



europaana

Regia

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

**D5.3 Quality Management**

<b>Deliverable number</b>	D5.3
<b>Nature</b>	R
<b>Dissemination level</b>	PP
<b>Delivery date</b>	Due Date : 31 October 2011 Actual Date :
<b>Status</b>	<b>Final</b>
<b>Work package number</b>	WP5
<b>Lead beneficiary</b>	BSB
<b>Author(s)</b>	Rahel Bacher, Claudia Fabian, Wolfgang -Valentin Iikas, Carolin Schreiber (BSB)



## Document history

### Revisions

Version	Date	Author	Changes
0.1	04/02/2011	Rahel Bacher, Carolin Schreiber, Wolfgang-Valentin Ikas, Claudia Fabian	Created initial version
0.2	07/03/2011	Wolfgang-Valentin Ikas	Input from the partners
0.3	02/05/2011	Rahel Bacher, Carolin Schreiber	Draft version
0.4	16/05/2011	Jérôme Boillet (BnF)	Major input
0.5	09/2011	Rahel Bacher, Carolin Schreiber	Input from project partners

### Reviews

Date	Name	Result
29/04/2011	Jérôme Boillet (BnF)	Additional notes

### Signature/Approval

Approved by (signature)	Date

Accepted by at European Commission (signature)	Date

## **1. Executive Summary**

This deliverable is dedicated to giving guidelines for optimal quality management for digitization projects in the field of medieval manuscripts and other prestigious objects. It is based on the comprehensive understanding of quality management in the international standards ISO 9000, 9001, 9004 and 19011. These guidelines are applicable both to inhouse digitization and to projects including external service providers. Based on the experience of the *Europeana Regia* project partners and other international expert organisations, it is also intended to harmonize and adapt current practice 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 institutions some procedures are still rather unfamiliar as they started digitization in the more recent past. Their experience is complemented by the results obtained in a questionnaire sent to leading institutions in the field of digitization, which was answered by 14 libraries in both Europe and the United States.

## Table of Contents

1. Executive Summary.....	3
Table of Contents .....	4
2. Aim and scope of WP 5 .....	6
2.1. Original Definition of the Work Package.....	6
2.2 Re-definition of the Name and Scope of WP 5 according to the Project Plan ...	7
2.3 New Content of Work package 5 according to the Project Requirements.....	7
3. External Influences on General Quality Aspects in Digitization .....	9
3.1 Customer / User requirements .....	9
3.2 Suppliers / Content Providers .....	11
4. Conceiving and leading a quality strategy on digitisation projects of manuscripts	13
4.1 Planning stage (instruction phase) .....	13
4.1.1 Definition of Quality Management Scope and Aims.....	13
4.1.2 Quality Staff .....	14
4.1.3 Planning of the quality tools and intervention measures.....	14
4.2 Pre-production, Test phase .....	14
4.2.1 Testing Scanning Instructions .....	15
4.2.2 Calibrating hardware (for technical details, see D4.1) .....	16
4.2.3 Testing the rendering of gold .....	17
4.2.4 Testing the integration of metadata to the numerical documents .....	17
4.2.5 Developing the Quality Assurance Plan .....	17
4.3 During Production (production phase).....	18
4.3.1 Preventive action .....	19
4.3.1.1. Training of staff.....	19
4.3.1.2. Survey of the physical condition of the objects by a conservator.....	19
4.3.1.3. Intermediate Tests.....	20
4.3.2. Quality parameters and checking operations on digital documents.....	20
4.3.2.1 Methods.....	23
4.3.2.2. Automated procedures .....	23
4.3.2.3 Manual quality control.....	24
4.3.2.4 Colorimetric control (see D4.1) .....	26
4.3.3 Audits with service providers .....	26
4.3.3.1. Preparation of the audit (3 weeks).....	26
4.3.3.2. Audit: visit to the site of production (max. 2 days) .....	26
4.3.3.3. After visit of the site of production (2 weeks) .....	27
4.3.4 Correction and Rework .....	27
4.3.5 Corrective Action.....	27
4.3.6 Release .....	28
5. Current practice of Quality management at the Europeana Regia Partner Libraries	29
5.1. Framework and organization.....	29
5.1.1. Seat of quality management .....	29
Bayerische Staatsbibliothek: .....	29
Bibliothèque nationale de France: .....	29

Bibliothèque royale de Belgique .....	29
Biblioteca Historica – Universitat de València .....	29
Herzog August Bibliothek Wolfenbüttel .....	30
5.1.2. Quality management staff.....	30
Bayerische Staatsbibliothek: .....	30
Bibliothèque nationale de France .....	30
Bibliothèque royale de Belgique .....	31
Biblioteca Historica – Universitat de València .....	31
Herzog August Bibliothek Wolfenbüttel .....	31
5.1.3. Equipment .....	31
Bayerische Staatsbibliothek: .....	31
Bibliothèque nationale de France .....	32
Bibliothèque royale de Belgique .....	32
Biblioteca Historica – Universitat de València .....	33
Herzog August Bibliothek Wolfenbüttel .....	34
5.1.4. Documentation of quality management .....	34
Bayerische Staatsbibliothek: .....	34
Bibliothèque nationale de France .....	35
Bibliothèque royale de Belgique .....	35
Biblioteca Historica – Universitat de València .....	36
Herzog August Bibliothek Wolfenbüttel .....	36
5.1.5. Timing of quality control.....	36
Bayerische Staatsbibliothek: .....	36
Bibliothèque nationale de France .....	36
Bibliothèque royale de Belgique .....	36
Biblioteca Historica – Universitat de València .....	36
Herzog August Bibliothek Wolfenbüttel .....	37
5.1.6. Time delay between digitization, quality control and release.....	37
Bayerische Staatsbibliothek: .....	37
Bibliothèque nationale de France .....	37
Bibliothèque royale de Belgique .....	37
Biblioteca Historica – Universitat de València .....	37
Herzog August Bibliothek Wolfenbüttel .....	37
5.2 Parameters of Quality Control.....	37
Bayerische Staatsbibliothek: .....	37
Bibliothèque nationale de France .....	39
Bibliothèque royale de Belgique .....	39
Biblioteca Historica – Universitat de València .....	39
Herzog August Bibliothek Wolfenbüttel .....	40
5.3. Methods and Workflow.....	40
Bayerische Staatsbibliothek: .....	40
Bibliothèque nationale de France .....	41
Bibliothèque royale de Belgique .....	42
Biblioteca Historica – Universitat de València .....	43
Herzog August Bibliothek Wolfenbüttel .....	43
References: .....	44
Closely connected deliverables of the Europeana Regia project:.....	45

## 2. Aim and scope of WP 5

### 2.1. Original Definition of the Work Package

In the project plan for Europeana Regia, WP 5 was formulated in the following way: "The objective of the work package is the integration of the digitized images to make them harvestable by *Europeana*". According to the description of work, the work package was defined as follows:

*"The work package includes different tasks:*

- *Integration of the images in each library's database*
- *Checking of the links between the images and the relevant metadata*
- *Checking of the OAI-data (after mapping)*
- *Preparation and updating of special subsets in OAI repositories.*

*Part of these tasks are automatical, after a first test period. But the participants will have to check precisely several pieces of information that cannot be verified automatically, because Medieval manuscripts aren't normal books. They lack of identifying items like title pages or authors. Sometimes, even the page numbers and the numbering can be faulty:*

- *pages left out and not numbered at all;*
- *pages cut out or otherwise removed and therefore the numbering skips (correctly) some numbers;*
- *numbers skipped, i.e. all pages are in place and none missing, but anyway the numbering jumps from one number to another not succeeding one;*
- *numbers appear more than once, thus violating the numbering scheme;*
- *pages, items or the like can be bound together with the "historical" manuscript pages, thus having to be included in the digitisation but violating the concordance of image number and numbered pages of the manuscript.*

*In manuscript research, it is common to refer to a manuscript page by its foliation (Medieval manuscripts are often foliated instead of paginated.) Foliation normally starts at the first page that really is written upon, but not necessarily, as sometimes the numbering starts at the first leaf of a certain material (usually the first parchment leaf, if there are extra leaves in paper). Sometimes even that can be different from one catalogue to another, e.g. one catalogue might count only leaves that have been written upon, another catalogue might include leaves bound together with the "original" manuscript leaves, another might "correct" the numbering that has been carried out in the manuscript whereas another one refers to the faulty numbering (stating these are faults)...*

*In digitisation centres, it is common to supply a running number as image names. These numbers have to be mapped to the conventional naming of the pages (leaves) of the manuscript. If now one of the before listed violations of the numbering scheme exists, the concordance cannot longer be compiled (semi-automatically) but manual interference is needed. This is necessary anyway as especially extra leaves need manual control.*

*As the combination of "faults" in the numbering, in the referencing by the catalogues or even in the digitisation process (pages not in the first run digitised but only added afterwards, thus disturbing the sequence) can appear in countless combinations, a careful examination of:*

- a) the numbering in the manuscript,*
- b) the digitisation results,*
- c) the source catalogues*

*has to be carried out, which will take a lot of time and requires a lot of knowledge about the manuscript culture and scientific habits.*

*The careful checking during the integration of the images could be used to acquire data that is referred to as "structural metadata". Common entries in structural metadata would be:*

- *numbering scheme(s);*
- *textual divisions;*
- *certain quotations like titles, heads, incipits, colophons;*
- *illustrations like initials, drawings;*
- *other important respectively eye-catching items like included materials*

*A list of terms that could (or even should) be listed already exists at <http://dfg-viewer.de/strukturdatenset/>; a TEI-compliant taxonomy of terms is provided at <http://diglib.hab.de/rules/concordance/structMD.xml>.*

The following deliverables were defined:

D5.1 – Draft report on quality control and quality check auto proceed (M12)

- D5.2 – Report on images online, accessible to *Europeana*: n° 1 (M18)
- D5.3 – Final report on quality control and quality check auto proceed (M22)
- D5.4 – Report on images online, accessible to *Europeana*: n° 3 (M24)
- D5.5 – Final report on images online, accessible to *Europeana*: n° 4 (M30)

In the course of project work, however, it became clear that this definition of WP 5 was not compatible with the technical layout of *Europeana* and aspects originally to be dealt with in this work package were already covered by other work packages of which they form a natural part. It seemed preferable not to split up these processes, but to integrate them into the digitization or metadata content work packages.

## **2.2 Re-definition of the Name and Scope of WP 5 according to the Project Plan**

Early in 2011 it became clear that it was necessary to refocus the content of WP 5 in the light of the current technical developments, the organisational decisions and new experience gained in the course of project work. During the plenary meeting in Paris on January 24th & 25th 2011, the project partners therefore agreed to propose a redefinition of work package 5 and a transfer of responsibility from the Royal Library of Belgium (KBR) to the Bavarian State Library (BSB). This seemed necessary since WP5 should be treated by a library with an established record both in mass digitization and specialized digitization projects; besides this work package is closely related to WP4, for which the BSB is responsible. The partner institution that seemed most suited to take over the coordination of WP5 was therefore the BSB in Munich.

A modification of the contents of the work package became inevitable when it became clear that only metadata and the link to the images will be harvested by *Europeana*, not the images themselves. Most aspects of quality management are therefore not influenced by *Europeana*, but must be carried out and ensured in their local contexts by the project partners. It is them who are responsible for the creation of digital content and its presentation, which is mostly done in the context of the digitization process as described in WP 4.

The decision to leave all digital images in the local presentation systems has also the effect that structural metadata linked to the images are not integrated into *Europeana*, but must be created, kept, and presented by the local content providers or holding libraries themselves. Existing technical solutions and standards of encoding structural metadata in the holding libraries can therefore be allowed to coexist as they are. For a common approach and to ensure a possible future transfer it seems more important to define a common list of mandatory elements of structural metadata and to propose a standardized XML encoding for them. This aspect is already treated in WP 2

Two more aspects of the original work package are covered in other work packages or by other partners:

- The integration of digital images in local databases is part of WP4.
- Checking the links between the images and the relevant metadata is the task of the aggregator.

## **2.3 New Content of Work package 5 according to the Project Requirements**

WP 5 (Quality management) within the scope of the *Europeana Regia* Project will discuss the application of the methods of quality management to the results and to the specific processes involved in the digitization of medieval manuscripts.

In the past decades, most libraries controlled only the quality of the digital images after production, i.e. the results of the digitization process. According to a survey report issued in spring 2011, in more than 95% of all of European digitization projects of library material, such quality checks are carried out.<sup>1</sup> Given the rapidly increasing scope of digitization projects, however, more comprehensive approaches have to be taken into account and to be applied. A widespread quality approach in the manufacturing industries is defined in the international standards ISO 9000, 9001, 9004 and 19011. According to these standards, quality management is defined as all coordinated activities to direct and control an

---

<sup>1</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/>>.

organization with regard to quality.<sup>2</sup> It therefore comprises not only the quality control after production, but includes all processes and methods involved in the production, here the digitization of images, starting with the planning stage. Quality management uses quality assurance and control of processes as well as products to achieve more consistent quality.

Documentation is an important factor; therefore quality manuals, quality plans, specifications, guidelines, documented procedures, work instructions and records of results achieved will be collected in this work package and closely scrutinized. In the partner libraries of *Europeana Regia*, the different results of the digitization process will therefore be reviewed and areas for continual improvement should be identified. Not only the experience of the ER partner libraries has been relied upon, but also a detailed questionnaire on all aspects of the digitization of medieval manuscripts (organisation, workflow, infrastructure, image parameters, preservation, quality control) was sent in March 2011 to more than 20 international organizations with an established record in this field of activity.

The *Europeana Regia* Project is especially suitable for the development of quality standards due to the very special material which will be digitized. Because of the extreme old age and value of the Carolingian manuscripts (some 1200 years) and some illuminated manuscripts of the 14<sup>th</sup> and 15<sup>th</sup> century as well as the high expectations of the main target group of users (see WP6), it is certainly one of the digitization projects where the highest quality standards should be employed. Of course, these will not be applicable in every manuscript digitization project, where time and resources available must be weighed against the gain in quality.

This comprehensive evaluation could then serve as a model of excellence for future digitization projects in this area.

The content of WP 5 is therefore dedicated to recommending quality documentation standards for digitization projects

---

<sup>2</sup> ISO 9000:2000 (E), p. 22.



### 3. External Influences on General Quality Aspects in Digitization

The interests of a number of partner organisations must be respected in the Project *Europeana Regia*. These include

- The European Commission as funding organisation,
- Europeana (the database to receive the metadata),
- Europeana Regia (the five project partners and their individual framework),
- Content providers (holding libraries, library networks),
- External service providers (e.g. in the creation of digital images),
- The European Library (aggregator),
- Users.

#### 3.1 Customer / User requirements

The general requirements of users of Europeana Regia have been examined in D6.3 (Report on Users' Requirements). The key findings of this study were:

##### Specific requirements of each category of public

*"The study confirmed the high level of interest **researchers** have for this project, but also their high expectations. Compared to what is available on other sites, their demands do not so much concern the need for innovative functions as they do the need for effective tools, speed of access and exhaustive information.*

*Scientific information: up-to-date summaries including the status of information pertaining to each manuscript; links to the catalogues in which the manuscript are listed and all relevant external scientific resources (publications, online articles, etc.), which users can view without having to leave the site; a very detailed description of the images: precise thesaurus, descriptions of the elements represented as well as the context of the illustration, explanation of the text/image relationship, etc.*

*Search functions: expert search by entry field and index covering every aspect of the manuscript: call number, author (with lists of authorities managing the different written forms), all or part of the title, iconographic thesaurus, incipit of the second and penultimate pages, codicological elements; navigation by corpus, genres, commissioners, authors, translators, production centres (with geolocalisation), periods, etc.*

*Presentation of results: list of results giving direct access to the document's complete record, with an illustration of the first page (cf. <http://www.e-codices.unifr.ch/>; option to set the number of results per page, etc.).*

*Viewing/consulting documents: powerful zoom function, preferably controlled by the user (and not using an enforced scale); magnifying glass; brightness/contrast settings and colour filters; 360° rotation; option to simultaneously view two pages of one or two manuscripts and to create a personalised mosaic; possibly 3D viewing. There was a difference of opinion with regard to the usefulness of e-books."<sup>3</sup>*

As specified in the project plan for *Europeana Regia*, the following target users have been identified for the project: learners, academics, and all those interested in medieval European culture. The project is considered to be especially relevant for:

- *art historians,*
- *historians,*

---

<sup>3</sup> Europeana Regia d6.3, p. 3.

- the general public interested in European history in relation to other cultures, art, royalty, or fine books, in local history or genealogy, in education, literature, and the survival of books over the centuries,
- university and secondary-school students of history, English, French, Italian, German, Latin, Theology, and the arts,
- creative industries,
- publishers of Art books,
- and artists, designers including electronic and web designers.

*The digitized manuscripts realized through Europeana Regia will also be included in virtual on-line exhibitions, turning-the-pages facsimiles, in order to attract the widest possible audience, and not to be reserved to specialists. Users' representative groups will be constituted in every participant library, in order to enhance the best possible adequation between the chosen technical solution, and users' needs."*

The expectations of the target group(s) must have a formative influence on all definitions of image quality aspects. The highest quality standards for digital images are formulated by scholars specialized in aspects of mediaeval culture (e.g. History, History of the Book, Palaeography, Philology, History of Art). The requirements for academic use of mediaeval 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>4</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 palaeography, a pretty high resolution image, close to 4491 x 3480 in size, is presented."*

For an evaluation of the resolution values recommended for medieval manuscripts, see also D4.1 "State of the art in image processing", ch. 5.5,<sup>5</sup> and users' requirements as analyzed in D6.3, ch. 4.

According to the information on user expectations, the main aspects of quality can be summarized in the following way:

- high optical quality of images,
- completeness of the data,
- faithfulness to the original,
- quality and reliability of metadata, in this context with an emphasis on working links.

Other user expectations are related to the integration and presentation of the images on the World Wide Web:

- search engine optimization,
- retrievability through metadata (see WP 2),<sup>6</sup>

<sup>4</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>5</sup> Cf. D4.1: State of the art in image processing: this report is concerned with optical quality parameters (resolution, colour depth), the relation of digital images to the original (dimensions, completeness, sequence), and aspects peculiar to manuscript digitization.

<sup>6</sup> Cf. D2.1: State of the art in metadata, especially in the field of Mediaeval and Renaissance manuscripts: this paper deals with the format of metadata and file naming conventions. D2.2: Catalogue requirement for metadata: determination of the fields to supply, in order to fulfil the requirements of "minimus level metadata" of not already existing, mapping of the existing data to the OAI format – Dublin Core; includes rules about the translation in English.

- unproblematic interaction with internet software,
- open access,
- reusability.

Especially adequate display options were discovered to be important for users of Europeana Regia:

**"Tools to view pages up close and more:**

Enlargement (considered "essential" with the highest rating by the BHUV study)

*The first level of enlargement of a page should keep a page within the limits of the screen in order to view the page in its entirety.*

*It should be possible to zoom in on the whole page and move it around the screen. The zoom function should be progressive, preferably "operated by the user" instead of enforced scales.*

*Magnifying glass to focus on a section of the user's choice.*

Contrast and brightness (considered "desirable" but not essential by the BHUV study).

*To be able to control the contrast and brightness of an image (include the tools, now standard, for retouching photos).*

*Include colour filters (Wood's lamp) to show up things that are poorly visible or even invisible to the naked eye.*

Function to rotate the page 360° to make it easier to read glosses and annotations (considered "desirable" but not essential by the BHUV study).

Scale and measurement (only mentioned in the BnF study)

*The scale of the document being viewed, both vertical and horizontal, should be permanently on display however much the image is enlarged.*

*An electronic ruler should be integrated in order to measure the distance between two elements on a given page.*

*Similarly, while it was not suggested, a tool for measuring areas within a page would likely be appreciated.*

Functions for correcting undulated or domed pages, like certain scanners. (only mentioned in the BnF study)

Tools for easily comparing pages and viewing from different angles: (only mentioned in the BnF study)

- ☞ *Putting non-consecutive pages in a same or different works side by side*
- ☞ *Transcriptions and translations of text should be included for a selection of works, alongside the original text*
- ☞ *Possibility to open two works on the same screen*
- ☞ *Mosaics of pages selected by the user."*<sup>7</sup>

### **3.2 Suppliers / Content Providers**

On the side of the content providers, i.e. the holding libraries, perhaps the most important decision to take is whether the institution feels competent to carry out the digitization process itself or if will rely on external service providers/subcontractors. Both options have been chosen within the project *Europeana Regia*; the first by the BSB, HAB, and KBR, the second by the BnF, which is also responsible for several French municipal libraries. In the case of the BHUV, digitization is carried out jointly by the library and an external service provider.

A quality control program is needed both for in-house projects and for projects where all arrangements or parts thereof are outsourced. An important difference is that in a partly or totally outsourced project the quality requirements have to be formulated before a contract is signed, due to their legally binding nature. When working with vendors, the institution must be able to state its requirements in clear, quantifiable and verifiable instructions and it must be able to recognize whether the product returned by the vendor is what was requested.<sup>8</sup> In-house projects can build up their quality control programmes step by step as a part of their project activities.

Although quality control is a crucial factor to ensure the best results, there is no standard way to guarantee a certain image quality at capture. Different source documents require different scanning

<sup>7</sup> D6.3, p. 17.

<sup>8</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).

processes as well as accompanying preservation measures, and these factors have to be considered when developing a quality control programme.<sup>9</sup>

---

<sup>9</sup> IFLA: "Guidelines for digitisation projects for collections and holdings in the public domain, particularly those held by libraries and archives"; March 2002, p. 21  
<<http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf>>.

## 4. Conceiving and leading a quality strategy on digitisation projects of manuscripts

### 4.1 Planning stage (instruction phase)

The first step in the realization of a project is the planning stage, which is presented in chapter 6 of D4.1 for digitization projects of medieval manuscripts (Workflow and Technical Aspects of Manuscript Digitization).

It is at this stage that a quality management strategy can be developed, which defines its organisation, methods and tools. To give an example: the definition of rules for the scanning and photographing process as such does not belong to the quality strategy. To be more precise, it is the definition of the means by which it can be assured that these rules will be observed and put to practice that is a central part of the quality strategy.

#### 4.1.1 Definition of Quality Management Scope and Aims

According to the requirements of the target audience / user group, the desired end product and production goals are defined. This includes, for example, acceptable levels of digital image quality based on both the attributes of the source documents and the capability of the digital imaging system used. Reference points should be defined, i.e. a decision is taken what the output of the digitization process should be judged against.

Dependent on these parameters and on the scope of the digitization project, a selection must be made as to what the quality control programme should include. Different levels of quality control are conceivable. For high-ranking projects (such as *Europeana Regia* and most projects concerning medieval manuscripts), every single page is submitted to an intellectual quality control. For less prestigious projects, especially the mass digitization of printed books, sampling at regular intervals may be sufficient. In the latter case, it is essential to define a sample procedure which is strictly followed; otherwise, it is not possible to obtain reliable results.

Another question is if quality control should be extended to *all* digital files produced (including derivative formats such as access files or thumbnail files) or only to the digital master files.

Quality control should be documented precisely to enable members of staff to answer occurring questions (for example on missing or inverted pages) without recourse to the original.

Methods of evaluation must be decided upon and prepared, e.g. for manual/intellectual control:

- on-screen evaluation,
- print-out evaluation,
- the use of targets for colour and grey scale,
- the use of resolution targets.

At this planning stage, a decision will be made if the originals should be digitized in-house or by an external service provider.

As an example for the first approach, the experiences gained by the Munich Digitization Centre (MDZ) of the BSB will be summarized. The MDZ has expertise in the extensive scanning of diverse collections of objects spanning over fourteen centuries. Digitization of valuable objects is always carried out in close collaboration with the Institute of Book and Manuscript Conservation (IBR) as an important partner.<sup>10</sup>

As an example for the second approach, the BnF will be chosen, which has selected the service provider Azentis for the digitization of its manuscripts.<sup>11</sup>

---

<sup>10</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>11</sup> See <<http://www.azentis.com/>>.

#### 4.1.2 Quality Staff

Questions of staff are addressed at this stage as well. Project management is defined; quality management should take a central position in the project organigram. Besides the project manager, a person responsible for quality management can be chosen, who is also responsible for audits.

It is essential to note that quality control is both cost- and labour-intensive: about 50 % of the time required for the scanning or photographing of the originals needs to be budgeted additionally for quality control. If quality-control is not budgeted during the early phase of a project, the consequences can therefore be serious.

Quality control should be carried out by special staff independent of the scanning process respecting in a "four eyes principle".<sup>12</sup> By this separation, the risk of overlooking errors can be minimized. For quality control, specially trained staff should be employed. Since quality control is an activity where attention and concentration plays an important role, the duration of shifts should be strictly limited to a maximum of four hours per day.

It is necessary to give clear instructions to keep the effort calculable and the reliability of results predictable. Written guidelines defining criteria for the rejection of images should therefore be developed.

Approximately 8% of men and 0.5% of women have inherited colour deficiencies. Staff employed for quality control should therefore be tested for colour blindness.

#### 4.1.3 Planning of the quality tools and intervention measures

Following the definition of the scope and aims of quality management, quality tools and intervention measures must be defined.

- A project/steering committee must be established, uniting the different project partners and team members, which is to meet regularly.
- definition of production indicators,
- provision of hardware for quality control,
- provision of software tools for the production workflow and quality control (e.g. Imatest for colorimetric tests),<sup>13</sup>
- ensuring appropriate environmental conditions for quality control (see below, ch. 4.3.2.3),
- documentation on the different processes of the digitization chain,
- training sessions for staff,
- visits to the site of the digitization chain or audits with the service provider.

➤ **To anticipate these elements is an important step towards reducing the risk of disturbances and errors during production.**

#### 4.2 Pre-production, Test phase

Before the digitization process can take place, a number of quality documents have to be developed. Once the decision has been made on who will be responsible for carrying out the digitization project (internal vs. external service provider), it is essential to test the digitization chain before the start of production.

It is at this stage that all quality elements which were defined in the instruction phase will be applied and tested by both the holding library and the external service provider. If the latter is been chosen, a quality assurance plan is developed with this partner.

---

<sup>12</sup> See Borowski, Marion *et al.* (Fraunhofer Institut Intelligente Analyse- und Informationssysteme): "Bestandsaufnahme zur Digitalisierung von Kulturgut und Handlungsfelder", p. 78, available at <[http://www.iais.fraunhofer.de/bkm\\_studie.html](http://www.iais.fraunhofer.de/bkm_studie.html)>.

<sup>13</sup> <<http://www.imatest.com/products/software>>

#### 4.2.1 Testing Scanning Instructions

Scanning instructions are a list of the standard procedures to be applied to all objects to be digitized or for a special subset of collections (such as medieval manuscripts). An example is the scanning instructions for manuscript projects at the BnF:<sup>14</sup>

- *glass plates or permeable plates are used for keeping the objects to be digitized flat,*
- *the original ordering and orientation of pages is preserved,*
- *no processing of images at a post-production stage,*
- *one scan per page (including front and back covers, empty pages, pages not foliated),*
- *images extending to two opposing pages will be photographed or scanned as one image,*
- *digitization of the entire page, no cropping of margins or edges,*
- *no digitization of details except when otherwise stated on the accompanying document,*
- *all images are taken at the same resolution,*
- *use of underlying paper covers if the page is damaged by cuts, holes, etc.*

The BSB is preparing a similar document though digitization is not performed by external service providers. The practice of the BSB differs in some aspects from the BnF, for example glass plates are never used on medieval manuscripts, and images extending to two opposing pages are not scanned as one image.

A selection of original documents is made which is typical of the specific digitization problems expected in this project. The application of the scanning instructions is then tested on the chosen digitization hardware. According to the results of this test, standard pictures can be defined and examples of acceptable images can be given, as well as examples of images which would be rejected. These can be summarized in a document which gives clear indications as to when an image is to be classified as nonconforming. In the case of a vendor, such information must be included in the specification. The document should be complete, detailed, and illustrated by images to make sure that it will be understood by an external service provider. An example from the BnF is the following extract from the "Guide de contrôle / numérisation des documents reliés: Marché de numérisation des documents opaques des départements spécialisés (Azentis)", which lists the criteria for the rejection of images in manuscript digitization projects:

<b>Cas n°9 : l'onglet tronqué des informations. : A rejeter</b>
---

La lecture de l'information doit primer sur la vitre. L'onglet aurait du être relevé pour que la prise de vue aille chercher l'information. Un commentaire product aurait du être saisi si besoin.
--

---

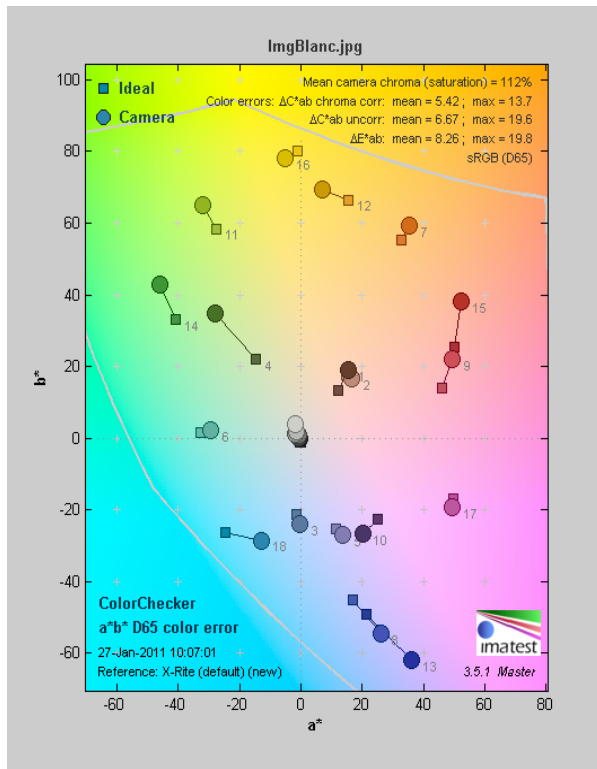
<sup>14</sup> BnF, "Demarche qualité Europeana Regia", Définition de l'organisation des prises de vues (summarized).



#### 4.2.2 Calibrating hardware (for technical details, see D4.1)

Both the display screens and the digitization hardware are calibrated during this stage. Tests are made for obtaining the colorimetric values conforming to the norms defined (e.g. the "Lab" values: L for light, a for the red-green axis and b for the yellow-blue axis)

A special software application can calculate the distances between the Lab values obtained by the shooting of colour targets and the standard Lab values. The image below illustrates an example obtained by imatest software:





At this stage, is also necessary to compare intellectually the digital images to the original manuscript. This comparison should be carried out under optimal viewing conditions (see below, ch. 4.3.2.3). The best results obtained will serve as a reference point during production.

#### 4.2.3 Testing the rendering of gold

In the case of manuscripts illuminated with gold, special digitization hardware can be used in order to enhance the brilliance of the respective pages. It should be noted, however, that these procedures can have a negative effect on the reproduction of other colours. Besides, the compressed file formats can have an effect on the brilliance of gold hues as well: JPEG is less appropriate than TIF.

#### 4.2.4 Testing the integration of metadata to the numerical documents

Different types of metadata can be distinguished (see D4.1, p. 38), e.g. TIFF tags and other metadata concerning the digital image.

This test is aimed at verifying whether the software systems of the libraries are compatible with the structure and organization of the data (automated control); then a visual control will be executed to check the correctness of the foliation and bibliographic data (manual control).

#### 4.2.5 Developing the Quality Assurance Plan

A quality assurance plan is a collection of documents which comprises the different guidelines (e.g. guideline on foliation, scanning instructions), i.e. the manuals. It defines the different processes in the digitization chain and specifies the procedures which must be put to practice. If an external service provider is employed, this plan is legally binding, developed collaboratively and continually evolving.

As an example, the "Plan Assurance Qualité" (quality assurance plan) developed by the BnF is quoted below. This plan lists the separate processes carried out by the service provider in the course of digitization; each process corresponds to a specific document.

SERVICE PROVIDER PROCESSES TO AUDIT	PROCEDURES AND/OR ACTIONS ASSOCIATED WITH THE REALISATION OF THE DU PROCESS		ORDER OF PRIORITY OF ASSOCIATED PROCEDURES			
			1	2	3	4
<b>P2</b> PREPARATION AND RECEPTION ON SITE OF DOCUMENTS	P2-2	Reception of original documents				
	P2-2	Stamping of BTA				
	P2-2	Coherence check BTA-Doc				
	P2-2	Storage of originals				
	P2-3	Treatment of incoherences				
	P2-3	Treatment of rejected images				
	P2-4	Archive storage of documents				
	P2-4	Handling of documents				
	P3-1	Preparation et Orientation of documents				
	P3-1	Environmental conditions				

<b>P3</b> <b>PRODUCTION AND POST-PRODUCTION</b>		for document handling				
	<i>P3-1</i>	Digitization of documents				
	<i>P3-1</i>	Treatment of images				
	<i>P3-1</i>	Creation of metadata				
	<i>P3-1</i>	Packaging of documents to be returned				
<b>P4</b> <b>CONTROL OF DATA PRODUCED, ORGANISATION OF DELIVERY AND DELIVERY TO THE BnF</b>	<i>P4</i>	Quality control				
	<i>P4</i>	Organisation and delivery of images				
	<i>P4</i>	Exchange procedures with the BnF				
	<i>P4</i>	Delivery of digital documents				
<b>P5</b> <b>TREATMENT OF NONCONFORMITY</b>	<i>P5-2</i>	Organisation of rescans of the images rejected by the BnF				
<b>P7</b> <b>MAINTENANCE AND CALIBRATION OF THE SCANNING STATIONS</b>	<i>P7</i>	Calibration of scan stations				
	<i>P7</i>	Maintenance of hardware				
<b>P8</b> <b>PLANNING AND FOLLOW-UP TOOLS -QUALITY MANAGEMENT</b>	<i>P8</i>	Organisation of planning and software tools				
	<i>P8</i>	Documentation management				
	<i>P8</i>	Follow-up of the audit				
	<i>P8</i>	Management of anomalies and nonconformant objects				

➤ It is during the test phase that the quality of the digital images produced in the final workflow is determined.<sup>15</sup>

#### 4.3 During Production (production phase)

Especially when working with vendors, it is important to carry out quality control routines as early as possible in a digitization project. Errors should be reported early, since the vendor may thus be prevented from continuing to replicate errors, be alerted to problem procedures or ill-trained staff. In addition, there is usually a cut-off date after which the vendor will no longer accept the correction of errors for free.<sup>16</sup>

Several tools and methods can be applied during production to evaluate the digitization chain as a whole:

- preventive action,
- checks on the digital images (automated checks, visual checks, and colorimetrics),

<sup>15</sup> For recommended procedures during this test phase, see Gertz, "Vendor Relations".

<sup>16</sup> See Gertz, "Vendor Relations".

- a synthesis: the audit.
- Corrective action: these are aimed at readjusting the quality of the digital documents and the general functioning of the digitization chain.

### **4.3.1 Preventive action**

According to the regulations laid down in the quality documents, the staff involved in the digitization process must be trained. In the case of external service providers, quality documents will be included in the contracts and specifications for this partner.

#### **4.3.1.1. Training of staff**

Two aspects have to be considered:

- Training of the scanning staff on the handling of manuscripts. Its aim is not only to make the staff aware of the integrity of the physical object but also to the possible effects of conservation measures on the handling and the digitization process. To give an example: binding strips will stiffen the manuscript and it will be more difficult to capture characters on the inner margin.
- Training of the quality staff, especially on acceptable image qualities.

#### **4.3.1.2. Survey of the physical condition of the objects by a conservator**

The physical condition of the objects chosen for the project is examined by a conservator prior to digitization. The results are entered in a checklist with detailed instructions for the scan operator how to handle each manuscript (see D4.1, ch. 6.2).

### **Scan-lists**

A second examination of the original manuscript, together with all available descriptions and catalogue entries, may lead to the production of a so-called scan-list. This list is a concordance between the following pieces of information:

- future image-scan-numbers,
- the corresponding page numbers with special attention to faults in the numbering systems of the manuscript (on which see D4.1, ch. 5.3.1),
- structural metadata (on which see D4.1, ch. 7.1.3),
- basic bibliographical metadata.

This document can be of central importance for the scanning process itself, since it makes the scan-operator aware of difficult areas in the original and gives instructions on how to handle them. The previous analysis of the original can be very time-consuming and requires a lot of knowledge about both manuscript culture and academic habits (e.g. numbering conventions); it therefore should be carried out by a member of staff with some expertise in this field.

If a document lacks page numbers, the first word on a page can be noted down as a catchword or marker. This approach is typically used for documents written in uncommon scripts (e.g. Chinese). With the aid of these concordances, several errors typical of the scan process can be avoided altogether or their frequency at least reduced:

- pages omitted,
- pages scanned twice,
- errors in the naming of files,
- delays due to questions on how to handle difficult sections of manuscripts.

Scan-lists are also very useful for post-processing quality control and can be used for the production of structural metadata for the internet presentation of the digital object. See below, for example, an Excel scan list produced for the ER project at the BSB:

shelfmark:	Clm 6393			
ID Num:				<i>Bibliographic data</i>
Title:				
Subtitle:				
Editor:				
Author:				
Pubplace:	s.l.			
Publisher:				
Pubyear:	s.a.			
<i>page - category</i>	<i>Image number</i>	<i>page number</i>	<i>toc-point of acces</i>	<i>remarks for Scan operators</i>
	-- STRUCTURE --		= <i>Beginning of toc</i>	
front				
	00001			
binding		cover outside	Front cover: late gothic leather binding, clasp and chain missing<lb>table of contents and shelf mark on the spine	
binding<lb>h	00002	cover inner side	Ex libris: Freising Cathedral library	
misc	00003	1r	pen trials: beginnings of prayers	
body				
holdings<lb>decLetter<lb>>section	00004		ownership mark: "Lib[er] sancte Marie et sancti / Corbiniani Frising[e]" (1v to 2r ) <lb>Initial<lb>decorative script as display script. (1v to 2r) <lb>beginning of text: Rufinus: Historia MonachorumRufinus: Historia Monachorum	
	00005	2r		
decLetter	00006	2v	decorated initial	
condition	00007	3r		page damaged: hole in upper right corner
	00008	3v		
	00009	4r		
	00010	4v		
decLetter	00011	5r	decorated initial<lb>Capitals as display script	
	00012	5v		
	00013	6r		

#### 4.3.1.3. Intermediate Tests

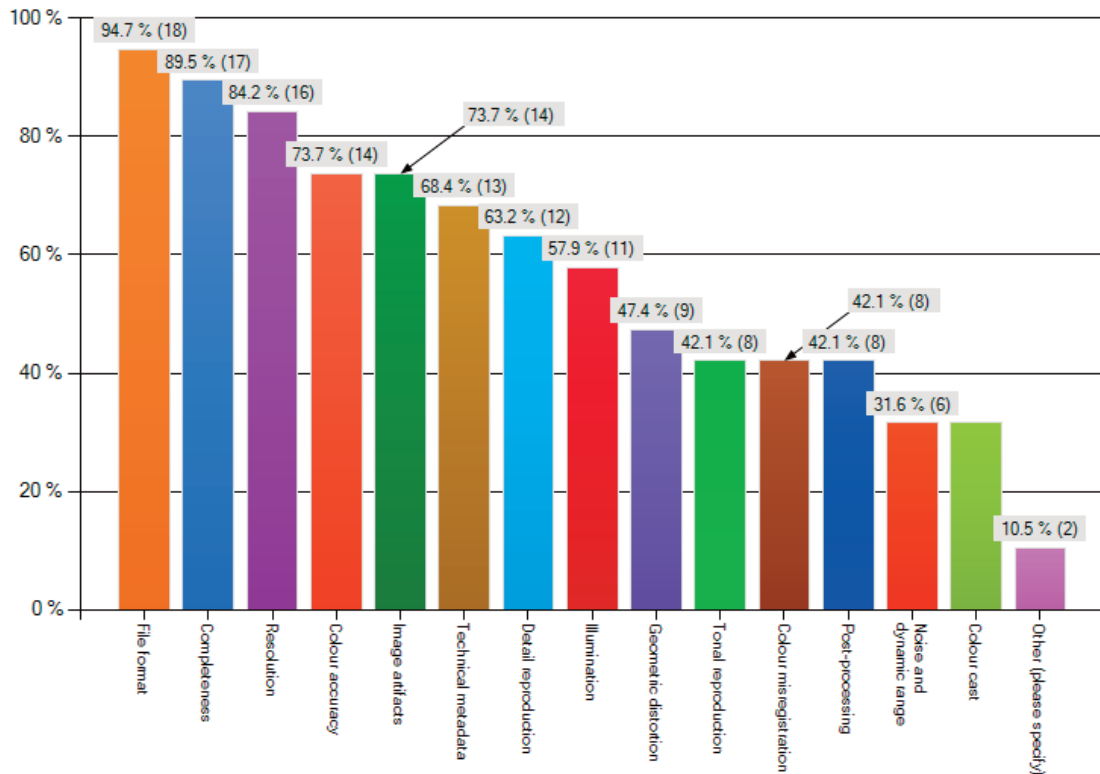
If part of the project consists of documents which are very difficult to digitize, intermediary tests can be planned to avoid dissatisfactory results.

#### 4.3.2. Quality parameters and checking operations on digital documents

According to the 2011 "Survey Report", the following parameters are checked in European digitization projects carried out by the National libraries:<sup>17</sup>

<sup>17</sup> Sturm, "Survey Report", p. 9.

### Which criteria do you check? (multiple answers possible)



Several of these parameters can be summarized under the heading optical quality (resolution, colour accuracy, image artifacts, rendering of detail, tonal reproduction, colour misregistration, noise and dynamic range, colour cast).

Among these parameters, the following aspects are more or less routinely checked in quality management workflows, as the answers of the *Europeana Regia* questionnaire have shown:

- completeness,
- optical quality,
- connection to metadata.

### Optical quality

#### Sharpness

- The sharpness of an image depends on the resolution chosen for the scanning process, but also on the correct focussing of the imaging device. Image sharpness can be evaluated by using so-called resolution targets (see D4.1, ch. 6.2.1).
- 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 and below, ch. 4.3.2.2).<sup>18</sup>

#### Colour and Tone

<sup>18</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.

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. 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.

For an objective evaluation of colour values, greyscale targets are used to assess adherence to the ideal RGB values. Each patch of a greyscale or colour target has a numerical digital value assigned to it. For example, for the perfect white patch, each RGB value should be close to 255. Therefore, ideal RGB values can be defined with reference to the colour and greyscale targets used. This is done, for example, in the *NARA Guidelines for Digitizing Archival Materials for Electronic Access*, where also the tolerance level for the numeric difference among the R, G, and B channels is specified. Another example from the Library of Congress:

*For color images with typical scene brightness that include white or black elements, similar values shall be provided for RGB (red, green, blue) renderings of white or black, i.e., a white area shall have values of r=243-250, b=243-250, and g=243-250 and a black area shall have values of r=5-12, b=5-12, and g=5-12.*<sup>20</sup>

However, it is recommended not to rely on the use of colour and greyscale targets alone, since an image may appear flat despite a perfect reproduction of the greyscale.<sup>21</sup> In addition, control bars fade over time; therefore they should be replaced periodically.

### Noise

Noise is defined as point-to-point light-intensity fluctuations in an image that are not part of the original object, i.e. in electronic imaging systems they are usually caused by the sensor and associated electronics. It can be measured by calculating the standard deviation of pixel count values in an image of a uniformly lit gray patch. The greater the standard deviation, the poorer the image quality.<sup>22</sup> Noise may also be caused by the target used, nonuniform lighting and other artifacts; these different noise sources should, however, be separated.

### Artifacts

Artifacts are defined as nonrandom or correlated light intensity fluctuations in an image that are objectionable because of their magnitude, e.g. distortions, streaks, dust, scratches or oversharpening.<sup>23</sup> Software tools for the measurement of artifacts are offered, for example, by imageXpert™ (www.imagexpert.com)

## Completeness and Sequence

Other aspects which need to be checked routinely are the completeness of the data, e.g. the inclusion of *all* elements of the manuscript (see D4.1, ch. 6.2.3), and their correct sequence (see D.4.1, ch. 6.2.4). This can be done most efficiently by using scan-lists (see above, ch. 4.3.1.2).

## Relation to metadata

The last aspect to be included in quality control is the relation of the digital object to its metadata, i.e. to bibliographical metadata, technical metadata, and structural metadata (see also D4.1, ch. 8).

---

<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> The Library of Congress: *Requests Proposals For Digital Images of Pictorial Materials*. National Digital Library Program, see <<http://memory.loc.gov/ammem/prpsal9/rfp9.pdf>>, ch. C.4.4.1.

<sup>21</sup> Rieger, "Establishing a Quality Control Program", p. 81.

<sup>22</sup> Williams, "An Overview", p. 73.

<sup>23</sup> Williams, "An Overview", p. 73.

#### 4.3.2.1 Methods

Three approaches can be distinguished in the identification of nonconformity, i.e. the non-fulfilment of a requirement, which typically takes place after production. These can either be carried out using automated procedures or done manually/intellectually, i.e. with book in hand or with the aid of documents prepared in advance. A relatively new approach is to include user feedback after the release of the data on the internet, which is now possible due to Web 2.0 technology. This quality control through user feedback cannot, however, replace control directly after production, since the cost and effort of rework would be high at this late stage after publication.

Some quality aspects, such as completeness, can be checked using automated procedures. For example, the expected number of images (e.g. from a scan-list, see above p. 9) can be checked against the actual number of files. Other aspects can only be controlled in a manual solution or in a mixture of manual and technical checking. Of course, the most careful procedures of manual quality control can only be used on high-ranking projects, since they are cost- and labour-intensive. In the case of manuscript digitization it is recommended to always check the result of the digitization process, i.e. the digital master files, image by image manually and intellectually.

Among the project partners of *Europeana Regia*, the BnF, BSB, and HAB use technical solutions. This seems to be uncommon in other institutions experienced in the digitization of manuscripts, since only three of the libraries who returned the *Europeana Regia* questionnaire use technical procedures in quality control.

#### 4.3.2.2. Automated procedures

##### Automated structure checks

In the case of mass digitization projects, an automated structure check is often considered sufficient before the release of the data for the internet; only random sample checks with manual/intellectual control of the image files are then executed.<sup>24</sup> These automated procedures are especially important where a complete, intellectual control of images is not feasible due to cost considerations or in the case of external service providers. They should be carried out immediately after the scanning process. In the case of the BnF, these automated checks imply the structure, the format, and the validity of the paths of the data received from the service-provider.<sup>25</sup>

Automated structure checks can be achieved by an analysis of the TIFF Image File Directory (IFD), also known as TIFF header (see also D4.1, ch. 7.1.2). This contains technical information on the image files themselves, e.g. their length and width in pixels, compression, resolution, size in bytes, production information, e.g. references to scanner and software, and some pieces of administrative information, e.g. on the author and the owner of the image.<sup>26</sup> Similar automated checks can include a coherence check for metadata or a comparison of metadata to the total number of image files.

##### Other automated procedures

To avoid errors which are due to inattention (such as keying errors), repetitive actions and standard procedures should be fully automated. This is why a number of software tools for the digitization workflow have been developed according to the specific needs of the institutions; two examples from Germany are the Munich ZEND software tool or the Göttingen GOOBI software. Examples of tasks carried out by the BSB-digitization software tool ZEND (see D4.1) are:

- generation of an object-specific digitization order (end user or library staff) with plausibility check,
- automatic Uniform Resource Name (URN) generation with a Uniform Reference Locator (URL) to the central link-resolving server of the German National Library,
- Open Archives Initiative (OAI) data provider,
- import of all necessary production data, such as bibliographic metadata via Z.39.50 interface from the Online Public Access Catalog (OPAC),

---

<sup>24</sup> See

<[http://www.bnf.fr/fr/professionnels/num\\_logistique/s.num\\_logistique\\_qualite.html?first\\_Art=non](http://www.bnf.fr/fr/professionnels/num_logistique/s.num_logistique_qualite.html?first_Art=non)>.

<sup>25</sup> See BnF: "Demarche qualité Europeana Regia", p. 1.

<sup>26</sup> For an example of automated checks on information contained in the TIFF header, see Riley and Whitsel, "Practical quality control", pp. 41-2.

- support (import, export, generation) of all current XML document formats, especially Text Encoding Initiative (TEI) and Metadata Encoding and Transmission Standard (METS), including the so-called DFG-VIEWER,
- cooperative interfaces, e.g. for indexing of digital objects,
- possibility to separate between public (open) and internal access, e.g. a digital reading room,
- PDF on the fly output of entire e-books via the Web
- customizable Really Simple Syndication (RSS) feeds, e.g., the daily production of digitized items,
- transfer of files to long term digital preservation archives,
- integration and transfer to other databases and internet presence,
- establishment of links to library catalogue entries,
- deletion of scan data on local servers at the end of the process.

## Checksums

A checksum is a fixed-size datum computed from an block of digital data for the purpose of detecting accidental errors that may have been introduced during its transmission or storage. The integrity of the data can be checked at any later time by recomputing the checksum and comparing it with the stored one. If the checksums match, the data were almost certainly not altered (either intentionally or unintentionally). For the purpose of integrity checking the most commonly used checksum algorithms are:

- MD5, Message-Digest Algorithm,<sup>27</sup>
- SHA-1, Secure Hash Algorithm,<sup>28</sup>
- CRC 32, Cyclic Redundancy Check.

Checksum(s) for each master-file should be computed during the first step of the digitization workflow: immediately after the scanning and before transferring the file to another storage place. It is also possible to compute checksums for entire directories (containers). At critical points in the workflow (e.g. subsequent to file transfers, before long term archiving, during delivery) checks should be run to confirm successful transmission and storage, i.e. the integrity of the files and the completeness of the directories. Due to the vulnerability of each algorithm (different files can have the same checksum), it is reasonable to rely on more than one algorithm to improve the data security which can be achieved by checksums. Risk management methods can help to set time and place for the integrity checkings according to data security requirements.

The BSB has just begun to use checksums. Detailed information on the experience with checksums will be available within several months.

## Software Solutions

For some image quality factors, software tools are available:

- signal-to-noise measurement and artifact (e.g. dust, scratches, streaks) detection tools,<sup>29</sup>
- The Modulation Transfer Function, in which the spread of light in the imaging process (line spread function) is measured. This is a mathematical function which describes the different values of contrast of the original object and its digital image. This is a more reliable and objective way to evaluate how well details are preserved.
- For missing and incomplete images software tools are available.

### 4.3.2.3 Manual quality control

Because of the complexity and individual properties of each manuscript, it is essential to carry out a visual quality control in the digitization chain. These specific properties of each original cause difficulties and limits to digitization which quality staff must be in a position to analyse and evaluate.

<sup>27</sup> For the definition of MD5 see RFC 1321: <<http://tools.ietf.org/html/rfc1321>>

<sup>28</sup> For the definition of SHA-1 see RFC3174: <<http://tools.ietf.org/html/rfc3174>>

<sup>29</sup> IFLA "Guidelines", p. 25, <<http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf>>.



## Hardware and Software, Environmental Conditions

As in the process of image production, the use of hard- and software must be planned and environmental conditions and hardware must be carefully controlled for all processes of quality control. For an effective onscreen quality control, the following factors are important:

- adequate hardware configuration (sufficient RAM capacity, graphics board, 17", 24-bit colour monitor),
- image display software.
- Monitor set-up (colour calibration): Gamma values (light intensity) should be at 2.2, white point (colour temperature) to cool white (5000° K).<sup>30</sup> Some digitization operations use monitors calibrated to the sRGB standard of 6500°K. However, the International Organization for Standardization generally recommends 5000°K over 6500°K when comparing images on a screen to hard-copy images. This calibration can be done with hardware (colorimeters) and accompanying software. Colorimeters are attached flat to the display's surface, shielded from all ambient light. The calibration software sends a series of colour signals to the display and compares the values that were actually sent against the readings from the calibration device.
- Room lighting: should also be at 5000° K. For professional purposes (the comparison of originals to digital images), colour-viewing booths, which provide metered daylight and balanced light levels are offered, e.g. by Gretag-Macbeth (The Judge II).
- If the quality control is done onscreen, a 100% (1:1) magnification should be used.

Practical advice concerning the hardware and Viewing conditions is summarized by Rieger:<sup>31</sup>

- *Leave the computer on for at least half an hour to allow the monitor display to stabilize*
- *Set the room lighting and then adjust the brightness and contrast controls on your monitor*
- *Clean your screen once a week according to the manufacturer's instructions*
- *Avoid turning the monitor on and off frequently*
- *Keep the windows mostly shaded*
- *Keep only half of the overhead lights on, preferably not those directly over the monitor*
- *Have the monitor face a wall rather than windows, to eliminate reflections*

Very detailed recommendations as to viewing conditions are given in the Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group: *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files* (2009), pp. 4-7, which are based on the international standards ISO 3664 Viewing Conditions- For Graphic Technology and Photography (which provides specifications governing viewing images on reflective and transmissive media, as well as images displayed on a computer monitor without direct comparison to any form of the originals) and ISO 12646 Graphic Technology – Displays for Colour Proofing – Characteristics and Viewing Conditions (which provides specific requirements for monitors and their surrounds for direct comparison of images on a computer monitor with originals).

Some aspects of image quality can be checked without recourse to the original manuscript, e.g.

- errors in the relation to bibliographical metadata,
- incomplete images,
- resolution problems (text illegible),
- images inclined with regard to their axis,
- wrong orientation of the pages,
- too narrow frames for scanning,
- images scanned in the wrong sense,
- insufficient visibility of the image (e.g. images too dark or too bright),
- inclusion of elements foreign to the original,
- hard shadows,
- missing paper covers,
- distortions,

---

<sup>30</sup> See Rieger and Whitsel, "Practical quality control", p. 48, Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, CA, 2000), pp. 67-70, ISO 3664, and ISO 12646.

<sup>31</sup> Rieger, *Moving Theory into Practice*, p. 68-9.

- illegible parts of the page which could possibly be included in the image (e.g. annotations in the inner margins).

An important instrument in this process is the above-mentioned scan-list, which is produced during a preliminary examination of the manuscript before the scanning process proper. Problems which can be detected with the aid of the scan-lists are, for example:

- incomplete digitization of the original documents (missing image-scans),
- errors in the sequence of digital images,
- errors in the relation between image-scan and metadata (e.g. wrong file-numbers),
- errors in the structure of the metadata.

In other cases, it may be necessary to consult the original, e.g. in the case of optical image quality such as problems with colour and calibration (see D4.1, ch. 6.2.1) as well as the relation of the images to the original (see D4.1, ch. 6.2).

#### **4.3.2.4 Colorimetric control (see D4.1)**

Also during production, colorimetric checks can be executed. The colour targets included in the digital image are analysed in ch. 4.2 and 4.2.2 of this report. The results must be compared to the standards defined during the test phase.

#### **4.3.3 Audits with service providers**

If an external service provider is entrusted with the digitization of the manuscripts, an important means of discovering non-conformant procedures are regular audits. The aims of audits with service-providers are

- to verify that the regulations and agreements as defined in the project specifications and quality assurance plan are put to practice (see ch. 4.2.5),
- to measure the efficacy of the procedures and to propose improvements if necessary.

Before an audit is carried out, the following steps must be taken:

- documentation of non-fulfilment of requirements as laid down in the specifications,
- development of indicators of production progress,
- quality control of digital documents (criteria for rejecting images in the manual/intellectual quality control, results of automatic structure checks).

An audit is structured in three phases:

##### **4.3.3.1. Preparation of the audit (3 weeks)**

- The aim, type and scope of the audit is determined,
- necessary reference material and documents are assembled, internal documents of the service provider are requested if necessary,
- study of reference documents,
- samples taken at random are checked as to coherence,
- results obtained by automated and visual checks are evaluated,
- production of a list of open questions,
- organisation of the visit.

##### **4.3.3.2. Audit: visit to the site of production (max. 2 days)**

- observation of the scan operators during the scanning process,
- talks to the project managers, based on a questionnaire concerning the processes under survey

#### **4.3.3.3. After the visit of the site of production (2 weeks)**

- processing of information,
- redaction and validation of the audit report

The audit report should include information on the following points:

- critical aspects and difficulties,
- track of the progress made,
- non-conforming elements,
- remarks,
- transmission of the audit report to the service provider,
- action plan and corrective action.

During production, a technical supervisor should regularly, but in irregular intervals control the imaging conditions to avoid "bad series".<sup>32</sup>

#### **4.3.4 Correction and Rework**

These two steps in quality management are defined as elimination of a detected nonconformity and action on a nonconforming product.

The rejection of an image should always be based on criteria laid down in a quality document produced before the scanning process proper, e.g. the scanning instructions. These two steps should take place immediately after the scanning process, i.e. before the images are deleted from the harddisks / servers attached to the local scan computers.

If nonconforming images are detected in the course of post production quality control, it is important to document them in order to be able to use these errors for an analysis of quality problems and for feedback on the production process itself.

Feedback is then given to the person responsible for the corresponding production process.

In the case of faulty image-scans or photographs, this will ultimately be the scan operator or photographer, who is responsible for the correction:

- faulty images are discarded,
- a new image is taken,
- the image is inserted into the series of image-scans at the appropriate place.

In the case of problems of the relation of the digital object with its metadata, correction can either be carried out

- by the person executing the quality control directly in the database,
- by other project collaborators or database managers.

#### **4.3.5 Corrective Action**

Corrective action is defined as all measures which aim to eliminate the cause of a detected nonconformity. These include:

- alterations in the workflow (e.g. because the necessary effort is too high, or the rate of non-conformant images is too high),
- the documentation of alterations.

In order to guarantee that corrective action is put to practice, visits to the digitization chain can be made, e.g. to make sure that the original documents are handled in an adequate way and to evaluate the possibilities and difficulties of improvements.

---

<sup>32</sup> Fraunhofer Institut Intelligente Analyse- und Informationssysteme: "Bestandsaufnahme zur Digitalisierung von Kulturgut und Handlungsfelder", available at:<[http://www.iais.fraunhofer.de/uploads/media/BKM\\_End\\_01.pdf](http://www.iais.fraunhofer.de/uploads/media/BKM_End_01.pdf)>.

If images rejected by the quality staff cannot be improved on, that is, if a better result cannot be obtained, feedback should be given to the quality staff in order to put them in a position to evaluate correctly the digitization results.

#### **4.3.6 Release**

Only after quality control has been completed is there the release of the digitized object, which in turn triggers the following processes, which are partly automated:

- production of an archive copy with the corresponding archive report for long-term preservation, which needs to be controlled intellectually by the project dispatcher,
- deletion of scan files,
- transfer of the URN and Resolving URL to the appropriate entry in the Union catalogue.

**👉 During production, quality management proposes methods and tools to anticipate, identify and resolve errors or deviations.**

## 5. Current Practice of Quality management at the Europeana Regia Partner Libraries

### 5.1. Framework and organization

#### 5.1.1. Seat of quality management

##### **Bayerische Staatsbibliothek:**

Both digitization and quality management are done in-house.

##### **Bibliothèque nationale de France:**

Digitization is performed by a service provider, but the digitization workshop is installed in-house.

From the beginning to the end of the digitization line, all processes have been defined. Each process is specified in a written document, which describes several proceedings (for instance, "Preparation and reception of documents" [P2] or "Production and post-production" [P4]). Some of these documents are drafted by the service provider, some by the BnF. Guidelines and technical references, such as "foliation guidelines" or "guidelines for digitization and control of bound and special collections at the BnF", are also included. The audits take place in conformity with these documents.

Audits are held once a year. Because the workshop is in-house, visits are frequent. An audit is structured in three phases:

1. Audit preparation:

- the aim, type and scope of the audit are determined,
- necessary reference material and documents are assembled, internal documents of the service provider are requested if necessary,
- study of reference documents,
- study of the results of checking operations,
- production of an open questions list.

2. Audit: visit to the site of production:

- observation of the scan operators during the scanning process,
- interview with the project managers, based on a questionnaire concerning the processes

3. After the visit:

- processing of information,
- redaction and validation of the audit report,
- transmission of the audit report to the service provider.

##### **Bibliothèque royale de Belgique**

Digitization is done in-house. An external audit concerning quality management with the company "Diadeis" was held in June 2011.

##### **Biblioteca Històrica – Universitat de València**

The documents included in the project are digitized jointly by the library and an external service provider (*Digitalizaciones Codoñer S.L.*). The following documents have been prepared for the service provider:

- Technical specifications of digitization,
- Instructions concerning archive nomenclature,
- Specifications for the assignment of technical metadata.

In the course of the project, the following series of documents has been drawn up to guarantee quality standards during the process:

##### *Scan-list:*

Before carrying out the digitization of a manuscript, the service provider is provided with a table in which the characteristics of the document and its state of conservation are detailed. The data collected in this table include: date of digitization, manuscript signature, total number of images to be digitized, page-numbering of the document and existing flyleaves, structural metadata, a page selected as a representative image of the document, information regarding the presence of miniatures and capital

letters to be considered in its description, data related to the physical state of the manuscript, and page-numbering errors.

*Quality control calendar:*

Table of the quality control calendar for the digitized images, listing the quality control sessions carried out by members of staff of the Historical Documents Library.

This table also indicates the number of times the monitor used to visualize the images was calibrated during the quality control process.

*Report on Problems:*

A report on the problems detected during the quality control of the digitized images includes the digitized image reference (page or manuscript section) and the problem or discrepancy detected. In this report the digitization technician must indicate if the problem has been solved and which measures were adopted for doing so.

*Conformity Report:*

This report is filled out by image quality control staff. This includes all the basic aspects of the process up to the final validation of the result obtained.

*Access data base (GEMA):*

This database allows for the follow-up of all the steps carried out in the digitization process of a manuscript as well as the creation of status reports, digitization statistics and incident reports.

Periodically (every 2-3 months, and always when it is deemed necessary, depending on the topics under consideration), follow-up and quality control meetings with all those taking part in the project are held in the offices of the *Servei de Biblioteques i Documentació de la Universitat de València* (Library and Document Services of the University of Valencia). A representative from the image digitization service provider also attends these meetings.

In addition, the library staff in charge of quality control maintain a direct relationship with the digitization technicians.

## **Herzog August Bibliothek Wolfenbüttel**

Digitization and quality control are done in-house.

### **5.1.2. Quality management staff**

#### **Bayerische Staatsbibliothek:**

A first quality check is carried out during and immediately after production by the scan operators, e.g. by a careful comparison of the digitization parameters as detailed on the printed digitization order slip produced by the ZEND software with the original object.

There are seven FTE involved in quality control at the BSB for all current digitization projects. These members of staff are responsible for individual digitization projects respectively.

Usually the members of the quality control team do not work as scan operators. The duration of their shifts is limited to a maximum of four hours per day. This is done in order to maintain concentration levels high. This limitation of shifts means that, on average, only one manuscript per day is submitted to quality control in very prestigious projects such as *Europeana Regia*.

Members of quality control are trained on the job by expert members of the team. There is a special instruction session given by the Institute of Book and Manuscript Conservation (IBR) on the handling of manuscripts and on the use of special equipment, such as book cradles, foam wedges, and cotton gloves.

If possible, staff is chosen who are able to read the texts on the digital objects, and who preferably have the language skills required (i.e. a good command of Latin in the case of the *Europeana Regia* project).

#### **Bibliothèque nationale de France**

The quality manager and the project manager are responsible for quality management. Digitization is first checked by scan operators of the vendor, then by the BnF. Two controllers and the quality manager are concerned with quality management. Checking a digitized manuscript of 400 images

requires about one hour. After a demonstration phase, the checkers are trained on the job and they are always supervised in their work. Each checker proposes images to reject and the person in charge of quality (sometimes in cooperation with the project manager) decides if the digitized object in question should be rejected or not.

### **Bibliothèque royale de Belgique**

Since August 2011 one person has been responsible for quality management. Altogether, two people are involved in quality control, that is the scan operator and the person responsible for quality management. A first check is executed by the scan operator directly after digitization, whereas the final quality control is carried out by an external person (post-graduate, experienced in the handling of manuscripts). Work shifts will be limited to two hours per week. At the end of each process, an oral review and an email containing the result of the quality check is sent to the project manager.

### **Biblioteca Historica – Universitat de València**

Quality management is carried out by staff from the Historical Documents Library and directed by the library director and the technical coordinator of the project.

The members of staff in charge of quality management are independent from those digitizing the images. The quality control staff consists of four librarians. The quality control sessions is carried out by each member of the team last on average 2-3 hours per session with a maximum duration of four hours.

Prior to the project start, material created by institutions with recognized expertise in quality management (guidelines and best practices manuals) was made available to the employees assigned the task of quality control.

After carrying out various sessions to exchange opinions and reach a consensus on the work procedures to follow with regard to quality control, written quality management guidelines were elaborated. These include evaluation parameters and process normalization methods.

Once a manuscript has undergone digitization and quality control, a file is filled out with the results, the final conformity, if applicable, and any relevant observations.

Should the quality control staff detect repeated problems or problems that have not been visibly improved upon, the staff responsible for digitization is asked to find solutions that will lead to better results.

### **Herzog August Bibliothek Wolfenbüttel**

Quality control is carried out by members of staff of the library's photography department (professionals). Digitization and quality control are performed by the same persons.

## **5.1.3. Equipment**

### **Bayerische Staatsbibliothek:**

#### *Hardware and software*

Quality control proceeds onscreen in the case of the ER project. It is done on colour calibrated monitors. At present, colour management is carried out by an external service provider,<sup>33</sup> who is also responsible for monitoring the colour profiles of the scanning hardware. The external consultant also provides targets for checking the accuracy of colour rendition, tonal values, and resolution, depth of field and scale, which are scanned together with each object.

Colour profiles of the scanning equipment are controlled monthly, using x-rite's ProfileMaker 5 software.<sup>34</sup> The calibration software compares the scan of the image to the colour information in the reference file to create an ICC profile specific to each scanner. Colour profiles of scanning equipment are routinely transmitted together with the digital images.

For the colour calibration of monitors, IT8.7/2 targets are employed together with the spectral photometer X-Rite Gretag Macbeth Eye One. The calibration is checked monthly.

- If there are problems with tonal values, histograms are used.

---

<sup>33</sup> <<http://www.digitalfoto-trainer.de/>>

<sup>34</sup> <<http://www.xrite.com>>

- The modular transfer function is used in problematic cases.
- Software for the detection and elimination of artifacts, however, is not employed in order to guarantee maximum faithfulness to the original.
- Check sum algorithms have been programmed and are being tested in other digitization projects.
- For Europeana Regia, the BSB's own digitization workflow tool ZEND is used; in other projects, the software Treventus ScanFlow™ is being tested.<sup>35</sup>

*Control of the environmental conditions (see D5.1, ch. 3.1.3)*

- Quality control of Europeana Regia manuscripts is carried out in a day-light protected room of the Scan Centre.
- A lux-meter is employed.
- Humidity is constantly surveyed and two humidifiers are employed.
- Digital images are viewed in 1:1 size in relation to the originals.

## **Bibliothèque nationale de France**

### *Hardware and software*

Quality control proceeds onscreen in the case of the ER project. It is done on colour calibrated monitors (Philips 240S, 24 inch Wide LCD Monitor). Colour targets ("GretagMacbeth ColorChecker"<sup>36</sup>) are used; greyscale or resolution targets are not used. The calibration software "Eye-One match 3" is used, the colorimetric control of hardware is done using "Imatest".<sup>37</sup> For checksum the algorithms "FileZilla Client"<sup>38</sup> and a BnF software, "EntryChain" are used. The digitization workflow tool is the BnF software "SIPIL".

*Control of the environmental conditions (see D5.1, ch. 3.1.3)*

- Quality control of Europeana Regia manuscripts is carried out in room lighting at 5.000 K.
- Digital images are viewed in 1:1 size in relation to originals.
- Monitors are calibrated.
- Colour viewing booths are used if necessary.
- Day-light is excluded.

## **Bibliothèque royale de Belgique**

### *Hardware and software*

Quality control is performed on-screen. Monitors are calibrated. The targets (colour, greyscale and resolution) have been provided by the manufacturer of the scan material. "X'n View"<sup>39</sup> software is used during the process to rename, resize, and re-orientate the pictures.

A recent external audit evaluated the optical parameters of the digitized results. The external audit included a test of the scanner with Certifi Media Inc. software,<sup>40</sup> based on the FC-1C target:<sup>41</sup>

<sup>35</sup>See <<http://www.treventus.com/products/scanflow.html>>.

<sup>36</sup><[http://xritephoto.com/ph\\_product\\_overview.aspx?ID=1192](http://xritephoto.com/ph_product_overview.aspx?ID=1192)>

<sup>37</sup><<http://www.imatest.de/?gclid=COPvori8xKsCFQIYzQodVyDW>>

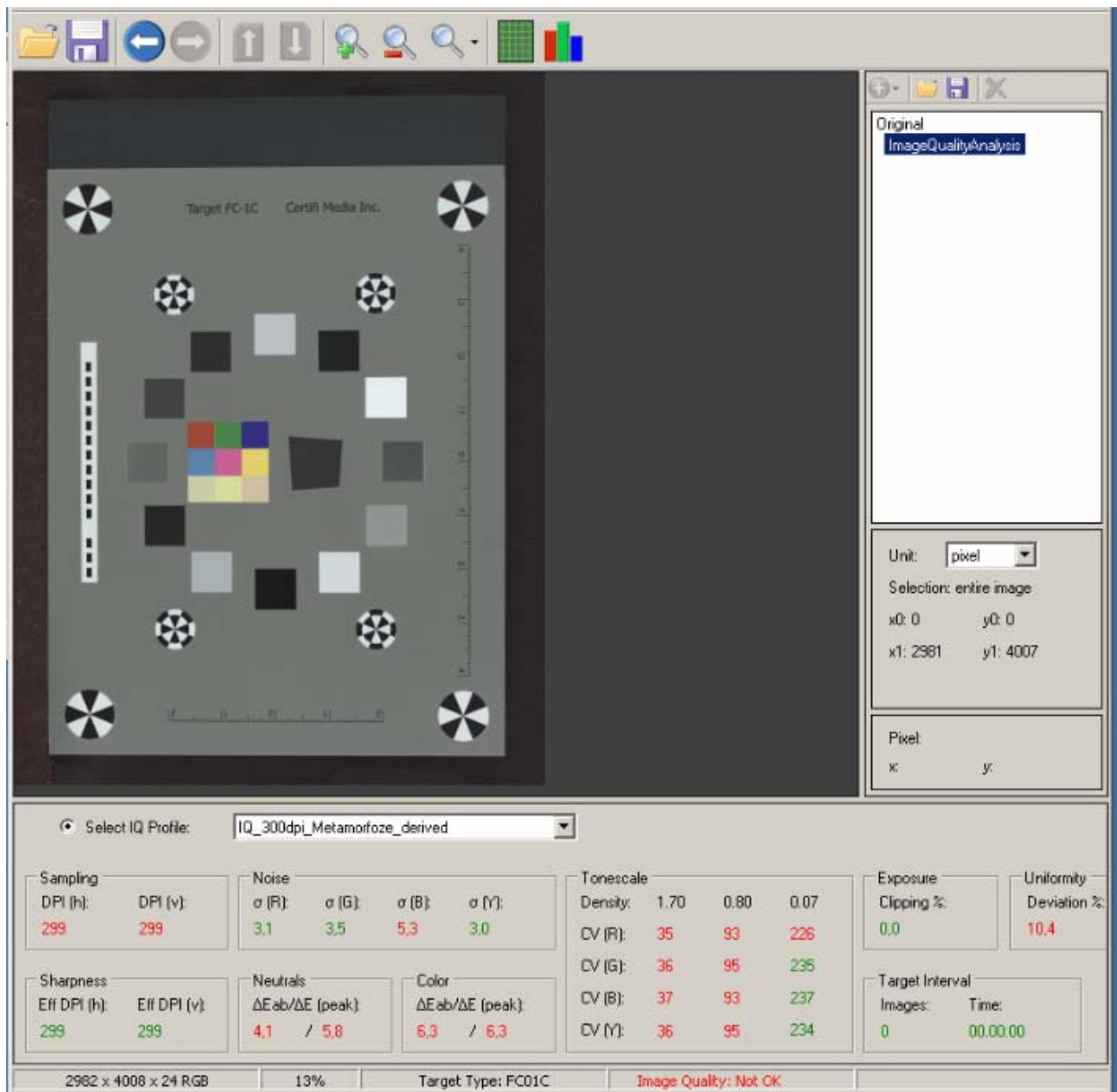
<sup>38</sup><<http://filezilla-project.org/download.php>>

<sup>39</sup><<http://www.xnview.com/>>

<sup>40</sup><<http://www.certifi-media.com/>>

<sup>41</sup><<http://www.certifi-media.com/Targets.aspx>>





The external audit launched by “Diadeis” resulted in the following findings:

- Sampling : 299 DPI, horizontal and vertical
- Sharpness : 299 DPI, horizontal and vertical
- Noise level:
  - $\sigma$  (R) : 3.1,
  - $\sigma$  (G) : 3.5,
  - $\sigma$  (B) : 5.3,
  - $\sigma$  (Y) : 3.0.
- Colour:
  - $\Delta E_{ab}$ : 6.3,
  - $\Delta E$  (peak): 6.3.
- Uniformity: 10.4 % deviation.
- Exposure: 0.0 % clipping.
- Size ratio: 1/1.

**Biblioteca Historica – Universitat de València**

*Hardware and software*

Monitors:

17" monitor, LCD, TFT, 75 Hz, 250 cd/m<sup>2</sup>, 1280x1024, 16.2 million colours.

Computers:

Processor: VGA QUAD CORE Q9400 750 GB 4096MB RWDVD; computer: XTREM QUAD CORE Q9400 (6MB, 2,66Ghz, 1333 MHz), 4096 MB RWDVD XTETIE24HQADNEW Dimm 4096 mb Pc 800 DDR2 (2x2GHZ); Graphics card: MSI 9500GT 1GB; Hard drive: SATA II 750GB

Colour targets:

A metric scale and a QpCARD 201 colour standard are used in the Europeana Regia project.<sup>42</sup>

Calibration and other software:

A calibration process was periodically carried out to adjust the configuration of the colour conversion of the monitor to a standard level. The monitor calibration hardware and the corresponding software used were *Spyder3Pro*.<sup>43</sup> The histograms of the images were viewed with Photoshop CS4.<sup>44</sup>

BookDrive Capture, version 4.1.2.0,<sup>45</sup> automatically checks the Modular Transfer Function; this is checked periodically with the EOS utility 1.1.0.2. The same programme also automatically checks the algorithms.

Digitization workflow tools:

- BookDrive Capture 4.1.2.0,
- Photoshop CS4,
- Irfanview 4.3.<sup>46</sup>

*Control of the environmental conditions (see D5.1, ch. 3.1.3)*

Quality control of the digitized images is based on visual onscreen verification of the master images stored in \*.tif files on a 1:1 scale (100%), with simultaneous comparison to the corresponding original in a controlled environment.

The monitor is protected from direct sunlight. The *Spyder3Pro* sensor is used as a photometer. The lighting is white light projected from above and controlled with the aid of *Spyder3Pro*.

## **Herzog August Bibliothek Wolfenbüttel**

*Hardware and software*

Quality control is performed on-screen using EIZO colorEdge CG 220 monitors. Monitors are calibrated using x-rite colormunki™ software.<sup>47</sup> Adobe Photoshop Lightroom is used as photo management software.<sup>48</sup> A batch histogram correction routine is implemented in Lightroom. Moreover, X-Rite ColorChecker is used.<sup>49</sup> White matching is performed by the ColorChecker function implemented in Lightroom. Digitization workflow is data base supported (software).

### **5.1.4. Documentation of quality management**

**Bayerische Staatsbibliothek:**

*General workflow documents and processes*

---

<sup>42</sup> <<http://www.qpcard.se/BizPart.aspx?tabId=31&catId=1&tci=100&prod=6&tci=146>>

<sup>43</sup> <<http://www.datacolor.eu/de/produkte/monitor-kalibrierung/spyder3pro/>>

<sup>44</sup>

<<http://success.adobe.com/de/de/sem/products/photoshop.html?kw=p&ssid=IGALF&skwid=TC|22767|photoshop%20CS-4||S|b|8824536182>>

<sup>45</sup> <<http://www.atiz.com/bookdrive-capture/>>

<sup>46</sup> <<http://www.irfanview.net/>>

<sup>47</sup> <<http://www.colormunki.com/>>

<sup>48</sup> <<http://www.adobe.com/de/products/photoshoplightroom/features/>>

<sup>49</sup> <[http://www.xritephoto.com/ph\\_product\\_overview.aspx?catid=28](http://www.xritephoto.com/ph_product_overview.aspx?catid=28)>

For the following steps within the digitization workflow, written documents with detailed descriptions of processes, which apply to all projects, have been prepared:

- planning of digitization projects and estimation of costs,
- condition survey of manuscripts to be digitized (see D4.1, ch. 7.2),
- digitization workflow,
- ZEND processes within the project workflow,
- several instruction papers for long term preservation of digital objects.

#### *Project-specific scanning instructions*

For each project, specific scanning instructions ('Scanleitfaden') are produced, which contain information on the hardware to be used and its setup, the scanning parameters, on the handling of books, frames to be used, and the treatment of covers and spine etc. It also includes information on the specific quality management measures to be applied, e.g. on the placement of colour and resolution targets, the number of test scans and the frequency of sharpness checks to be executed. It also contains information on the processes for correction and rework.

#### *Guidelines for quality control*

Specific guidelines for quality control in the digitization of rare objects are documented in guidelines entitled „Leitfaden Qualitätsprüfung von Images von Handschriften und Alten Drucken im ZEND-ToC-Editor (ohne OCR)" (Guidelines for image quality control in digitization projects of manuscripts and early printed books).

#### *Scan-lists*

In some cases, scan lists for each object to be digitized are prepared, which include, for example, a concordance of digital image scan file names to page numbers and hints on problematic areas of the original.

#### *Administration excel files*

The results of quality control for each digital object are documented in the so-called "administration" excel-files. These files include information on the dates of quality control, release, and long-term preservation of files. At the same time, they record faults in the digital files that need correction and rework.

#### *Treatment of (gilt) book covers, edges, and spines*

The complex workflow for the digitization of these book parts has been laid down in a separate process paper.

### **Bibliothèque nationale de France**

Quality documents:

- written documents describing BnF processes,
- guidelines for digitization and control of bound and special collections at the BnF,
- foliation guidelines,
- guidelines for manual handling of documents,
- guidelines for audits,
- audit reports,
- control slip: computer file sent to the service provider, indicating rejected and validated documents, and the reasons for rejection.

As scanning instructions, a transfer slip (computer file) and a written paper (see example below) are produced.

### **Bibliothèque royale de Belgique**

As yet, no in-house quality documents about the digitization of manuscripts are available. A scan-list (Excel) has been established for the scan-process.

### **Biblioteca Historica – Universitat de València**

To exercise proper control over the digitization process and control management, the following documents have been developed:

- digitization guidelines,
- quality control manual,
- technical specifications for digitization,
- archive nomenclature instructions,
- specifications for assigning technical metadata,
- specifications for assigning structural metadata,
- specifications for assigning descriptive metadata for illustrations,
- digitization tables,
- quality control checklists,
- report for problems detected in the quality control of digitized objects.

At the beginning of the project, and in accordance with the indications specified therein, a list of necessary requirements was provided to the staff in charge of digitization.

At the same time, each manuscript has a digitization table indicating the characteristics of the document and information on its state of conservation.

### **Herzog August Bibliothek Wolfenbüttel**

Digitization procedures follow the *DFG-Richtlinien zur Digitalisierung*.<sup>50</sup> Internal rules apply to the database-supported workflow (including, but not limited to name conventions for file names, resolution and data formats for master files and for Internet conversion).

#### **5.1.5. Timing of quality control**

##### **Bayerische Staatsbibliothek:**

All digital objects are submitted to quality control as soon as the images are available on the ZEND workflow tool. This is normally the case on the next working day. In the case of very precious objects which are stored in special safes, these have to be transported back after scanning and have to be fetched again the following day for quality control.

##### **Bibliothèque nationale de France**

The results are submitted to quality control one week after delivery. First automated structure checks are carried out, and then visual control takes place. The latter takes place before the documents are online and is exhaustive: all images are checked. It is still possible to reject them in a period of six months after the online release.

##### **Bibliothèque royale de Belgique**

The digitization results are submitted to quality control after scanning. The quality review process has been drafted recently and therefore has been applied only to a small number of manuscripts (regarding the situation at the 1st September 2011).

A check of a limited sample of images was part of the external audit.

### **Biblioteca Historica – Universitat de València**

Depending on the size and number of folios in each manuscript, there is a time delay of 10 to 15 days from the global digitization process and subsequent treatment of the images to the delivery for quality control, which is carried out by library staff.

---

<sup>50</sup> 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)>.

### **Herzog August Bibliothek Wolfenbüttel**

Manual/intellectual quality control is performed immediately after digitization and prior to the publication of digital files.

#### **5.1.6. Time delay between digitization, quality control and release**

##### **Bayerische Staatsbibliothek:**

The release of images takes place immediately after quality control if no faults in the digital objects are detected. Delays may occur in cases when corrections and rework are required.

##### **Bibliothèque nationale de France**

Due to the fact that the BnF works with a service provider, there is always a delay between digitization and delivery and the control at the BnF. After delivery, the BnF has 8 weeks for the automated structure checks and the visual control. After this, release will not take more than three weeks.

Quantities of scanned documents are very important at the BnF. So, the digitization line (and the tasks involved) is very fragmented and the different operations on digital files take some time.

##### **Bibliothèque royale de Belgique**

Time delays are not to be expected.

##### **Biblioteca Historica – Universitat de València**

The process of digitization, quality control and validation of results is carried out sequentially. Should errors be detected or if any of the images need to be corrected, the process may be delayed, following a sequence of priorities based on the digitization order, thus giving preference to the correction of errors detected during the quality control of any given manuscript.

At the beginning of the project, delays occurred due to the necessity of making changes in the digitization process to achieve optimal final results in accordance with the established specifications.

### **Herzog August Bibliothek Wolfenbüttel**

There is no significant delay between digitization and quality control. Data are rapidly released online.

#### **5.2 Parameters of Quality Control**

##### **Bayerische Staatsbibliothek:**

###### *Technical parameters*

The following technical parameters are subject to quality control:

- file format and size,
- completeness of files,
- conversion errors from TIFF to JPEG.
- TIFF headers are only checked for the correct registration of colour profiles and colour ranges.
- The object is compared to bibliographical metadata in the ZEND software.
- The naming of files. The barcode system used by the ZEND workflow software reduces the risk of manual keying errors.

## Europeana Regia: Kooperative digitale Bibliot...

Augustinus, Aurelius: De magistro - BSB Clm 6822, . [S.] Frankreich 2. HXCBelBe 8. Jh.

urn:nbn:de:bvb:12-bsb00066386-3

## Digitalisierungsauftrag (Fotostelle)

Signatur:	Clm 6822
Kurzinfo:	Clm 6822
ID	bsb00066386

## Digitalisierungsvorgaben

Name der ersten Seite*	bsb00066386_00001.tif
Auflösung	400 ppi
Farbtiefe	Farbe

\*Wenn der Rücken mitgescannt werden soll, beginnt die Digitalisierung mit dem Rücken und die Imagazählung mit Null. Beispiele: bsb00000001\_00000.tif oder 2003-03-04-08-19-54\_00000.tif

## Lieferinformationen (Umfang vollständig!)

Ausdruck			CD-ROM
Original-	vergrößert	verkleinert	DVD
größe	DIN A4	DIN A3	Mikrofilm

## Lieferung der Tiff-Dateien an

Abholer/cher der Scanner:
---------------------------

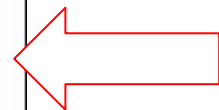
München, den 20.10.2011      Unterschrift: \_\_\_\_\_

## Rückgabevermerk (Dienstleister)

Anzahl der Scans: \_\_\_\_\_

Datum: \_\_\_\_\_

Unterschrift: \_\_\_\_\_



Digitization order slip with barcode system to avoid keying errors

## Optical image quality:

In a prestigious project like Europeana Regia, each image is controlled manually and intellectually by a direct comparison to the original. The following parameters are always checked:

- resolution,
- colour accuracy,
- tonal reproduction,
- noise,
- critical areas of manuscripts (e.g. illumination, use of gold, silver and purple),
- geometric distortion.
- Treatment of artifacts. If they can be removed from the surface of pages, they are scanned in a separate image at the end of the image sequence. If not, they are left in place.
- Special attention is paid to the legibility of text passages. The text should be legible also near the inner book folds in tight bindings, but of course text loss is sometimes inevitable.
- Completeness and the correct sequence of pages are checked against the original (page by page). It is checked if pages have been scanned with an appropriate angular orientation. This

also depends on the original: if a page has been bound at an angle in the original, it should also appear oblique in the digital copy.

- The position of frames and backing sheets according to project specifications are checked.
- Errors of foliation are recorded. If there has been no accidental omission of pages, these are not corrected.

## **Bibliothèque nationale de France**

### *Coherence and completeness:*

- comparison of bibliographic information via BTA,
- check of pagination / foliation (according to the "guide de foliotation et pagination"),
- completeness.

In this process, user comments or product reviews are included.

### *Control of image quality (visual rendition of image including colour rendition):*

- quality of images and reproduction (definition, contrast, potential saturation etc.),
- quality degradation between original document and digitized file (colour rendition, completeness, truncation, corrugation, shadows, etc.),
- comparison with the original under identical viewing conditions,
- homogeneity with regard to the treatment of the entire set of digitized documents,
- cropping errors,
- image too bright / too dark.

## **Bibliothèque royale de Belgique**

### *Technical parameters:*

- file format,
- completeness of files,
- analysis of technical metadata,
- correct naming of files.

The validity of paths and links cannot be checked as long as the ingestion of manuscripts on the KBR web-site is not achieved.

### *Optical parameters*

during the quality process and approved by the external audit:

- resolution,
- colour accuracy,
- tonal reproduction,
- image artifacts,
- noise and dynamic range,
- critical areas of manuscripts (e.g. illumination, use of gold, silver and purple),
- geometric distortion.

The following parameters have to be checked during the final quality control:

- completeness of images,
- orientation of images,
- sequence of images,
- errors in foliation.

## **Biblioteca Historica – Universitat de València**

### *Technical parameters*

The technical and structural metadata are included in a \*.txt file which is checked during the quality control process. The completeness of files is checked.

An analysis of technical metadata (e.g. TIFF headers) is carried out by the digitization technician. The validity of paths and links and the correct naming of files are also submitted to quality control.

### *Optical image quality*

The following parameters are checked:

- resolution,
- colour accuracy,
- tonal reproduction,
- image artifacts,
- noise and dynamic range,
- focus and pixeling,
- geometric distortion,
- critical areas of manuscripts (e.g. illumination, use of gold, silver and purple): The difficulties in digitizing folios illuminated with gold or silver led to the adoption of new methods in the process of image capture, taking into account any repercussions on the final result and any significant effects on the reproduction of the remaining colours.

### *Completeness of images*

It is assured that there is a margin around the image of the complete folio. Other aspects are:

- orientation of images,
- sequence of images,
- errors in foliation.

## **Herzog August Bibliothek Wolfenbüttel**

### *Technical:*

Controls involve file formats, completeness of files, and correctness of filenames. Path validity and links are also checked in a database supported mode.

### *Optical:*

Resolution, colour accuracy and colour rendering are checked.

### *Completeness:*

Completeness of images, orientation and sequential order of images are controlled.

### *Content:*

The foliation of manuscripts is typically checked prior to digitization by a member of staff of the manuscripts department rather than by scan operators or photographers.

## **5.3. Methods and Workflow**

### **Bayerische Staatsbibliothek:**

#### *General*

One expert in the library's scanning centre is responsible for overall quality management; this person has a background in image editing in the printing sector; he or she may also instruct other members of staff if necessary.

In the BSB, quality control in the most prestigious variant of digitization projects, i.e. manuscript projects, depends on careful manual/intellectual control of every single image. Quality management is carried out by members of staff of the library's scanning centre. There is usually no overlap between scan operating and quality control staff.

Qualified user feedback which reaches the library's departments via e-mail, or the knowledge management tool QuestionPoint is more relevant in mass digitization projects of printed books, in which no extensive quality control takes place before the publication on the internet. Should user feedback indicate quality problems also in the Europeana Regia project, it will be considered; justified complaints about image quality will then lead to correction and rework.

#### *Correction and rework*

If quality staff detect a fault in the digital object, a handwritten correction slip is added to the object documentation. The original is returned to the scan operator in charge of the object, who takes a fresh scan under the same production parameters as the original scan.



The new image is then inserted in the correct position of the digital file; file names are corrected if needed.

This corrected file is subjected to quality control a second time, when, however, only the passages which have been reworked are considered. Once rework has been successfully completed, the digital object is released and the master files are transferred to long-term preservation. As the last step of quality control, the archival report within the ZEND workflow tool is checked to see if there have been any irregularities during the transfer to long-term preservation and the digital files are deleted from the local scanning stations.

## **Bibliothèque nationale de France**

### *General*

Digital objects are checked against the originals (book in hand). Automated methods and manual/intellectual control are used.

Qualified user feedback which reaches the library's departments via e-mail is included as well (during the six months after the online release of the documents, the service provider can correct the error).

### *Correction and rework*

The digital documents can be rejected after having been controlled. The BnF cannot correct them in-house. Hence it is the task of the provider to make corrections within two months. If corrections are required, the digitized document is entirely rejected. After correction and renewed delivery by the provider, all controls are done again.

The Europeana Regia project is a part of an important contract for special collections digitization. Quality management tools were created for this contract. A project board is held every month. It is the decision-making authority for quality problems. A long test phase was established before starting the production to calibrate the processes and define quality expectations with the provider. But tests can be done during the production phase for particularly difficult manuscripts.

Example for scanning instructions:



## Fiche d'accompagnement de numérisation

Num lot répertoire : <input type="text" value="0"/>	Cote de l'original: Français 256
Num doc. numérique: <input type="text" value="0"/>	Conditionnement : Sans
	Programme: Europeana Regia Charles V

Format max: A3

Dimensions: 400x290

Type de numérisation: Couleur

Ouverture: 120°

Nombre d'images: 408

Foliotation/pagination:

Nombre de feuillets ou pages: 199

Foliotation moderne

Nombre de feuillets np au début: 1

Nombre de feuillets np à la fin: 1

Consignes de manipulation:

Consignes de prise de vue:

- Ms très ancien et très précieux ; prière de le manipuler avec précaution.

- prises de vue de la reliure (8 images - les tranches sup. et inf. doivent être présentées verticalement ainsi que la gouttière).
- Manuscrit comprenant des dorures qui doivent impérativement être restituées avec éclat + couleurs.
- premiers feuillets foliotés A, B

*Visa du conservateur responsable de la numérisation :*

**Bibliothèque royale de Belgique**

*General*

Digital objects are checked against the originals (book in hand). Automated methods are used (for example software for the renaming process and the right orientation of the sheets). User feedback (via email, personally) will be considered.

#### *Correction and rework*

In general, the organisation of the quality control is done in the department of manuscripts of the KBR. An external audit has recently checked the resulting images, in terms of quantitative analysis (MTF, deltaE, sharpness, sampling, etc.).

For the moment, rework and corrections have been processed internally, via the same workflow as the regular production (status update in the scan List, re-scan of the page[s] concerned).

### **Biblioteca Historica – Universitat de València**

The detection of errors in the digitized images that must be corrected by the digitization provider have led to the establishment of a corresponding protocol described in the quality management manual as well as in the digitization guidelines (section 3.3.5.). Basically, the process comprises the following steps:

- a) Issuing a form listing the problems detected.
- b) Having the digitization provider make the appropriate corrections within 7-10 days.
- c) Revision of the corrections.
- d) Issuing a quality control checklist with a positive assessment of the digitized document.

### **Herzog August Bibliothek Wolfenbüttel**

#### *General*

At the HAB, there is manual/intellectual control for manuscripts projects; user feedback is considered as well.

#### *Correction and rework*

Digitization errors (optical errors, technical faults) are typically diagnosed and corrected immediately after scanning, prior to file release, by members of staff of the digitization department. Prior to their publication in the Digital Library, digital files are checked by members of staff of the manuscript department as well. If errors are discovered by a comparison to the original document, these are reported directly to the scan operator in charge of the object. Typically, errors are corrected immediately. If necessary, the original object is consulted (for example in the case of missing pages). Errors which are reported by library staff or users subsequent to the publication of digital files are corrected as well. In order to guarantee stable or persistent URLs, image numbering must be conserved in case of retroactive corrections. When this is not feasible, for example when duplicate pages need to be deleted, the identifier of the deleted page is redirected to the correct image. This type of corrections requires a close cooperation of the scan operators and the „digital library“ department.

## References:

Borowski, Marion *et al.* (Fraunhofer Institut Intelligente Analyse- und Informationssysteme): "Bestandsaufnahme zur Digitalisierung von Kulturgut und Handlungsfelder", available at <[http://www.iais.fraunhofer.de/bkm\\_studie.html](http://www.iais.fraunhofer.de/bkm_studie.html)>

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 (FADGI) - Still Image Working Group: *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files* (2009).

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).

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>>

ISO 3664, (*Viewing conditions – graphic technology and photography*), International Organization for Standardization, Washington, DC.

ISO 9000 (*Quality management systems - Fundamentals and vocabulary*), International Organization for Standardization, Washington, DC.

ISO 9001 (*Quality management systems - Requirements*), International Organization for Standardization, Washington, DC.

ISO 9004 (*Quality management systems - Guidelines for performance improvements*), International Organization for Standardization, Washington, DC.

ISO/TR 10013 (*Guidelines for quality management system documentation*), International Organization for Standardization, Washington, DC.

ISO 12646 (*Graphic technology – displays for colour proofing – characteristics and viewing conditions*), International Organization for Standardization, Washington, DC.

Kenney, Anne R., Oya Y. Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, CA, 2000).

Library of Congress: Requests Proposals For Digital Images of Pictorial Materials. National Digital Library Program, available at: <<http://memory.loc.gov/ammem/prpsal9/rfp9.pdf>>.

Maier, Gerald, "Qualität, Bearbeitung und Präsentation digitaler Bilder", in: Hartmut Weber, Gerald Maier, eds., *Digitale Archive und Bibliotheken: Neue Zugangsmöglichkeiten und Nutzungsqualitäten* (Stuttgart: 2000), p. 129-177.

*Moving Theory into Practice: Digital Imaging Tutorial* (Cornell University Library/Research Department), available at: <<http://www.library.cornell.edu/preservation/tutorial/index.html>>

Riley, Jenn and Kurt Whitsel, "Practical quality control procedures for digital imaging projects", in *OCLC International Digital Library Perspective* 21.1 (2005), p. 40-48.

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://imagendigital.esteticas.unam.mx/PDF/Guides.pdf>>

Sitts, Maxine K., ed. for the Northeast Document Conservation Center: *Handbook for Digital Projects: A Management Tool for Preservation and Access* (Andover, MA: 2000).

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), <http://www.dlib.org/dlib/february01/thaller/02thaller.html>

### **Closely connected deliverables of the Europeana Regia project:**

D2.1: State of the art in metadata.

D2.2: Catalogue requirement for metadata.

D4.1: State of the art in image processing.

D6.3: Report on users' requirements, following meetings with users' representatives and online survey.

## Europeana Regia Questionnaire:

This questionnaire was sent to more than 20 institutions with experience in digitization projects of 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