WP2 / T2.1 to update information on the available content and also to describe the initial readinness of particular Content Partners for the Intersystem Communication.

Above all we have targeted the numbers of available documents and pages, existing collections and their attributes and descriptions. In the technical area we have focused on the existing digital repositories and their structure, physical availability of the data elements and the way of access, data and metadata formats, ability to share the metadata and the way of sharing.

The questioning was led individualy with each of the Content Partners as during the preparation phase of the ENRICH project it became obvisous that the particular situation within each ENRICH Content partner‘s institution is diametrically different: different digitisation approaches and processes are used producing different digitisation outcomes, different formats and standards are used to preserve the outcomes and different tools are used to manage and manipulate the digital documents. In such a heterogenous environment targeting particular partners individually is the only possible effective way of cooperation.

Though the results of this survey show that most of the Content Partners are (or soon will be) ready to cooperate within the project as the most basic principles of cooperation were fulfiled. Where these were not fulfiled a discussion will be led during the T2.3 to find a solution.

Even in the heterogenous environment the interpretation of the survey results enabled their classification of the Contents Partners into 4 groups that will use the same concepts of cooperation. Detailed description of each way of cooperation is also included in this document.

The information gathered is a good source of information for the subsequent dialogue led within the T2.3. It is possible to propose particular steps to be undertaken with Content Partners during the subsequent tasks.

2 Shortcuts

MNS – Manuscriptorium

MAN - manuscript
PRN – print
INC – incunabulum
CHR – charter
JOU – magazines/periodic
MAP – maps
DRW – old drawings

GUI – graphic user interface
API – application programming interface

Other and not explained shortcuts are the names of our partners.

3 Terms

Digital Document, Document

Digital form of the original document. It contains data (typically image files), descriptive metadata (usually structured bibliographical description), structural metadata and other types of metadata. Structural metadata create the digital binding of the document, and the digital document can not exist without them.

Descriptive Metadata

Descriptive metadata describe properties of the original document. Usually it is the structured bibliographical description. Descriptive metadata can use standard formats (TEI P5, MASTER, MARC 21, UNIMARC, etc), or proprietary formats, and generally we can even think of any structured texts as descriptive metadata.
Each document which is going to be imported into Manuscriptorium has to contain descriptive metadata at least in the scope of Evidence record (see further the definition of "Evidence record" for these purposes).

Metadata about Structure, Structural Metadata

Metadata about structure create the digital binding of a document. They describe

- the structure of the original document, including information about so called applicable pagination/foliation
- existence and location of accessible digital representations

For example in the case of image document it concerns this particular information:

- the consequence of pages/folia and their applicable pagination/foliation – enables the orientation in the digital document
- information about existing quality levels
- information enabling construction of links to location of particular image files of all the available quality levels (this information is often accessible directly in the form of URL, but it holds generally that it does not need to be expressed in such an explicit form)

From the above mentioned information it is obvious that these „binding information" create the basic condition for existence of a digital document in the sense of definitions explained in this document, because standalone images do not create a document, and they still remain a mere sequence of numbered images without any chance for a subsequent convenient usage.

Technical Metadata about the Process of Digitization

Technical metadata describe technical aspects of digitization, and principally influence the quality of the digitization process outcomes. The usability of image data is directly determined by the quality of this type of information.
Although this information is crucial from the point of view of the digitization process, availability of this type of information is no principal condition for accessing documents in MNS.

Digital Binding

Digital binding is created by linkage information being contained in metadata about structure. Digital binding enables the composition of a digital document itself, and the subsequent usage of the documents.

Import of Documents into Manuscriptorium

Speaking about the import of documents in relation to Manuscriptorium, we perceive it as dividing of metadata from data (of images) in such a way that data (images) remain being managed by Content Partners, and just metadata are physically imported into Manuscriptorium. Metadata are further processed inside the system for purpose of presentation in homogeneous user interface of this platform. Based on internal processing of these metadata and end-user interaction the system generates outputs into GUI in such a way that data themselves (images) are transferred directly from the Content Partner (repository) into the client (browser) of the end-user. The composition of the final digital document is not accomplished before displaying in user browser and the document is composed using the digital binding information.

Repository, Data Repositry

This term we use for any interface enabling direct access to individual data components of a digital document by means of URL, and the protocol HTTP. In the simplest version it can be a web server directory structure filled with data (image files). The structure has to have its image in structural metadata.

Connector

Connector is a tool enabling the conversion of existing input metadata of digital documents according to needs of the Manuscriptorium system. It is a part of input interface of this system, and it is constructed in such a way that it can not overreach to the managing system of a partner.

This approach enables to divide the data production and management from the data presentation, and to optimize both of these tasks. It ensures also that modifications of existing processes are not required on workplaces of individual Content Partners.
The most important part of the connector are conversion routines for transferring descriptive and structural metadata.

Connectors process **existing** information.

With regard to the demands of creating a connector (and its conversions) these are prepared for partners producing/managing relevant amount of digital documents only (to rationalise the time demands and financial costs).

Off-line Connector

The connector receiving metadata in an other way than by the means of the protocol OAI-PMH.

On-line Connector

The connector cooperating with the harvester of MNS which conveys metadata to the input of the connector using the protocol OAI-PMH.

M-Tool

M-Tool is an application serving partners of Manuscriptorium, who create new metadata for existing image data. It enables to create **descriptive and structural metadata**, and consequently also digital documents in the sense of definitions mentioned above.
M-Tool currently exists as a stand-alone application, and it is going to be created in an on-line version during the work on project ENRICH. It's functionality is going to be updated in regards with results of concrete WP (mainly WP5).

Talking about M-Tool in this deliverable, it is not important which version we have in our mind unless mentioned otherwise.

M-Can

M-Can is an on-line application enabling the visual checking of readiness of digital documents to be imported into Manuscriptorium. It enables to upload metadata prepared by M-Tool, and to check their presentation by the means of tools being used also in Manuscriptorium. Using the application the verified documents are able to be passed on to the import into Manuscriptorium.

Harvesting Interface

The interface of a digital library accommodated by the protocol OAI-PMH ensuring transferring metadata into the Manuscriptorium system.

Harvester

An engine ensuring the data gain from harvesting interface.

The Type of an Document

Under the term "the type of an document" we understand the document classified from the point of view of physical properties of the pattern into the categories manuscript, print, incunabulum, charter, map, etc.

Collection

Under the term "collection", we understand the classification of documents among any criteria (content criteria, provenance but also the type of document, etc.).

MASTER+

DTD based on MASTER project DTD adapts the complete and unchanged msDescription, on the same level it parallelly uses elements for recording technical and structural metadata. It has been developed in cooperation of NKP and AIP (http://digit.nkp.cz/MMSB/1.1/msnkaip.xsd).

Evidential Record

Simplified for the purposes of this deliverable: It is an elementary descriptive record of a document containing at least identifying information of this extent:

- home country of the physical document
- the name of the library where the document is stored
- the geographical entity (city) of the library location where the document is stored
- the shelf mark of the library where the document is stored

Each document going to be imported into Manuscriptorium has to dispose of descriptive metadata at least in the extent of the evidential record from the technical point of view.

4 Intersystem Communication – Preparation

The way of communication among individual systems/projects, and the accessibility of documents itself are influenced mainly by (for explaining used definitions see above):

- the quality of metadata about structure
- the accessibility of metadata about structure
- the data accessibility in the repository

The quality of work of a end-user working with the documents integrated into the platform Manuscriptorium is influenced mainly by:

- the quality of image and other types of data
- the quality of descriptive metadata
- the quality of connectors (more exactly the quality of conversion processes applied in the particular connectors above the metadata Content Partners)

From the point of view of Intersystem Communication in T 2.1 we are focusing on **gathering and preliminary analysis of all accessible information about properties of both the mentioned metadata groups and about ways of data availability and organization in our partners´ repositories**. The output is this deliverable (D2.1).

Gathering of information is based on principles usually applied during connecting documents of the new partners into the Manuscriptorium system. The communication is lead individually with particular partners because the first anticipatory questionnaires manifested that **approaches to solving of the historical digitization problems are such diverse in particular institutions that we could very hardly find a commonly operative working method** which we could apply on all our partners. Individual questioning enabled to formulate more precisely questions clearing specific details of digitization projects of our partners.
The preliminary analysis being the result of questioning (and the part of this deliverable) is addressing each partner separately, and apart from gathering accessible information it also gives names to concrete emerged problems. We will have to resolve these problems during m3-m6 within T2.3 (Dialogue for preparing the technical conditions for inter-system communication).

An important benefit of this announcement is also **the suggestion of steps necessary to do in T2.3** to reach the readiness of particular Content Partners for document implementation in the Manuscriptorium platform in concordance with principles of this (Manuscriptorium) project.

4.1 Results of Questionnaires – In General

We managed to gain information from **all 10 Content Partners** (IMI originally acting in the role of a Content Partner can not (against the original assumption) provide documents of partner Library of the Lithuanian Academy of Sciences which should have contributed content through IMI mediation).

We can judge from the results of questionnaires that within one partner institution there can be diverse approaches both to creating metadata and to organizing the repository for various types of documents (manuscripts, prints, charters...). The basic unit which properties require to be described is not always a set of documents managed by partner institution, but in some cases the basic units are the different collections in a partner institution.

5 Intersystem Communication – Readiness of Content Partners

The level of readiness of particular Content Partners differs strongly. Generally we can say that the more advanced and better founded digitization projects are, the easier the implementation of their outputs into the Manuscriptorium platform will be. In the case of advanced projects laying emphasis to utilization of perspective and longevity standards and approaches (for processing and storing of data and metadata) we can say that the basic requirements for documents implementation into the common platform have been accomplished, and we can straight get on with the creating of connectors (rules tuning for obtaining metadata, definition of conversion processes). Generally we can say this about collections whose metadata are accessible by the means of OAI, and which use common "independent" formats for data storing.

With partners who are beginning with digitization projects or who use a special or non-standard approaches or proprietary tools it will be necessary to finish the preparation for fulfilling basic requirements for documents implementation into Manuscriptorium. It regards mainly the preparation of particular repositories, and the existence of structural metadata.

Although the formats and levels of information differ in descriptive metadata, here we do not expect principal problems because the standard and proved approach of Manuscriptorium is the adaptation of input interfaces (connectors) to individual input metadata properties. Of course it requires the cooperation with the individual Content Partner in the field of mapping the information structure.

5.1 Criteria of Readiness Valuation of a Content Partner for Entering into Intersystem Communication

We can say that if a Content Partner can enter into Intersystem Communication leading to the import of his digital documents into MNS, he has to:

1) prepare the repository in the sense of the definition cited above
2) dispose of quality metadata about structure
3) dispose of descriptive metadata in the extent of the evidential record (technical criteria)
4) be able to pass metadata into the MNS system

The major part of the following text deals with the conclusions concerning the readiness of our partners to fulfill these criteria.

5.2 BNCF

			Notes
Format of descriptive metadata	UNIMARCslim		
Format of metadata about structure	MAG		
Numbers of accessible documents	MAN	**216**	**67845** actually accessible images
	MAN (Galileian collection)	Actually undefined	About **80000** images that will be accessible in the development of the ENRICH project
	Handwritten MAPS	**137**	**233** actually accessible images
	Historical printed MAPS	**810**	**3765** actually accessible images
	Historical printed BOOKS	**377**	**159381** actually accessible images
	Partially digitised historical printed BOOKS	**52096**	**211618** actually accessible images of cover, title page, table of contents and variable significant pages of each one of the partially digitized volumes
Samples handed over	Yes		
Repository	yes		Fully described
Mining interface	yes, OAI		at present for MAG and DC, not UNIMARC (a new profile will be created): http://teca.bncf.firenze.sbn.it/OAI/servlet/OAIHandler
Other information			

5.2.1 Metadata

BNCF provides its metadata in DC and MAG; by the projection of information we can get all the information required for import of documents into Manuscriptorium.

5.2.2 Descriptive metadata

The new UNIMARC-slim profile prepared specially for ENRICH project needs will carry the primary bibliographic description (descriptive metadata). In addition there will be available the "dc:identifier-like" information in the profile records to enable Manuscriptorium to paralelly link end-users directly to your digital library.

Details of preparation of new possible UNIMARCslim profile in the OAI interface will be discussed more detaily during the T2.3.

5.2.3 Metadata about structure

Root URL: http://teca.bncf.firenze.sbn.it/TecaFrontEnd/servlet/readImg?parameters
where parameters are couples param=value.
A concrete complete URL is possible to mine in MAG:
mag:file[@Location="URL"]/@xlink:href
e.g..:
http://teca.bncf.firenze.sbn.it/TecaFrontEnd/servlet/readImg?RisIdr=BNCF0002981479&usage=3

Parameters and domains of valid values		
Param	Value	sense
RisIdr	Mined in MAG	resource identifier
Usage	[1,2,3,4]	Qualitative level of the image [master copy, internal use copy, public copy, preview]

We managed to acquire images where usage=3, the usage=4 will be also made available during the ENRICH project.

5.2.4 Repository

It is fully prepared in accordance with the basic principles of MNS (verified by sample metadata).

5.2.5 Planned steps in T2.3

- To settle a temporal framework during T2.3 when there will be metadata at our disposal according to above mentioned changes (e.g. UNIMARCslim via OAI).
- To agree a final way of import both descriptive and structural metadata into MNS (will MAG structural information be part of UNIMARC slim profile or will be harvesdted separately?)
- To map properties of "xlimage software" in the relation to the posible integration of documents into native GUI of MNS.

5.3 BNE

			Notes
Format of descriptive metadata	DC unqualified		The same format for all collections Some elements are qualified
Format of metadata about structure	MASTER+ DTD		Created using M-Tool, for more information see below
Numbers of available documents	MAN	183	
	PRN	3020	
Samples handed over	None		
Repository	At present does not exist		For information see bellow
Other information			

5.3.1 Metadata

BNE references to online GUI with hte samples – representative samples of records levels with bibliographical descriptions are not at our disposal at present.

Two possible ways of cooperation:

1. The nonexistence of metadata about structure leads to the assumption to use M-Tool. There are **descriptive** metadata which should not be re-created using M-Tool. So M-Tool will be used for creating of evidential records with structural metadata. These records will be imported into MNS.

   At the same time a connector for descriptive metadata will be created. It will enable to update the existing records in MNS with detailed description information gained from DC records of BNE. The connector will also prepare the possibility to actualize MNS on the bases of changes being done with primary metadata BNE.

   As soon as the connecter will exist, it will be possible to import all related bibliographical records into MNS. When the work on M-Tool will go on, digital documents will be added to these records. In case of need we will proceed conversely. It will depend on the fact which way will be prepared sooner.

2. We will find out the possibility of exporting existing metadata provided online by DigiTool.

Information for auto classification according to documents types has not been verified. (No samples are at our disposal).

5.3.2 Repository

The contemporary presentation does not fill the function of the repository, because it does not permit the transfer of images using unique URL. DigiTool used in BNE for image presentation to the user. It transfers images using a script, and this script is accessible just within the application.

5.3.3 The Planned Steps in T2.3

1. To find out an alternative for the repository using one of the next two methods. The choice of them depends on DigiTool system possibilities in BNE, and technicians of BNE should make it after the consultation with AIP concerning the future implementation of documents into MNS. Possible methods of solution:
    a. To create a new repository in the form of a web server with JPEG file directories in one or more quality levels (the ideal variant from the point of view of MNS/ENRICH);
    b. To configure DigiTool in such a way that the script would return complete JPEG (or any other generally supported) images based on URL; we can create qualities subsequently "virtually" by changing parameters (we can suppose that this second option would be a high workload both for the system, and for web server).

2. To check up the ability of M-Tool to product structural metadata for the selected solution, as soon as we will know the solution of the point 1; in the case that M-Tool will not be able to encompass the chosen structure, the conclusions of T2.3 for WP5 will be proposals of changes in M-Tool.
3. To collect representative data samples at the same time. These samples will be used like the input for connectors' preparation in WP5.
4. To arrange the way of connecting full texts data which are in the course of preparation in PDF format.

5.4 BUTE

			Note
Format of descriptive metadata	MASTER, eventually TEI P5;		using M-Tool
Format of metadata about structure	MASTER+		using M-Tool
Numbers of available documents	PRN	about 130	
	JOU	will be supplied	
Samples handed over	yes		Samples of repository solution have been given, Descriptive records are not yet at our disposal (they are going to be created using M-Tool)
Repository	In preparation		
Other information	Only complete digitalized documents are going to be imported into MNS.		

5.4.1 Number of available documents

- There are more than 6000 pages representing about 130 chapters in old books available for sharing and enrichment.
- BUTE expects (but without committment yet) to provide about 100 additional articles (about 600 pages) from the early periodical entitled Journal für Fabrik, Manufaktur, Handlung und Mode, published in 1794, dealing with a variety of topics from engineering through trade and fashion in German language. This replaces Commentarii Academiae Scientiarum Imperialis Petropolitanae, (Russia) 1.tom.1726-14.tom.1744/[17]46 proposed earlier, due to the greater interest in the German language.

5.4.2 Metadata

In the case of BUTE samples are not necessary. Descriptive metadata will originate in the extent which makes M-Tool able to write for PRN documents. We have to consider which types of specific information will be necessary to write in M-Tool concerning JOU documents. It might be necessary to revise M-Tool according to conclusions.

M-Tool will have to make possible adding definite information for automatic classification according to types of documents.

5.4.3 Repository

The final repository structure has not been defined. The presented and described way of realization proofs the fully understanding of principles and advantages of documents implementation into MNS.

5.4.4 The planned steps in T2.3

1. As soon as we will have final versions of repository structure, BUTE in cooperation with AIP will verify the ability of M-Tool to product structural metadata for selected solution; in case any changes to the output of M-Tool will be required these changes will be an input to WP5.
2. Should M-Tool be able to generate metadata without any changes, BUTE in cooperation with AIP will verify suggested solution on chosen testing document.
3. An internal meeting is planned during March in Prague.

5.5 DSP

			Notes
Format of descriptive metadata	CHR	CEI schema	M-Tool will be used to create all the metadata for MAN / INC.
	MAN	MASTER, event. TEI P5	
	INC	MASTER, event. TEI P5	
Format of metadata about structure	CHR	CEI schema	M-Tool will be used to create all the metadata for MAN / INC.
	MAN	MASTER +	
	INC	MASTER +	
Numbers of available documents	CHR	40000	
	MAN	305	
	INC	386	
Samples handed over	CHR	yes	CHR – complete samples as a basis for next work (WP4, WP5); MAN and INC for guidance (Not all the levels).
	MAN	partially	
	INC	partially	
Repository	CHR yes		Ready for CHR, In preparation for MAN/INC
OAI interface	OAI-PMH is planned for needs of intersystem communication for the CHR		
Other information	Only digitalized documents are going to be imported from DSP into MNS.		

5.5.1 Metadata - Generally

In the case of CHR we have at our disposal a representative sample suitable for following WP5.

The insufficiency of the minimal record (from the point of view of ENRICH) is as follows:

- It contains only "idno" information – it means that it does not contain any identification information enabling to set up the identification information in the way described in the paragraph "Evidential Record" (at least library name, settlement, idno, country) . This type of extreme minimally record will be not able to be imported into the MNS platform
- It does not contain any information enabling the creation of URL for images.

Samples in MS Word for MAN a INC is the material for WP5 for a possible discussion about a set of information writable in future M-Tool.

M-Tool will have to make possible the adding definite information for automatic classification according to file types, and CHR will have their own connector.

5.5.2 Metadata about Structure

Metadata about structure CHR are fully suitable, and in accordance with principles of import of documents into MNS. URL of images of CHR documents:
http://www.monasterium.net/pics/id_collection/file_with_image

where:

http://www.monasterium.net/pics/ is the root for all the documents
id_colection is the identificator (which we can not see in the samples!)
file_with_image is the information directly contained in metadata:
//witList/witness/figure[not(figDesc)]/graphic/@url

The absence of id_colection in samples handed over is the last fact preventing the closing of DSP / CHR like a repository ready for the cooperation with MNS.

Metadata MAN and INC will be created using M-Tool, but at the time we have no information about a repository structure enabling contingent future corrections.

5.5.3 Planned steps in T2.3

- To find out id_colections in metadata.
- To gather a final documentation for the planned repository for INC and MAN.
- To agree when will be prepared
    - the OAI interface CHR for ENRICH
    - the repository for INC / MAN

5.6 KU

			Notes
Format of descriptive metadata	MASTER		
Format of metadata about structure	Information about structure does not exist.		The intention is to use the element "locus" in a MASTER/TEI P5 description.
Numbers of available documents	MAN	100	
Samples handed over	yes		
Repository	It does not exist.		
Other information	We can expect that KU will import into MNS also separate catalogue records for not yet digitalized documents.		

5.6.1 Descriptive metadata

Metadata contain information enabling the classification according to the type of the document also in other collections:
`objectDesc/@form`
and
`msItem/@class`

5.6.2 Metadata about structure

The intention to use "locus", concretely the attribute "`locus/@facs`", seems to be a realistic solution from the technical point of view - especially if the information from the rest of the record enables to create an URL. We have to verify this as soon as KU makes a progress in the preparation of the repository.

5.6.3 Repository

For data management the tool Fotoware will be maybe used.

5.6.4 Planned Steps in T2.3

1. The way on how to import the metadata into the Manuscriptorium will be agreed based on results of discussion on possible replacement of the descriptive MASTER DTD part of the MASTER+ format (used in Manuscriptorium as an internal format) with the more advanced TEI P5 DTD.
2. During T2.3 a description of the data repository that will be in conformance with the Manuscriptorium import principles should be created and also a description on the way of how the structural metadata will carry all necessary information should be created. Regarding the repository AIP will be at disposal for consultations if required.

5.7 NULI

			Note
Format of descriptive metadata	MAN	MARC 21	NULI is going to be converted into TEI P5 during the project
	PRN	MARC 21	
Format of metadata about structure	MAN	Do not exist	
	PRN	We do not know	
Numbers of available documents	MAN	1100	
	PRN	452	
Samples handed over	MAN	yes	In the case of MAN we have 3 levels of records at our disposal, in the case of PRN just one.
	PRN	yes	
Repository	Does not exist		
Other information	Only digitalized documents are going to be imported from NULI into MNS.		

5.7.1 Descriptive metadata

We will not know the definitive form of MAN until NULI will be converted into this format after WP3. We can derive the structure of PRN according to the sample. Differences in comparison with the standard MARC 21:

- The MARC format has been adapted to accommodate special Icelandic usage of personal names in tags 100, 600 and 700.
- In tag 03X there is 039 that indicates record for www.saganet.is.
- Last, subject words for www.saganet.is are in 690.

PRN can be the basis for the best-practice (it is brief enough), and the connector for NULI/PRN.
With respect to the existence of separate connectors for both of these collections the classification according to documents types will be executed automatically according to the used connector.

5.7.2 Metadata about structure

For MAN we suppose the same method like by KU.

For PRN we suppose that the record of in Marc language itself hardly can carry structural information. M-Tool enabling creation evidential records (for the definition of this term see above), and metadata about structure will be at the disposal for NULI.

At the same time a connector for descriptive metadata will have to be created. It will make possible to complement evidential records in MNS by detailed information gained from

MARC 21 records. The connector will also ensure the possibility of actualization of information MNS on the basis of changes executed in primary metadata in MARC 21.

As soon as the connector will exist, we will be able to import all bibliographical notes into MNS. During the process of work on M-Tool we will add to these records digital documents. Also an inverse process is possible – it depends on which way will be ready sooner.
By the means of M-Tool we can optionally create metadata about the structure for MAN – see the description by KU.

5.7.3 Repository

It is solved in the same way like KU.

5.7.4 The Planned Steps in T2.3

The same like KU.

5.8 SAM

			Notes
Format of descriptive metadata	MASTER		During the project SAM is going to be converted into TEI P5
Format of metadata about structure	Do not exist		
Numbers of available documents	MAN	400	
Samples handed over	yes		
Repository	Does not exist		
Other information	Only digitalized documents are going to be imported into MNS by SAM.		

Other information is similar like in the case of KU and NULI (see above).

5.9 ULW

			Notes
Format of descriptive metadata	DC		
Format of metadata on structure	DjVu		all documents (data and metadata) at present available using DjVu only
Numbers of available documents	MAN	87	
	music manuscripts	14	
	DRW (old drawings)	287	
	PRN	172	
Samples	yes		
Repository	yes		At present based on the format DjVu
Harvesting interface	http://www.bibliotekacyfrowa.pl/dlibra/oai-pmh-repository.xml		
Other information	Only digitalized documents are going to be imported into MNS by ULW.		

5.9.1 Metadata

Dublin Core can be harvested, and an id enabling to create an URL (directly the URL itself) is contained in dc:identifier.

The information about the document type is carried by an element dc:type.

The information about the shelf mark is probably in dc:source, it is missing in some documents or it is a part of another element. We will have to find this element.

5.9.2 The Integration into DjVu MNS

The usage of a proprietary format (DjVu) is generaly not in accordance with the principles of MNS and with attaching documents into MNS. A possibility of an alternate approach will be discussed.

5.9.3 Planned steps in T2.3

- To find a possible alternate way of sharing documents (preferably without the usage of DjVu) and prepare a description of the repository and the way of transfering necessary metadata on structure,
- To verify accessibility of information about shelf mark, which is necessary for the construction of identification in MNS.

5.10 UZK

			Note
Format of descriptive metadata	MAN	MASTER (modified)	During the project UZK is going to convert into TEI P5
	INC	ISTC	
Format of metadata about structure	MAN	MASTER	A part of MASTER
	INC	eBind	A separate file
Numbers of available documents	MAN	400	
	INC	1214	
Samples handed over	MAN		
Repository	exists		Fully described
Harvesting interface	Plans for introducing a profile for TEI P5 (MAN) and possibly ISTC		
Other information	Only digitalized documents are going to be imported into MNS by UZK.		

5.10.1 Descriptive metadata

UZK is going to realize transfers into TEI P5 itself.

UZK is ready to offer metadata of INC via OAI in the format which is going to be recommended within ENRICH or it is possible to prepare for the format ISTC in the a connector Manuscriptorium system (the second possibility was discussed recently and most probably will be used during the project; OAI interface will be used).

5.10.2 Metadata about structure and repository

UZK is ready to cooperate according to the principles of MNS.

5.10.3 The planned steps in T2.3

- To settle the possibility to use the OAI interface for testing the harvester of MNS (first DC will be used)
- To settle the way of connecting of INC (the final format for the connector)

5.11 VUL

			Notes
Format of descriptive metadata	UNIMARC		will migrate to MARC 21 during the project
Format of metadata about structure	Do not exist		
Numbers of available documents	MAN	1300	
Samples handed over	yes		Partially like screenshots of UNIMARC Records and also real exports
Repository	In preparation		
Other information	Only digitalized documents are going to be imported into MNS by VUL.		

5.11.1 Repository

Complies to the principles of MNS. The existing DjVu documents are also going to be ready for the repository using JPEG files.

5.11.2 Metadata

Note: Following way of combined use of M-Tool with an connector is possible. Based on the most recent information the M-Tool may be replaced by possible existing structural metadata processing.

The nonexistence of metadata about structure leads to the assumption to use M-Tool (it is not possible to use UNIMARC/MARC 21 to carry the necessary structural information). There are descriptive metadata which will not be re-created using M-Tool. M-Tool will be used for creating the most simplified records (with the basic identification information) with structural metadata. These records are going to be imported into MNS.

At the same time a connector for descriptive metadata will have to be created. It will make it possible to complement records in MNS produced by M-Tool by detailed information gained from MARC 21 records of VUL. The connector will ensure the possibility to update MNS accordingly to the changes performed above primary decriptiove MARC 21 metadata.

As soon as the connector will exist, it will be possible to import all the bibliographical records related to documents available for ENRICH into MNS. The structural metadata produced by M-Tool will update the content already available in MNS. Also an inverse process is possible – it depends on which way will be ready sooner.

Information about the document type exists in encoded information of the field LDR.

5.11.3 Planned Steps in T2.3

1. To define the final structure of the repository
2. To agree a timeframe when the repository will be ready for the ENRICH project
3. To check the existence of structural metadata and the possibility of its processing within the connector

6 The Classification of Partners according to the Way of Intersystem Communication

As soon as particular Content Partners fulfill "Criteria of Readiness Valuation of a Content Partner for Entering into Intersystem Communication" described in Chapter 5.1., they will be ready to collaborate with Manuscriptorium in one of the subsequent ways, according to that how they will transmit metadata into the Manuscriptorium system:

- Way of M-Tool
- Way of M-Tool + off-line connector
- Way of Off-line connector
- Way of On-line connector
- (Way of M-Tool + on-line connector)

The way mentioned in the parenthesis is not going to be realized by any of present Content Partners.

Yet today we can summarize on the basis of questioning what way is going to be used by particular partners, or better said, by particular collections.

The classification may be a subject of change in the case we will find a more suitable approach with the individual partners during T2.3.

6.1 The Way of M-Tool

All types of metadata – both structural and descriptive – are created manually for concrete documents using the M-Tool application.

In its new version it will be adapted to new demands in the area of produced structural and descriptive metadata. In the area of descriptive metadata we count with necessary corrections. This need will be caused by specific properties of various types of described documents. In the area of structural metadata we count with necessary corrections caused by different praxes of the repository organization by particular partners.

Some partners will be able to use the existing version of the M-Tool application immediately; some other will have to wait for a new on-line version – the result of T5.1 (it depends mainly on the final repository organization of partners/collections and on demands on the cataloguing of various types of documents).

The work on M-Tool will be consequently followed by the work on the M-Can application which enables to verify the final digital documents, and to offer them for the final import into MNS.
This way is going to be used by:

- BUTE
- DSP – MAN
- DSP – INC

6.2 The Way of M-Tool + off-line Connector

This way is similar like the previous one, and it is suitable for partners who have a larger amount of catalogue records which substantiate the creation of connectors. The use of the connector will make possible to connect directly with the partner's source of primary information, and will not be necessary to create duplicate descriptive metadata in M-Tool. M-Tool will be used for the creation of structural metadata + descriptions in the extent of the evidential record (for the definition see above) meanwhile the detailed description is processed by the connector..

In MNS it is possible to supply the outputs of M-Tool equipped by these brief descriptions by identifiers. These identifiers enable to join the document with descriptive information processed by the connector.

The connector is going to be optimised and adapted to the properties of input metadata to provide – where possible – zero-loss conversions.

Descriptive metadata will be transferred into the input of the connector by the way agreed later within T2.3, probably by the means of FTP, where Content Partners will upload exports of their metadata.

Structural metadata with evidential records will be checked out, and transmitted in the same way like in the case of the Way of M-Tool.

This way is going to be used by:

- BNE
- VUL

6.3 The Way of Off-line Connector

This way is suitable for partners who manage existing structural and descriptive metadata for such an amount of documents which rationalise the creation of the connector.

The connector is going to be optimised and adapted to the properties of input metadata to provide – where possible – zero-loss conversions.

Metadata will be transferred into the input of the connector by the way agreed later within T2.3, probably by the means of FTP, where Content Partners will upload exports of their metadata.

This way is going to be used by:

- KU
- NULI
- SAM

6.4 The Way of On-line connector

It is a similar way like the Off-line connector, but this one uses OAI-PMH for communication. At present all the partners having the OAI interface implemented will contribute with a relevant amount of documents rationalising the creation of connectors.

This way is going to be used by:

- BNCF
- DSP – collection CHR
- ULW
- UZK

7 The Content Description of Particular Collections

BNCF

Document type: manuscripts, handwritten maps
Number of provided documents: over 53 636
Number of images: ca. 522 842
Date (time interval from- to-): 10th – beginning of the 19th century
Short description of content:

1) **Manoscritti In Rete**: The set contains the rarest and most consulted manuscripts owned by the BNCF (i.e. the Messale Ottoniano of the X Century, which is the oldest manuscript in the BNCF, the Palatino 556, also named Lancelot, Filarete's treatises on architecture, etc.).

2) **Bibliotheca Universalis II**: The set contains manuscripts of English and French travellers in Tuscany from the end of the XVII Century to the beginnings of the XIX Century. The manuscripts concern subjects related to the *Grand Tour* theme.

3) **Carte Geografiche II**: The set is composed by 250 handwritten maps and portolani (XV-XVII) Century, containing the handwritten maps of the cartographer Luigi Giachi.

4) **Bibliotheca Universalis I**: The same subjects as in **Bibliotheca Universalis II** but related to printed books (between the bibliographical units are included 5 periodicals for about 4300 issues)

5) **Carte Geografiche I:** The set is composed by printed geographical maps, charts and military maps (XVII-XIX Century) belonging to the Palatino antiquarian collection (that is one of the two main historical collections –the other one is the Magliabechiana- from which the BNCF was born).

6) **Magliabechi:** As remarked above the Magliabechi historical collection (together with the Palatino collection) constitutes the historical beginnings of the BNCF. It consists of printed books, mainly of XVI, XVII and XVIII century: of this historical collection were digitized and electronically catalogued more than 50000 bibliographical units corresponding to about 1/3 of the whole original size of the collection. The digitizations in general concern the oldest books, but are not complete digitisations, since regard cover, title page, table of contents and variable significant pages of each one of the partially digitized volumes. Catalographic records and relative images are accessible via OPAC (i.e. you can recover them by a search using the field shelfmark = Magl and then selecting the digitized schedule).

7) **Galileian Collection**: A particular consideration must be settled on the galileian collection. The BNCF holds one of the most important collections of manuscripts for the history of science, containing almost all the manuscripts of Galileo Galilei and a relevant number of manuscripts of Galilei's followers (Castelli, Cavalieri, Torricelli, Viviani), their correspondence, and the correspondence with the most important European scientists of the time (Mersenne, Gassendi, Stenone, Peiresc). The collection was restored, catalogued (even though with not complete uniformity) and digitized: it consists of more than 80000 images (the exact number is not available). Actually it is not accessible by OPAC and only a minimal part (15 units) is accessible from the old version of the BNCF website using a detailed but complex and old system of queries: http://alpha9.bncf.firenze.sbn.it/cgi-galileo/makeQuery.cgi?rigamenu=Galileo. A goal of the ENRICH project could be the implementation of this resource inside the structure of the actual digital collections of the Library, producing accessibility via OPAC and by the OAI interface.

***Conclusion:* From content point of view the partner is fully able to be integrated into the Manuscriptorium**

BNE

Document type: manuscripts, historical maps, early printed books, prints, drawings, engravings, photographs
Number of provided documents: 3203
Number of images: over 1 000 000
Date (time interval from- to-): to 1936/1939
Short description of content:

1) **Obras Maestras [Masterworks]:** The most important works holded at the National Library of Spain. It includes manuscripts, prints, drawings, engravings, maps, photographs, posters and music from the IXth to the XXth century.

2) **Hispanoamérica:** Chronicles, dictionaries (historical and biographical), geographic descriptions and studies about the missions in Hispanic-America and the South of the United States. Prints from the XVth to the XXth century.

3) **Mapas de España [Maps of Spain]:** Atlas and maps from the holdings of the Department of Maps of the National Library. The ones of the Mendoza collection as well as the manuscripts ones are of special interest.

4) **Hemeroteca Digital [Newspaper Digital Library]:** Historical press published between 1772-1933 (more than 500.000 pages digitised)

Other:

5) **Engravings:** The collection titled Iconografia Española [Spanish Iconography] is made up of engravings, drawings and photographs of relevant Spanish celebrities from the XVth to the XIXth century as well as German engravings of the XVth and XVIth centuries, standing out the works of Durero and some of the most important German painters of his time. The perfect technics used, both xylographic and chalcographic, and the quality in the representations of the diferent subjects treated are of special importance.

6) **Niños de la guerra [War Children]**: 1171 drawings made by children during the Spanish civil war.

***Conclusion:* From content point of view the partner is partially able to be integrated into the Manuscriptorium**

BUTE

Document type: incunabula, early printed books (chapters) and journals (articles)
Number of provided documents: ca. 130
Number of images: ca. 5000
Date (time interval from- to-): ca. 1450-1800
Short description of content: As a minimum we will provide 5000 pages submitting chapters from early books published in the 15th-17th centuries, and articles from an early Russian scientific journal from the 18th century. Additional documents are still under selection on the basis of historical value, the number of images contained, interest of potential users. The languages will be Latin, Hungarian, French, German. The subject fields will include engineering sciences, natural sciences (Mathematics, Physics, Biology) , History, Law, general culture (e.g. cook book), Theology

***Conclusion:* From content point of view the partner is fully able to be integrated into the Manuscriptorium**

DSP

Document type: charters, manuscripts, incunabula
Number of provided documents: over 40 691
Number of images: over 100 000
Date (time interval from- to-): to ca. 1600
Short description of content:

1) The collections selected for ENRICH are part of a larger collection of charters. The collections of available documents can only be enumerated. We plan to build a OAI-set of collections available for ENRICH identified by the setName "ENRICH". The approximately 40.000 charters to be part of Manuscriptorium originate to a great extent from monastery archives in Lower Austria, Vienna and Upper Austria but there are also charters kept by the state and federal state archives which come from former monasteries. They range chronologically from the 9th to the 18th century and are important sources for the early history of the mentioned regions. Detailed information on each object of the collection is available, including images and/or secondary data like text summaries, full texts and glossaries.

2) The collection of manuscripts covers 300 books from the beginning of the 13th to the 19th century – about 120 dating from the Middle Ages – and represents an important holding of the Diocese Archives St. Pölten. They originate from the former Augustinian monastery in St. Pölten, from several parishes and other monasteries in Lower Austria and consist especially of Biblica and Liturgica. Most of them are richly illustrated (e. g. Hs 1, Antiphonar ~ 1486 which contains illustrations of monks of St. Pölten) and therefore of historical and art historical value.

3) Furthermore the Diocese Archives keeps 386 incunabula and early printed books which mainly cover liturgical, historical, philosophical and canon law issues; they date from the 1470s to the 16th century. Over 270 incunabula are in their original binding which make them singular for scientific studies. – The descriptions of the manuscripts and incunabula are not published yet; together with the images ENRICH will make this data accessible for the first time.

***Conclusion:* From content point of view the partner is fully able to be integrated into the Manuscriptorium**

KU

Document type: manuscripts
Number of provided documents: ca. 100
Number of images: ca. 2000
Date (time interval from- to-): 12th century – 18th century
Short description of content:

The Arnamagnæan Manuscript Collection, named after the Icelandic scholar Árni Magnússon (1663-1730), comprises some 3000 items, the earliest dating from the 12th century, the latest from the 18th century, the majority of them Icelandic. Just over half of these items were transferred to Iceland in the period 1971-1997. Remaining in Copenhagen are approximately 1400 manuscripts and fragments. About half of these are Icelandic and include histories of the kings of Norway and Denmark, religious texts or translations from Latin and other languages. The other half consists of Danish, Norwegian and Swedish manuscripts, along with approximately one hundred of continental provenance. We hope to have XML catalogue records for all these items by the end of the project period.

***Conclusion:* From content point of view the partner is fully able to be integrated into the Manuscriptorium**

NULI

Document type: manuscripts, early printed books, secondary printed books
Number of provided documents: 1552
Number of images: 373 000
Date (time interval from- to-): to ca. 1900
Short description of content:

Saganet gives Web access to digital images of about 220.000 manuscript pages and 153.000 printed pages. The material consists of the entire range of Icelandic family sagas. It also includes a very large portion of Germanic/Nordic mythology (the Eddas), the history of Norwegian kings, contemporary sagas and tales from the European age of chivalry. A great number of manuscripts contain Icelandic ballads, poetry or epigrams. Most manuscripts, on vellum and paper, and printed editions and translations of the Sagas as well as relevant critical studies published before 1900 are included.

***Conclusion:* From content point of view the partner is fully able to be integrated into the Manuscriptorium**

SAM

Document type: manuscripts, charters
Number of provided documents: 400
Number of images: ca. 45 000
Date (time interval from- to-): 12th – 19th century
Short description of content: The Arnamagnæan Manuscript Collection, named after the Icelandic scholar Árni Magnússon (1663-1730), comprises some 3000 items, the earliest dating from the 12th century, the latest from the 19th century, the majority of them Icelandic. Just over half of these items were transferred to Iceland in the period 1971-1997 and are kept in the Stofnun Árna Magnússonar í íslenskum fræðum (SAM). The other half that remains in Copenhagen is kept in Den Arnamagnæanske Samling (KU). Of the 3000 codices and fragments of codices in the original collection some 1600 were repatriated to Iceland along with some 1300 charters, 5800 apographs of charters and 140 manuscripts that had been kept in the Royal Library in Copenhagen; to that we can add some 100 manuscripts that have been bought or donated to the Institute in Reykjavík in the past 40 years. As several of the 1600 manuscripts consist of fragments from more than one original codex the number of manuscript records will in due course not be 1600 but around 2400. We have now some 1600 MASTER descriptions on different levels of completion. Within ENRICH I expect that we will more or less work within that number. We want to transfer them all into TEI-P5 format, add information to the minimum detailed descriptions and link them to images as far as our number of present available images plus the images made within the next two years will allow us. The processing of the images will cost us considerable work. Of course we would like to have at least minimal XML catalogue records for all our manuscripts by the end of the project period, but the workload we can manage within the timelimit will decide how much we can do.

Conclusion: **From content point of view the partner is fully able to be integrated into the Manuscriptorium**

ULW

Document type: manuscripts, early printed books
Number of provided documents: 560
Number of images: ca. 70 000
Date (time interval from- to-): to 1800
Short description of content: The digital collection of ULW contains at the moment 87 manuscripts, 14 music manuscripts, 287 old drawings, and 172 early printed books. It is a selection of the most valuable physical collections of the ULW that is the largest historical manuscript library in Poland.

Conclusion: **From content point of view the partner is fully able to be integrated into the Manuscriptorium**

UZK

Document type: manuscripts, incunabula
Number of provided documents: 2534
Number of images: 584 500
Date (time interval from- to-): to ca. 1600
Short description of content:

1) **CEEC**: Systematic and complete digitization of all medieval codices available in the Erzbischöfliche Diözesan- und Dombibliothek Köln, augmented by ca. 20 codices form various libraries within the bishopric.

2) **vdIb**: Platform to provide a common interface to different collections of incunabula. Currently hosting incunabula form 3 libraries: Universitätsbibliothek Köln, Erzbischöfliche Diözesan und Dombibliothek Köln, Herzog August Bibliothek Wolfenbüttel; in preparation: Universitätsbibliothek Mannheim.

***Conclusion:* From content point of view the partner is fully able to be integrated into the Manuscriptorium**

VUL

Document type: charters, manuscripts, historical maps, graphics, drawings
Number of provided documents: ca. 1300
Number of images: ca. 3000
Date (time interval from- to-): to 19th century
Short description of content:

1) **Parchments**: The collection of parchments now belonging to the Manuscripts department of the Vilnius University Library is not big – it consists of less than hundred units. These are remains of the collection of parchments, that once belonged to the former Public Library of Vilnius (larger part of it is now stored at the Vilnius Library of the Science Academy), also documents, that were displaced from the Kaunas University, or came to our library in some other ways. Among our parchments there are brevets on land property and other muniments signed by the Polish Kings and Grand Dukes of Lithuania (Sigismund the Old, Sigismund Augustus, Zygmunt Vaza, Stanislas August and others); also documents of land property and buildings sale, rescripts sustaining or ratifying privileges given by previous rulers. There are also some documents, verifying knighthood of individual persons, a number of Pope‘s bulls, pardons, nominations, the documents of property of some monkhoods and other.

2) **Manuscripts**: Vilnius University Library Manuscript Department contains 543 books of court acts of the Grand Duchy of Lithuania of 1540-1845. The county courtbooks of the Grand Duchy of Lithuania take a special place among the collections of VU Manuscript Department. In the middle of XVI c. the administration of the Grand Duchy of Lithuania was reformed, and the county courtbooks became the most important source for investigation of the history of Lithuania, Poland and Byelorussia. Practically the county courtbooks are like metrics, i.e. inventories of documents of separate Lithuanian regions, as any significant document drawn up in manors was necessarily included in those metrics. The accumulated information evidences the activities of courts and reflects a broad political, social and economic panorama of Lithuanian history, mode of life of past generations, public and

individual psychology. Richness of the accumulated information allows us to research these unique directories that disclose changes of life standard. These books are important as a source of material culture, and as documents that supplement other sources. Chronologically broad documentary material is interesting and important to the history of language, historical geography and genealogy research. Books are written in the old Byelorussian and Polish languages.

3) **Cartography collection**: V. Mincevičius (1915–1992) – a priest, journalist, translator and collector, who had lived in Italy (and was buried there), donated a collection of maps (332 map items/317 titles) to the University Library. The greater part of the collection consists of old cartography. There are maps and city plans of exceptional value: XVI-XIX c. world, Europe, different regional maps and city plans created by eminent authors and publishers such as C. Ptolemaeus, W.J.Blaeu, G. Mercator, S. Münster, A. Ortelius, J. Hondius and others. The map collection is an integral and perfectly formed collection mostly representing the Kingdom of Poland and the Grand Duchy of Lithuania with neighbouring countries and states, there are also several European states and city plans as well as masterly facsimile reproductions (18 items.)

4) **Pieces of graphic art**: One of the most notable portrait engravers of the XVIII c. Lithuania was Hiršas Leibovičius (Herszek Leybowicz), an artist at the Radziwill family estate in Nesvizh, the Grand Duchy of Lithuania. Throughout 1745-1758 he re-created on copper plate engraving one hundred and sixty-five portraits of the Radziwill family from the picture gallery of the Nesvizh Castle. In 1758 M.F. Wobe published an album named "Images of the noble family of Radzivill" ("Icones Familiae Ducalis Radvilianae"). This album was reissued in Peterburg in 1875. Separate portrait pages printed in Peterburg are held in Vilnius University Library.

Conclusion: **From content point of view the partner is fully able to be integrated into the Manuscriptorium**

All ENRICH full Content Partners (with exception of NKP):

Number of provided documents: **over ca. 104 106**
Number of images: **over ca. 2 705 342**

Conclusion: **All ENRICH Partners (with exception of NKP) are from content point of view able to be integrated into the Manuscriptorium, most of them fully, some partially.**

Documents provided by NKP:

Number of provided documents: **over ca. 2 836**
Number of images: **over ca. 1 100 000**

Conclusion: **From content point of view fully able to be integrated into the Manuscriptorium**

All documents provided within the ENRICH project:

Number of provided documents: **over ca. 106 942**
Number of images: **over ca. 3 805 342**

Conclusion: **All ENRICH Partners are from the content point of view able to be integrated into the Manuscriptorium, most of them fully, some partially**

8 Conclusions

This document describes the actual state in m3. Within the ENRICH project the working paper of this document is going to serve for description of the actual state updated and agreed during T2.3. At the end of T2.3 this document will become the final documentation of the agreed Intersystem Communication with particular partners. Collected samples and particular technical information serving within WP4 and WP5 will also be a part of this final documentation for Intersystem Communication.

8.1 Metadata Samples

From the beginning of the process of questioning we try to collect representative metadata samples, and we have several reasons why to do it:

- They enable to verify information provided in questionnaires, and eliminate potential misunderstandings
- They are an important source of information for WP4 and WP5 (optimalization for end-users + optimalization for users–partners, it means M-Tool, convertors, best-practice)

These are the reasons why the representative samples are going to be further demanded.

8.2 External Technical Documentation

As soon as the preparation phase of T2.3 will be finished by all the partners, all the technical information from questionnaires and subsequent communication will be summarized in relevant technical documents creating a part of the final documentation of Intersystem Communication.