



**EUROPEANA REGIA**  
**CIP- ICT- PSP 2009 - 250560**

**D4.1 State of the Art in Image Processing**

<b>Deliverable number</b>	D4.1
<b>Nature</b>	SP
<b>Dissemination level</b>	CO
<b>Delivery date</b>	Due Date : 30 June 2010, extended to 15 May 2011 Actual Date : 12 May 2011
<b>Status</b>	<b>Final</b>
<b>Work package number</b>	WP4
<b>Lead beneficiary</b>	BSB
<b>Author(s)</b>	Rahel Bacher, Claudia Fabian, Wolfgang-Valentin Ikas, Carolin Schreiber (BSB)



## Document History

### Revisions

Version	Date	Author	Changes
0.1	13/05/2010	Wolfgang-Valentin Ikas, Claudia Fabian	Created initial version
0.2	01/07/2010	Wolfgang-Valentin Ikas	Input from the partners
0.3	27/07/2010	Wolfgang-Valentin Ikas, Claudia Fabian	Corrections from T. Delcourt (BnF)
1.1	28/01/2011	Rahel Bacher, Claudia Fabian, Wolfgang-Valentin Ikas, Carolin Schreiber	New version as required by the Reviewers
1.2	15/04/2011	Rahel Bacher, Carolin Schreiber	Input from other digitising institutions (questionnaire)
1.3	21/04/2011	Rahel Bacher, Carolin Schreiber	Input from colleagues at the BSB (Markus Brantl and Irmhild Schäfer)
1.4	26/04/2011	Rahel Bacher, Carolin Schreiber	Input from the partners (Elisabeth Freyre, BnF, and Stefanie Gehrke, HAB)
1.5	29/04/2011	Rahel Bacher, Wolfgang-Valentin Ikas, Claudia Fabian, Carolin Schreiber	Input from the IFLA Standing Committee on Rare Books and Manuscripts
1.6	11/05/2011	Rahel Bacher, Carolin Schreiber	Input from T. Delcourt (BnF)

### Reviews

Date	Name	Result

### Signature/Approval

Approved by (signature)	Date

Accepted by at European Commission (signature)	Date

## 1 Executive Summary

This deliverable is dedicated to summarizing the present procedures applied by *Europeana Regia* project partners in order to harmonize and adapt them among the partners. The BnF, BSB, and the HAB have been able to gather a considerable amount of knowledge based on experience and are thus able to assist the BHUV and the KBR, to which some procedures are still rather unfamiliar as they began digitisation more recently. Their experience is complemented by the results obtained in a questionnaire sent to leading institutions in the field of digitisation, which was responded to by 14 libraries in Europe and the United States.

## Table of Contents

<b>1</b>	<b>Executive Summary .....</b>	<b><u>44</u></b>
	<b>Table of Contents .....</b>	<b><u>55</u></b>
<b>2</b>	<b>Introduction .....</b>	<b><u>77</u></b>
<b>3</b>	<b>Presenting the Project Partners in <i>Europeana Regia</i>.....</b>	<b><u>1040</u></b>
<b>4</b>	<b>Comparable International Projects .....</b>	<b><u>1343</u></b>
<b>5</b>	<b>Specific Aspects of <i>Europeana Regia</i> .....</b>	<b><u>1545</u></b>
<b>6</b>	<b>Aspects of Image Quality in Manuscript Digitisation .....</b>	<b><u>1646</u></b>
6.1	General Aspects: Optical Quality Parameters .....	<u>1646</u>
6.1.1	Resolution.....	<u>1646</u>
6.1.2	Colour Depth.....	<u>1747</u>
6.2	General Aspects: Relation to the Original.....	<u>1848</u>
6.2.1	Colour and Tone .....	<u>1848</u>
6.2.2	Dimensions .....	<u>2049</u>
6.2.3	Completeness.....	<u>2020</u>
6.2.4	Sequence.....	<u>2120</u>
6.3	Aspects Peculiar to Manuscript Digitisation .....	<u>2124</u>
6.3.1	Due to the Physical Condition of the Manuscript .....	<u>2124</u>
6.3.2	Appropriate Backing of Translucent Materials .....	<u>2222</u>
6.3.3	Conservation Aspects and Digitisation .....	<u>2322</u>
6.3.4	Other Factors Peculiar to Manuscripts.....	<u>2423</u>
6.4	Effects of these Factors .....	<u>2525</u>
6.5	Summary: Recommended Quality Parameters for the Digitisation of Manuscripts.....	<u>2726</u>
<b>7</b>	<b>Workflow and Technical Aspects of Manuscript Digitisation.....</b>	<b><u>2827</u></b>
7.1	Digitisation Inhouse v External Service Providers.....	<u>2827</u>
7.2	Condition Survey of Manuscripts .....	<u>2927</u>
7.3	Preparation of Manuscripts for Digitisation (Step 1).....	<u>3028</u>
7.4	Digitisation Process/ Scanning or Photography (Step 2) .....	<u>3029</u>
7.5	Long-Term Preservation (Step 3) .....	<u>3330</u>
7.6	Display and Integration of Images (Step 4).....	<u>3633</u>
7.6.1	Display Formats.....	<u>3633</u>
7.6.2	Reusability in other Media .....	<u>3734</u>
7.6.3	Integration of Images .....	<u>3734</u>
7.6.4	Persistent Identifiers (PI) .....	<u>3936</u>

<b>8</b>	<b>Relation of Images to Metadata.....</b>	<b><u>4138</u></b>
8.1	Different Types of Metadata.....	<u>4138</u>
8.1.1	Bibliographical Metadata .....	<u>4138</u>
8.1.2	Technical Metadata .....	<u>4138</u>
8.1.3	Structural Metadata .....	<u>4542</u>
8.2	Creation of Links to Metadata .....	<u>4643</u>
	<b>Appendix: Naming conventions of Partners: .....</b>	<b><u>4845</u></b>
	<i>BSB</i> .....	<u>4845</u>
	<i>BnF</i> .....	<u>4845</u>
	<i>KBR</i> .....	<u>4845</u>
	<i>HAB</i> .....	<u>4845</u>
	<i>BHUV</i> .....	<u>4946</u>
	<b>Literature .....</b>	<b><u>4947</u></b>

## 2 Introduction

### *Document History*

This paper was first submitted one month after the start of the project, summing up the discussions during the project kick-off-meeting on 18 and 19 January 2010. The aim of this first version was to collect and transmit to *Europeana Regia* the experience that had been gained by the project partners in other projects. Results from the work on *Europeana Regia* could not be reported because at this stage the project had just started. After one year, a revised version of this first paper is now presented, which incorporates the experience gathered over the first twelve months of the project. At the same time, the expertise of other leading institutions which have carried out a substantial number of digitisation projects will be taken into account. To achieve this, a questionnaire was developed and sent to about 20 leading institutions in the field of manuscript digitisation worldwide ('*ER* questionnaire', see list below, pp. 48-9). The answers given by these co-operating institutions are included in the corresponding sections of this paper. A first draft of this version was discussed on 27-29 April in Madrid at the mid-term meeting of the IFLA Standing Committee on Rare Books and Manuscripts, which was dedicated to the production of IFLA guidelines for the digitisation of rare books and manuscripts.

### *Challenges of Digitisation in General*

- Digitisation is a relatively new method.
- Digitisation is quite expensive.
- It requires technical expertise; few libraries are experienced in all aspects. Experienced partners and competence centres are needed.
- For the time being only large libraries are in a position to purchase their own equipment, train specialised staff and ensure the long-time preservation of digital material.
- Digitisation depends on the state of technology at a given time. It is obvious that in the future, technical advances will lead to better results; technical equipment may quickly become outdated. It is therefore very important to be aware of the state of the art and technical developments in order to obtain the best possible results.
- Expectations of users rise and today they include presentation on the World Wide Web, colour images with high image resolutions and exact colour values, structural metadata and perhaps a transcription of texts.

### *Challenges of Digitisation in Manuscript Projects*

*Europeana Regia* is a project that deals exclusively with digitisation from original medieval manuscripts, the earliest of which date from the eighth and ninth centuries. Compared to printed materials, the digitisation of medieval manuscripts poses specific and challenging problems in a number of areas.

The most important difference to printed materials is that manuscripts are unique historical objects, whereas printed sources are in principle available in a number of originally identical copies. Even in cases where there are several manuscripts that transmit the same text, each single copy is a unique manifestation of its appropriation and contextualisation in a specific historical situation, as well as a specific material object (with, for example, peculiar script, paper, decoration and layout). It is therefore not only the content of a given manuscript that is interesting to scholars and to the public, but also the unique characteristics and history of the codex, its early ownership, later history and provenance, as documented, for example, by

ownership marks, annotations or its specific bindings. This unique character of medieval manuscripts alone justifies the production of high-quality secondary formats. Only a small part of the corpus of medieval manuscripts has been published in print and so for many manuscripts digitisation is the first publication.

For a number of scholarly disciplines, the analysis of the fine details of the physical make-up of manuscripts is indispensable. Art historians, palaeographers, historians and linguists depend on the availability of high-resolution images tailored to their needs. For the same reason, colour images are needed in these disciplines, for example for the study of illuminations and initials (art history), various colours of ink that help to differentiate various hands of scribes (palaeography), or different shades of parchment or paper which help to understand the composition of the book (codicology).

These requirements must be taken into consideration when deciding on the parameters of image quality for the digitisation of medieval manuscripts. In some cases special techniques such as multi-spectral imaging will be necessary to capture details that cannot be detected by the human eye alone. The exchange of information about these relatively new techniques will enable the partners to enlarge their repertoire.<sup>1</sup>

Special technical challenges caused by their material have to be addressed as well. Early medieval European manuscripts were written not on paper but on parchment, and are often bound very tightly in heavy wooden covers. It may therefore be impossible to open manuscripts up to an opening angle of 180°. The surface of the parchment may show distortions and convolutions which must be reproduced. Preservation aspects play an important role and have a major impact on the digitisation process itself. The qualities and characteristics of these unique originals may vary within one and the same project and may require individual treatments and processes. The digitisation of medieval manuscripts must therefore necessarily differ from the mass digitisation of printed modern material.

In the process of manuscripts digitisation, the requirements for image quality must be carefully weighed against the demands of manuscript preservation and conservation. Developing the optimum compromise between these two aspects is essential.

### *Aim and Scope of this Report*

The topic of this report is to summarize the present procedures applied by the *Europeana Regia* project partners in order to harmonize and adapt them among the partners. The BnF, BSB, and the HAB have been able to gather a considerable amount of knowledge based on experience and are thus able to assist the BHUV and the KBR, to which some procedures are still rather unfamiliar as they began digitisation more recently.

Most digitisation projects are not only funded by the holding institutions themselves but also by third parties. These are often the Ministries of Culture or other national agencies for the support of scientific projects and infrastructure. In recent years, several funding organizations in member countries of the European Union have summarized the experience gained so far in the field of digitisation and have issued guidelines in which detailed quality parameters are defined. Most of these digitisation standards were influenced by the 'Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives', issued in 2002 by the IFLA (International Federation of Library Associations) and ICA (International Council on Archives), which had been invited to draft these for UNESCO.<sup>2</sup>

In Germany, the most important funding organisation for library digitisation projects is the DFG (Deutsche Forschungsgemeinschaft, German Research Association). For projects

---

<sup>1</sup> See below, pp. 24-25, on infrared photography to capture watermarks.

<sup>2</sup> <<http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf>>



funded by the DFG there are mandatory parameters for digitisation which are specified in detail.<sup>3</sup> There seems to be a high awareness of these standards in Germany, since they were quoted in all answers to the *ER* questionnaire from this country. In Spain, there are guidelines for contracts with service providers, issued by the Subdirección General de Coordinación Bibliotecaria.<sup>4</sup> In addition, there is a regional standard for the members of the Consortium of University Libraries in Catalonia.<sup>5</sup> It looks as if in other member states of the EU national or regional standards for digitisation either do not exist or are not as widely known.

The recommendations contained in these standards (and their counterparts in countries outside the EU)<sup>6</sup> will be included and summarized in the corresponding sub-chapters of this report.

Apart from the identification of problems and limitations of digitisation projects, the aim of this paper is to summarize important new insights and eventually to suggest best practice for manuscript digitisation. This will be achieved by intensive collaboration among leading European libraries with expertise in digitisation.

---

<sup>3</sup> Cf. Deutsche Forschungsgemeinschaft, Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS): *DFG-Praxisregeln ‚Digitalisierung‘* (April 2009), available at <[http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln\\_digitalisierung.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung.pdf)>. The *DFG-Praxisregeln* are updated regularly; the current version is authoritative.

<sup>4</sup> Available at <<http://contrataciondelestado.es/wps/portal/plataforma>>.

<sup>5</sup> ‘Estándares de digitalización: requerimientos mínimos’, available at: [http://www.recercat.net/bitstream/2072/97497/3/1010EstandardsRequerimentsMinims\\_Public\\_spa.pdf](http://www.recercat.net/bitstream/2072/97497/3/1010EstandardsRequerimentsMinims_Public_spa.pdf)

<sup>6</sup> Similar national guidelines exist in the United States, issued by the U.S. National Archives and Records Administration (NARA) *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*, available at <<http://www.archives.gov/preservation/technical/guidelines.pdf>> and in Switzerland (,‘Freiburger Richtlinien’: *Richtlinien für die Digitalisierung mittelalterlicher Handschriften*, herausgegeben vom Kuratorium für die Digitalisierung mittelalterlicher Handschriften (Codices Electronici Confoederationis Helveticae)).

### 3 Presenting the Project Partners in *Europeana Regia*

Some of the partners had gained valuable experience through previous projects. Because this experience is the basis for the *Europeana Regia* project, a selection of these projects will be briefly described:

#### *Bibliothèque nationale de France (BnF):*

- The BnF offers access to the 'Gallica' Digital Library. Started in 1997, today it contains more than 820,000 digitised objects: more than 115,000 books, 3,500 journals (= 530,000 volumes), 110,000 drawings, 5, 000 maps, 4, 000 manuscripts, 2,000 sheets of music, etc.
- In 2009, a pilot project on Chinese and Central Asian manuscripts from Dunhuang was completed. The project was managed by the British Library in co-operation with the National Library of China.
- In 2007 a new programme for the digitisation of manuscripts was launched. It concerns every kind of manuscript material such as historic and illuminated manuscripts as well as literary manuscripts by important authors such as Marcel Proust, Gustave Flaubert and Victor Hugo. Every year 30,000 pages are digitised.
- A comprehensive programme started in 2008 has achieved the digitisation of the 150 manuscripts of the 'Roman de la Rose' preserved in French libraries.

#### *Bayerische Staatsbibliothek (BSB):*

In the BSB, the Munich Digitization Center (MDZ) founded in 2004 is responsible for digitisation projects for manuscripts and early printed material.

- VD 16 Digital. The aim of this project, which was started in 2006, is to digitise one copy of every sixteenth-century German edition (retrospective national bibliography). The ultimate scope of this project is about 40,000 volumes. The project started in 2006 with manually operated scanners and shifted in 2007 to the first project using automated book scanners (ScanRobots®) for digitising early printed books. This was carried out in a development partnership between the Munich Digitization Center (MDZ),<sup>7</sup> the Institute of Book and Manuscript Conservation and the service provider Treventus.<sup>8</sup>
- Incunabula: The BSB holds about 20,000 examples of 9,742 incunabula editions; one copy of every edition is currently being digitised with funding from the DFG.
- Digitisation of blockbooks, i.e. very rare books made from whole page copies of wooden boards, mainly from the fifteenth and early sixteenth centuries. This project is focused on the digitisation and in-depth cataloguing of 92 blockbooks available in Bavarian collections; it is funded by the DFG.
- Another project funded by the DFG deals with manuscripts from convents in Southern Germany. The digitisation and cataloguing, in combination with the scientific

<sup>7</sup> For the Munich Digitization Center (MDZ), see <<http://www.digitale-sammlungen.de/index.html?c=startseite&l=en>>; a list of publications in English is available at <<http://www.digitale-sammlungen.de/index.html?c=dokumente&l=en>>.

<sup>8</sup> Brantl, Markus / Ceynowa, Klaus / Fabian, Claudia / Meßmer, Gabriele / Schäfer, Irmhild: 'Massendigitalisierung deutscher Drucke des 16. Jahrhunderts – Ein Erfahrungsbericht der Bayerischen Staatsbibliothek', *Zeitschrift für Bibliothekswesen und Bibliographie* 56/6 (2009), pp. 327-338.

exploration of medieval manuscripts, is jointly carried out by BSB, Bayerisches Hauptstaatsarchiv and the Westfälische Wilhelms-Universität in Münster.

- Reichenau – Memory of the World: The manuscripts from the island convent of Reichenau in Lake Constance are particularly high-ranking documents with regard to their age and artistic quality. In 2003, ten manuscripts from the Reichenau (three of them owned by the BSB and seven by other libraries) were included in the World Document Heritage. The Memory of the World programme was initiated in 1992 by UNESCO in order to safeguard historically and culturally important documents and to make them available to the general public.
- Corvini: The eight Corvinus manuscripts in the BSB have also been included in the World Document Heritage programme.
- E-books on Demand: Complete digital copies of books from the period from about 1501 to 1900 can be obtained via eBooks on Demand. This service was initiated with EU support in thirteen libraries in eight European countries. After the termination of funding by the EU, document delivery now remains the responsibility of the BSB, which is enlarging the services.
- Digitisation on demand (DoD) for manuscripts has been available since 2005. The Digitization on Demand service is available for all materials in the rich holdings of the BSB. The expenses are covered by user fees. The technical administration is the same as for the projects financed by the DFG. Documents in different languages and from various periods and subject areas are available; a software tool (ERATO) was developed for the Document Delivery Service of older books.<sup>9</sup>
- Google: In 2008, through a public-private partnership, Google began to digitise the complete stock of printed documents in the BSB that are not subject to copyright restrictions. The project will proceed for a period of several years and will cover more than a million titles; at present, Google has digitised over 450,000 books.
- Co-operation with publishers of facsimiles (for example the Quaternio, Lucerne). Digital photographs serve as a basis for facsimile production; moreover, a complete digital copy is always put at the disposal of the holding library.
- Exhibitions: Increasingly, exhibitions make use of digitised copies, for example in the 'BSB Explorer', a device in which exhibition visitors can turn the pages of a digital copy, displayed on a large screen, through the movement of their hands and arms. The device was funded by the Siemens Kulturstiftung.
- Preservation copies: In the interests of the holding library, especially in the context of a loan to an exhibition, objects can be digitised, following the same procedures as Digitisation on demand (DoD).

#### *Herzog August Bibliothek, Wolfenbüttel (HAB):*

- Digitisation of the print production of the former University of Helmstedt.
- Festival Culture Online - 17th-century German Imprints of Baroque Festival Culture. Digitisation of 314 early prints in co-operation with the University of Warwick/British Library.
- Incunabula digitisation in a co-operative enterprise between the University of Cologne, Cologne University Library and the HAB. This project aims to digitise and publish on the web a comprehensive corpus of 1,000 incunabula.
- Digitisation of approximately 2,000 prints listed in the standard bibliography of Baroque authors by Gerhard Dünnhaupt.
- Digitisation of two manuscripts of the so-called 'Sachsenspiegel' by Eike von Repgow.
- Reconstruction and investigation of late medieval monastery libraries in Lower Saxony.

<sup>9</sup> For more information on ERATO, see <<https://erato.digitale-sammlungen.de/index.html?sprache=1>>.

The other partners have had little or no experience so far.

Twelve out of the fourteen institutions which responded to the *ER* questionnaire had experience in the digitisation of manuscripts (cf. appendix).

## 4 Comparable International Projects

When the *Europeana Regia* project started in 2010, few comparable projects had been undertaken worldwide. The following projects are at least partly comparable:

- Digitisation of the Codices Palatini germanici  
(<http://www.ub.uni-heidelberg.de/helios/digi/palatina.html>)
  - Comprises 848 manuscripts, administered by the University Library of Heidelberg.
  - Funded by the Manfred-Lautenschläger-Stiftung, no co-operation with partners.
  - Presentation: on the website of the University Library and for the images via the 'HeidiCon' database, a tool developed by Heidelberg University Library.
- Co-operation project among five libraries in Baden-Württemberg
  - Five partners: University Library of Heidelberg, University Library of Freiburg, University Library of Tübingen, Württembergische Landesbibliothek Stuttgart, Badische Landesbibliothek Karlsruhe, coordinated by the University Library of Heidelberg.
  - Funded by the Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg.
  - Aim: digitisation of precious and unique sources (for example charters from the Palatinate (Kurpfalz), manuscripts from the monasteries of Salem, Petershausen and Reichenau, and *Alba amicorum* of students of Tübingen University)
  - Integration into a regional educational platform is intended: 'Leo-BW-Landeskunde entdecken, erleben, erforschen online'.
  - Start: January 2010.
- CEEC (Codices Electronici Ecclesiae Coloniensis)  
(<http://www.ceec.uni-koeln.de>)
  - Digitisation of about 400 medieval manuscripts.
  - Administered by the Diözesan- und Dombibliothek Köln, no co-operation with partners, started in September 2000.
- CESG (Codices Electronici Sangallenses)  
([http://www.cesg/unfr.ch/de/index.htm](http://www.cesg.unifr.ch/de/index.htm))
  - Digitisation of about 400 medieval manuscripts, some more than 1,000 years old.
  - Administered by the Stiftsbibliothek Sankt Gallen, no co-operation with partners.
- e-codices  
(<http://www.e-codices.unifr.ch/de>)
  - Developed from the CESG project; today the Library of St Gall is still the main contributor. So far 722 manuscripts are available.
  - Partners: 30 Swiss libraries
  - Funded by the Andrew W. Mellon Foundation, the Foundation Martin Bodmer, the Swiss Electronic Library *et al.*
  - Aim: to make available all Swiss medieval manuscripts and some modern manuscripts available via the Internet
  - Dispersed manuscripts in Swiss libraries will be added to the project and digitised.
  - Dispersed manuscripts which have left Switzerland (for example to Stuttgart, Karlsruhe, Berlin, Vatican, etc.) will be added to the project in the future.

- Parker Library on the Web  
([http://parkerweb.stanford.edu/parker/actions/page?forward=about\\_project](http://parkerweb.stanford.edu/parker/actions/page?forward=about_project))
  - Partners: Corpus Christi College, Stanford University Libraries, Cambridge University Library
  - Aim: Digitisation of the 559 manuscripts in the Parker Library
  - Placed on a subscription-only interactive web application.
  - Funded by the Andrew W. Mellon Foundation
  - Concluded in 2010.
  
- Digital Scriptorium  
(<http://www.scriptorium.columbia.edu/>)
  - Not a digitisation project, but an image database that unites scattered resources digitised by many institutions
  - Aim: enrichment of descriptions by images, correction of errors, combination and bundling of knowledge
  - Presently viewable on the DS website are records for 5,300 manuscripts.
  - Still ongoing, started in 1997.

## 5 Specific Aspects of *Europeana Regia*

The main differences between previous projects focused on manuscripts and *Europeana Regia* can be summarized as follows:

- *Europeana Regia* is based on manuscripts. Most large-scale projects developed so far have been based on printed materials (i.e. incunabula, early prints, prints of modern times).
- *Europeana Regia* is exclusively concerned with the digitisation of entire manuscripts, not single parts thereof (such as illuminated pages, marks of ownership or bindings).
- *Europeana Regia* is a collaborative project among five partners in four countries. The results of these partner institutions will be fully integrated into the context of *Europeana* and the *ER* website established by the BnF.
- Two additional partners (British Library, Biblioteca nazionale di Napoli) will contribute supplementary material in the future; other partners may follow.
- *Europeana Regia* thus virtually unites collections which are physically separated.
- *Europeana Regia* is one of the largest projects in the field of the digitisation of manuscripts or early printed material. 874 manuscripts, i.e. 307,000 images, will be digitised. It is concerned with twice as many manuscripts as CESG and CEEC and as many as the Heidelberg project.

## 6 Aspects of Image Quality in Manuscript Digitisation

### 6.1 General Aspects: Optical Quality Parameters

#### 6.1.1 Resolution

As outlined above (p. 8), high resolution images are important for most aspects of manuscript study. Therefore, not only must the legibility of the text contained on the manuscript page be guaranteed, but there must also be sufficient resolution for research on the historical and material aspects of the manuscripts (such as marks of ownership, watermarks, codicological features, art historic details). The requirements for academic use of medieval manuscripts have been described by Thaller 2001 in the following way:

*'A digital object is called paleographic in our terminology if the quality allows the user to access all the information that can be derived from the original with the unaided eye. In medieval codices it is important to be able to read the text. It is also important, however, to be able to see if in the lettering there is a recognizable change in the way the pen was held, thus indicating a change of authorship.*

*Finally, a digital object is enhanceable if the digital version provides access to information that cannot be extracted from the original with the unaided eye. Image enhancement may, for example, make erasures legible again.'*<sup>10</sup>

Thaller concludes that, for academic purposes, the following minimum resolutions are required:

*'Professional manuscript work cannot be done on screens with a resolution of less than 1024 x 768. [...] 1200 x 1024 is considerably superior as soon as we go beyond the plain reading of manuscripts. ... For rare cases of detailed professional work, specifically in the area of paleography, a pretty high resolution image, close to 4491 x 3480 in size, is presented.'*

- For technical reasons, the resolution a bookscanner in a fixed production environment can provide is always the same, depending on its CCD (Charged coupled device, linear or array, the distance to the object cannot be modified). In the case of digital cameras, however, the spatial resolution depends on the camera/object distance. In order to guarantee resolution above a predefined minimum, it is therefore important to calculate the maximum object size for the real resolution of its CCD.<sup>11</sup>
- To give an example: If an original of size A4 is digitised at a resolution of 300 ppi, an image pixel corresponds to an area of about 0.1 x 0.1 mm of the original, and the total size of the image is about 2000 x 3000 pixel.<sup>12</sup>
- Although high resolution is in principle desirable, the storage place for high-resolution images must be taken into account. A restraint on high resolution is the fact that the size of an image file depends on the number of pixels in the image multiplied by the bit depth. However, the price of electronic storage media has been constantly decreasing. Similarly, limitations of data transfer velocity are rapidly disappearing by emerging technology.
- A minimum resolution of 300-400 ppi is advised for colour or shades of grey digital images by the DFG and by the Ministerio de Cultura of Spain.<sup>13</sup> A higher resolution does not make sense for most materials since it is sufficiently detailed for the relevant

<sup>10</sup> Manfred Thaller, 'From the Digitized to the Digital Library', *D-Lib Magazine* 7 (2001), <<http://www.dlib.org/dlib/february01/thaller/02thaller.html>>

<sup>11</sup> Vgl. DFG-Praxisregeln 'Digitalisierung', 2.2.1.1.

<sup>12</sup> 'Freiburger Richtlinien', p. 18.

<sup>13</sup> Praxisregeln, p. 6; 'Estandàres de digitalizaci3n', p. 4.



information to be captured by the human eye. Only in special cases, such as the examination of the structure of paper or illuminated manuscripts, is a higher resolution needed.

- For cases in which images with a higher resolution may be required for future use, digitisation at a higher resolution than that which is considered sufficient for present purposes is recommended. Otherwise, the entire digitisation process would need to be repeated from scratch. The U.S. National Archives and Records Administration therefore recommends scanning at a spatial resolution 1.5 times above the desired final format.<sup>14</sup> If 300 ppi is considered sufficient for present uses, it is therefore advisable to digitise at a resolution of at least 400 ppi.
- Current practice in manuscript digitisation projects according to the *ER* questionnaire is scanning at a minimum of 300 ppi, with the most common range for archival copies between 400 and 600 ppi.
- The resolution of an image can be evaluated by using so-called resolution targets. On the target there are smaller and smaller elements, each consisting of black bars of precisely defined width and spacing. These individual elements are numbered and organized into groups. The Resolution Target can also be used to make statements about the sharpness at different resolutions, in order to find the optimum resolution for scanning and to differentiate the sharpness at different points in the scanned area. A common standard for resolution evaluation is the USAF 1951-Resolution Target or the ISO 12233 Photography, Electronic Still Picture. While these resolution targets can serve as generic measures, the Modulation Transfer Function (MTF) measures the extent to which light spreads in the imaging process more precisely. The smaller the spread of light, the greater the resolution (see also D5.1).<sup>15</sup>

### 6.1.2 Colour Depth

As outlined above, p. 8, digitisation in colour is strongly recommended for medieval manuscripts. Only in exceptional cases, such as digitisation from secondary formats, is it acceptable to scan bitonally or with grey levels only.

- Bitonal scanning operates at a colour depth of 1 bit per pixel. A given pixel has either the value 1 (= black) or 0 (= white). Grey levels are digitised with 8 bit per pixel (256 levels). Colour pictures use a combination of 3 × 8 bit (3 × 256 levels). Technically, a colour picture is a superposition of three grey levels where the colour values are additively superimposed. Hence, 16.7 million colour levels are obtained.
- Camera and bookscanner hardware works internally always with a higher bit rate. Sometimes an output of up to 48 bit colour depth is possible. However, these so-called 'RAW'-Images cannot be rendered at this enhanced resolution on normal monitor screens.
- It may be necessary to store these images in certain cases, e.g. where a scanned image is subject to secondary editing, since tone corrections may be conducive to a loss of colour depth.
- On the other hand, archiving of 48 bit images is not easily justified. Specifically, 48 bit images require twice the storage place than 24 bit images, and this may cause problems in the case of very large archives. Hence, the use of colour depth in excess

<sup>14</sup> NARA *Technical Guidelines*, p. 8.

<sup>15</sup> Don Williams, 'An Overview of Image-Quality Metrics', in *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, CA: Research Libraries Group: 2000), pp. 72-73. On commonly used resolution targets, see also Oya Y. Rieger, 'Establishing a Quality Control Program', in *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, CA: Research Libraries Group: 2000), pp. 61-97, at 74.

of 24 bit should be limited to cases where the best possible rendering is essential for academic reasons or in order to avoid extensive post-processing.<sup>16</sup>

- In most manuscript digitisation projects covered by the *ER* questionnaire, 24-bit colour was used for manuscript materials; two institutions used only 8-bit colour images.

## 6.2 General Aspects: Relation to the Original

### 6.2.1 Colour and Tone

The treatment of colour is one of the most complex aspects of modern digitisation. Problems begin with the different sensitivity of scanner sensors (which transform optical wavelengths to electric impulses), the output of which is never identical, but always bears a specific fingerprint in its colour spectrum. To avoid such variation, individual scanner colour profiles for each station must be developed, i.e. mathematical operations modify the peculiar results of each scanner towards the standardized RGB (Red, Blue, Green) colour space,<sup>17</sup> or the CIE-LAB (Commission Internationale de l'Éclairage) system of colour values.<sup>18</sup> This procedure has to be repeated periodically, since ageing of the hardware can lead to variation in the individual colour profiles over time. As is obvious for these reasons, no change of equipment should take place during the scan process, that is, every manuscript should be scanned at one and the same scanner station, even if digitisation takes more than one day. The International Color Consortium (ICC) standard profiles, based among others, on CIE-LAB, are intended to represent colour consistently across devices and platforms. They should therefore always be stored together within the images.

Faithfulness to the original can only be judged if the display hardware (e.g. colour monitors) is also calibrated by individual monitor colour profiles. This can be achieved by the use of modern calibration software.

Closely related to colour is tone, i.e. the light intensities of a given object. Tone values determine how dark or light an image is, as well as its contrast. Tone is considered the most important image quality metric because the effectiveness of all other image quality metrics assumes that the tone reproduction is satisfactory.<sup>19</sup> Many image viewing and manipulation software programs allow the creation of histograms, which provide a graphical representation of the distribution of tones in an image. The horizontal axis of a histogram displays the distribution of tones running from dark (0) to light (255). The vertical axis shows in relative terms how much of the image is assigned to each tone. A histogram that shows any clipping in highlights or shadows may indicate a system's limited dynamic range - the loss of certain colour values during capture or image editing.<sup>20</sup>

<sup>16</sup> Vgl. DFG-Praxisregeln „Digitalisierung“, 2.2.1.2. 24-bit colour depth is also recommended by the Biblioteca Nacional de España and the Consortium of University Libraries of Catalonia (Biblioteca Digital Hispánica: 'Proceso de Digitalización en la Biblioteca Nacional de España', p. 10, available at: <[http://www.bne.es/es/Catalogos/BibliotecaDigital/docs/ProcesoDigitalizacionBNE\\_10012011\\_version3.pdf](http://www.bne.es/es/Catalogos/BibliotecaDigital/docs/ProcesoDigitalizacionBNE_10012011_version3.pdf)> and Consorci de Biblioteques Universitàries de Catalunya: 'Estandàres de digitalització: requeriments mínims (actualitzado en octubre de 2010)', available at: <[http://www.recercat.net/bitstream/2072/97497/3/1010EstandardsRequerimentsMinims\\_Public\\_sp.pdf](http://www.recercat.net/bitstream/2072/97497/3/1010EstandardsRequerimentsMinims_Public_sp.pdf)>)

<sup>17</sup> For a short introduction to colour theory, see Oya Y. Rieger, 'Establishing a Quality Control Program', p. 64.

<sup>18</sup> <<http://www.cie.co.at/>>

<sup>19</sup> Don Williams, 'An Overview of Image-Quality Metrics', in *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, CA: Research Libraries Group: 2000), pp. 72-73.

<sup>20</sup> Oya Y. Rieger, 'Establishing a Quality Control Program', p. 79.

For calibration operations, colour targets (or colour charts) and greyscale targets are employed. A colour target is a hard-copy image containing a range of colour patches of known values that can be compared either objectively or subjectively with the original and then used to calibrate the device or create colour profiles. For checking the colour correctness additional scans should be done with a colour target. Common colour targets are, for example the Kodak Q-60 Color Input Target (IT8) for ANSI IT8.7/1 (transmission) and IT8.7/2 (reflection) Standards (or ISO 12641),<sup>21</sup> or Gretag Macbeth™ ColorChecker®.<sup>22</sup>

Other less problematic aspects of optical image quality are the saturation of colour values (also referred to as chroma, the intensity or purity of the colour), and halos or flares (stray light in the optical system) and flaws such as Moiré effects (interference patterns).

Light exposure is an important factor in the production of good colour images. If there is no efficient lighting, the colours will not be correct. On the other hand, light is another damaging factor for the original (see below, pp. 29-30). The so-called 'colour temperature' varies according to the light source used. The use of standardized light (with a colour temperature of 5000 K) is therefore important not only in the digitisation process, but also in the course of quality control (see D5.1, p. 8).

Pages on which gold, silver and purple were used are particularly difficult to reproduce. Purple can appear brown or grey. Gold may appear without reflexes and silver black because of corrosion. For a good reproduction of gold, the original must be very carefully lit. To achieve a correct reproduction of colour values, experimenting with different light sources at different angles or with hardware (bookscanner or digital camera) or software solutions may be necessary. It is recommended to use mobile light stands and permanent lighting with the aid of D 50-lamps (with a temperature of 5.000°K) in order to guarantee a homogenous reflection of light towards the objective. The light source should be equipped with softboxes (a box of translucent fabric mounted over the lamp) in order to guarantee homogenous light and to protect the original against damage. The reflective properties of gold vary regionally, depending on its specific chemical composition. The so-called shell gold from Asia (made from gold powder suspended in gum arabic) is even more difficult to reproduce. The best results are obtained when the surface has hair cracks or other signs of damage and wear; the light is then reflected several times and the surface in the digital image appears golden. A flawless patch of shell gold is more difficult to reproduce, since the light is reflected in a homogenous way. This reflection leads usually to white patches, often with cyan rims and artefacts in the digital image.

Coloured areas on textiles are particularly problematic; on this material, the light is better absorbed and scattered; a colour shift is observable in some cases.

Some of the partner libraries in the *ER* project have gathered valuable experience in this area which is available to their partners. The BnF, for example, offers to do this kind of photographic work for French partners. A book with silver script which the Bibliothèque Municipale de Reims will contribute to the project will be physically transported to the BnF for digitisation.

The *ER* questionnaire revealed that there is no common or standard practice in the use of colour charts in manuscript digitisation projects. Some institutions use no colour targets at all, others embed colour charts in every single image. The most common practice is to include a colour target on one selected page of each volume. Since the production of targets has not

---

21

<http://www.kodak.com/global/en/professional/products/films/catalog/kodakProfessionalQ60Targets.html>.

<sup>22</sup> For a list of common colour targets and their peculiarities, see Rieger, 'Establishing a Quality Control Program', p. 80. On Colour Management in general, see Gerald Maier, 'Colormanagement bei der Farbdigitalisierung von Archivgut: Grundlagen, Hintergründe und Ausblick', in: *Digitale Archive und Bibliotheken*, ed. Hartmut Weber and Gerald Maier (Stuttgart, 2000), pp. 179-199.

been standardized so far, the wide range of different products used by different institutions makes a comparison of colour values across institutions difficult.

### 6.2.2 Dimensions

Another aspect to consider in the relation of image and original is the original format and its representation on the screen. Linear scales should therefore always be digitised at least once together with the original in order to be able to assess the size of the original. In order to achieve faithfulness to the original, the orientation of pages within one binding unit should not be altered by image processing or the turning of the object; nor should the sequence of images be altered in comparison to the original. Folded pages pose a special problem in this context (e.g. the question of sequence of the digital images).

For the same reason, there should be no post-production modification of the images, such as cropping. Although the tight bindings of manuscripts may lead to distortions in the image, no attempt to correct these by using software should be made.

A related issue is the question whether images of single or double pages should be taken. In some cases, the display of two pages may be preferable for a global impression of the opened book. From the point of view of conservation, however, it is recommended to take single page images, since they are easier to capture from books with tight bindings: the codices need not be flattened if modern camera equipment is used. A second argument is related to the requirements of modern display options, such as 'turning-the-pages-devices', for which single page images are needed. If necessary, facing pages can be virtually united for the web presentation.

Eight out of eleven institutions in the *ER* questionnaire opted for single page images.

### 6.2.3 Completeness

As outlined above, the individual history of each manuscript is of interest to the scholar and the public. Much information on the early history of a manuscript can be found outside the text pages, e.g. through marks of ownership on flyleaves, shelfmarks on a book's spine, or stamps executed in blind tooling on leather bindings. But the intrinsic artistic value of the binding (e.g. precious covers in metal or ivory, decorated covers) has also to be considered. Manuscripts should therefore always be digitised as completely as possible, that is, from cover to cover, including flyleaves, empty pages and pastedowns. Images of the bindings (the front and back cover and their interiors) should also be added. If possible, the spine of the volume and its edges should be included in the standard digitisation process, especially since these parts of the book must be digitised if 3D reproductions of the entire book are intended.

Since the authentic reproduction of the original is desirable, digitisation with 'wide frames' is recommended. This means that on all images, the edges of the open codex remain visible and a narrow surrounding rim is included to show that nothing was added or omitted; there should be no cropping of pages. This may contribute to create a global impression of the entire codex even on the screen.

A complete set of all pages, including empty pages and the verso of illuminated pages, should also be displayed on the Internet. In the past, some institutions tended to omit such images from the versions displayed on the Internet, although they were included in the set of images intended for long-term preservation. If the presentation of the digital images

comprises sufficient structural metadata, however, such pages could be easily identified and skipped by users not interested in them.<sup>23</sup>

A difficult question which has to be addressed at the beginning of each digitisation project is whether 'alien objects' (such as loose pages of later date or protective paper on illuminations) should be digitised along with the original pages of the manuscripts. The former may be valuable indicators of ownership or of the use of manuscripts in later centuries.

According to the *ER* questionnaire, most institutions digitise the book covers, flyleaves and pastedowns. Only few institutions generally include the edges and spine.

### 6.2.4 Sequence

The digitisation of watermarks (for late medieval and Renaissance paper manuscripts) as well as of folded pages and of 'alien objects' raises the question of sequence: while folded pages can (and should) be integrated into the image sequence to reflect the correct sequence of pages of the original, watermarks, in contrast, are digitised in a different process and as separate images, which means that they cannot be integrated as easily into the image sequence. They may be added after each page or at the end of the original copy or in a separate file to be addressed by a specific URN/URL. The group of *ER* project partners opted in favour of the latter option.

Another issue is the sequence of spine and edges. A good solution may be to arrange spine and edges clockwise (i.e. in the order spine, upper edge, vertical edge, lower edge) and insert these pictures after those of the back cover in the Internet presentation. It would then be possible to show the colour target at the very end of the digital images of each codex. Further information based on new experience will be given in D5.3.

For alien objects, a good solution could be to take two images, one together with these later additions, and one without.

## 6.3 Aspects Peculiar to Manuscript Digitisation

### 6.3.1 Due to the Physical Condition of the Manuscript

Western medieval manuscripts differ in their physical format from modern printed books in a number of significant aspects. They sometimes display

- large formats
- different formats of pages in one binding unit
- fragile areas, e.g. layers of painting
- damages like tears or wrinkles
- microbiological infestation, that may even be a health hazard for the staff and a problem for technical equipment
- transparent areas from thin layers of parchment
- page curvature
- reinforcement strips in the inner or outer folds of quires
- tight bindings

These peculiarities may be an obstacle to optimal image quality. Especially tight bindings require an elaborate scanning system, since otherwise text may be lost in the inner folds of quires.

---

<sup>23</sup> On recommended elements in structural metadata, see below, pp. 42.

A special challenge is the reproduction of palimpsests, that is, parchment where the manuscript text has been scraped off for re-use. Here, multi-spectral imaging may be needed (see below, pp. 23-24).

Special care is also required because errors in pagination occur frequently in medieval manuscripts. Since pagination is not an original feature in them, the question arises whether such errors should be corrected or not before scanning. If it is decided to correct the faulty original numbering system, it should be taken into consideration that older literature may refer to such faulty numbering systems. It is therefore recommended to retain the older faulty foliation or pagination, together with the corrected numbering system. Ideally, both also should be visible in the structural metadata.<sup>24</sup>

Examples for such problems with pagination are:

- pages not numbered
- numbers skipped
- pages cut out or removed after pagination
- numbers appearing more than once
- pages or items bound together with the historical manuscript pages
- several empty pages
- numbered flyleaves

Complicated numbering schemes may necessitate the preparation of concordances between page numbers and digital image numbers before the digitisation process (see D5.1, p. 11).

### 6.3.2 Appropriate Backing of Translucent Materials

The recording medium of manuscripts and books, whether paper or parchment, has at least some partial transparency/translucency and often contains holes. Frequently, therefore, items on the reverse side or the following page of the document to be scanned may be picked up by scanners (giving the scan a 'pseudo-palimpsest' character). In order to minimize this effect, it is recommended to back each page that is being scanned with a neutral sheet of paper inserted between the target page and the following page. However, there is incomplete agreement whether the backing sheet should be white, black, off-white or cream-coloured. Black backing has the advantage of reliably blocking out any *verso* features which can shine through in the case of a white backing. It has also been argued that black backing reduces the contrast of the digitised image, especially when the digital image is recorded with a low dynamic range of 8 bit for black and white or 24 bit for colour. On the other hand, black backing can modulate the appearance of the background in the scanned image and look like a letter or a note. Off-white backing materials have been recommended to match the colour and texture of the original recording medium in order to minimize the impact of the backing medium on the colour hue of the background and the margin.<sup>25</sup> If the paper is beige, however, users might not be able to detect existing holes. If the paper is white, a distracting impression may result.

---

<sup>24</sup> In an example from the Kongelige Bibliotek of København (GKS 2232 4<sup>o</sup> Guaman Poma, Nueva corónica y buen gobierno, 1615), both numbering systems are displayed, see <<http://www.kb.dk/permalink/2006/poma/438/en/text/?open=id2975498>>.

<sup>25</sup>NARA, *Technical Guidelines*, S. 42.

### 6.3.3 Conservation Aspects and Digitisation

Manuscripts are unique, often priceless physical artefacts, consisting of precious and rare materials. Frequent conflicts between the requirements of image quality and preservation aspects are therefore inevitable. The holding institutions must carefully weigh the arguments on both sides; the more precious and rare the original, the more caution and effort will be applied. Once digitised, however, the digital copy may also serve to ensure the preservation of the original, since the result of the digitisation process may serve as a surrogate for the use of the often fragile originals. Digitisation is also an occasion for the secondary production of safety microfilms of the original, which are considered to be more durable than digital formats.

A prominent example where the institutions responsible for manuscript digitisation projects differ is the use of glass plates on the original to keep it flat and thus avoid distortions. The BSB<sup>26</sup> and the Swiss e-codices project consider the risk of damage to pages (especially those containing illumination) and to the spine too high and therefore never use glass plates. As the *ER* questionnaire has shown, several other institutions do use glass plates for the sake of picture quality.

Another example is the unbinding of manuscripts for digitisation purposes, which was previously practiced when facsimiles were produced. This is no longer recommended because valuable information may be irrevocably destroyed. Although photographic and lighting techniques have advanced considerably, they can only partially compensate for the loss of image quality when manuscripts are tightly bound.

It goes without saying that due to the age and value of the original material, many automatic procedures which are applied in mass digitisation projects for printed material must be used with caution. Examples are automatic methods of turning and holding pages: static charges cannot be used on parchment. On the other hand, aspiration by an air stream for fixing and turning pages may be more preservative than manual handling. Some practices are precluded, e.g. automatic document feeding.

A number of the conditions under which digitisation is carried out may cause damage to the original, e.g. through exposure to light and/or through inadequate climatic conditions (humidity, temperature). Consequently, the relevant international standards have to be respected. Another important aspect is the handling of the object through the scan operator, e.g. the turning of pages, or the opening of the book, which may damage the spine if it is opened too wide.<sup>27</sup>

Preventive measures must therefore be taken to avoid preservation problems caused by the digitisation process. The necessary first step is a preliminary examination of each manuscript. Trained conservators should then take the decision whether the manuscript is suitable for digitisation at all and if so which measures have to be taken to avoid damage to the original. Such measures may include preservation actions before digitisation. Most institutions in the *ER* questionnaire consider this a standard procedure. Other measures could be -

- the presence of trained conservators during digitisation (however, this is practiced only by about half of the institutions which participated in the *ER* questionnaire)
- control of environmental conditions (relative humidity, air temperature, lighting) during digitisation and transports
- use of special equipment, e.g. book cradles, foam wedges, cotton gloves

<sup>26</sup> For technical and conservation aspects of digitization at the BSB, see Brantl and Schäfer, 'Massendigitalisierung'.

<sup>27</sup> On the risks resulting from the digitisation process, see Gerhard Banik, 'Risiken bei der Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut aus konservatorischer Sicht', in: Hartmut Weber und Gerald Maier, eds., *Digitale Archive und Bibliotheken* (Stuttgart, 2000), pp. 311-324.

- book specific instructions to the scan operator: definition of the maximum opening angle, information on delicate parts of the object, appropriate handling (see also below, p. 29-30).

### 6.3.4 Other Factors Peculiar to Manuscripts

Manuscripts are unique documents. As a result, information which is not of general importance for printed books is relevant, for example, on details such as:

- watermarks (important for dating, provenance)
- stamp marks (provenance)
- composite manuscripts: manuscripts of different ages and/or origins bound together (provenance)
- annotations of readers (reception)
- illumination (provenance)

Even though the ideal of digitisation is the creation of an 'identical facsimile', one has to be aware of the fact that there are always aspects of the original which are lost in a digital image, making it necessary to fall back on the original manuscript for research purposes in certain cases. Some information cannot be recorded by digitisation alone, for example, the quire structure of a manuscript cannot be shown in a two-dimensional image.

In other cases the reproduction of difficult details may only be possible by multi-spectral imaging: high-resolution images in different light wavelengths may report details that are invisible to the unaided eye, for example -

- infrared light highlights ink containing carbon from crushed charcoal in medieval inks or watermarks
- ultraviolet light picks up ink containing iron
- CAT scan and X-ray technology can help to decipher read brittle papyrus scrolls

As the *ER* questionnaire revealed, the use of such multispectral imaging is still rare; only one institution had made experiments with UV light. Another example is the special workflow for infrared photography at the BSB, which will be presented in the following excursus.

#### Excursus:

#### **Special workflow for infrared light photography (thermography) for watermarks used in the BSB blockbook project**

##### *1) Condition Check*

First of all, it has to be decided whether the manuscript in question is suitable for thermography. The most important criteria are:

- material (paper manuscript)
- condition

##### *2) Preparing the Documentation for Thermography*

First, it has to be decided which pages contain watermarks. This can be done with the help of an electroluminescent panel. The preliminary check avoids unnecessary thermography procedures for pages without watermarks.

##### *3) Photography: Technical Aspects*



Thermography is based on temperature differences of different parts of the object. When putting the page in contact with a warm plate, there is initially a temperature difference between the thinner layers of paper in the area of the watermarks and the other parts of the paper sheet.

In order to capture the watermark image at the time of its optimal visibility during the warming-up process, a time sequence of images is recorded. The best picture is then selected by a one-by-one inspection of the image frames. Since the warming up takes place rapidly, it is not possible to catch the best picture by manual intervention during the warming up process.

It is mandatory to include a ruler in the recorded image, because accurate sizing is very important for comparison of watermarks. Only by a comparison with dated watermarks included in the standard literature (Piccard, Briquet) can watermarked papers be dated.

#### 4) Post-Processing

After the scanning process, the pictures require suitable image processing. The angular orientation of the picture is important. The chain lines should be at a right angle to the image frame, the motif should be orientated according to standard viewing conventions (for example, for a unicorn 'head at top, feet at bottom of the image'). After quality control (compare WP 5), the pictures are imported to the project homepage and the watermark database WZIS.<sup>28</sup> According to the BSB internal rules, the of every picture file must contain the shelfmark and the folio number.



*Example: a watermark digitised by thermography*

### 6.4 Effects of these Factors

As is obvious, the special care required in the handling of these unique originals has immediate effects both on the speed and the cost of the digitisation process. In comparison to other materials, the digitisation of medieval manuscripts requires more time and effort and the use of special equipment (e.g. book cradles, special light sources or humidifiers for preservation purposes).

<sup>28</sup> For the Watermark Information System project, see < <http://www.landesarchiv-bw.de/web/50960> >

As outlined above, the interests of preservation and image quality must be balanced. Due to the physical peculiarities of manuscripts and aspects of preservation as outlined in ch. 6.3, image quality may still be inferior to the standards that can be achieved with less challenging modern printed material.

Whereas some central aspects and standards of digitisation, such as quality parameters, must be determined at the outset of each project, extensibility remains an important issue. This is especially important in the case of metadata, which may become more detailed and refined as a project proceeds: complex structural metadata, the transcription of texts, etc., which may be added when (users') need arises, should be mentioned in this context.

## 6.5 Summary: Recommended Quality Parameters for the Digitisation of Manuscripts

- Digitisation of manuscripts from the original should always be in colour. Digitisation in black and white would result in the loss of important data.
- Digitisation in black and white can therefore only be recommended for mass digitisation of manuscripts from black and white microfilms.
- A minimum resolution of 400 ppi is recommended. For existing images,, a resolution of 300 ppi may be sufficient. In justified cases (e.g. for small details such as watermarks or for large formats), a resolution up to 600 ppi may be required.
- Digitisation should include metrical and colour targets.
- The use of colour management software (ICC profiles) is recommended.
- No change of equipment should take place during the scan process.
- All manuscripts should be digitised in one sequence of operation.
- The books must be digitised entirely, including front and back covers, fly-leaves and empty leaves. If feasible (i.e. if conservational aspects or format are not obstacles), the spine and the edges should be included.
- Watermarks should be digitised in separate files and in a special workflow.

## 7 Workflow and Technical Aspects of Manuscript Digitisation

Since many institutions have gathered substantial experience of digitisation projects over the last decade, standardized workflows have been established in a growing number of institutions. All of the 14 institutions which returned the *ER* questionnaire have developed standard digitisation procedures for manuscripts, which, of course, vary in details. In the following section, the major steps of a workflow for digitising manuscripts are described.

### 7.1 Digitisation Inhouse v. External Service Providers

Digitisation can be carried out inhouse by the holding library or by an external service provider. The advantages of inhouse digitisation are:

- growing experience of the institution
- equipment can be used for subsequent projects
- no dependency on external services (danger of insolvency, price increase, fluctuation of staff)
- no need for the production of technical specification and contracts
- lesser risk of damage to the originals (e.g. during transportation) or theft,
- normally no insurance necessary
- no or low transport costs
- conservators at hand
- trained librarians at hand for creation of and connection to reliable metadata
- quality control can be carried out by the library's own trained staff
- short-time unavailability of originals
- control over all procedures, including data transfer

Of course, it may not be possible for all holding libraries, especially smaller ones, to build up their own digitisation departments. In such cases, it may be necessary to rely on experienced external service providers. The advantages of using vendors in digitisation projects in general are summarized in Gertz 2000:

*The primary benefits of working through vendors are financial and technical.*

- *The institution does not have to devote space to scanning, nor does it need to convert its space (possibly including construction) to suit electrical and other technical requirements.*
- *The institution does not constantly need to purchase the latest equipment and software. The vendor is responsible for keeping up with the times.*
- *The institution does not have to deal with hiring, training of sophisticated specialist skills, and management of staff.*
- *The vendor and not the institution copes with costly equipment breakdowns, downtime, and correction of errors.*
- *The institution benefits from the vendor's economies of scale and high productivity.*
- *Finally, the price is stated up front.*<sup>29</sup>

It should be noted, however, that very few commercial service providers have gained sufficient experience with manuscripts or rare material; it is therefore recommended to turn to digitisation competence centres or other experienced libraries for advice on potential project partners. Cost and effort advantages on the production side must be weighed against conservation and quality aspects. In addition, the effort necessary in choosing a vendor and

---

<sup>29</sup> Janet Gertz, 'Vendor Relations', ch. 8 in: Maxine K. Sitts, ed. for the Northeast Document Conservation Center: Handbook for Digital Projects: A Management Tool for Preservation and Access (Andover, MA: 2000).

the development of detailed specifications (including conservational aspects),<sup>30</sup> external audits during productions and post-production quality control must not be underestimated (see D5.1).

According to the *ER* questionnaire, the overwhelming majority of institutions with experience of manuscript digitisation carry out these projects inhouse (13 out of 14 responding libraries). This is a very high rate compared to the overall use of service providers in digitisation projects in Europe, which is about 55%.<sup>31</sup> The preference for inhouse digitisation is undoubtedly due to the special value and often fragile condition of medieval manuscripts. Transporting manuscripts to external service providers must be carefully planned; appropriate environmental conditions must be guaranteed (see below, p. 29). Some partners (e.g. the BnF) have accommodated external service providers in their libraries so that the manuscripts do not need to be transported and can be taken care of by the library staff. At the same time, closer oversight of the vendor work is possible.

## 7.2 Condition Survey of Manuscripts

This important step should take place before a project is agreed on, since the selection of objects included in the digitisation project will depend on the results of this condition survey. The condition survey of the originals has a high impact on the feasibility of the digitisation project as a whole, on the expected effort, time and cost structure of the project.

On the basis of the results of the condition check, it is decided if conservation treatment (e.g. the stabilisation of tears, loose bindings etc.) is necessary before scanning or whether the volume is suitable for digitisation at all. The survey should be carried out by a conservator.

- The condition survey can lead to the detection of hitherto unnoticed preservation problems, such as fragile parchment or paper, chemical reactions in colour particles and ink corrosion.
- Peculiarities such as gilded pages are noted.
- The characteristics of the binding must be evaluated (e.g. its material, tight bindings).
- The extent is determined; the size of the object is measured and noted.
- According to the condition survey, picture capture techniques will be recommended, together with the necessary additional material such as supports and book cradles.
- Object-specific instructions are then given to the scan operators on the appropriate handling of the object and on the use of supports and other equipment. Cf., for example, the sheet used for these procedures at the BSB (no use of glass plates, no loosening of the binding)

---

<sup>30</sup> For details on vendor relations, including useful hints on their selection, see Gertz 2000.

<sup>31</sup> Marlene Sturm: 'Processing Digitized Materials in European Libraries: A Survey Report (March 2011)', p. 7, available at: < <http://www.kb.nl/pr/pers/download/>>.

**List of items to be inspected prior to digitisation by the IBR (Institute of Book and Manuscript ConseRvation)**

Name of project: \_\_\_\_\_

Date of inspection:  
Member of IBR staff involved (name):

Please note: In the field "Instructions for the scan operator" he or she ist to be provided with all relevant facts in order to enable him/her to work with the appropriate (very diligent) care, such as:  
 - IBR staff has to be present (initially/at all times)  
 - High ranking manuscript ("Tresorhandschrift") | Specific characteristics of the binding | Damages of all kind | Details on how the book has to be placed on the scanner (e.g. with foam book rests)  
 - Illumination(s)  
 - Leaf gilding(s)  
 - Paper slips attached? (To be scanned as well: yes/no)  
 - Air humidifier necessary?  
 - etc. ...

No	Shelfmark	Digitisation and microfilming for security reasons	Extent (leaves/pages)	Format	Fit for scanning (Yes/No)	Opening angle: 90   90-110   180 degrees	Instructions for the scan operator	Pertaining to which part(s) of the original: e.g. entire work, binding(s), folio/page, etc.

All partners in the *ER* project and eleven out of thirteen institutions in the ER questionnaire carry out a condition check on the originals at the beginning of a project. This corresponds to the results of a survey report carried out by the Koninklijke Bibliotheek in The Netherlands: it is customary in over 90% of all European digitisation projects.<sup>32</sup>

### 7.3 Preparation of Manuscripts for Digitisation (Step 1)

- Before the scanning or photography process, it is necessary to check as early as possible if appropriate bibliographical metadata are available or must be generated. From this stage, databases (such as the BSB-ZEND, on which see below, pp. 43-44) can be used to support the workflow.
- The checking of the foliation is common practice in all partner institutions and also in the institutions which returned the *ER* questionnaire. This checking is important because problems might occur if several foliations exist or if single leaves are missing. Some manuscripts contain no foliation at all. In this case, a foliation in pencil should be added if the condition and rarity of the manuscripts do not prevent it.

### 7.4 Digitisation Process/Scanning or Photography (Step 2)

#### a) Pre-Processing

- Choice of appropriate scanning/camera hardware in line with the condition check of the manuscript. For medieval manuscripts, high resolution digital cameras are recommended. Most institutions which participated in the *ER* questionnaire use only digital cameras (of a wide range of manufacturers and often together with special support solutions) for manuscript digitisation.
- If cameras only are used, there is normally no appropriate software solution supporting the digitisation workflow. It may therefore be preferable to use special

<sup>32</sup> Sturm, 'Processing Digitised Materials', p. 7.

bookscanning devices with cameras which offer software support for the production workflow.

- Choice of appropriate auxiliary material (e.g. book cradles, foam wedges) in line with the condition check of the manuscript. For fragile originals such as medieval manuscripts, a number of special solutions have been developed. Examples are the 'Wolfenbüttel Book Reflector',<sup>33</sup> the 'Grazer Buchtisch' (which also exists in transportable form),<sup>34</sup> or the new Cobra book scanner by Microbox.<sup>35</sup>



Grazer Kameratisch<sup>36</sup>



<sup>33</sup> E.g. for early printed books <<http://www.hab.de/bibliothek/rw/buchspiegel/index-e.htm>>.

<sup>34</sup> See Manfred Mayer, 'Digitalisierung mittelalterlicher Handschriften an der Universitätsbibliothek Graz', available at <[http://cool.conservation-us.org/iada/ta99\\_185.pdf](http://cool.conservation-us.org/iada/ta99_185.pdf)> and "The Traveller's Conservation Copy Stand", see

<<http://orbislibrorum.at/pdf/Beschreibung%20Traveller%20VS%20neu.pdf>>.

<sup>35</sup> <<http://www.book2net.net/de/buchscanner/book2net-v-scan-cobra.html>>.

<sup>36</sup> Image from <<http://www.ub.uni-heidelberg.de/helios/digi/palatina.html>>.

*Cobra book scanner*<sup>37</sup>

- Specific environmental conditions during the scanning process must be guaranteed. For parchment manuscripts, an air temperature of 18-21 °C and 50-60% of relative humidity are recommended. Humidifiers and air conditioning should be used if appropriate. It is very important to avoid sudden changes of climate: a maximum change of 2 °C and 3% of relative humidity per 24 hours are acceptable according to the International Standard DIN ISO 11799:2005-06. These conditions must be observed not only during the digitisation process, but also during transports, which may necessitate special packaging and measuring equipment. There seems to be a lack of awareness about the importance of these factors in some institutions, especially of humidity, which has a greater impact on the degradation process than, for example, temperature.<sup>38</sup> According to the ER questionnaire, only seven out of thirteen institutions controlled humidity.
- Special lighting sources should be used to avoid damage by light exposure. Today, ultraviolet and infrared filters are a basic requirement. In recent years, the cold light emitting diode (LED) lighting in scanning systems has replaced conventional fluorescent tubes. In addition to these lighting techniques, electronic flash can be used directly or indirectly, with a Pyrex dome as a filter and burst protection. For example, two flashlights with diffusers (Soft Boxes) are recommended by the Swiss e-codices project.<sup>39</sup> A very short exposure to flashlight leads to better photographic results and guarantees a high frequency of image taking.
- Handing over of the manuscript to the scan operator, accompanied by a conservator if necessary.

*b) Scan Operating*

- Digitisation of single pages is the preferred option in most cases, especially since this is possible even when the manuscript is tightly bound (see above, p. 20). Special supports and camera tables today make possible the digitisation of codices which can only be opened to an angle of 45°.
- Sometimes, however, the digitisation of two facing pages may be preferable, e.g. for purposes of orientation and coherence. If two pages facing each other are digitised as one image, however, problems in indexing may result because a system to differentiate the left and the right page must be developed.
- The use of paper/cardboard behind the page to be photographed is recommended if there are damaged areas of parchment (see above, p. 22).
- If the scope of the project is rather small and if all objects are of similar size, the use of the same hardware for all objects to be digitised is strongly recommended. Different colour profiles for each scanning station or camera may otherwise lead to differences in image quality according to the hardware used.

*c) Post-Processing*

- quality control of images (see WP 5)
- rework of images which fail to meet the quality standards defined for the project or which are missing
- generation of structural metadata (see D5.1 and WP 2)

---

<sup>37</sup> Source: <<http://www.book2net.net/en/book-scanner/book2net-v-scan-cobra.html>>.

<sup>38</sup> See Banik, 'Risiken', p. 313.

<sup>39</sup> 'Freiburger Richtlinien', p. 17.



- web presentation, integration of image links, e.g. to the library's catalogue
- manuscripts returned to the stacks

### 7.5 Long-Term Preservation (Step 3)

Before and during the process of digitisation some important parameters are set that are crucial for the long-term preservation/digital preservation<sup>40</sup> of the created digital objects:

- file format of the master file
- preservation metadata
- archival storage

---

<sup>40</sup> Long-term preservation/digital preservation can be defined as the 'series of managed activities necessary to ensure continued access to digital materials for as long as necessary'. See Digital Preservation Coalition: *Preservation Management of Digital Materials: The Handbook*, available at <<http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts>>.

### a) File Format of the Master File

Choosing the appropriate file format for the archival storage of the image masters with the aim of digital preservation in mind is a challenging task and depends on several factors that have to be assessed by every institution individually:

- required storage capacity
- image quality
- long-term sustainability, e.g. the openness of the used standard functionality<sup>41</sup>

Every institution should take these aspects into account and clarify its own requirements before a final decision is taken (e.g. low storage consumption,<sup>42</sup> no or lossless compression, support by current applications, degree of standardization, number of copies, etc.). For the time being there is no single standard file format that can be recommended for every institution. Nevertheless, some formats for image master files are considered by the preservation community to be more adequate for long-term preservation than others:<sup>43</sup>

1. TIFF (uncompressed / lossless)
2. JPEG2000 (lossless)<sup>44</sup>
3. PNG.

The hierarchy of the list above reflects the wide range of file formats used in memory institutions. According to the 2011 'Survey Report', most European national libraries store uncompressed TIFF files for long-term preservation.<sup>45</sup>

---

<sup>41</sup> Gillesse, Robèrt: *Alternative File Formats for Storing Master Images of Digitisation Projects*, available at [http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/alternative\\_file\\_formats\\_for\\_storing\\_masters\\_2\\_1.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/alternative_file_formats_for_storing_masters_2_1.pdf).

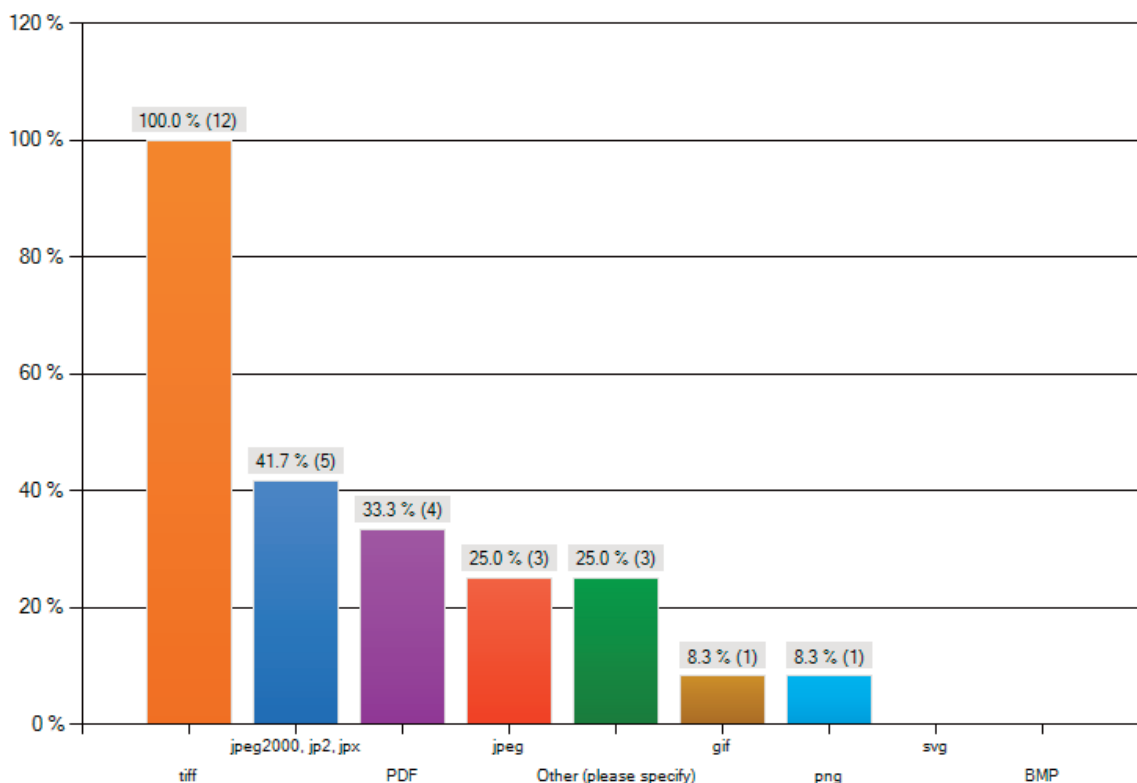
<sup>42</sup> The IMPACT Storage Estimator can help to assess the amount of storage and storage media needed for the output of digitisation projects: <http://www.impact-project.eu/uploads/media/IMPACT-ise-4.xls>.

<sup>43</sup> See Federal Agencies Digitization Initiative Still Image Working Group, *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*, p.67, available at: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image-Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf); Gillesse, *Alternative File Formats*, and *Recommended Data Formats for Preservation Purposes in the Florida Digital Archive*: <http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>.

<sup>44</sup> Use of JPEG2000 may be problematic as there are some known problems with colour spaces and ICC profiles, see [http://www.dpconline.org/component/docman/doc\\_download/526-jp2knov2010vanderkniff](http://www.dpconline.org/component/docman/doc_download/526-jp2knov2010vanderkniff), and <http://jpeg2000wellcomelibrary.blogspot.com/2010/12/guest-post-loc-response-to-discussion.html>

<sup>45</sup> Sturm, 'Survey Report', p. 16.

**Which file formats do you use for the long-term-storage of your digitised material?  
(multiple answers possible)**



### b) Preservation Metadata

The recording of adequate metadata for digital preservation is necessary as it enables institutions to plan and conduct preservation measures (migration, emulation) and also to document the context of the digital objects in order to guarantee their management in the long term (e.g. provenance, events in the lifecycle, legal aspects, fixity information).<sup>46</sup> It is strongly recommended to use XML as the format and the already existing metadata standards for all types of metadata, for preservation metadata PREMIS (on which see below, pp. 39-41), and METS as a container format for all metadata and structural information.<sup>47</sup>

### c) Storage Media

It is not recommended to use optical media like CD-R, CD-ROM or DVD for storing digital master files. For reasons of longevity and data integrity (bitstream preservation), it is strongly recommended to use redundant hardware systems like RAID-5 for the storage of the image master, and additional copies should be kept on tape storage systems. Redundancy, regular back-ups and periodical renewal of storage media are crucial for preserving the bitstream permanently. Ideally this task can be handed over to a competence centre for the long-term preservation of digital data, e.g. a university computing centre. To ensure the usability of the archived data it is fundamental to maintain the connection of the data with the corresponding metadata. In any case, file naming conventions must be defined by each institution before

<sup>46</sup> For further information on the importance of preservation metadata, see the first chapter of Caplan, 'Understanding PREMIS', <<http://www.loc.gov/standards/premis/understanding-premis.pdf>>.

<sup>47</sup> METS: Metadata encoding and transmission standard <<http://www.loc.gov/standards/mets/>>.

the start of the digitisation process (see also below, pp. 45-6), and they should never be changed in order to avoid problems within the native system.

These are the most important aspects in the field of archival storage that have to be taken into consideration before and during the process of digitisation; digital preservation as a whole is a complex challenge and comprises many more activities.<sup>48</sup>

## 7.6. Display and Integration of Images (Step 4)

### 7.6.1 Display Formats

Since the TIFF files produced for archival purposes are unsuitable for display on the web due to their size, other formats must be derived from the original files for presentation purposes. Usually, compressed formats such as JPEG<sup>49</sup> are created for these purposes out of the digital master files.

The advantage of compressed files is that, due to their reduced size, they need less storage space and reduce reaction time in browsers and other software. For Internet publication, JPEG, PDF and PNG formats are preferable due to their wide distribution. GIF has severe limitations of colour depth and is limited to rendering bitonal grey levels. However, no strict limitations should be imposed, since these formats, as well as related formats which are supported by customary browsers, can be generated later from the master file. Hence, decisions can be revised, and modifications of rules can be taken into account.

The *ER* questionnaire showed that for these presentation formats, most institutions use files with a resolution between 150 and 300 ppi; according to the 2011 'Survey Report', the preferred format is JPEG or PDF.<sup>50</sup>

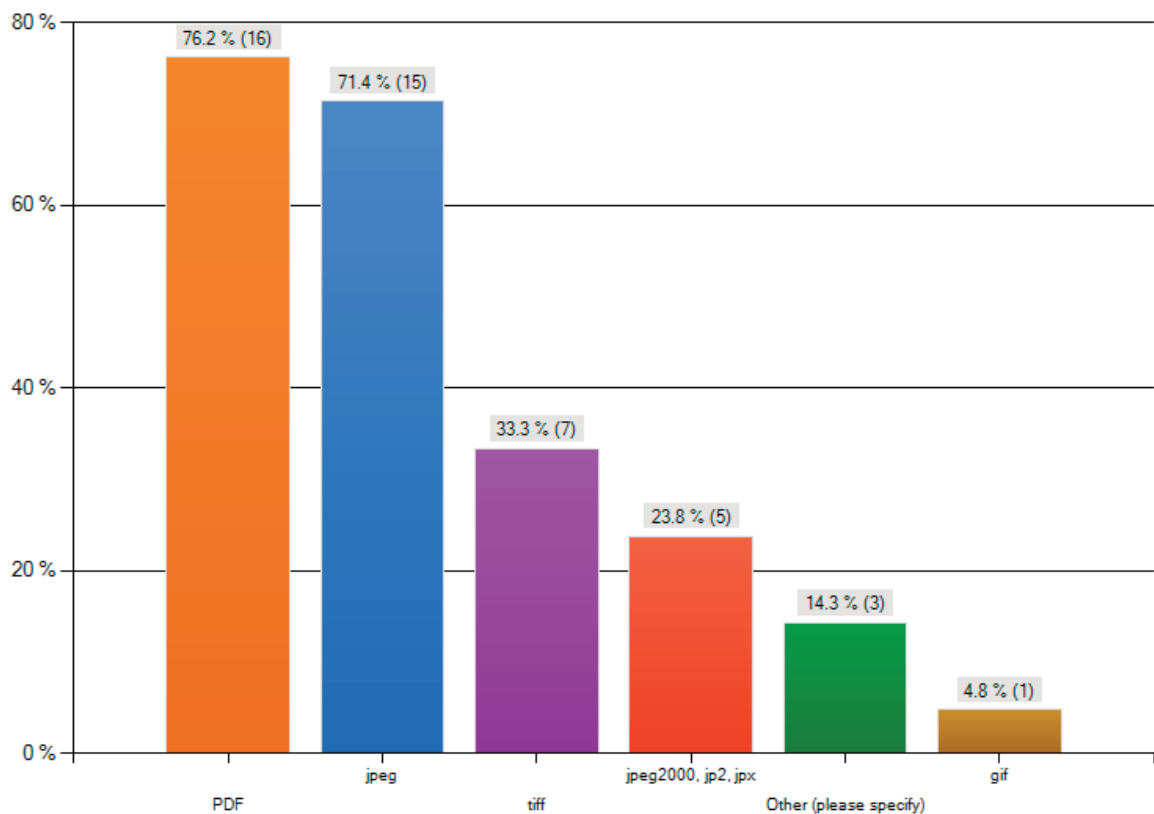
---

<sup>48</sup> See the ISO Standard OAIS (Open Archival Information System):  
<public.ccsds.org/publications/archive/650x0b1.PDF>.

<sup>49</sup> <http://www.w3.org/Graphics/JPEG/itu-t81.pdf>.

<sup>50</sup> Sturm, 'Survey Report', p. 14.

### Which file formats do you use for access? (multiple answers possible)



### 7.6.2 Reusability in other Media

In the interest of sustainability, projects funded by public institutions should also be adapted to be reused for other projects in different media, for which the images usually have to be post-processed:

- print. For reuse in print, images must be available in sufficiently high quality (high, resolution, colour correctness); special image processing with the aid of colour scales is necessary.
- 3D-applications
- 'turning-the-pages' applications (e.g. in virtual exhibitions)
- selected images

### 7.6.3 Integration of Images

#### *Display Options*

Apart from general optical quality parameters, user expectations play an important role in choosing appropriate functional solutions for displaying the digitised images. Easy and intuitive handling of all software tools is a basic requirement. Present-day users expect increasing comfort in the use of images on the web, e.g.

- original image scale and derivative sizes (large, medium, small, thumbnail)
- zooming functions
- a good compromise between the size of files and a high resolution

- citable (persistent URL/URN) at least on the level of binding units; preferably even to single images or details

The results of a digitisation project are almost always displayed on the websites of the home library (e.g. digital collections with local viewers). Sometimes, funding organisations have developed their own standardized viewers, e.g. the German DFG-viewer.<sup>51</sup>

For *Europeana Regia*, a project website is under construction. In the context of *Europeana Regia*, the partner institutions will in addition continue to use their own customary viewers. Besides these, different technical solutions may be used to display individual manuscripts, e.g. 'turning-the-pages' devices or even 3D-solutions.

Another important aspect is the question of which image shall be shown as thumbnail or teaser, e.g. for use in the Europeana database. Where medieval manuscripts are concerned, it is advantageous for the user to see on a thumbnail the beginning of the text, which is almost never contained in the first image of the series. This gives a better impression of the original than the first image, usually of a cover of a manuscript or the folder of a fragment, because these may be much later additions to the original or are not specific in most cases. Whereas the ZEND does not allow the choice of any image other than the first for a thumbnail, other technical solutions, for example that used in the UNESCO portal World Digital Library,<sup>52</sup> are more flexible. This issue will be discussed with the project partners and with *Europeana*; new results will be reported in D5.3.

### Dissemination

For the sake of visibility, the exchange and transmission of digitised images and links, or at least the corresponding metadata (including links), becomes an important issue. Whereas in early digitisation projects, the digitised images were often published only on CD-ROM, publication on the World Wide Web has today become standard. A format for the exchange of metadata must therefore be used which is sufficiently general to be widespread and at the same time allows special modifications tailored to the needs of different user communities. The Open Archive Protocol (OAI) can today be considered an international standard for such an exchange of metadata,<sup>53</sup> together with all other XML-based metadata formats which are supported by the OAI protocol (e.g. MARCxml, MABxml, EAD, METS/MODS, TEI P5, see WP2 and 3). The dissemination of metadata (and sometimes images) is possible through -

- The Registry of Digital Masters. This joint DLF/OCLC venture, a subset of the OCLC's World Catalog,<sup>54</sup> has set the standard for the creation of metadata for digital objects in the United States.
- Portals
  - International and national portals for digital content in general (e.g. *Europeana*<sup>55</sup>)
  - Portals specializing in historical manuscripts and books, e.g.
    - the CERL Portal<sup>56</sup>
    - *Manuscripta Medievalia*, the German national manuscripts portal<sup>57</sup>
    - *Manuscriptorium*, an EU-funded manuscript digital library within the ENRICH project<sup>58</sup>

<sup>51</sup> <<http://daten.digitale-sammlungen.de/~db/0003/bsb00037298/images/index.html?id=00037298&fip=eayayztsewqeyayxssdasyztsqrseayaxs&no=3&seite=11>>.

<sup>52</sup> <<http://www.wdl.org>>.

<sup>53</sup> <<http://www.openarchives.org>>.

<sup>54</sup> < <http://www.oclc.org/de/de/digitalregistry/default.htm>>.

<sup>55</sup> [www.europeana.eu/portal](http://www.europeana.eu/portal).

<sup>56</sup> <<http://cerl.epc.ub.uu.se/sportal/>>.

<sup>57</sup> <http://www.manuscripta-medievalia.de>.

- e-codices (see above, p. 13-14)
- *digital scriptorium* (see above, p. 14)
- Library catalogues (such as union catalogues) and databases:
  - Karlsruhe Virtual Catalogue<sup>59</sup>
  - Worldcat<sup>60</sup>
- Lists distributed by means of RSS feeds, e.g. for the use of libraries, museums, archives. An RSS document (which is called a 'feed', 'web feed', or 'channel') includes full or summarized text, plus metadata such as publishing dates and authorship. They are used to subscribe to timely updates from favoured websites or to aggregate feeds from many sites into one place.

#### 7.6.4 Persistent Identifiers (PI)

At the beginning of digitisation, the importance of being able to quote a digital resource was underestimated. In fact, the possibility to cite correctly is essential for the use of an Internet source for academic purposes. The introduction of new citation rules was necessary because digital resources in the World Wide Web cannot be cited in the same way as printed books, CDs or other offline media. If a document is online, it is necessary to access it from other documents or databases. Therefore it has to be unambiguously addressed.

A positive effect of the citation rules for the Internet is that they refer to physical images. This means that referencing becomes definite. The target of citation should therefore be the individual image. For this, it is necessary that the naming conventions for individual files should respect a numerically ordered sequence (for example 0001, 0002, 0002a, 0003 etc.; image 0002a has been added). For new technical possibilities such as details and zooms, new citation rules must be developed in accordance with rising user expectations.

When applied to manuscripts, this is an advantage because manuscripts often contain no pagination at all or a confusing pagination (for example, contradictory paginations, missing leaves, jumps, etc.). Hence a citation by content should be complemented by a formal citation. Moreover, images which do not belong to the corpus can be addressed in this way, for example, images of the cover, spine, watermarks, etc.

Apart from this granularity, the persistence of a resource is the most important aim. For this reason, various methods have been developed to guarantee the long-term availability of documents, for example PURL (Persistent Uniform Resource Locator), URN (Uniform Resource Name), DOI (Digital Object Identifier) and Handle, also known as Persistent Identifiers (PI). In Germany, the most widespread one is the URN hosted by the German National Library. The basic idea of PIs is the strict separation of the identification of an object by an unambiguous string and its location. PIs are used instead of URLs (Uniform Resource Locators) to identify the digital objects and it is only as a second step that this identifier is resolved to the corresponding URL. PI services require a technical and organisational infrastructure to guarantee their long-term availability. This means that technical effort and manpower must be invested. Consequently, a strict selection of document classes for which PIs are used is needed.

However, various problems attach to the handling of URNs. Theoretically, once a URN has been allocated to a digital object, it should never be changed. In the case of faulty images or earlier and inferior versions of image scans, however, it has been common practice to replace these inferior images by the superior image when available in order to minimize

---

<sup>58</sup> <http://enrich.manuscriptorium.com>.

<sup>59</sup> <[www.ubka.uni-karlsruhe.de/kvk.html](http://www.ubka.uni-karlsruhe.de/kvk.html)>.

<sup>60</sup> <[www.worldcat.com](http://www.worldcat.com)>.

storage space and effort. The impact of this practice on the quality of persistent identification remains to be discussed.



## 8 Relation of Images to Metadata

### 8.1 Different Types of Metadata

#### 8.1.1 Bibliographical Metadata

- A fundamental issue in the field of bibliographical metadata which has not been agreed upon is the relation of metadata for digital objects to the metadata for the corresponding originals. In the context of OCLC's *Registry of Digital Masters*, separate catalogue records for digital objects are recommended.<sup>61</sup>
- In the *Europeana Regia* project, two sets of metadata will be available: concise metadata v. full academic records (see WP 2 & 3).
- Only concise metadata sets will be harvested by *Europeana*. These are delivered by the content providers in different formats, e.g. XML-based standards (e.g. TEI/P5, EAD), national and international library formats (e.g. MARC and MAB).
- The conversion of these different formats, their mapping onto ESE is the subject of WP 2 and 3
- The conversion will be carried out not by the project partners themselves but by external aggregators

#### 8.1.2 Technical Metadata

##### a) TIFF Header

'The TIFF image format contains a flexible, extendable metadata structure based on a header that contains numerically tagged data fields. The baseline tags are listed and described in Part 1 (Baseline TIFF) of the TIFF 6.0 specification, where they are described as Required Fields. These 36 tags cover mainly technical metadata that any TIFF-specification compliant application must be able to read. The specification defines four image types -- bilevel, grayscale, palette-color, and full-color -- each of which has different baseline tags. The specification also lists 60 extension tags that applications should but are not required to read. Additional private tags may be reserved by software developers for specific applications; many of these have become very public and are widely adopted.'<sup>62</sup>

Examples of the technical information typically contained in TIFF headers are:

- the length and width in pixels
- sampling
- compression
- resolution
- size in bytes
- production information, e.g. references to scanner and software
- the date of creation

The header should also contain information on whether and how the images have been subject to digital post-processing (such as cropping, rearrangements of the image sequence, etc.).

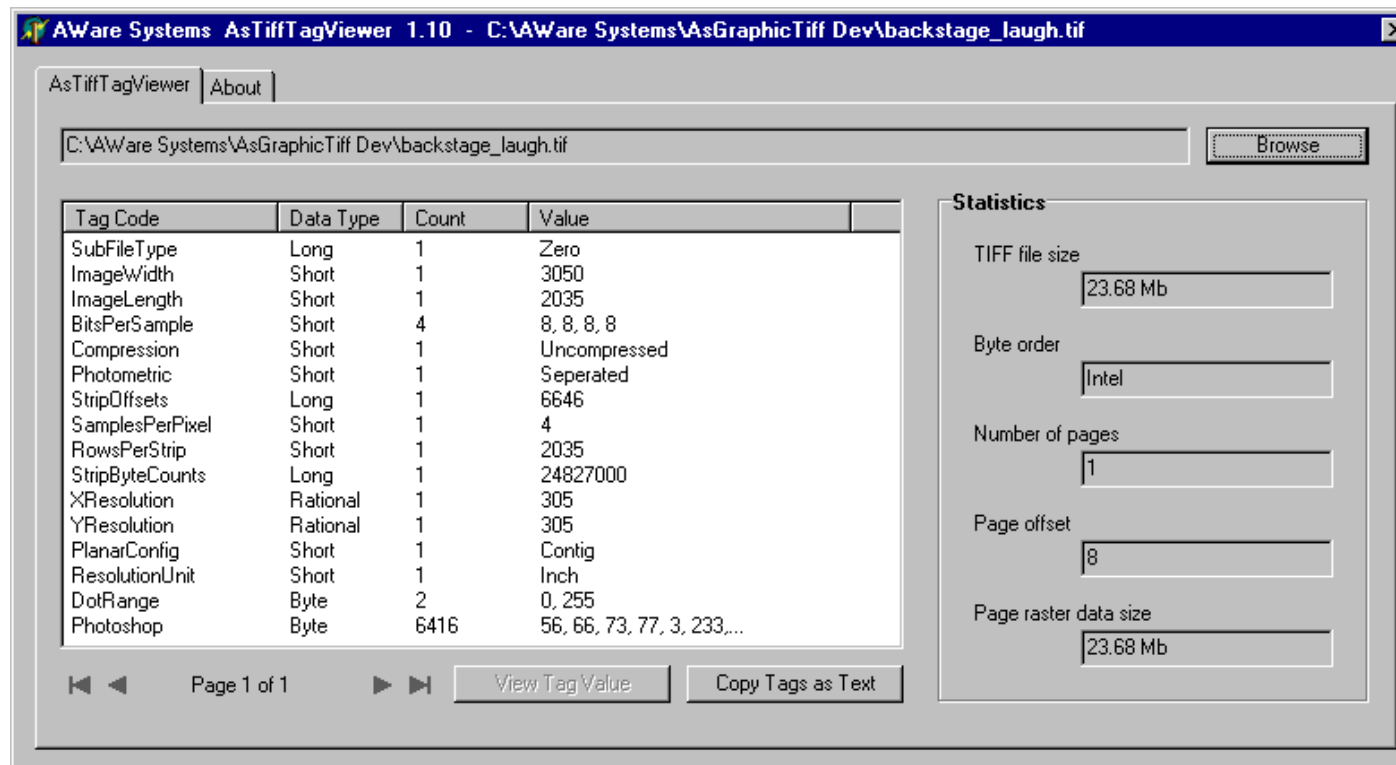
<sup>61</sup> On the metadata used for the US Registry of Digital Masters, see Registry of Digital Masters: *Record Creation Guidelines* (Version 2., May 2007), available at: <http://old.diglib.org/collections/reg/DigRegGuide200705.htm>.

<sup>62</sup> <<http://www.digitizationguidelines.gov/term.php?term=tiffheader>>; more detailed information is available at [http://www.digitalpreservation.gov/formats/content/tiff\\_tags.shtml](http://www.digitalpreservation.gov/formats/content/tiff_tags.shtml)>.

Some points of administrative information, which can be registered in the TIFF header as well, should be recorded, e.g. on -

- the creator of the file
- the owner of the image and copyright (name of the holding institution)

In order to view the TIFF headers of a given file, software tools are needed: these show the set of TIFF tags that constitute the first page. It shows known tags as well as proprietary undocumented tags ('private tags' or 'custom tags'):<sup>63</sup>



## b) PREMIS

In the context of digital preservation, an emerging standard is PREMIS (PREservation Metadata: Implementation Strategies),<sup>64</sup> which contains the following elements (example from the Biblioteca Nacional de España):<sup>65</sup>

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
- <premis:premis version="2.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
```

```
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:premis="info:lc/xmlns/premis-v2">
```

```
xsi:schemaLocation="info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premisv2-
```

<sup>63</sup> Example from < <http://www.awaresystems.be/imaging/tiff/astifftagviewer/screenshots.html>>.

<sup>64</sup> For details on PREMIS, see <<http://www.loc.gov/standards/premis/>>.

<sup>65</sup> Example from the 'Proceso de digitalización en la Biblioteca Nacional de España', <[http://www.bne.es/es/Catalogos/BibliotecaDigital/docs/ProcesoDigitalizacionBNE\\_10012011\\_version3.pdf](http://www.bne.es/es/Catalogos/BibliotecaDigital/docs/ProcesoDigitalizacionBNE_10012011_version3.pdf)>p. 22-3.

**0.xsd»>**

```

- <premis:object xsi:type=>premis:representation> xmlID=>VC_002307-006>
- <premis:objectIdentifier>
<premis:objectIdentifierType>899$j</premis:objectIdentifierType>
<premis:objectIdentifierValue>VC/2307/6</premis:objectIdentifierValue>
</premis:objectIdentifier>
- <premis:preservationLevel>
<premis:preservationLevelValue>full</premis:preservationLevelValue>
<premis:preservationLevelDateAssigned>20070529</premis:preservationLevelDateAssigned>
</premis:preservationLevel>
<premis:originalName>VC_002307-006</premis:originalName>
</premis:object>
- <premis:object xsi:type=>premis:file>>
- <premis:objectIdentifier>
<premis:objectIdentifierType>File</premis:objectIdentifierType>
<premis:objectIdentifierValue>VC_002307-006_0001</premis:objectIdentifierValue>
</premis:objectIdentifier>
- <premis:preservationLevel>
<premis:preservationLevelValue>full</premis:preservationLevelValue>
<premis:preservationLevelDateAssigned>20070529</premis:preservationLevelDateAssigned>
</premis:preservationLevel>
- <premis:objectCharacteristics>
<premis:compositionLevel>0</premis:compositionLevel>
<premis:size>1234567</premis:size>
- <premis:format>
- <premis:formatDesignation>
<premis:formatName>image/tiff</premis:formatName>
<premis:formatVersion>6.0</premis:formatVersion>
</premis:formatDesignation>
</premis:format>
- <premis:creatingApplication>
<premis:creatingApplicationName>Omniscan</premis:creatingApplicationName>
<premis:creatingApplicationVersion>11.0</premis:creatingApplicationVersion>
<premis:dateCreatedByApplication>20090102</premis:dateCreatedByApplication>
</premis:creatingApplication>
- <premis:objectCharacteristicsExtension>
- <mix:mix xmlns:mix=>http://www.loc.gov/mix/v20>
xsi:schemaLocation=>http://www.loc.gov/mix/v20
http://www.loc.gov/standards/mix/mix20/mix20.xsd>>

```

```

- <mix:BasicDigitalObjectInformation>
<mix:byteOrder>big endian</mix:byteOrder>
- <mix:Compression>
<mix:compressionScheme>Uncompressed</mix:compressionScheme>
</mix:Compression>
</mix:BasicDigitalObjectInformation>
- <mix:BasicImageInformation>
- <mix:BasicImageCharacteristics>
<mix:imageWidth>5530</mix:imageWidth>
<mix:imageHeight>3210</mix:imageHeight>
- <mix:PhotometricInterpretation>
<mix:colorSpace>RGB</mix:colorSpace>
</mix:PhotometricInterpretation>
</mix:BasicImageCharacteristics>
</mix:BasicImageInformation>
- <mix:ImageCaptureMetadata>
- <mix:ScannerCapture>
<mix:scannerManufacturer>Zeutschel</mix:scannerManufacturer>
- <mix:ScannerModel>
<mix:scannerModelName>OS 10000-90 TT</mix:scannerModelName>
<mix:scannerModelSerialNo>52008</mix:scannerModelSerialNo>
</mix:ScannerModel>
</mix:ScannerCapture>
</mix:ImageCaptureMetadata>
- <mix:ImageAssessmentMetadata>
- <mix:ImageColorEncoding>
- <mix:BitsPerSample>
<mix:bitsPerSampleValue>8</mix:bitsPerSampleValue>
</mix:BitsPerSample>
<mix:samplesPerPixel>3</mix:samplesPerPixel>
</mix:ImageColorEncoding>
</mix:ImageAssessmentMetadata>
</mix:mix>
</premis:objectCharacteristicsExtension>
</premis:objectCharacteristics>
<premis:originalName>VC_002307-006_0001.tif</premis:originalName>
- <premis:storage>
- <premis:contentLocation>
<premis:contentLocationType>filepath</premis:contentLocationType>

```

```

<premis:contentLocationValue>VC_002307-006</premis:contentLocationValue>
</premis:contentLocation>
<premis:storageMedium>HD 001 Alta</premis:storageMedium>
</premis:storage>
- <premis:relationship>
<premis:relationshipType>structural</premis:relationshipType>
<premis:relationshipSubType>is included in</premis:relationshipSubType>
- <premis:relatedObjectIdentification RelObjectXmlID="VC_002307-006">
<premis:relatedObjectIdentifierType>899$j</premis:relatedObjectIdentifierType>
<premis:relatedObjectIdentifierValue>VC/2307/6</premis:relatedObjectIdentifierValue>
<premis:relatedObjectSequence>1</premis:relatedObjectSequence>
</premis:relatedObjectIdentification>
</premis:relationship>
</premis:object>
</premis:premis>

```

### 8.1.3 Structural Metadata

(see WP 2 & 3)

In the case of manuscripts, structural metadata are an important aid -

- to display the structure and contents of the manuscript
- to make navigation and orientation within the often complex codices easier

Elements which should be represented in these structural metadata are -

*a) Mandatory Elements:*

- image numbers

*b) Recommended Elements:*

- numbering scheme(s) with all page numbers

*c) Desirable Elements (which may be added at a later stage):*

- textual divisions, such as beginnings and ends of texts
- important quotations from the manuscript (e.g. titles, headings, incipits, explicits, colophons)
- illustrations (initials, drawings, illumination)
- important other elements (e.g. marks of ownership, glosses, ... ) which, of course, may vary according to users' interests

The terminology which should be used is still open to discussion; no regional, national or international standards have been created. It is an important question in this context, to which extent existing terminology, which was usually developed for printed material, should be modified to accommodate manuscript-specific peculiarities. Examples of such terminology are:

- METS <sup>66</sup>
- the BSB Strukturdatenset
- the TEI-compliant terminology of HAB <sup>67</sup>

## 8.2 Creation of Links to Metadata

Several project partners have been able to develop and put into practice automatic procedures for several aspects of the workflow mentioned so far. One important aspect covered by these automatic procedures is the creation of links to the different sets of metadata. Examples of such software solutions are the BSB-ZEND, and refNum, developed by the BnF.

*Example: BSB-ZEND (Zentrale Nachweis- und Erfassungsdatenbank, Central Digital Asset Management System)*

In order to standardize production processes, a software tool with different document and workflow management components was developed by the MDZ in 2003. ZEND controls the entire production process from the preparation to the automated transfer of the digital master files to archival storage. ZEND is an electronic publishing system for (retro-) digitisation based on two open-source software modules, namely a LAMP system (Linux, Apache, MySQL and PHP / Perl) and the EXtensible Markup Language (XML) publishing framework Cocoon with the Lucene/Solr-search platform. It combines different components of the document, web content and workflow management, and creates a comprehensive production environment. In daily production, ZEND is very successful, as the over 500,000 processed files prove (including in the Google project). It is able to -

- generate a digitisation order
- support the scanning process
- manage metadata:
  - import a selection of bibliographical data, which are obtained from the Bavarian Union Catalogue via a Z39.50 interface or manual addition of metadata
  - export bibliographical metadata in different formats (e.g. Dublin Core, METS)
  - create structural metadata (ToCs, pagination, which can also be used for quality control)
  - administer technical metadata (e.g. Mime-type, size, colour profiles)
  - store legal information (copyright, access rights)
  - store administrative data (e.g. provenance, date, history, project)
- generate a URN (which is added to the Union catalogue entry):
  - Example: urn:nbn:de:bvb:12-bsb00001119-6 [bvb:12 indicates the id-number of the library within the Bavarian Union Catalogue – BibliotheksVerbund Bayern]
  - The assignment is carried out locally, but the administration of the URN and the central link resolving is part of the duties of the DNB (Deutsche Nationalbibliothek = German National Library in Frankfurt).
- fetch digital images from the production servers inhouse and, if available, also those of external service providers

<sup>66</sup> <http://www.loc.gov/standards/mets>.

<sup>67</sup> <http://diglib.hab.de/rules/concordances/structMD.xml>.

- delete digital images on the production servers
- compile 'digital books' from them which can be leafed through virtually
- create formats for web presentation automatically (JPG, PDF, etc.)
- organize the long-term storage of image data on external servers
- hand over data to the OCR workflow (if necessary)
- give feedback to the local library system once a digital object is available and add the URN of the digital object to the catalogue entry

## Appendix: Naming Conventions of Partners:

As most partners in *ER* already have naming conventions, which fit and are consistent within their respective workflows, they should not be changed in order to avoid problems within the native system. To this end, a concordance was produced by the end of month 3, informing partners about the different conventions.

### *BSB*

- The filenames are based upon a specific eight digit running number, indicating the digitised document: e.g. bsb00001119.
- This ID is automatically created and assigned by the Zentrale Erfassungs- und NachweisDatenbank (ZEND) of the Munich Digitization Center as part of the digitisation process. The filename also indicates the running sequence of the images, hence bsb00001119\_00001.tif for the first image.
- Example: urn:nbn:de:bvb:12-bsb00001131-6
- Within the digital presentation the persistent link to single pages is given in the following way: <[http://daten.digitale-sammlungen.de/bsb00045283/image\\_21](http://daten.digitale-sammlungen.de/bsb00045283/image_21)>.

### *BnF*

- The whole digital document (1 original manuscript = 1 digital document) receives a number ID: 8 digits, the first 3 identifying the department. For instance, 060 or 084 identifies 'Département des manuscrits' and a complete ID can be 06000216. All the digital images are kept in a directory named after this number ID: 06000216. Inside this directory, all TIFF files are named with a sequential numeric ID beginning with T + 7 digits.
- On *Gallica*, each image has its own persistent URL. For instance, in Français 12595, folio number 20's URL is:  
<<http://gallica.bnf.fr/ark:/12148/btv1b60002167.item.f45>> (45th image file of the document).

### *KBR*

- The naming of the digitised item consists of a simple running number of 8 digits, separated by a slash after the second and the fifth digit, e.g. 00/000/001 for the first document number and so on.
- In the directory where the image data are stored, the metadata, the TIFF files, and the files for web display are maintained. Files are numbered with a 4 digit number in scanning order. Manuscripts can show the original pagination (e.g. 1r, 1v), added to the scanned number with an underscore, e.g. 00/000/001/0006\_1v.jpg.

### *HAB*

- Images are numbered in a row, starting from 00001. Numbers have 5 digits.
- Images of completely digitised manuscripts are stored in one folder. The path to the folder includes the material type, here 'mss' for manuscripts and the shelfmark. The shelfmark of the digitised book is converted according to a short set of rules:
- Spaces and dots are converted to hyphens
- There cannot be two hyphens in a row. These would be collapsed to one hyphen



- There could not be a hyphen at the very end
- Format signs are transformed into the letter 'f'
- The collection name will be represented in short form and in lowercase

Examples:

Cod. Guelf. 50.2 Aug. 4° will become 50-2-aug-4f

Cod. Guelf. 3 Weiss. will become 3-weiss

Full paths: <<http://diglib.hab.de/mss/79-weiss/00009.jpg>>

The names of all images that belong to one manuscript are referenced in one file named 'facsimile.xml'. This file would contain metadata like the concordance between image numbers and 'real' page numbers like '3r'. This file is placed in the same folder.

Example: <<http://diglib.hab.de/mss/79-weiss/facsimile.xml>>

## *BHUV*

All digitised images of manuscripts are included in the same folder named 'ms', where each document has a folder labelled with its shelfmark (consisting of a number of 4 digits) and which contains all the images as well as a file '\*txt' with the technical and structural metadata.

The images are identified by a name formed by the library code (uv), the manuscript shelfmark (i.e. ms\_0227), the image sequence (i.e. 0001) as well as, finally, a reference to the part of the manuscript (i.e. \_f5r) and the file extension (.tif).

Examples:

uv\_ms\_0375\_0001\_ea

uv\_ms\_0375\_0002\_fg1

uv\_ms\_0375\_0003\_fg2

uv\_ms\_0375\_0011\_f1r

uv\_ms\_0375\_0048\_ep

uv\_ms\_0375\_0049\_el

uv\_ms\_0375\_0050\_cc

## Literature

(Minor Internet sources are quoted only in footnotes):

Biblioteca Nacional de España: 'Proceso de digitalización en la Biblioteca Nacional de España' (January 2011), available at:  
<[http://www.bne.es/es/Catalogos/BibliotecaDigital/docs/ProcesoDigitalizacionBNE\\_10012011\\_version3.pdf](http://www.bne.es/es/Catalogos/BibliotecaDigital/docs/ProcesoDigitalizacionBNE_10012011_version3.pdf)>.

Brantl, Markus/Schäfer, Irmhild, 'Massendigitalisierung von Handschriften und Alten Drucken. Herausforderungen und Erfahrungen an der Bayerischen Staatsbibliothek', *Symposium on Digital Imaging of Ancient Textual Heritage: Technological Challenges and Solutions. Helsinki, Finland, 28-29 Oktober 2010. Centre of Excellence (Academy of Finland)*, pp. 175-197. Available at: <<http://www.eikonopoiia.org/files/Eikonopoiia-2010-Proceedings.pdf>> (English version in print).

Brantl, Markus/Ceynowa, Klaus/Fabian, Claudia/Meßmer, Gabriele/Schäfer, Irmhild: 'Massendigitalisierung deutscher Drucke des 16. Jahrhunderts – Ein Erfahrungsbericht der Bayerischen Staatsbibliothek', in: *Zeitschrift für Bibliothekswesen und Bibliographie* 56/6 (2009), pp. 327-338.

Caplan, Priscilla (for the Library of Congress): *Understanding PREMIS: an overview of the PREMIS Data Dictionary for Preservation Metadata*, available at:  
<<http://www.loc.gov/standards/premis/understanding-premis.pdf>>.

Consorti de Biblioteques Universitàries de Catalunya: 'Estandàres de digitalització: requeriments mínims (actualitzat en octubre de 2010)', available at:  
<[http://www.recercat.net/bitstream/2072/97497/3/1010EstandardsRequerimentsMinims\\_Public\\_spa.pdf](http://www.recercat.net/bitstream/2072/97497/3/1010EstandardsRequerimentsMinims_Public_spa.pdf)>.

Deutsche Forschungsgemeinschaft, Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS): *DFG-Praxisregeln, 'Digitalisierung' (Stand April 2009)*, available at:  
<[http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln\\_digitalisierung.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung.pdf)>.

Digital Preservation Coalition: *Preservation Management of Digital Materials: The Handbook*, available at:  
<<http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts>>.

Digital Scriptorium: 'Imaging standards', available at:  
<[https://www1.columbia.edu/sec/cu/libraries/bts/digital\\_scriptorium/technical/image\\_contribution/standards.html#reftargets](https://www1.columbia.edu/sec/cu/libraries/bts/digital_scriptorium/technical/image_contribution/standards.html#reftargets)><[https://www1.columbia.edu/sec/cu/libraries/bts/digital\\_scriptorium/technical/image\\_contribution/standards.html#reftargets](https://www1.columbia.edu/sec/cu/libraries/bts/digital_scriptorium/technical/image_contribution/standards.html#reftargets)>.

Federal Agencies Digitization Initiative Still Image Working Group: *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*, available at:  
<[http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image-Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf)>.

Florida Digital Archive: *Recommended Data Formats for Preservation Purposes in the Florida Digital Archive*, available at:  
<<http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>>.

'Freiburger Richtlinien': *Richtlinien für die Digitalisierung mittelalterlicher Handschriften*, herausgegeben vom Kuratorium für die Digitalisierung mittelalterlicher Handschriften (Codices Electronici Confoederationis Helveticae).

Gertz, Janet, 'Vendor Relations', ch. 8 in: Maxine K. Sitts, ed. for the Northeast Document Conservation Center: *Handbook for Digital Projects: A Management Tool for Preservation and Access* (Andover, MA: 2000), available at:  
<<http://www.ncsi.iisc.ernet.in/raja/is214/is214-2006-01-04/dman.pdf>>

Gillesse, Robèrt: 'Alternative File Formats for Storing Master Images of Digitisation Projects', available at:  
<[http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/alternative\\_file\\_formats\\_for\\_storing\\_masters\\_2\\_1.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/alternative_file_formats_for_storing_masters_2_1.pdf)>.

IFLA: *Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives*, available at:  
<<http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf>>

Maier, Gerald, 'Colormanagement bei der Farbdigitalisierung von Archivgut: Grundlagen, Hintergründe und Ausblick', in: *Digitale Archive und Bibliotheken*, ed. Hartmut Weber and Gerald Maier (Stuttgart, 2000), pp. 179-199.

Mayer, Manfred: 'Digitalisierung mittelalterlicher Handschriften an der Universitätsbibliothek Graz', available at:  
<[http://cool.conservation-us.org/iada/ta99\\_185.pdf](http://cool.conservation-us.org/iada/ta99_185.pdf)>.

Registry of Digital Masters: 'Record Creation Guidelines' (Version 2., May 2007), available at: <<http://old.diglib.org/collections/reg/DigRegGuide200705.htm>>.

Research Libraries Group/Digital Library Federation: *Guides to Quality in Visual Resource Imaging* (issued 2000 by the Council on Library and Information Resources), available at:  
<<http://imagedigital.esteticas.unam.mx/PDF/Guides.pdf>>.

Rieger, Oya Y., 'Establishing a Quality Control Program', in: *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, CA: Research Libraries Group: 2000), pp. 61-97.

Sturm, Marlene: 'Processing Digitized Materials in European Libraries: A Survey Report (March 2011)', available at: <<http://www.kb.nl/pr/pers/download/>>.

Thaller, Manfred, 'From the Digitized to the Digital Library', *D-Lib Magazine* 7 (2001), available at: <<http://www.dlib.org/dlib/february01/thaller/02thaller.html>>

U.S. National Archives and Records Administration (NARA): *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*, available at <<http://www.archives.gov/preservation/technical/guidelines.pdf>>.

Williams, Don, 'An Overview of Image-Quality Metrics', in *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, CA: Research Libraries Group: 2000), pp. 72-73.

### Europeana Regia Questionnaire:

This questionnaire was sent to more than 20 institutions with experience of digitisation projects for medieval manuscripts. The following institutions kindly replied by the middle of April 2011:

Barcelona, Biblioteca de Catalunya

Frankfurt am Main, Universitätsbibliothek

Fribourg (CH), e-codices, University of Fribourg

Göttingen, Niedersächsische Staats- und Universitätsbibliothek  
Helsinki, National Library of Finland  
Kopenhagen, Det Kongelige Bibliotek  
London, The British Library  
Lund, Universitetsbiblioteket  
Lyon, Bibliothèque municipale de Lyon  
Madrid, Biblioteca de la Universidad Complutense  
Madrid, Ministerio de Cultura  
Stuttgart, Württembergische Landesbibliothek  
Wolfenbüttel, Herzog-August-Bibliothek  
Yale, Beinecke Rare Book and Manuscript Library