Digitisation – Standards

Jeannette Frey, Alexandre Lopes BCU Lausanne

LIBER-EBLIDA Digitization Workshop 2011
The Hague, October 5-7, 2011
Standards... which standards?

- Standards are set by experts (not by users)
- Users may de facto prefer other « standards »
- Standards of today (October 7, 2011)
- Technology changes – standards evolve
- But in most cases... use a standard!
Standard elements of life cycle approach

- Project planning
- Selection of the items
- Preparation for the digitisation process
- Digitization
- Storage of the Digital Master
Life cycle approach and more…

- Metadata, resource discovery
- Publishing on the Web
- Delivery formats
- Go social: reuse and re-purposing
- Get user feedback
- Improve the user experience
- Digitize more!
Standard elements of life cycle approach

- Project planning
- Selection of the items
- Preparation for the digitisation process
- Digitization
- Storage of the Digital Master Material
Project planning

- Use a project method:
  - Prince2 (UK),
  - HERMES (CH)

- recommandation, no standard
Standard elements of life cycle approach

- Project planning
- Selection of the items
- Preparation for the digitisation process
- Digitization
- Storage of the Digital Master Material
Material selection

• Explicit what is the aim of your digital collection! Create a policy (NISO collection principle 1)
Standard elements of life cycle approach

- Project planning
- Selection of the items
- Preparation for the digitisation process
- Digitization
- Storage of the Digital Master Material
Preparation of the materials

- Cataloging (use your library metadata standard, mostly MARC)
- Analyse the physical constrains of the materials (size, opening angle, fragile, unsafe) → scanner types, resolutions
- In-house or outsource
- If in-house, select hardware and software
- If outsource, tender...
Standard elements of life cycle approach

- Project planning
- Selection of the items
- Preparation for the digitisation process
- Digitization
- Storage of the Digital Master Material
Digitisation process

- No standard process
- JISC Digital Media, http://www.jiscdigitalmedia.ac.uk/ guidance
- Document the technical procedure - when it’s being done! not « later »!
Digital file formats: text master

• You should do fine with:
• XML (with the used character encoding recorded in the encoding declaration of XML) conforming to a published Document Type Definition (DTD).
• TEI (Text Encoding Initiative)
• PDF/A (ISO Standard)
Digital file formats: still image master

• **You should be fine with:**
  • TIFF 8 bit greyscale or 24 bit colour, 600 dpi (photographs) 2400 dpi (slides).
  • !! check resolution of negatives and slides first, don’t try to capture more than the useful information!!
  • PNG, GIF, JPEG/SPIFF (Still Picture Interchange File Format), JPEG 2000 if you can’t do TIFF.
Digital file formats: video master

- **You should be fine with:**
  - RAW AVI (uncompressed), highest suitable resolution
  - If you need to use a compressed format: MPEG (MPEG-1, MPEG-2 or MPEG-4)
  - Framesize of 720x576 pixels
  - Framerate 25 frames per second
  - 24-bit colour
  - PAL colour encoding
Digital file formats: audio master

• You should be fine with the uncompressed formats:
  • Microsoft WAV
  • Apple AIFF
  • 24-bit stereo at 48/96 KHz sample rate (guidelines AES and IASA)
• Use MP3 if you must use a compressed format
Standard elements of life cycle approach

- Project planning
- Selection of the items
- Preparation for the digitisation process
- Digitisation
- Storage of the Digital Master Material
Storage of the digital master

- Do always have at least 2 copies!
- At least on harddisk or harddisk of servers
- Have a migration strategy (including for the metadata)
Life cycle approach and more...

- Metadata, resource discovery
- Publishing on the Web
- Delivery formats
- Go social: reuse and re-purposing
- Get user feedback
- Improve user experience
- Digitize more!
Metadata standards

• Metadata about the physical object digitized
• Metadata about the digital object created during the digitization process (digital master)
• Metadata about the delivery formats
• Metadata about new resources created using the digital objects

• **Metadata about the physical object:**
  • Museums: SPECTRUM, CDWA
  • Archives: ISAD(G), ISSAR(CPF), EAD
  • Libraries: MARC
Descriptive Metadata

• For discovery and interpretation of the digital object
• **Dublin Core** (minimum requirement)
• **Europeana Semantic Elements (ESE)** normative specification (24.1.2011):
  • Dublin Core + 13 specific elements: isShownby, country, dataProvider, hasObject, language, object, provider, rights, type, unstored, uri, userTag and year.
Administrative Metadata

• For managing the digital object and provide information about its creation etc.
• Technical metadata
• Source metadata
• Digital provenance metadata
• Rights management metadata
• No single standard!
• NISO draft data dictionary for digital still images (NISO Z39.87-2002 AIIM 20-2002 Data Dictionary)
Preservation Metadata

- 16 basic metadata elements of RLG (1998)
- OAIS model (framework)
- PREMIS data dictionary (LOC project)
Structural Metadata and collection level description

• Describes the logical or physical relationship between the parts of a compound object.

• METS (Metadata Encoding and Transmission Standard)

• IMS Content Packaging Specification for composite learning resources.

• Collection level description: item is part of a collection (scope, format, ownership, restriction of access)

• Dublin Core Collections Application Profile, work of MICHAEL Project
Resource discovery

• Good metadata → interoperability
• For human needs... and for robot needs!
• Provide Dublin Core
• Or DC.Culture schema
• IEEE Learning Object Metadata (LOM) for learning objects
• To share beyond the own community (libraries are a community), go semantic with
• RDF
• SKOS
• OWL
Harvesting, searching, alerting

- OAI-PMH (standard) harvesting
- Z39.50 (standard) searching and retrieval
- SRW/SRU (Search/Retrieve Web Service)
- RSS specifications
Life cycle approach and more…

- Metadata, resource discovery
- Publishing on the Web
- Delivery formats
- Go social: reuse and re-purposing
- Get user feedback
- Improve the user experience
- Digitize more!
Publishing

- **MINERVA** quality principles for quality cultural websites (guidelines)
- Accessible using Web browser, use HTML or **XHTML and HTTP 1.1 protocol (standard)**
- Accessible through mobile devices (use HTML/XHTML, Cascading style sheets (CSS) and Document Object Model (DOM) (recommandation of the W3C)
- Accessible for desabled: W3C Web Accessibility Initiative (WAI) recommandations
Life cycle approach and more...

- Metadata, resource discovery
- Publishing on the Web
- Delivery formats
- Go social: reuse and re-purposing
- Get user feedback
- Improve the user experience
- Digitize more!
Delivery formats (1)

- **Identification**: Uniform Resource Identifier (URI) (w3c standard), OpenURLs (standard ANSI-ISO Z39.88-2004), DOI.

- **Document format**: XHTML 1.0 or HTML 4 through use of SGML or XML conforming to other Schemas, PDF

- **Still images**: JPEG/SPIFF (standard)
Delivery formats (2)

- **Video for download:** MPEG-1, AVI, WMV or Apple Quicktime
- **Video for streaming:** Microsoft Advanced Systems Format (ASF), WMV or Apple Quicktime format.
- **Audio for download:** MPEG Layer 3 (MP3 - standard), RealAudio (proprietary) or Microsoft WMA, 256 Kbps for near-CD quality.
- **Audio for streaming:** MPEG Layer 3 (MP3), RealAudio (RA) or WMA.
Life cycle approach and more...
Go social: reuse and repurposing

• Use of open standards can make it easier for 3rd parties to re-use materials
• Especially in the field of learning objects: IEEE Learning Object Metadata (LOM) standard
• OAI-ORE effort
A word on digitization of out-of-commerce works

- Copyright last 70 year after death of author
- If you want to digitize out-of-commerce works, consider the new Memorandum of Understanding: Key Principles on the Digitisation and Making Available of Out-of-Commerce Works, recently signed by LIBER and EBLIDA in Brussels.
Tools

- Plenty of them, impossible to make a tour!
- Software tools (see for example IMPACT tools for text recognition, or DocWorks)
- Decision support tool (see IMPACT tools for strategy)
- Workflow tools using XML and SOAP, Goobi
- Professional tools related to hardware (with scanners)
Tenders…

• Eager to write a precisely formulated tender?