

# **Metadata Dictionary for Public Broadcasting Phase 3 Test Implementation Plan & Timeline from Task Team E**

## **Team E Members**

Nancy Baldacci/American Public Television, David Felland/Milwaukee Public Television, Steven Heard/Public Interactive, Bea Morse/PBS, Steven Vedro/Consultant, Wisconsin Public Television. Additional participants: Marcia Brooks (Project Director)/WGBH, and Alison White (Project Officer)/Corporation for Public Broadcasting.

## **Plan Overview**

This plan defines the learning objectives, processes, timeline and budget required to conduct Test Implementations of the PB Core (formerly referred to as the PBMD), as known in Phase 3 when this plan was authored.

Note that the Test Implementation rounds to be conducted in Phase 4 will be informed by the recommendations in this plan; however, the final execution will be proposed and led by a sub-contractor, working in close collaboration with Marcia Brooks/WGBH and Alison White/CPB.

## **Statement of Purpose**

The Test Implementation process is meant to prove that PB Core data can be inputted into a separate system, and then usefully searched upon and extracted from the system. This is to be accomplished by either beginning the digitizing of an un-archived collection, or to extract, translate and import program content data between software systems commonly used in public broadcasting. The end result of test implementation would be an application profile for a particular use – i.e., to extract data from Orion and map to the PB Core.

The RFC (request for comments) process will inform the test implementation process, with participants in the test implementation process formed from a sub-set of the RFC respondents.

## **Problem Statement/Learning Objectives**

Test sites will be representative of two varied user profiles: an organization just beginning to catalogue its assets, and a specific/limited scope, i.e. determining what it would take to map an existing system with the PB Core. The end result will be an application profile for particular use, i.e. to extract data from PBS' Orion, and map it to the PB Core. The most compelling test will be to prove that existing data can be extracted and translated. While learning objectives are listed below, the test implementation will be organized by existing project use case scenarios, in order to best test different aspects of different data in varied contexts.

- Is the PB Core registry and User's Guide effective?
- Does test or actual implementation of the PB Core require the creation of a new or temporary data repository within an organization?
  
- How difficult is it to map the records from an existing database to the PB Core? What software or other tools, efforts are necessary? What is the cost of creating and using these tools or conducting this process? Is coding the data elements in HTML and XML necessary?

- How effective is the PB Core when it's used as a data interoperability/translation tool (i.e. when data is extracted from one software system, mapped to the PB Core, then exported into another software system by being mapped back to those data fields)?
- Can the PB Core be used to initiate a digitizing/archiving process, or would it be likely to require a more intuitive user interface? Is the PB Core effective in providing a complete markup of rich media essence, or are more fields required?
- How effective is the PB Core in directly assisting asset or record discovery?
- What might need to be tested in a subsequent phase?
- When fields don't match up in test implementation, determine if/to what extent participants would be expected to revise their own systems to add fields.

## **User Profiles**

Interviews were conducted with key constituents to help identify learning objectives for the test process. Each interview exists as a separate report. A test participants matrix was developed, to map test's learning objectives to participants. See Participants List below; what do the projects do, and why are they appropriate to test implementation of the PB Core.)

## **Test Methodology**

The test implementations will be organized by use case scenarios, conducted with key constituents, in order to test different type and amounts of data as follows:

- Existing collection and data extraction/translation system
- Start-up project to test "does this work" for someone just starting out. Criteria: some documentation of the collection is required, even if on index cards.

It is presumed that separate instruction paths and questions will be required for each.

Other possible approaches to organize the test implementation process:

- By "program-level"
- By "clip" level

This a test to extract participant organizations' data in the form of the PB Core Participants take the PB Core, attempt to apply it to their own existing system, or input into a new system. Then, conduct searches against it to see if it crashes. Experts would be required to determine the cause of the crash. Participants need to be willing to compile and track problems encountered. It is anticipated that a hosted central repository may be required (and to possibly serve as a data validation center), as well as a central metadata expert to serve as a resource for test participants as they encounter questions or issues. All test participants will be expected to complete Step 1, with the assumption that the process will winnow further participants in Step 2 and beyond.

Step 1 (Inherent in Survey process)

- Conduct intellectual exercise to see how PB Core and participants' system(s) map together, by attempting to map portion of existing assets metadata to PB Core. Based on tester's needs: <Estimated duration: 4 hours>
  - Identify "easily matching" elements
  - Identify "useless" elements

- Identify "required additional" elements
- Process should reveal a decision point to determine:
  - If the process has failed and can't be advanced further, or,
  - That there's enough "cross-mapping" to proceed to step 2.

Step 2 – A test to define how much existing data maps to the PB Core; to determine who needs to map to which element. <Estimated duration for this step: 1-2 days>

- Use existing software extraction and translation tool to export out of existing database
- Map PB Core across existing metadata. Run duplicate databases (what tool does that require?)
- Dependent on how much structured data participant organizations have, each organization's expert needs to extract their data.
- Process should reveal a decision point to determine:
  - If the process has failed and can't be advanced further, or,
  - That there's enough data to reveal further findings and proceed to step 3.

Step 3

- Create a new temporary database (a "mock PB Core") to separate production system from development system. Whatever tool participants use, create a new set of fields.
- Use existing software extraction and translation tool to export out of existing database.
- Put the two databases side-by-side, input new fields/data to both to see what works in each system.
- Identify how much of participants' existing data maps.
  - Identify "easily matching" elements
  - Identify "useless" elements
  - Identify "required additional" elements
  - What is missing? How long does it take?
- Process should reveal a decision point to determine:
  - If the bulk is map-able, within a subset of participants' work
  - If there aren't that many fields to fill out, assess if it's worth going forward.
  - How much of the PB Core needs to be used to share the PB Core (RFC process will show that different communities use different fields.)

Step 4 – Determine if the PB Core is sufficient for full mark-up.

There may be an additional Step 4, to conduct a cost benefit analysis to weigh the time involved vs. the amount of time/money saved.

## Task List

- Develop Users Guide
- Develop invitation letter to participants
- Determine database requirements (engine=Sequel?)
- Identify "help desk" participants

## Test Environment/Equipment Needed

Further work is needed to define realistic workflow and cost, to be done in consultation with Test Implementation sub-contractor in Phase 4.

## **Test Monitoring**

### **Evaluation Measures**

Further work is needed to define the data to be collected, analysis and monitoring processes, and testing criteria, to be done in consultation with Test Implementation sub-contractor in Phase 4.

### **Test Implementation Participants**

Site selection criteria objectives are to invite projects/participants that are as limited in scope as possible, while being as illustrative as possible.

- NPR Content Depot or NPR Archives
- WGBH DAM Project
- PBS Orion Database (MP recommends mapping Moyer's NOW to PODS and NPR databases)
- AMIA Moving Image Collections (MIC) Project
- KUED "Utah Collections" Multimedia Encyclopedia
- Kentucky Network Digitization Project
- Annenberg Digitization Project
- Minnesota Public Radio Archive Project
- CPB-funded independent production TBD
- Milwaukee Public Television
- Research Channel
- United Streaming

### **Test Budget**

Items for consideration:

- Fee paid to test site(s)
- Repository development and hosting
- Software
- Web support/conferencing
- Equipment/Infrastructure
- Software and data extraction/translation experts as part of the project, and possibly a metadata expert
- A help desk – technical support and assistance from people who run databases

### **Test Timeline**

Duration is to be no longer than three months. Initial test implementation is slated to occur from mid-October through December 24, 2003, with three small groups using a vetted PB Core.