Standards for Archives

by Morag Boyd

Information Standards

EDITOR'S SUMMARY

Formal standards and professional practices characterize modern archival administration, increasing consistent archival description and interoperable metadata as well as the authenticity and reliability of the archives themselves. The International Council of Archives' General International Standard Archival Description identifies 26 data elements to describe archives, being extended for the semantic web. Archives in the United States follow three sets of archival description standards. *Describing Archives: A Content Standard, 2nd Edition* (DACS), used together with the other standards and with MARC, describes archival materials and authority records about material creators. It stresses principles for arrangement, order and hierarchical organization. The *Encoded Archival Description* (EAD) contains elements to describe archival materials and interrelationships. Like DACS, it stresses respect de fonds, keeping records together in original order. *Encoded Archival Context-Corporate Bodies, Persons, and Families* (EAC-CPF) describes information about people and organizations reflected in an archive. It was adopted by the Society of American Archivists in 2011 and has been used to derive 6.6 million EAC-CPF records from EAD finding aids and authority records. Archival descriptions are complex and unique. Using standardized and required descriptive elements and special search interfaces would maximize the advantage of EAD encoding and extend opportunities for data sharing between institutions.

KEYWORDS

archives

archival science

standards

data curation

interoperability

Encoded Archival Description

Morag Boyd is associate professor in the Ohio State University Libraries. She can be reached at boyd.402<at>osu.edu.

Records, papers and manuscripts have been collected, organized and retained by institutions for their long-term value probably for as long as there have been records. As for other information institutions, the quantity of materials in archives along with their diverse user communities requires systematic handling of the collections. Modern archival administration has been marked by its consolidation around recognized professional practice and the formal adoption of standards. In particular, standards for the encoding and content of archival description have led to more consistent and interoperable archival metadata.

To be clear about the scope of archival standards, archives are defined by the Society of American Archivists (SAA) as follows:

> … materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control; permanent records [1].

Archives vary widely in format and in the nature of their content, representing the range of human activities that generate documents in the broadest sense. They are collected in repositories (themselves often referred to as *archives*). An essential role of these institutions, as noted by the International Council of Archives (ICA), is to ensure that an archive is authentic and reliable and that the context of creation of has been documented [2]. Standards-based practice for the description of archives is one way that an archival repository can fulfill this duty.

International Context

Internationally, the theory and principles of archives have different histories in countries and regions. Despite these differences, the ICA reached

agreement on a core set of 26 data elements for the description of archives with the publication of the *General International Standard Archival Description* - ISAD(G) in 1993, now in its second edition [3]. ISAD(G) is both a schema – defining a set of elements for description – and a content standard, as it also offers guidance for how to provide data within the element set. There are six required elements:

1. Reference code
2. Title
3. Name of Creator
4. Dates of Creation
5. Extent of the Unit of Description
6. Level of description

For example, the required element title has the purpose "to name the unit of description" and the rule of completing the title begins "[p]rovide either a formal title or a concise supplied title in accordance with the rules of multilevel description and national conventions."

Currently, the ICA's Expert Group on Archival Description (EGAD) is working on extending these concepts into the semantic web through developing a "conceptual data model for archival description that identifies and defines the essential components of archival description and their interrelations in order to further shared international understanding, facilitate the development of the next generation of archival descriptive systems, further regional, national, and international collaboration, and promote collaboration with allied cultural heritage communities." [4] This work is delving deeper into resolving some differences among archival communities of practice to achieve a shared model.

In the United States, there are three core archival description standards: *Describing Archives: A Content Standard, 2nd Edition* (DACS) [5] , the *Encoded Archival Description* (EAD) [6] and *Encoded Archival Context-Corporate Bodies, Persons, and Families* (EAC-CPF) [7] These standards are the national application and extension of ISAD(G).

Describing Archives: A Content Standard, 2nd Edition (DACS)

The American archival community uses *Describing Archives: A Content Standard, 2nd Edition* (DACS) as the content standard and statement of

principles for arrangement and description of archives. DACS is used in conjunction with the encoding standards EAD and EAC-CPF.

DACS covers both description of archival materials and archival authority records that represent the people and organizations that created the materials. There are 25 archival elements covered by DACS, which are a refinement of the ISAD(G) 26 elements. Many of these elements are encoded into more than one EAD tag, as these tags are often broken into sub-elements that collectively represent a single DACS element. DACS can also be used with other encoded standards, including MARC.

These are the required elements for a multilevel description:

- Reference Code Element (2.1)
- Name and Location of Repository Element (2.2)
- Title Element (2.3)
- Date Element (2.4)
- Extent Element (2.5)
- Name of Creator(s) Element (2.6) (if known)
- Scope and Content Element (3.1)
  Note: In a minimum description, this element may simply provide a short abstract of the scope and content of the materials being described.
- Conditions Governing Access Element (4.1)
- Languages and Scripts of the Material Element (4.5)
- Identification of the whole-part relationship of the top level to at least the next subsequent level in the multilevel description. This may be done through internal tracking within a particular descriptive system; if so, the output must be able to explicitly identify this relationship.

Each subsequent level of a multilevel description should include all of the elements used at higher levels, unless the information is the same as that of a higher level or if it is desirable to provide more specific information [7, Chapter I] .

In addition to providing rules for the data to supply in an archival description, DACS also lays out the principles for arrangement and description of archival collections. One of the key principles is *respect de fonds*, or keeping collections separate from those originating from other sources and keeping them in original order whenever possible. The

principles also emphasize the importance of arranging or identifying logical groupings of materials and that the description reflects these groupings. This principle emphasizes the importance of hierarchical arrangement. Hierarchy is a key organizational technique, with context and description inherited from higher levels of description, allowing the whole and the parts to be described and understood in context. This approach can be contrasted with item-level description in which the description for one document is able to stand alone, as one commonly sees, for example, in library catalogs.

The final principle of archival description in DACS is that "the creators of archival materials, as well as the materials themselves, must be described" [8]. DACS directs archivists to identity all entities significant in the creation of the materials, to provide biographical information about the entities (particularly as it relates to the materials) and to use a standardized form to represent the names. DACS does not itself provide guidance for formulating names; rather, it recommends using other standards and tools such as the Library of Congress Name Authority File [http://id.loc.gov/authorities/names.html] and *Resources Description and Access* (RDA) instructions for constructing authorized access points.

DACS can be, and often is, used to create MARC-encoded descriptions for library catalogs. Library catalog records are typically only the front matter of the finding aid, with archival repositories depending on links to the full EAD finding aid to provide access to the components of a collection as well as longer descriptions.

Encoded Archival Description (EAD)

The *Encoded Archival Description* (EAD) is a metadata transmission standard for archival materials. Introduced in 1998, EAD recently went through a major revision; EAD3 was released in 2015. The Library of Congress is the official maintenance agency of EAD, and the SAA Standards Committee steered the recent revision by the SAA Technical Subcommittee for Encoded Archival Description (SAA TS-EAD). Although developed and maintained in the United States, the standard has been implemented in many other areas of the world.

EAD is expressed in XML with a published schema and DTD (data type definition) and is the set of elements that can be used for the description of archival materials and the relationship between the elements. EAD is designed for the intrinsic nature of archives, particularly in the American archival tradition. A finding aid, previously produced as an unstructured document, provides description of an archive as an aggregation of materials. This approach to description is intended to represent the context of creation and use of the materials, following the principle of *respect de fonds* – that records should be kept together in the original order as created or organized by the original creator whenever possible. EAD is an expression of ISAD(G)-defined elements for the components of a finding aid as structured metadata.

It can be helpful to think about physical boxes, folders and items in understanding EAD, but keep in mind that archives can include digital items, objects, artwork and more. Moreover, physical and intellectual order do not need to be the same. The components of an EAD encoded finding aid include the front matter with elements such as a scope and contents note which summarize the entire collection. At the lower levels of the hierarchy, more specific description can be provided at that level, without the need to repeat information from higher levels. For example, a series named "Plays" means that each box within that series – and each folder within the box – does not need to include the word plays for the user to understand that is what the item is.

Example:

```
c01 level="series">
<did>
<unittitle>...</unittitle>
</did>
<c02 level="file">
<did>
<container localtype="box"> 3 </container>
<container localtype="folder"> 18 </container>
<unittitle>Parent-Teacher Association of Fondsville</unittitle>
<unitdate unitdatetype="inclusive" normal="1959/1972">1959-1972</unitdate>
```

```
</did>
</c02>
<c02 level="file">
<did>
<container localtype="box"> 3 </container>
<container localtype="folder"> 19 </container>
<unittitle>Pasta and Politics Club</unittitle>
<unitdate unitdatetype="inclusive" normal="1967/1975">1967-
1975</unitdate>
</did>
</c02>
</c01>
```

Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF)

Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF) is the XML schema for expressing information about the people and organizations represented in archives. EAC-CPF was adopted by the SAA TS-EAD in 2011 and is jointly maintained by SAA and the Staatsbibliothek zu Berlin. Like EAD, EAC-CPF grew out of an IAC standard, the *International Standard Archival Authority Record for Corporate Bodes, Persons, and Families* [9]. The 2014 tag library included many elements, including many focused on relationships and contexts, as these are attributes valued by the archival community. For example:

```
<functionRelation functionRelationType="controls">
<relationEntry>Establishment and abolishment of schools</relationEntry>
<descriptiveNote>
<p>The second responsibility of the Department is to control the
establishment and abolishment of schools.</p>
</descriptiveNote>
</functionRelation>
```

EAC-CPF is in the early stages of adoption. The Social Networks and Archival Context (SNAC) [10] project has demonstrated the potential of the standard for interoperability and the usefulness of extracting and using information about organization and people from EAD-encoded finding aids. SNAC is hosted at the University of Virginia and has received support from U.S. National Endowment for the Humanities, the U.S. Institute for Museum and Library Services and the Andrew W. Mellon Foundation.

The SNAC prototype derived 6.6 million EAC-CPF records from existing data in EAD-encoded finding aids and authority records. These entities are then linked to related EAD finding aids in over 4000 repositories that contributed their holdings. A researcher can search for an entity and then identify archival collections that include documents created by or about that entity (Figure 1).

FIGURE 1. SNAC record for Vannever Bush, 1890-1974;
http://socialarchive.iath.virginia.edu/ark:/99166/w6cv4jx3

Conclusion

Descriptions of archives reflect the nature of these types of information resources; the descriptions tend to be complex, rich and lengthy. It is not uncommon for a finding aid to be thousands of words long. Professional theory and practice led to more consistent content in finding aids, but until the arrival of XML in the mid-1990s, there was limited ability to encode this rich data fully. International standards focused attention on the required elements of archival description, and EAD was developed as an encoding standard designed for archives. Specialized search interfaces that can take fuller advantage of the EAD encoding are certainly a benefit of standard-based archival description. However, projects like the Social Networks and Archival Context are demonstrating the incredible opportunities for re-using and building connections between the data held in many institutions. Just as archives are complex, so are the relationships among them and their many potential uses by researchers. Through early and collaborative adoption of XML and well-documented specialized implementations, the archival community has created the capacity for engagement in the emerging linked-data environments. ■

Resources Mentioned in the Article

[1] Pearce-Moses, R. (2005). Archives. In *A glossary of archival and records terminology*. Chicago: Society of American Archivist. Retrieved from www2.archivists.org/glossary/terms/a/archives

[2] What are archive [sic]? (2016). Retrieved from www.ica.org/en/what-archive

[3] *ISAD(G): General International Standard Archival Description.* (2000). (2nd ed.). Stockholm, Sweden: International Council on Archives. Retrieved from www.ica.org/en/isadg-general-international-standard-archival-description-second-edition

[4] ICA EGAD strategic work plan. (2012). Retrieved from www.ica.org/en/egad-strategic-work-plan-0

[5] Society of American Archivists. (2015). *Describing archives: A content standard*. (2nd ed.). Retrieved from www2.archivists.org/standards/DACS

[6] Encoded Archival Description official site. Retrieved from www.loc.gov/ead/

[7] EAD-CPF: http://eac.staatsbibliothek-berlin.de/about.html

[8] Statement of principles in DACS. Retrieved from www2.archivists.org/standards/DACS/statement_of_principles

[9] *ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families*. (2004). (2nd ed.). Stockholm, Sweden: International Council on Archives. Retrieved October 20, 2106, from www.icacds.org.uk/eng/ISAAR%28CPF%292ed.pdf

[10] SNAC: Social networks and archival content [website]: http://socialarchive.iath.virginia.edu/