PB Core — the Public Broadcasting Metadata Initiative: Progress Report

Alison M. White
Corporation for Public Broadcasting, USA
awhite@cpb.org

Alan Baker
Minnesota Public Radio, USA
abaker@mpr.org

Marty Bloss
National Public Radio, USA
mbloss@npr.org

Paul E. Burrows
Media Solutions, OIT, University of Utah USA
pburrows@media.utah.edu

Efthimis N. Efthimiadis†
The Information School, University of Washington, USA
efthimis@u.washington.edu

Marcia Brooks, David MacCarn, Thom Shepard
WGBH Boston, USA
{marcia_brooks, dave_maccarn, thom_shepard}@wgbh.org

Cate Twohill
Public Broadcasting Service, USA
ctwohill@pbs.org

Abstract

PB Core is the result of the public broadcasting metadata initiative (PBMI). It is an effort of the public radio and television broadcasters to develop a schema for the description of their assets. PBMI is under the auspices of the Corporation for Public Broadcasting. The paper discusses the user-centered development of the schema, the elements of the PB Core, the application profile, and the feedback and evaluation process of the schema.
Keywords: *Public Broadcasting Metadata Initiative, Dublin Core, PB Core, Media Asset Description.*

1. The Need for Public Broadcasting Metadata

As public broadcasting endeavors to maintain our value and values in a dramatically altered media environment, we know we must do three things: develop and deliver content across multiple platforms, strengthen our editorial and service partnerships, and engage in more efficient methods of conducting our new and legacy activities.

The recent convergence of IT capabilities with those of radio and television broadcasting has caused us and our constituents to appreciate that our prized editorial output (video clips, audio interviews, transcripts, etc.) can be understood as a series of digital assets, that can be identified, exchanged and distributed using an advanced digital infrastructure. Our ability to network – to exchange rich media content – within and across our newsrooms, production suites, satellite and terrestrial distribution systems, etc., and even with our educational and community partners (schools, libraries, museums) has never been greater. We have been afforded a tremendous

† To whom correspondence should be addressed.

opportunity for cultural relevance and operational efficiency.

In a public broadcasting system made up of hundreds of independent licensees, the challenges of organizing universal processes for asset appraisal, digitization, rights clearance, preservation, etc. are myriad, perhaps overwhelming. We did understand, however, that the foundation of any future effort in this direction would be a single, shared protocol for identifying and describing our rich media assets.

The Public Broadcasting Metadata Initiative (PBMI) is a cross-organizational, multi-disciplined effort to establish a standard for all public broadcasting content (radio and television), in order that metadata might be more easily exchanged between colleagues, software systems, institutions, community partners, individual citizens, etc. The PBMI will be a "touchstone," a single, streamlined standard to which other database structures, including those of PBS, NPR, major producing stations, and other asset/content management systems will be "mapped." It can also be used as a guide for the onset of an archival or asset management process at an individual station or institution.

The project has been extant since January of 2002, and during its first two phases of CPB Future Fund support, a team of individuals representing public broadcasting's key institutions and endeavors, along with subject matter experts (see appendix for list of participants) has worked to:

- Develop consensus regarding project objectives and timeline;
- Recognize and codify the way our constituents use our content and content information. (Developed use cases based on interviews with producers, broadcast operation staff, educators, website creators, etc.);
- Examine relevant metadata standards in the media and library communities, to ascertain their applicability to our content and constituencies;
- Make information about the PBMI available via numerous conference presentations and a project website;
- Contribute and combine the substantial metadata work already performed at key institutions in public broadcasting (PBS, NPR, WGBH, KUED, MPR);
- Form a preliminary consensus regarding a single set of metadata protocols - the Public Broadcasting Core (PB Core) Metadata, Preliminary Version.1.

2. What Alternatives Were Available

The main goal of the PBMI is to create a schema that is easily understood, implemented and adopted by the Public Broadcasting community at large. PBMI embarked in a detailed review of existing metadata standards that are used for the description of rich media assets. These included standards that deal with the descriptive, administrative, and educational aspects of the assets. In general, while many of the metadata standards discussed below are in development, the Dublin Core Element Set has remained stable since its 1.1 revision in 1999 [1]. Additions and other changes to the Dublin Core model come in the form of recommendations and application profiles, but the basic core of 15 elements remain unchanged. So we have built our model upon the Dublin Core that provides a solid foundation that is extensible, scalable, and easy to understand.

The standards that were considered were OAIS, SMEF-DM, MARC, METS and MPEG-7, as well as the educational standards SCORM, LOM, IMS. These are briefly discussed below.

OAIS: Reference Model for an Open Archival Information System [2] is a framework and reference architecture for digital preservation.

SMEF-DM: Standard Media Exchange Framework - Data Model [3] is an end to end broadcast production model, workflow oriented. Our assets may involve domains or materials not exclusive or even related to broadcasting, such as CD-ROM, DVD, books. Metadata was determined to describe assets as objects or files. However, SMEF mandates a specific workflow with limited options. For example, assumptions are made on the order of activities. Our experience is that productions have many different workflows that must be accommodated.

MPEG-7: "Multimedia Content Description Interface" is a highly structured standard focusing on multimedia. Our model does not preclude a station adopting MPEG-7 because the PB Core is based on the Dublin Core model and will map to MPEG-7. On the other hand, MPEG-7 is narrowly focused on multimedia, not on the wide range of other media or materials that will be found in a producing station's repository. See e.g., Hunter [5,6], Agnew [7].

MARC: The MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form [8]. MARC requires a cataloging skill set that is not likely to be found in most public broadcasting stations. Our model insists on the integrity of each asset (version or format of the content). Dublin Core crosswalk maps to key fields in MARC http://www.loc.gov/marc/dccross.html.

METS: Metadata Encoding and Transmission Standard [9]. The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library.

SCORM: The Sharable Content Object Reference Model [10]. This is an application profile "to provide a comprehensive suite of e-learning capabilities that enable interoperability, accessibility and reusability of Web-based learning content."

IEEE LOM: IEEE 1484 Learning Objects Metadata. A Learning Object is defined as any entity, digital or non-digital, which can be used, re-used or referenced during technology-supported learning [11]. The mapping of LOM to Dublin Core is available at [12].

IMS Global Learning Consortium. IMS Meta-Data v1.2.2 [13] The IMS initiative originated in higher education but it now involves stakeholders in corporate and government training, K-12, and continuing education. The IMS learning consortium develops learning technology interoperability specifications. IMS initially set out to produce a unified specification covering metadata, content, administrative systems, and learner information. This proved to be too large a specification and IMS broke it up into component parts, with separate working groups developing each, and each being released separately.

SCORM, LOM, IMS, and DCMI education (DCEd) are examples of standards for creating, storing and serving educational metadata. The above schemas have many commonalities and there is an effort to increase interoperability among them. For example, SCORM uses LOM vocabulary. All schemas could be mapped to qualified Dublin Core elements. Extensions to our model as well as value lists (element types) allow for incorporating some of these needs.

3. Why develop PB Core?

Many parties have asked us why we did not adopt and adapt metadata schemas already in existence or in development. For several reasons, the existing standards were not appropriate to our needs. Basically, alternative schemas were either too cursory in their descriptive capabilities or far too ponderous.

An implementation project, such as the Public Broadcasting Metadata Initiative Project, generally finds that no one metadata standard completely meets its needs for descriptions of media essence. General standards, like Dublin Core, are often folded into domain- or sector-specific standards, such as MPEG-7 for multimedia and IEEE/LOM for educational resources. New elements may be devised which meet local needs not covered by any existing standards. The Public Broadcasting Core can be thought of as an application profile whose schema combines elements from multiple standards, with application-specific constraints (as in the use of specific controlled vocabularies or structured values). The PB Core must be understandable and usable by all public broadcasting entities, from the smallest local NPR radio station to the largest public television producers of national programming.

The PBMI's primary interest is in data exchange, data crosswalks, and interoperability, not necessarily in creating a complete metadata model that can be exploited by digital asset management systems for comprehensive, original cataloging and markup of essence. The Project desires to facilitate the sharing of metadata and the discovery of valued assets. The PB Core is intended to be "simple," but not "simplistic." Furthemore, the PB Core should be considered as a starting point that may accommodate metadata extensions of interest to specific communities and users.

Consequently, the Project undertook a path that would reflect the Public Broadcasting industry's needs and wants regarding media assets by gathering together representatives from public broadcasting and growing a consensus. The unique quality of public broadcasting, both television and radio, is its local ownership and local ties to its surrounding communities. In a parallel fashion, the Public Broadcasting Metadata Initiative was designed to tap into the various local constituencies and develop a metadata core from "grassroots" origins, rather than by administrative edict.

The Project conducted a detailed "needs assessment" of public broadcasters. Such measures are revealing and often unmask and articulate conditions, issues, needs, and desires that otherwise are dismissed or forgotten. By applying user-centered techniques PBMI was able to discover a wide spectrum of needs and applied the most appropriate metadata elements.

4. The Process for Assessing the Need and Gathering User Requirements

Public broadcasters have always endeavored to engage in complex and robust relationships with their constituents, whether those are viewers, listeners, educators, community leaders, etc. We have always provided extensive outreach for our broadcast content, with particular emphasis on the needs of K-12 teachers and lifelong learners. Today, with the advent of the Internet, that outreach is more significant and successful than ever before. As mentioned above, we also have an extremely complex structure; as opposed to our media counterparts, who increasingly concentrate their ownership and control of media outlets, very little of public broadcasting's operations are centralized. We have innumerable systems for producing and tracking our content, and our institutions are structured in a variety of ways, often based on who holds the broadcast license.

In order to ascertain the metadata needs of our "external" users – constituents – and "internal" users – local and national staff – we first created a list of users, and then double-checked this "strawman" with the core PBMI working group. A "User Requirements Team" was formed from within the working group. Using the now-modified user list, they set out to create a series of Use Case Scenarios. During this process, the "User Requirements Team" interviewed a large number of stakeholders, including national program distributors, local station broadcast operations and IT staff, a K-12 "learning object" consortium, an independent television production company, a television graphic artist, and "interactive" specialists (web and TV).

The interviews provided very useful feedback that helped define aspects such as the levels of granularity for the description of assets, the specificity with respect to the

number of elements, type of information to be described, such as rights, and encoding standards, e.g., XML. For example, what emerged from the interviews was a clear division between full-program metadata (such as title, format, date), which serves the needs of national distribution and local broadcast operations, and fragment, or clip-level data, which serves the needs of producers, educators, and website programmers. Most use case participants felt that it was critical to have a simple, intuitive set of metadata elements, with extensions for particular constituencies, e.g., K-12 curriculum-correlation, or graphics creation, so that the maximum number of assets could be identified and retrieved by the greatest number of individuals and institutions.

There was a great deal of concern about rights management, without which future business and service models crumble. Several interviewees felt that the working group should also determine standards for metadata exchange, such as XML.

5. The Process of Refining the PB Core

A powerhouse of motivated and opinionated experts was assembled to contribute to the Public Broadcasting Metadata Initiative Project. The members were drawn from a variety of communities related to public broadcasting:

- National public television organizations and program distributors
- National public radio organizations and program distributors
- Online Internet-based resource organizations
- National program producers
- State and regional network organizations
- Community radio and TV licensed stations
- University radio and TV licensed stations
- Educators
- Metadata subject matter experts

The initial work of the members for the Public Broadcasting Metadata Initiative Project lasted seven months. The overarching goal of the group was to recommend usable metadata fields that would facilitate the exchange of program and resource information between public broadcasting communities and other interested parties. Guiding our work process was the question, "How would a particular metadata element ultimately contribute to the discovery of public broadcasting's intellectual content by various end-users"? The objectives of the Working Group were to:

- Develop and refine user requirements for a sharable metadata element set.
- Review existing metadata schemas to determine their applicability to the public broadcasting arena, to identify gaps and overlaps, and to incorporate the most germane while discarding the least useful or confounding.
- Determine the scope and breadth of a usable metadata schema that was consensus-built, extensible, and interoperable with other asset management systems and databases.
- Draft a preliminary application profile of the public broadcasting core metadata of descriptors and their usage.
- Present the PB Core to the public broadcasting community for review and comment.
- Refine and revise the PB Core prior to release and publication.

In the seven-month time period, two full meetings of the entire Working Group were conducted, as well as follow-up committee work.

- First Meeting: 2002-4-24&25
- Committee Work: 2002-Summer, PB Core Review Team and User Requirements Team
- Second Meeting: 2002-9-12&13
- The Boston Summit: 2002-10-16,17,18

These activities led to an intensive three-day work session in Boston (2002-10-16,17,18), where the Public Broadcasting Metadata Core was refined and honed by the PB Core Review Team.

Before the Boston Summit, the PB Core Review Team had surveyed existing metadata dictionaries from various authorities and organizations, including those in use by several public broadcasting groups. A total of 467 separate metadata elements were compiled, which spawned 2335 recommendations for grouping and collapsing the elements into the most relevant. From these recommendations, a total of 249 working metadata elements and their qualifiers were selected.

The work of the PB Core Review Team at the Boston Summit combined redundant elements, discarded the less relevant, and debated the appropriate application of preferred metadata within the dictionary. The Summit yielded a preliminary draft of 58 metadata elements and their qualifiers that were most appropriate to public broadcasting and related communities. (For details see http://www.utah.edu/cpbmetadata/PBCore)

6. The Public Broadcasting Core Elements

Many of the 58 metadata elements selected for the Public Broadcasting Core of metadata descriptors were drawn from the Dublin Core Metadata Initiative. Others were retained from existing public broadcasting digital asset management systems in development. Still others were drawn from additional working groups.

The PB Core Elements could be placed into three categories or clusters:

- **Content:** 20 elements describing the actual intellectual content of a media asset or resource.
- **Intellectual Property:** 9 elements related to the creation, creators and usage of a media asset or resource.
- **Instantiation**: 29 elements that identify the nature of the media asset as it exists in some form or format in the physical world or digitally.

Table 1 reviews the 58 elements and qualifiers currently under consideration by the Public Broadcasting Metadata Dictionary Project. The Registration Authorities listed represent the agency of responsibility for the long term integrity and viability of particular metadata elements and associated qualifiers:

- DCMI: Dublin Core Metadata Initiative
- DC-Ed: DCMI Education Working Group
- ViDe: Video Development Initiative
- [PBCore]: Corporation for Public Broadcasting as Interim Steward
- [MPR]: Minnesota Public Radio as Interim Steward

Table 1. Recommended Metadata Elements of the Public Broadcasting Metadata Initiative Project

Element Name	**Registration Authority and Element Definition**
01.00 Title	DCMI: A name given to a resource, as well as any other title(s) that would be useful in uniquely identifying a resource and that would facilitate discovery and retrieval.
01.01 Title.Alternative	DCMI: An Alternative Title is used in order to identify an asset or resource that has a title similar to the proper title, but which further assists in discovery and retrieval.
01.02 Title.Series	[PBCore]: A Series Title is one specifically identified by the video or audio production agency and is named as such in order to facilitate discovery and retrieval, as well as to more accurately reflect how a resource's title fits into a hierarchy of proper titles that are used to describe it.
01.03 Title.Episode	[PBCore]: An Episode Title is one specifically identified by the video or audio production agency and is named as such in order to facilitate discovery and retrieval, as well as to more accurately reflect how a resourceis title fits into a hierarchy of proper titles that are used to describe it.
01.04 Title.Program	[PBCore]: A Program Title is one specifically identified by the video or audio production agency and is named as such in order to facilitate discovery and retrieval, as well as to more accurately reflect how a resource's title fits into a hierarchy of proper titles that are used to describe it.
02.00 Creator	DCMI: An entity primarily responsible for making the content of the resource or asset. May be a person, business, organization, group, initiative or service.
02.01 Creator.Role	[PBCore]: Unlike print resources, there is no single role, such as author, that is commonly understood to have primary responsibility for the intellectual content of many resources, such as audio, video or film assets. In such cases, creators can include many different roles deemed to have primary responsibility for the creation of the essence, such as the instructor for a video course, the interviewee from a video history program, or the director of a feature film.
03.00 Subject	DCMI: The topic(s) of the intellectual content of a resource or asset. Contains controlled values and uncontrolled values (keywords). Use the Description element for more free-form text descriptions of a resource.
04.00 Description	DCMI: An account of the intellectual content of the resource. Descriptions are more free-from text entries when compared to the controlled vocabularies associated with the Subject element.
04.01 Description.Abstract	DCMI: As an account of the content of the resource, the qualifier Abstract is a short narrative summary of the topic of the resource. Provides additional supplied text by experts that adds color or insight to the description of the resource or asset not otherwise identified in the more specific content related fields. Anecdotal comments welcomed.

Element Name	Registration Authority and Element Definition
04.02 Description.Table of Contents	DCMI: As an account of the content of the resource, the qualifier Table of Contents is used for partial or full listings of subunits of the resource. Use the Table of Contents to identify other descriptive information such as: Composers and Works contained in a program; Cue Sheets; Play Lists; Rundowns; Edit Decision Lists (EDLs) (unformatted); Content Flags; Index of Sections or Segments; Formal Table of Contents.
04.03 Description.ProgramRelatedText	[PBCore]: As an account of the content of the resource, the qualifier ProgramRelatedText identifies other audio and textual representations of the main audio or language presentation mode for a resource or asset.
05.00 Publisher	DCMI: An entity responsible for distributing or making a resource available to other end-users and communities. May be a person, business, organization, group, initiative or service. Some resources may not have a publisher or distributor, and thus will not have an entry under Publisher.
05.01 Publisher.Role	[PBCore]: The Role that is played by a specific Publisher or Publishing entity is identified.
06.00 Contributor	DCMI: An entity responsible for making contributions to the content of the Resource, but whose contribution is secondary to any entity specified in the Creator element (for example, film editor, screenwriter, narrator). Examples of Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
06.01 Contributor.Role	[PBCore]: The Role which a Contributor plays is identified here. Use this element and qualifier to identify important production credits for a resource, e.g., producer, director, writer, special thanks, funding agencies, programmers, designers, graphics, instructional design, etc.
07.01 Date.Created	DCMI: The creation date for a resource or program.
07.02 Date.Issued	DCMI: Date of formal issuance (e.g. publication) of a resource for general public consumption.
07.03 Date.AvailableStart	DCMI: A specific start date for a resource's availability. May refer to start dates for the availability of a program that is broadcast locally, regionally, nationally or internationally.
07.04 Date.AvailableEnd	DCMI: A specific date that a resource’s availability has come or will come to an end. May refer to end dates for the availability of a program that is broadcast locally, regionally, nationally or internationally.
08.00 Type	DCMI: The nature or genre of the content of the resource, or the purpose for which the asset was created and made available.
08.01 Type.Form	[PBCore]: A format or program category for a resource.
08.02 Type.Genre	[PBCore]: The nature or genre of the content of a resource.
09.01 Format.Physical	[PBCore]: A physical manifestation of a resource as it may exist as a format or carrier that occupies physical space dimensions.
09.02 Format.Digital	DCMI: A digital instantiation of a resource that may or may not have existed originally in an analog, physical form. Digital media formats may be expressed as formal Internet MIME types or as other means of expressing the format of a digital resource.
09.03 Format.Identifier	[MPR]: Identifying information about the format of a resource.
09.04 Format.FileSize	ViDe: Measures the storage requirements or file size of a digital resource in Bytes, Kilobytes, Megabytes or Gigabytes to provide the most meaning to the end user.
09.05 Format.AudioBitDepth	[PBCore]: For a program or resource, this qualified element measures an audio signal in a number of bits and answers the question, ‘How Much’ data is allocated to a digital sampling of an audio signal. Provides information important for identifying retrieval and playback/display requirements for a resource.

Element Name	Registration Authority and Element Definition
09.06 Format.AudioChannelConfiguration	[PBCore]: Indicates the number of audio channels configured for the playback of a resource.
09.07 Format.AudioDataRate	[PBCore]: Expressed as amount of data per second and indicates how much data is delivered through a particular delivery pipeline for every second.
09.08 Format.AudioSamplingRate	[PBCore]: Measured in kiloHertz for a program or resource, this qualified element quantifies 'How Much' data is allocated to a digital sampling of an audio signal. Provides information important for identifying retrieval and playback/display requirements for a resource.
09.09 Format.ImageAspectRatio	[PBCore]: Indicates the ratio of horizontal to vertical proportions in the display of an image or moving image.
09.10 Format.ImageBitDepth	[PBCore]: For a program or resource, this qualified element measures a still or moving image in terms of the number of bits in a sample, and answers the question, How Much data is allocated to a digital sampling. Provides information important for identifying retrieval and playback/display requirements for a resource.
09.11 Format.ImageChannelConfiguration	[PBCore]: Indicates the number of image channels available in a resource. May be most appropriate for digital files, like QuickTime in which multiple video tracks can be encoded in a single file.
09.12 Format.ImageColorCode	[PBCore]: Indicates the color or lack of color in an asset. Does not measure the specific color metrics of a image or moving image.
09.13 Format.ImageDataRate	[PBCore]: Expressed as amount of data per second and indicates how much data is delivered for an image or moving image through a particular delivery pipeline for every second.
09.14 Format.ImageFrameRate	[PBCore]: Indicates the frames per second found in a resource's playback or display.
09.15 Format.ImageFrameSize	[PBCore]: Indicates the horizontal and vertical resolution of a format type. May be expressed in pixels, pixels per inch, or in the case of ATSC digital TV, a combination of pixels measured horizontally vs. the number of lines of image data stacked vertically (interlaced and progressive scan).
09.16 Format.TimeStart	[PBCore]: Indicates a time stamp representing the beginning point for the playback of a resource. Use in combination with Format.Duration to identify a sequence or segment of a resource that has a fixed start time and end time.
09.17 Format.Duration	ViDe: Describes the duration in time units for a resource, if that resource has an identifiable, linear start-to-end playback. Format.Duration does not describe the time required to utilize a resource in a setting, but is rather a strict playback time, TimeStart to TimeEnd.
09.18 Format.Standard	[PBCore]: The standard refers to an overarching architecture for underlying media formats.
09.19 Format.Type	[PBCore]: The Qualifier of Type is hierarchically a subset of the values found under Format.Standard and describes specific kinds of media formats found for each media standard.
09.20 Format.Encoding	[PBCore]: This proposed element with qualifier is designed to offer a single element with which the various media standards and their collected format types can be identified for a particular resource.
10.00 Identifier	DCMI: An unambiguous reference or identifier for a resource within a given context. Best practice is to identify a resource by means of a string or number corresponding to an established or formal identification system.
11.00 Source	DCMI: A reference to another resource from which the present resource is derived.
12.00 Language	DCMI: The primary language of the intellectual content of the resource, usually expressed by the audio track. If other, alternative audio and textual representations of the main audio or language presentation mode exist for a resource or asset, describe that information in the Language.Usage element.
12.01 Language.Usage	[PBCore]: The qualifier Language.Usage identifies the existence of other audio and textual representations of the main audio or language presentation mode for a resource or asset.

Element Name	Registration Authority and Element Definition
13.01 Relation.Type	[PBCore]: Relation.Type identifies a second resource that is related to the primary resource. It defines the relationship between the second resource and the primary resource. While the primary resource is described by the rest of the asset management's database record, the second resource is described using the Relation field.
13.02 Relation.Identifier	[PBCore]: Identifies a second resource related to the primary resource by using a specific numbering or labeling scheme to call out the related resource. Used in combination with the Relation.Type element to cross reference the type of relation with a unique identifier for that relation.
14.01 Coverage.Spatial	DCMI: Identifies the extent or scope of the resource's content from a spatial or geographical perspective of the intellectual content of a resource. Coverage.Spatial is used for geographic coordinates of maps and map-like images (e.g. aerial maps or map-like images concatenated as a video file) or to associate place names or logical jurisdication for a resource.
14.02 Coverage.Temporal	DCMI: Identifies the extent or scope of the resource's content from the perspective of the temporal or time characteristics of the intellectual content of a resource. CoverageTemporal is used for date and time-based events, designated numerically for precision searching, where the time element is critical for identification and use of the resource.
15.00 Rights.Usage	[PBCore]: Information about rights held in and over the resource, particularly in what manner the resource will be used, eg., broadcast, web, PDAs, or education/classroom.
15.01 Rights.Reproduction	ViDe: Statements or references about rights held in and over a resource, specifically regarding the rights to reuse, repurpose or reproduce a resource.
15.02 Rights.Access	ViDe: Access information about rights held in and over a resource. Rights.Access indicates either 'open access' or 'restricted access.' These two options are used as flags to trigger certain actions. For example, metadata records with 'restricted access' will not be exposed for mining by OAI initiatives.
16.01 Audience.Level	DC-Ed: A general statement describing the education or training sector. Alternatively, a more specific statement of the location of the audience in terms of its progression through an education or training sector or level.
16.02 Audience.Rating	[PBCore]: Designates categories of users for whom the resource is intended or judged appropriate. Standard ratings have been crafted by the broadcast television industry which are useful.
18.00 Annotation	[PBCore]: General field to be used to append helpful information for the metadata markup team about an asset and its metadata.
19.00 Location	[PBCore]: Identifies the location of a specific format or instantiation of a resource. Usually a text string describing where in the physical world the resourceís physical format resides. This is not an ID number, but a location identifier. Used when the actual physical form is being indexed in an asset management system.
99.00 Special Extensions	Extensions are additional descriptions for media resources that have been crafted by organizations outside of the PBCore development initiative. These extensions fulfill the metadata requirements for these outside groups as they identify and describe their own types of media with specialized, custom terminologies. Eg: DC-Ed Audience; Audience.Mediator; Standard; Standard.Identifier; Standard.Version IEEE LOM InteractivityType; InteractivityLevel; TypicalLearningTime

7. The Application Profile

The 58 elements are delineated by 15 attributes according to the modified ISO 11179 Specification and Standardization of Data Elements [14]. The full accounting of the specification is too large a document to include in this paper.

1. Element Number
2. Element Name
3. Version of the Element

4. Element Label
5. Definition
6. Namespace Identifier
7. Registration Authority
8. Language of the Element
9. Obligation in Usage
10. Data Type
11. Maximum Occurrence
12. Encoding Schemes
13. Restricted Values
14. Examples
15. Usage Guidelines

PBMI's interest is in data exchange, data crosswalks, and interoperability, not necessarily in creating a complete metadata model that can be exploited by digital asset management systems for comprehensive, original cataloging and markup of essence. Consequently, the primary desire of PBMI is to facilitate the sharing of metadata and the discovery of valued assets. Within the Application Profile, issues of concern to PBMI are:

- Who will serve as the real registration authority that takes responsibility for the declaration and maintenance of our newly defined, custom elements and their qualifiers that are not already part of a standard?
- Who will publish versions of the Core and its updates? Provide documentation? Provide Technical Support?
- How will we monitor adoption and compliance?
- How will we measure successful implementation?

The Project recognizes that it needs to remain focused on the fact that the Working Group is not a body of "standards makers." Rather, we are "real life implementers" who are tasked with generating effective solutions in order to service the efficient and widespread delivery of public broadcasting's intellectual content. Similar to our day-to-day business, we are engaged in applied and practical solution-making.

Like many other groups debating the application of metadata schemes, the Project remains conflicted in how best to match metadata descriptors with various instantiations of essence and assets. The question of embracing a "one-to-one" relationship between a metadata record and its associated essence or subscribing to a "one-to-many" relationship between a metadata record and the various instantiations of its essence still plagues the PBMI Project. Compelling arguments have been presented on both sides of the issue. We are hopeful that the next phase of our project, a Request for Comments, will assist us in sorting out a solution.

8. Feedback and Evaluation Mechanisms

To a great extent, the work of the Public Broadcasting Metadata Working Group has modeled an unheard-of process – coordination and consensus across vastly different institutions, on a topic of extreme detail and importance. The Preliminary PB Core is ready to be reviewed and tested.

During the next several months the Working Group will be asked to engage in an even more difficult process – a mid-course evaluation.

The group will be divided into task teams, and through research, interviews, conference calls, and "thought papers," will address the following issues and objectives:

- determine that the PB Core is sustainable over time (including knowing its lifespan, form, cost, etc. and how stations and producers can be made to comply with the protocol);
- ensure that the PB Core's strategic value is understood and acknowledged by senior management who will need to support it;
- devise a plan to integrate the PB Core into the day-to-day operations within local and national content infrastructures, especially the PBS Next Generation Interconnection System and NPR's Content Depot.

It is our assumption that these difficult questions will be answered in a manner that leads the project to the RFC (Request for Comments) process, and then test implementations in typical metadata scenarios.

The RFC process will include other public broadcasting production, IT and broadcast operations staff, key software vendors serving the industry, standards organizations, partnering institutions, etc.

Test implementations of the PB Core, still to be determined, will likely include radio, television and website production collaborations, tape libraries, national program distribution systems, as well as national producers of content. Consideration will be given to additional test participant(s) whose products, services and initiatives are used by, and/or relate to public broadcasting stations and organizations.

9. Next Steps

The PBMI process has illuminated for participants and observers alike the critical need for a new, "advanced networking" approach toward conducting our core activities. We must change our institutions and infrastructures, even our funding models, to reflect a new spirit of exchange, collaboration and consolidation. Certainly, without Internet-like standards for descriptive and administrative metadata, rich media file formats, file exchange, etc., we will not be able to keep pace with changes in the media environment, nor will we advance our public service mission.

References

[1] Dublin Core Metadata Element Set, Version 1.1: Reference Description. Retrieved April 15, 2003, from http://www.dublincore.org/documents/dces/

[2] OAIS: Reference Model for an Open Archival Information System. Retrieved April 15, 2003, http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html

[3] SMEF-DM: Standard Media Exchange Framework - Data Model. Retrieved April 15, 2003, http://www.bbc.co.uk/guidelines/smef/

[4] MPEG-7: "Multimedia Content Description Interface" ISO/IEC JTC1/SC29/WG11. Retrieved April 15, 2003, from http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm

[5] Hunter, J. "A Proposal for the Integration of Dublin Core and MPEG-7", ISO/IEC JTC1/SC29/WG11 M6500, 54th MPEG Meeting, La Baule, October 2000. Retrieved April 15, 2003, from http://archive.dstc.edu.au/RDU/staff/jane-hunter/m6500.zip

[6] Hunter, J. ViDE Video Access Group, "An Application Profile which combines DC and MPEG-7 for Simple Video Description", February 12, 2002. Retrieved April 15, 2003, from http://archive.dstc.edu.au/RDU/staff/jane-hunter/publications.html

[7] Agnew, G. (2003) A Tale of Two Schemas: Mapping Dublin Core to MPEG7. Retrieved May 15, 2003, from http://gondolin.rutgers.edu/MIC/text/how/mpeg7DC_agnew.pdf

[8] MARC: Machine Readable Cataloging. Retrieved April 15, 2003, from http://www.loc.gov/marc/

[9] METS: Metadata Encoding and Transmission Standard. Retrieved April 15, 2003, from http://www.loc.gov/standards/mets/

[10] SCORM: The Sharable Content Object Reference Model. Retrieved April 15, 2003, from http://www.adlnet.org/

[11] IEEE Learning Technology Standards Committee (LTSC). Learning Object Metadata. Draft Document v3.6. Retrieved April 15, 2003, from http://ltsc.ieee.org/doc/wg12/LOM3.6.html

[12] Sutton, S.A. (1999) IEEE 1484 LOM mappings to Dublin Core: *Learning Object Metadata: Draft Document v3.6*, IEEE Learning Technology Standards Committee (LTSC), 5 September 1999. Retrieved April 15, 2003, from http://www.ischool.washington.edu/sasutton/IEEE1484.html

[13] IMS Global Learning Consortium. IMS Meta-Data v1.2.2. Retrieved April 15, 2003, from http://www.imsproject.org/metadata/index.cfm

[14] ISO 11179 Specification and Standardization of Data Elements. Retrieved January 11, 2003, from http://www.diffuse.org/meta.html#ISO11179.

Appendix: Public Broadcasting Metadata Initiative Project Participants

At Large

Marcia Brooks, WGBH (Project Director)
Dennis Haarsager, KWSU
Amy Rantanen, WGBH
Alison M. White, Corporation for Public Broadcasting

Advisors

Grace Agnew, Rutgers University (AMIA)
Judy Brown, University of Wisconsin, DOD Academic CoLab (SCORM Project)
Efthimis N. Efthimiadis, The Information School, University of Washington, Seattle, WA

Working Group

Alan Baker, Minnesota Public Radio
Nancy Baldacci, American Public Television
Sharon Blair, AMIA Local Television Task Force
Marty Bloss, National Public Radio Public Radio Satellite Service (NPR PRSS)
Paul E. Burrows, Media Solutions, OIT, University of Utah
Brian Callahan, WHRO (*former participant)*
Michael Connet, onCourse
David Felland, Milwaukee Public Television
Tom Handy, KWSU (*former participant)*
Steven Heard, Public Interactive
Rob Holt, NPR Online
Dave Johnston, PBS Online
Ann Lootens, WGBH
Dave MacCarn, WGBH
Chuck McConnell, NETA/OSBE
Bea Morse, Public Broadcasting Service (PBS)
Robin Mudge, onCourse (*former participant)*
Lesley Norman, David Grubin Productions (*former participant)*
Meg O'Hara, WNET
Tim Olson, KQED
Marilyn Pierce, Public Broadcasting Service (PBS)
Richard Ruotolo, Public Radio International (PRI)
James Steinbach, Wisconsin Public Television
Michael Tondreau, Oregon Public Broadcasting
Brent Trinacty, Public Radio International (PRI) (*former participant)*
Cate Twohill, Public Broadcasting Service (PBS)
Steven Vedro, Consultant
Tracy Vosburgh, WPSX (Penn State University)
Michael Yoch, NPR Online (*former participant)*
Art Zygielbaum, Nebraska Education Television

Additional Support Provided By

Scott Bridgewater, National Public Radio Public Radio Satellite Service (NPR PRSS)
Carrie Lowe, Public Broadcasting Service (PBS)
Thom Shepard, WGBH
and other staff at many of the participating organizations listed above