Fundamental principles of digitization of documentary heritage
(information drawn from the non-exhaustive list of sources in the endnote)

Objective
The purpose of this text is to collate in one document, the basic information needed to understand the requirements of digitization. By no means intended to be exhaustive, it seeks simply to provide a primer, especially for developing countries, to guide or inform on digitization. It sets out the most basic elements to be taken into consideration before embarking on the process of digitizing national documentary heritage. The bibliography lists some of the international source documents that are currently available online. These may disappear over time. The sources cited should not necessarily be interpreted as having been endorsed by UNESCO or reflecting the Organization's views.

Definition
Digitization is the creation of digital objects from physical, analogue originals by means of a scanner, camera or other electronic device. It is undertaken as part of a process that includes:
- Selection
- Assessment, including of needs
- Prioritization
- preparation of originals for digitization
- metadata collection and creation
- digitization and creation of data collections
- submission of digital resources to delivery systems and repositories.

This process is accompanied along the way by management, including intellectual property rights management and quality control, and evaluation at the end. These steps are essential to ensure that the digital object remains accessible in the long-term.

Fundamentals
Generally, digitization in itself is not a method of preservation of documentary heritage although it does help to protect precious documents from excessive handling. It allows the preservation of a facsimile of the document (not the document itself) and ensures multiple access, with due respect for intellectual property and other rights, to the content.

Documents with text and images that are to be kept as originals after digitization may be digitized to a higher or lower level of detail. The reason for making the digital copies will define the choice of resolution and bit depth which may have an impact on costs, production flow as well as the long-term preservation.

For audio and video documents, however, digitization is the only viable method for long-term preservation, because, apart from carrier instability, dedicated replay equipment is rapidly vanishing. The time window left for the replay of audio and video originals may only be 15 years, which adds urgency to the situation.

Audiovisual documents must be digitized with appropriate digital resolution, and that capture resolution must equal or exceed the quality of original as, in the long-term, the digital master file will be the only version available.

Film preservation is adopting digitization out of necessity as manufacture of analogue photochemical film is fading.

Digital master files are a facsimile of the original document; they must not be enhanced, restored, or otherwise altered.

Digitized materials should enjoy the same intellectual property copyright protection level as the original.

Primary consideration should be accorded to accessibility and traceability of information when digitizing important

items. This is accomplished through the development of complete and detailed metadata, as well as documentation of the process including technical specifications and resolution settings. Digital conversion and metadata creation should be synchronized as far as possible thus enabling permanent access to all the relevant attributes of the object. Digitization is also not a method for the intellectual property protection of documentary heritage; digitization raises intellectual property issues, which should be identified and managed. Digitization should take into consideration and be done in respect of intellectual property law.

When digitizing material, preference should be given to openly defined formats which are widely used and supported. They should be suitable for long-term preservation have a long life expectancy and be easily migrated as when the need arises. Ideally, multiple copies should be made and stored in professional repositories at different locations and regularly checked.

The digital master that is created should not be enhanced or altered. It should represent the original as closely as possible.

The budget allocation for digitization must allocate funds for long-term preservation. The digitized collection will require permanent management to ensure its safety. It will also have to adapt to new preservation standards and practices that take into account technological developments.

Digitization strategy

Before starting to digitize it is necessary to formulate a digitization strategy. This document should provide a clear description of the purpose and objectives of the exercise; identify its scope (still images, audiovisual materials, manuscripts, etc) and the volume (large-scale digitization, selective digitization); identify strategic approaches (cooperation between institutions, domestic resources, outsourcing); and define methods of digitization and applicable standards.

The strategy may contain also information on the institutional ecosystem that ensures digitization and digital preservation, as well as a sustainable source of funding for these activities. Partnership and engagement should be incorporated as a means of leveraging capacity and support for the digitization strategy whose outcomes should also be presented to the public. The policy which is derived from the strategy should include principles that serve as the conceptual foundation as well as helping to prioritize projects.

Originals shall be maintained and protected under appropriate preservation conditions for as long as they remain usable.

The strategy should also include an intellectual property policy providing guidelines on the management of the intellectual property issues that arise in relation to digitization processes.

Main steps in digitization

Planning; pre-digitization; digital conversion; post-digitization processes.

- The planning process includes:
  o Identification of material to be digitized and rights related thereto.
  o Assessment of resources needed.
  o Decisions on standards.
  o Definition of methods and timing of quality control.
  o Assessment of risks, including current and future drawbacks

- The pre-digitization process includes:
  o The selection of materials to be digitized.
  o Quality control of the objects to be digitized - an assessment of their state of preservation and need of cleaning.

- o Prioritisation of digitization.
  - o Any treatment that may be required or possible.
  - o The collection of metadata (especially descriptive and structural metadata).
  - o Bibliographic and archival preparation.

- Digital conversion includes:
  - o Digitization.
  - o Availability of professional equipment.
  - o Quality control.
  - o The creation of digital masters from which access copies are made.

- The post digitization process includes:
  - o Control of metadata related to long-term preservation.
  - o Submission of information to delivery and repository systems, data collection and management.
  - o Making digitized copies and metadata available online.
  - o Assessment and evaluation of the project.
  - o Quality control.

Intellectual property rights

In the digitization process, it is important to ensure respect for intellectual property rights from several points of view:

- Rights the owners of the original source materials retain in their materials.
- Rights or permissions the collection developers have to digitize content and/or make it available and/or make other uses; and.
- interests or permissions the users of the digital collection are given, to have access to and make subsequent use of the materials..

It is important to ensure that intellectual property and economic or and moral rights, as well as other rights and interests, including the rights and interests of indigenous peoples and local communities in traditional knowledge and traditional cultural expressions, are respected and are not infringed. It is also important to take into account the rights and interests of beneficiaries of limitations and exceptions to copyright and of the general public in regard to access to digitized material in the public domain.

Adherence to standards

Adherence to internationally recognized standards and best practices facilitates interoperability and long-term access to digitized content. Digitization policy should set out the digitization specifications with a view to preservation, access and anticipated use.

Descriptive Metadata

Metadata, in a broader sense than descriptive, plays a significant role in the preservation, provenance, exchange, compatibility, and long term sustainable access of and to digital information. Most of this metadata is created for specialist needs or specific types of materials (technical metadata) or communities of expertise (e.g. linguistic, music, etc). All such metadata should be created in accordance with the technical standards or user community defined guidelines, and all processes of creating, maintaining, and sustaining digital information should be mindful of protecting this data in a usable form.

Digital preservation

This includes the processes needed to maintain access to information and the essential elements that make up digital objects. If access cannot be maintained the information carried by the object is effectively lost. Preservation is ensured through appropriate management practices which may be conducted in-house or may be outsourced. It is essential to identify the various stakeholders involved and their responsibilities for preservation and management along with the procedures to be implemented as well as the associated costs, etc,

Digital Access

Access of digital information should be guaranteed for everyone as a human right. Digital infrastructure and technology should consider digital divided and sustainable accessibility.

Digital repositories

Trusted digital repositories should be established to house digitized resources, ensure their authenticity and long term accessibility.

Economic aspects

On average, digitization will require 1/3 of project time and funding. Organising and describing materials, creating indexing and making the collection searchable by users will require the remaining 2/3rds of the project time and funding. It must be noted, however, that the long-term preservation of digitized objects is a permanent additional cost factor.

Matters of urgency

- Prioritise the digitization of unstable and at risk carries (acetate film, brittle paper etc).
- Transfer of analogue and digital single carrier audio and video contents to digital files.
- (and other matters to be identified).

<u>Digitization Guidelines: a selective list</u>

UNESCO

Guidelines for the preservation of digital heritage:
http://unesdoc.unesco.org/images/0013/001300/130071e.pdf

Risks Associated with the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections - Strategies and Alternatives
http://unesdoc.unesco.org/images/0014/001477/147782E.pdf

Towards an Open Source Repository and Preservation System
http://unesdoc.unesco.org/images/0015/001547/154761e.pdf

External sources

British Library
http://www.bl.uk/aboutus/stratpolprog/digi/digitisation/index.html
Digitization policy

Cornell University Library- Moving Theory into Practice: Digital Imaging Tutorial (In three languages: EN, FR, ESP)
http://www.library.cornell.edu/preservation/tutorial/index.html

National Library of Australia - Digitisation Guidelines
http://www.nla.gov.au/standards/digitisation_guidelines

Contains sections on the following topics
- Image capture standards
  Image capture standards used in collection material digitisation.
- Images for Web delivery - standards
  Standards used in production of derivatives for Web delivery.
- Digital capture equipment
  Capture devices used and guidelines for selection of a device suitable for capture of particular types of collection materials
- Workflows for still image digitisation
  Workflows used by the National Library in its collection digitisation program
- Metadata for images created by the National Library of Australia
  Guidelines for extracting or creating metadata describing images created through the digitisation process.
- Persistent identifier scheme for digital collections at the National Library of Australia
  Outline of the National Library's naming scheme for digital objects.
- Care and handling guidelines
  Guidelines on the care and handling of material to be digitised.

IFLA- Guidelines for Digitization Projects

http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf

International Association of Sound and Audiovisual Archives (IASA)
IASA Standards, Recommended Practices and Strategies

- Principles for the long-term preservation of the audio heritage.
  http://www.iasa -web.org/sites/default/files/downloads/publications/TC03_English.pdf

    Also available in French, German, Spanish, Italian, Swedish, Russian and Chinese
- Guidelines for the production and preservation of digital audio objects
  http//www.iasa-web.org/de/tc04/audio-preservation
- Selection criteria for analogue and digital audio contents for transfer to data formats for preservation purposes
  http://www.iasa-web.org/sites/default/files/downloads/publications/taskforce.pdf

    Under preparation: Guidelines for the production and preservation of digital video objects

FIAF - Some Considerations when Setting Up a Digitization Workflow
- http://www.fiafnet.org/commissions/TC%20docs/Setting%20up%20a%20digitization%20workflow-%20FIAF%20tech%20paper%20V1%201.pdf

Minerva Digitisation Guidelines: a selected list

http://www.minervaeurope.org/guidelines.htm

Federal Agencies Digitization Guidelines Initiative

http://www.digitizationguidelines.gov/

Digital Library of Georgia Digitization Guide

http://dlg.galileo.usg.edu/AboutDLG/DigitizationGuide.html?Welcome

North Carolina Digitization Guidelines

http://www.ncecho.org/dig/digguidelines.shtml

IFLA digitization guidelines (2002)

http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf

University of Virginia library

http://guides.lib.virginia.edu/content.php?pid=40437&sid=297543

Intellectual Property Related Resources

WIPO Guide on Managing Intellectual Property For Museums (an updated version is under preparation)

http://www.wipo.int/copyright/en/museums_ip/

WIPO Database of Existing Codes, Guidelines and Practices

http://www.wipo.int/tk/en/databases/creative_heritage/index.html

WIPO Creative Heritage Project

http://www.wipo.int/tk/en/culturalheritage/

WIPO Publication "Intellectual Property and the Safeguarding of Traditional Cultures: Legal Issues and Practical Options for Museums, Libraries and Archives"

http://www.wipo.int/tk/en/publications/1023.pdf

ISO 13028 - Implementation guidelines for digitisation of records.

ENUMERATE

http://www.enumerate.eu/en/about_enumerate/

***

Other Miscellaneous Sources

 SMU Central University Libraries
- Digitization Guidelines and Procedures - Best Practices for Digitization
  http://www.lib.unc.edu/instruct/artstor/artstor_digitization_guidelines.pdf

Washington State University
http://digitalwa.statelib.wa.gov/newsite/best.htm

National Information Standards Organization (NISO)
- A Framework of Guidance for Building Good Digital Collections 3rd edition
  http://www.niso.org/publications/rp/framework3.pdf

Yale University Library
http://www.library.yale.edu/dpip/documentation/digguides.html
This provides a gateway to the following sites:

- California Digital Library Digital Library Building Blocks - Best Practices and Standards
- California Local History Digital Resources Project Handbook
- New York State Archives Imaging Production Guidelines 2006
- University of North Texas Libraries - Scaiming Standards by Type of Material S
- University of Maryland Libraries - Best Practices Guidelines for Digital Collections
- Western States Digital Imaging Best Practices
- Picture Queensland Image Digitisation Manual 2007
- Picture Queensland Digitisation Resources

Abigail Sattler at Liberty University asked the Archives discussion list for pointers to a good set of standards for digitizing print materials. The group provided her with following list, which I'm publishing here (unedited) with Abigail's permission:
- http://memory.loc.gov/ammem/ftpfiles.html
- http://www.diglib.org/standards/bmarkfin.htm
- http://www.cyberdriveillinois.com/departments/library/who_we_are/bestpractices.html
- http://www.loc.gov/preserv/
- http://dublincore.org/documents/dces/
- http://www.rlg.org/en/pdfs/RLG_desc_metadata.pdf
- http://www.niso.org/framework/Framework2.html

- http://www.bcdlib.tc.ca/tools.html
- http://www.nla.gov.au/padi/
- http://www.knaw.nl/ecpa/grip/
- http://www.cdpheritage.org/digital/index.cfm
- http://www.ncecho.org/Guide/toc.htm
- http://nedcc.org/welcome/pubs.htm
- http://www.ohiomemory.org/om/links2.html
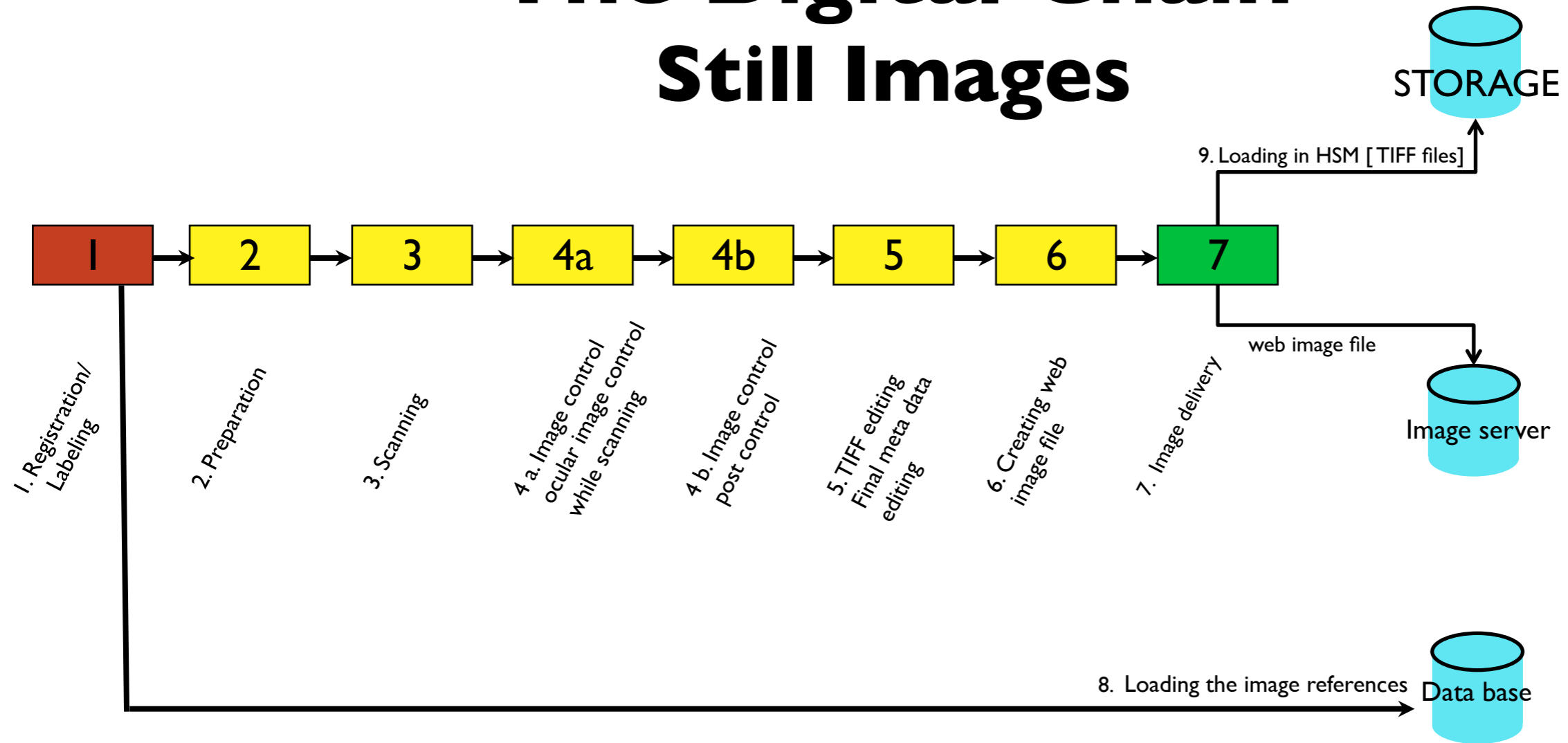- http://www.njdigitalhighway.org/image_requirements_libr.php
- http://images.library.pitt.edu/i/imls/docs/guidelines/Image_Quality.pdf
- http://www.unc.edu/projects/diglib/docs/dig_guidelinesQR011.pdf
- http://digitalcollections.smu.edu/all/cul/docs/culdigitizationguidelines.pdf

• **THINGS TO CONSIDER WHEN PLANNING DIGITIZATION PROJECT:**

• **Project Scope** – define goals and requirements, evaluate user needs, identification and evaluation of options, cost-benefit analysis, etc.

• **Selection** – criteria, identification of best copy, process, approval, etc.

• **Preparation** – archival/curatorial assessment and prep, records description, preservation/conservation assessment and prep, etc.

• **Descriptive systems** – data standards, metadata schema, encoding schema, controlled vocabularies, etc.

• **Project management** – plan of work, budget, staffing, training, records handling guidelines, work done in-house vs. contractors, work space, oversight and coordination of all aspects, etc.

• **Access to digital resources** – web delivery system, migrating files and metadata to web, etc.

• **Legal issues** – access restrictions, copyright, rights management, etc.

• **IT infrastructure** – determine system performance requirements, hardware, software, database design, networking, data/disaster recovery, etc.

• **Project Assessment** – project evaluation, monitoring and evaluation of use of digital assets created, etc.

• **Digital preservation** – long-term management and maintenance of files and metadata, etc.

Taken from NARA's Digitization guidelines.

The Digital Chain
Still Images

STORAGE

9. Loading in HSM [TIFF files]

| 1 | 2 | 3 | 4a | 4b | 5 | 6 | 7 |

web image file

Image server

1. Registration/ Labeling

2. Preparation

3. Scanning

4 a. Image control ocular image control while scanning

4 b. Image control post control

5. TIFF editing Final meta data editing

6. Creating web image file

7. Image delivery

8. Loading the image references

Data base

Originally designed at the Swedish National Archives

The Digital Chain
Sound and Audiovisual Documents

11. Loading to storage environment

Storage (OAIS) (incl. archival backup)

IF OUT-SOURCED
In-house quality control

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

[in practice done together]

Distribution file

Access server

1. Registration/cataloguing/documentation Numbering (PI)

2. Selection of best copy

3. Cleaning carrier restoration and repairs

4. Replay equipment: Selection and Alignment
• Equalisation and decoding
• Corrections for Errors Caused by misaligned Recording Equipment
• Speed assessment and adjustment

5. Replay of original media, including listening

6. Analogue to digital conversion

7. Sound and Audiovisual Editing

8. Recording coding history and preservation Metadata Creation, collection identification metadata

9. Creation of files
• Archival format
• Commercial use formats
• Distribution formats

10. Sound and Audiovisual delivery

Data base

12. Loading Cataloguing information, collection management information and other related data to the collection management systems and databases