DIGITIZATION PRACTICAL CONSIDERATIONS

By

Bright Joshua

e-mail: bjoshuamail@gmail.com

Mobile: +265999699983

Overview

- Introduction to E-records Management
- Aspects of Digitization
- Digitization project management
- Digitisation Needs Assessment, Sorting & Prioritization
- Digitization practical considerations
- Digitization techniques

What is a record?

- A document created or received and maintained by an agency, organization, or individual in pursuance of legal obligations or in the transaction of business

What is an Electronic Record?

- An electronic record is information recorded by a computer that is produced or received in the conduct of official business.

Types of e-records

- Text
- Images
- Data
- Voice
- Graphic

Creation of e-records

- Personal computers
- Shared computer servers
- Shared servers with centralised control
- Shared servers using EDRMS
- Through Digitization

Note:

- When there is little or no control over how electronic records are created and used, inconsistency can lead to difficulties finding and retrieving information.

- More control leads to more consistency, but it is important to ensure that the users' needs are well Served. *(IRMT, module 3)*

Attributes of e-records

- **Content**
  - What the record says.
- **Context**
  - Appearance and structure of the content
- **Structure**
  - Information about the creator and the purpose of creation

Qualities of E-records

Accurate

Reliable

Authentic

Trustworthy

Advantages of keeping e-records

- One document can be accessed by multiple users at a time.
- Saves storage space.
- Provides easy retrieval and backup
- Cost effective in term of saving tonner and stationery costs
- Allows instant sharing regardless of location

Disadvantages of keeping e-records

- Undetected change, alteration or deletions.
- Technology obsolescence
- Disasters
- Hacking
- Theft
- Viruses/ Malware/ Trojan horses

Digitization

What is Digitization?

- Digitization is the creation of digital objects from physical, analogue originals by means of a scanner, camera or other electronic device.

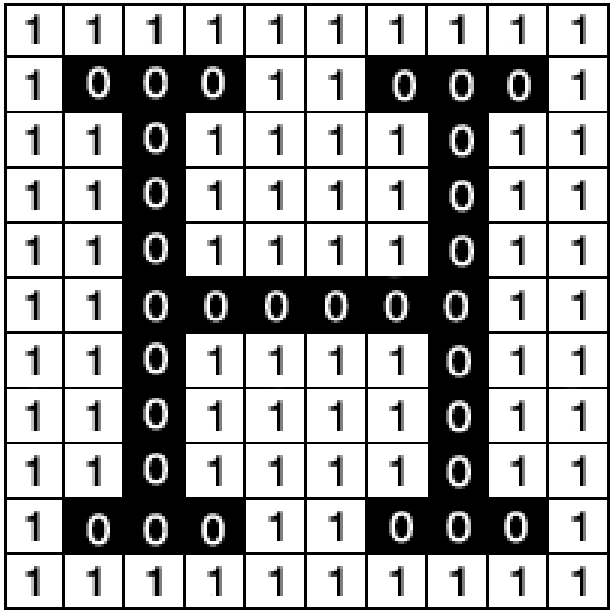
What are digital images?

- Digital Images are electronic snapshots taken of a scene or scanned from documents, such as photographs, manuscripts, printed texts, and artwork. *Cornell University Library (2009)*

- The digital image is sampled and mapped as a grid of dots or picture elements called pixels.

- Each pixel is assigned a tonal value (black, white, shades of gray or color), which is represented in binary code (zeros and ones).

- The binary digits ("bits") for each pixel are stored in a sequence by a computer and often reduced to a mathematical representation.
- The bits are then interpreted and read by the computer to produce an analog version for display or printing.



- **Pixel Values:** As shown in this bitonal image, each pixel is assigned a tonal value, in this example 0 for black and 1 for white

What is Digital Imaging?

- Digital imaging or digital image acquisition is the creation of digital images, typically from a physical scene.

- The term is often assumed to imply or include the **processing**, **compression**, **storage**, **printing**, and **display** of such images.

- Done using digital camera and scanning.

Digital image processing

- This is the use of computer algorithms to perform image processing on digital images.

- As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing.

- It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing.

Compression

- This is the reduction of image file size for storage, processing, and transmission.

- The file size for digital images can be quite large, taxing the computing and networking capabilities of many systems.

Image Compression Types

(a). lossy

(b). lossless

Lossless & Lossy compression

- Lossless compression, abbreviates the binary code without discarding any information, so that when the image is "decompressed" it is bit for bit identical to the original.

- Lossless compression is preferred for archival purposes due to no loss of fidelity as opposed to lossy compression.

- Lossy methods are especially suitable for natural images such as photographs in applications where minor loss of fidelity is acceptable to achieve a substantial reduction in bit rate

Why digitize?

- To facilitate access
- For preservation
- To prevent loss of information due to
    - Obsolescence
    - Deterioration
    - Handling
    - Theft or destruction

Why digitize cont'd?

- For exhibits, publications, and web use
- For researcher/patron requests
- To support current business processes

What Does it Involve?

- Selection
- Needs Assessment
- 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.

DIGITIZATION PROJECT MANAGEMENT

What is Project Management ?

- **Project management** is the discipline of initiating, planning, executing, controlling, and closing the work of a team to achieve specific goals.

Project Planning

- Clearly articulate the goals and deliverables of the project.
- Conduct formative evaluation to validate the initial goals and deliverables of the project.
- Identify what work needs to be done to accomplish the goals and deliverables of the project.
- Break down the work into manageable sub-tasks, and identify dependencies between the sub-tasks.
- Estimate and allocate the time and resources required to successfully complete each sub-task.
- Create a project plan that includes an estimated timetable for the completion of the sub-tasks and estimates the resource requirements for each sub-task.
- Identify key milestones and deliverables in the project.

Project Implementation

- Once the project has begun, monitor completion of tasks, sub-tasks, and milestones on the project plan.

- Regularly review and update the project plan as new and more detailed information about scheduling and resource allocation becomes available.

- Conduct additional formative evaluation to revalidate and, if necessary, modify the project's goals, deliverables, and the project plan.

Project Review

- After the final milestone in the project has been reached, review and document the project's progress, and identify any changes that were required to the project plan, goals, or deliverables.

- Conduct summative evaluation to determine the success of the project.

- Articulate the findings of the summative evaluation in a report that captures the lessons learned from the project.

Qualities of Good Digitization Projects

1. **Substantial design and planning**
   - Users functional requirements identified
   - Written project plan in place

2. **Appropriate level of staffing**
   - In-house
   - Outsourcing
   - Collaboration

3. **Evaluation component**

- To identify and refine project goals
- Assess progress toward project goals
- Determine the quality of project results
- Measure impact of the project
- Show accountability

4. Markets itself

- Disseminates information about its progress and outcomes
- Communicates its activities and broadcasts the availability of its deliverables as widely as possible

5. **Considers the entire lifecycle of the digital collection**

- A plan for maintaining master objects to ensure their persistence over time should be developed
- Investment in the collection's long-term management and sustainability
- Periodic checking of objects, regardless of storage medium, should be periodically checked for accessibility and usability

Note:

- The long-term archiving and preservation of digital materials is a difficult and expensive undertaking that requires substantial resources and serious institutional commitment.

Digitisation Needs Assessment, Sorting & Prioritization

Needs Assesment

- Helps in determining the resources required to undertake the project.

- At this stage, the nature, condition, volume and time to digitize the collections is determined.

- Involves consultation with relevant stakeholders within and outside the organization to decide on what should be done to reach the project goals.

Sorting & Selection

- Sorting involves organizing collections into logical order.

- This is done by arranging all records into their respective functional/ subject series, volume numbers, and the dates covered.

- In a situation where records are disorganized, sorting includes classification of records into appropriate series using an approved filing system (filing index) for the organization.

- During sorting, a **Records Retention and Disposal Schedule** tool is used to identify collections that are due for disposal so that they should not be selected for digitization

- All collections deemed due for disposal should be listed down.

Prioritization

- Strategic records
- Physical structure
  - deterioration rate
  - Fragility
- Most accessed collections
- On-demand

DIGITIZATION PRACTICAL CONSIDERATIONS

Digitization Practical Considerations

- Collections
- Objects
- Metadata

Collections

- A digital collection consists of digital objects that are selected and organized to facilitate their discovery, access, and use. Objects, metadata, and the user interface together create the user experience of a collection.

Qualities of Good Collections

1. Explicit
2. Well described
3. Curated
4. Broadly available
5. Respects intellectual property rights
6. Mechanisms for monitoring usage
7. Interoperable
8. Integrates workflows
9. Sustainable

Explicit

- digital collection is created according to an explicit collection development policy that has been agreed upon and documented before building the collection begins.

- Collection builders should be able to refer to the mission statement of their organization and articulate how a proposed collection furthers or supports that mission.

Well described

- Collections should be described so that a user can discover characteristics of the collection, including scope, format, restrictions on access, ownership, authenticity and integrity.

- Collection description is a form of metadata that serves two purposes:
  - help people discover the existence of a collection
  - help users understand what they are viewing

- **Good collection description contains**
  - scope and extent of the collection
  - names and contacts of the creating agency
  - terms and conditions of use
  - restrictions on access
  - copyright status of collection
  - contact points for questions and comments

Curated

- Curation is concerned with the lifecycle management of a resource from the time it is created or obtained until it is purposely disposed of.

- Curation encompasses a set of activities that include active **data management**, **archiving**, and **digital preservation.**

Broadly available

- This principle encompasses three attributes:

Availability

Usability

Accessibility

Respects intellectual property rights

- The collection should incorporate the organization's copyright policy.

- Intellectual property rights must be considered by responding to these questions:
  - what rights the owners of the original source materials retain in their materials**?**
  - what rights do the collection developers have to digitize content and make it available**?**
  - What rights are the users of the digital collection given**?**

- Fundamental treaties governing international copyright are:
  - **Berne Convention (Paris, 1971)-** *for the Protection of Literary and Artistic Works.*
  - **WIPO Copyright Treaty (WCT) (Geneva, 1996)** – *recognizes computer programs as copyright protected literary works*
  - **WIPO Performances and Phonograms Treaty (WPPT) (Geneva 1996)** – *rights of performers and producers for authorizing the fixing of performances and phonograms and making them available to the public*

Mechanisms for monitoring usage

- Digital collections should be evaluated periodically to monitor usage, assess service effectiveness, demonstrate return on investment, inform collection development, inform strategic planning, and support funding requests.

- To obtain a clear picture of the value of a digital collection is to answer the question:

  “Who is using what, how, and why?”

Interoperable

- Collection services should share their metadata with external search engines.

Integrates into users workflows

- A good collection integrates into the workflows of staff and end users.

- This allows:
  - The operational staff performing these activities to be the best advisors as to how to make them more intuitive and less time consuming

  - Users to contribute without significant additional effort and use a digital collection more comfortably if they can access it from a familiar environment

Sustainable

- Digital collections containing resources of long-term value should be sustained and archived permanently to ensure access.

- There must be a clear understanding of the long-term **obligations** necessary to ensure a sustained digital collection.

- There must be a clear understanding of the long-term obligations necessary to ensure a sustained digital collection.

OBJECTS

Objects

- Digital materials created after digitization

Qualities of Good Objects

- i. Internationally accepted format
- ii. Preservable
- iii. meaningful and useful
- iv. Have a global unique identifier
- v. Authenticated
- vi. Associated metadata

Internationally accepted format

- A good object exists in a format that supports its intended current and future use.

- Should be exchangeable across platforms, broadly accessible, and formatted according to a recognized standard

Internationally accepted formats examples

CATEGORY	Format	Requirements
Printed matter and manuscripts -	Uncompressed TIFF or Lossless compressed JPEG2000 **Access**- PDF or JPEG	300-600 ppi @ 8 or 24 bits per pixel
Pictorial materials	Uncompressed TIFF or Lossless compressed JPEG2000 **Access**- JPEG	300-600 ppi @ 8 or 24 bits per pixel
Sound recordings (Music or Speech)	WAVE or AIFF **Access**- MP3, windows Media, QuickTime	Sampling Rate 44,100 KHz 16 bits per channel
Analog, videotapes & DVDs	MPEG-2 or MPEG-4	

Note:

- Formatting depends on the institution

- If possible encourage authors-
  - To create works in specified formats
  - Include rich metadata
  - Unlock associated digital rights management

Preservable

- An object should not raise unnecessary barriers to remaining accessible over time despite changing technologies.

How do we ensure access of these objects?

THROUGH:

Digitization

Migration

Emulation

Migration

- Involves transforming objects so they can move between technical regimes as those regimes change

- Occurs at all levels, as objects are moved across
  - Media as media evolve (CD to DVD);
  - Software products as the products become outmoded (Old spreadsheet to new version)
  - Formats or encodings as new standards emerge (JPEG to JPEG2000).

Emulation

- Involves reproducing on contemporary systems the computer environment in which digital objects were originally created and used.

- Particularly appropriate for executables and complex multimedia objects such as interactive learning modules.

Note:

- Migration and emulation should be seen as complementary approaches. Neither is appropriate for all types of materials.

- If possible, choose formats that are non-proprietary and do not contain patented technologies.

- Formats that are widely used and have published specifications are most likely to have migration paths.

Meaningful & Useful

- Objects should be coherent, meaningful, and usable outside of the context in which it was created.
  - Metadata should be self-contained
  - Format and any technical requirements necessary for its use should be readily apparent.
  - Must carry with it a clear statement of acceptable users and uses.

Have a global unique identifier

- Objects should be named with a persistent, globally unique **identifier** that can be resolved to the current address of the object

- An identifier is a name assigned to an object according to a formal standard, an industry convention, or a local system providing a consistent syntax.

- Identifiers should be:
  - Scalable
  - **Consistent**
  - Actionable
  - **Persistent**
  - Conform to international standards

Authenticated

- Authentication is the act of determining that the object conforms to its documented origin, structure, and history, and that the object has not been corrupted or changed in an unauthorized way.

- Authentication is done using

  Documentation of digital provenance

  Watermarking

  Fixity checking or checksums

Associated metadata

- Objects should have structural metadata to document the relationships between components of the object and ensure proper use of the components.

META-DATA

WHAT IS METADATA?

- Metadata is structured information associated with an object for purposes of discovery, description, use, management, and preservation.

“Data about Data”…

Types of Meta-Data

- **Descriptive metadata**
- **Administrative metadata**
- **Structural metadata**

Descriptive metadata

- Helps users to:
  - find and obtain objects
  - distinguish one object or group of objects from one another
  - discover the subject or contents

Administrative metadata

- Helps collection managers to:
  - keep track of objects
  - rights management
  - Preservation

Structural metadata

- Helps to:
  - Build relationships between documents & among objects
  - enable users navigate complex objects

Metadata Building Self questions

- What is the purpose of the digital collection?
- Where will the metadata be stored (embedded or in external system or both)
- What are the goals and objectives for building this collection?
- Who are the targeted users?
- Are the materials to be accessed at the collection level or as individual items, or both?
- Do multiple versions or manifestations of the object need to be distinguished from each other?
- Will the collection or its objects have metadata before the digital collection is built?
- What metadata standards are you going to use?
- How rich a description is needed?

Qualities of Good Metadata

- Supports interoperability
- Uses authority and content standards
- includes a clear statement of conditions and terms of use
- supports long-term access, curation and preservation
- Determines relationship to identifiers of digital objects

SECURITY OF E-RECORDS

Security of E-records

- Disaster preparedness plan developed
- Backup and recovery of records to protect against information loss
  - A minimum of 3 copies
  - Stored at 3 different locations
- Servers stored in a well secured purposefully built room

Security of E-records (Cont'd)

- Security should aim to minimize unauthorized addition, modification, alteration, erasure of records.
- Unauthorized access should be prevented at all cost.
- ERMS should produce logs
- **Strong** passwords should be assigned
- Develop e-records security policy

Security of Digital Records (cnt'd)

- levels of user access and rights should be defined
- Firewall and Antivirus kept up to date
- Personnel trained in safeguarding sensitive or classified electronic records
- Write procedures for the care and handling of direct access storage media

Security of Digital Records (cnt'd)

- Prohibit smoking and eating in the storage building
- Avoid the use of magnetic media not specifically designed for purposes of long term storage

Selection of Storage Medium

Consider:

- The authorized life of the records, as determined by the appropriate record series.
- The maintenance necessary to retain the records
- The cost of storing and retrieving the records
- The records' density
- Portability of the medium
  - Should run on multiple equipment
  - Easy transfer information

Retention & Disposition of E-records

Retention & Disposition of e-records

- Retention and disposal rules determine how long should records be kept and how to dispose of them.

- The retention and disposal schedule shows the specified retention timeframe as approved by management.

Importance of Retention & Disposal Schedules

- Reduces the chances of destroying valuable records or retention of wrong records
- Provides evidence for transparency and accountability during the disposal process
- Provides a pre-defined criteria and clearly articulated processes for destroying records.

How to appraise and destroy records

- Have your records retention schedule officially approved by senior management
- Consider special circumstances which may lead to the delay or stop destruction
- Ensure *all* copies of records scheduled for destruction are destroyed
- Ensure records are destroyed in a confidential and non-recoverable manner
- Ensure a documentary audit trail is in place to record processes from selection through to confirmation of destruction.

DIGITIZATION PRACTICALS

Digitization Techniques

- Documents
- Film
- Images
- Sound