Building Standardized Digital Collections:
ResCarta Tools, a Demo

John Sarnowski and Samuel Kessel

The ResCarta Foundation, Inc.
`info@ResCarta.org`
`http://www.rescarta.org`

Abstract. ResCarta Tools are a suite of open source software applications which can assist in the creation of standardized digital objects. ResCarta Tools are open and modular in their design. Modules for creating digital objects store the metadata in Library of Congress METS/ MODS/MIX XML formats. Collection and indexing modules create LUCENE indexes for high speed fielded and full text retrieval of objects. The tools have been used to create digital collections from a variety of analog and digital sources. Collections can be hosted on the web using Apache TOMCAT and the ResCarta WEB application, which provides inline metadata using COINS. Integrating the use of DLESE OAI is done using the Collection Manager METS XML data. The tools have been used by small public libraries to host a dozen pamphlets and aerospace manufactures to host tens of thousands of documents and millions of pages.

Key words: Search, open source, METS, MODS, MIX,.

1   Introduction

In its funding request for the Making of America 2, the member libraries stated, "The emerging national digital library has the potential to elevate resource sharing to a new level, as it will be possible for users anywhere to find and use entire books, journal articles, and primary source materials directly over the Internet."

They also warned, "However, this potential will be realized only if the library community agrees to new practices and standards that will allow digital library materials to be easily located and used. Without community standards, each library will store its electronic content in a proprietary format in proprietary computer systems.

They went on to further state "To create a national digital library, it will be necessary to define: a) community standards for the creation and use of digital library materials and; b) a national software architecture that allows digital materials to be shared easily over the network."

The ResCarta Foundation, a non-profit organization, was founded to encourage the development of a single set of open community standards and open source implementations of those standards.

Our goal is to create, through collaboration, digital content production standards and open source applications that allow users to access disparate digital collections in a simple, user friendly process, leading to interoperability.

2   Conclusion

ResCarta tools are professional grade, open source software applications that produce imagery which passes best practices testing, provides the end user with simple forms to create archive quality metadata in accepted Library of Congress XML formats, creates collection level metadata and provides web applications for the end user to host or have hosted their full resolution archive without the need for html coding. The collections produced can be validated from time to time with a checksum validation tool. Hosted collections using the ResCarta Web application can provide inline metadata for consumption by other open source tools like Zotero and collection level XML files to produce extract files for supporting open source OAI/PMH delivery of metadata.

Existing collections were made from city directories[1], newspaper archives[2], microfilmed journals[3], and born digital electronic documents[4].

3   The Demonstration

Starting with a scanner and a digital camera, we will produce raster images in TIFF and JPEG formats. The ResCarta Metadata Creation tool will be used to add metadata to the images. The ResCarta Data Conversion tool will convert these images along with their respective metadata elements into digital objects with unique identifiers, embedded metadata in MODS/MIX. Existing PDF multipage documents (Normal, Image and Text and Image Only) will also be converted to digital objects. The ResCarta Textural Metadata editor will be used to verify the full text data from the PDF sources. Then the ResCarta Collection Manager will be used to gather these objects and other preexisting digital objects into a collection and generate a METS XML file. An APACHE TOMCAT web server will be installed, the objects will be indexed and a fully functional website will be produced from the collection. An OAI/PMH service will be setup for distribution of the collections metadata. The collection will have each items checksum verified with the ResCarta Checksum Verification tool. Attendees will be given handouts with technical details and a disk with the software tools and sample data. The handouts will allow them to follow along, as the raster images become information on the web in less than twenty minutes. Questions will be taken from the audience.

---

[1] `http://rescarta.lapl.org:8080/ResCarta-Web/jsp/RcWebBrowse.jsp`

[2] `http://murphylibrary.uwlax.edu/Racquet/jsp/RcWebBrowse.jsp`

[3] `http://www.lib.pu.ru/dcol/jsp/RcWebBrowse.jsp`

[4] `http://www.ama-assn.org/ama/pub/about-ama/our-history/`
  `ama-historical-archives.shtml`