

## **WRLC Digital Collections Production Center (DCPC) Guidebook**

The DCPC Guidebook provides information to WRLC libraries for creating digital collections through WRLC's Digital Collections Production Center (DCPC).

### **DCPC Services**

The mission of the DCPC is to promote the development of digital collections among WRLC libraries by providing a shared facility to consolidate project management, information technology, and digital conversion expertise in a production environment. The DCPC provides staff and systems to support five primary services, in conjunction with library staff:

- plan and manage digitizing projects
- design digital collections
- design and enter descriptive and other metadata,
- scan or otherwise input materials according to specifications
- convert machine-readable finding aids if available.

The DCPC is particularly suited to working collaboratively with libraries on small to medium-sized digitizing projects. These projects often include a relatively high degree of internal variety within the collection, requiring individualized handling and interpretation (consistent with the WRLC digital library framework). The DCPC's current in-house scanning operation is best-suited for conversion of the most common types of special collections materials: single-leaf flat text, graphic materials, black-and-white or sepia photographs, and slides.

DCPC staff can also work with libraries on projects where it may be more appropriate to outsource the digital conversion to a service bureau, either because the materials are well-suited to batch processing or because they require specialized equipment (materials such as microforms, audio or video, or very large quantities of printed matter). In such cases, DCPC can still provide the project management, design, and descriptive metadata services, and can load the resulting files into the ALADIN digital library.

### **Developing digital collections**

Developing digital collections for the ALADIN digital library requires efforts of the owning library and WRLC's library services, DCPC, and information technology departments.

Staff at the owning library perform these activities:

- Select collections for digitizing projects
- Obtain copyright permission as necessary
- Work with DCPC staff to define the project plan and schedule

- Provide descriptive information about the project collection and items, in electronic format if readily available
- Prepare physical items for transport to WRLC
- Review the preview site for the digital collection once developed by WRLC
- Create a collection level record in the WRLC online library catalog
- Maintain and enhance metadata on an ongoing basis.

WRLC's Library Services and DCPC staff perform these activities:

- Work with the library to define the project plan and schedule
- Design the digital collection header and search options based on characteristics of the collection
- Test and select options for scanning or other input methods
- Scan or otherwise input the selected items
- Design metadata template in conjunction with library staff
- Create or convert descriptive and administrative metadata for individual digital objects
- Develop a preview site for review by library staff
- Provide access and instructions for library staff to edit and enhance the metadata records as desired.

WRLC's Information Technology staff perform these activities:

- Maintain information technology resources (hardware, software, network) and staff to develop and operate the digital library
- Store digital object files and metadata on WRLC servers and provide access for library staff to retrieve the master files from the server
- Provide digital collection files on CD if requested
- Maintain digital collections for long-term use, including storing metadata and digital objects in formats which support migration to future technologies

### **Selecting collections**

There are several different kinds of related collections which play different roles and have different characteristics:

- Archival (physical) collection: original material housed in library special collections or archives department, made available onsite for use by researchers
- Digital collection: digital representation of some or all items in the archival collection, made available for online searching and display of individual digital items
- Online exhibit: a distinct subset of digital items selected and presented online, usually with new descriptive material and often a predefined sequence of presentation.

The DCPC is designed to support development of digital collections. The digital collection is not necessarily intended to completely replicate the archival collection -- the most serious researchers will always need to work with the original materials. Rather, the digital collection is intended to provide online access to meaningful subsets of the archival collection, to promote awareness of the full archival collection, and, especially important, to bring archival materials to the attention of searchers who may not have known to search there.

The characteristics of a good digital collection may be somewhat different than the characteristics of a good archival collection. When selecting collections for digital conversion, libraries should consider such factors as:

- Does the collection contain visual materials (photographs, graphics) or audio/video materials that would make good use of online display or performance capabilities?
- If the collection is largely text, would the content support and benefit from key word searching, i.e. by OCR conversion or transcription?
- Has the collection been processed sufficiently to include basic descriptive information about each item, or can such information easily be derived from the items themselves?
- Does the library have the right to digitize and disseminate the materials in the collection? If the copyright is held by other parties, can permission be obtained in a timely manner?
- Is the collection of a manageable size or can it be divided into meaningful subcollections for online use?
- Does the collection complement other WRLC digital collections and expand the scope of a particular topic?

The project should be planned to make a meaningful collection or subcollection available to users in the shortest time possible. Plans for large collections may involve multiple phases for multiple subprojects, or later phases for providing additional descriptive metadata. Often this will require focusing first on certain materials within the original collection and/or certain kinds of readily-available descriptive metadata, with a plan to

add items or metadata on an ongoing basis. DCPC staff will work closely with the owning library to define the project or subprojects.

### **Project planning**

During the course of each digitizing project, a number of project planning documents will be developed, including:

- Project agreement
- Rights and permissions statement
- Material characteristics and scanning specifications
- Metadata scheme and template

The Project Agreement includes information about:

- Owning library (including contact person)
- General description of the physical archival collection
- Description of the digital collection(s) (formats and quantities, relationship to the physical collection)
- Delivery and retrieval plans for the original materials
- Special handling requirements if any
- Conversion schedule

WRLC requires the owning library to affirm that it has the rights to digitize the material in the collection. In cases where works to be digitized are covered by copyright, the owning library is responsible for obtaining permission to digitize and disseminate the works from the copyright holder. A copy of the permissions document must be provided to WRLC along with the materials.

The Colorado Digitization Project has developed an excellent summary of the “Legal Issues to Consider When Digitizing Collections” located on their website at [http://www.cdpheritage.org/resource/legal/legal\\_issues.html](http://www.cdpheritage.org/resource/legal/legal_issues.html)

### **Metadata specifications**

Each WRLC digital collection contains four categories of data:

- Collection record: A bibliographic record in the shared WRLC libraries catalog, describing the archival collection and linking to the digital collection (MARC format)

- Collection header: A home page for the digital collection providing a basic description and search options, with a durable URL which can be referenced from other websites (HTML format)
- Item metadata: Descriptive information about each item and administrative and technical data describing the specifications for the digital item (Dublin Core format)
- Finding aid (optional, if available): Descriptive text about the collection and its provenance, with detailed information about the contents of the collection (Encoded Archival Description (EAD) format)

#### Descriptive metadata template

DCPC staff will work with the library staff to define a metadata template which identifies the descriptive metadata to be entered for each item in the digital collection. DCPC staff or library staff (depending on the project) will create and enter data for which there is readily-available source information, i.e. derived from the original piece in hand, from an existing index or finding aid, or already in machine-readable form. Library staff can enter additional metadata later if they choose.

For each project or subproject, DCPC staff and library staff will identify:

- Any metadata that may already be available in machine-readable form. DCPC staff will work with WRLC's IT staff to develop specifications and programs for loading that data.
- Common or standard subject headings to be assigned to all items in that collection (usually two or three headings). This is to facilitate combined searching with other collections or with the libraries' online catalog.
- Other possible categories for other individual headings (e.g. personal names of photo subjects), the source for that data, and any controlled vocabulary that should be used. Usually only two or three additional major headings will be created for any individual item.
- Other descriptive data that might be available or possible, such as a descriptive note or a transcription of manuscript. When feasible, DCPC will create transcriptions for very brief manuscript items such as letters. The text of existing machine readable transcriptions will be indexed for keyword searching where appropriate.

#### Administrative metadata

DCPC staff also create administrative metadata for each item (digital object), including:

- Date scanned
- Scanning location (WRLC)
- Scan device
- Resolution

- Bit depth
- Compression
- Dimensions
- Display file name
- Master file name

### **File naming convention**

WRLC uses a standard file directory and file name convention for all DCPC collections. Digital object files are stored on the WRLC server under a URL consisting of:

- <http://www.aladin.wrlc.org/dbdata/>
- /Collection name/
- /display/ or /master/
- An individual item identifier corresponding to the item's original Box# Folder# Item# and Page # (if any). The actual source location would be used if applicable, or an identifier would be constructed according to this structure.

For example, a display image for a single photo from Catholic University's Powderly collection would be identified as:

<http://www.aladin.wrlc.org/dbdata/powderly/display/b1-a05-001.jpg>

Each digital object may be accessed from any other Web source by using that URL to identify the object.

### **Scanning specifications**

WRLC has adopted the following standards for the 3 types of images created for each item:

- Master image  
Format: TIFF  
Compression: None  
Resolution: 300 dpi  
Bit-depth: 24-bit color  
Dimensions: Actual
- Display image  
Format: JPEG  
Compression: Medium  
Dimensions: Manuscripts: 800 pixels wide for both landscape and portrait.  
Graphic materials and typescripts: 600-800 wide (landscape),  
300-400 wide (portrait).  
Clippings: vary with the original size.
- Thumbnail image  
Format: JPEG  
Compression: Medium  
Dimensions: 150 pixels wide

Some materials require a different approach. Any variations from the standard WRLC specifications will be documented as part of the project plan.

### **Finding aids**

In addition to developing and maintaining digital collections, WRLC maintains a searchable collection of finding aids in Encoded Archival Description (EAD) format. Finding aids may be converted and loaded into the WRLC digital library either with or without a corresponding digital collection.

If an individual finding aid is available for a given digital collection, DCPC staff will work with the library to determine how best to convert the finding aid. DCPC is not staffed to perform extensive or complicated EAD tagging, but may occasionally convert a simple finding aid as part of developing the digital collection. WRLC is working on a feature to allow library staff to enter and tag finding aid data.

If a database of finding aids is available in machine-readable form, WRLC staff will work with library staff to develop an automated conversion of the existing finding aids.

### **Summary**

DCPC staff are available to help member libraries plan, design, and create digital collections. For more information, contact the Manager, Digital Collections Production Center at [dcpc@wrlc.org](mailto:dcpc@wrlc.org).