Metadata Dictionary for Public Broadcasting
Phase 3
Compliance-System Integration Plan from Task Team F

Introduction

The activities of the Integration and Compliance Team (Task Team F) of the Public Broadcasting Metadata Initiative Working Group are summarized in the Phase 3 Interim and Final Project Narratives, submitted under separate cover.

This plan describes how the work of the Public Broadcasting Metadata Initiative, and its resulting PB Core, are related to public broadcasting's broad and long-term data exchange, storage and integrity needs. Institutions, software systems, vendors and content distribution services that would be affected by the requirement to be compliant with the PB Core are listed.

Included are key insights gained from interviews that task team members conducted represent the perspective of:

- National distribution initiatives
    - TV program distribution: PBS Next Generation Interconnection System (NGIS) and Enhanced Interconnection Optimization Project (EIOP)
    - National Radio program distribution: NPR PRSS Content Depot
- Independent producer (with previous museum experience)
- Station: large, small, EIOP participant
- Software vendor

It is anticipated that the development of the PB Core in a usable form (such as making the data elements available in XML DTDs, or Extensible Markup Language Document Type Definitions, in a permanent online "registry") is necessary for both test implementations and "compliance".

Integration

It is anticipated that the results of the impending test implementations in Phase 4 will greatly inform how the PB Core will be integrated into key local operations (e.g. BO&E, Development, Web), as well as national initiatives (e.g. PBS NGIS, NPR Content Depot, PBS Orion, etc.) The anticipated software systems and technical infrastructures that are impacted by the integration of, and compliance with the PB Core are listed below. (Note: Reference "Entities/Software Systems/Technical Infrastructures Affected by PB Core Compliance Requirements" below.)

Existing Need for Metadata

- National TV Program Distribution - Metadata is vital to the success of PBS' NGIS and EIOP (Enhanced Interconnection Optimization Project). For NGIS and EIOP, metadata will be used for internal and distribution purposes, and therefore, a small subset of the PB Core will be used to identify most assets at both PBS and member stations (with other fields being related to archival functions and will be of interest to producers and archivists). A metadata standard embraced by all within the PBS community is necessary to assure timely and economical transfer of program materials, with the goal to reduce duplicative ingesting of program content by moving it as far upstream as possible (e.g. to the program producer or post-production facility). *All parameters of a program's content must be fully understood when the program arrives at PBS, and at the member station.*
- National Radio Program Distribution – NPR is working on four major projects that each have their own metadata needs (NPR library, production management, asset management for NPR Online, and a smaller version of asset management for Distribution). The needs of each project are all different, and metadata for each will certainly exceed the elements covered within the Public Broadcasting Metadata Initiative.
- Independent Producer – Metadata is absolutely critical. Conversations with other producers indicate a great interest in metadata, and an understanding of its tremendous potential.
- Large Station – Having a better structured and more complete metadata for finished shows would be helpful.

Challenges

- National TV Program Distribution - The big challenge is to find ways to move data between PBS and stations, and between systems at each station.
- National Radio Program Distribution –
    - The real challenge is coming up with a metadata model that is clear and understood by everyone once it's put into use. Concern is that standards appear simple until they are actually implemented. For example, vocabulary -- descriptors mean different things to different people, especially when there are a myriad of disciplines involved. Descriptors may seem clear until they need to be understood within a specific context.
    - An additional challenge is to identify as many elements as possible that are common to all projects so that data flow can be as simple and smooth as possible.
- Independent Producer – Minimize the time and effort to input data; the process needs to be automated as much as possible. The freelance producer quickly becomes disassociated with his/her production, since he/she doesn't own the project. Therefore, presently, many notes and other data taken in the field do not follow the assets once they leave the producer's possession. As a result, without metadata directly attached to the asset, considerable information concerning each production is often lost.

- Large Station –
    - Would need to see benefits described to be convinced they exist.
    - It's a challenge to get people to use any structure consistently. Every project does its own thing.
    - There's too much re-keying between systems, but the main problem is lack of regional network programming and the fact that (PBS's) Orion is not done.
    - Don't know how the PB Core can address the station's needs, since the main problem seems to be lack of implementation of long established concepts, not lack of concepts.
- Software Vendor –
    - There will be outlying factors from the fields that have been identified once testing and implementation begins. Expect an abundance of initial issues, then a tapering off of questions until a policy, technology, or process change occurs.
    - Vendors will not only need to map to the PB Core, they will also need to map to each other (e.g. ProTrack/Harris). ProTrack and Harris will be faced with the challenge of mapping to a PB Core that is not a one-to-one mapping, so there's a need to ensure that information is not lost in the mapping process.
    - There is concern about global identifiers; PBS will be using UMIDS while APTV is potentially going in another direction. Need to understand where global identifiers get established.
    - What potentially needs to be tracked for DAM from a digital rights perspective would probably be different from what ProTrack would need to track; ProTrack would be interested in the rights associated with broadcast, while a DAM system would be interested in all aspects of rights.

Cost Considerations

- National TV Program Distribution - It is anticipated that adherence to a metadata standard will not add further expense than those producers are already incurring for providing data now.

Potential Savings/Added Value

- National TV Program Distribution - Significant savings will be realized downstream by the elimination of ingest functions at PBS and each member station.
- National Radio Program Distribution – Assets will gain more value the further they move down the production or distribution chain.
- Independent Producer - Metadata, such as GPS data acquired by the camera, may ultimately have value many years later when that particular scene is viewed in a final production.
- Small Station –
    - Metadata accuracy is imperative to ease workload.
    - Centralization, if planned properly, is almost always beneficial.
    - Various existing systems (e.g. programming, logs, tape library, membership, underwriting, pledge scheduling, Web, etc.) can be

mapped to the PB Core if structured correctly, allowing data to be moved more easily between systems, saving manual data translation time.

Suggested Approaches

- National TV Program Distribution – While in general metadata should always move along with the asset, some applications would require them to be separate. For example, EIOP will provide control and monitoring services to member stations; assets will reside at the local station, while metadata needed for control and monitoring would need to reside at the control point (PBS).
- National Radio Program Distribution – There are valid reasons to send metadata separately from the assets, as well as to have it "locked" to the asset. In fact, at NPR, metadata needs to be decoupled in terms of both geography and time. Often, metadata is created well before the essence is is acquired or created; disconnecting metadata from the essence is therefore critical to NPR's operations.
- Independent Producer –
    - Metadata needs to be firmly attached to the asset. A method needs to be developed that ensures the metadata follows the assets as they go through the shoot/edit/dub/distribution process. Metadata should follow the asset and its components at all times and should never be thrown away. Without metadata directly attached to the asset, considerable information about each production is often lost.
    - Methods to easily read the metadata at any stage need to be devised.
- Large Station – Have PBS fix NOLA and get regional program distributors to use Orion.
- Software Vendor –
    - It's important to detail why a field in the PB Core was dropped and where the dropped fields of information might be expressed or included in the existing PB Core; fields that are open to interpretation definitely need to be clearly defined. As testing and implementation begins, distinct reasons on why fields weren't included will need to be provided, and whether those fields need to be added to the PB Core will need to be determined.
    - There should be an ongoing "referee team" to answer questions that aren't accounted for in the existing PB Core.
    - There should be strict processes in place for documenting and identify needed actions to be taken.
    - The implementation team should not consist completely of people from academia, but rather, those familiar with testing and implementation, to ensure there is a clear understanding of what works and what doesn't.
    - Vary use cases.
    - Test an actual production (e.g. Dragontails) to face the challenge of how episodes are labeled. Dragontails is a good example of how episodes are labeled and then segments from the show are repackaged and sent back out with a different episode number, which provides no relation to the original content.

- Involve vendors in the test process to ensure they have captured how information needs to be encoded.
- As much information should be included in the fields, to allow for one export function and one import function for vendors.
- Whatever is presented to the end users needs to be as simplistic as possible.

Compliance

The constituents listed above provided the following insights as to how to assure compliance with the standard at the national and local level, including stations, national distributors, regional distributors, producers, etc. The Phase 4 marketing plan is anticipated to play a significant role in further disseminating information to key constituents about the use of, and adherence to, the PB Core.

- National TV Program Distribution –
  - Will ultimately be assured, as it is now, by monetary penalties. However, full migration to the use of metadata may be slowed somewhat, as many ingest functions may initially need to be maintained until PBS is assured that producers are fully compliant with standards.
  - PBS is trying to encourage software and equipment manufacturers to resolve many of the interface and translation issues that currently exist
  - For stations who choose to go with EIOP, If things become more centralized, stations will be able to rely on the central authority for compliance. They'll simply become users of a remote system.
- National Radio Program Distribution –
  - Because public broadcasters are thought to be more thorough than many of their commercial colleagues, it's thought that software vendors tend to underestimate the amount of metadata public broadcasters will need to capture, store or process. NPR recognizes that it, and similar organizations, will need to play an important part in influencing vendors. Discussions with vendors are complicated by different needs of different constituencies, e.g. radio and tv often have vastly different needs. Further, different disciplines within an organization (e.g. producers and engineers) have differing needs and interpretations.
  - It's not necessary to identify every element within a metadata standard and NPR does not always intend to do so; users should be allowed to use "subsets" without impairing the validity of any standard. It is not yet certain whether or how to identify fields that have been left blank intentionally or by accident.
  - Clear policy and guidelines should be established early and then adoption will most likely come quickly. Compliance/enforcement may be best accomplished by encouraging producers to comply with all standards and policies; convincing them that by doing so, assets and programs will become easier to find, use and distribute, and thereby become more valuable and sought after. Compliance with agreed-upon standards may never be achieved; there may always be some

doubt as to whether the metadata associated with an asset is complete and/or accurate. To some extent, audit or quality controls steps may always need to be interjected somewhere in the workflow.

- Creating metadata should be as transparent and as automatic as possible. Metadata should be added upstream as far as possible.

- Independent Producer –
  - Producers tend to put non-critical activities at the bottom of the list; unless the process of adding metadata is very simple, metadata will end up at the bottom of the list and never be fully implemented. Going back to add data later at the end of the day would be secondary to planning the next day's activities, and much information would simply be missed or forgotten. It is vital to produce metadata as early in the production process as possible. The process must be easy, non-intrusive, and intuitive. Any process that is distracting or time consuming will be avoided, and will result in failure.
  - A big step forward would be for camera manufacturers to enable input of metadata right at the camera as assets are being captured, (Note: it is believed that Sony has a camera that allows metadata to be added at time of shoot.) Examples of metadata added at time of shoot include: adding GPS information (latitude/longitude data, time, date, etc.); and other descriptive information – possibly by using verbal recordings at the camera.
  - Equipment or software should prompt the user by asking for specific information. Each discipline (e.g. producer, editor, distributor) will need to use only a portion of any metadata standard, and certain "subsets" should be recognized as valid contributions. Some method (possibly developed by equipment manufacturers/software vendors) should be used to recognize that some data fields have been intentionally left blank and not just ignored.
  - Large Station – It's hard to know what would compel various individuals and organizations to comply when I don't know what "it" is. (Note: this input has implications for the Phase 4 marketing plan.) If it's what's been seen passed out from PBS, or from WGBH's metadata project, it's difficult to imagine how even a large station could make practical use of it. The first thing needed would be practical and affordable tools that could comply with the structure, while interfacing with ours station's everyday tools.
  - Small Station – Don't know what would compel various individuals and organizations to comply when I don't have any prior knowledge of the Public Broadcasting Metadata Initiative.

Note: The PBS EIOP project is expected to dramatically affect participating stations' operations, and accordingly, their need to be compliant from a broadcast perspective. In EIOP's more centralized scenario, as users of a remote system, participating stations will be able to rely on the central authority (PBS) for compliance.

Entities/Software Systems/Technical Infrastructures Affected by PB Core Compliance Requirements

Software Systems – Technical Infrastructures

- Web content management software
    - Public Interactive
- Programming, traffic, development/underwriting software
    - Myers Information Systems/ProTrack
    - Target Software/Team Approach
    - Memsys
    - Trac Media
    - Allegiance Software (Traffic and corporate support system for radio/TV; Main Fundraising, Web membership)
    - Scout
    - Raiser's Edge
    - Triveni Digital (PSIP)
    - Harwood Consulting (Radio traffic, billing, accounting)
- Vendors from EIOP
    - Omneon – video servers
    - Microsoft – operating systems, databases, etc.
    - Miranda – operations and monitoring
    - Omnibus – automation and control
    - Accenture – analysis, design and project management
    - Broadview – traffic and scheduling software
    - Intel – microprocessors
    - SES Americom – satellite transport and monitoring

Other Institutions

Any system or company using a system that has data fields containing information about public broadcasting programming, at the full-length or clip program level:

- Production companies
- National distributors/funders
    - CPB Program funding
    - PBS
        - Orion program information
        - NGIS
        - EIOP
        - PBS.org CMS
    - Content distribution/subscription services
        - Research Channel - eLinks
        - Teachers Domain
        - Portal Wisconsin
        - Digital Well

- Related organizations with which public broadcasting interfaces/will share content
    - Libraries
        - Public
        - University
        - Government (Library of Congress)
    - Museums
    - Historical Societies
    - Archives
    - Educational institutions and organizations
        - AIT
        - Ames
        - United Streaming

Decision Points

Having included in this document a representative sample of the needs and challenges of various constituencies, the following list is offered to represent the points at which decisions about integrating and complying with the PB Core will likely be made.

To clarify, decision points happen over time in different parts of different organizations, and similarly, they will vary in size and importance. Regardless, they are likely to inform how and when resources can be prioritized, and they are tied to implications for maintenance and sustainability of the PB Core.

- **National Programming** – Adoption will be driven by PBS's requirements for program submission compliance (i.e. as is currently the case for PBS TOS <Technical Operating Specifications>).
- **NGIS/EIOP**
    - While local implementation of PB Core is critical, it can only be guided by an imposed national system. PBS's adoption of PB Core (and mapping EIOP metadata to it) will drive adoption.
    - When locally produced programs are broadcast along with nationally distributed content (e.g. from PBS), PB Core markup(s) are required for consistent metadata records that will match locally produced and nationally distributed programs.
    - Stations that subscribe to EIOP will be faced with fewer decisions points. Programming metadata is inevitable and inarguable.
- **Locally Produced/Acquired Programs** – Need to create a markup for existing internal systems, and standardize content labeling internally in a way that's consistent with workflow.
- **Local Integration of Internal Production Processes** – When production processes are integrated and the PB Core is not sufficient for them, metadata markups will still ensure PB Core compliance. IT, engineering and station management will need to know where workflow is impacted (e.g. in order to create interfaces for producers to enter production metadata at the initial point of logging). There's a need for data integration at every level of the organization, as well as every level of public broadcasting and beyond. The emerging mindset will need to build in the assumption that others will be

utilizing locally produced programming, and therefore, the groundwork must be laid when content is shot.

- **Content Consortia Participation/Content Partnership with Non-Broadcasters** – A good metadata protocol will enable every local station to search, locate and leverage a vast collection of public broadcasting content, which will enable mission-consistent public service partnerships and new business models. Content partners will assume public broadcasting's ability to do so.
- **Facility Changes/Software Purchases** - Adoption is driven by PBS's requirements for compliance (i.e. as is currently the case for PBS TOS <Technical Operating Specifications>). Vendors will need to implement PB Core or build the middleware map in order to retain PBS and station customers.

Conclusions

The main activities and desired outcomes of the upcoming final project phase (4) will be directly informed by the common themes that emerged from constituent interviews represented above:

- The Request for Comments rounds and Test Implementations will address:
  - The need for an easy-to-use standard, in order to ensure adoption
  - The known sub-sets of information required for typical discipline and constituent-specific content exchange scenarios
  - Whether or not to package metadata with its asset (or to do both)
- The Marketing Plan will address:
  - The specific need for increased awareness and understanding of the PB Core in the station community
  - What compliance means for target adopters
  - The benefits of adoption to all users of the PB Core

Efforts will be made to map criteria, decisions and measures for success, in order to enhance the delivery of services, and/or to identify and prioritize where cost savings can be realized (e.g., the potential establishment of service bureaus for metadata tagging of content targeted for multiple uses).