



Lifelong
Learning
Programme

the
Leonardo
da Vinci
project

2013-2015

'Training the trainers'

Training equipment
for teaching in
digitisation

also
known
as

*Sofia's
Toolbox*





Sofia's toolbox means to support trainers who teach (new) staff in memory institutions working on the field of digitisation. It consists of introductory texts, exercises and experiences.

The toolbox is a product of the project Training the trainers. Dissemination of knowledge on digitisation. The presented experiences come from the partners in this project.

*Some of the topics in the toolbox are:
selection and prioritisation,
project management
preparation
data collection and management*

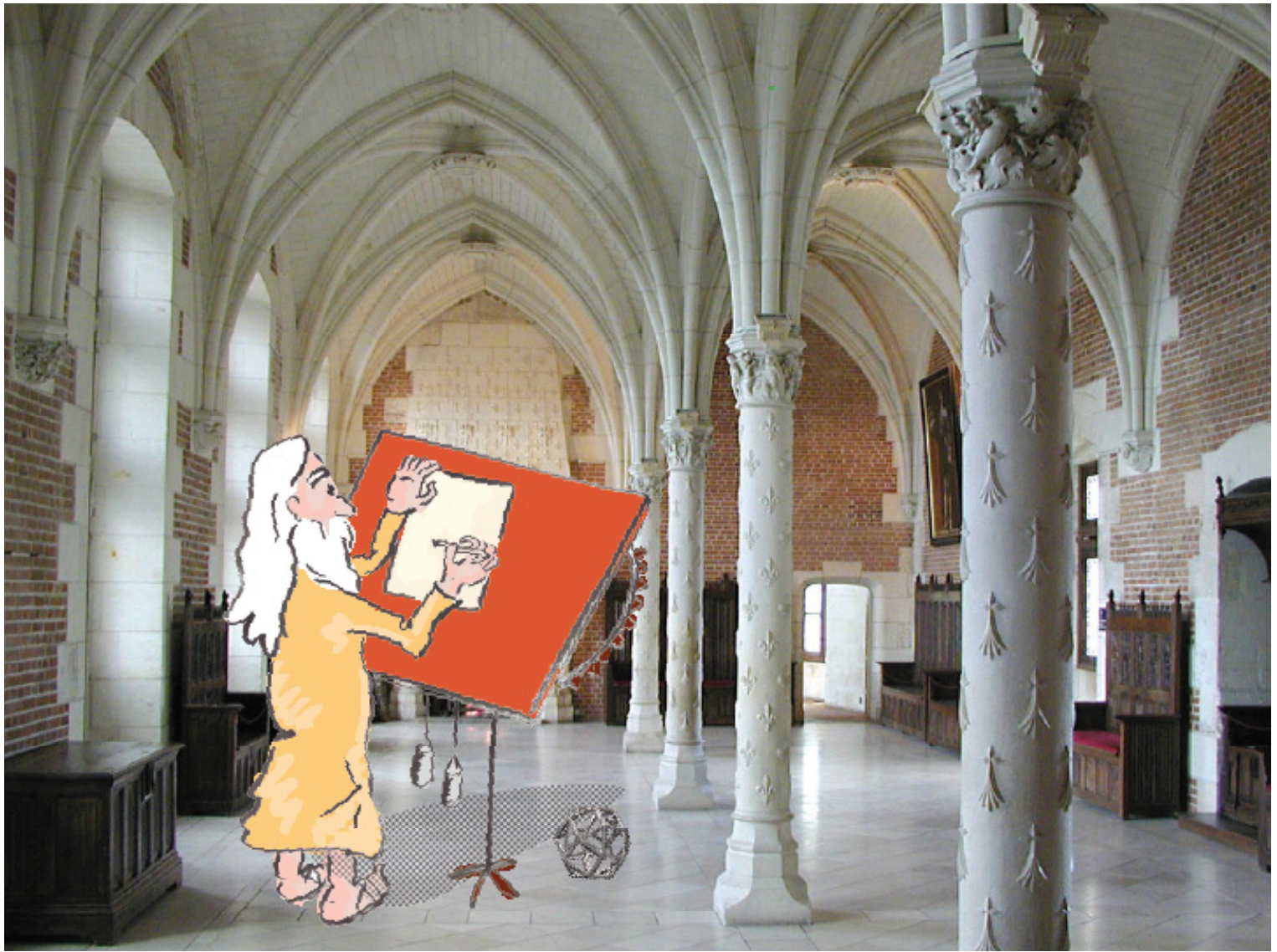


Contents

Chapter 1 <i>Introduction</i>	5
Chapter 2 <i>Selection</i>	9
Chapter 3 <i>Project management</i>	17
Chapter 4 <i>Preparation</i>	23
Chapter 5 <i>Data collection using oral history as a method</i>	29
Chapter 6 <i>Digitisation (capture and editing)</i>	37
Chapter 7 <i>Metadata</i>	65
Chapter 8 <i>Providing access</i>	71

Partners in this project are:

University of Borås, Swedish School of Library and Information Science (Sweden)
Budapest History Museum (Hungary)
Centre for Image Research and Diffusion, Girona City Council (Spain)
Municipal Archives of Gävle (Sweden)
Municipal Archives of Schiedam (Netherlands)

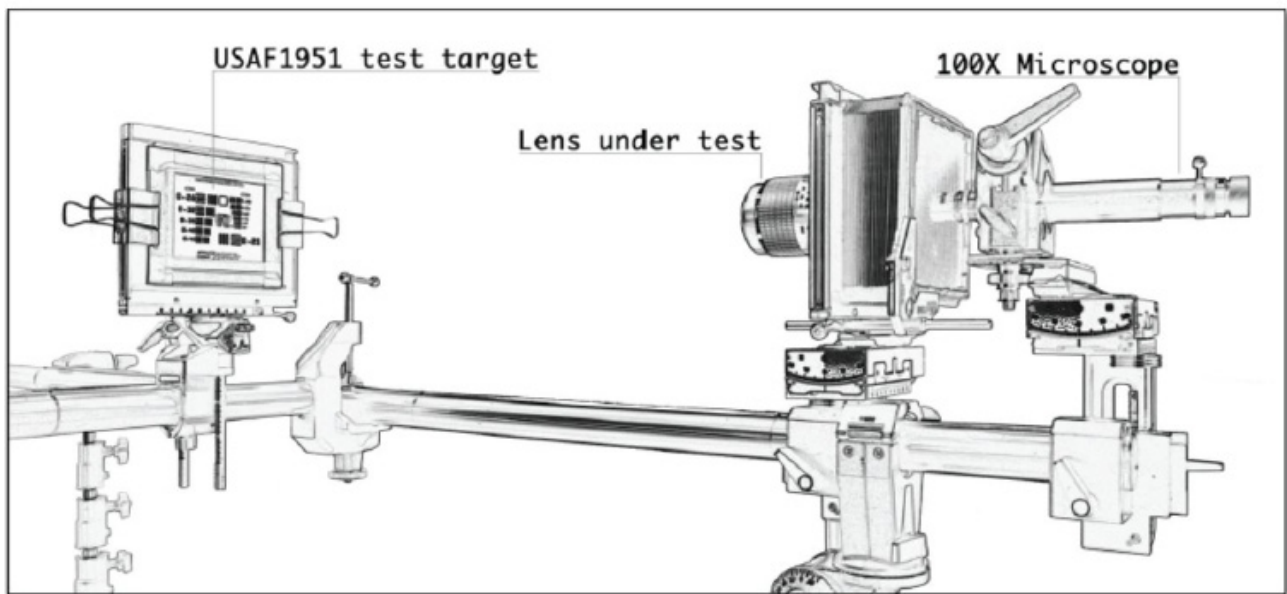


1515, Italy

Chapter 1. Introduction

A pedagogical toolbox

This toolbox means to support trainers who teach staff in memory institutions working in the field of digitisation. It consists of introductory texts, exercises and experiences. It provides the basic content for understanding the digitisation of the cultural heritage with the aim of being helpful to professionals in charge of digitisation projects and those who need to train employees without previous experience in digitisation. For this purpose, we have collected and made available a set of concepts that come from the most relevant practice papers existing currently, in order to provide useful insight on being trained in digitisation. The toolbox is a product of the Leonardo da Vinci European project: **Training the trainers. Dissemination of knowledge on digitisation**. The purpose of this project is to provide pedagogical resources for trainers on how to manage and teach digitisation. The target groups are the staff responsible for the training of co-workers and those who need to transfer this knowledge. It is also targeting professionals working in archives, libraries, and museums who need to acquire practical knowledge on digitisation, as well as new staff and volunteers responsible for digitisation and those who need to acquire practical knowledge. Lastly, it is aimed for teachers in the field of digitisation of cultural heritage who need to transfer and acquire this knowledge.



The guidebook is structured in 8 different chapters. Each chapter contains an introductory text to explain the scope of the chapter and sometimes some explanations regarding the main concepts. There is also a section with specific instructions for trainers on how to deal with exercises proposed for trainees followed by a section dedicated to specific experiences related to the chapter subject that can be helpful to better understand the content. The last part is the resources list which includes an online bibliography and web resources related to digitisation. In the text below we provide a brief summary of each chapter.

- 1. Introduction.** Chapter 1 (this chapter) is an introductory text to define the scope of this toolbox and to specify the target group and to explain how to deal with it. It also includes a definition of digitisation, to better understand the approach of this toolbox.
- 2. Selection.** Chapter 2 aims to establish some selection criteria when digitising with the idea that in the selection process, institutions of all sizes and types aim to concentrate on parts of their collections that are best suited to digitisation as cultural institutions cannot afford to preserve or to digitise every object in their repositories.
- 3. Project management.** Chapter 3 is focused on project planning and describes a suggested layout for the work process behind a digitisation project.
- 4. Preparation.** Chapter 4 is focused on how to prepare originals in a digitisation process. You can learn about decisions to make when the originals are damaged, dirty, etc. or how to facilitate the handling of originals if the digitisation itself might be harmful.

5. **Data collection using oral history as a method.** Chapter 5 is a special chapter, as it is dedicated to Oral History. Even if it is not a digitisation technique itself, it is linked to many digitisation projects.
6. **Digitisation (capture and editing).** Chapter 6 aims to provide an overview of the general concepts concerning digitisation as well as the best procedures to obtain good results in terms of quality and fidelity to the original records. It contains an introductory text in which the basic concepts to understand digitisation are explained. To achieve good results, it is absolutely necessary to understand these concepts and the characteristics related to digital objects.
7. **Metadata.** Chapter 7 is focused on documentary and technical metadata concerning video files. It is a very specific chapter, as it is mainly focused on the experience of Schiedam Municipal Archive.
8. **Providing access.** Chapter 8 aims to provide some strategies to make use of material from a collection. It is mainly focused on XML and EAD standards.

What is digitisation?

When it comes to digitisation, it is important to start with the definition of the term. We use the same word to refer to the mechanical conversion of a physical object to digital information and to refer to the process of transformation of the cultural heritage in digital form. In this process, technological, methodological, and organizational issues are involved.



Borås (2015)

Digitisation includes a large number of activities, many of which are unavoidable. The definition of the U.S. Federal Agencies is exhaustive and precise: “Digitisation is a complete process that broadly includes: selection, assessment, prioritization, project management and tracking, preparation of originals for digitisation, metadata collection and creation, digitising, quality management, data collection and management, submission of digital resources to delivery systems and into a repository environment, and assessment and evaluation of the digitisation effort “[USA, Federal Agencies Digitisation Guidelines Initiative (FADGI) 2009]. Having this definition as a premise, we develop this toolbox in order to train the trainers in the digitisation of the cultural heritage.

Among the activities mentioned above, we can group those that correspond to the preparatory phase, those corresponding to the digitisation process, and those involving metadata management. All these activities must be taken into account for the project management.

Preparation. The starting point for digitising has to do with preparative works related to the originals, the equipment, and the environment. It is necessary to provide a list of objects to digitise, grouped by container (usually boxes), to prepare materials according to their physical characteristics, and to assign a registration number to the original materials. It is also important to consider planning the outflow of records from stores, the needs for cleaning and restoring, and to previously organize materials physically and intellectually.

Digitisation. For digitising we must consider the device performance tests, the calibration of capture devices, the provision of materials to the scanner or camera, the capture, the file’s name, the file’s encryption, the assignation of formats and the compression scheme, the file processing, and the quality control. It is important to define the digitisation parameters based on the characteristics of the original material and to define the post capture adjustments needed (cropping, sharpening, colour balance, etc.). The recommended parameters depend on the characteristics of the originals. A good reference for different kind of records can be found in the paper Technical Guidelines for Digitising Cultural heritage Materials: Creation of Raster Image Master Files. FADGI, 2010. We should also decide which

interim storage to use (external hard drive, DVD, etc.) for the capture process, considering that the final destination is a repository.

Metadata management. Cataloguing works are focused on the record's contents and also on the physical objects. A key aspect of the cataloguing process, is the software for managing a digital collection. We should also consider technical metadata, as the digitisation process generates metadata about file formats, file sizes, equipment used, bit depth, etc. Those metadata are important for the management of the repository and for the quality control process.

Besides technical issues, there are many questions that arise when starting a digitisation project. They are not necessarily strictly related to technical aspects and it is convenient to have a clear idea of why we digitise a specific part of the cultural heritage. So, every single digitisation project should start with a series of questions like: is there a real demand from users? Who will use these records? How will people access the records? Will they access them through the web, an intranet, or social channel? Are there different ways of accessing different levels of quality? Do you need specific software to access more advanced presentations? Are there any risks associated with the access? Are there limitations on copyright or image rights? (etc.). It is very convenient to ask these questions beforehand because a digitisation project is an important strategic factor for any cultural institution and cannot be reduced to a technical issue.



1515, still Italy.

Chapter 2. Selection

Contents

1. *Introduction*
2. *Instructions for trainers*
- 2.1. *Determining selection criteria*
- 2.2. *Selection based on selection criteria*
3. *Exercises*
- 3.1. *Exercise 1: Define selection criteria*
- 3.2. *Exercise 2: How to deal with special types of originals?*
4. *Experiences*
- 4.1. *Experience 1. Digitising within the Framework of the Public Works Programme at Museum Aquincum*
- 4.2. *Experience 2. Digitising the Poster Collection of Museum Kiscell*
- 4.3. *Experience 3. Digitising the Architectural Collection of Museum Kiscell*
- 4.4. *Learning results based on the experiences*
5. *Resources*

1. *Introduction*

Selection for digitisation, shapes the online collections by cultural heritage institutions. In the selection process, institutions of all sizes and types, aim to concentrate on parts of their collections that are best suited to digitisation and also meet their visitors' needs. Most cultural institutions cannot afford to preserve or to digitise every object in their repositories. Some items are not worth preserving or digitising, whether they are texts, photographs, sound recordings, or any other genre.

Proper selection decisions are essential in planning a digitisation project. The first step of every digitisation project is planning. The more time you spend planning your project, the more effective your work will be done. The decision criteria is determined by the main goals of the project to be implemented, however, the decision may be influenced by financial and technical barriers, copyright laws and regulations, specific demands from researchers and other users, political or ideological concerns, national, regional or other directives, as well as activities of similar projects in other institutions.

In the planning phase, stakeholders of the given digitisation project have to carefully assess the physical nature and content of the original materials, the intellectual property rights connected to them, and the requirements for a technically suitable and cost-effective method for digitising. Selection of materials to be digitised work best within a framework of priorities for preservation and digitisation that carefully considers scale and sustainability.

2. *Instructions for trainers*

Chapter 2 aims to draw the attention of trainees to some problems and concerns related to selecting material to be digitised. With the following exercises trainers can help to make their trainees aware of the importance of defining selection criteria and planning the process of digitisation. In each exercise a different aspect of selection is highlighted.

In the first exercise, trainees should decide about the selection criteria and think about aspects like target group, importance of accessibility, and have to make decisions about preparation of the originals. The second exercise deals with the issue of digitising special objects like "3D" museum objects which also require a different approach to digitisation.

First, trainees should read the short introductory text below about the main criteria related to the issue of selection. For exercise 1, it's recommended to work in two groups where group 1 is elaborating case 1 and the other group works on case 2. After the work is done, the two groups should discuss their results together.

2.1. *Determining selection criteria*

Selection criteria should reflect the objectives of the project. It is very important to pay attention to the following minimum criteria:

- The project should focus on archival, library or museum material, which has no or limited access.
- Digitising should provide easier and wider access for high interest materials.

- Take into account the condition of the source material!
- Provide a digital copy to protect vulnerable materials!
- Obey the laws relating to copyright and intellectual property!
- Always take into account the costs of digitization!
- Examine the original material from the point of view of publishing it online!

These criteria should be cleared with the stakeholders of the digitisation project. The accepted criteria must be documented in full detail, so that they are always accessible for all participants of the project.

2.2. Selection based on selection criteria

Once you have determined the criteria for selecting the material to be digitised, we should start sorting:

- All material should be examined to ascertain whether it fulfils the criteria you specify. If there is a condition which does not correspond to the given material, stakeholders should prepare a minute about it. In cases where relevant materials remain outside criteria specified before digitisation, stakeholders need to reconsider the criteria again and we need to define new ones.
- A detailed description of the record produced from materials selected for digitisation should be recorded in a central database. In this phase of the project, we collect each item of the material selected for digitising for the first time together. This is a good opportunity to create a database that contains every aspect of characterization. This will help to ensure the successful implementation of the project and the appropriate handling of original materials or to determine the original location of the materials to be digitised.

3. Exercises

3.1. Exercise 1: Define selection criteria

Case study:

Invoices from the Municipal Executive Board of Gävle, year 1860-1946. The invoices are bound in book format and contain economic transactions of the municipality of Gävle and therefore they are of significant relevance for the history of the municipality.

1. Analyse the originals before planning a digitisation project!

- What is the historical importance of the books?
- Who might be interested in researching the invoices?
- What kind of data is extractable from the historical source?

2. Make a plan for digitising the originals!

- What is the purpose of the digitisation? Is preservation the main reason for digitising or is the main goal to give access to a wider public?
- Who is the main target group of publishing the digitised records? Professional researchers, interested non-professionals, or others?
- Decide what to publish on the internet! Will you publish scanned originals or data extracted from the originals imported into a database?

3. Choose a method for digitising! Define selection criteria!

Case 1: You aim to digitise all the originals without exception and published them online.

- The invoices are bound in large format books where some parts of the relevant information aren't visible because of the subsequent binding. What kind of actions do you have to take in order to properly scan all the sheets?
- As you have to rely on the help of a restorer in the preparatory phase of scanning, you should give him/her instructions about how to prepare the originals for scanning. (see Chapter 4 of Sofia's toolbox)
- Find a method for digitising which doesn't affect the condition of the originals!
- How do you publish the digital data after scanning? Make a decision about extracting metadata from the digital records! Decide whether you are using OCR for a complete extraction of information or not! If not, select the types of metadata you want to publish! (see Chapter 7 of Sofia's toolbox)

Case 2: You don't want to take restoration actions and aim to publish only the information that is contained in the books.

- Make a selection of the information you want to record in a database to be published!
- Make a plan for extracting the metadata from the originals! Make a decision about the three W's: who, when and what!

Discuss choosing from the above mentioned methods. What are the advantages and disadvantages of each method? Do you have similar experiences with other types of originals?

3.2. Exercise 2: How to deal with special types of originals?

Case study: digitising a Textile Collection (for a good example see: <http://www.clothestellstories.com/>, especially: <http://www.clothestellstories.com/index.php/displaying-clothes/digital-display-kongedragter>)

Textile objects are very sensitive to any damages during manipulation as they are very light-sensitive. Therefore, objects of this kind should be digitised — that means mostly photographed — only one time! Selection, preparation, and planning is thus an essential phase in digitising a textile collection.

See the example photos: 1. Woman's dress, ca. 1908; 2. Ceremonial attire, worn by Hungarian noblemen. You can work with the example photo or choose an item from your own collection!

- Define what you are considering as a “record”: a full dress with accessories, the full dress only, or one part of a garment!
 - Inspect the condition of the object to be digitised! The presence of a restorer is necessary for digitising textile objects.
 - Proper installation is essential in planning the digitisation of a textile collection. Decide how you are planning to digitise a full dress:
 - Make a list of tools and accessories needed for the process!
 - Do you digitise each part separately or the full dress in one, or both?
 - How do you install each part? Do you use a dress-stand?
 - Specify from which positions you want to digitise a full dress (e.g. front view, rear view, side view etc.)!
- Which details are required to be digitized to document the object properly?

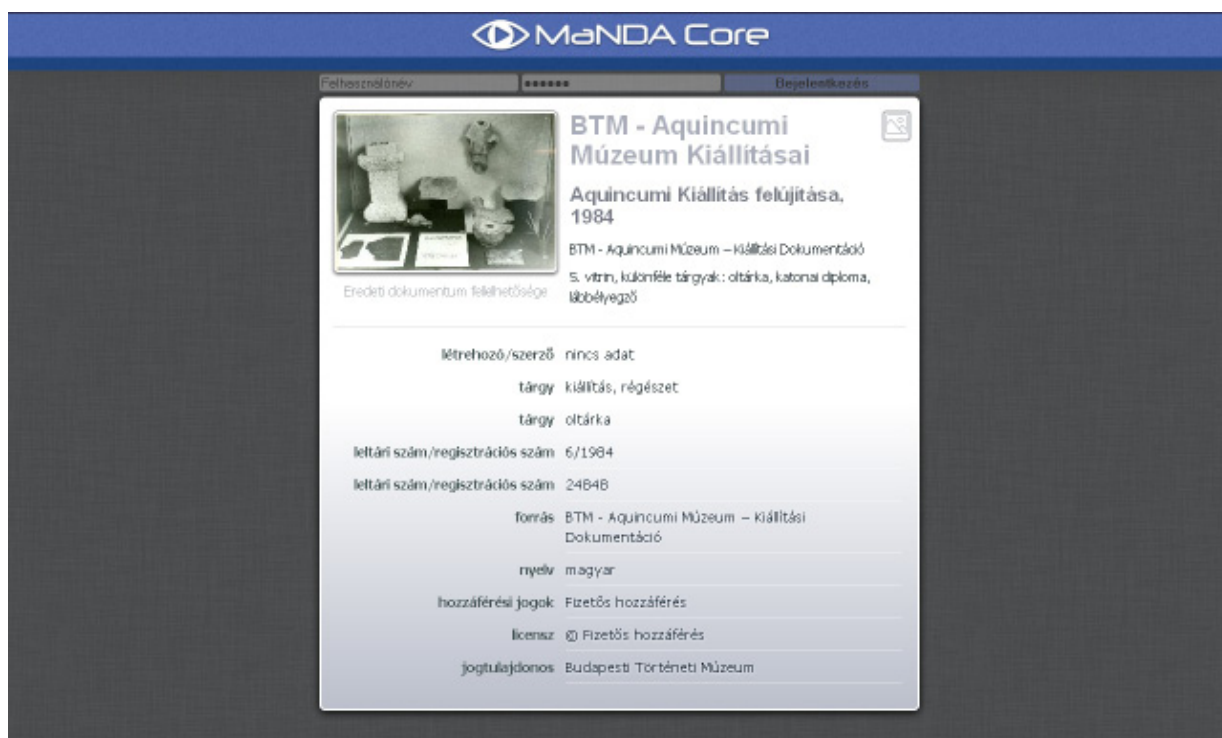
4. Experiences

4.1. Experience 1. Digitising within the Framework of the Public Works Pro- gramme at Museum Aquincum

4.1.1. Introduction

The comprehensive digitisation programme of Budapest History Museum was started in December of 2014 with the signing of an agreement with the Hungarian National Digital Archive (MANDA).

Digitisation work within the Public Works Programme had to be planned and organized within a relatively short



time. At the first stage of planning we asked and answered the following questions:

- What should be the task to be performed?
- Where and how is this task going to be carried out?
- What kind of technical equipment is required?
- How to schedule the project? What is the amount of digitised material we can expect?

Before selecting the material to be digitised, we had to consider many aspects of the structure of the museum's collections, but also the tasks related to publishing the digitised data.

It was also important in the planning phase to select a digitisation staff with appropriate skills and qualifications. Historical and archaeological knowledge and experiences were essential criteria for choosing participants of the Public Works Programme in Digitising Cultural Heritage.

4.1.2. The main aspects considered in the planning phase

- The purpose of the project, defining realistic goals
- Give priority to material with restricted accessibility for the public in the selection
- Make an accurate estimation of the costs of the digitisation project
- Obey the regulations relating to copyright and intellectual products
- Prepare a work plan where different workflows are separately defined
- Make regulations for each working process. For all of the separate processes there should be a person with specific responsibilities
- Always consider the condition of the material to be digitised (time needed for preparation, cleaning, digitising, and in some cases also for repackaging should all be taken into account)
- If possible, classify the original material from the aspect of both the needs of researchers and ordinary visitors in terms of publishing the digital material online:
 - Preparation, cleaning
 - Selection (classification – settings)
 - Before digitisation you should provide a suitable environment for both the material to be digitised and for the staff digitising it. Staff should be supported with the appropriate hardware and software
- Make plans for releasing the originals, for processing time, and also for transporting the originals back to the repositories
- Naming, storing, and uploading files
- Saving, storing files on servers or other types of data media
- Define criteria of digitisation based on the specific needs of different originals (f. e. photo, glass negative, painting)
- Always choose a digitising method which doesn't risk the condition of the original
- If possible, digitise a document or an object once, in the highest possible resolution – you can process and compress this digital original according to your needs
- Always take care of keeping and saving digital records
- Introduce a digitisation workbook
- Document the process of digitisation

4.1.3. Evaluation of the project

Based on state resources launched in the framework of the Public Works Programme, the digitisation project done in Aquincum Museum was a viable and extremely efficient job from the institutional point of view, because the basic objective could be completed at a minimum cost (equipment, scanner procurement, providing space for digitisation). Within one year the amount of digitised cultural heritage objects added up to 20.000 pieces. Digitisation was also connected to a partial revision of selected collections. The result of the project is a digitised material which is also suitable for use in research from outside and the digital data meets both the needs of museum staff and visitors' demands. The project was also beneficial for participants of the Public Works Programme because all staff working within the framework of this programme found a job before the final completion of the project.

4.2. Experience 2. Digitising the Poster Collection of Museum Kiscell

4.2.1. Introduction

During 2008, the collection of over 5000 posters in Museum Kiscell was fully digitised. The posters' dimensions vary between 30x47 and 95x126 cm, and their digitisation was performed by using a professional scanner of 1118 mm. (Contactor: Arcanum Database Publishing House. Sponsor: The National Fund for Culture)

4.2.2. Problems and requirements of scanning the posters

The scanner lens, which is extremely fragile, must be constantly observed and cleansed. The lens quickly collects dust and this affects the quality of the digital image. Scanning as a method for digitising is very well suited for all 2-dimensional items (photos, posters, documents, architectural plans, etc.) In case of deteriorated items (posters), the scanning process requires the cooperation of 2-3 individuals inside a room of at least 20m². In every case the employment of a protective light filtering foil is required, as the strong scanner light can harm the poster paper.

4.2.3. Digitisation of deteriorated/ fragmented posters

It is possible to scan deteriorated posters, conditional on their degree of fragmentation. The process depends on the size of the deteriorated areas: the broken pieces will be set on pieces of carton (1-3 mm thick) of the same dimensions as the poster, and then will be covered by protective light filtering foil of the same size. This assemblage is then inserted in the scanner for digitisation.

4.2.4. Learning results

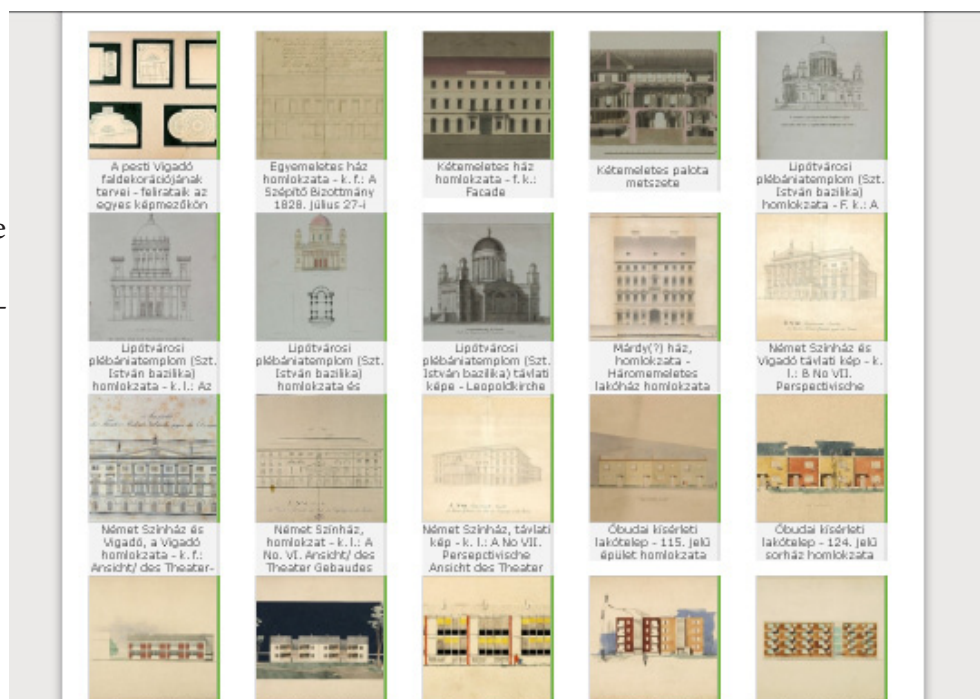
The above-mentioned method is applicable for deteriorated posters. The quality of the scanned image depends on the precision with which the broken poster pieces are set on the carton. In cases when the size of the poster does not allow scanning, the digitisation will be performed by taking photos. However, in this case, the quality of the image will not match the one obtained through scanning.

4.3. Experience 3. Digitising the Architectural Collection of Museum Kiscell

4.3.1. Introduction

The architectural collection includes several workswith the largest unit of the collection being a nearly 35,000-piece collection of architectural plans with relevance to Budapest. The earliest pieces are from the end of the 18th century, but the largest part of the collection is originated from the 19th century and contains plans of the State Opera, the Matthias Church, the Fisherman's Bastion, and the Buda Royal Palace. Significant elements of the collection are design collections of the most important 19-20th century Hungarian architects' legacies. The Architectural

Collection is one of the most researched collections of the museum, art historians and architects working on renovations of historical buildings rely on information extractable from historical plans. Therefore, it's very important to have a digitised collection, not only from the point of view of accessibility, but also of preservation. That's why we've



decided to digitise the whole collection.

4.3.2. *Selecting the material*

The first question to answer before we started digitising was, whether we are selecting material for digitisation in order of importance or whether we are following the order of location in the repositories (objects are topographically ordered on each shelf). At the beginning of the project we digitised the plans in order of importance. This required constant coordination between the curators and the staff working on digitisation. In the beginning, many aspects had to be discussed in detail. As the scanning routine was established, scanning staff worked more and more autonomously, even more quickly, so we decided to change the procedure and follow the order of the repositories. The digitisation process went more effectively this way, allowing the digitising staff to work smoothly even without the guidance of the curators.

4.3.3. *Restore or preserve?*

Although the optimum preference would be if all the objects could be digitised in a restored state, we don't have the resources to follow this intent consistently. Objects in poor condition have to be restored and in cases where preservation is sufficient, we select this method. In addition, there might be cases, where the object itself is in good condition, but a smaller part of it is contaminated. Because of this specific reason we've purchased a cleaning eraser and we have learned how to work with it so that the digitising staff can perform the cleaning as well.

A similarly small treatment of the object is straightening the folding lines on plans. Scanning is distorted if the paper has sharp folding lines, but that does not necessarily require restoration interventions. We have learned how to straighten the lines with the help of a bone knife or a similar device.

4.3.4. *Scanning*

In our case, digitisation serves registry purposes and for that a medium-quality image would be suitable. In the museum, digital images are often ordered by researchers for publication purposes, so it is important to obtain with a one-time scanning (and thus without moving the original more than once) a high definition image. In addition, we had to constantly monitor the colour correction during scanning. An end result is not satisfactory if we get a yellowish or reddish tone on the digital image, therefore, we had to continually monitor the colour settings and often needed to re-calibrate the scanners. To constantly have noise-free scans, we had to properly maintain the scanner. Cleaning could be performed by scanning staff, but we periodically ordered the cleaning services of the provider of the scanners.

4.3.5. *Record management and processing*

Before starting digitisation we created a database (in excel format) for internal use. However, there were objects with missing data or incomplete information. The appropriate method for scanning in this case would have been if record keeping and scanning could run parallel, but this method requires at least two people working on digitisation. For us this was not possible due to financial barriers and the lack of suitable space. Thus, the same person scans the items and after scanning he/she enters the missing data by analysing the digital image. The digital object scanned in the original size and resolution contains the same image information as the original object.

4.3.6. *Selecting the technical characteristics*

We have chosen to record digital images in a form which can be easily converted in any database in the future. Thus, the relevant data is collected in excel. The scanned images of 300 or 400 ppi are in tiff format. Before we started scanning we had to choose the resolution of the output image (300, 400 or even 600 ppi). The higher resolution gives a better image quality, but it also means slower scanning, and it's difficult to open these files on normal and often outdated desktop computers used in the museum. Large files also require a large amount of storage space in our servers. Experience has shown that larger format architectural plans are very rarely required for special uses like giant posters or other large prints, so we have decided that the scanned images should have a resolution of 300 or 400 ppi. In case we receive such a request (i. e. someone wants to use an enlarged image for decoration purposes), we re-scan the original.

4.3.7. *Storing images*

We store digital images in TIFF format, 300 ppi resolution. In case of large-scale architectural plans, normal desktop computers have trouble opening these files, so from each unit within the collection we created a reduced-size jpg image with batch conversion, which is easier to handle. This image still shows visible details of the original, like in-

scriptions, architectural drawings sizing etc. The on-line published images are converted into a lower resolution jpg.

4.4. Learning results based on the experiences

- Planning a digitisation project and defining selection criteria is a crucial moment of the whole process
- One of the most important factors of selecting material for digitisation is to give priority to material with restricted accessibility for the public in the selection
- Always keep in mind that the results of a project should facilitate accessibility and also serve preservation of cultural heritage

5. Resources

Selection for Digitisation: Factors to consider. Link:

http://www.carli.illinois.edu/sites/files/digital_collections/documentation/Selection-of-materials_20120221.pdf

Digitisation and Archives. Canadian Council of Archives Homepage. Link:

http://www.cdncouncilarchives.ca/digitization_en.pdf

Selection Policy for Digitisation Projects of Dartmouth College Library. Link:

<http://www.dartmouth.edu/~library/digital/about/policies/selection.html?mswitch-redir=classic>

Homepage of the North Carolina Digital Heritage Centre. Link:

<https://www.digitalnc.org/about/participate/select/>

Analysing Selection for Digitisation. By Bart Ooghe and Dries Moreels. In: D-Lib Magazine, September/October 2009, Volume 15 Number 9/10. Link:

<http://www.dlib.org/dlib/september09/ooghe/09ooghe.html>

Guidance for selecting materials for digitisation. By Paul Ayris. Link:

http://eprints.ucl.ac.uk/492/1/paul_ayris3.pdf

Task Force to establish Selection Criteria of Analogue and Digital Audio Contents for Transfer to Data Formats for Preservation Purposes. By Majella Breen, Gila Flam, et al (Ed.).

International Association of Sound and Audio-visual Archives Homepage. Link:

<http://www.iasa-web.org/task-force>

Guidelines for Digitisation Projects for collections and holdings in the public domain, particularly those held by libraries and archives, March 2002. The International Federation of Library Associations and Institutions (IFLA) homepage. Link:

<http://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines.pdf>

Selecting Traditional Library Materials for Digitisation. Report of the CUL Task Force on Digitization, Cornell University. Link:

<https://www.library.cornell.edu/colldev/digitalselection.html>

Strategies for Implementing a Mass Digitisation Program. By Erik Moore. In: Practical Technology for Archives, Issue 3. Link:

http://practicaltechnologyforarchives.org/issue3_moore/



2015, Hungary...

Chapter 3. Project management

Contents

- 1.1 Introduction
- 1.2 Project vocabulary
- 2. Instructions for trainers / Project process
 - 2.1. Overall description of the work process for a digitisation project
 - 2.2 The Project Pyramid
 - 2.3 Value Analysis
- 3. Exercise: Practicing Project Management, A Case Study
- 4. Resources

1.1 Introduction to Project Management

The focus of this chapter is on project planning and describes a suggested layout for the work process behind a digitisation project. Most (but not all) digitisation happens within the confines of a project, sometimes with internal and sometimes external funding. Nonetheless, understanding the project form and how to manage it becomes vital for those partaking in it.

1.2 Project vocabulary

Project Manager

The project manager's role is to oversee and handle the project, from initial brief to closing the project. As such, the trainer is the natural party to fill this role in a digitisation project. A project manager reports directly to the sponsor at pre-determined times and at all milestones.

Sponsor

In the private sector this is often a client, but in these types of projects, the sponsor is most often the individual creating the project brief and bringing the budget to the project. As such, the sponsor can be either an individual or an organisation, depending on the project.

Milestone

A milestone is a check-point that defines an important and measurable project event, such as a stage objective. The milestone is results or criteria that are defined in advance which can be fulfilled at a certain time. The purpose of the milestones are to be a support in planning, follow-up and management of the project, where the project manager report to their sponsor, who in turn decide on whether the project shall continue as planned or change direction.

Roles

It is important to set up clear roles where it is clearly stated who are responsible for each task. This does not necessarily mean that the specific person must carry out the activity, but that they are responsible in seeing to that the work is completed.

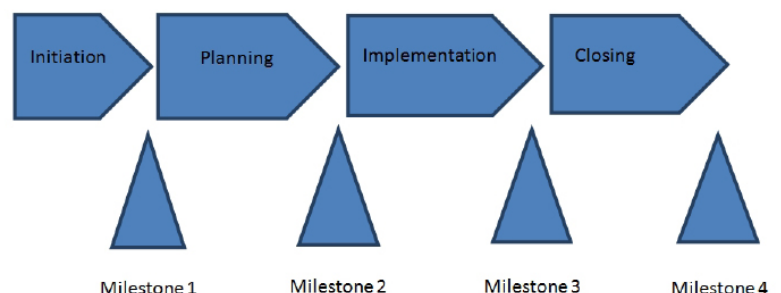
2. Instructions for trainers / Project process

2.1. Overall description of the work process for a digitisation project

The following presents the work process, in which we also highlight the role and responsibility of the project manager (most often the trainer). The description of the process should be developed to a more complete instruction later on, directed towards project managers and participants in a digitisation project where documents can be linked or be available as attachments. It is very important that information and instructions concerning the process are carefully and clearly coordinated.

Project stages

Initiation of a digitisation project entails that a private individual, archive, or other organisation



prepares a project brief. This brief should include an initial description of the background, context and comprehensive objective for the project, where central resources, methodology, and an overall time frame is included. In order for the project to succeed, it is important that the formulated objective is clear, realistic, and has measurable goals. The sponsors, project managers, and other participants must all agree on these goals. The project objectives must be formulated so that they are easy to measure and understand. Initially, the project's main objective is formulated and is then broken down into smaller partial objectives. At this point, it is important that the project managers communicate with the project sponsor and coordinate so that all parties are in agreement on the expected outcome of the project.

Planning entails that the project manager/sponsor draws up a project plan for carrying out the digitisation project. The plan should include purpose, contents, scope, and a more detailed methodology. One of the most important project management activities is to draw up a financial budget for the project. The project manager is responsible to monitoring and controlling the budget, and needs to identify all cost factors into each of the project activities. This type of project planning establishes the foundation of the entire project.

Because of this, it is important to invest a good amount of time on the planning of the different steps of the project and distribute the available resources to those activities that must be carried out. In order to create effective project management, it is important to have clear objectives that all participants interpret the same way and to have a continuous follow-up of the objectives. This is where the project managers have a vital task that requires good communication skills, and if possible, it is an advantage if the project participants can be a part of the planning. This can result in motivated co-workers, which in turn makes implementing the plan easier.

Creating a well-structured project plan requires that it is broken down into smaller activities, where each activity is a task with a specified result that is carried out during the project. An activity has expected time duration and an expected cost and resource requirement. For example, it could be acquiring resources, manning the project, following-up and managing changes, maintaining running communication, and evaluating the project. Another important part of planning is establishing a schedule or time-table that includes all activities.

Implementation entails that the project participants together with the project manager carry out research, data collection, and digitising etc. according to the project plan. As in all projects, it is possible that plans change due to unexpected events. This is why the project plan must be a living document, able to be changed in the event something within the project pyramid (see section 2.2) changes. This could be anything, from dwindling resources, to a change in the time frame given. Here it is important that the project manager takes responsibility and continuously follows up the plan and is able to discuss any conversions with concerned parties.

Closing entails that the digitisation project is finalised and reported to the project sponsor, who signs off that the project has been closed and finished.

2.2 The Project Pyramid

Every project must, by necessity, work within the confines of three distinct and mutually exclusive definitions – those of Time, Cost and Quality. This trinity of definitions is commonly known as the project pyramid, because through these definitions, the project takes shape. Each project must, before commencing, decide upon the ramifications that are acceptable so as to give the project manager a grasp of how to prioritize. However, it is also a feedback instrument, allowing the project manager to inform the sponsor of potential upcoming problems with keeping within the confines, and also acts as a guidance tool in the absence of the sponsor.

2.2.1 Definitions of Time, Cost and Quality

Time

Most often measured in how long a project can take to fulfil, time should not be confused with how many full-time resources are associated with a project. Most often, if the project is part of a larger portfolio, the deadline can either be set, or will at least affect other projects inside the portfolio if it needs to change. Time can also be a factor that influences the budget. Many economic sponsors of projects work inside certain periods (between 2010-2020 for example) and thus the deadline can be “hard” in that it defines the end of all possible resources for the project regardless.

Cost

The most fluid of descriptors, Cost can be defined by a set budget of money, but also in the form of other resources, such as number of full-time employees, what hardware can be used (perhaps even time-limited), what raw materials are available and so on. To make things easier, fluid resources are often translated into their economic value to give project managers and sponsors an easier attribute to work with, but such values are often subdivided themselves (for example the total budget can be 450'000€, but out of these, 300,000€ are bound into human resources in the form of full-time employees, leaving 150'000€ for non-HR related expenses).

Quality

The final deliverable of the project is often defined by its Quality. If the deliverable is a number, it becomes easy to define the Quality of a project. With 120,000 photos to digitise, anything above that is achieving over 100% quality (or it could mean that you achieve 100% faster or for less cost) and anything less is a lower level of quality. In a project that is supposed to deliver three parts, (cost analysis, value analysis and proposal), these can be weighted (cost analysis 55%, value analysis, 25%, proposal 20%) as a way to determine both priority but also what the minimum deliverable level of quality is (with a minimum of 55%, it becomes apparent that the cost analysis is both the priority and the minimum deliverable, while with a minimum of 20%, each part by itself is enough for a minimum deliverable, but the priority is still the cost analysis).

2.2.2 Confines

The most common method of dividing the project pyramid is through setting a level of acceptable confines for how much Time a project can absorb, how high the Cost can become and what targeted level of Quality the project should end with, but also a priority level, to help guide the manager. A small project might have a set confine of three months (Time) and a budget of 450,000€ (Cost) to achieve a complete analysis and solution to a problem (Quality). Some confines are more fluid, such as setting the timespan to be “three to six months” or denoting what is a minimum acceptable delivery (a financial cost analysis) and suggested goal delivery (complete cost and value analysis with proposed solution). Such variable confines are tools, to help the project manager work independently from the sponsor but still within the acceptable limits. To allow a project to move forward efficiently, it can therefore be useful to use variable confines – that said, not all projects have such a fluid potential.

2.2.3 Priorities

Creating a priority matrix for the project pyramid allows both the project manager and the sponsor to be guided by these priorities when making decisions. Priority matrixes are defined by setting the order between Quality, Time and Cost and applying it to all decisions. With a priority of QTC (Quality before Time, Time before Cost), it becomes apparent that if a decision leads to a higher quality, it is acceptable that it takes more time or costs more, or if a high cost can mean delivery on or before set time, that is preferable. A project where the priority is TCQ (Time before Cost, Cost before Quality) would accept a lesser project delivery to meet the three month deadline first, but could accept a higher cost too, if that was necessary to achieve the minimum delivery in time.

2.3 Value Analysis

Once you have a set of confines defined by the project pyramid, you can analyse what resources you have and make an analysis of expected outcomes. A constantly evolving value analysis is a powerful tool for highlighting problems before they stop the project delivery completely, or to course correct and achieve a higher quality. The value analysis is dependent on an existing project pyramid and associated priorities, but with that in place, the value analysis provides great insight.

The main idea behind a value analysis is to define a “value” to each transaction inside the project, having every activity defined by its smallest part. By defining what the value and associated cost is, it allows us to extrapolate if the budgeted cost, time, and quality are achievable. It also allows us to see what out of these three must change if anything.

2.3.1 Example of value analysis as a tool

Using digitisation of photos as an example, you can set a cost per digitised photo by analysing the process. The total cost (often first measured in time and then converted to €) to acquire photo originals, set them up for scanning, digitally remaster them, deliver both digital copies and return physical originals, can be divided into a cost per photo. If 1000 photographs take 3 hours to deliver from archive to digitisation area, and each hour costs 10€ in labour and 5€ in operational costs, the cost of delivery is 4.5€. If a single photograph takes 45 minutes to scan and digital remas-

ter, the cost rises to 7,545€ per photo (7,5€ for the 45 minutes of scanning plus 4,5€ for delivery). With a budget of 450,000€ and setup costs of 50,000€, that leaves 400'000€, divided by the cost per photo of 7,545€ it tells us that we will be able to roughly digitise approximately 53'000 photos which will take roughly 160 hours to deliver and about 39,750 hours to scan and digitally remaster.

If the goal is to digitise 120,000 photos inside of three months, we know that with 160 man-hours per month we will need roughly 83 people doing the digitisation full-time to meet the Time deadline (three months) and Cost limit (450,000€) but with a lowered quality (53,200 scanned photographs vs the goal of 120,000).

2.3.2 The Value of Value Analysis

An early estimation of this means that we can immediately inform the sponsor about this, so that the sponsor can make appropriate prioritised decisions. Perhaps, with the above example, it is acceptable to raise the budget to allow for more photos to be digitised, but not to change the time factor, which means that the project would have to hire more people to do more in the same timeframe – to have this information and decisions in front of the sponsor in the first week compared to the last week which would mean a world of differences.

By analysing the value and making predictions for the long term, changes can be made early in the project to course correct, at a lower cost and with a higher reliability throughout. Through constant evaluation of the value analysis, such course correction can be achieved through the entire life of the project, giving the manager, the sponsor and the team sufficient time to achieve full quality of delivery, or at least to achieve as high a quality as the priorities would allow.

The value analysis also shines a spotlight at what is cost effective. Optimizing the delivery might save 35% of the time, but when that translates to 50 hours in total, it can be worth more to optimize for a 1% time reduction in the scanning department instead, which translates into a total of 399 hours, or almost 8 times as efficient. Value analysis allows us to determine what changes are better, or lead to greater effects, thus allowing us to optimize and prioritize at the same time.

3. Exercise: Practicing Project Management, A Case Study War Children

Today's Europe has a joint history and legacy of various local, national and international conflicts. During WW II, nearly 2,000,000 children were evacuated from their homes. In Sweden alone, approximately 70,000 children were evacuated from Finland to Sweden during the period 1939-1945.

In this exercise, you are provided with a grant of 25,000 € from your sponsor and your project brief asks you to create a project with the purpose to collect, digitise, link and make accessible memories and narratives of a population movement of war children in your country. The main objective of the project is to collect oral narratives, and a partial objective is to transcribe source material for the identification of war children. An additional partial objective is to trace where these individuals live today, a third is to carry out interviews, and a fourth partial objective is to transcribe these interviews. You have six months to close the project and present your work to the sponsor.

Here are some questions to discuss:

- How would you organise your project? How should the routine/work process be designed?
- What do you think are the most important success factors in this project?
- How does a limited Cost, Time or Quality factor change how the project works?
- How does changes in the project pyramid affect the project?
- How would project co-workers be chosen?
- Collaboration with others can be beneficial in a project like this. Would you work with other actors? Who?

What are the prospects?

4. Resources

For Project Management

http://managementhelp.org/plan_dec/project/project.htm

<http://www.pmi.org/PMBOK-Guide-and-Standards.aspx>

For Project management triangle

https://en.wikipedia.org/wiki/Project_management_triangle

<https://support.office.com/en-US/article/The-project-triangle-8C892E06-D761-4D40-8E1F-17B33FDCF810>

Useful Project Management tools

<https://www.projectplace.com/> - all-in-one project management tool

<https://asana.com/> - extremely simple, free for small projects

<https://trello.com/> - more visual project management tool, also free

<https://basecamp.com/> - last but not least, a very simple but effective tool



“Sofia, we need the toolbox!”

Chapter 4. Preparation

Contents

- 1. Introduction to preparation**
 - 1.1 Make a plan for preparation and preservation**
 - 1.2 Evaluate status of the original records**
 - 1.3 Identify difficulties and risks**
 - 1.4 Consult a professional conservator-restorer**
 - 1.5 Complete with a proper conservation system**
- 2. Exercise**
 - 2.1 Exercise: Preparation Planning**
 - 2.2 Instructions for trainers**
- 3. Experiences**
 - 3.1 Introduction**

1. Introduction to preparation

This chapter will focus on how to prepare originals in a digitisation process. Here you can learn more about decisions to make when the originals are damaged, dirty, etc. or how to facilitate the handling of originals if the digitisation itself might be harmful.

1.1 Make a plan for preparation and preservation

When you start a digitisation project it is necessary to make a plan for how to handle the originals throughout the whole digitisation process, without damaging them. The condition of the originals decides the preparation and how the digitisation process will be performed.

The preparation process gives a good opportunity to evaluate the needs for conservation and restoration. You may decide that highly valuable originals should be set up by a conservator-restorer to ensure that the objects are carefully handled throughout the whole process. This time and cost will need to be factored in to the project. This also gives you a reason to consider how thorough you should be in the preparation process. How much time is reasonable to spend on each step in the preparation process? The preparation work should be efficient, but at the same time gentle and careful to the originals. In that way, the process can be as effective as possible having better quality and being more economic. A good preservation plan is realistic and practical.

How can you achieve balance between the benefits of producing a digital copy, against the risk of damaging the original during the digitisation process? For example, if you have a book with valuable content, but the binding of the book makes it difficult to digitise; is it acceptable to deconstruct the binding to make the pages easy to digitise? Does the value lie in the book as an object or in the content?

1.2 Evaluate status of the original records

It's important to get as much information as possible from the originals before we take any actions. The original should be evaluated with consideration to the condition, quality, value and extent of the material. In what condition are the records? What materials and techniques have been used? What is the historical and/or cultural significance of the originals?

In many cases, the evaluation should be made by a conservator or someone with similar experience. The conservator-restorer can evaluate the technical structure and stability of the documents, identify difficulties and risks, and find solutions for the originals. The decisions of a conservator are based on knowledge and experience.

The condition may vary in addition to the material and technique, the age of the original, and how the original has been stored and handled. The material, technique and format, can be for example:

- Paper or parchment documents such as books, manuscripts, maps, plans etc.

- Still photography such as paper photographs, nitrate negatives, acetate negatives, glass negatives, slides
- Motion pictures
- Audio recordings
- Artefacts made by stone, pottery, metal, wood, etc.
- Textiles

1.3 Identify difficulties and risks

There are some limits concerning documents and digitisation; some documents are not suitable for digitisation, they might be too fragile or have a difficult structure. Therefore, it's important to identify if there are any difficulties and risks. For example, if the originals are broken, dirty, wrinkled, rolled, folded, with tears, or attacked by microorganisms etc.

Anyhow, there might be ways to solve the problems. Sometimes we only need to adjust the records to fit the digitisation process by using tools and facilities in the digitisation such as materials to support and fix, for example transparent materials (polyester film) etc. Even if the originals are in very bad condition there are advanced methods for conservation and restoration to be used.

1.4 Consult a professional conservator-restorer

When difficulties and risks have been identified it might be necessary to take actions. There are many different techniques when it comes to restoration, it all depends on the original. A professional conservator-restorer can make great improvements to the condition of the originals. The conservator-restorer can be a specialist in different areas, such as conservation of paper, textiles, artefacts or photographs.

Depending on how much time and money is allocated to spend on the originals, there is a wide range of methods when it comes to cleaning, strengthening, repairing, flattening, disinfecting etc. Most of all these techniques need to be performed by a professional. It might be expensive, but the result is that it's easier and safer to digitise the originals and the original document will have a longer life. Besides the positive effects for the originals, another aspect is that the preparation might be necessary to ensure that the expensive scanner equipment isn't destroyed. The use of methods such as cleaning and disinfecting can give the scanner equipment a longer life.

1.5 Complete with a proper conservation system

There are some consequences of the preparation process that may need to be handled after the digitisation. Has the preparation process caused any damage to the material? For example if you had to deconstruct the book to be able to scan it properly, should it be taken care of by a bookbinder? Or does the nature of the media require any other special considerations? For example, should nitrate negatives be stored in a special way or should they be destroyed?

The preparation gives you a good opportunity to evaluate the storage of the originals. In a careful preparation process you have already planned from the beginning how the whole process will be performed. That means that you have planned how the originals will be handled and stored after the digitisation. You can easily, for example, replace acidic storage materials such as boxes and folders with materials of known quality that meet national/international standards. Many of these acid-free boxes can be adjusted to perfectly fit the size and format of the originals. There are also methods to use polyester film, which is totally transparent and not deteriorating to the original, both for preservation and to be able to touch and move the document in the digitisation process.

3. Exercise

3.1 Exercise: Preparation Planning

Invoices from the Municipal Executive Board of Gävle, year 1860 - 1946. Number of items: 333.

Contains all major and minor economical transactions for most activities in the municipality of Gävle. For example: transactions connected to roadwork, construction projects, funding, salaries of municipal employees (polices, nurses, teachers) etc.

Discuss this example (or a similar example from our own collections) with help from the questions below. How would you plan the preparation and the capturing process? What difficulties do you see, and how can these be

solved?

Evaluation of the original records

- Historic and cultural significance
- Material and technique
- Condition

Preparation

- What actions do you think are needed in the preparation of the books to make them fit the digitisation process? Different options?
- What are the difficulties? How can they be solved?
- Do you need help from a professional conservator-restorer?

Preservation

- How will the original records be preserved and stored after the digitisation?

3.1 Instructions for trainers

In this exercise you get an example from The Gävle Municipal Archive on how to plan the digitisation of large format books. The documents are handwritten originals in leather binding. If you want to, you might use similar examples from your own collections.

We call this exercise a case study, where you discuss the example and try to find different solutions. There are no definitive answers to the questions in this exercise, instead you should try to evaluate different methods and discuss advantages and disadvantages in the methods.

Here are some of the difficulties to identify, and then try to solve:

- The binding is damaged in many of the books. Will the binding be even more damaged in the digitisation process? Do you need a conservator-restorer to take care of this?
- It's sometimes difficult to read the text in the middle of the books: it runs down the binding. Do you need to deconstruct the books before digitising? If you decide to do that, how will the originals be stored after the digitisation? Do you need a book binder to restore the binding?
- Many pages in the books are of different size, will this be a problem in the digitisation process?

You can also discuss different digitisation techniques and try to find out the most appropriate solution.

3. Experiences

3.1 Introduction

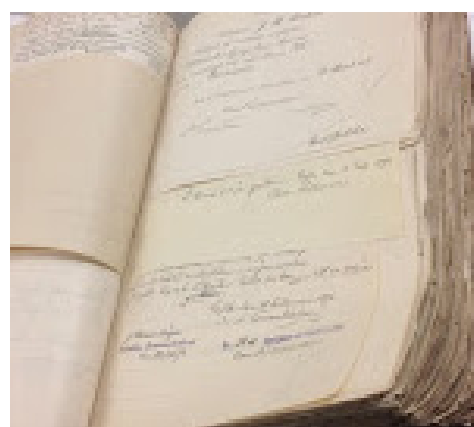
In December 2009, the Municipal Archives of Gävle started a digitisation project together with the Municipal Labour Market Unit of Gävle. The aim of the project is to digitize approximately 80,000 negatives, most of them glass negatives, from a photographic collection at the municipal archives.

The photos were taken by the female photographer Berna Roth (1885-1956). She had her own portrait studio in Gävle between the years 1911 to 1956. The photos are portraits of both inhabitants of Gävle as well as visitors from other cities and even visitors from abroad.

Berna was known to be very careful and accurate in her composition of the photos.

In the digitisation project, the Labour Market Unit is responsible for the preparation, digitisation and registration of the negatives. The glass negatives are dirty, fragile and difficult to handle. They have to be sorted, cleaned, digitised, registered, packed and stored properly.

In the project there are two trainers, Erica and Connie, responsible for how the work is performed. To make sure that the whole process is performed properly and consistently, it has been important to establish a workflow schema with



instructions for the participants. Most of the participants in this project have never worked with digitisation and preparation of photographs. That makes it even more important to have easy and clear instructions that show the different stages in the process.

In his chapter about preparation, we show the workflow document and the instructions for sorting and cleaning the glass plates. The other instructions have not been added in this chapter.

3.2 Experience 1; “Workflow in the Berna Roth Project”

1. Sorting and cleaning (Connie’s group)

Further reading in document “Instructions for sorting and cleaning”

- Start with the box with lowest number.
- If there are several glass negatives with the same number: add the letters A, B, C, D. Write the letter on the information leaf after the number; use a lead pencil.
- Clean the glass negative according to the instructions. When finished, write your initials in the column “Cleaned” on the checklist.

2. Scanning (Connie’s group)

Further reading in document “Instructions for scanning”

- Scan the glass negative
- Save the image as JPEG, in a map with a number from the checklist.
- Save in the catalogue O:\Arbetskatalogermediaforum\arb_Berna
- Put the glass negative in an acid free envelope and sign with the glass negative number. Arrange the envelopes with glass negatives in numerical order in acid free archival boxes, the lowest number first. Check the numbers with the previous box to make sure the order is correct.
- Sign “Checklist”

3. Registration (Erica’s group)

Further reading in document “Instructions for Registration in Excel”

- Write glass negative number and volume number on the “Checklist Reg” from the “Checklist” from Media.
- Produce copies of the diary.
- Register in Excel.
- Sign the “Checklist Reg” under “Registration”.

4. Inspection (Erica’s group)

Further reading in document “Instruction for Control and final inspection”

The control is performed with a contact sheet. Use “Checklist Reg”

- Enter additional information (metadata) that is not listed in the diary on the checklist.
- Sign “Checklist Reg”

5. Final inspection (The Municipal Archives)

Further reading in document “Instruction for Final inspection”.

The final inspection is made when the Excel data has been imported to CALM (archival database) and the digital image is loaded up into Calm.

- Check spelling, dates, etc.
- Add additional information (metadata)

6. Long term storage of the negatives (The Municipal Archives)

- The boxes are placed in the acclimatized archive at the municipal archive.

3.3 The document “Instructions for sorting and cleaning”

Sorting and cleaning of glass negatives

Sorting:

Always wear cotton gloves. Don't touch the surface of the glass negatives; only touch the edges of the glass.

Start with the box with the lowest number and take out all the glass plates and place on a soft surface.



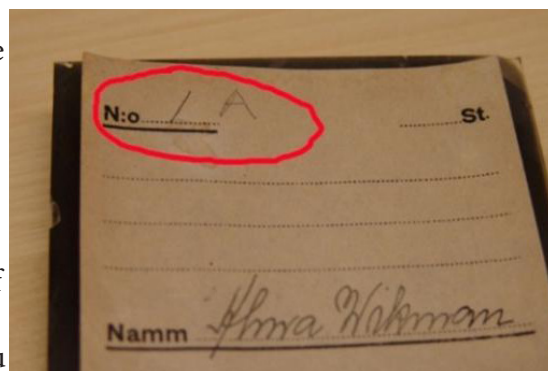
NOTE! There may be broken glass negatives; try to keep the pieces together.

Don't touch the emulsion side (matte finish side with the information sheet).

When the box is emptied, it can be discarded.

If the information sheet is detached from the glass; make sure that you put it back on the emulsion side of the glass.

If there are several glass plates with the same number: add the letters A, B, C, D etc. Write the letter on the information leaf after the number; use a lead pencil.



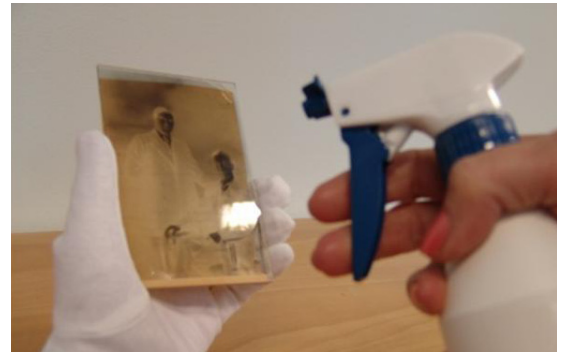
Cleaning:

Wet cleaning (molecule polish) should only be used on the glossy side of the glass. Spray the molecular polish and then wipe with a tissue. Some glass negatives are retouched with red paint; never remove the paint. You can use a tissue with molecular polish and wipe all parts except the paint.



Before you put the glass negative on the cleaned scanner, blow with compressed air on both sides of the glass negative. Keep container straight and at a 20 cm dis-

tance from the plate to avoid that the glass gets wet.



Things we use:

Pencil, Faber-Castell 8B. For labeling the information sheet attached to the glass negative.



“Scander” Molecule polish wet spray. For cleaning, dirt-repelling and antistatic treatment.



Tissues: “Durable” Dry clean. Lint-free cloths.

Compressed air: “Durable” Powerclean 400.



4. Resources

HENDRIKS, Klaus B., Thurgood Brian, Iraci Joe, Lesser Brian, Hill Greg. Fundamentals of Photograph Conservation. A Study Guide. Lugus Publications. Toronto, Canada. 1991

SHWEIDLER, M. The Restoration of Engravings, Drawings, Books and Other Works on Paper, Los Angeles, The Getty Conservation Institute, 2006.

TACÓN CLAVAÍN, J. La restauración en libros y documentos. Técnicas de intervención. Ed. Ollero y Ramos. 2009. Madrid.

NORTH EAST DOCUMENT CONSERVATION CENTER: <https://www.nedcc.org/free-resources/preservation-leaflets/overview>

UNESCO: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/digitization_guidelines_for_web.pdf

Glossary

Acid-free paper: According to international archival standards, acid-free and lignin-free paper is recommended to use when preserving documents, artwork, artefacts, etc. for long periods. The paper's pH should be neutral or basic pH (7 or slightly greater).

Acidic storage: Acidity and lignin chemicals cause paper to yellow and deteriorate over time. If archival records and cultural artefacts are stored in these kind of boxes and envelopes it will over time effect the records and artefacts; they might yellow and deteriorate. Instead, use acid free paper, boxes and envelopes (see above Acid-free paper)

Conservator/Conservator-restorer: a professional responsible for the preservation and/or restoration of artistic and cultural artefacts and documents. Conservators possess the expertise to preserve cultural heritage in a way that retains the integrity of the object, building or site, including its historical significance, context and aesthetic or visual aspects. Strategies that prevent damage, carrying out conservation treatments, and conducting research. If the conservator is also a restorer he/she has the skills to perform different kind of restoration activities such as repairing, reinforcing of materials and reintegrating missing parts of documents/objects.

Disinfection: refers to the destruction of microorganisms, particularly fungi, in this case microorganisms that have attacked historical documents, artefacts etc. Conservators use antimicrobial agents that are applied to the objects to destroy microorganisms, and through this process stop and prevent the deterioration of the historical documents and objects.

Polyester film: there are polyester films, brands like Mylar and Melinex, which are suitable for archival storage and are used by archives and museums worldwide. These polyester films are PH neutral, and do not deteriorate over time. This is because the films are free of plasticisers, catalysts and solvents and therefore ideal for protection and conservation.

Technical structure: conservators-restorers refer to technical structure of an original document or artefact, as an analysis of the physical materials that has been used. For example, what kind of paper a document consists of (fibre, technique), what kind of ink that is used on the paper, how the book binding is constructed etc.



“Have no fear, the toolbox’s here!”

Chapter 5. Data collection using oral history as a method

Contents

- 1. *Introduction*
 - 1.1 *Collection management*
 - 1.2 *Using oral history as a method*
- 2. *Instruction for trainers*
 - 2.1 *Data collection*
 - 2.2 *Oral history*
- 3. *Exercises*
 - 3.1 *Exercise: discussion*
 - 3.2 *Exercise: DPI*
 - 3.3 *Exercise: questions*
 - 3.4 *Exercises: camera*
 - 3.5 *Exercises: transcription*
 - 3.6 *Exercise: summary*
 - 3.7 *Exercises: analysis*
- 4. *Resources*

1. Introduction

In the first part of this chapter you learn more about the acquisition, selection, numbering, and scanning of documents. In the second part you learn more about Oral History. Oral history is chosen as a method of how to acquire (or even create) new material (film, documents and photo's related to each other).

If you have a theme in your contemporary history that has not been documented, you can start an Oral History project. To do an Oral History properly it consumes a lot of time and effort. A (useful) definition of Oral History is: the systematic collection of living people's testimony about their own experiences.

Oral History is always co-created: individual, informal, dialogic narratives created in the encounter between the historian and the narrator (interviewee). Oral History is not meant for fact-finding, because the memory is not a depository of information. Have in mind that memorizing is an ongoing process of elaboration and or of reconstructing of meaning (interpretation). The narrative form always contains more layers of meaning and more directions of interpretation (also beyond the conscious intention of the narrator).

Oral History is a listening art. Look at your narrator and ask open questions and encourage the narrator to go on with his/ her life story.

Make a transcription of the interview and collect photos and (ego)documents concerning the theme of your Oral History project. Make sure that the interviewee understands the aim of your project and the status of his/her interview and documents before they sign a consent form. Make your interview, transcription, photos and documents accessible (in time) for the public or for researchers.

1.1 Collection management

Cultural institutions like museums, archives and libraries are not rigid organisations; they stand in the middle of society and are in motion. Collection management is core business. Collection management contains the tasks: acquisition, preservation and sometimes restauration, numbering and scanning.

Acquisitions

Acquiring a new collection is done in different ways in the different institutions. Acquisitions can be bought or given by someone from the public, an institution, or organisation. Archives get next to these sources the biggest acquisitions through legal obligations; the civil municipal records. These kind of archives are sometimes already under guardianship, even if they are not public yet. Since the date of the (mandatory) disclosure is already known, the digitalisation can be prepared in advance. The public can get the data directly on the moment when the data become public.

Selection

Archives and museums do not always accept everything that is given to them, mostly they make a selection, before they accept an acquisition.

Archives also make a selection out of the archive that is accepted. They don't keep entire piles of papers, only the important papers will be preserved for the coming generations.

Number after selection

An acquisition has to be cleaned and preserved when it has entered the institution. The new object, book or archive gets a number, a description and a permanent place on a shelf by the employee of the institution. Then the object or archive is uploaded to the database.

Taking pictures or Scanning

A museum-employee takes pictures from the new object. The object has a number, description and name. Then it can be uploaded to the database of the museum.

When the new archive is placed in the database, another employee of the archive makes scans of the papers. The papers are scanned in Tiff-format 300dpi. Text is also scanned in PDF because pdf-documents have a search option. For example lists, address' - and yearbooks are treated this way.

Pictures as collection-items

Archives also get images as a collection. Images can be drawings, prints, paintings or pictures. The format of those pictures can also differ from glass negatives, postcards, slides and digital photos. The pictures are scanned in Tif-format 300dpi. To make the pictures less heavy and easier to work with -for instance to put on a website- they are also scanned in JPEG.

Each scan is given a unique number, and is digitally stored on a common (hard) disk. Jpegs and Tiffs have the same number. The place where it is stored has a logical structure and is easy to find; the structure and folders have names that correspond with the unique name and number of the originals. The municipality makes a mirror backup and stores it in a remote location.

Databases

Several database systems for objects or archives are in use, to mention a few: Adlib Q&A (question and answer), AB-SArcheion, Memorix, Mais Flexis, Atlantis and Adlib.

They all have their own pros and cons. unfortunately they are not easily made compatible.

Restoration

Sometimes objects, book or documents are not in a good condition, preservation is not enough and something has to be done. Some big museums or archives have a restorer in their staff, but mostly a restorer has to be sought outside the organisation.

1.2. *Using oral history as a method*

Glossary:

Oral History: the systematic collection of living people's testimonies about their own experiences

Interviewee: narrator, the person who answers the questions

Transcription: the written interview

Tripod: tool where the camera is fixed on

Oral history

Choose an important theme or period in contemporary history that has not been documented yet and of which there are not a lot of documents. Are there still some people who can talk about this period? Find as many people as possible.

Research

First you have to do your research. Oral History is not an easy job. It's not a way to relax and let someone else do the talking. Try to get as much information as you possible can. Read as much as you can about this theme and the people involved. If there's nothing about your city then read about a similar city. What do you really want to know? Oral History is not a good method to check dates and facts. It's more about experiences, emotions and thoughts. But

if there is no archive and there are no documents at all, you have no choice and Oral History is the only method to reconstruct this part of history.

Select interviewees

If you have read all the relevant information and you know everything there is to know, you can start to form the questions. Think of the information you want to get from your interviewees. What do you want to know? What is important and what is already known from literature? Select a number of people you want to interview. Choose them carefully and make sure they have different backgrounds and roles. Think of age, sex, religion, wealth, education and occupation. Make the variation between the people as big as possible. Make a small selection of 5 or 7 people. To do a scientific oral history project costs a lot of time and effort. Make sure you treat all the interviews and interviewees in the same way and give them an equal amount of time and attention.

Questions

Think of a good set of questions you want to ask every interviewee. Make a standard list of questions you want to ask. The advantage of standard questions is that at the end of the project you can compare interviewees. Think of how you pose your questions. Make open questions. Closed questions on which you can only get a “yes” or a “no” are not very useful. Oral history is meant to be storytelling, telling the story of your life.

Questions starting with a What, When, Where, Why or How are generally good questions.

The best way of telling a story is to do it chronologically. If you have enough time, start with their childhood and end with the present. It seems a like overkill to start with childhood but when you do you get an overall idea of the circumstances in which your storyteller grew up. Was his family poor or well off? Did he have an authoritarian father? Interviewing in such a way you can notice if the theme you focus on is important for your interviewee. Is it a key experience or is it just one of the things that happened in his or her life? And was it important only to some individuals or did it affect the whole city?

Make contact

If you know what focus you want in an interview and which questions you want to ask, it's time to contact your interviewees. If you ask someone to have an interview with you, tell them what your purpose is. Tell them it's a scientific project about a specific theme. That the interview is audio visual recorded and meant to be kept for long time use in a depository and made accessibly for the use of research.

Sometimes people are scared of the words “scientific” and “long time use”. Try to explain these things as simple as possible, without lying. Save the paperwork for later, because that's a sort of contract and another thing that could possibly scare people as well.

Camera

Buy a good digital camera and think of the following options. The camera is for inside use. The camera needs to be put on a tripod, because you will not be able to hold it still for so long. The camera must have a battery, but must also be able to run with an electric cable. The memory of the camera must be able to hold a film of 2 hours. Very important: be sure it has a time code!

Test

Test your camera before you are going to use it! Do not test it on your daughter on a swing in the garden, because that's not your aim. Try it on an adult sitting in a chair inside the house.

Put the camera on the tripod. Be sure there's approximately 70- 100 centimetre between interviewer and interviewee. You should be able to see upper half of the body. The interviewee can use his/her hands when he/she talks.

Transcription:

It is important to make a transcription of an interview. Having a transcription has some advantages: the written interview is searchable with any word you like and the interview can be stored in the depository in two formats: digital and on paper.

Making a transcription takes a lot of effort and is time-consuming. To write down one hour of an interview will take roughly 4 to 6 hours of work. When you transcribe, it's important that you use the time code very carefully. Write down the time code every new paragraph. If someone wants to hear a certain paragraph out of the original interview he/she needs the time code to find this paragraph.

There are different ways for transcribing an interview if you put down the meaning of what someone says it consu-

mes less time than if you put down literally every word someone says. If you choose the last method you get sentences that are not complete, that contain the same words twice or more, or silences and a lot of..uhm, uhm. People tend to speak as they think, so sentences can change in another direction halfway or never end.

This kind of transcribing gives linguistic scientist the opportunity to use your interviews for other reasons and give them a second life in another context.

Automatic Speech Recognition (ASR)

You can also try to make an alignment and use keywords and the time code. Speech recognition nowadays is not very accurate. For the moment it is not possible to make a whole readable transcription of an interview. ASR is not very time-consuming and therefore can be helpful.

Analyse deeply

If you know a lot about your theme or period then you will be able to analyse the interviews that are taken. Is your interviewee really involved? What role does your interviewee play? Is he/she a leader, a follower, or a bystander? What about the facts that your interviewee tells you? An interviewee will seldom lie; he or she will tell what he or she thinks is true. Oral History is not meant for fact-finding because the memory is not a depository of information. Have in mind that memorizing is an ongoing process of elaboration and or of (re)constructing meaning (interpretation). Portelli states: a story can be factually wrong, but psychologically true.

Photos and documents

At the end of the interview you can ask whether the storyteller has some photos or documents which have to do with the theme or period you focus on. This material can be useful in your research. It can also be useful if you plan to write a book or organize an exhibition.

The photos and documents do not have to come physically in the archive; you can make a scan and return the original to the storyteller. A decent digital version is good enough for research.

Your project consists out of several storytellers who all have their interviews on film, a transcription of the interview, photos and documents. Give the project a name and an archive number and each storyteller also a sub-number. Place it in the depository and make it accessible.

Consent

To be able to use an interview for an article, book, or exhibition, you need consent from the interviewee. In this consent you state that you or your organisation is the author of the interview. This allows you to show the interview with the portrait of the interviewee. This is the right to own ones image. It is very important that the interview is accessible. If it cannot be accessible right away, make the agreements crystal clear. It must be accessible in 10 years or after the death of the interviewee or someone involved. It's useless to hold or have an interview that is not accessible. Go to the juridical department if necessary.

For an example in Dutch of the municipal archive Schiedam, see on the website of DANS:

<https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:41097>

2. Instruction for trainers

2.1 Data collection

Here the students learn more about the core business of archives and museums: collection management. This part does not go very deep. Professionals might want to skip this part of the chapter. The exercises speak mainly about the decisions organisations makes and why.

2.2 Oral history

Not all exercises are useful for all kind of students. Depending on the level of the students, you should choose the exercises needed. Some interviewers are historians with little knowledge of technical aspects, while others might be students who have no problems with the equipment but are not familiar with recent history and the context in which the interviewees are speaking about.

If you start an Oral History-project you have to discuss what method you choose to obtain the outcome you want. Make sure that all the interviewers do their interviews in the same way. So you can compare the results and draw the

right conclusions.

3. Exercises

3.1 Exercise: discussion

Discuss what kind of archive or collection of museum-objects is useful for your organisation. Does everyone share your opinion? Why not?

3.2. Exercise: DPI

Do you always make a scan in Tiff-format 300dpi? Or are there exceptions to the rule? When do you make an exception and why?

3.3. Exercise: questions:

Take the theme or period you want to know more about and make questions for one of the interviewees. Make open questions and begin each question with the word: What, When, Where, Why or How.

3.4 Exercises: camera

3.4.1 Camera 1 discussion

Take two different cameras and 1 mobile phone. You are going to film and your colleague is going to be a human test subject. He or she is going to talk for 3 minutes. After 3 minutes you stop and take the other camera or the mobile phone.

Compare the cameras and the mobile phone. What is the best result? Which one has the best image and the best sound? Discuss whether you think it is worthwhile to save an interview made by a telephone?

3.4.2 Camera 2: distance:

Take the best camera and let your colleague or someone else be a human test subject. Let him or her talk for 10 minutes. After 3 minutes you alter the situation: (strong) light behind the person, (strong) light next to the camera, strong light from one side, strong light in the face of the interviewee. Move the person close to the camera (less than 50cm), a little further away (80 cm away) and far away (over 150 cm away).

Compare all the results. Can you also see the hands of the interviewee when he/she is talking?

What distance gives the best result? What kind of light gives the best result?

3.5. Exercises: transcription

3.5.1 Method 1

Take an interview that has not been put into transcript yet. Make a transcription of one hour of this interview and follow the first method. Write down the time code with every new paragraph. Make complete sentences; leave the silences and uhm's out. Construct it smoothly.

3.5.2 Method 2

Take the same interview and transcribe again the same hour of interview by the second method. Note every word or sound. It's not necessary that the interview only contains whole sentences. Write down also the small errors in language, write down the silences with (...) and write down all the uhm's. Make sure you still note the time code with every new paragraph.

3.5.3 Compare and discuss:

Which method was the easiest? What method do you prefer and why? Let other colleagues read your transcriptions and discuss which method is the best. Decide what method you are going to use in your project or in your institution.

3.6 Exercise: summary

Make a summary of the interview while using all the important subjects' keywords.

Note that you may need to use a keyword for a subject although this word is not used in the interview nor by the interviewer or by the storyteller.

3.7 Exercises: analysis

3.7.1 Analysis

Think about the interview you have transcribed and answer the following questions.

Are the questions clear to the interviewee? (not multiple interpretations)

Is the interviewer very talkative? Does the interviewer encourage the interviewee to go on?

The theme of the interview should be quite clear. But can you also distinguish a second story?

The interview always contains more layers of meaning and more directions of interpretation (also beyond the conscious intention of the narrator).

3.7.2. Analyse deeply

If you know a lot about your theme or period then you are able to analyse an interview. Is it true what the interviewee says? Has he/she been in the circumstance? Is it his own experience or does he talk about someone else?

What's his/her role? Can the story and facts he or she is telling be placed in the general history that is known about that theme or period? Remember what Portelli says

4. Resources

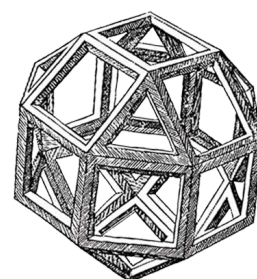
Introduction:

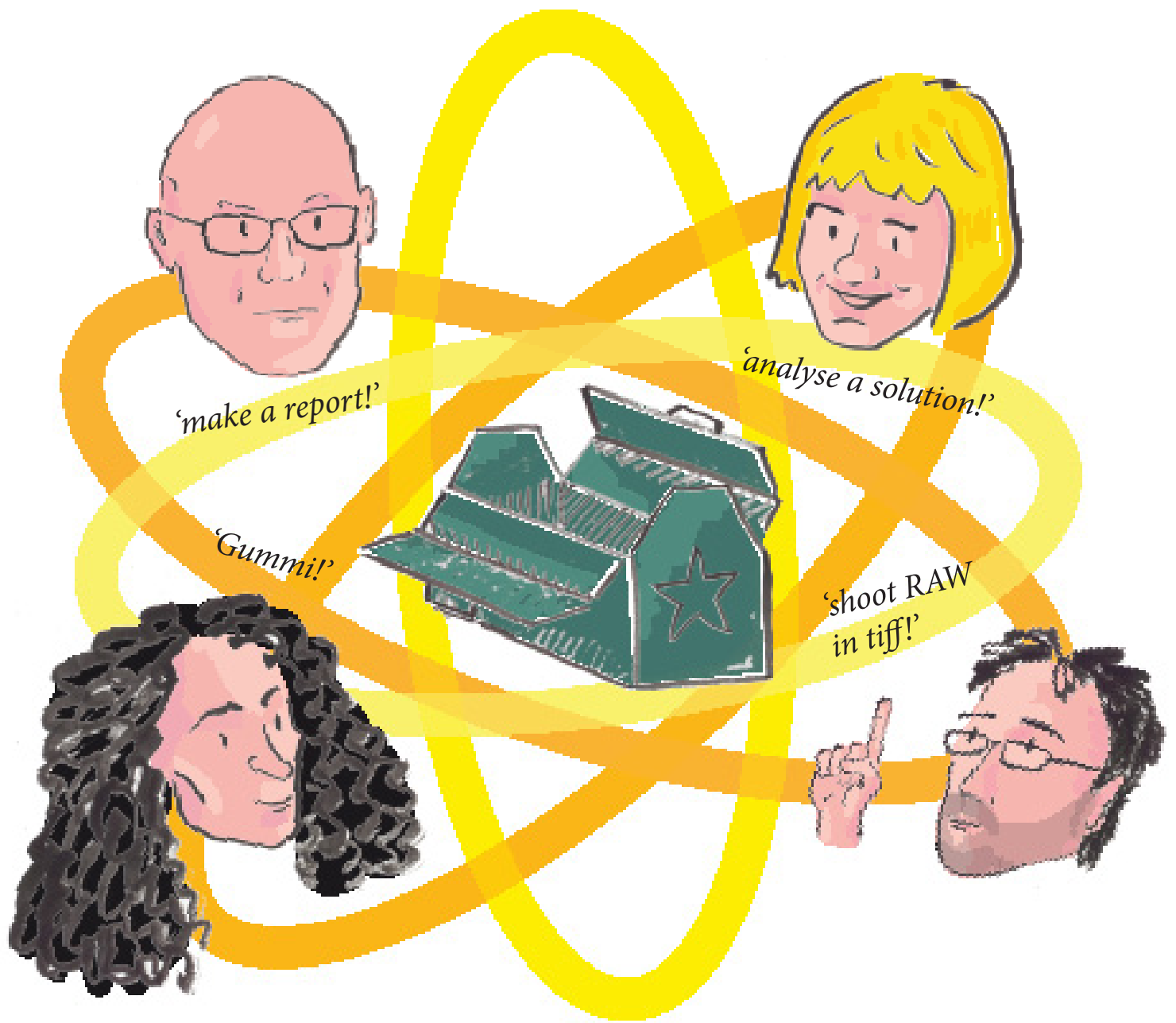
<http://www.oralhistory.org/about/principles-and-practices/>

http://dohistory.org/on_your_own/toolkit/oralHistory.html

enhanced publication:

http://www.clarin.nl/sites/default/files/Heuvel_Paper_interviews_2010.pdf





Chapter 6. Digitisation (capture and editing)

Contents

- 1. *Introduction*
- 1.1 *Image-based records*
- 1.2 *Audiovisual concepts*
- 2. *Instructions for trainers*
- 3. *Exercises*
- 3.1 *Exercise 1. Create your digitisation protocol*
- 3.2 *Exercise 2. Analysing graphic formats*
- 3.3 *Exercise 3. Analysing video formats*
- 3.4 *Exercise 4. Analysing image technical metadata*
- 3.5 *Exercise 5. Analysing video technical metadata*
- 4. *Experiences*
- 4.1 *Experience 1, on digitising photographic records: EuropeanaPhotography project*
- 4.2 *Experience 2, on digitising sound records: The Nuclear Claims Tribunal of the Republic of the Marshall Islands*
- 5. *Resources*

1. Introduction

This chapter aims to provide an overview of the general concepts concerning digitisation as well as the best procedures to obtain good results in terms of quality and fidelity to the original records.

It contains an introductory text in which the basic concepts to understand digitisation are explained. All of them are concepts concerning image-based records and audiovisual records. To achieve good results, it is absolutely necessary to understand these concepts and the characteristics related to digital objects. So, we define and explain: spatial resolution, bit depth, colour mode, colour profiles, graphic formats, audiovisual formats, etc.

The main part is the exercises. In each exercise (there are five of them) you will find specific instructions on how to organise the training and also specific instructions for trainees. Three exercises are related to image-based records and two of them are related to audiovisual records.

There is also some complementary information that may be useful to accomplish with the aim of this chapter. Section 4 is dedicated to experiences on digitisation. In this section we explain an experience concerning digitisation of photographs and another one explaining the digitisation of sound records. In section 5 we list the main resources that can be consulted for doing the exercises we propose.

Even if it is not touched-on in this chapter, it is worthy to mention that it is highly recommended to have a good technical knowledge on equipments and facilities in order to create the best conditions for digitising. It is also important to define the digitisation parameters based on the originals, to decide the metadata to be included during the capture process, and to define the post captures adjustments (cropping, sharpening, colour balance, etc.).

1.1 Image-based records

In the text bellow we explain the main concepts related to image-based records. This text may be helpful to work on the exercises prepared for chapter 6.

Adobe RGB (1998). An abstract colour space designed to reproduce in CMYK but from RGB values, with a range that can represent 50% of the CIELab space.



Digitisation equipment at CRDI facilities. Ajuntament de Girona

Artefacts

Set of alterations of the image which compromise the level of quality in whatever manner. Noise is an artefact, as is also the aliasing that occurs when the sampling rate is insufficient for the spatial frequency of the digitised image and is manifested in the presence of information that doesn't exist in the analogue original (typically pixels of colour) in the outlines of elements that form the image.

Bit depth

Number of bits used in representing a single pixel of the image (1 bit, 8 bit, 16 bit). The moment the electrical charge is transformed into numerical values, via the analogue-digital converter (ADC) in an operation known as quantisation, is when the bit-depth, i.e. the number of bits that represent each pixel, is decided. The greater number of bits, the more likely the tonal values of the image will be represented, but also the larger the volume of the final file. Thus, we can say that bit depth determines the tones that can be represented between the two extremes of minimum and maximum brightness. For example, 1-bit only allows the representation of a black and white image; 4-bit allows the representation of up to 16 tones, a number which is insufficient to display a continuous tone and it is what leads to the appearance of bands on the image. It is from 8-bit and above that the representation of the image presents a correct gradation of tones. We should always think in terms of 16-bit for images intended for permanent preservation and especially for those in which further processing will be done, such as images that will be published.

CMYK

RGB mode can be transposed into CMY, a colour system formed by the subtraction of a particular light by an object. The primary colours subtracted are cyan, magenta and yellow, complementary to the primary additive colours. It is used in printing, where black (the key colour) is added, in what is known as a CMYK model.

Colour

The generic representation of colour is produced by encoding the resulting values of the demosaicing process. Along with tone, colour is the other most important qualitative factor, since the human eye is an excellent judge of colour.

Colour management

It consists of a set of tasks for the equalisation in the interpretation of colour in order to give specific meaning to the RGB and CMYK numbers so that they maintain the consistency of the original colour. For this a colour profile was created.

Colour profile

A reference space of RGB or CMYK values with the values represented in the colour models independent of output devices, such as CIE XYZ and CIE Lab models. It can represent a specific device or an abstract colour space.

Colour temperature

It depends on the wavelengths that make up a particular light (light can have a predominance of red, green, yellow or blue, according to the predominant wave length).

Dynamic range

It is the ratio of the tonal values included between the brightest and darkest point of a scene that can be captured by a scanner or a camera.

Graphic formats

Formats are the structures that give sense to the bits that make up a particular digital resource and, consequently, allow its reproduction in a particular technological environment. Two main types can be identified: raster (bitmap) formats, represented through a matrix of pixels, and vector formats, which represent objects through geometric figures. The former are those that relate to photography.

Graphic formats structure

There are different structures for raster formats, but generally they always have a header containing metadata, and the set of bitmap data, i.e., the information for each pixel forming the image. The level of metadata can vary greatly depending on the format, but in all cases there is a minimum of information that allows the version of the format to be identified and the image to be reproduced in particular environments.

ICC profiles

Reference standards for devices. These are standard profiles created by the International Colour Consortium intended for the description of ranges of scanners, digital cameras, monitors, and printers.

Image quality

The concepts explained above are fundamental to understanding the digital image but in order to analyse the digital image in qualitative terms, we need to introduce other concepts that can help to focus our attention on the quality of the pixels. We have mentioned resolution in reference to the ability to represent detail.

JPEG

The technical specifications of the JPEG define it as “a family of algorithms for encoding-decoding for continuous

tone images with a dataflow architecture to maintain and represent the compressed data.” Therefore, this definition makes it clear to us that JPEG is not a proper format but a compression system for continuous-tone images. But under the common name of JPEG we find the formats developed by the Joint Photographic Experts Group (JPEG) to facilitate the flow of JPEG data; the exchange within a wide range of technological platforms; and applications with the aim of being highly compatible with all kinds of applications and environments. These formats include JPEG_EXIF, JFIF, SPIFF and JTIP.

JPEG characteristics

- Representation of gray-scale and true-colour images.
- In the standard compression it transforms RGB values into YCbCr. In the case of gray-scale it uses only the Y component.
- There are four modes: sequential JPEG (Baseline), progressive JPEG, hierarchical JPEG, lossless JPEG (un-compressed).
- Compression is based on the DCT and entails losses of information. Applications allow the applied compression ratio to be determined, which can range from 12:1 to 100:1.
- It is not compatible with high depth, 16-bit, with the exception of JPEG with lossless compression.
- The metadata which the JFIF format includes is minimal and nothing in comparison to the high performance of the JPEG-EXIF. The technical metadata is included in the APP0 frameworks. Additional frameworks are ignored by applications that don't recognise them and, therefore, those applications can also interpret the format and display the image. This circumstance makes it a highly compatible format.
- ISO 10918-1 Standard.
- The technical specifications are published at <http://www.jpeg.org/public/jfif.pdf>

JPEG 2000

The format was developed by the Joint Photographic Experts Group (JPEG) in order to overcome the weaknesses of the JPEG format and bring new features to the image files. The first draft of the first specification was published in 1999 and, since then, the 12 parts which constitute the JPEG2000 family, characterised by its encoding, have been developed. The values that the format possesses have caused some heritage institutions to back its adoption to the detriment of other formats with a longer tradition in the heritage field, especially TIFF. Its implementation is slow and it doesn't have the predominance that it has in other fields, such as in medical images. Although it is compatible with most image editing software, it has compatibility problems with Web browsers.

JPEG2000 characteristics

- Representation of bitonal, gray-scale, palette-colour and true-colour images.
- Wavelet compression (DWT - Discrete Wavelet Transform), with or without losses (optional).
- Compatible with high depth, 16-bit.
- It allows the storage of multiple resolutions in a single file.
- It allows the creation of different resolutions in specific areas of the image, a function much appreciated for medical images.
- Processed slowly due to its complexity. For this reason it has not been adopted en masse by the digital camera industry.
- In low and medium resolution, the quality is superior to JPEG.
- At high resolutions it avoids the effect of blurs, but on the other hand loses definition.
- The standard comprises 12 parts. The parts of the format applicable to the photographic image are: JP2_FF, JPEG 2000 Part 1; JPX_FF, JPEG 2000 Part 2; JPM_FF, JPEG 2000 Part 6
- The structure of the JPEG 2000 family is always based on the JP2 (Baseline) which is extensible. The basic component of the structure is known as the box which contains different types of fields at the same time. The boxes may contain image data or metadata.
- There is a minimum of required metadata, such as basic image data and the information relating to colour. However, it's an extensible format that can accommodate metadata in XML and different schemas such as the IPTC and the Dublin Core, amongst others.
- ISO / IEC 15444-1 Standard.
- In principle, JP2 is free of patents, although the situation is not entirely clear.
- The technical specifications are published online, but they have to be paid for. However there is a version of the draft specifications from the year 2000 viewable at: <http://www.jpeg.org/public/fcd15444-1.pdf>

MTF

The MTF is the calculation of the spacial frequency response of an imaging system and is explained by the contrast in relation to the low frequencies, based on the idea that high frequencies mean greater detail.

Noise

It consists of the appearance of random fluctuations which are undesirable in an image. It is inevitable when receiving the signal from an electronic device. This can be caused by the sensitivity of the system, the exposure time and temperature.

OECF (Opto-Electronic Conversion Function)

It relates the optical densities of the original with the numerical values of the digitised image.

Raster or bitmap image

An image formed by a matrix of pixels, each with a particular value and a specific location within that matrix. Pixel is the contraction of the term “picture element”. It is the smallest element of a raster image with its own representation.

It is a mathematical concept, but has its equivalence in the analogue reproduction by devices.

RGB

The RGB model is an additive colour system, which are those that are formed by the sum of the various primary colours: red, green and blue. It is used in monitors, projectors, cameras, and scanners.

SNR (Signal to Noise Ratio).

To measure the noise of an electronic device, the signal received is compared with the noise obtained.

Spatial resolution

The number of samples that make up the image. When the reference is given in inches, we talk about ppi (pixels per inch), dpi (dots per inch) or lpi (lines per inch). In photography, the concept of resolution refers to the ability of an image to represent detail, a result of the processing of the spatial frequencies in the capture process. The calculation of detail is achieved by the interpretation of the MTF, a concept which applies equally to chemical and digital photography.

SRGB

The most popular abstract colour space, with a limited colour range designed for the web and therefore to be displayed on the screen.

TIFF

Format created by the Aldus Corporation in 1986, in an attempt to overcome the lack of a standard for the image. In 1987 it passed into the hands of Adobe which created new versions: version 4.0 in 1987, 5.0 in 1988, and 6.0 in 1992. Over the years, the format has established itself as the de-facto standard and has added multiple features (extensions). The bid to create a standard RAW format led to the emergence of DNG, and the need to overcome the 4 GB barrier resulted in the BigTiff. But despite the usefulness of the new formats, the benchmark version remains that from 1992.

TIFF characteristics

- Representation of bitonal, gray-scale, palette-colour and true-colour images.
- The use of a compression system is optional. Accepts different systems, such as the LZW, RLE and JPEG.
- Compatible with high depth, 16-bit.
- It is designed to be extensible and to evolve according to new requirements that might arise.
- It allows unlimited inclusion of private or specific information (metadata).
- Versions 5 and 6 are compatible with most applications.
- The structure of this format has the distinctive feature of containing the directories where the image tags (metadata) are stored.
- Metadata. The mandatory elements of the TIFF baseline for images in gray-scale or true-colour apply only to 11-12 of the total 32, not counting those corresponding to extensions.
- TIFF is a de-facto standard, the TIFF / IT is an ISO 12639:2004 standard and the TIFF / EP is an ISO 12234-2 standard.
- The technical specifications are published at:
<http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf>

Tone

It is the most important attribute of the image, and the most decisive in the evaluation of the observer. Its evaluation is complex but an objective assessment can be made based on the knowledge of the dynamic range, deviations of light and the OECF (Opto-Electronic Conversion Function).

White balance

It consists in tuning the white point of the camera depending on the type of light. The colour of the light is determined by the colour temperature.

1.2 Audiovisual concepts

In the text below we explain the main concepts related to audiovisual records. This text may be helpful to work on the exercises prepared for chapter 6.

ADC (analog-to-digital converter)

An ADC is a device used to convert from a continuous-time and continuous-amplitude analog signal to a discrete-time and discrete-amplitude digital signal.

Advanced Authoring Format (AAF)

This is a container format for professional use which stores audio and video information and, along with this, encapsulates the metadata needed for its reproduction. It is a format optimised for sharing content across different platforms and applications.

Advanced System Format (.asf)

This is a container format developed by Microsoft that can contain different data streams including audio and video. It was designed primarily for streaming, i.e. to deliver video over the Internet. It is associated with the video and audio formats Windows Media Video and Windows Media Audio.

Aspect ratio

This refers to the ratio of the rectangle that forms the image. In the analogue video era, the ratio was always 4:3. When digitalising analogue videotapes, this ratio must be maintained so as not to distort the original image. In the last decade of the 20th century, a 16:9 ratio, that it to say a more panoramic ratio, became prevalent. The 16:9 ratio predominates in digitally created video.

Audio Video Interleave (.avi)

It is an audio and video container format developed by Microsoft and is considered to be one of the de-facto standards for video storage. The audio and video stream can be encoded using different video codecs such as MJPEG or MPEG-2 or audio codecs such as AC3 or MP3.

Betacam SP

Sony Betacam SP format, launched in 1986, came to become the de facto standard for television. This format improved its predecessor, the Betacam format. It used metal particles that were more stable than the iron oxide particles of the previous formats. It recorded on separate tracks luminance and chrominance signals.

Betamax

The Betamax format (often called Beta) was developed by Sony in 1975. Sony Betamax competed during the early years of the expansion of home video, especially in the market for movies distribution. Although it offered a better quality than its main competitor, the JVC VHS, over the 80s of the twentieth century was losing market share, until 1988 when the company Sony decided to bet for VHS. It was continued manufacturing until 2002.

Bit Depth or Quantification

Each pixel is given a value, which determines the brightness and colour information. The greater number of bits assigned to quantify this value, the greater the number of tones that may be rendered. This progression is exponential. If we take as an example a grey-scale between the black and the white, an 8-bit encoding can represent 256 shades (2⁸), while a 10-bit encoding can represent 1,024 shades (2¹⁰).

Bit rate

This is the number of bits that must be transferred per second for a digital video file to play correctly. A quality video encoded in high resolution and with low compression, will need a high rate of bit transfer in order for it to be played, and will therefore require powerful processors and a considerable bandwidth.

Digital video compression

Digital video files contain lots of information that must be processed, transferred, and stored. This has led to the need to develop systems able to reduce the number of bits in a video sequence. A compressed video format can achieve a reduction in the number of bits contained in an uncompressed video file by tens or even hundreds of times.

DV (Digital Video)

The DV format appeared in 1995 with the purpose of applying digital technology to the recording of video onto magnetic tapes in cameras, but later it also began to be used for other storage devices such as hard drives. The DV-CAM and DVCPRO formats (and their high definition versions), used for recording video in a professional environment, or MiniDV, for the domestic environment, belong to the same family.

Divx

It originated as a proprietary video format, based on the MPEG-4 standard, for the renting of films. In 2000, the Divx Inc. Company was founded and, in the following years, it worked on an open source version, OpenDivx, which would result in one of the most popular codecs for the distribution of audiovisual content over the Internet, Xvid

(based on MPEG-4 ASP).

FFV1 (FF video codec 1)

This is a lossless intra-frame codec, meaning that it encodes each frame separately. It is part of the open source Ffmpeg project. It is a codec which meets the needs of long term storage so it should be taken into account by archives. It may be associated with containers such as AVI and Matroska.

Flash Video (.flv)

It is a container format developed by Macromedia and acquired by the company Adobe Systems (2005), used mainly to deliver video via the internet. It is associated with the Adobe Flash multimedia player and is commonly used with the Sorenson Spark, VP6 and, more recently, the H.264 codecs.

Frames per second

Just like analogue video, digital video gives the impression of a moving image by relaying a series of static images at high speed. In analogue video, the number of images (frames) per second was 25 for the PAL system and 30 for the NTSC system. The digital video standard keeps the same frames per second ratio.

H.264/MPEG-4 AVC

H.264 is a digital high compression codec developed by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG). The goal of the project H.264/AVC was to create a standard that was able to develop a good image quality with substantially lower bit rates than previous standards (MPEG-2, H.263, etc.). This codec has been adopted in various sectors like satellite television or mobile phone video.

Hi8

Hi8 format was evolved from the Video8 format in order to allow the semi-professional video editing. It was launched in 1989 by Sony. Hi8 format offers a higher resolution than its predecessor. Video8 tape can be played in a Hi8 player but not vice versa

Lossless and lossy compression

A distinction is usually made between formats that employ lossless compression systems (lossless) and those that present losses of information (lossy). Obviously, the former is preferable but the technological resources of archives and their storage capacities do not always allow this.

Material Exchange Format (MXF)

This is a subtype of AAF for digital video. It has been developed by the SMPTE (Society of Motion Picture and Television Engineers) and supports a large number of different data streams encoded with various codecs. It has established itself as a standard for the archival storage of digital video.

Matroska (.mkv)

This is an open-source and cross-platform container format which is associated with free software applications. It claims to be an audiovisual container format for universal use. It supports different video codecs such as MPEG2, MPEG4-ASP (DivX6 or XviD), H.264 (X 264) and Theora, and audio such as MP3, AC3 or DTS.

Motion JPEG 2000 (MJ2 o MJP2)

Based on the JPEG 2000 set of standards for digital image coding, Motion JPEG 2000 is used for the coding of moving image sequences and digital audio. An MJ2 video sequence can be formed by frames encoded separately using JPEG 2000. Motion JPEG 2000 has been established as the standard for the encoding of digital cinema.

MPEG-2

This can be both a container format and a video codec. The MPEG-2 standard includes standards for multiplexing of video, audio and other data streams. It was initially developed to serve as a basis for the digital television standard and was later adopted as the standard for the distribution of DVD-Video.

MPEG-4

It's a multimedia format of the MPEG (Motion Picture Experts Group). MPEG-4 defines a series of tools for coding audiovisual and multimedia objects. Its applications include interactive television, multimedia mobile telephone and internet video streaming. In its part 10, it defines the AVC (Audio Video Coding) compression codec equivalent to the H.264 standard.

Ogg (.ogg)

This is a container format that encapsulates audio, video and other data streams such as text or metadata. Because it is open source, several codecs developed for Ogg, like Theora (video) and Vorbis (audio) have been incorporated into a large number of media players.

Quicktime (.mov, .qt)

This is a video format developed by Apple. It is cross-platform, i.e. it can run on different operating systems and supports a good number of codecs. It is capable of encapsulating various audio, video, and multimedia content streams.

It is widely used as the format for professional video editing.

Resolution

Each digital image or frame is formed by a series of points of different intensities of brightness and colour. These points or pixels form a grid of vertical and horizontal lines. Resolution is the number of pixels the image is divided into or, in other words, the number of rows of pixels (X) by the number of columns (Y). There are different standards of resolution according to the characteristics of the format and the display medium: 720x576, 1280x720, 1920x1080, etc.

Sampling rate

The sampling rate is the number of samples taken per second of the analogue signal in order to convert it into a digital signal. The sampling rate is measured in hertz. The ITU-R BT 601 standard, the first standard for digital television, established a sampling rate of 13.5 MHz for the luminance signal and two of 6.75 MHz for the chrominance (colour difference) for reproduction in the 4:3 screen format. This standard has been superseded by other standards for the 16:9 screen ratio and for higher resolutions requiring a higher sampling frequency.

Scanning

In video technology the image is scanned onto the screen in horizontal lines, which follow each other at high speed. Due to the technological limitations of the first televisions, each image wasn't formed in a single scan but rather, first the even numbered lines were scanned, then the odd (or vice versa). This system is known as interlacing. Digital video in standard definition (SD) for a 4:3 screen ratio keeps the interlaced system. The other system is progressive scanning whereby the whole frame is constructed in a single scan and is used mainly in digital high definition (HD) video.

Subsampling

In the analogue video signal, brightness and colour are transmitted separately in three channels (YUV signal): a brightness or luminance signal (Y) and two colour chrominance signals (UV). This is the same for digital video and is represented as YCbCr... If the total numbers of samples for the three signals are represented as 4:4:4, some forms of subsampling are represented as 4:2:2, 4:2:0 or 4:1:1, i.e. a lower number of chrominance samples are taken for each luminance sample. It is possible to reduce the number of colour samples to maintain the brightness information without perceiving changes to the image because the human eye is less sensitive to colour than it is to variations in light intensity.

S-VHS

S-VHS format, launched in 1987 by JCV, was developed as a professional alternative to VHS. It offered better quality than its predecessor (400 lines of resolution in front of 240 lines of VHS). It recorded chrominance and luminance signals separated.

Although it had some impact as a format for video editing, its duration in the market was relatively short, mainly because of the emergence of digital formats. The S-VHS recorders are able to play VHS tapes but not vice versa.

U-Matic

One of the first formats that extend video to the wide public was the U-Matic, developed by Sony and launched in 1971. That format used $\frac{3}{4}$ inch tape. Thanks to its affordable cost and small size, was widely used in semi professional audiovisual productions. There were three versions: LB (Low Band), HB (High Band) and SP (Superior Performance).

VC-1

It was published in 2006 as SMPTE 421M. Its most popular implementation is Windows Media Video 9. It is based on DCT (like MPEG and H.263 family). This format emerged as an alternative to standard codec H.264 / MPEG-4 AVC. It serves both encoding interlaced and progressive video sequences.

Both HD DVD and Blu-ray adopted the VC-1 codec.

VCR (Videocassette recorder)

It's a device that records analog video and audio in a magnetic tape videocassette.

VHS

VHS was the most successful format in the market of home video. This format was launched in 1976 as a product of the JVC Company. In the early years of its existence, it mainly competed with Sony's Betamax format (and less with the Philips Video 2000), but ended imposing and became the standard format for movies distribution until the consolidation of DVD.

Video8

The first videotape format in 8mm was appeared on the market in the mid-80s of last century. One of the 8mm cameras more successful was the Sony Handycam, launched on the market in 1985. One of the virtues of 8mm format, were the small size of the camera that made it very manageable. 8mm formats successors have been Hi8 and Digital

8.

VP8

It was developed by On2 Technologies. It appeared as a competitor to other codecs such as H.264 and Ogg Theora. In 2010 On2 Technologies was acquired by Google that released the VP8 codec and promoted it for use with HTML5 in the framework of the free software WebM. WebM is based on the VP8 video codec and Vorbis audio codec encapsulated in a Matroska container.

VTR (Video tape recorder)

It is a tape recorder designed to record video material on magnetic tape. The early VTRs were open reel devices.

Windows Media Video

This is a proprietary video format from the Microsoft Company mainly used for video streaming. The latest versions of WMV allow the encoding of high definition video. This format is associated with Windows Media Audio and with the .asf container format.

2. Instructions for trainers

Here you are the trainers' instructions for the 5 exercises proposed for chapter 6. In each exercise you will find specific instructions on how to organise the training and also specific instructions for trainees. Three exercises are related to image-based records and two of them are related to audiovisual records.

Exercise 1. Create your digitisation protocol

Instructions for trainers:

The aim of the practice is to create a digitisation protocol for an institution of the GLAM sector (Galleries, Libraries, Archives and Museums). Trainees should decide about technical parameters, graphic formats for capture and delivery, some basic metadata to be embedded, and also about the way that images will be delivered. After making a decision about all these issues they should check their answers with the trainer and also with the rest of the group. There are no wrong answers even if there should be coherence with choices. So, discuss the different options you had and explain and justify every choice.

For doing this practice the trainees can read or consult:

- Section 1.2 of chapter 6.
- Besser, Howard; Sally Hubbard (ed.); Deborah Lenert (ed.). Introduction to imaging. Rev. ed. Getty Research Institute, 2003. http://getty.edu/research/publications/electronic_publications/introimages/index.html
- Federal Agencies Digitisation Guidelines Initiative. Technical Guidelines for Digitising Cultural Heritage Materials: Creation of Raster Image Master Files. 2010. http://www.digitisationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf
- Digitisation of photographic materials. Guidelines. Nationaal Archief (Holland). 2010. http://www.nationaal-archief.nl/sites/default/files/docs/guidelines_digitisation_photographic_materials_1.pdf

Exercise 2. Analysing graphic formats

Instructions for trainers:

The aim of the practice is to improve the trainees' knowledge on graphic formats. The trainees will be asked to answer a questionnaire and, after that, you should organise a debate based on these 3 questions we provide.

For doing this exercise the trainees can read or consult:

- Section 1.2 of chapter 6.
- Arms, Caroline, R.; Carl Fleischhauer. Digital formats: factors for sustainability, functionality, and quality. Washington: Library of Congress, 2005. http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf
- Frey, Franziska. File formats for digital masters. Guides to Quality in Visual Resource Imaging. Council on Library and Information Resources, 2000. <http://www.diglib.org/pubs/dlf091/dlf091.htm#visguide5>
- National Archives (UK). Selecting File Formats for Long-Term Preservation. 2008. <http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>

Debate:

After analysing the graphic formats, trainees should answer the following questions:

- Which format do you choose for capturing?
- Which format do you choose for preservation?

- Which format do you use for online dissemination?

They should explain the technical reasons of their answers and discuss their choices with the group.

Exercise 3. Analysing video formats

Instructions for trainers:

The aim of this practice is to improve the trainee's knowledge on digital video formats. In this exercise we will find a questionnaire with 5 possible answers that are 5 combinations of formats and codec. The trainees have to choose one or more of these answers for each question.

For doing this practice the trainees can read or consult:

- Section 1.3 of chapter 6.
- The website: Sustainability of Digital Formats. Planning for Library of Congress Collections <http://www.digitalpreservation.gov/formats/>
- FADGI. Digital File Formats for Videotape Reformatting. http://www.digitizationguidelines.gov/guidelines/video_reformatting_compare.html

Debate:

After analysing the audiovisual formats and codec, you should answer the following questions.

- Which format do you choose for preservation?
- Which format do you use for online dissemination?

Explain the technical reasons of your answers and discuss your choice with the group

Exercise 4. Analysing image technical metadata

Instructions for trainers:

The aim of this exercise is to learn how to analyse embedded metadata. From this metadata, trainees should be able to understand how images are captured and what their main characteristics are. We used the Jhove software to extract metadata and after that, we selected the metadata we considered most important to get the goals of this exercise. The trainees will be asked to answer a questionnaire and after should organise a debate based on the questions we provide.

For doing this exercise the trainees can read or consult:

- Section 1.2 of chapter 6.
- Guidelines for handling image metadata. V. 2.0. November 2010. Metadata Working Group, MWG. http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf
- ANSI/NISO Z39.87. Data dictionary – technical metadata for digital still images. NISO, 2006. <http://djuv.org/docs/Z39-87-2006.pdf>

Debate:

After analysing technical metadata and answering the questionnaire, trainees should answer and debate the following questions:

- Is the file format appropriate for archiving?
- Do you know what mime type is? You can look for it on the internet.
- Could you explain what a 300 ppi resolution means?
- How do you calculate the size in Mb?

Exercise 5. Analysing video technical metadata

Instructions for trainers:

The aim of this exercise is to know the main characteristics of a digital video file from analysing its technical metadata. We used the free software to extract technical metadata of a video file. Working on the questionnaire will aid us to understand the meaning of each metadata.

For doing this practice the trainees can read or consult:

- Section 1.3 of chapter 6.
 - Jisc Digital Media website. Introduction to Digital Video. <http://www.jiscdigitalmedia.ac.uk/guide/introduction-to-digital-video>
 - Mike Coyne and Mike Stapleton. The Significant Properties of Moving Images. JISC, 2008. Chapter 4. http://www.webarchive.org.uk/wayback/archive/20140616090254/http://www.jisc.ac.uk/media/documents/programmes/preservation/spmovimages_report.pdf
 - Goldsmith, Barbara. Digitising Video for Long-Term Preservation: An RFP Guide and Template. New York University Libraries, 2013. Glossar7 (p. 40). <http://library.nyu.edu/preservation/VARRFP.pdf>
- Debate:

After analysing technical metadata, trainees should answer the following questions.

- Is this file format appropriate for preservation?
- Would you choose this file format to disseminate video on internet?
- Do you think the resolution and the aspect ratio are correct? Would it be better on a high resolution (1920 x 1080) and a 16:9 aspect ratio?

Explain the technical reasons of your answers and discuss your choice with the group.

3.1 Exercises

Instructions for trainers:

The aim of the practice is to create a digitisation protocol for an institution of the GLAM sector (Galleries, Libraries, Archives, and Museums). Trainees should decide about technical parameters, graphic formats for capture and deliver, some basic metadata to be embedded, and also about the way that images will be delivered. After making a decision about all these issues, they should check their answers with the trainer and also with the rest of the group. There are no wrong answers even if there should be coherence with choices. So, discuss the different options you had and explain and justify every choice.

For doing this practice the trainees can read or consult:

- Section 1.2 of chapter 6.
- Besser, Howard; Sally Hubbard (ed.); Deborah Lenert (ed.). Introduction to imaging. Rev. ed. Getty Research Institute, 2003. http://getty.edu/research/publications/electronic_publications/introimages/index.html
- Federal Agencies Digitisation Guidelines Initiative. Technical Guidelines for Digitising Cultural Heritage Materials: Creation of Raster Image Master Files. 2010. http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf
- Digitisation of photographic materials. Guidelines. Nationaal Archief (Holland). 2010. http://www.nationaal-archief.nl/sites/default/files/docs/guidelines_digitisation_photographic_materials_1.pdf

Instructions for trainees:

From the questionnaire bellow you should create a digitisation protocol for your institution. The aim is to make a decision about technical aspects concerning digitisation in order to make clear what you expect from digitisation outcomes. It will be important to debate within the group the different options chosen by everyone.

QUESTIONNAIRE

Image resolution:

- We require a minimum resolution of 3,000 pixels across long side for negatives and prints.
- We require a minimum resolution of 7,000 pixels across long side for negatives and prints.
- We require a minimum resolution of 1,280 pixels across long side for negatives and prints.

Bit depth

- We require a bit depth of 16 bits per channel.
- We require a bit depth of 16 bits per channel. Otherwise 8 bits per channel is fine.
- We require a bit depth of 16 bits for grayscale and 8 bits for RGB.

Colour profile:

- Convert the file to sRGB IEC61966-2.1.
- Convert the file to Adobe RGB(98)
- Convert the file to Pro Photo.

Sharpening:

- Apply only capture sharpening.
- Apply a final sharpening.
- No sharpening is required.

Image capture format:

- Shoot RAW file only if DNG is allowed. Otherwise transform your camera RAW file to DNG before editing.
- Shoot camera RAW. Otherwise, shoot directly with your camera's JPEG file.
- Shoot directly with JPEG.
- Shoot directly with TIFF.

File delivery format:

- Deliver the image files in TIFF format, compressed into LZW.
- Deliver the image files in uncompressed TIFF format.
- Deliver the image files in DNG and JPEG.
- Deliver the image files in camera RAW and JPEG
- Deliver the image files in JPEG_EXIF, minimum compression.

Embedded metadata:

- Creator.
- Creator, Copyright owner.
- Creator, Copyright owner, Title.
- Creator, Copyright owner, Title, Description.
- Creator, Copyright owner, Title, Description, Keywords.

File naming:

- A unique identifier for all records of the photographic services. It is a sequential numbering string.
- The name reflects the physical location of the photographic record in our archive.
- Images get filename of the negative to have direct access to the negative
- A trigram to identify collection name, then a letter to identify single or part of a group of images, last part with the image ID and file extension. Example: ACA-F-011535-0000.jpg
- All digitised images have the same prefix. Each project has its own project number. Within the project each object or collection has its own number. Each image has a 6 character number. Example: DIGI_0007_0001_017369
- You will create your own filenames schema.

File transfer method:

- Deliver your images on a physical support: HD, DVD or CD.
- Deliver your images via FTP
- Deliver your images via WeTransfer or Dropbox
- Send your images via email.

3.2 Exercise 2: ANALYSING GRAPHIC FORMATS

Instructions for trainers:

The aim of the practice is to improve the trainees' knowledge on graphic formats. The trainees will be asked to answer a questionnaire and after you should organise a debate based on these 3 questions we provide.

For doing this exercise the trainees can read or consult:

- Section 1.2 of chapter 6.
- Arms, Caroline, R.; Carl Fleischhauer. Digital formats: factors for sustainability, functionality, and quality. Washington: Library of Congress, 2005. http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf
- Frey, Franziska. File formats for digital masters. Guides to Quality in Visual Resource Imaging. Council on Library and Information Resources, 2000. <http://www.diglib.org/pubs/dlf091/dlf091.htm#visguide5>
- National Archives (UK). Selecting File Formats for Long-Term Preservation. 2008. <http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>

Instructions for trainees:

The selection of a format is a key factor for preservation. So, it is necessary to assess the technical characteristics of each format. Library of Congress, USA National Archives, UK National Archives, Library and Archives of Canada,

and many other pioneer institutions in the field of digital preservation have established some criterion to analyse formats. We use some of these criterion to assess the 4 graphic formats we propose: TIFF, JFIF, JPEG2000, and a camera RAW.

You should answer the following questionnaire in order to analyse different aspects related to these formats. Think carefully on each criterion because, after the questionnaire, you will be asked to participate in a debate.

QUESTIONNAIRE

1. Widespread use and for a long period of time

TIFF

JFIF

JPEG2000

RAW

2. Technical specifications available in the public domain

TIFF

JFIF

JPEG2000

RAW

3. Wide range of supporting software tools

TIFF

JFIF

JPEG2000

RAW

4. Use of non-compressed images

TIFF

JFIF

JPEG2000

RAW

5. It offers metadata support for preservation

TIFF

JFIF

JPEG2000

RAW

6. Compatible with high bit depth

TIFF

JFIF

JPEG2000

RAW

7. It doesn't require large memory disc

TIFF

JFIF

JPEG2000

RAW

8. It allows the inclusion of derivatives in a single file

TIFF

JFIF

JPEG2000

RAW

9. It may contain a large number of colour spaces and profiles

TIFF

JFIF

JPEG2000

RAW

10. Suitable as a format for access and website

TIFF

JFIF

JPEG2000

RAW

11. It is considered a preservation standard

TIFF

JFIF

JPEG2000

RAW

12. Unencumbered by patents for its compression schemes

TIFF

JFIF

JPEG2000

RAW

DEBATE

After analysing the graphic formats, you should answer the following questions:

- Which format do you choose for capturing?
- Which format do you choose for preservation?
- Which format do you use for online dissemination?

Explain the technical reasons of your answers and discuss your choice with the group.

QUESTIONNAIRE with the right answers (for trainers)

1. Widespread use and for a long period of time

TIFF

JFIF

JPEG2000

RAW

2. Technical specifications available in the public domain

TIFF

JFIF

JPEG2000

RAW

3. Wide range of supporting software tools

TIFF

JFIF

JPEG2000

RAW

4. Use of non-compressed images

TIFF

JFIF

JPEG2000

RAW

5. It offers metadata support for preservation

IFF

JFIF

JPEG2000

RAW

6. Compatible with high bit depth

TIFF

JFIF

JPEG2000

RAW

7. It doesn't require large memory disc

TIFF

JFIF

JPEG2000

RAW

8. It allows the inclusion of derivatives in a single file

TIFF

JFIF

JPEG2000

RAW

9. It may contain a large number of colour spaces and profiles

TIFF

JFIF

JPEG2000

RAW

10. Suitable as a format for access and website

TIFF

JFIF

JPEG2000

RAW

11. It is considered a preservation standard

TIFF

JFIF

JPEG2000

RAW

12. Unencumbered by patents for its compression schemes

TIFF

JFIF

JPEG2000

RAW

3.3 Exercise 3: ANALYSING VIDEO FORMATS

Instructions for trainers:

The aim of this practice is to improve the trainee's knowledge on digital video formats. In this exercise we will find a questionnaire with 5 possible answers that are 5 combinations of formats and codec. The trainees have to choose one or more of these answers for each question.

For doing this practice the trainees can read or consult:

- Section 1.3 of chapter 6.
- The website: Sustainability of Digital Formats. Planning for Library of Congress Collections <http://www.digitalpreservation.gov/formats/>
- FADGI. Digital File Formats for Videotape Reformatting. http://www.digitizationguidelines.gov/guidelines/video_reformatting_compare.html

Instructions for trainees:

The election of a format is a key factor for preservation. So, it is necessary to assess the technical characteristics of each format. Library of Congress, USA National Archives, UK National Archives, Library and Archives of Canada, and many other pioneer institutions in the field of digital preservation established some criteria to analyse formats.

A format is a particular way of codifying and structuring digital information in order to store it in a file. There are formats that can contain different information streams such as video, audio, graphic images or metadata. They are called container formats and are especially important for digital video storage. Container formats should be differentiated from codecs, which constitute the set of algorithms used in the encoding of digital video and audio. The same container format can support different codecs.

You should answer the following questionnaire in order to analyse different aspects related to these container formats and codecs. Think carefully on each criterium because, after the questionnaire, you will be asked to argue for your choice.

QUESTIONNAIRE

Widespread use and for a long period of time

- Quicktime – Uncompressed, 10-bit
- MXF – JPEG 2000
- Matroska – FFV1
- FLV – VP6
- WMV – WMV9

Technical specifications available in the public domain

- Quicktime – Uncompressed, 10-bit
- MXF – JPEG 2000
- Matroska – FFV1
- FLV – VP6
- WMV – WMV9

Wide range of supporting software tools

- Quicktime – Uncompressed, 10-bit
- MXF – JPEG 2000
- Matroska – FFV1
- FLV – VP6
- WMV – WMV9

Use of lossless codec

- Quicktime – Uncompressed, 10-bit
- MXF – JPEG 2000
- Matroska – FFV1
- FLV – VP6
- WMV – WMV9

It offers metadata support for preservation

- Quicktime – Uncompressed, 10-bit
- MXF – JPEG 2000
- Matroska – FFV1

- d. FLV – VP6
- e. WMV – WMV9

It doesn't require large memory disc

- a. Quicktime – Uncompressed, 10-bit
- b. MXF – JPEG 2000
- c. Matroska – FFV1
- d. FLV – VP6
- e. WMV – WMV9

Suitable as a format for access and website

- a. Quicktime – Uncompressed, 10-bit
- b. MXF – JPEG 2000
- c. Matroska – FFV1
- d. FLV – VP6
- e. WMV – WMV9

It is considered a preservation standard

- a. Quicktime – Uncompressed, 10-bit
- b. MXF – JPEG 2000
- c. Matroska – FFV1
- d. FLV – VP6
- e. WMV – WMV9

Unencumbered by patents for its compression schemes

- a. Quicktime – Uncompressed, 10-bit
- b. MXF – JPEG 2000
- c. Matroska – FFV1
- d. FLV – VP6
- e. WMV – WMV9

QUESTIONNAIRE with the right answers (for trainers)

Widespread use and for a long period of time

- f. *Quicktime – Uncompressed, 10-bit*
- g. MXF – JPEG 2000
- h. Matroska – FFV1
- i. FLV – VP6
- j. WMV – WMV9

Technical specifications available in the public domain

- f. *Quicktime – Uncompressed, 10-bit*
- g. *MXF – JPEG 2000*
- h. *Matroska – FFV1*
- i. *FLV – VP6*
- j. *WMV – WMV9*

Wide range of supporting software tools

- f. Quicktime – Uncompressed, 10-bit
- g. MXF – JPEG 2000
- h. *Matroska – FFV1*
- i. *FLV – VP6*
- j. *WMV – WMV9*

Use of loseless codec

- f. *Quicktime – Uncompressed, 10-bit*
- g. *MXF – JPEG 2000*
- h. *Matroska – FFV1*
- i. FLV – VP6
- j. WMV – WMV9

It offers metadata support for preservation

- f. *Quicktime – Uncompressed, 10-bit*

g. *MXF – JPEG 2000*

h. *Matroska – FFV1*

i. *FLV – VP6*

j. *WMV – WMV9*

It doesn't require large memory disc

f. *Quicktime – Uncompressed, 10-bit*

g. *MXF – JPEG 2000*

h. *Matroska – FFV1*

i. *FLV – VP6*

j. *WMV – WMV9*

Suitable as a format for access and website

f. *Quicktime – Uncompressed, 10-bit*

g. *MXF – JPEG 2000*

h. *Matroska – FFV1*

i. *FLV – VP6*

j. *WMV – WMV9*

It is considered a preservation standard

f. *Quicktime – Uncompressed, 10-bit*

g. *MXF – JPEG 2000*

h. *Matroska – FFV1*

i. *FLV – VP6*

j. *WMV – WMV9*

Unencumbered by patents for its compression schemes

f. *Quicktime – Uncompressed, 10-bit*

g. *MXF – JPEG 2000*

h. *Matroska – FFV1*

i. *FLV – VP6*

j. *WMV – WMV9*

DEBATE

After analysing the audiovisual formats and codecs, you should answer the following questions.

Which format do you choose for preservation?

Which format do you use for online dissemination?

Explain the technical reasons of your answers and discuss your choice with the group

3.4 Exercise 4: Analysing image technical metadata

Instructions for trainers:

The aim of this exercise is to learn how to analyse embedded metadata. From these metadata, trainees should be able to understand how images are captured and which their main characteristics are. We used the Jhove software to extract metadata and after that, we selected the metadata we considered most important to get the goals of this exercise. The trainees will be asked to answer a questionnaire and then you should organise a debate based on the questions we provide.

For doing this exercise, the trainees can read or consult:

- Section 1.2 of chapter 6.
- Guidelines for handling image metadata. V. 2.0. November 2010. Metadata Working Group, MWG. http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf
- ANSI/NISO Z39.87. Data dictionary – technical metadata for digital still images. NISO, 2006. <http://djvu.org/docs/Z39-87-2006.pdf>

Instructions for trainees:

A digital file contains an important set of technical metadata which is necessary for the software to interpret the content described in a file. The main characteristics are reported in the embedded metadata. This metadata is also useful for archivists to better understand the images it contains and the specific technical parameters for each image.

It is also useful to understand how images are captured and sometimes how they are processed.

For this exercise, we selected the metadata we considered most important to understand the goals of the exercise. We used the Jhove software to extract metadata and after that we made the selection.

You should now answer the following questionnaire in order to analyse the metadata provided for this exercise. Think carefully on each question because, after the questionnaire, you will be asked to participate in a debate.

METADATA:

JhoveView (Rel. 1.10, 2013-06-10)

Date: 2015-04-28

Reporting Module: TIFF-hul, Rel. 1.7 (2012-08-12)

Size: 24316382 bytes

Format: TIFF

Version: 6.0

MIME type: image/tiff

Compression Scheme: uncompressed

Image Width: 3297

Image Height: 2452

Colour Space: RGB

Device Manufacturer: Phase One

Device Model Name: P40+

Scanning Software: Capture One 6 Windows

Sampling Frequency Unit: inch

X Sampling Frequency: 300

Y Sampling Frequency: 300

Bits per Sample: 8, 8, 8

Samples per Pixel: 3

Date Created: 2013-01-24

QUESTIONNAIRE

1. Which is the spatial resolution?
 - a. 3297x2452
 - b. 300 x 300
 - c. 8, 8, 8
2. Which is the bit depth?
 - a. 8 bits
 - b. 8 bits per channel
 - c. 16 bits
3. Which is the most precise way to report about the graphic format?
 - a. TIFF
 - b. image/tiff
 - c. TIFF 6.0
4. Which is the file size in Mb?
 - a. 23746
 - b. 2,3
 - c. 23
5. It uses lossless or lossy compression?
 - a. Lossless
 - b. Lossy
 - c. No compression is applied
6. What is the camera model used for digitisation?
 - a. Phase One
 - b. Phase One P40+
 - c. It is not a camera, it is a scanner

7. What is the date of digitisation?
 - a. 2013-06-10
 - b. 2012-08-12
 - c. 2013-01-24

DEBATE

After analysing technical metadata and answering the questionnaire, you should answer and debate the following questions:

Is the file format appropriate for archiving?

Do you know what mime type is? You can look for it on the internet.

Could you explain what a 300 ppi resolution means?

How do you calculate the size in Mb?

QUESTIONNAIRE with the right answers (for trainers)

1. Which is the total image resolution?
 - a. 3297x2452
 - b. 300 x 300
 - c. 8, 8, 8
2. Which is the bit depth?
 - a. 8 bits
 - b. 8 bits per channel
 - c. 16 bits
3. Which is the most precise way to report about the graphic format?
 - a. TIFF
 - b. image/tiff
 - c. TIFF 6.0
4. Which is the file size in Mb?
 - a. 23746
 - b. 2,3
 - c. 23
5. It uses lossless or lossy compression?
 - a. Lossless
 - b. Lossy
 - c. No compression is applied
6. What is the camera model used for digitisation?
 - a. Phase One
 - b. Phase One P40+
 - c. It is not a camera, it is a scanner
7. What is the date of digitisation?
 - a. 2013-06-10
 - b. 2012-08-12
 - c. 2013-01-24

3.5 Exercise 5: Analysing video technical metadata

Instructions for trainers:

The aim of this exercise is to know the main characteristics of a digital video file from analysing its technical metadata. We used the free software to extract technical metadata of a video file. Working on the questionnaire will aid us to understand the meaning of each metadata.

For doing this practice the trainees can read or consult:

- Section 1.3 of chapter 6.
- Jisc Digital Media website. Introduction to Digital Video. <http://www.jiscdigitalmedia.ac.uk/guide/introduc->

tion-to-digital-video

- Mike Coyne and Mike Stapleton. The Significant Properties of Moving Images. JISC, 2008. Chapter 4. http://www.webarchive.org.uk/wayback/archive/20140616090254/http://www.jisc.ac.uk/media/documents/programmes/preservation/spmovimages_report.pdf
- Goldsmith, Barbara. Digitising Video for Long-Term Preservation: An RFP Guide and Template. New York University Libraries, 2013. Glossar7 (p. 40). <http://library.nyu.edu/preservation/VARRFP.pdf>

Instructions for trainees:

In any digitisation project, it is necessary to know the main characteristics which define your digital video files. Before starting to digitise, we will have to decide several parameters like resolution, bit depth, aspect ratio, etc. and program its values in our capturing devices and software. The election of those values will determine the quality, size, and sustainability of the video file resulting of the digitisation process.

To make this practice, we will extract technical metadata from a file and we will analyse it by answering a questionnaire.

Finally, we will answer three questions in order to encourage debate about the results of the practice.

After opening a video file with MediaInfo, you will see following metadata

General #0

Complete name : C:\Users\psaavedra\Videos\Cavalcada_reis_1985.avi
Format : AVI
Format/Info : Audio Video Interleave
Format/Family : RIFF
File size : 3.02 GiB
PlayTime : 14mn 16s
Bit rate : 30.3 Mbps
Video #0
Codec : Sony DV
Codec/Family : DV
Codec/Info : Sony Digital Video (DV) 525 lines at 29.97 Hz or 625 lines at 25.00 Hz
PlayTime : 14mn 16s
Bit rate : 27.6 Mbps
Width : 720 pixels
Height : 576 pixels
Display Aspect ratio : 4/3
Frame rate mode : CFR
Frame rate : 25.000 fps
Standard : PAL
Resolution : 24 bits
Chroma : 4:2:0
Interlacement : Interlaced
Audio #0
Codec : PCM
Codec/Family : PCM
Codec/Info : Microsoft PCM
Codec settings : Little / Unsigned
Codec settings, Endianness : Little
Codec settings, Sign : Unsigned
Bit rate : 1536 Kbps
Channel(s) : 2 channels
Sampling rate : 48.0 KHz
Resolution : 16 bits
Coherency/PlayTime : 856360

QUESTIONNAIRE

From these metadata, choose the correct answer

1. Which is the image resolution?
 - a. 720 x 576
 - b. 1280 x 720
 - c. 1920 x 1080
2. Which is the container format?
 - a. AVI
 - b. DV
 - c. PCM
3. Which is the video codec?
 - a. AVI
 - b. DV
 - c. PCM
4. Which is the chroma subsampling?
 - a. 4:2:2
 - b. 4:2:0
 - c. 4:1:1
5. It uses lossless or lossy codecs?
 - a. Lossless
 - b. Lossy
6. How much bit information has to be processed per second?
 - a. 25 fps
 - b. 24 bits
 - c. 30,3 Mbps
7. The scanning system is:
 - a. Progressive
 - b. Interlaced
 - c. Constant
8. Which is the standard system?
 - a. NTSC
 - b. PAL
 - c. SECAM
9. How many audio channels?
 - a. Stereo
 - b. Mono
 - c. Surround sound
10. What is the video definition?
 - a. High definition (HD)
 - b. Low definition
 - c. Standard definition (SD)

QUESTIONNAIRE with the right answers (for trainers)

1. Which is the image resolution?
 - a. 720 x 576
 - b. 1280 x 720
 - c. 1920 x 1080
2. Which is the container format?
 - a. AVI
 - b. DV
 - c. PCM
3. Which is the video codec?
 - a. AVI
 - b. DV
 - c. PCM

-
4. Which is the chroma subsampling?
 - a. 4:2:2
 - b. 4:2:0
 - c. 4:1:1
 5. It uses lossless or lossy codecs?
 - a. *Lossless*
 - b. Lossy
 6. How much bit information has to be processed per second?
 - a. 25 fps
 - b. 24 bits
 - c. 30,3 Mbps
 7. The scanning system is:
 - a. Progressive
 - b. *Interlaced*
 - c. Constant
 8. Which is the standard system?
 - a. NTSC
 - b. *PAL*
 - c. SECAM
 9. Which audio channels?
 - a. *Stereo*
 - b. Mono
 - c. Surround sound
 10. What is the video definition?
 - a. High definition (HD)
 - b. Low definition
 - c. *Standard definition (SD)*

DEBATE

After analysing technical metadata, you should answer the following questions.

Is this file format appropriate for preserving digital video in the archive?

Would you choose this file format to disseminate the video on internet?

Do you think the resolution and the aspect ratio are correct? Would it be better in a high resolution (1920 x 1080) and a 16:9 aspect ratio?

Explain the technical reasons of your answers and discuss your choice with the group

4. Experiences

4.1 Experience 1, Digitising photographic records

EuropeanaPhotography project

EuropeanaPhotography was a quite unique consortium and project, putting together some of the most prestigious photographic collections from archives, public libraries, and photographic museums covering 100 years of photography. They range from 1839 with the first example of images from Fox Talbot and Daguerre to the beginning of the Second World War (1939).

It was funded within the Competitiveness and Innovation framework programme 2007-2013. It lasted 36 months, from the 1st of February 2012 to the 31st of January 2015. It was dedicated to the selection, digitisation and enrichment of 450,000 early photographs (1839-1939), to be provided with metadata to the digital cultural heritage database Europeana.

The project outcomes actually improved the profile of Europeana under the content and multilingual points of view. Europeana has an impressive mass of images that usually represent cultural objects like paintings, sculptures, and archaeological artefacts, while photographic images, the early ones in particular, were underrepresented. Europeana-Photography filled this gap in.

The consortium included 19 partners' representatives of 13 Member States. They are leading private and public photographic agencies as well as public institutions and museums.

The core activities of EuropeanaPhotography can be summarised as follows:

- Selection of images to be digitised
- Digitisation
- Indexing and cataloguing
- Multilingual enrichment and mapping of metadata
- Provision of content to Europeana

Content selection and digitisation were preliminary operations at the basis of all tasks to follow, such as: multilingual enrichment (using the multilingual vocabulary, tailored to the needs and characteristics of EuropeanaPhotography), aggregation of metadata in MINT (allowing for metadata mapping), publication of content on the project server, and ingestion of digital content by Europeana.

The project was structured in 7 work packages (WPs), led by different partners. One of these work packages was dedicated to digitisation.

Concerning digitisation, the initial objective was the selection and definition of standards with regard to digitisation



and image quality, to be used throughout the project term by all providers, in order to guarantee a high and constant quality level.

Furthermore, the work package leader, CRDI, assisted content providers in adapting the most recent technologies and procedures for preparation of historical photos and the implementation of digitisation techniques.

The key objective was to not only align partners to the selected technical, scientific and methodological standards, but also to help them mold their own workflows to these project standards and guidelines.

To this end, they produced a set of recommendations and factsheets, and organised three seminars: (Girona, May 2012), (Barcelona, September 2013) and (Barcelona, January 2014).

- **Factsheet 1. General recommendations for digitising the photographic heritage**, by Carles Mitjà and Beatriz Martínez. These recommendations were produced from the lectures and feedback during the workshop about digitisation (Girona, May 2012). The workshop aimed to provide an overview of the available systems for digitisation of original photographic materials as well as the procedures to obtain the best results in terms of image quality and fidelity to the original.

- **Factsheet 2. Practical recommendations for digitising the photographic heritage**, by Bruno Vandermeulen. These recommendations were produced from the technical visit in KU Leuven, in the framework of the content seminar (Leuven, April 2012). The factsheet provides some practical recommendations on digitisation.

- **Guide. Input profiling with a colourchecker**, by Bruno Vandermeulen. This guide was produced by one of the partners to give advice in colour management, a critical issue for digitisation. The guide is focused on the Colour-Checker from X-Rite, an industry standard colour reference chart used for colour control and evaluation. It gives you a step by step guide through the different stages of creating a custom input profile and apply this on RAW images.

Finally, digitisation as such went hand in hand with the collection and inclusion of digitisation metadata in the part-



ners' databases. This data was to be integrated in the descriptive information later to be provided with each image. To ensure a correct application and to support the implementation of good practices, a final quality check of the digitised images was performed. The digitisation process was carried out in line with the existing standards to deliver high-quality files. The standards and guidelines are a good guarantee for image quality. After digitising, images had to be tested for compliance with the documented specification of each partner. Image quality assurance had to consider the following issues: correct file naming, size of images, acceptable borders and margins on images, image completeness, correct graphic files, image characteristics (resolution, bit-depth, colour space, etc.) .

The most significant results of the digitisation work package of this project are:

- Digitisation and local cataloguing of over 451,178 photographs at the end of the project.
- Publication of 2 deliverables: D3.1.1 Digitised material (first release) and D3.1.2 Digitised material (second release).
- Common agreement in the best way to organise a digitisation project. Mainly following FADGI guidelines (2009).
- Common agreement in the technical parameters for digitisation. Mainly following FADGI guidelines (2010)
- Three different factsheets/guides concerning digitisation procedures.
- Organisation of five different training sessions related to digitisation.
- Two surveys: the experience of digitising / self image quality control.
- Report on digital image quality control (published as an annex of D 3.1.2)



All these resources are published on the European Photography website: <http://www.europeanaphotography.eu/> And the digitised images are available at European platform: http://www.europeana.eu/portal/search.html?query=%3A*&rows=24&qf=PROVIDER%3AEuropeanaphotography&qt=false

At the end of the project, partners of European Photography founded Photoconsortium, an international consortium for photographic heritage that has been established as a non-profit association with a democratic structure, political independence, and unlimited duration. The purpose of the association is the promotion and enhancement of the culture of photography and the photographic heritage. To achieve its statutory goals, it aims to promote, organize, and manage: conferences, exhibitions, awards and training courses as well as to carry out publishing activities. <http://www.photoconsortium.net>

4.2 Experience 2, Digitising sound records

The Nuclear Claims Tribunal of the Republic of the Marshall Islands

The sound and video archive of the Nuclear Claims Tribunal of the Republic of the Marshall Islands is composed of 495 cassette tapes containing a total of 483 hours of sound, and 40 VHS videotapes totalling 63 hours and 43 minutes of moving image and sound. The cassette tapes are dated between August 1989 and October 2007, whereas the videotapes cover the period from September 1999 to July 2002. The languages of the records are English and Marshallese.

The Girona City Council was informed about the existence of this archive and the responsible party of the municipal archive understood the need to digitise it as a part of the world heritage. They were aware of the vulnerability of magnetic tapes to the physical degradation and technological obsolescence that exposed those documents to the risk of disappearing.

To understand the value of the mentioned documentation, we have to place it within its historical context. After World War II, between 1946 and 1958, the US military conducted several nuclear tests in the Marshall Islands: 67 nuclear tests with a power equivalent to 7000 times the force of the Hiroshima bomb. Those nuclear tests caused in the population many calamities: forced migrations, famine and diseases due to radiation. In 1988, after the independence of the country (1986), the Nuclear Claims Tribunal of the Marshall Islands was established, and in 1991 the Court applied a program of compensation for personal injury directly caused by nuclear testing program, such as cancer and other diseases.

The cassette tapes contain the statements of witnesses, lawyers, and judges involved in the claim process for the effects of nuclear testing in the Marshall Islands. The videos contain also images of the hearings.

In October 2012, Girona City Council signed an Agreement with the Nuclear Claims Tribunal of the Republic of the Marshall Islands for the custody of the audio and video records generated by such Tribunal. Under this Agreement, the Municipal Archive of Girona City Council agreed to reproduce all the material in digital format, store securely and preserve indefinitely the digital copies of the records, assuring unfettered access to such records to the Nuclear Claims Tribunal. The Agreement also provided that, once the process was completed, original records were to be returned in the same condition in which they were received.

First materials arrived at Girona in September 2012 and on the 28th of February 2014 all tasks were completed. They consisted of the organisation, physical description, and digitisation of video and audio materials. Since then, digital copies have been kept in the repository of the Municipal Archive of Girona in safe conditions and the original materials have been returned to the Marshall Islands.

Digitisation

The capture interface Mbox Pro, together with the software Avid Pro Tools 9, was used for the digitization of the cas-



Cassette tapes in **provisional** archive boxes



Repository of the Municipal Archive of Girona

sette tapes.

As for the recording equipment, a Tascam CD-A750 professional tape deck and a Sony TCM-200DV cassette recorder were used, depending on the cassette type.

The audio files generated have the following characteristics:

- Format: Wave (PCM)
- Channels: 2 (stereo)
- Sampling rate: 96 KHz
- Bit depth: 24 bit

It should be noted that cassette tapes were not originally recorded in the same way, the same playback speed, or the same number of channels. Three main different types of recordings were determined:

- Stereo recordings (2 channels) at a speed of 4.76 cm/s
- Mono recordings (1 channel) at a speed of 2.4 cm/s



CASSETTES: Tascam CD-A750 tape deck



CASSETTES: Software Avid Pro Tools 9

- Stereo recordings (2 channels) at a speed of 2.4 cm/s

Cassette tapes recorded in stereo at 4.76 cm/s were reproduced for their digitisation in a professional tape deck, specifically a Tascam CD-A750 model.

Cassette tapes recorded in mono at 2.4 cm/s were reproduced for their digitisation in a cassette recorder Sony TCM-200DV.

Cassette tapes recorded in stereo at 2.4 cm/s were digitised with a Tascam deck (in this cases, speed was adjusted with the software Pro Tools).

5. *Recources*

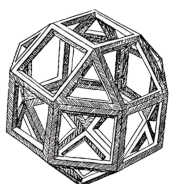
For image-based records:

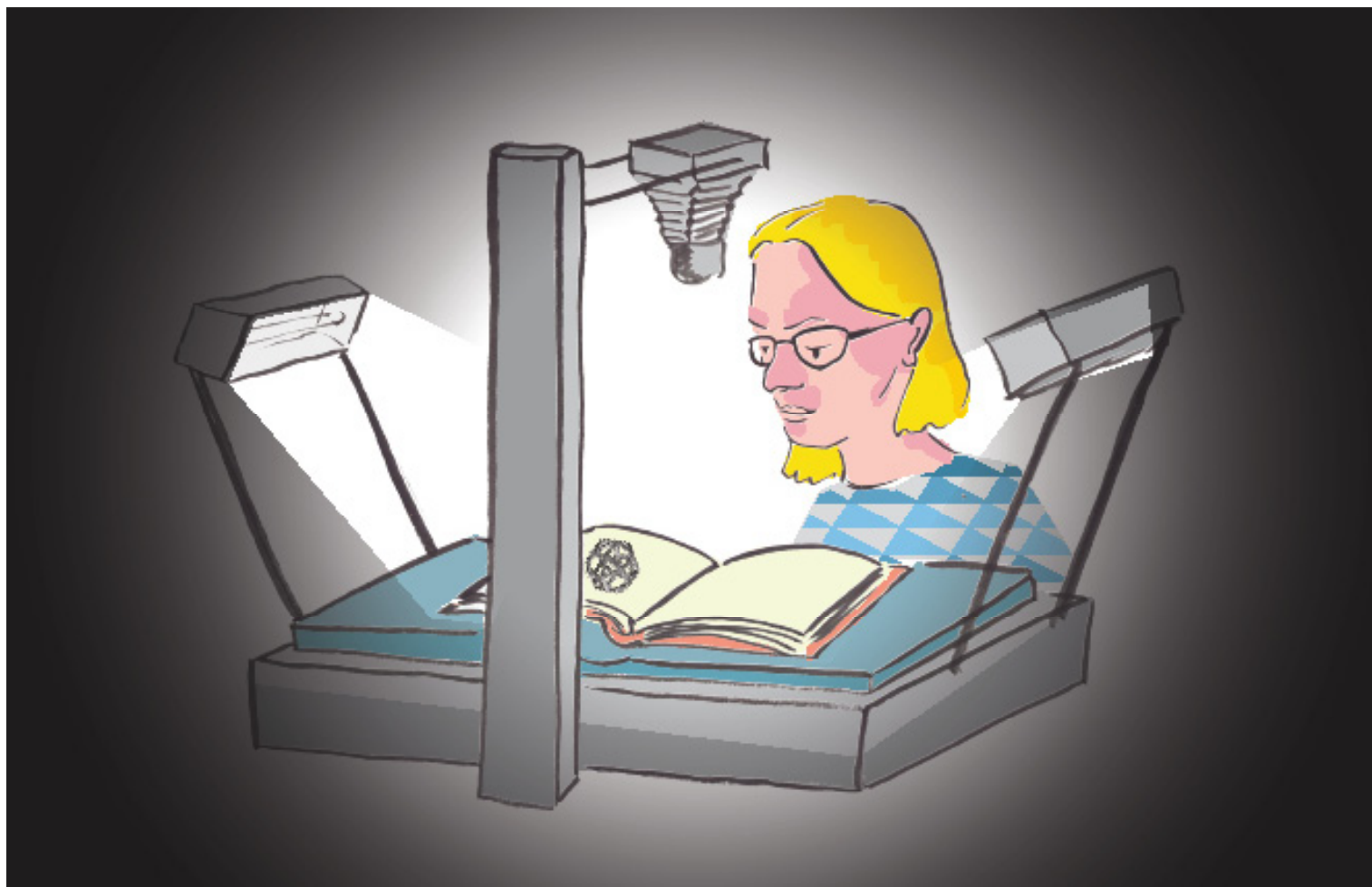
- ANSI/NISO Z39.87. Data dictionary – technical metadata for digital still images. NISO, 2006. <http://djuv.org/docs/Z39-87-2006.pdf>
- Arms, Caroline, R.; Carl Fleischhauer. Digital formats: factors for sustainability, functionality, and quality. Washington: Library of Congress, 2005. http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf
- Besser, Howard; Sally Hubbard (ed.); Deborah Lenert (ed.). Introduction to imaging. Rev. ed. Getty Research Institute, 2003. http://getty.edu/research/publications/electronic_publications/introimages/index.html
- Digitisation of photographic materials. Guidelines. Nationaal Archief (Holland). 2010. http://www.nationaal-archief.nl/sites/default/files/docs/guidelines_digitisation_photographic_materials_1.pdf
- Federal Agencies Digitisation Guidelines Initiative (FADGI). Technical Guidelines for Digitising Cultural Heritage Materials: Creation of Raster Image Master Files. 2010. http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf
- Frey, Franziska. File formats for digital masters. Guides to Quality in Visual Resource Imaging. Council on Library and Information Resources, 2000. <http://www.diglib.org/pubs/dlf091/dlf091.htm#visguide5>
- Guidelines for handling image metadata. V. 2.0. November 2010. Metadata Working Group, MWG. http://http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf
- National Archives (UK). Selecting File Formats for Long-Term Preservation. 2008. <http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>

For audiovisual records:

- Federal Agencies Digitisation Guidelines Initiative (FADGI). Digital File Formats for Videotape Reformatting. http://www.digitizationguidelines.gov/guidelines/video_reformatting_compare.html
- Goldsmith, Barbara. Digitising Video for Long-Term Preservation: An RFP Guide and Template. New York University Libraries, 2013. Glossar7 (p. 40). <http://library.nyu.edu/preservation/VARRFP.pdf>
- JISC Digital Media website. Introduction to Digital Video. <http://www.jiscdigitalmedia.ac.uk/guide/introduction-to-digital-video>
- Mike Coyne and Mike Stapleton. The Significant Properties of Moving Images. JISC, 2008. Chapter 4. http://www.webarchive.org.uk/wayback/archive/20140616090254/http://www.jisc.ac.uk/media/documents/programmes/preservation/spmovimages_report.pdf

-
- Sustainability of Digital Formats. Planning for Library of Congress Collections <http://www.digitalpreservation.gov/formats/>





Keep the metadata in mind, even while scanning.
Don't daydream about Surströmming!

Chapter 7. Metadata

Contents

1. *Introduction*
2. *Instructions for trainers*
3. *Exercises*
 - 3.1. *Exercise 1, Metadata that come with the hardware*
 - 3.2. *Exercise 2, Files and metadata*
4. *Experience*
5. *Resources*

1. Introduction

Metadata are structured data about objects. There are four main types of metadata:

- descriptive metadata, required to identify and find objects,
- structural metadata, these lay the solid relationship between individual objects that make up a unit,
- technical metadata, information on the realisation of the object,
- administrative metadata, these focus on control and management of objects.

In general, the more information has to be added to an object, the richer the context and thus the potential uses. It is important to establish by whom a particular description text has been added and for what purpose. Metadata, as part of the digital heritage, remain permanently accessible.

There are many standards for recording of metadata. The difference between the standards is dependent on the function of the metadata. Descriptive and structural metadata vary greatly by domain and / or material. When choosing a certain standard you should consider use outside the domain of the own institution, most importantly on the Internet.

For new intended uses internal standards used can be converted (mapping). For specific purposes there are specialised exchange standards, such as A2A (Archives tot Archives) or more generic ones, such as Dublin Core. Mapping always means loss of information.

In order to be able to use the meaning (semantics) of content descriptions as widely as possible, while describing one should always answer the basic user questions: who, what, where, when and how.

Which metadata do we use while sharing digital photo's with the public?

Experience teaches us that the target group for our photo viewer is mostly interested in place names and family names. Local archives are very place determined, we must therefore be very precise about making an accurate description of where and when the photo was taken. The public is more interested in the place (the exact location and city quarter), time (at least which year) and persons that were portrayed than in the question who the author of the photo is or what technique is used (calotype, carbon print, albumin print etc.). For the best search results, we describe this data extensively. Each photo gets an exclusive number and we also describe the photographer of the photo, its royalty rights, if needed the agency of company the photographer was working for, and in what format the image was taken.

We own about 228,000 photo's drawings, prints and maps, of which some 40,000 images now can be seen on our photo viewer that is available on our website: www.schiedam.nl/Def/Gemeentearchief/Archieven-en-Collecties/Beeldbank.html. In this photo viewer an accurate description is given of the image with all relevant information. We show the photo in our photo viewer without a watermark, but in low quality (150 dpi).

In the Media Conversion Centre (the MKC) in Fränsta, Sweden, and in the Lantmätieret project (see www.lantmatieret.se) Deja Vue software was used but this did not meet the expectations of the MKC. Maybe our Leonardo group can make a list of software that meets our standards while still easy to use. If we start to use the same software we can all benefit. Communication is easier, we can share our knowledge while using it, make deals with the enterprises providing the software and use the same metadata.

In order to do so it is recommended to use a general standard as the Dublin Core or the Metadata system of the City Archives of Rotterdam, both are based upon the international standards as ISAD(G), EAD and ISAAR(cpf). The Dublin Core Schema is a small set of vocabulary terms that can be used to describe web resources (images, web pages, etc.), as well as physical resources such as books or CDs,

and objects like artworks. The full set of Dublin Core metadata terms can be found on the Dublin Core Metadata Initiative (DCMI) website.

ISAD(G) (General International Standard Archival Description) defines the elements that should be included in an archival finding aid. It was approved by the International Council on Archives (ICA/CIA) as a standard to register archival documents produced by corporations, persons and families.

EAD (Encoded Archival Description) is an XML standard for encoding archival finding aids, maintained by the Technical Subcommittee for Encoded Archival Description of the Society of American Archivists, in partnership with the Library of Congress.

ISAAR, an International Standard Archival Authority Record, is a form of authority control record, standardised by the Committee of Descriptive Standards of the International Council on Archives.

A digitisation project demands a selection strategy. The obvious starting point should be the existing regulations and instructions of the organiser, although these may not be sufficient when it comes to digitisation. Therefore, it is useful to establish a specific selection strategy. To begin with, it is crucial to have an adequate acquisition strategy. The better quality of the acquisitions, the better the quality of the digitised product will be. For example, it is crucial to gain proper metadata when accepting a new acquisition. Moreover, when requiring new acquisitions, one can gain opportunities to make an impact on the author when it comes to technical quality, how the author stores the images, and the how the author creates metadata such as registers and descriptions. This facilitates forthcoming acquisitions and in the long run facilitates the digitisation process.

When it comes to establish a selection strategy for digitising projects there are several aspects to consider.

Selection criteria

Documentary criteria

In what way do the records reflect the author's activity?

Are the photos unique or common?

Do the records change over time?

Provenance and value criteria

Is the provenance of the photos known?

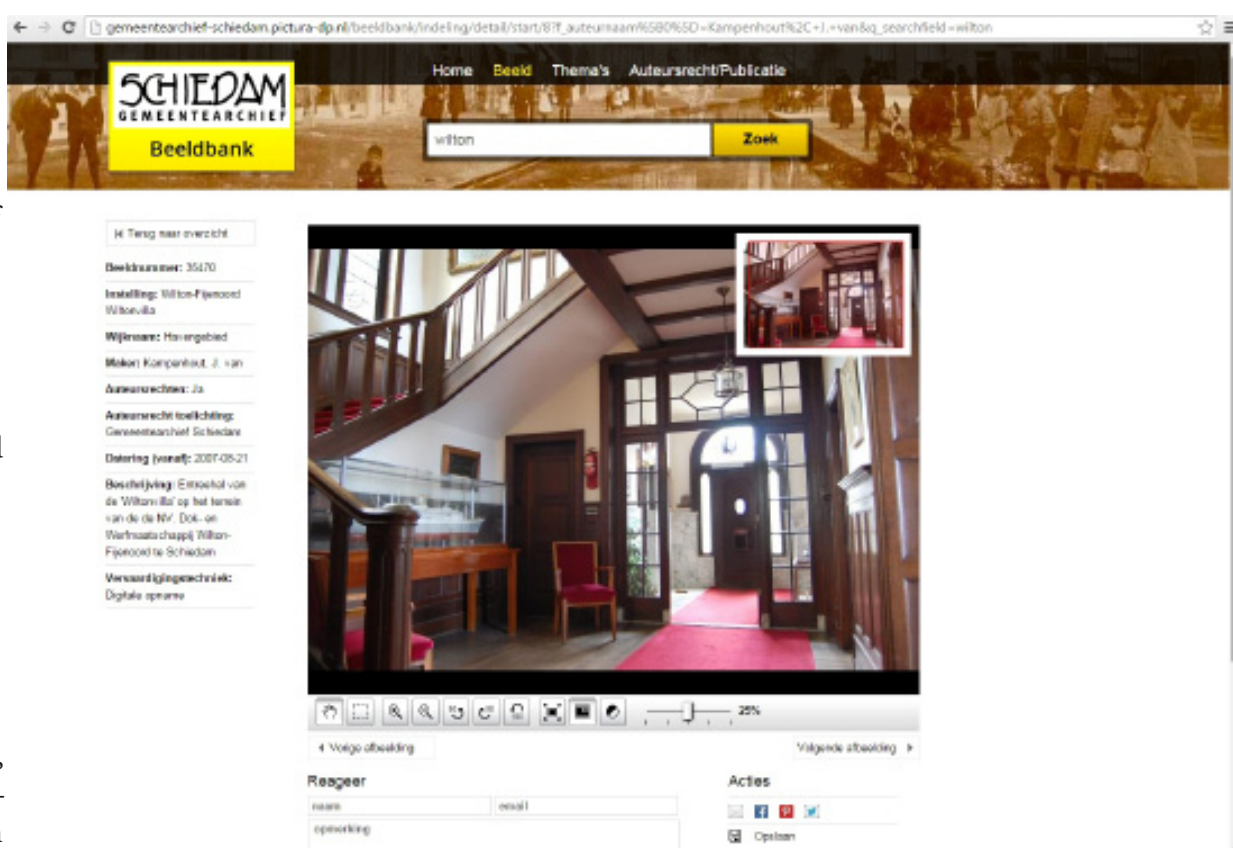
Is it possible to say how representative the photos are?

How accessible is the information in the photos?

Is there important or rare photo historical information regarding technique and/or method?

Technical criteria

What is the physical condition of the photos?



What technical abilities does the institution have for storage of the originals and copies? Should the copies be accessible in the future as technology evolves?

What technical (image) quality should the copies have?

What format should be chosen? Consider that linked open data is an important way to make the image useful. This means that the digital copies can be published in structured, readable, open formats that can be linked to other material.

What kind of metadata is adequate?

How does the technical context look? Will the user be able to find the images? Will they be published in a structured format? Will a standardised format be used?

Dissemination: How can the images be as well as not be used?

Economic criteria

What costs for preservation and system maintenance may occur?

Will there be, in the aspect of accessibility and dissemination, any costs or limitations due to questions of copyright?

2. Instructions for trainers

In the exercise, there is an emphasis on adjustable metadata (as distinguished from non-adjustable metadata).

Key to exercise 3.1

It is very important to bear in mind that adjustability of settings may vary from camera to camera. Relevant settings are:

Format

Bit rate

Duration (obviously)

Height

Width

Audio bit rate

3. Exercises

3.1 Exercise 1, Metadata that come with the hardware

Depending on the camera used you can adjust settings. The camera used in chapter 5 to record the interviews, a Samsung HMX-U10, has its own technical metadata as shown below. Before you start to record, you have to make a choice about the settings of the camera and therefore about which metadata you need.

Which settings can be changed by the user?

General

Complete name : \\synarch01\E-Depot\Archief_0590\Kinderuitzendingen videos\mevr J. Herlaar-Winkelman.mov

Format : MPEG-4

Format profile : QuickTime

Codec ID : qt

File size : 3.04 GiB

Duration : 2h 23mn

Overall bit rate : 3 032 Kbps

Encoded date : UTC 2009-11-30 11:52:52

Tagged date : UTC 2009-11-30 15:59:24

Writing library : Apple QuickTime

Video

ID : 2

Format : AVC

Format/Info : Advanced Video Codec

Format profile : Main@L3

Format settings, CABAC : No

Format settings, ReFrames : 2 frames

Format settings, GOP : M=2, N=24

Codec ID : avc1

Codec ID/Info : Advanced Video Coding

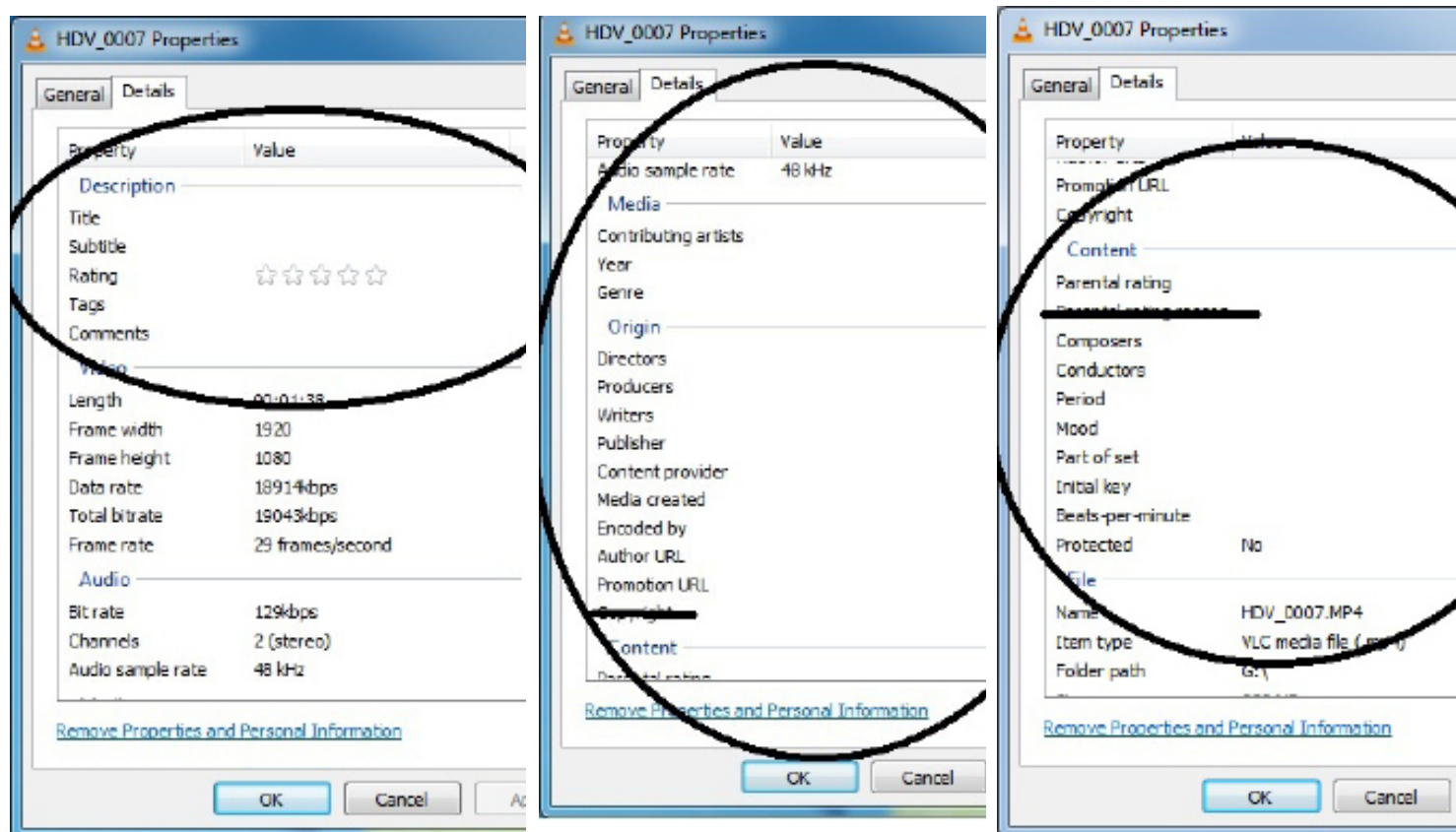
Duration	: 2h 23mn
Bit rate	: 2 918 Kbps
Width	: 720 pixels
Clean aperture width	: 703 pixels
Height	: 576 pixels
Clean aperture height	: 576 pixels
Display aspect ratio	: 1.85:1
Original display aspect ratio	: 5:4
Clean aperture display aspect ratio	: 16:9
Frame rate mode	: Constant
Frame rate	: 25.000 fps
Standard	: PAL
Colour space	: YUV
Chroma subsampling	: 4:2:0
Bit depth	: 8 bits
Scan type	: Progressive
Bits/ (Pixel*Frame)	: 0.281
Stream size	: 2.92 GiB (96%)
Language	: English
Encoded date	: UTC 2009-11-30 15:43:09
Tagged date	: UTC 2009-11-30 15:59:24
Colour range	: Limited
Colour primaries	: BT.601 NTSC
Transfer characteristics	: BT.709
Matrix coefficients	: BT.601
Audio	
ID	: 1
Format	: AAC
Format/Info	: Advanced Audio Codec
Format profile	: LC
Codec ID	: 40
Duration	: 2h 23mn
Source duration	: 2h 23mn
Bit rate mode	: Constant
Bit rate	: 110 Kbps
Nominal bit rate	: 128 Kbps
Channel(s)	: 2 channels
Channel positions	: Front: L R
Sampling rate	: 48.0 KHz
Compression mode	: Lossy
Stream size	: 112 MiB (4%)
Source stream size	: 112 MiB (4%)
Language	: English
Encoded date	: UTC 2009-11-30 15:43:09
Tagged date	: UTC 2009-11-30 15:59:24

3.2 Exercise 2, Files and metadata

Once a file is created you can observe its metadata. To do this you need to turn to the properties. This can usually be done with a right click.

See images on the right.

Now try to change some of the adjustable metadata, for example “title”.



4. *Experience of the Public Works Programme for Digitisation, Aquincum Museum, Budapest.*

Metadata is indispensable elements of digitisation and essential parts of a digital record. Dealing with metadata is one of the most important tasks before publishing anything on the internet.

In the first phase of our digitisation project within the framework of the Public Works Programme for Digitisation coordinated by MANDA (Hungarian Digital Archives), we carried out the recorded data of existing inventory books in a database. This data was uploaded to the database (SQL) of the digital pictures. The database itself has its own web interface, where we could publish the uploaded files — texts, images, documents — with a single click. Publishing on the internet happened of course with a reduced resolution stock of images, and so through the web link anyone could have access to the search engine of the database.

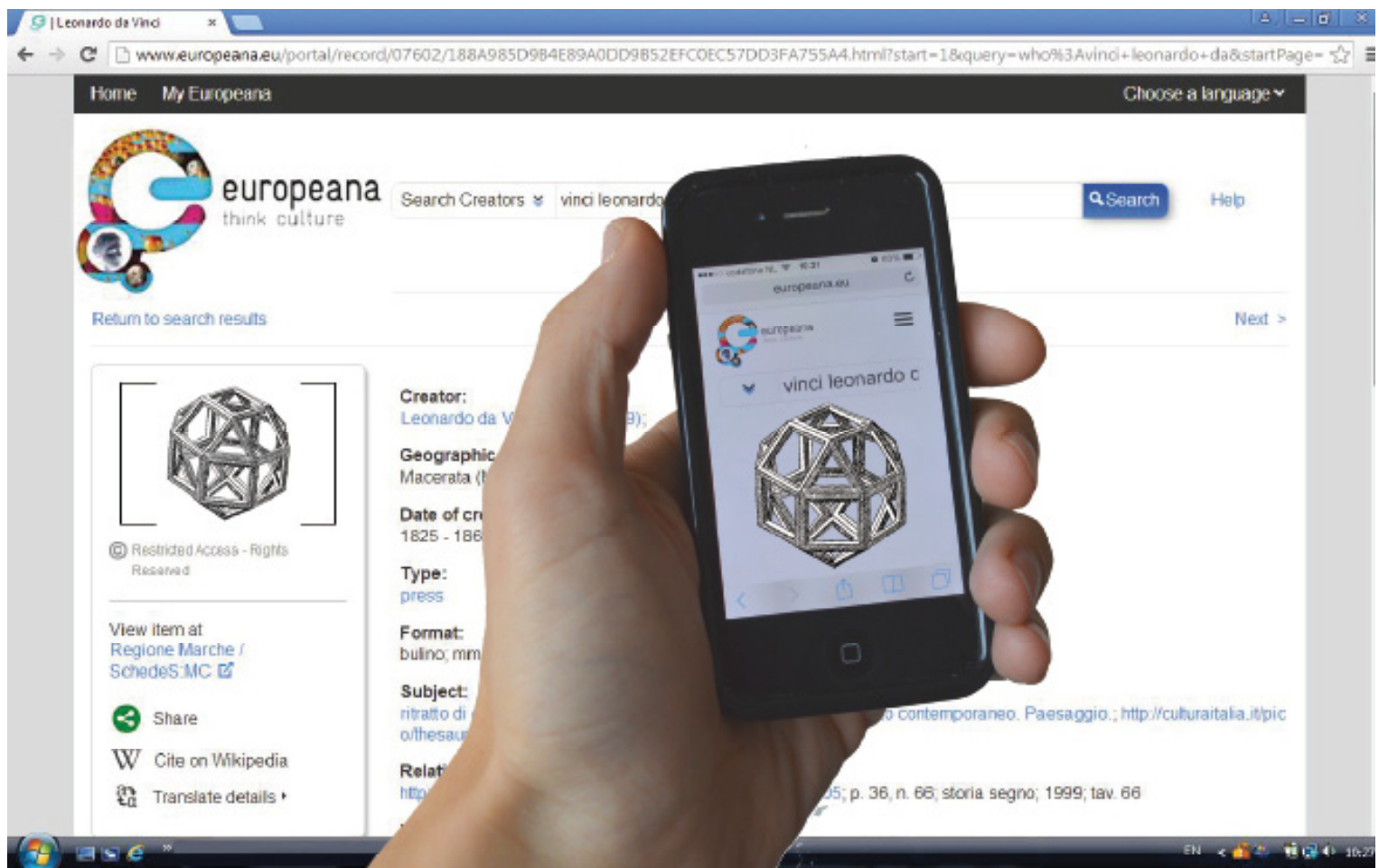
The agreement between the two institutions — MANDA and Aquincum Museum — included the transfer of the full amount of digitised data to the database developed by MANDA. Interoperability between the two systems was provided by exports in HTML and XML. In this process two collections with 7000 photos from the museum's database were converted. Because the the museum's online search engine didn't require metadata for images, all the images transferred to the host database had to be re-attached to the existing data set.

As the above example shows, transferring data between different databases requires preliminary consultation of the partners, possibly prior to the start of the project, in the planning phase. Defining the exact input and output contents and the expectations of both parties is essential.

5. Resources

Links:

- <http://www.apex-project.eu/>
- <http://www.archivesportaleurope.net/>



Succes! The image can now be found online, thanks to the toolbox.

Chapter 8. Providing access

Contents

1. **Introduction**
2. **Instructions for trainers**
3. **Exercises**
 - 3.1 **Exercise 1: XML**
 - 3.2. **Exercise 2: E(ncoded)A(rchival)D(escription)**
 - 3.3 **Exercise 3: A(rchives)P(ortal)E(urope)**
 - 3.4 **Exercise 4: Create your own digital collection**
4. **Experiences**
5. **Recourses**

1. **Introduction**

Access is the ability to make use of material from a collection, usually subject to rules and conditions (based on the definition in ISAD(G): General International Standard Archival Description. 2nd edition). Materials should not only be accessible now but also –more importantly- in the long term. Long term accessibility can be problematic because of different reasons, for example script typefaces.

Out of date file formats pose a modern challenge.

Access is traditionally provided for by professionals but also volunteers may work on it. Internet has boosted the possibilities for crowdsourcing. To find and identify materials one can revert to descriptions but also to images.

Finding aids are the tools to give access to archives. On a European level they are gathered in the Archives Portal Europe (APE) using Encoded Archival Description (EAD), which is a form of extensible Mark-up Language. EAD is in line with ISAD(G), a standard provided for by the International Council on Archives.

Within APE, one can present both archival description and images.

2. **Instructions for trainers**

The exercises of Chapter 8 help to show trainees the basics of publishing finding aids and digital information in general. The emphasis is laid on practice, and work in smaller groups (2-3 persons) is recommended. Exercise 1 deals with XML but is only meant to give a limited impression of it.

Exercise 4 helps trainees to plan a project in publishing materials.

3. **Exercises**

3.1. **Exercise 1: XML**

XML = extensible mark-up language. XML uses elements to mark-up text.

For example: Budapest, Gävle, Borås, Girona, Schiedam.

These are the cities from our project, but how does a computer know that?

With XML we can mark the meaning of the word, like this:

```
<city>Budapest</city>
```

```
<city>Gävle</city>
```

Note that the element begins and ends with a tag. And: the end tag contains a '/' preceding the name of the element.

Elements have to be put in order like a box in a box. That is: like Russian dolls.

Like this:

```
<country>Hungary
```

```
<city>Budapest</city>
```

```
</country>
```

The following is wrong (it is not 'a box in a box'):

```
<country>Hungary
```

```
<city>Budapest
```

```
</country></city>
```

XML with correct syntax is called well-formed.

Now you are going to make a well-formed xml file. Copy the following:

```
<country>Austria
```

```
<city>Salzburg
<person>Wolfgang</person>
</city>
</country>
```

Paste and save it as an xml file, which is as a file with extension .xml. Using Windows, Notepad is suitable for this. However, it is then necessary to explicitly save the file as an utf-8 encoded text file. For this exercise Notepad is fine but if you want to do more serious XML editing it is recommendable to use an editor dedicated at XML editing. oXygen(?), EditiX and XML spear are plausible choices, of which the last one is free

Now include your own city and your own name. Remember:

- begin and end with a tag

- the order of elements is 'box in box'

When you have finished: save the file again and open it with double click.

What do you see? Is it correct?

If you made a mistake in the syntax you'll be notified by your browser.

Now you've made a well-formed xml file!

3.2 Exercise 2: E(ncoded)A(rchival)D(escription)

The official website on EAD is hosted by the US Library of Congress.

Go to <http://www.loc.gov/ead/>

Go to the EAD tag library (with examples and crosswalks). Under "EAD elements by tag name" you find a full list of all elements that are allowed in EAD. Look around in this list. Do you see the examples?

Under "Appendix A. Crosswalks" you'll find a conversion table from ISAD(G) to EAD. Have a look in this list.

With this link you reach an example finding aid (from Schiedam).

To edit EAD files you might want to use an EAD editor. ICA Atom is a

ICA-AtOM is a fully web based archival description application that is based on International Council on Archives (ICA) standards. AtOM is an acronym for Access to Memory

Unfortunately in Schiedam we have not been able to implement ICA Atom successfully. Instead of that we use XME-TAL, which is a general XML editor.

3.3 Exercise 3: A(rchives)P(ortal)E(urope)

You can publish your finding aids on the Archives Portal Europe (APE). On the project website APEx you'll find extensive information. If you want to contribute to the Archives Portal Europe contact your "country manager".

After you've been admitted as a contributor you can upload your finding aids in EAD. For testing there in the content checker. When you are satisfied with a finding aid you can publish it with the content manager.

APE allows you to see images in finding aids. So you can show what you have described.

Let's look at a description of pictures in the exemplary finding aid.

```
<c04 level="file"> <did> <unitid>2</unitid>
```

```
<unittitle>Photographs</unittitle> <unitdate normal="1943/1956?">1943-1956</unitdate> <physdesc>4 entities</physdesc>INSERT
```

```
</did>
```

```
<odd> <p>N.B. Only digitally accessible, see picture numbers 34125-34128.
```

```
<lb/> </p> </odd></c04>
```

To show images with your archival description you need to add and element <daogrp> instead of INSERT in the above. Note the crucial hyperlink.

```
<daogrp linktype="extended"> <resource linktype="resource" label="start"/><daoloc href="http://images.memorix.nl/shi/thumb/500x500/e-4653107c-39b0-3035-27ad-44b4e968746e.jpg&#8221; linktype="locator" label="thumb"/><daoloc href="http://gemeentearchief-schiedam.pictura-dp.nl/beeldbank/weergave/record?id=4653107c-39b0-3035-27ad-44b4e968746e&#8221; linktype="locator" label="reference"><daodesc>
<p/></daodesc></daoloc><arc linktype="arc" show="embed" actuate="onload" from="start" to="thumb"/><arc linktype="arc" show="new" actuate="onrequest" from="thumb" to="reference"/> </daogrp>
```

The last thing you do is upload your EAD finding aid to APE. The described item is shown as a thumbnail.

Finding aid with picture in APE

3.4. Exercise 4: Create your own digital collection

In this exercise you are curating a diverse collection of digitised objects. Here are some examples:

<http://gyujtemeny.imm.hu/site.php> (collection database of the Museum of Applied Arts in Budapest)

<https://www.google.com/culturalinstitute/u/0/collection/hungarian-national-museum?projectId=art-project> (Collection of the Hungarian National Museum on Google Art Project)

<http://hungaricana.hu/en/databases/> (Hungarian Cultural Heritage Portal)

Your task is to make a plan for publishing a set of digital objects and give access to a wider public. By doing this, you have to consider the following issues:

- Aim of publishing the collection
- Target group
- Type of collection: thematic collection / “highlights” from your collection / give a complete overview?
- Create your own online interface or use available platforms?

Examine these aspects and make your own plan! According to your plan put your own digital collection together from objects accessible on Google Art Project (user gallery, needs an existing google-account) or Europeana (MyEuropeana).

Read more about these projects here:

<https://www.google.com/intl/hu/culturalinstitute/about/partners/>

<http://pro.europeana.eu/>

4. Exercises

Experiences about publishing content in the Google Art Project (Budapest)

Museum Kiscell has joined the Google Art Project in 2015. The project contains three main parts: first, the Google Street View mapping of the museum building, second, a collection of representative objects from the museum's collections, and lastly, the possibility to compose virtual exhibitions from uploaded items. Work has begun with uploading digital images and metadata collected in a Google spreadsheet (for the method of uploading images and metadata see: <https://support.google.com/culturalinstitute/partners/answer/4618073>). This was a very easy and effective method for us, because the museum is collecting metadata in excel which could be easily converted to a Google-spreadsheet. This way you can import aggregated data to the online interface of Google Cultural Institute. The uploaded content was verified by Google and the revised collection and virtual exhibition is scheduled to be published on the Google Art platform in December 2015.

The Google Art Project is a good way to give access to a wider public and also to introduce our institution to future visitors, especially to tourists from other countries.

Experiences for chapter 8 by Gävle

Experiences (add to chapter 8, Providing Access):

Sofia: In Schiedam, I learned more about archival access on the internet. For example, we had a seminar and workshop about Archives Portal Europe. This internet portal provides access to information on archival material from different European countries as well as information on archival institutions throughout the continent. It was fascinating to see how many finding aids and images from all over Europe that can be searched and found in one place. In Schiedam, I also appreciated the seminars and workshops about Encoded Archival Description (EAD), an XML standard for archival finding aids. It opened up my view of how data and metadata can be standardised in a digital environment.

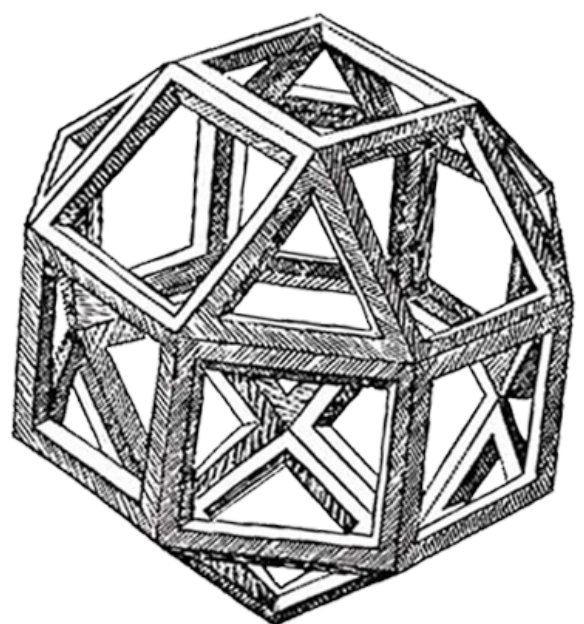
Lessons learned in Gävle

Caroline, Schiedam: During my visit to Gävle I saw a lot of different institutes and met a lot of interesting people from Sweden and from the other countries with whom I spoke about several topics. In Sweden I learned more about digitisation and especially about the problems of digitisation. Apparently, everyone has more or less the same problems. I think that (sadly enough) most institutions are still in the phase of gathering the problems and analysing them rather than thinking of solutions. At the moment, there are not so many solutions and everyone is trying but doesn't know the right answer.

Jan, Schiedam: Another positive point I noticed is what meaningful work can do for people. I felt this at the reintegration project where the negatives of Berna Roth were being processed. The people really took an interest and appreciated the work done by Roth, and in that way felt they were working on a project that wasn't just a job. The same can be said about the employees at the Railway museum.

In my experience it is important to realise that in working with archive material, you're working with pieces made (or processed) by 'real' people and in a small way you are connecting with them.

One thing I maybe missed during the visit is a point which, in my opinion, can be as important as digitisation of cultural heritage, namely the communication to the public. We should think about the best way to let our public know about the process of digitisation. After all, it's their heritage too.



Notes



With the support of the Lifelong Learning
programme of the European Union

