PBCore
Public Broadcasting
Metadata Dictionary Project

CPB Public Broadcasting Metadata Dictionary Project

Test Implementation Summary

August 2004

Prepared by:
Alan Baker
Minnesota Public Radio
abaker@mpr.org
(651)-290-1320

EXECUTIVE OVERVIEW

In January of 2002, the Corporation for Public Broadcasting called together a team of public radio and television producers and managers, archivists, and information scientists to develop a single, shared protocol for identifying and describing the rich media assets produced by public broadcasters and their content partners. This document summarizes the results from the test implementation process, which used the version of the Public Broadcasting Metadata Dictionary (PBCore) that was published in draft form in February, 2004.

As the test implementation phase was getting started, another project team was concluding a request-for-comment process in which they polled 47 participants from many areas within public broadcasting and related fields about the need for a shared public broadcasting metadata dictionary, and asked them to comment on the current draft version of the PBCore. The participants in this survey overwhelmingly expressed a belief that there was a real need for the PBCore, and that the proposed PBCore met this need. Results from that survey are available at:
http://www.utah.edu/cpbmetadata/resources/index.html#RFC_Summary. The results from the RFC and test implementations informed the changes made to finalize PBCore version 1.0.

The task for the test implementation team was to use the draft version of the PBCore in several real-world scenarios. We tested three scenarios with the following goals:

Test A:
>   To determine how effective the PBCore is when used as a data inter-operability/translation tool. (i.e., when data is extracted from a content producer's system, mapped to the PBCore, then imported into the national distribution system. This was done both for television and radio)

Test B:
>   To determine whether the PBCore can be used as the metadata model for a digitizing/archiving process, and in what form the PBCore would be most effective in facilitating a complete "markup" of rich media assets

Test C:
>   To map the PBCore to several other emerging standards in order to determine how difficult it is to map the metadata fields from an existing metadata dictionary to the PBCore data element set

TEST TOOL DEVELOPMENT

It was necessary to create a set of tools for test sites to use. Two development projects were undertaken by WGBH.

The first project was to create a DTD that defined how to represent the PBCore metadata in XML. All four inter-operability/translation test sites used this DTD to export and import metadata records.

WGBH also created a special cataloging tool that would allow the Test B site to catalog a collection of video assets using a FileMaker Pro database built around the PBCore metadata set and recommended controlled vocabularies.

Both of these tools were created specifically for these tests and were not intended for use among general public broadcast stations or their partners beyond the test phase of this project.

PARTICIPATING TEST SITES

Given the need to conduct the tests in a limited time frame, selected test sites already had familiarity with the PBCore and/or was able to appoint selected staff with expertise to enable them to quickly carry out the tests with minimal training and support.

Test A: Interoperability/exchange between national production and distribution partners
- Television production partner, WGBH
- Television distribution partner, PBS
- Radio production partner, Minnesota Public Radio
- Radio distribution partner, National Public Radio

Test B: Content description for archives or digitization/preservation projects
- Kentucky Educational Television and the University of Kentucky

Test C: Mapping the PBCore to other established or emerging metadata schemes
- Grace Agnew

TEST METHODOLOGY AND REPORTING MECHANISM SUMMARY

Before the testing launch date, each test site was given a launch packet, delivered electronically, which contained all the necessary documentation and tools required to conduct each test. The launch packet contained the following:
- An overview of the project, and the objectives for the specific testing that site was to carry out
- A document that detailed the methods to be employed

- A spreadsheet to track labor and other resources used
- A questionnaire to be filled out after the completion of testing
- A list of links to online PBCore resources to help the test team understand how to properly apply the PBCore
- A simplified Excel spreadsheet overview of the PBCore elements and modifiers
- Sample PBCore XML documents and a draft PBCore DTD
- For the Test B site, which required a mark-up or cataloging tool, we provided a FileMaker Pro database for use in describing their materials.

Testing began at each test site with a kickoff conference call between test partners facilitated by the testing subcontractor. There were also regular conference calls throughout the run of each test that allowed the participants to discuss the progress of the testing with each other and the test subcontractor.

Upon completion of testing, the test subcontractor held a final conference call to gather final impressions from each test team. This call allowed the test subcontractor to answer questions that came up during testing, and provided a way to discuss issues that were brought up on test site questionnaires. Reponses to the test site questionnaires and e-mail messages exchanged during testing were all drawn from to create the summarized Lessons Learned section below.

LESSONS LEARNED

There was a sincere belief by all participants in this test phase that their organizations, and public broadcasting at large, need something like the PBCore to be widely available (and in use by a majority of public broadcasters) to enable them to capitalize on opportunities in the future. These range from simple labor cost savings to new business opportunities.

While technology for production and broadcasting has changed drastically over the past several years, broadcasters have not had affordable digital asset management solutions available to them. Many organizations have delayed acting to solve pressing problems related to the storage and handling of their content, especially material that is born digital and resides on production and broadcast servers. This delay has created a situation in which many broadcasters have a critical need for solutions that will enable them to share content with their production and distribution partners, as well as help ease the burden on their current digital production, broadcast, and archive systems.

Educating the industry is critical
Perhaps the No. 1 lesson learned in this test phase was how important educating potential PBCore users will be in achieving an adoption rate that will help sustain the PBCore. Test participants said that the cultural change necessary for

organizations to adopt the PBCore is likely an even greater challenge than overcoming the technical learning curve. Institutional support for the PBCore is needed *throughout* organizations attempting to benefit from this development, so that people charged with implementing the PBCore have the resources necessary for successful implementation.

Relevance to daily work
It was clear during the testing that an organization's desire and ability to use the PBCore metadata standard is closely tied to the benefits they believe will accrue to them as a result. PBCore implementation efforts should help stations see real benefits in day-to-day labor savings and the ability to pursue new opportunities.

Understanding project scope
Knowledgeable professionals are needed to consistently apply the metadata fields in the PBCore to ensure the usability of the data they gather. While every test site had plenty of technical capacity for manipulating the data as XML, the procedure of mapping the metadata to existing or future internal system databases called for skills that were not readily available in all cases. Both of the national distribution partners involved in the testing are developing next-generation systems for the distribution of content to their member stations. Many test participants expressed the strong desire for these systems to have built-in support for materials defined using the PBCore.

Clarify the PBCore's intended use
Some test participants and others who are interested in the PBCore expressed the need to better understand PBCore, and where it can best be put to use. Implementation efforts should include a complete vision statement for what the PBCore is, its intended uses, and how organizations can implement it.

Simplify the language
Nearly all the test participants felt that simplifying and explaining the language within the PBCore is a must. Some test participants requested simplification of the language used to label elements and modifiers; there was also a unanimously expressed desire for the PBCore documentation and examples to use terminology that is more production personnel friendly. This information informed the development of PBCore version 1.0.

Support minimal implementation
The PBCore model needs to contain a built-in way to support organizations that want to be compatible but can only provide minimal information about their content. Test participants believed that any organization should be able to start wherever their experience level and internal systems currently allow, with the option to grow into a more comprehensive use of the PBCore over time. Organizations will be able to implement more of the PBCore as the standard evolves, as their internal experience and expertise grow, and as their internal systems gain the ability to support the PBCore's structured metadata schema.

Develop solid technical tools and documentation

All of the sites involved in the exchange of data believe that there must be a robust Document Type Definition (DTD) that dictates and explains the proper way to represent a PBCore record in XML. Detailed comments within the DTD will help technical implementers be consistent in how they import and export metadata. Grace Agnew suggested that we create an XML schema to represent the proper structure of PBCore XML. An XML schema would be superior to a DTD in representing relationships within a PBCore metadata record.

A majority of test sites also requested industry-specific examples of full metadata records and their expression in XML, since they felt that the current examples were too generic to help all audiences. These new examples should be added to the PBCore Web site, and will be a great help to those trying to implement the PBCore in various locations. Several test teams suggested that television and radio should have examples created specifically for them; ideally, there would be sites devoted specifically for radio and television to make navigating the examples and application guidelines easier.

The PBCore mapped successfully and fairly easily to several other existing or emerging metadata standards. There are, however, competing needs for those wishing to use the PBCore. We need to support a high level of technical sophistication in the schema design, but also need to present it in a way that general users can understand. This increases the need for support and training for technical implementers, and also increases the need for some basic cataloging or mark-up tool to be available to organizations with minimal technical know-how on staff.

Several test sites also requested two tools:
- A simple tool that would help organizations map existing data fields to PBCore metadata fields.
- A cataloging utility to create PBCore metadata records and export those records in PBCore-compliant XML.

Support of national organizations is critical

Both national distribution partners involved in the testing phase are in the process of designing next-generation distribution platforms for their members. Enthusiastic support from these organizations and other leaders in the industry is required to ensure that producers realize benefits of adoption and labor savings in using data delivery channels that ensure PBCore compatibility.

TEST SUMMARIES

Test A: Television

Test Focus:

Interoperability/exchange between national production and distribution partners

Test Partners:

WGBH and PBS

Testing Methodology
1. Evaluate PBCore XML representation, its associated DTD, and map metadata fields from existing repository to PBCore
2. Create PBCore XML records for a collection of multimedia content and document the process
3. Publish those records in PBCore DTD compliant XML
4. Distribute records to distributor test partner
5. Ingest records, manually or via an automated process, to the distribution system's metadata management tool
6. Test searching across these records, to determine ease of access and benefits or drawbacks to using the PBCore as a system to enhance discovery of materials
7. Summarize the work done and lessons learned in test results questionnaire provided by test subcontractor
8. Participate in post-test follow-up phone interview with subcontractor

Test A: Radio

Test Focus:

Interoperability/exchange between national production and distribution partners

Test Partners:

Minnesota Public Radio and National Public Radio

Testing Methodology
1. Evaluate PBCore XML representation, its associated DTD, and map metadata fields from existing repository to PBCore
2. Create PBCore XML records for a collection of multimedia content and document the process
3. Publish those records in PBCore DTD compliant XML
4. Distribute records to distributor test partner

5. Ingest records, manually or via an automated process, to the distribution system's metadata management tool
6. Test, searching across these records, to determine ease of access and benefits or drawbacks to using PBCore as a system to enhance discovery of materials
7. Summarize the work done and lessons learned in test results questionnaire provided by test subcontractor
8. Participate in post-test follow-up phone interview with subcontractor

Test B:

Test Focus:

Content description for archives or digitization/preservation projects

Test Partners:

Kentucky Educational Television (KET) and The University of Kentucky

Testing Methodology

1. Select collection to be described using the PBCore
2. Set up tool (provided by CPB team) to create PBCore compliant records for these assets
3. Work with testing subcontractor to describe selected collection as accurately and completely as possible
4. Deliver collection of records for evaluation
5. Investigate how easy it is to find assets or information contained in records that have been created using the PBCore
6. Summarize the work done and lessons learned in test results questionnaire provided by test subcontractor
7. Participate in follow-up interview with test subcontractor

Test C:

Test Focus:

Mapping the PBCore to other established or emerging metadata schemes

Test Site Participant:

Grace Agnew, chosen for this test because of her extensive knowledge of the latest developments on the metadata standards that are specifically designed to be used with rich media assets

Testing Methodology

1. Evaluate PBCore elements and modifiers for equivalencies in other metadata standards
2. Evaluate PBCore presentation and XML formatting for applicability to assets in participant's field

3. Create crosswalk mapping of elements and modifiers from the PBCore to other standards
4. Summarize process and challenges for future implementations of similar crosswalks
5. Summarize the work done and lessons learned

Test Summary

This test was fairly simple in design, but required a working knowledge of the newest metadata schemes in order to create the crosswalks between the many metadata standards.

Mappings to PBCore were created for the following eight metadata standards:
- Moving Image Collections (MIC)
- Dublin Core (DC)
- Qualified Dublin Core (now called DC Terms)
- SCORM
- MPEG-7
- SMPTE
- MODS
- METS

The mapping was done in an Excel spreadsheet, which listed the PBCore-equivalent data elements for each data element in these metadata schemes. Also noted were some basic guidelines for application, as well as any mapping considerations that came up as the mapping was being done.

To guarantee that the mapping process was complete, and that all application considerations were fully considered, Grace created a PBCore metadata record for a video asset. This record was then expressed in XML, conforming to each of the standards that was mapped.

Some of the key findings from this process:
- The PBCore maps well to most existing or emerging standards
- Several schemes do a better job then PBCore of presenting the relationships of data elements within a metadata record, yet the process of creating records using those standards and manipulating the XML is quite complicated
- The SMPTE metadata dictionary, which has been suggested as a dictionary that could be used instead of the PBCore, lacks some key descriptive elements that are critical to use within public broadcasting
- There are good models for extending the PBCore in the future to include administrative metadata (meta metadata) and structure maps