



ECP-2008-DILI-538002

EuropeanaTravel

Best practice examples in library digitisation

Deliverable number	<i>D2.2</i>
Dissemination level	<i>Public</i>
Delivery Date	<i>March 2011</i>
Status	<i>Final</i>
Author(s)	<i>Karmen Štular Sotošek National and University Library, Ljubljana, Slovenia</i>



eContentplus

This project is funded under the eContentplus programme¹, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

¹ OJ L 79, 24.3.2005, p. 1.

Table of the Content

Executive Summary	3
1. About this Document	4
1.1 The purpose of this document and its relationship with other Europeana Travel documents	4
1.2 Input for this document	6
2. Digital Content Creation and Access to the Digital Content	8
2.1 Best practice examples of capturing and managing images	8
2.1.1 Image capture standards	8
2.1.2 Digital capture equipment	11
2.2 Best practice examples of handling the materials	12
2.3 Best practice examples of metadata generation	12
2.4 Best practice examples of Optical Character Recognition (OCR)	14
2.5 Best practice examples of access to the digital content	16
3. Managing Digitisation Projects	18
3.1 Best practice examples of the strategic documentations on digitisation	18
3.2 Best practice examples of workflow management	19
3.3 Best practice examples of quality assurance	22
3.4 Best practice examples of cost modelling for digital preservation	25
4. User Focus	27
4.1 Best practice examples of user-orientated digital content and user evaluation studies	27
5. Summary	30
References	31

Executive Summary

This is the second of two documents to be prepared and published as a result of the “Implementing digitisation” (WP2) work package within the *EuropeanaTravel* project. This document deals with best practice examples in digitisation. It presents a selection of the most recent best practices in the field of digitisation, those that have been recognised by the project partners over the duration of the project. As well as performing digitisation to a high and consistent standard, we have recognised many best practices, covering the capturing and management of images, handling the material, metadata generation, OCRing, access to the digital content, workflow management, strategic planning, quality assurance, cost modelling and user evaluation that are presented in this document. It also takes into account the best practices in the field of digitisation that had been established and presented in the framework of related projects, such as TELPlus, IMPACT, EOD, etc.

This document is public so that other cultural institutions have the chance to benefit. This practical document addresses the typical development questions of those who wish to digitise material.

1. About this Document

1.1 The purpose of this document and its relationship with other *Europeana Travel* documents

This is the second of two documents to be prepared and published as a result of the “Implementing digitisation” (WP2) work package within the *EuropeanaTravel* project. This document deals with best practice examples in digitisation. It presents a selection of the most recent best practices in the field of digitisation – those that have been recognised by the project partners during the duration of the project. Sixteen institutions from fourteen European countries have contributed their digital content, which will be made available through Europeana. As well as performing digitisation to a high and consist standard, we have recognised many best practices covering the capturing and management of images, handling the material, metadata generation, OCRing, access to the digital content, workflow management, strategic planning, quality assurance and user evaluation that are presented in this document. It also takes into account the best practices in the field of digitisation that had been established and presented in the framework of related projects, such as TELPlus, IMPACT, EOD, etc.

This document is public so that other cultural institutions have the chance to benefit. This practical document addresses the typical development questions of those who wish to digitise material.

The central document is divided into three parts:

In the first one we present examples of best practices in the field of creating digital content by the *EuropeanaTravel* project partners. The primary objective of this chapter is to support the best practices for the capture and storage of the archival digital records of selected library items. It presents the best practice models concerning the capture of digital images, the creation of metadata and the generation of OCR text. The correct handling of library materials during the digitisation process is an important component of these processes; the chapter therefore also refers to best practice in the field of handling the materials. This document has been compiled with the awareness that digitisation means a more advanced access to the content, so this chapter ends with best practice examples of access to digital content.

The second part covers examples of best practices in managing a digitisation project. We display sample workflows supporting the digital library development. We recapitulate the best strategic documents of the partners – those intended for a quality digitisation process. We summarize the best procedures to guarantee the quality digitisation of library material. This part also includes best practice examples of cost modelling for digital preservation.

The third part is devoted to user orientation in planning digitisation. We present examples of best practices that also include the users in the planning of digitisation, trying to satisfy them with appropriate digital content and access to it. This chapter describes the basic purpose and significance of digitisation and a digital library. Technology is not enough; we need cooperation with users; we need international cooperation with cultural institutions and partnership with others (publishers, et al.).

This document is connected with the D2.1 deliverable, which was prepared as part of WP2 – *“Infrastructure for quality control”*. Ensuring the quality of the digitisation processes and following standards, recommendations and instructions in this field are crucial for good practice in digitisation process management. The contents of the document are also connected with the WP1 deliverables, namely D1.1. *“Document describing possible standards for discussion in the workshop”*, D1.2. *“Synthesis of digitisation and metadata plans including standards to be applied”* and D3.1. *“Report on metadata standards”*.

1.2 Input for this document

This document is based on the use of quantitative research methods. By analysing the responses to the questionnaire that was the basis of the WP1 workshop of EuropeanaTravel, we have got precise descriptions of carrying out the processes of digitisation in different institutions and in the various countries.

At the same time, we have analysed documentation referring to the digitisation guidelines in the libraries of those partners that sent us completed questionnaires and documents on digitisation as well, or that would allow us to examine their internal official records on digitisation designed for quality digital content. Obviously in most cases we were only able to include and analyse the documents that were available in English.

The *EuropeanaTravel* project brings together 19 national and regional libraries, research institutions and foundations who have shared their know-how and best practices.

List of *EuropeanaTravel* project participants

1	National Library of Estonia	RR	Estonia
2	National Library of Finland	UH.NLF	Finland
3	National Library of Latvia	LNB	Latvia
4	National Library of Poland	NLP	Poland
5	Austrian National Library	ONB	Austria
6	Slovak National Library	SNK	Slovakia
7	National and University Library	NUK	Slovenia
8	EDL Foundation	EDL	Netherlands
9	Eremo srl.	EREMO	Italy

10	University College London	UCL	United Kingdom
11	National Library of Wales	NLW	United Kingdom
12	Lund University Library	LUB	Sweden
13	National Library of The Netherlands	KB	The Netherlands
14	University Library of Regensburg	UREG	Germany
15	Moravian Library in Brno	MZK	The Czech Republic
16	University and Regional Library of Tyrol, University of Innsbruck	UIBK	Austria
17	University and National Library of Debrecen	DE	Hungary
18	Trinity College Library	TCD	Ireland
19	State and University Library of Lower Saxony	UGOE	Germany

Additional sources of information were the findings of other projects, such as the TELplus project (eContentplus, 2007-2009), Digitisation on demand (eTEN, 2006-2008), EOD project - eBooks on Demand (Culture, 2009-2013) and IMPACT (FP7, 2008-2012), as well as the European Library and Europeana practical experience.

Last but not least, some of the information also came from desktop research in literature and documents in this field, which are listed at the end of the deliverable. In researching best practices, we monitored the current literature and in certain cases, where we did not find suitable practices with our partners, we also focused on the practices of other cultural institutions.

The author wishes to thank the external referees who reviewed the document during the preparation process and made many useful suggestions, which influenced the document's final contents.

2. Digital Content Creation and Access to the Digital Content

The primary objective of this chapter is to support the best practices for capturing and storing archival digital records of selected library items. This section describes the production of the digital datasets, focusing on the technical standards. It presents the best practice models of the capture of digital images, the creation of metadata and generation of OCR text. The correct handling of library materials during the digitisation process is an important component of these processes; the chapter therefore also refers to best practice in the field of handling the materials. This document has been compiled with the awareness that digitisation means a more advanced access to content and that “digitisation = access”, so this chapter ends with best practice examples of access to digital content.

2.1 Best practice examples of capturing and managing images

2.1.1 Image capture standards

A lot of materials and documents about image capture are currently available. The primary goal of image capture in each library is to produce a digital image file that is highly representative of the original physical item. Libraries achieve this in one of two ways: the capture is done in-house or is outsourced. Regardless of who performs the digitisation processes, all institutions strive for processes that are as standardised as possible. Standardising the capture ensures the future usability of the electronic image files. The only way to ensure quality and achieve the primary goal of digitisation is to use standardised processes.

Among the different standards for the capture and management of images, we should primarily mention the best practice example of the quality standards set and adopted by the *EOD project*² partners.

² EOD project- eBooks on Demand - A European Library Network, financier: European Commission, Culture programme. The project is a continuation of the eTEN DOD project (Digitisation on demand), which was carried out from 2006 to 2008. „eBooks on Demand” is a network of currently 30 libraries from 12 European countries providing their holdings for digitisation on demand. The quality standards laid out in the mentioned document were agreed between the project participants of the eTEN DOD project with the focus on how to best handle digitised document delivery. It might not be applicable to mass digitisation or other digitisation projects.

Best Practise Example: Quality Standards for EOD, EOD project, Culture program, Version 2.2. (22/07/2009), <http://books2ebooks.eu>

Quality Standards for EOD covers all aspects of book capture with elements that are obligatory (must), recommended (should) or optional (nice to have) for EOD members. The fact that this document includes three-levels of options for monitoring standards is what makes it so important. This document is undoubtedly the best when it comes to systematic and simple recommendations for full informational capture. While the document only prescribes actions for digitising books, the selection and systematic nature of the standards presented therein make it useful regardless of the type of material that is being digitised. The recommendations cover the following processes and material features: preparation, completeness, supplements, book cover, empty pages, disordered and missing pages, resolution, colour depth, image formats, compression, image processing, file naming, the delivery of files, OCR, formats for delivering the file to customers and processing time.

The best example of imaging standards and requirements for all media undergoing digitisation and digital archiving is the document from the Trinity College Library, Dublin entitled Digital Resources and Imaging Services; Digital Imaging Standards Policies V.6, The University of Dublin, June 2009.

The primary objective of this standards requirements document is to support the best practices for the capture and storage of an archival digital record of selected library items to be included as a new 'virtual' or digital collection within the existing domain of Trinity College Library, Dublin.

Best Practise Example: Digital Resources and Imaging Services; Digital Imaging Standards Policies V.6, The University of Dublin, June 2009.

This document represents best practice because it acquaints the professionals involved in digitisation projects with the sets of well-defined stages in capturing and managing images. The document also anticipates potential difficulties and problems in producing quality digital copies in the various stages and proposes corrective procedures to eliminate them. The biggest value of the document lies in both its practical nature and in the holistic and systemic view presented.

The paper includes images and offers an in-depth discussion of all the stages and difficulty levels of individual approaches to image capture and processing:

<p>Image Structure</p> <ul style="list-style-type: none"> File Types Spatial Resolution Bit Depth Colour Mode Colour Management Colour Profile Rendering Intent 	<p>Physical/File Structure</p> <ul style="list-style-type: none"> Cropping Backing File Saving Reference Targets Capture 	
<p>Calibration/Characterization</p> <ul style="list-style-type: none"> Reference Targets Imaging Devices Monitor Quality Assurance Image Quality Metadata Quality Image Processing 	<p>Metadata</p> <ul style="list-style-type: none"> Rights Technical 	<p>Processing</p> <ul style="list-style-type: none"> Overview Physical Adjustments Tonal Adjustments Aim Points Sharpening Artefact Removal Colour Correction Watermarking

The document contains detailed technical specifications regarding the size of digital objects and the possible formats for creating quality digital surrogates that are as similar as possible to the physical originals. The value of the document also lies in the fact that it anticipates the potential anomalies that could appear in the digitisation process and lists processes for the correction and elimination of such anomalies.

This excellent document contains advice on how to:

- Standardize the image structure for the capture of all electronic media files that are designated part of the libraries' digital collections.
- Standardize the capture and storage criteria for the digital assets to ensure the future usability of the electronic image files, supporting the widest possible audience.
- Recognize the inherent value of the digital records and provide the necessary resources for adequate long term management and preservation.
- Develop processes and procedures to support the perpetual storage and retrieval of non-human readable digital information.
- Manage and protect the libraries' public and private digital assets from unapproved use and dissemination.

2.1.2 Digital capture equipment

Other than image capture standards, our attention should also be directed to the issue of digital capture equipment. The library makes use of a variety of digital capture devices, including flatbed scanners, film/slide scanners, microfilm/microfiche scanners and digital cameras.

Based on our analysis of our partners' documents and the analysis of documentation concerning similar projects we found that this field of digital capturing, which depends on the commercial providers of the hardware and software equipment, is only discussed in passing and within the context of other digitisation processes. We therefore searched elsewhere for examples of best practice in this field and found the best guidelines for choosing capture device on the *National Library of Australia* website.

Among our best practice examples, the Australian document especially deserves attention because of its simplicity and conciseness in using sophisticated and complex digital technology. The *National Library of Australia* briefly introduces the most basic instructions, adapted to even technically inexperienced staff, regarding suitable equipment for individual types of material: e.g. which materials require a large format scanner, which materials are captured using exclusively digital cameras and what kind of

equipment is used for materials smaller than A3. The short paper recommends equipment for faster and better digital capture.

Best Practise Example: Digital capture equipment, National Library of Australia, <http://www.nla.gov.au/digital/capturedevice.html>

2.2 Best practice examples of handling the materials

Care of library collections is paramount and correct handling is an important part of the digitisation process. The correct handling is surveyed prior to digitisation and minimal handling during the digital capture process is preferred (Guidelines for choosing a capture device, <http://nla.gov.au/digital/capturedevice.html>). Naturally this field is closely connected to the field of digital capture equipment, since it is very important how the material is processed using the prescribed equipment.

Again, the most generic guidelines can be found in the regulations of the *National Library of Australia*. They comprise general rules (e.g. wash hands regularly, always have plenty of room in your workspace, etc.) as well as detailed instructions for individual special collection materials – maps, manuscripts, photographic items and artworks. These guidelines are intended as a starting point for handling materials and therefore represent best practice examples.

Best Practise Example: Care and handling guidelines for the digitisation of Library materials, National Library of Australia, (http://www.nla.gov.au/digital/care_handling.html)

2.3 Best practice examples of generating metadata

Metadata, most generically defined as »data about data« is an essential ingredient needed to support almost all the current approaches to digital-collection

interoperability and aggregation (The Whole Digital Library Handbook, 2007, str.303). To be founded and included in the portal or website, all digital collections and their items require a metadata record in a structural way.

The European Library offers its partners a practical Handbook, in which a special chapter covers the subject of providing and managing collection descriptions and metadata requirements. The European Library Handbook is a tutorial that assists national libraries in making their collections available via The European Library portal. It provides assistance for partners in choosing their access protocol, submitting their collection descriptions and making their metadata compliant with The European Library Application Profile.

The European Library has been a leading organisation in the digital library environment and its practices will be of relevance to many other libraries and institutions dealing with digitisation. Its handbook therefore holds an important place among our best practice examples.

Best Practice Example: The European Library Handbook, section: Working with The European Library,
http://www.theeuropeanlibrary.org/portal/organisation/handbook/handbook_en.html

Understanding metadata generation within the framework of The European Library Handbook is based on the concept of "core metadata" for simple and generic resource descriptions, also known as the fifteen-element "Dublin Core", which achieved widespread dissemination as part of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and has been ratified as IETF RFC 5013, ANSI/NISO Standard Z39.85-2007 and ISO Standard 15836:2009.

From the perspective of the Dublin Core community (Dublin Core Metadata Initiative, <http://dublincore.org/metadata-basics>, (last accessed 2010-11-12)), the metadata landscape is currently characterized in terms of four "levels" of interoperability:

1. shared term definitions: interoperability among metadata-using applications is based on shared natural-language definitions;
2. formal semantic interoperability: interoperability among metadata-using applications is based on the shared formal model provided by RDF, which is used to support Linked Data. As defined in Wikipedia, the term "Linked

- Data" describes "a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs [Web addresses] and RDF.";
3. description Set syntactic interoperability: applications are compatible with the Linked Data model and, in addition, share an abstract syntax for validatable metadata records, the "description set";
 4. description set profile interoperability: the records exchanged between metadata-using applications also follow a common set of constraints, use the same vocabularies and reflect a shared model of the world.

2.4 Best practice examples of Optical Character Recognition (OCR)

Automated text recognition, carried out by Optical Character Recognition (OCR) engines frequently does not produce satisfying results for historical documents. Recognition rates are poor or even useless. The IMPACT project³ pushes innovation in OCR technology and language technology for historical document processing and retrieval and shares the expertise to build capacity in digitisation across Europe. The project develops innovative tools to enhance the capabilities of OCR engines and the accessibility of digitised text and lays down the foundations for mass-digitisation programmes (<http://www.impact-project.eu/about-the-project/concept/>).

The IMPACT Decision Support Tools document, which was compiled in 2010 as part of the IMPACT project, contains practical instructions and guidelines for people dealing with the digitisation of text-based historical material. The document presents various types of OCR use, standards and methodologies. The document also offers insight into the history of the automatic translation of text-based images, contains a detailed description of how OCR works, explains when the use of OCR is recommended and when automated text recognition is best avoided. The document presents a selection of the best practices in OCR use by presenting the main factors that can influence OCR results. Special emphasis is placed on raising awareness regarding the most common problems with OCR, which can be successfully avoided if we are aware of them before the beginning of the project.

³ IMPACT – FP7, 2008-2012

Best practice example: IMPACT Decision Support Tools/ OC2, Impact project, 2010.

The document also presents two main types of post-correction in development. The first one is automatic post correction with the help of specialised dictionaries treating historical language.

Another experimental field that shows some encouraging results is collaborative correction, whereby users help in correcting digitized documents. The full text of a digital document is shared with a resource's users, along with a toolkit that allows them to correct the original OCR results. The IMPACT Project is building a sophisticated "carpeting" tool for collaborative correction, whereby a great many characters and words with suspect OCR results can be corrected and filed. The more characters a user corrects, the more the tool's ability to recognise characters grows – so the users will be developing the OCR engine.

All of these features of the IMPACT Decision Support Tools document demonstrate its high theoretical and especially practical value, bringing a range of new additions and innovations to the current practice of OCRing. Involving users in correcting digitised documents in particular reflects a wider social engagement in achieving better results for historical documents. That is why this document represents the best practice in the field of Optical Character Recognition.

There are several projects and tools in existence for collaborative correction of OCR results. There are a number of projects now that focus on creating or correcting transcripts by using innovative crowd sourcing methods in addition to or instead of OCR. One such project is Transcribe Bentham, a participatory project based at University College London. Its aim is to engage the public in the online transcription of original and unstudied manuscript papers written by Jeremy Bentham (1748-1832), the great philosopher and reformer.

Best Practice Example: Transcribe Bentham. London : University College London, 2010-2011; <http://www.ucl.ac.uk/transcribe-bentham/>

The project uses volunteers to transcribe manuscripts rather than attempting to use OCR. This is why we highlight this project as another successful method of avoiding errors in OCRing, allowing us to focus our efforts not on improving the technology, but on engaging the public in preserving cultural heritage. Another reason this project deserves attention is because it was highly successful in gaining media support in the United Kingdom and worldwide, which is the only guarantee for attracting public attention.

2.5 Best practice examples of access to digital content

Digital files can provide extraordinary access to library materials. We found that the majority of the EuropeanaTravel project partners provide free and open access to all the materials that are digitised⁴. They offer their digital collections on the library website or on portals designed for this purpose. One of the examples of good practice that we could focus on are portals built or co-designed by the Austrian National Library, the University of Innsbruck, the National and University Library, Ljubljana and the Moravian Library in Brno. As a best practice example, we focus on the Digital Library of Slovenia – dLib.si. This portal, built by the National and University Library of Slovenia, is one of the best, especially due to the following properties: ease of use, interactivity and all the characteristics of web 2.0. The portal is distinguished particularly by a high level of user interactivity, which is reflected in both the basic digital collections search and in support services (e.g. My dLib.si, Tags, etc.). The portal offers searching and browsing options and is available in Slovenian and English. It provides access to various types of material:

- texts: journals (articles, scientific articles), books, Slovenian Research Agency reports, academic achievements;
- photos: art library, photographs, sheet music, posters, postcards, manuscripts, maps;

⁴ Some partners use a combination of free access and restricted access within the library (mainly due to copyright legislation demands) or remote access for authorised users only.

- multimedia: virtual exhibitions, sound recordings.

Best Practise Example: Digital Library of Slovenia - dLib.si. Ljubljana : National and University Library, 2005-2011; <http://www.dlib.si>

Leading EU cultural institutions, including the *EuropeanaTravel* partner projects, offer digital contents via the Europeana multimedia library. Europeana provides direct access to digitised books, audio and video material, photographs, images, maps, manuscripts, newspapers and archival documents that are part of the European cultural heritages. Visitors to www.europeana.com have the opportunity to browse various virtual collections of European cultural institutions in their own language and all in one place. Because of this last feature, this joint portal of the EU cultural institutions deserves its place among best practice examples of access to digital content, since it offers easy access to the cultural heritage of the EU nations to every EU citizen.

Best Practice Example: European; [http:// www.europeana.com](http://www.europeana.com)

3. Managing digitisation projects

This chapter covers the field of project management in digitisation projects by focusing on the following topics: strategic planning process, designing the workflow, quality assurance and digital preservation costs. Of course there are other aspects of management involved in project management, but they are discussed in other chapters of this document (e.g. management risk, communicating, the procurement of the hardware and software infrastructure, user needs, etc.).

3.1 Best practice examples of the strategic documentation on digitisation

The success of the digitization project will depend upon the effectiveness of the planning (The Whole Digital Library Handbook, p. 313). Comprehensive digitisation and the construction of a digital library also include previously prepared tactical and long-term plans in the form of documents such as the digital library development strategy, the priority list for digitisation, a description of the quality control procedures in the library and other document(s) concerning the digital library, digitisation etc., which can take the form of individual documents or can be combined together in a single document.

Most documents of this type at the disposal of the EuropeanaTravel project partners are only available in their national languages; we were therefore only able to examine the documents that are available in English and were recognised as the best during the project implementation. One such document was prepared in the National and University Library of Slovenia and it is available also in English.

Best Practice Example: Digital Library of Slovenia Development Strategy - dLib.si 2007-2010/. - Ljubljana : National and University Library, 2007; http://www.dlib.si/v2/documents/pdf/strategy_dks.pdf

The purpose of the Digital Library of Slovenia Development Strategy – dLib.si – for the period from 2007 to 2010 is to project the way to compile a comprehensive digital library of Slovenia. With this document, the National and University Library responded to European Union initiatives, as well as to the new and changed requirements of users and the rapidly advancing technology that is providing us with a new and modern way of offering library services. The main result of the strategy is a modern and comprehensive digital library that is user-friendly and easy to use, while also enabling the permanent preservation of the most priceless cultural heritage collections and making them available to the wider public. The main topics discussed in the strategy are:

- the construction of a comprehensive digital library,
- establishing an organisational structure for the support of the Digital Library of Slovenia,
- improved quality of services,
- the preparation of a national plan for the digitisation of library material,
- the digitisation of national collections,
- offering services through the dLib.si web portal.

The document also contains a schedule for the implementation of all development activities required for the realisation of the strategy.

The value of the Slovenian strategy is represented by the holistic and systemic view presented, so it deserves a place among other best practice examples.

3.2 Best practice examples of workflow management

At this point, you should create a detailed workflow manual for all the team members and any sub-contractors. These manuals will contain clear instructions that will help staff in their responsibilities, ease the introduction of new staff and act as a basis for your quality control. It is likely that you will find you need to amend and update these manuals throughout your project to take account of lessons learned through experience (JISC Digital Media, 2008).

One of the characteristics of a good digitisation workflow is that it is documented in clear steps. We found this example of a digitisation workflow containing a few steps in the Trinity College Library, Dublin.

Best Practise Example: Digital Resources and Imaging Services; Digital Imaging Standards Policies V.6, Dublin : The University of Dublin, June 2009.

The following workflows are used by Trinity College Library, Dublin:

- Document Selection
 - Material Review / Project Scheduling - DRIS
 - IT Impact Assessment
 - Conservation Review
- Metadata Prep
 - Project naming
 - Library catalogue review
 - Metadata research
- Scanning
 - Material transport from the library to the DRIS scanning lab
 - Material Preparation
 - Scanning
 - Document Return to the library
 - Data moved from the scanner PC to processing computers
- Processing
 - Project naming and file folder organization
 - Image processing
 - LAB colour space conversion
 - Apply sharpening (unsharp mask)
 - Convert to final colour space (Adobe 1998)
 - Apply metadata
 - De-skew/crop
 - Save final master Tiff

- Surrogate generation
 - High Resolution full size JPG
 - Apply visible TCD watermark
 - Screen viewable small JPG
 - Apply visible TCD watermark
- Data backup
 - Departmental Optical backup
 - Tiff Archival master
 - Hi Res JPG
 - Low Res JPG
 - Archival Optical Backup
 - Tiff Archival master
- Metadata
 - Generate project headings and structure in the Image Repository
 - Upload Images Into the Repository
 - Input Metadata
 - Select QA Images (randomly)
 - Approve Project (via auto generated email)
- Project Close
 - Input archival backup into the library catalogue
 - Move archival media into the library storage
 - Web Page Update
 - Communications
 - Clean data from the scanning and processing computers.

Naturally, every project manager who creates a detailed workflow faces the following question: “Should we be doing the digitisation work in-house or should we be outsourcing?” The document containing the most information on the advantages and shortcomings of outsourcing and in-house digitising is available in the JISC Digital Media advice.

Best Practise Example: To Oursource or to Digitise In-house?, JISC Digital Media, 11. November 2008; <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/to-outsource-or-to-digitise-in-house>

This document is a good starting point for anyone trying to decide whether to divide the labour among team members, involve any sub-contractors or take a hybrid approach, dividing the labour and combining the benefits of internal and external resources.

3.3 Best practice examples of quality assurance

Quality Assurance is a key part of any digitisation workflow. The basic principles of ensuring quality control procedures for digitisation purposes are comprehensively presented in the S. Gray article entitled “Quality assurance and digitisation projects”. This article is unique in that it provides a systematic and comprehensive overview of this indispensable part of the efficient management of library content digitisation.

Best Practise Example: Stephen Gray, Quality assurance and digitisation projects. Published in: Managing a project / Digitising analogue media, JISC, Digital Media, 14 November 2008, <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/quality-assurance-and-digitisation-projects/>

The quality assurance procedures should be established during the planning stages and implemented throughout the entire duration of the project.

The quality of the digital objects acquired in the digitisation process can only be defined with regard to its intended use. A digital object that is suitable for a certain purpose is not necessarily also suitable for other purposes. There are a number of issues influencing the choice of standards for the digitisation project. One of the main issues is whether the project aims to create a perfect replica of the original or merely to represent and convey the informational content of the original. Another important question is how the digital object will be accessed. It is therefore impossible to set a single general standard to ensure the highest or most acceptable quality. Considering that each individual digitisation project will have its own unique goal, the standards of quality assurance in the digitisation processes will have to be set in accordance with the goal of each project. While it is true that, in the digitisation of similar collections, similar

standards of quality assurance are used, these standards only apply to most of the projects, never to all of them.

The accepted criteria and standards of quality should be determined, quantified and unanimously accepted in the planning stage. They should be based on an assessment of the user needs and the initial test of the work progress. The accepted criteria must be documented and included in the project specifications.

There are a number of factors that can influence the quality of digital objects, such as:

- the condition of the original
- the equipment used in the digitisation processes
- the skill of the operator
- resolution
- the post-digitisation procedures (the optimisation or re-mastering of digital objects)
- the choice of format or the compression algorithms used.

Quality standards are often the result of a compromise between the available options to ensure a high level of quality and reduced expenses. For example, a higher resolution will result in a larger file, requiring more storage space and a longer period of time for data transfer. Similarly, quality checks of individual digital objects will give better results than batch checks of content samples, but it will also result in more expenses.

S. Gray (2008) on quality assurance: »Quality assurance (QA) can be considered an attitude to work, rather than an external testing system. Everyone involved in the project needs to take responsibility for ensuring quality at all times. QA should be pervasive and can usefully be considered in four-layers:

1. Process QA :

Errors in the digitisation process are usually out of the digitisation operator's control and should be addressed by the head of the project. Errors in the process usually relate to one process or several processes documented within the project.

2. Automated QA :

The system for quality assurance must be as objective as possible. The digitisation equipment must be calibrated in accordance with the specified

standards and the workflow should be automated wherever possible (for example: some metadata can be automatically collected from the management system.)

3. Personal Checking :

Despite well-planned quality assurance processes and regardless of the amount of automated processes, the results must still be checked before they are handed in.

4. User fault reporting :

Once the work has been handed in, it is considered good practice to provide users with a way to report any faults to the digital collection manager so they can be amended.

The most extensive and well documented system for quality assurance in the digitisation processes was developed at Trinity College Dublin and their document “Digital Resources and Imaging Services” encompasses every aspect of digitisation, from descriptions of the image structure, file structure and calibration to metadata, the processing of copies and the workflow. They have a variety of quality control functions integrated within the digitization workflow to ensure consistent, efficient and error free work processes. They have a dedicated quality assurance program (based loosely on the six sigma process), which reside completely outside of the digitization and metadata cataloguing workflows. This quality assurance activity includes random audits on a statically significant portion of each imaging project and measures image quality attributes, file and system structure errors, metadata accuracy and completeness, and a variety of other relevant attributes. Audit failure mediation not only includes file error correction but also workflow and process reviews and corrective actions⁵.

Best Practise Example: Digital Resources and Imaging Services; Digital Imaging Standards Policies V.6, The University of Dublin, June 2009.

⁵ More on the quality assurance in D2.1 deliverable of the project Europeana Travel: Infrastructure for quality control, Europeana Travel, [Internal document], eContentplus, July 2009.

3.4 Best practice examples of cost modelling for digital preservation

This part covers best practices in the field of digital content with regard to life cycle costing. A number of experts believe that preservation is digitisation's greatest weakness. The complexity of preservation treatment can also be seen as an economic activity.

With regard to this, there is a general principle that needs to be considered from a long-term preservation perspective:

“The potential costs of ongoing storage and preservation need to be balanced against the potential costs of re-digitising content. While it may seem absurd to plan a large-scale digitisation project on the understanding that the process may have to be repeated at some point in the future, the uncertain costs of long-term preservation, as well as ongoing improvements in digitisation techniques, may mean that this will ultimately be the most efficient or cost effective approach.” (IMPACT Decision Support Tools, p. 114)

In past years, a great deal of attention was directed towards the economics of digital preservation. Several international projects and a Task Force all developed various cost models aiming “to develop a set of economically viable recommendations on the adoption of digital preservation strategies.” (IMPACT Decision Support Tools, p. 115)

The best practice example of cost modelling for digital preservation is LIFE, a three-phase UK project (LIFE, LIFE2 and LIFE3, Life Cycle Information for e-Literature)⁶. The project has attempted to model digital preservation costs over five, ten and one hundred years. Long-term modelling of digital preservation costs for cultural heritage materials and improved forecasting and control of the lifecycles of digital collections are the features that place the findings of this project among best practice examples. LIFE (which has just completed its third phase) delivers a predictive costing tool that will significantly improve the ability of organisations to plan and manage the preservation of

⁶ LIFE (Life Cycle Information for E-Literature) is a funded by the Joint Information Systems Committee (JISC) and Research Information Network (RIN) and is a collaboration between University College London (UCL) and the British Library.

digital content. The tool is currently available in two trial versions – as a web application and an Excel model. Another advantage of this tool is that it offers an insight into the big picture, since it takes into account the widest range of contextual factors (e.g. content profile, organizational profile, inflation, staff cost, hardware costs, etc.) that could influence long-term costs.

The LIFE model and tool make an important contribution to approaching this subject by providing cost estimates for the lifecycle of digital collections. LIFE provides strategic dimensions and a practical way to determine the cost of digital lifecycles. “Knowing the relative costs is essential for choosing the correct repository and preservation system, where the future financial consequences of mistakes can be serious.” (Rosenthal D.S.H. in Hole *et al.*, 2010).

Best Practise Example: LIFE Spreadsheet and online tool in beta phase available for evaluation; <http://www.life.ac.uk/blog/2010/06/11/life3-model-beta-available-for-evaluation>

4. User Focus

4.1 Best practice examples of user-orientated digital content and user evaluation studies

When talking about the best practices in the fields of digital content creation and access to digital resources, we must mention the users of digital material. Moreover, the best practices in all the previously mentioned fields are only the best if they offer direct or indirect quality services to the end user. In connection with this, J. Penka (in: *The Whole Digital Library Handbook*, 2007, p. 308) writes: "It is by understanding and focusing on the users needs (...), rather than simply adopting the latest technology, that libraries can look holistically at their reference offerings and build adaptable, goal oriented systems. It is critical to define the target audience and understand the context and conditions of those using digital reference services."

Research among the EuropeanaTravel project partners has shown that even when starting a digitisation project, cultural institutions are already considering the end users. They deal particularly intensively with promoting new digital content to (potential) users.

We also discovered the best example of including users in all stages of digitisation in one of our partners. The National Library of Wales's Digitisation Strategy 2008/09 - 2010/11 document shows that their entire process of digitisation is strongly user-oriented. It is clear from the document that digital services and collections exist to meet the needs of everyone who can benefit from them. Moreover, the digital services and collections that are the subject of their digital strategy are not only aimed at existing users, but also at reaching other potential users. The document contains a selection of progressive ways to establish interaction with users (particularly through the use of Web 2.0 elements). It presents examples and ways in which the library supports the educational and academic community in a digital environment and with digital collections and is an active partner in the learning process. The National Library of Wales digital strategy anticipates the active participation of users in the selection and formation of digital collections, both in selecting material and in the final stage, where users evaluate the digital collections.

Best Practice Examples: Digitisation Policy and Strategy
The National Library of Wales's digitisation policy and strategy, 2009
(<http://www.llgc.org.uk/index.php?id=information>)

In addition to analysing the needs and expectations of users prior to starting digitisation projects, special attention should be given to user feedback and the evaluation of the existing offer of digital collections. When searching for the best examples of the user evaluation of a digital collection, JISC Digital Media⁷ directed us to the NINCH Guide to Good Practice⁸, the Assessment of the Project through User Evaluation, which presents various timeless methods for the evaluation of digital collections: computer logging of user interaction, electronic questionnaires, observing and tracking, interviewing and focus group discussions.

The contribution poses a large number of questions that must be answered with the help of evaluation, e.g.: Who uses the digital collections?, What are the formal and informal learning outcomes of the digital collections and services?, Are they satisfied with it?, Is the resource easy to navigate? Does it attract users?, etc.

Best Practice Example: NINCH Guide of Good Practice, Assessment of Project by User Evaluation, Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, and the National Initiative for a Networked Cultural Heritage (NINCH); <http://www.nyu.edu/its/humanities/ninchguide/XII/>

⁷ JISC Digital Media exists to help the UK's FE and HE communities embrace and maximise the use of digital media and to achieve solutions that are innovative, practical and cost effective. <http://www.jiscdigitalmedia.ac.uk/>

⁸ The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials, by the Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, and the National Initiative for a Networked Cultural Heritage (NINCH)

The significance of this contribution also lies in the fact that it contains detailed arguments about why evaluation is an extremely valuable component of digitization work. Only through the evaluation by our users can we find out how the digital resources we create are actually being used. Research and evaluation will help indicate how we can enhance the functionality and usability of the digital resources we are creating.

If user focus is the rightful focus of digital libraries, then we should highlight the latest work on user perspective. This document is an IFLA publication, combining the theoretical and practical views of various experts from different parts of the world, as well as the findings of EU projects such as ATHENA, CERL, DC-Net, DPE, MICHAEL, TEL and Europeana.

Best Practice Examples: DIGITAL library futures : user perspectives and institutional strategies / edited by Ingeborg Verheul, Anna Maria Tammaro and Steve Witt. - Berlin : De Gruyter Saur, cop. 2010. - (IFLA publications, ISSN 0344-6891 ; 146)

Knowledge of this publication is key to compiling any digital library or digital collection, since it is based on three ideas: a) technology is not enough; b) we need cooperation with users; c) we need international cooperation with cultural institutions and partnership with others (publishers, et.al) (Digital Library Futures, p. 11).

Knowledge of this work and its basic findings is recommended to anyone working in the field of digital libraries, since it provides an insight into the key elements of user perspective: user experience and what users want and how they use it. The message of this publication is clear: the only way a digital library can ever be successful is if it puts the user in the centre of its development and operation. At the same time the publication also contains the following fundamental truth about the user experience: "If we want to focus on the users, we need to focus on the behaviour of information and on the interface design."

Last but not least, another reason why this publication deserves attention is that it also deals with the success rate of meeting the users' digital needs in the light of a new partnership between different cultural institutions (libraries, museums and archives), as well as the public and private sector, in particular partnerships with publishers.

5. Summary

These best practice examples have been written to give an overview of the points that must be considered when planning and carrying out digitisation. The document presents examples of best practice in the following fields: capturing and managing images, handling the material, metadata generation, OCRing, access to the digital content, strategic planning, workflow management, quality assurance, cost modelling for digital preservation, user-orientated digital content and user evaluation.

Because of the differences in planning different types of project, these best practices examples could be a checklist for some people. But for others it could be more of a list of points to be aware and suggestions how to deal with these points once they become important for the project.

It is important that these examples of best practices are taken as a starting point for further discussion rather than as an attempt to speak the last words on the subject.

This document mainly takes into account documents in English. It is the author's wish that this document would encourage people with valuable documents concerning digitisation and digital libraries that are only available in their national languages to publish these documents in English, thus making them available in the common European sphere of exchanging knowledge and ideas.

““Best practice” is not always achievable, although there are bottom lines and “good” or “better” practices.” (Hargreaves, 2008)

References

Annex 1, **Description of Work, ECP-2008-DILI-538002, EuropeanaTravel**. [Internal document]. eContentplus. Version 27702/2009.

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. <http://brtf.sdsc.edu>, (last accessed 2010-12-11).

Care and Handling guidelines for the Digitisation of Library Materials http://www.nla.gov.au/digital/care_handling.html, (last accessed 2010-10-11).

Digital Capture Equipment, National Library of Australia; <http://www.nla.gov.au/digital/capturedevice.html>, (last accessed 2010-10-11).

DIGITAL library futures : user perspectives and institutional strategies / edited by Ingeborg Verheul, Anna Maria Tammaro and Steve Witt. - Berlin : De Gruyter Saur, cop. 2010. - (IFLA publications, ISSN 0344-6891 ; 146).

Digital Library of Slovenia Development Strategy - dLib.si 2007-2010/. - Ljubljana : National and University Library, 2007; http://www.dlib.si/v2/documents/pdf/strategy_dks.pdf, (last accessed 2010-10-11).

Digitisation Policy and Strategy. The National Library of Wales's digitisation policy and strategy, 2009, (<http://www.llgc.org.uk/index.php?id=information>), (last accessed 2010-09-01).

Digital Resources and Imaging Services. Digital Imaging Standards Policies V.6. Trinity College Library, Dublin, The University of Dublin, June 2009.

Dublin Core Metadata Initiative, <http://dublincore.org/metadata-basics>, (last accessed 2010-11-12).

Gray, Stephan. Quality assurance and digitisation projects. Published in: Managing a project / Digitising analogue media, 14 November 2008 (JISC Digital Media, <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/quality-assurance-and-digitisation-projects/>), (last accessed 2009-07-15).

Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives. IFLA, Preservation and Conservation section, March 2002. www.ifla.org.sg/VII/s19/pubs/digit-guide.pdf (last accessed 2009-06-06).

Guideline on records digitisation. United Nations, Archives and Records Management Section. <http://archives.un.org/unarms/en/unrecordsmgmt/unrecordsresources/guideline%20on%20records%20digitisation.htm>, (last accessed 2009-06-06).

Hargreaves, J. Building Digital Collections, JISC Enriching Digital Resources, November 2008. http://www.jisc.ac.uk/search.aspx?keywords=jisc+best+practice+hargreaves&collection=default_collection&type=adv (last accessed 2010-12-12).

Hole B., Lin L., McCann P. and Wheatley P. LIFE³: A Predictive Costing Tool for Digital Collections. Paper submitted for iPres September 2010. <http://www.life.ac.uk/3/documentation.shtml>, (last accessed 2011-01-06).

Icelandic National Digital Library, Policy for the retroactive digitization and preservation of digital objects, National Library of Iceland, June, 2008.

IMPACT Decision Support Tools/ OC2, Impact project, FT7, 2010.

Impact – Improving Access to Text, FT7, <http://www.impact-project.eu/about-the-project/concept/>, (last accessed 2010-07-01).

Infrastructure for quality control, EuropeanaTravel, [Internal document], Contentplus, July 2009.

JISC Digital Media, **Project Management for a Digitisation Project**, 2008. <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/project-management-for-a-digitisation-project/>, (last accessed 2010-11-11).

JISC Digital datasets, 2009.

<http://www.jisc.ac.uk/whatwedo/programmes/digitisation/scopingstudy/digitaldatasets>, (last accessed 2010-11-11).

Korb, Joachim. Survey of the availability of digitised images for OCR, TELplus project (ECP-2006_DILI-510003), Deliverable D 1.1, 24 January 2008.

Korb, Joachim. Survey of existing OCR practices and recommendations for more efficient work, TELplus project (ECP-2006_DILI-510003), Deliverable D 1.2, 31 July 2008.

LIFE Spreadsheet and online tool in beta phase available for evaluation; see <http://www.life.ac.uk/blog/2010/06/11/life3-model-beta-available-for-evaluation>, (last accessed 2011-01-05).

NINCH Guide of Good Practice, **Assessment of Project by User Evaluation, Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, and the National Initiative for a Networked Cultural Heritage (NINCH)**, 2002; <http://www.nyu.edu/its/humanities/ninchguide/XII/>, (last accessed 2010-01-16).

Quality Standards for EOD, Version 2.2. EOD Project, Culture Program. (22/07/2009), <http://books2ebooks.eu>, (last accessed 2011-01-16).

Park, Jung-Ran. Metadata quality in digital repositories: a survey of the current state of the art, Cataloguing & Classification Quarterly, 47/2009, 3-4, 213-228.

Price Tags of Digital Preservation Policy Choices, Conference report, The Hague, 16. September 2010.

Reerink, Henriëtte. Practice, organisation and quality control of digitization projects, *Liber Quarterly*, 13/2003, 154-163.

Report on existing standards applied by European Museums. Athena project WP3, <http://www.athenaeurope.org/index.php?en/1/home>, (last accessed 2011-02-16).

Guidelines for digitising library material; Instructions for handling library material during the process of digitisation. National and University Library. Ljubljana. (http://www.nuk.uni-lj.si/dokumenti/2010/pdf/smernice_za_digitalizacija_koncna.pdf), (last accessed 2010-09-01).

Specification for quality control (Version 1.1). United States Government Printing Office (GPO), March 2006. www.gpoaccess.gov/about/reports/qc-spec-v1-1.pdf, (last accessed 2009-06-06).

Specification for the Europeana Semantic Elements. V 3.1, 25/02/2009., http://dev.europeana.eu/provide_content.php (last accessed 2009-06-06).

Survey of existing OCR practices and recommendations for more efficient work, TELPlus project, <http://www.theeuropeanlibrary.org/portal/organisation/cooperation/telplus/documents/TELplusD1%20Final.pdf>, (last accessed 2010-11-11).

Technical Guidelines for Digital Cultural Content Creation Programmes. Version 2.0, September 2008. (Minerva Knowledge Base. Digitising Content Together). <http://www.minervaeurope.org/interoperability/technicalguidelines.htm>, (last accessed 2009-06-06).

To Outsource or to Digitise In-house?. JISC Digital Media, 11. November 2008; <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/to-outsource-or-to-digitise-in-house>, (last accessed 2011-02-11).

Transcribe Bentham. University College London. 2010-2011. <http://www.ucl.ac.uk/transcribe-bentham/>, (last accessed 2011-03-15).

The European Library Handbook, section: Working with The European Library, http://www.theeuropeanlibrary.org/portal/organisation/handbook/handbook_en.html, (last accessed 2010-09-27).

The Whole Digital Library Handbook, edited by Diane Kresh for the Council on Library and Information Resources, American Library Association, Chicago, 2007.

Verheusen, Astrid. **Mass digitisation by libraries: issues concerning organisation, quality and efficiency,** *Liber Quarterly* 18/2008, 1, 28-38.