



Digital Preservation in State Government Best Practices Exchange 2006

<http://statelibrary.dcr.state.nc.us/digidocs/bestpractices/index.html>

Summary of the “Metadata” Sessions **Compiled by Kristin Martin, Digital Metadata Manager/Documents Cataloger**

During the exchange sessions on metadata, multiple projects and types of metadata were discussed. Speakers from the Library of Congress, North Carolina State University, and the states of Arizona, North Carolina, Texas, Utah, and Washington shared their experiences. Types of metadata discussed included: descriptive metadata for digital publications, archival materials, and geospatial data; preservation metadata; and structural metadata. Participants came from a variety of backgrounds and sophistication in regards to methods for metadata creation. Talk focused on the selection of metadata schema, standards for content, converting metadata from one format to another, and automated creation and control of metadata.

Three major themes emerged from these sessions. First, quality control of metadata is a major hurdle, especially with metadata created by such a wide variety of individuals with varying levels of training. Even among professional staff, the semantic meaning of metadata may change depending upon the user creating the metadata, and this problem has increased as institutions coming from different background, such as libraries, archives, and museums, have begun to collaborate with one other to create a single repository for digital information (Arizona called this the “semantic battle”). With the explosion of digital information, many institutions need to rely on existing metadata which may not have followed standards. States are combating this problem through educational efforts aimed at the metadata creators to “fight dirty metadata,” clean-up and normalization efforts after data has been collected, and the development of templates, best practices, and input guidelines. Second, the desire for rich high-quality metadata must be constantly weighed against trying to develop standards that are simple and inexpensive. Dublin Core is a preferred descriptive standard for state publications because of its universality and simplicity, but states recognize that adding input guidelines and requiring the use of controlled names and controlled vocabulary, while adding the complexity of creating the metadata, greatly increases the probability of the user being able to find the information. Requiring content creators to use a complicated standard may lead to less cooperation and less deposit of digital materials by content creators, but may in fact be necessary for certain types of information, such as geospatial data. Collecting data through automated methods may help gather more data but can add to problems of quality control. Third, the selection of any one standard to hold metadata is less important than the ability to transform the data from one standard to another by building crosswalks. Flexibility in accepting metadata about an object is necessary if collecting from a number of sources, what then is important to successfully move the metadata into a common standard without losing information, or the building of crosswalks to allow for successful searching across multiple metadata schema.

There were also themes in the metadata exchange sessions that were universal across the Best Practices Exchange. Participants stressed the need to get some sort of project underway, even if all of the tools are not optimal. As an example NC ECHO recommended institutions store preservation in a spreadsheet as a first step if they are lacking the capability to store data in XML. Participants identified the need to gain buy-in from the content creators and policy

makers to understand the importance of metadata in preserving and providing access to digital government information. Not only would this help in the immediate term by making the creation and collection of high-quality metadata easier, it could help lead to a creation of legal mandate for the preservation of digital state government assets. Finally, the importance of working collaboratively with other institutions to develop systems and standards was consistently described by all who spoke. Collaborative work has its headaches (consider the “semantic battle”) but it provides richer solutions than institutions could create on their own.