

Build It, Share It, Keep It Safe: Using Open Source Tools to Create Digital Archives

John Sarnowski
Director, The ResCarta Foundation
313 15th Ave N, Onalaska, Wisconsin 54650
+1 (608) 566-5966
john.sarnowski@ResCarta.org

ABSTRACT

This is an introductory tutorial on the use of open source and free software to create, validate, index, search, display and maintain a digital archive of various materials including photographs, full text searchable books and searchable oral histories.

Attendees are encouraged to bring sample digital files (Tiff, PDF, and WAV) that they might use in their own use in their own digital collections so they can start to use these tools.

1. TUTORIAL LEVEL

Introductory level

2. DURATION

Three hours

3. OUTLINE OF THE CONTENT

The tutorial will cover the following topics:

Introduction to Creating Digital Collections: The introduction will give an overview of the processes required to create, maintain, and display digital collections.

File Formats: An overview of TIF, PDF, JPG, WAV formats and variations.

Metadata: Metadata standards of MODS/METS/MIX/AudioMD from the Library of Congress will be discussed and created for text, photographs and audio sources.

Data Conversion: The processes creating and converting data are complex. These issues will be discussed in detail, and a reasoned, simplified approach will be provided.

Textual Metadata and Transcriptions: Optical Character Recognition (OCR) for textual materials and Automatic Audio Transcription (AAT) software will be demonstrated and explained.

Collection Management: Digital objects will be created and collected using XML based metadata and storage structures. The archive created will be validated digitally.

Web Servers and Websites: A webservice and a web application will be created and configured to search and display the digital objects.

OAI/PMH: Finally, the archive created will be exposed to metadata harvesting by the creation and configuration of an OAI/PMH server.

4. INTENDED AUDIENCE

The intended audience of this tutorial includes professionals generating digital materials from analog sources who prefer to use open standard formats for storage. Professionals include librarians, archivists and researchers who are interested in the creation of open, sharable archives without the need for complex, costly computer systems.

5. EXPECTED LEARNING OUTCOMES

The tutorial participants will become acquainted with

- Preparing materials for digitization
- Digitizing at certain resolutions and bit depths and the consequences of those choices
- Developing documented workflows and their importance
- Adding metadata to image and audio files
- Creating full-text searchable documents and transcriptions of audio content.
- Avoiding Conversion SNAFUs
- Setting up a basic web server with security
- Creating and modifying a web application to expose an archive
- Building and configuring an OAI-PMH Server
- Validating the contents of an archive over time

6. BIOGRAPHY OF THE PRESENTER

John Sarnowski is a director of The ResCarta Foundation, Inc. with more than 20 years of experience in building digital collections. He was responsible for creating millions of digital objects for learned societies, libraries and major corporations as the director of Imaging Products at Northern Micrographics. Projects included "The Making of America", JSTOR, and Historic Pittsburgh. He is a member of ALA, WLA and IS&T.