

Describing Entities and Identities: The Development and Structure of Encoded Archival Context—Corporate Bodies, Persons, and Families

KATHERINE M. WISSER

*Graduate School of Library and Information Science, Simmons College, Boston,
Massachusetts, USA*

In January 2011, the Society of American Archivists fully endorsed the standard Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF). This article details the development of the standard, from its conceptual beginnings in 1998 to its dissemination and adoption in 2011. It provides an overview of the general structure of EAC-CPF and discusses variables that were considered important in the design of the standard. It concludes with a reflection on the strength of international participation in the development and review of the standard in order to ensure that EAC-CPF would be applicable across many different boundaries.

KEYWORDS archival description, contextual metadata, Encoded Archival Context—Corporate Bodies, Persons and Families, standards development

In 1998, Encoded Archival Description (EAD) introduced the archival community to the world of structured encoding of information that was traditionally created through narrative format. Because the finding aid was ill suited for the MARC-encoded format due to its multiple-level descriptive practice, the creators of EAD sought other potential encoding structures that could more easily mimic the descriptive structure of the traditional finding aid. More than 10 years later, it is hard to imagine the American archival scene without EAD. As archivists have adopted EAD, they have increasingly gained

Address correspondence to Katherine M. Wisser, Graduate School of Library and Information Science, Simmons College, 300 The Fenway, Boston, MA 02115-5898, USA. E-mail: wisser@simmons.edu

comfort with the use of eXtensible Markup Language (XML) and have begun to think about information in highly structured and manipulable ways.

Contextual information about record creators has been a traditional component of archival description. The three Dutchmen, Muller, Feith, and Fruin wrote the first archival descriptive standard in the *Manual for the Arrangement and Description of Archival Materials*. They describe the central role that creator description plays in understanding the material being described (2003). In outlining the principles of description, the first edition of *Describing Archives: A Content Standard* (DACCS) includes this emphasis: “An important aspect of understanding archival materials is the description of the context in which they were created” (2004, p. xv). To that end, DACS includes a chapter (10) on the description of creators, outlining the kinds of information that can be recorded for persons or families and for corporate bodies. This chapter, however, is written to facilitate a narrative description of creators only.

International standards have taken the description of creators further than DACS. The *International Standard for Archival Authority Records—Corporate Bodies, Persons, and Families* (ISAAR[CPF]) (2nd ed., 2003) outlined detailed instructions on the creation of separate authority records with contextual information about creators. ISAAR[CPF] does not exist in isolation of the descriptive standards for archival materials. The *ISAD(G): General International Standard for Archival Description* (2nd ed., 2000) and DACS both reference the possibility of creating separate records that describe creators rather than incorporating that description into the description of materials. DACS states, “Information about creators of archival materials can be captured and maintained in a separate system of archival authority records that are linked to the archival descriptions rather than being embedded within them. This approach reflects the model created by the International Council on Archives whereby the *General International Standard for Archival Description* (ISAD(G)) provides rules on description and the *International Standard Archival Authority Records for Corporate Bodies, Persons and Families* (ISAAR[CPF]) governs the creation of information about creators” (86). ISAD(G) makes numerous references to the ISAAR[CPF] standard in structuring 3.2, Context Area, and reference rule 1.14, which outlines the intention of the development of ISAAR[CPF]: “Because of the importance of access points for retrieval, a separate ICA standard . . . has been developed. ISAAR[CPF] gives general rules for the establishment of archival authority records that describe the corporate bodies, persons, and families that may be named as creators in descriptions of archival documents” (9).

Encoded Archival Context—Corporate Bodies, Persons and Families (EAC-CPF) was designed to address the possibility of separate creator description and to strengthen the duality that exists in archival description between context and content. Using the structure of ISAAR[CPF], EAC-CPF

provides a structured environment to describe creators and subjects of archival and other cultural heritage materials. EAC-CPF was initially developed as a beta standard (EAC Beta) and released in 2004. At that time, international implementers worked with the standard and provided significant input on the standard for the eventual development of EAC-CPF. In addition, the eXtensible Mark-up Language (XML) community continued developing the XML standard. Both the input from repositories using EAC Beta and the changes in XML are reflected in the current standard. Therefore, the current contextual metadata standard reflects a closer alliance with ISAAR(CPF) takes advantage of capabilities in XML and represents a philosophical neutrality in terms of the relationship between entities and identities necessitated by international understandings.

THE WORK OF THE EAC WORKING GROUP

EAC-CPF was long in development prior to the establishment of the EAC Working Group in 2006. When the Working Group met for three days in 2008 in Bologna, Italy, for its single full face-to-face design meeting, it had the benefit of 10 years of work and solid documentation by the Contextual Information Initiative. Through the effort of Richard Szary, Wendy Duff, and Daniel Pitti, two meetings took place in 1998 and 2001 to outline the goals of a standard for encoding and exchanging authoritative information about the context of archival materials. The first meeting, held at Yale University in 1998, resulted in action items such as the identification of existing archival authority initiatives, the development of an ad hoc international coordinating body to track work being done, and the development of a preliminary SGML/XML document-type definition for contextual information.¹ The second meeting, held at the University of Toronto in 2001, had an even more significant impact on the eventual development of EAC-CPF. At that meeting, a set of tenets was formulated that provided a backbone for the current standard. The Toronto Tenets: Principles and Criteria for a Model for Archival Context Information is a document that “defines principles and criteria for designing, developing, and maintaining a representational scheme and communication structure for archival context information.” The Tenets outline the justification for a model, including a refinement of contextual information to focus on the description of creating entities. By 2008, it was believed that that refinement needed to be more explicit and the name of the standard was changed to Encoded Archival Context—Corporate Bodies, Persons and Families rather than the generic Encoded Archival Context. This was done to ensure that other contextual information, such as the description of functions, events, and concepts could be accommodated in their own standards or in already existing structures.

The Toronto Tenets outlined “Definitions and Uses” such as what is archival context information and what is not context information, the role that contextual information plays in archival description, and its value as an independent resource. The “Structure and Content” section of the Tenets identified the potential for either an integration of information or an independent system that is dependent on linked data. It outlined a model whereby an instance describes a single entity, covering the full range of information that can be recorded with minimum sets of elements describing an entity, and a model that supports linking descriptions to digital or other surrogate representations. The Tenets also stipulated that the model would be expressed as an XML-compliant platform but left open the possibility of other approaches.

The Society of American Archivists presented the formal charge to the EAC Working Group in 2007. At that time, the working group consisted of 13 individuals from the United States, Canada, England, Scotland, France, Germany, Italy, Sweden, and Australia. The working group received grant support from the Gladys Kriebel Delmas Foundation for a face-to-face meeting, and with additional support from OCLC Research and the National Library of Australia and significant support from IBC (Istituto per I beni artistici culturali e naturali) of the Regione Emilia-Romagna, and the generous hosting by the Archivio di Stato di Bologna, the working group met for three days in May 2008. Prior to the face-to-face meeting, a call for comments and feedback on EAC Beta was distributed through various email lists and separate projects submitted feedback. That feedback was analyzed and digested for the working group to review prior to meeting. Following the three-day meeting, the working group was invited to participate in a day and a half conference in Bologna, *Standard e formati di scambio per l'interoperabilità dei sistemi archivistici* (Standards and exchange formats for interoperability among archival information systems), sponsored by the Istituto per I beni artistic culturali e naturali, Bologna, Italy.

With the work that was already accomplished and well articulated in the Toronto Tenets, the EAC Working Group had a head start in analyzing the feedback regarding EAC Beta, which was released in 2004. As part of the initial discussion, five principles of design were agreed upon:

- 1) Keep it simple.
- 2) The schema should have a direct relationship to ISAAR(CPF).
- 3) Parsimony is a design principle that will make the EAC schema interoperable and data-base friendly but also extensible.
- 4) Avoid doing things in the schema just for presentation.
- 5) If you can't explain it, don't do it.

Another aspect that the working group established early was the intended relationship with other standards, in particular EAD. Acknowledging that the

concept of “context” was broader than just creator description, it was early decided that, based on our charge, we should explicitly narrow the coverage of the standard to intentionally limit the outcome of the standard. Therefore, the title of the standard needed to include “Corporate Bodies, Persons, and Families” in order to explicitly identify the scope of the standard. This was particularly important given the recent completion of the development of the *International Standard for the Description of Functions* (ISDF), which constitutes another area of contextual information that can be described separately. It was envisioned that other contextual metadata standards could be developed under the umbrella of EAC.

GENERAL STRUCTURE OF EAC-CPF

The EAC-CPF model is relatively straightforward and should be familiar to those who have encountered other metadata standards in XML. This section outlines the significant aspects of the model, but it does not account for every single element within the standard.

The basic model consists of two sections: <control> and <cpfDescription>. The <control> element contains information about the electronic document and is parallel to the “header” elements in other standards. Within <control>, though, the information is much more directly parsed and includes six required and five optional elements. Each element represents a separate piece of information related to the ability to manage the electronic record. The required elements include <recordId>, <maintenanceAgency>, <maintenanceHistory>, <maintenanceStatus>, <languageDeclaration>, and <sources>. Optional elements include <publicationStatus>, <conventionDeclaration>, <otherRecordId>, <localControl>, and <localTypeDeclaration>. It should be noted that these last two optional elements provide the extensibility within <control> for individual implementers to manage their records. This control structure has some significant differences from standards such as EAD, which focuses on a smaller number of elements that contain a variety of information. For instance, there are three elements in EAC-CPF for actions done to the record, whereas EAD bundles that information in <profiledesc> with some other actions recorded in <revisiondesc>.

The element <cpfDescription> contains the contextual information about the identity being described. It includes three elements: <identity>, <description>, and <relations>. The <identity> element is the only element required in <cpfDescription>, and it represents the name control section of the record. Within <identity>, there is a required element to establish the type of identity (corporate body, person, or family with the element <entityType>) the <cpfDescription> represents and one or more names by which that identity is known (with the element <nameEntry>). These

names are separated into one or more <part> elements and have an element <useDates> available for any date information associated with the name.

The elements <authorizedForm> and <alternativeForm> allow the record creator to identify, according to a set of rules, those names that are authorized or variant. The advantage to this model is that multiple, controlled vocabularies with various constructions of names can be included in the same record. For example, Pablo Picasso's Library of Congress Authorized heading is "Picasso, Pablo, 1881-1973." From the Union List of Artist Names at the Getty Institute, the heading is "Picasso, Pablo." The use of the <authorizedForm> element allows the record creator to identify the vocabulary that the heading has been derived from and include multiple different authorized headings in the same record.

The identity section also includes a <nameEntryParallel>, which allows for multiple linguistic versions of a name to be bundled together in those situations in which multilingual representations are a component of the name (e.g., the Library and Archives Canada has a parallel French name Bibliothèque et Archives Canada; both are considered to be official names of this corporate body). The <nameEntryParallel> element consists of two or more <nameEntry> elements.

The designator <description> consists of 11 elements. The ISAAR(CPF) standard outlines eight areas for description, and with the removal of the extensible element <localDescription> and the separation of functions and occupations into separate elements (<function> and <occupation>), <languageUsed> is the only element not represented in ISAAR(CPF). Several of the elements within <description> have pluralized elements in order to facilitate the bundling of multiple instances (i.e., <functions>, <languagesUsed>, <legalStatuses>, <localDescriptions>, <mandates>, <occupations>, and <places>). The EAC Beta model envisioned <description> as containing both formal and informal descriptive elements. While that language is no longer used, the distinction between elements has not been removed. The six elements referenced above are intended to be populated with terms that are derived from controlled vocabularies. Other less formal descriptive elements, including <biogHist>, <generalContext>, and <structureOrGenealogy> are structured for narrative descriptions represented in a choice of format-specific elements such as <p>, <list>, and <outline>.

The last section of the <cpfDescription> is <relations>. The ability to establish relationships within EAC-CPF is perhaps one of the most exciting aspects to the standard. There are three different types of relations defined: <cpfRelation>, <functionRelation>, and <resourceRelation>. Another significant departure from the EAC Beta model occurs in the relations section of the schema. In EAC Beta, the schema included elements to describe resources and functions. In EAC-CPF, any attempt to describe the thing that is being related to the identity has been eschewed. Instead, there are three different

ways that the identity, resource, or function can be portrayed. First, there is the option to include `<relationEntry>`, which contains a textual description but is generic. Second, there is an `<objectBinWrap>` element that allows for the inclusion of base64-bit encoded binary representations. Third, and perhaps the most exciting for the archival field, there is an `<objectXMLWrap>` element that allows for the declaration and inclusion of other namespaces, such as EAD, MODS, or METS. Whereas in EAC Beta, the schema included a host of EAD elements, with this move toward using XML capabilities to its fullest, it is not necessary to repeat the elements within the EAC-CPF schema. It streamlines the schema and also opens up the possibility of using a wide range of XML standards in EAC-CPF records. Each relation established must include an attribute that identifies the relationship type. The three relationships allowed contain unique relationship type values depending on whether it is another identity, function, or resource.

DEFINING ENTITIES AND IDENTITIES

As has been noted, the standard to be developed needed to focus on the description of corporate bodies, persons and families, creators and/or subjects of archival, and other cultural heritage materials. The working group grappled with the challenges that that description would present. It was necessary to address the issues of entities and identities in order to design a standard that would adequately represent the complexities that are present in human characters. For instance, the standard needed to accommodate the straightforward and the complex, the official and physical embodiment, and the imaginary and the shared. The first step was to provide definitions and to associate terminology with those definitions. Therefore, an entity is defined as the physical embodiment of the unit being described (the real person), whereas identity is defined as a single representation of that entity. Entities can have multiple real or imagined identities and identities can be shared by multiple entities (e.g., the shared pseudonym of Franklin W. Dixon or the Office of the Presidency). While some of the more complex models considered were not thought to be common or to be necessary to handle in a majority of the records that will be created, it was considered important to have a standard that could accommodate those complexities when they arose.

At the same time, the issues of entities and identities can be seen through a variety of philosophical lenses. The working group acknowledged that not all archival traditions are the same and that the issue between whether or not each identity should be described separately and then linked or that multiple identities should be described in a single record was not something that needs to be dictated by the standard. The design, though, needed to be able to toggle between the representations of multiple identities in order to facilitate aggregation. The design was specifically intended to facilitate

aggregation and every attempt was made to ensure that aggregation, even with local extensibility, could take place.

Built into the model, therefore, is the ability to represent a single identity with a single EAC-CPF record and, using a relationship structure, link records that represent multiple identities for the same entity (e.g., Bill Clinton the person, Bill Clinton as president, Bill Clinton as governor). Another choice is to use a specific element, `<multipleIdentities>`, to incorporate the description of more than one identity for an entity in a single record. The element `<multipleIdentities>` may contain two or more `<cpfDescription>` elements, which allow for each identity to be represented by a single `<cpfDescription>`. The primary difference between these two models is the kind of control over the record, but in actuality they are interchangeable. Going from the single identity-record model to the single-record-multiple-identities model would require the generation of a new `<control>` section that reflects the conversion. Original information included in the `<control>` section from the individual records could be maintained in the `<maintenanceHistory>` `<descriptiveNote>` so that no information would be lost. Similarly, to convert the other way, the `<control>` would be identical in each new EAC-CPF record with a note in `<maintenanceHistory>` identifying the conversion. This stance represents a philosophical neutrality. The standard does not require any one philosophical position about multiple identities to sacrifice that position. At the same time, the design also facilitates aggregation.

A final note about the design and aggregating records: the working group acknowledged that multiple records, particularly in different languages, could be created for the same identity. In order to allow for multiple linguistic representations, a bundling element, `<alternativeSet>`, was included that would allow records constructed for the same identity to be included together without preferencing any single linguistic version. For example, if the Bundesarchiv in Berlin creates a record for Otto Von Bismark that is expressed in German and the Bibliothèque Nationale de France also creates a record for Von Bismark that is in French, these two records can be bundled together by an aggregator using the `<alternativeSet>` element. The use of `<alternativeSet>` requires no changes to the original individual EAC-CPF records, but a `<control>` section is created for the aggregation of these records.

RELATIONSHIP WITH ISAAR(CPF)

Adherence to ISAAR(CPF) was an important early decision of the working group. With minimal areas of deviation, the EAC-CPF structure mirrors the structure of ISAAR(CPF), while at the same time including an extensible design that allows for implementation modification that would not decrease

shareability of records among aggregators. As outlined in ISAAR(CPF), the second edition of that content standard provides a closer alliance to the types of information that would be important in any kind of shareable metadata. Other changes in the second edition of the standard confirm the implementation of the standard across international boundaries: “Elements and rules within the four new areas have been structured and drafted with the aim of supporting clear understandings of concepts and better practices in the documentation of the context of archives. In addition, this edition contains a section describing how archival authority records can be linked to archival materials and other resources, including ISAD(G)-compliant archival descriptions” (5). This model provided a framework from which the control, name formation, description, and relationship sections of EAC-CPF could be shaped. Where EAC-CPF deviates from ISAAR(CPF) is within the description elements (i.e., the separation of functions and occupations and the addition of `<languageUsed>`). These departures were a result of input from the user community once a penultimate draft was released in August 2009 and do not constitute enough of a difference to make it impossible to create an ISAAR(CPF)-compliant EAC-CPF record.

XML ADVANCES AND RELATIONAL DATABASES

From the initial release of EAD in 1998 and the development of the EAC-CPF schema in 2009, XML experienced a parallel development and increasing sophistication. The EAC-CPF schema reflects that increasing sophistication. Using the addition of XML namespace inclusion and extensive use of xlink linking-language instead of generating link attributes are just some of the ways in which EAC-CPF reflects the current state of XML technologies.

Additionally, the design of the schema was intended to facilitate the storage of information in a relational database. In order to accomplish this, the design vacated mixed-content models extant in other metadata standards, and wherever possible eliminated the use of attributes in favor of elements. This would not have been possible without the use of inline style capabilities, so a `` element was included. However, overall, the standard consists of elements that can either contain parsible character data or other elements, not both.

CONCLUSION: THE INTERNATIONAL EMPHASIS OF THE STANDARD AND ITS STRENGTH GOING FORWARD

In constructing the working group for the development of EAC-CPF, the Society of American Archivists concurred with working group members that an

international emphasis was essential in its development. It was clear that international participation in the development of the standard was an important development as had been observed by increased international participation in the ongoing maintenance and development of EAD and by the active testing and use of EAC Beta by international audiences. In fact, EAC Beta enjoyed a much greater exploration by non-American archival institutions, such as the LEAF project, People Australia, and other initiatives. International engagement has been a facet of the development of EAC-CPF since the meeting in Toronto and is reflected in the membership of the working group that developed the standard: 14 members representing 9 countries. From the initial survey for feedback on EAC Beta representing a broad spectrum of the international arena, through detailed input from institutions spanning the globe during its final review period, attendance to Webinars hosted through OCLC Research, through the descriptive components that allow for multiculturalism (e.g., <languageUsed>) and through continuing engagement in making this standard accessible through multiple translations of the Tag Library, EAC-CPF is modeling international cooperation and engagement in standards formation.

ENDNOTE

1. Documents from the Contextual Information Initiative can be retrieved from <http://www.library.yale.edu/eac/goalsandworkplandraft.htm> [Viewed September 25, 2011].

REFERENCES

- Describing archives: A content standard*. Chicago, IL: Society of American Archivists, 2004.
- ISAAR (CPF) International Standard Archival Authority Record for Corporate Bodies, Persons and Families*. 2nd Edition. International Council on Archives, Adopted by the Committee on Descriptive standards, Canberra, Australia, 27-30 October, 2003. [http://www.icacds.org.uk/eng/ISAAR\(CPF\)2ed.pdf](http://www.icacds.org.uk/eng/ISAAR(CPF)2ed.pdf)
- ISAD(G): General International Standard Archival Description*. 2nd Edition. International Council on Archives, Adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999.
- Muller, S., J.A. Feith, J. A., & Fruin, R. *Manual for the Arrangement and Description of Archives*. Chicago: Society of American Archivists, 2003. (First English edition, 1940).
- “Toronto Tenets: Principles and Criteria for a Model for Archival Context” Report from Toronto Archival Context Meeting, March 2001, Appendix B Statement of principles: the Toronto Tenets, <http://www.library.yale.edu/eac/torontotenets.htm>

Copyright of Journal of Library Metadata is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Copyright of Journal of Library Metadata is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.