Metadata Schema Report

Mikaela Slade

LIS 882 – Metadata for Digital Objects

Fall 2020 Semester

Dominican University

Encoded Archival Description as a Metadata Schema

**History**

In 1995, Daniel Pitti introduced Encoded Archival Description at the Society of American Archivists’ annual meeting. It first began development in 1992 at the University of California, Berkeley as a response to the growth of the Internet. EAD was “developed specifically for encoding archival finding aids for presentation of the World Wide Web” (Dow, 2009). Previously, paper finding aids were used but online access tools allowed archival systems to increase access to archival holdings. Finding aids are individual archival descriptions and are the primary way of searching and accessing archival content.

Archivists were able to expand access somewhat from paper finding aids in the 1970s and 1980s due to the development of MARC AMC (Machine-readable Cataloging for Archives and Manuscript Control). This enabled archivists to make collection-level summary descriptions available in online catalogs and “disseminate information about their holdings more widely via national bibliographic systems,” but the full, detailed descriptions were still only available in printed forms (EAD Application Guidelines, 1999). Although MARC AMC allowed archival records to be searched similarly to published materials, it wasn’t able to accommodate more than a summary. EAD hoped to better represent the richness of archival description data.

**The Significance of EAD**

XML was released in 1998 by the W3C and was quickly embraced by EAD users and shifted their work from SGML to XML. XML and other markup technology can support documents with unique and in part mutually exclusive kinds of data, as well as “many-to-many relations” and the order of data elements has “minimal if any significance”, making it flexible to work with and semantically rich (Pitti, 2012, p .690). XML is also widely adopted and supported. Currently, many archival institutions publish finding aids online using HTML and PDF formats. While HTML and XML are both markup languages, HTML is used for displaying data and XML is used for transporting it (“HTML vs XML,” 2019). Daniel V. Pitti (2012), one of the major developers and proponents of EAD, states that XML is “the superior representation” for “the purposes of communicating information, as opposed to its creation and maintenance” (p. 690). Partially due to the adoption of XML, the 2nd edition of EAD was released in 2002.

Encoded Archival Description is an international digital communication standard for the coding of archival records. It is a standard for communicating archival descriptions and not a standard for the content like ISAD(G). Unlike libraries, archives and archival records document people, organizations, or other entities that perform activities or functions and books/journals, etc. are the product of those activities. These records “serve as instruments for carrying out or documenting the performance of activities” (Pitti, 2012, p. 686). Archives have a responsibility on behalf of those entities to preserve and describe those functions. Due to this, archives focus on fonds, rather than individual items, and each is unique. Fonds are a collection with a hierarchical description of records that share a common origin, also known as provenance. EAD is a standard created specifically designed for archival practices.

**How EAD is Used**

EAD is both hierarchical and multilevel. Finding aids begin at the top level with the most descriptive components and amount of data and the data and description decrease with each subsequent level. It has four basic components: fonds, series, file, and item. The description of the collection begins with a description of the whole and then describes components of the whole based on shared characteristics, which in turn can also be components. Description at any level applies exclusively to that level. Description at each level inherits description of the parent level.

Fonds is the top level, the broadest category, and contains information about the entire collection. The top level has three main sections: the EAD Header, Front Matter, and Archival Description. The EAD Header contains control information like EAD identifier, which is a unique designation for identifying and tracking EAD document files, and File description (title information, edition, etc.). The second section, Front Matter, is an optional section containing information for use in formal publication. The final section, Archival Description, represents logical semantics and structure for description (like headings and paragraphs).

Other important EAD elements include <dsc> “Description of Subordinate Components” and <did> “Descriptive Identification”. “Description of Subordinate Components” is a “wrapper element that bundles information about the hierarchical groupings of the materials being described” (*Encoded Archival Description Tag Library - Version 2002*, 2006). It includes components that can be simple, <c>, or enumerated, <c01> through <c12>. Each component is repeatable, recursive, and “all components of description are available at each level of descriptive analysis” (Pitti, 2012, p. 692). The “Descriptive Identification” element is “a required wrapper element that bundles other elements identifying core information about the described materials in either Archival Description <archdesc> or a Component <c>” and “this grouping ensures that the same data elements and structure are available at every level of description within the EAD hierarchy” (*Encoded Archival Description Tag Library - Version 2002).* This assists retrieval in resource discovery.

**EAD Example**

This example is taken from the Encoded Archival Description official site Description Tag Library’s entry for “Descriptive Identification” (2006) shows how the <dsc> and <did> elements function as well as demonstrating the hierarchical and multilevel nature of EAD:

<dsc type="combined">

<c01 level="series">

**<did>**

<unittitle>Series 1: Correspondence,</unittitle>

<unitdate type="inclusive">1943-1978</unitdate>

<physdesc><extent>2.5 linear ft. </extent>(5 document boxes)</physdesc>

**</did>**

<scopecontent>[...]</scopecontent>

<c02 level="subseries">

**<did>**

<unittitle>Subseries 1.1: Outgoing Correspondence, </unittitle>

<unitdate type="inclusive">1943-1969</unitdate>

<physdesc><extent>0.75 linear ft.</extent></physdesc>

**</did>**

<c03 level="file">

**<did>**

<physloc audience="internal">B:14:D</physloc>

<container type="box">1</container>

<container type="folder">1</container>

<unittitle>Abbinger-Aldrich</unittitle>

<physdesc><extent>14 letters</extent></physdesc>

**</did>**

</c03>

</c02>

</c01>

</dsc>

**Resources**

* Encoded Archival Description Tag Library - <https://www.loc.gov/ead/tglib/index.html>
* Frequently Asked Questions about EAD – <https://www2.archivists.org/groups/encoded-archival-standards-section/frequently-asked-questions-about-ead-and-ead3>
* RLG EAD Advisory Group’s Best Practice Guidelines for EAD - <https://www.oclc.org/content/dam/research/activities/ead/bpg.pdf>

**Conclusion**

EAD is a metadata schema that was built for archivists as a way to bring finding aids online. As a standard it is flexible, semantically rich, and able to accommodate multiple levels of description. Despite that, many archival organizations, especially in America, have yet to adopt it. Yakel and Kim (2005) found that archival systems are less likely to want to adopt a universal standard than libraries or similar institutions because of the unique nature of archival holdings. However, there are significant benefits to implementing standards such as increasing discoverability, sharing, and publication tools. In fact, archival institutions are much more likely to use EAD if part of a consortia that utilizes it. The future of EAD lies in building off these advantages and embracing technologies that allow for more dynamic and “relational” systems (Pitti, 2012).

**References**

Dow, E. (2009). Encoded Archival Description as a Halfway Technology. *Journal of Archival Organization*, *7*(3), 108–115. <https://doi.org/10.1080/15332740903117701>

*EAD application guidelines for version 1.0: Setting EAD in context*. (1999). Society of American Archivists. <https://www.loc.gov/ead/ag/agcontxt.html>

*Encoded archival description tag library—Version 2002*. (n.d.). EAD Official Site; Library of Congress. <https://www.loc.gov/ead/tglib/index.html>

*Frequently asked questions about EAD and EAD3* . (n.d.). Encoded Archival Standards Section; Society of American Archivists. https://www2.archivists.org/groups/encoded-archival-standards-section/frequently-asked-questions-about-ead-and-ead3

HTML vs XML. (2019, April 22). *GeeksforGeeks*. <https://www.geeksforgeeks.org/html-vs-xml/>

Pitti, D. V. (2012). Encoded Archival Description (EAD). In M. J. Bates (Ed.), *Understanding information retrieval systems management, types, and standards* (pp. 685–697). essay, CRC Press, Taylor & Francis Group.

RLG EAD Advisory Group. (2002). *RLG Best Practice Guidelines for Encoded Archival Description.* Retrieved from <https://www.oclc.org/content/dam/research/activities/ead/bpg.pdf>

Yakel, E., & Kim, J. (2005). Adoption and diffusion of Encoded Archival Description. *Journal of the American Society for Information Science & Technology*, *56*(13), 1427–1437. <https://doi-org.dom.idm.oclc.org/10.1002/asi.20236>

PBCore as a Metadata Schema

**History**

At the beginning of the 21st Century, Public Broadcasting as an industry was rapidly changing at the onset of digital technologies. In 2001, the Corporation for Public Broadcasting met and founded a project known as the Public Broadcasting Dictionary Project (PBMD) (Rubin, 2012, p. 2). The goal was to adopt a single set of metadata protocols known as Public Broadcasting Core (PBCore). This group spent more than two years looking at existing metadata standards as well as the needs of their projected users. PBCore was developed based off the Dublin Core elements in addition to elements from other schemata and local digital asset management systems.

PBCore 1.0 data dictionary was published in 2005 and was presented as a “starter kit” for asset and content management and even audiovisual preservation. Public Broadcasting groups were slow to embrace PBCore at first but outside groups such as moving image and media archivists took notice and began to implement it in their own institutions. Even during 2007-2009, when funding from the CPB was stagnant, communities of users kept PBCore active. PBCore became “large enough to support itself through information sharing and ‘unofficial’ websites” (Rubin, 2012, p.6). PBCore 2.0 was released in 2011 after a renewed push and re-allocated funding from CPB. Most recently, in May 2017, the WGBH Media Library and Archives received a grant to pursue the PBCore Development and Training project, “designed to develop tools, methodologies, and training workshops to make PBCore more accessible to archivists and public media organizations” (WGBH Media Library & Archives, 2017?).

**The Significance of PBCore**

Although there are other metadata schemas that can handle audiovisual collections, PBCore was designed specifically for Public Broadcasting, which is a distinctive industry itself. Unlike traditional media sources, “the unique quality of public broadcasting, both television and radio, is its ownership and local ties to its surrounding communities” (While et al., 2003). Public Media can range from the small and local to large and national: “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”. The Public Broadcasting Metadata Initiative created PBCore to provide a standard for the Public Broadcasting community in order that metadata could be more easily identified, shared, and organized. The goal of PBMI was to “create a schema that is easily understood, implemented and adopted by the Public Broadcasting community at large.”

Another important aspect is that the majority of those who work in Public Broadcasting are not information professionals and do not have cataloging knowledge. While examining existing multimedia standards such as MPEG-7 and MARC, PBCore developers realized that there was often a high technical barrier to entry. The creators sought PBCore to be “’simple’ but not ‘simplistic’” so it could be implemented by a variety of users and purposes (While et al., 2003). While identifying PBCore’s community of users and their potential needs, the creators observed: “a clear division between full-program metadata..., 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”. PBCore was created by “implementers, not standard makers,” intimately familiar with Public Media. They understood that PBCore must be flexible enough to be used at all levels and rigid enough to be shared with outside groups. The true advantage of PBCore is as a clear metadata standard that improves interoperability both inside and outside an organization.

**How PBCore is Used**

PBCore is used by more than just Public Broadcasting groups because it is good for describing digital media. It can include file URLs, can serve as syndication format like RSS or ATOM but in more detail, and contains rich metadata on provenance and location, essential aspects of digital preservation (Rubin, 2012, p. 3). With 15 containers, 82 elements, and available use of 49 XML attributes, PBCore 2.1 is able to describe and categorize media items that can then be retrieved. PBCore elements are divided into three categories: Root, Asset, and Instantiation. Asset elements are, broadly, descriptors for intellectual content such as title, subject, coverage, etc. Instantiation elements are “technical metadata about the physical or digital representation of the AV asset” (Elements, (n.d.)). Not all elements are mandatory or non-repeatable but some. Similarly, some sub-elements may be required, and most elements have the option of attributes for added detailed description. Importantly, elements must be in the correct order for validation.

Each PBCore XML document has only one root element: “It encloses all the other elements and is therefore the sole parent element to other elements” (*Elements*, n.d.). Examples of root elements include: pbcoreCollection, pbcoreDescriptionDocument, and pbcoreInstantiation. These reflect the three main ways of structuring a PBCore XML document (WGBH Media Library & Archives, 2017?). The most common way is by containing all the elements within a pbcoreDescriptionDocument. This allows users to describe intellectual content of an asset using asset-level sub-elements. The second way to form a document is by using the root element pbcoreCollection where one or many Description Documents are contained within the element for serialization. Finally, the third way is to use pbcoreInstantiationDocument as the root element by which only instantiation elements may appear. This can be used as a means of recording technical metadata about digital files in a collection and can be used to track physical information about archival collections for which little to no content information is available.

**Examples**

All examples are taken from WGBH’s PBCore 2.1 Handbook (2017?).

**Example 1 – Description Document**

<?xml version="1.0" encoding="UTF-8" standalone="no"?>

<pbcoreDescriptionDocument xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html"

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html

https://raw.githubusercontent.com/WGBH/PBCore\_2.1/master/pbcore-2.1.xsd">

<pbcoreIdentifier source="PBCore Handbook">00001</pbcoreIdentifier>

<pbcoreTitle>Hamlet</pbcoreTitle>

<pbcoreDescription>Filmed production of Shakespeare's Hamlet</pbcoreDescription>

</pbcoreDescriptionDocument>

**Example 2 – Collection Document**

<?xml version="1.0" encoding="UTF-8" standalone="no"?>

<pbcoreCollection xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace"

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace

https://raw.githubusercontent.com/WGBH/PBCore\_2.1/master/pbcore-2.1.xsd">

<pbcoreDescriptionDocument>

<pbcoreIdentifier source="TGIF">19993</pbcoreIdentifier>

<pbcoreTitle titleType="Series">Sabrina the Teenage Witch</pbcoreTitle>

<pbcoreDescription>Fragment of a recorded episode of a television series

about a teen witch and her wacky aunts</pbcoreDescription>

</pbcoreDescriptionDocument>

</pbcoreCollection>

**Example 3 – Instantiation Document**

<?xml version="1.0" encoding="UTF-8" standalone="no"?>

<pbcoreInstantiationDocument

xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html"

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html

https://raw.githubusercontent.com/WGBH/PBCore\_2.1/master/pbcore-2.1.xsd">

<instantiationIdentifier source="PBCore

Handbook">00001</instantiationIdentifier>

<instantiationLocation>Library of Congress</instantiationLocation>

</pbcoreInstantiationDocument>

**Resources**

The website, PBCore.org is an expansive resource that covers Schema, Resources, and Community.

* PBCore Website: <https://pbcore.org/>
* PBCore Elements: <https://pbcore.org/elements>
* PBCore Handbook: <http://pbcore.org/handbook>
* PBCore Google User Group: <https://groups.google.com/g/pbcore-talk?pli=1>

**Conclusion**

PBCore is a successful audiovisual metadata standard because it was built with the users in mind. Since Public Broadcasting has such a wide user base, PBCore had to be flexible and powerful without requiring a high-level of technical training. Due to its structure, PBCore can be used in multiple ways for a variety of purposes and was designed with interoperability in mind. It has an active user group that spans outside the Public Broadcasting sphere. It is a solid schema for cataloging, preservation and publication of audiovisual material and archives.

**References**

*Elements*. (n.d.). PBCore Metadata Standard. Retrieved from https://pbcore.org/elements

*PBCore metadata standard*. (n.d.). PBCore Metadata Standard. Retrieved from <https://pbcore.org/>

*PBCore talk—Google groups*. (n.d.). Retrieved from https://groups.google.com/g/pbcore-talk?pli=1

Rubin, N. (2012). The PBCore metadata standard: A decade of evolution. *Journal of Digital Media Management*, *1*(1), 55–68.

WGBH Media Library & Archives. (2017?). *PBCore: PBCore Handbook.* Retrieved from <http://pbcore.org/handbook>.)

While, A., Baker, A., Bloss, M., Burrows, P., Efthimiadis, E., Brooks, M., MacCarn, D., Shepard, T., & Twohill, C. (2003). PB Core--the Public Broadcasting Metadata Initiative: progress report. *International Conference on Dublin Core and Metadata Applications, 0*, pp. 213-222. Retrieved from <https://dcpapers.dublincore.org/pubs/article/view/749/745>.