

Digitizing Archives

Karen Coyle

April, 2008

Planning

Planning before digitizing is the key to
success

Step 1: Selection

- ◆ Value of the content (historical, current scholarship)
- ◆ Usability of digital version (e.g. search capabilities)
- ◆ Nature of the collection (can it be partially digitized?)

Potential users

- ◆ Faculty, researchers
- ◆ Schools
- ◆ Government and business
- ◆ Scholars abroad

Potential uses

- ◆ Is there sufficient context for study?
- ◆ Will the files be usable?
Will the formats be ones scholars can work with?

Material types

- ◆ print and manuscripts (text, maps, etc.)
- ◆ photographs and artworks
- ◆ sound, video, film

Different material types
have different challenges
and costs

Portioning

- ◆ If a whole collection is very large, is there a portion that could represent it?

Retrieval capabilities

- ◆ Is there already sufficient metadata to assure retrieval?
- ◆ If no metadata exists, does the library have the skills to create it?
- ◆ What other types of retrieval are possible for this data type? (OCR and Keyword indexing)

Delivery capabilities

- ◆ Are there legal barriers to open access? (©)
- ◆ Are there technical barriers? (e.g. bandwidth)

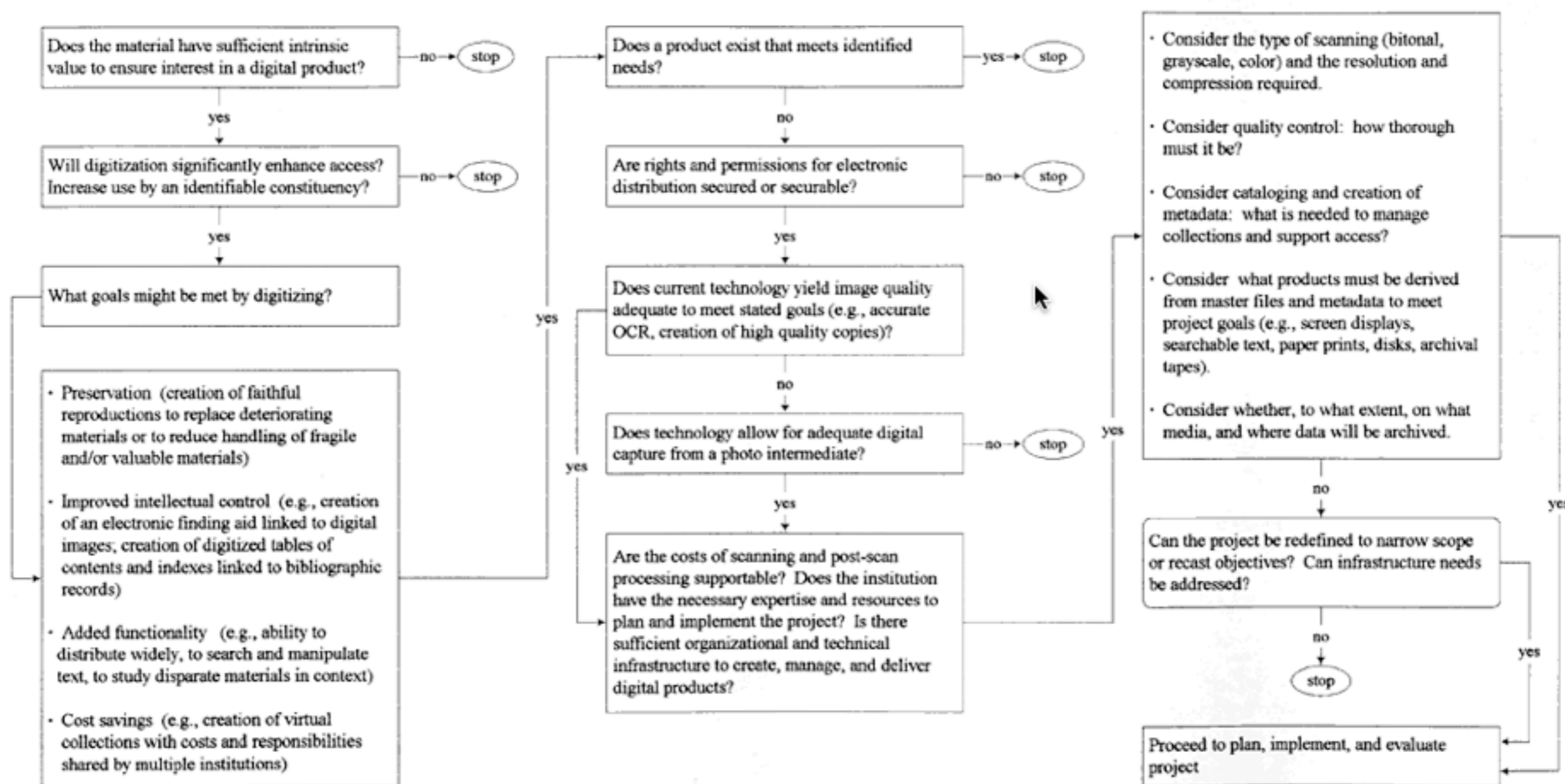
Trade-offs

- ◆ Digitize more at a lower quality
- ◆ Digitize less but at a higher quality

Digitization context

- ◆ Rapidly deteriorating materials
- ◆ One-time funding
- ◆ 'There will always be another day'
- ◆ Other digitization projects
- ◆ Possibilities for cooperative effort

SELECTION FOR DIGITIZING: A Decision-Making Matrix



SChapman, DHazen, JBarrell, JMerrill-Oldham, 5/12/97
 OHarvard University Library

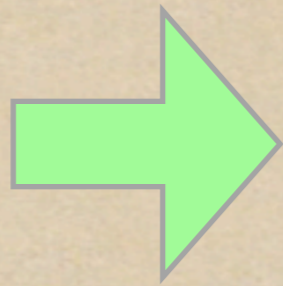
Finally....

Step 2. Digitization Workflow

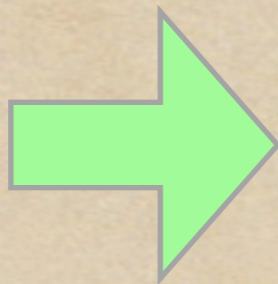
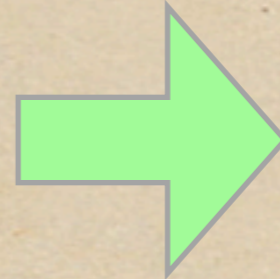
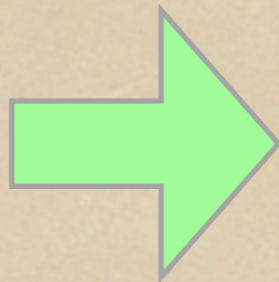
Workflow



Workflow

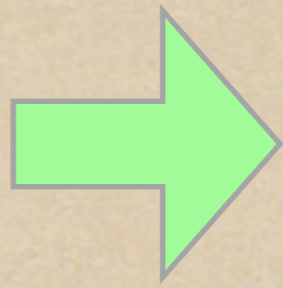


Digitization



Documentation

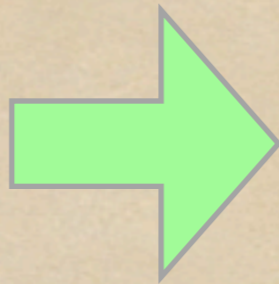
Workflow



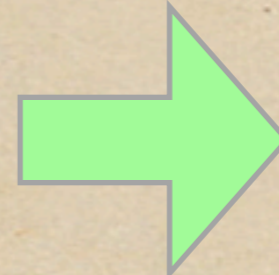
Digitization



Original



Digital copy




Digital access and
preservation

Digitization Workflow Software

CONTENTdm [OCLC - Digital Collection Management]

http://www.oclc.org/us/en/contentdm/default.htm

UMLet 7.1 free download. UML... EIFL: eIFL-FOSS CONTENTdm [OCLC - Digital C...

 **OCLC™** The world's libraries. Connected.

United States (English) [Change](#) [About OCLC](#) [News and Events](#) [Contact us](#)

Home WorldCat Community and Membership Products and Services Programs and Research

◀ Digital Collection Management


CONTENTdm

- At a glance
- Overview
- Success stories
- Take a tour
- Free evaluation
- Ordering
- Support, training and documentation

[View recent Web sessions](#)

[Subscribe to OCLC updates and offers](#)

CONTENTdm™
Digital Collection Management



AT A GLANCE

- Easy to use—Templates, tutorials, comprehensive help files, batch processes and customizable out-of-the-box Web interfaces help you build and post collections efficiently.
- Standards-based—CONTENTdm supports numerous industry standards including Dublin Core 1.1, VRA, XML, HTTP and OAI.

[View complete at a glance >>](#)

[Download the brochure](#)

NEWS

- [CONTENTdm Featured Collections: March 2008](#)
- [CONTENTdm Featured Collections: February 2008](#)
- [CONTENTdm Featured Collections: January 2008](#)

[More news >>](#)

Digitization Workflow Software

“Digitization Assets Factory” (DAF)
Bibliotheca Alexandrina

<http://wiki.bibalex.org/DAFWiki/>


Main Page - DAFWiki

http://wiki.bibalex.org/DAFWiki/index.php/Main_Page

contentdm

Main Page - DAFWiki

Main Page - DAFWiki



Digital Asset Factory

navigation

■ Main Page

■ Key Features

■ Downloads

■ Recent changes

■ Random page

■ Forum

■ Publications

■ Contact Us

downloads

■ Binaries

■ Source Code

Done

article

discussion

view source

history

Main Page

About DAF

DAF v2.0 provides all the necessary tools required to manage the whole process of a digitization workflow, including its various Phases, User management, file movement and archiving. It provides the flexibility to manage multiple simultaneous projects with a diversity of materials, covering books, journals, newspapers, manuscripts, unbound materials, audio, video, and slides.

The system allows easy integration of any tool used to perform functions of the workflow, such as the OCR, image processing, etc. It can be integrated with the current tools used at your organization.

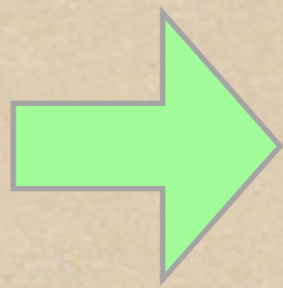
DAF v2.0 is highly reliable and can be configured for large and challenging digitization projects. The system is fully deployed in BA digitization laboratory and has been working smoothly for over a year.

Getting started

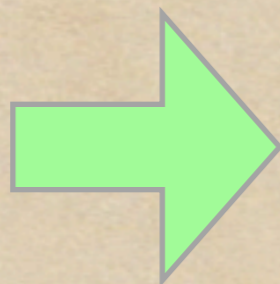
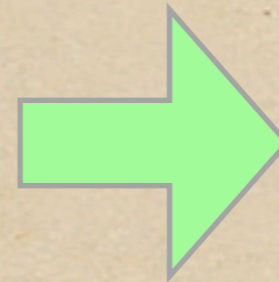
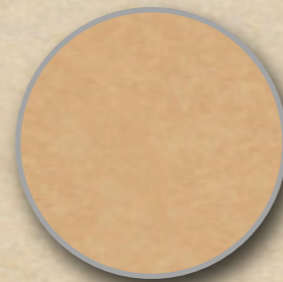
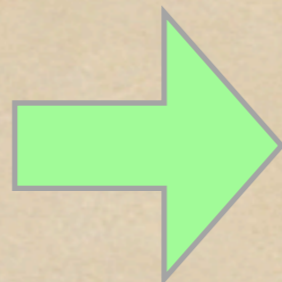
Documents

- [Java Doc](#)
- [Data Model](#)
 - [ERD](#)
- [DAFv2 Documentation](#)
- [DAFv2 User Manual](#)
- [Publications](#)

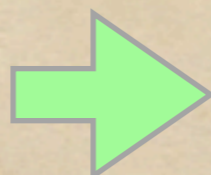
Workflow



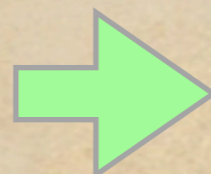
Digitization



Documentation



The original resource



The digital object

documenting original

- ◆ decide goals for metadata (identification, retrieval...)
- ◆ determine level of cataloging
- ◆ use identifiers when possible
- ◆ create metadata (cataloging) as early as possible in the process so it can be used during digitization steps

Documenting digital copy

- ◆ Document technical aspects
- ◆ Use standards where possible

Very complex!

Some standards

- ◆ Digital Image Collection Standards,
University of California, Digital Library

<http://www.cdlib.org/inside/diglib/guidelines/bpgimages/>

“These guidelines primarily define approaches for creating digital master files to facilitate affordable reprocessing.”

Some standards

- ◆ PREMIS, Preservation Metadata, Library of Congress

<http://www.loc.gov/standards/premis/>

“A data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials.”

Some standards

- ◆ METS, Metadata Encoding and Transmission, Library of Congress

<http://www.loc.gov/standards/mets/>

“...a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library...”



IFLANET

International Federation of Library Associations and Institutions

[Search](#)
[Contacts](#)

Electronic Collections

SITE SEARCH

Google™

Search

☒ IFLA
 ☐ www

IN THIS DOCUMENT:

[Collections](#)
[Bibliography](#)
[Periodicals](#)
[Conferences](#)
[Organizations](#)
[Projects: General](#)
[Projects: Americas](#)
[Projects: Europe](#)

SUGGESTIONS TO:

IFLANET@ifla.org



Home

DIGITAL LIBRARIES: Resources and Projects

COLLECTIONS

The World Digital Library will make available on the Internet, free of charge and in multilingual format, significant primary materials from cultures around the world, including manuscripts, maps, rare books, musical scores, recordings, films, prints, photographs, architectural drawings, and other significant cultural materials. The objectives of the World Digital Library are to promote international and inter-cultural understanding and awareness, provide resources to educators, expand non-English and non-Western content on the Internet, and to contribute to scholarly research.

<http://www.worlddigitallibrary.org/project/english/>

Charles Templeton collected a wide range of music memorabilia. Included among his treasures is a collection of some 22,000 pieces of sheet music from late nineteenth and early twentieth century America. *The sheet music illustrates a broad spectra of music genres, from the ragtime of Scott Joplin to the dixieland of W. C. Handy to the smooth ballads of Irving Berlin to the stirring patriotic anthems of John Phillips Sousa and George M. Cohan to the early roots of big band sounds.*

<http://library.msstate.edu/ragtime/>

BIBLIOGRAPHY

Anderson, Elizabeth, et. al.

Digitizing Legacy Documents: A Knowledge-Based Preservation Project. *Fermi National Accelerator Laboratory, September 1998.* PDF:

<http://www.ifla.org/documents/libraries/net/tm-2056.pdf>

Andersson, U.

Short version of the Sesam report: philosophy and rules concerning electronic archives and authenticity. *In: Proceedings of the DLM-Forum on Electronic Records*

The End

Questions?