

SNAC Cooperative

Rationale for an Archival Description Cooperative

While the library community has a long history of cooperatively describing library holdings and sharing the descriptive data, the archival community does not. This is arguably based on a fundamental difference in the nature of the holdings of the two professional communities: many and perhaps most of library holdings are not unique, while most of the archival holdings are. While the archival holdings are largely unique, they are nevertheless inextricably interconnected in ways that provide opportunities for cooperating that will significantly benefit both the archival community and the researchers, scholars, and members of the public that use their holdings.

A. Library Cooperative Description

The nature of library holdings and the economics of describing and providing access to the holdings provide compelling incentives for libraries to cooperate with one another in description. The majority of library holdings are copies of published works, and many libraries collect copies of the same published works. This basic fact has led libraries to seek economies in sharing the labor and expense of cataloging the works commonly held. The fundamental characteristic of cooperative cataloging is that cataloging experts in many libraries cooperatively create and maintain pools of data that are shared with one another and others.

Library cooperative cataloging emerged in the 1970s with the emergence of OCLC.¹ While OCLC has over the years developed a wide variety of services, its core service has been to aggregate cataloging data from around the world that participating libraries can share. Rather than copies of the same works being cataloged over and over again in holding libraries, one library catalogues its copy and shares the data with other libraries.

Cataloging involves a variety of specialized activities, one of which is the creation of entries or access points in the cataloging of each work. Prominently among the access points are authors and subjects of works. Such access points are not unique in the cataloging of each work, as authors may write more than one work, more than one author can co-author a work, many works are about the same subject, and one work may be about many subjects. All of this to say that the authors and subjects employed in the cataloging of library resources represent an excellent candidate for cooperatively creating and maintaining, as they are reused over and over.

Because the names of authors and subjects are reused, libraries have traditionally maintained a separate list or file for creating and managing them. The separate list has traditionally been called an authority file. The authority file's central function is to ensure that the same *authorized* name or term is used for each distinct author or subject. Any given

¹ Beacher Wiggins. "The Program for Cooperative Cataloging," in the *Proceedings of the Taxonomic Authority Files Workshop*, Washington, DC, June 22-23, 1998 (<http://researcharchive.calacademy.org/research/informatics/taf/proceedings/wiggins.html>).

author may have many different names, and some authors may have the same name. Similarly with subjects, many different terms may represent the same or similar topics. Thus authority control, as the activity is called, involves mapping from the names and terms that are not used (that are not authorized) to the names and terms that are used (that are authorized).

Until relatively recently, each library maintained its own authority files in relation to its catalog of holdings. The emergence of computing and network technologies, though, provided a means to begin sharing authority work and data. The fact that authority work is one of the (if not most) time-consuming cataloging activities was ample incentive. In 1977, the Library of Congress began cooperating with the U.S. Government Printing Office in the creation and maintenance of the Library of Congress Name Authority File (LCNAF).² This joint venture was named the Name Authority Cooperative (or NACO). In the following decades, the program expanded to include a large number of other participants, eventually became international.

In 1998, the Library of Congress, the Deutsche Nationalbibliothek and OCLC began exploring the possibility of aligning national authority files with one another. The three entities formed the Virtual International Authority File Consortium (VIAF) in 2003, and in 2007 the Bibliothèque nationale de France joined. Since 2007, additional national and program contributors have continued to join. In 2012, VIAF became an OCLC service. Today there are thirty-four national libraries and other organizations contributing data (more if one includes the national libraries contributing through NACO). While VIAF does not support the maintenance of the contributed authority data – each of the contributors maintains its own authority file – it does provide an unprecedented reference resource that invaluablely contributes to the efficiency of authority work.

The overlap in library holdings – copies of the same publications – and the overlapping components in the description of the holdings – the controlled access points – provides the library community with common work that can be effectively cooperatively performed and shared. Computer and network technologies present the means to cooperate. The library community is chronically behind in the description of holdings, because the amount of work needed exceeds the resources needed to complete it, and thus economic pressure provides the motivation. The archival community has the means to cooperate, the same means as the library community, and it has the same economic incentive – more work than resources to complete it – but is there sufficient, sharable work to be cooperatively performed?

B. The Social Context of Records

The holdings of each archive are largely unique. Sometimes copies of records are made, and copies of the same records may be found in different archives, but records are not systematically massed produced, with copies being held in common by many archives. The responsibility for describing the records themselves in any given archive is therefore not sharable work. Copy cataloging has no significant role to play in archives. But if descriptions of the records themselves cannot be usefully shared, are there components of the description that overlap the holdings of archives that would provide an economic motivation for archives to share the work of creating and maintaining them?

² Ibid.

The records in archives document human activity in all of its broad and diverse forms. National, state, and local governments, businesses large and small, hospitals and doctors' offices, non-profit organizations, educational institutions at all levels and of all types, religious organizations, artists, farmers, poets, writers, and so on. Records document organizing and organized activity³. All of the activities exist within a social context. The people documented in all of the records are interrelated with one another. Thus the records reflect, among other things, the context in which they were created: the work and activities being performed, why it was being performed, the lives being lived, the interactions of the people working and living, and the influence of one person on another, sometimes from the recent or even distant past.

While the records themselves are largely unique, the documentary content of the records necessarily overlaps because the lives of the people documented in the records overlap, in the present as well as into the past. Records are the product of and document a vast global social-historical network of humans living and working together in organized, purposeful ways.

Major components of archival description involve describing the *context* in which records are created, accumulated and used: the creators of records, the activities in which they engaged, and the objectives pursued. Once the records are no longer in the originating environment, once they have become archival or historical records, those interested in using the records as historical evidence need to understand the originating context in order to understand and interpret the records.

C. Centrality of Context in Archival Description

Traditional archival descriptive practice combines the description of records, and of the people, functions, and activities documented in the records in a single apparatus, the finding aid. Each finding aid contains a description of all records held in a particular repository that are created by the same corporate body, person, or family (a collection or *fonds*) and descriptions of the people, functions, and activities documented in the records. The finding aid enables users to discover, locate, identify, and understand historical records.

Describing all of the records with the same origination is based on two fundamental and closely related principles of archival description that emerged in the nineteenth century: the Principle of Provenance, and the *Respect de fonds*. The principles state that records created or accumulated by the same corporate body, person, or family are to be kept together and not intermingled with records with a different origin. Practice of the principles has among its objectives preserving the original context (administrative, legal, and social) and the interrelation of the documents within that context.

Application of the principles traditionally has been expressed in a hierarchical description. Hierarchical description provides the most economic way to apply the principles. Thus the finding aid begins with a description of the entire collection and then provides an analysis of component sets of records (generally grouped by the activity or function they document,

³ The earliest forms of records can be dated to around 8000 BCE, the use of tokens to keep track of agricultural products. By the fourth millennium BCE, assemblages of tokens were being bound together using strings connected by bullae bearing seals, and in hollow clay envelopes with markings on the outside indicating the number and types of tokens contained within, and with seals. It is noteworthy that these early forms of record keeping predate written language. See Denise Schmandt-Besserat, *How Writing Came About* (Austin: University of Texas Press, 2006).

such as correspondence or minutes of meetings; thus maintaining the interrelation of the records). Components of the components may then be similarly analyzed, with the description sometimes, but not always, terminating with the description of individual files or even items within files. The depth and detail of analysis and description is determined by both intellectual and economic criteria.

The record description components of finding aids is quite similar to the description of library and museum objects: name or title of the collection or component of the collection; name of the creator; date of creation; statement of extent (ideally how many records and how much information); language or languages of the information; and associated places and subjects. If not for the collective, hierarchical approach to the description, the archival description would not be unfamiliar to library and museum professionals. The archivists, however, describe the *context*, and frequently in great detail.

Description of the context involves a number of different descriptive components. First and foremost, it involves description of the creator. The name of the creator is provided, of course, but also biographical or historical information about the creator's existence. For persons, life dates will be given, if known. In narrative or timeline form, his or her biography will be provided. While frequently the biographies may be brief, even sparse, they may also be quite long and include the occupations, places of birth and death, employers and persons known of or known by the person are provided, major events in the person's life (education, marriage, children), and other significant events and achievements in the person's life. For corporate bodies, the description may include names, the dates of establishment and disestablishment, organizational relations to parent bodies, to preceding and following bodies, a brief or detailed history of its existence, the functions for which it is responsible, and prominent individuals that affiliated with it or with whom it interacted. Families follow a pattern similar to persons and corporate bodies, though their descriptions may involve the naming and biographical description of prominent members of the family. The biographical and historical information, the social and organizational context, the description of activities performed and the ends sought are all archival descriptive context.

The contexts of records overlap with one another because the lives of the people documented in them overlap. The corporate body, person, or family that created one set of records is found in the records created by other corporate bodies, persons, or families, in the same archival institution, but also in records held in other archival institutions. The functions, activities, and occupations documented in the records are also not unique, as many of the same functions, occupations, and activities have been and are performed by many different people and organizations, in a wide variety of times and places.

Description of context provides archivists the common work that can be effectively cooperatively performed and shared, with the means to achieve new economies in the processing and description of records, and many more benefits, for archives and the users of archives.

D. Libraries: Linked Authority Control

In the 1980s, various libraries began what came to be called "linked authority control." In the world of manual cataloging, libraries maintained both a card catalog and a separate file or files for managing names and subject terms used for author and subject access to the library holdings. As described above, the access points and the descriptions of the holdings have a many-to-many relation with one another. The same person or corporate body may

create more than one work, and a work may have more than one creator. The same holds true with subjects: a single subject may be addressed in multiple publication, and one publication may address multiple subjects. With card catalogs, the authors and subjects were typed on each catalog card. For multiple entries in the catalog, there would be multiple cards (or sets of cards for more detailed descriptions), since there would be one card (or set) for each entry point. Each author and each subject associated with a work would have a corresponding almost identical card or cards for each work. The only difference between the cards would be that each author or subject would appear at the top of the card, reflecting the point at which it was to be alphabetically filed in the catalog. And if a name or subject term was revised for any reason, the library would retrieve the cards, retype them with the revision, and then re-file them in the catalog. All in all, an exceptionally inefficient, labor-intensive technique!

With the advent of machine-readable cataloging (MARC) and public "online" catalogs, the need for the multiple cards disappeared. But the names and subject terms were still redundant, since all MARC records with the same authors or subjects retained the same names or terms. To address this remaining redundancy, the community developed linked authority systems.

The emergence of relational database technology inspired linked authority systems. Relational database technology optimizes the maintenance of many-to-many relations of which the names and terms and the descriptions of publications are a clear example. Using relational technology, the name of an author or a subject term could be recorded once, and then linked to description of the publication. If a name or term needed to be revised, then it was revised once and only once, and the changes instantly reflected in the catalog. This innovation was quite dramatically more efficient than the manual and even the early digital cataloging environments, as it eliminated the last remaining redundancy. Perform the work once in one place, and use it over and over in many places.

Archival description, as currently practiced, involves the same many-to-many redundancies as the manual and early library catalogs: the same corporate bodies, persons, and families, functions and occupations, and when subject analysis is also provided, subject terms and geographic names, appear in many different finding aids. Clearly linked systems of description would economically benefit archives. The principle technology that enabled linked authority systems in libraries, relational database technology, is available to archives. However, the archives must have underlying communication standards that enable consistently representing and sharing the data in machine-readable form. For the library community, the MARC bibliographic and authority formats have served this role. What are the comparable archival communication standards?

E. Archival Description Standards

With the release of *Encoded Archival Description* (EAD)⁴ in 1998, many archival repositories around the world began to convert their finding aids into a machine-readable form. Based on the International Council on Archive's (ICA) *General International Standard Archival Description* (ISAD(G))⁵, EAD has become a standard for computer representation and communication of archival finding aids.

⁴ <http://www.loc.gov/ead/>

⁵ ISAD(G) Second edition: <http://www.ica.org/download.php?id=1687>

Both ISAD(G) and EAD reflect the traditional archival practice of combining the description of records and the description of context in a single apparatus, the finding aid. In 1996, two years before the release of EAD, ICA released the first version of *International Standard Archival Authority Records—Corporate Bodies, Persons, and Families* (ISAAR (CPF)).⁶ One of its principal objectives was enabling the separation of record and creator descriptions, which is to say, the development of linked systems. In such a linked archival description system, the record description in finding aids and the description of the corporate bodies, persons, or families associated with the records would be *maintained independently* and *interrelated* as appropriate. For research users, the interrelated components of the description would be brought together to form a complete description, in exactly the same manner that data is brought together at the time of use in library catalogs.

In order to separate and maintain independently the corporate body, person, and family contextual data from the description of the records, the archival community needed a communication standard that would enable it to use computer and network technologies to build linked archival description systems. With the release of Encoded Archival Context – Corporate Bodies, Persons, and Families (EAC-CPF) in 2010, based on the ICA standard ISAAR(CPF), the archival community had the final necessary component needed to separate description of corporate bodies, persons, and families from record description.⁷ With relational technologies and the allied standard EAD, EAC-CPF provides the last essential component for developing linked archival description systems that support sharing work and data. The archival community is thus positioned to begin realizing the economies of cooperative description.

F. Archive and Library Authority Control

The existence of national and international library cooperative authority control programs (such as LCNAF and comparable national programs, and VIAF) make it reasonable to explore whether it would be feasible for the archival community to collaborate with the library community in the use of these existing programs rather than establish its own program. Many of the corporate bodies, persons, and families, after all, are common to the holdings of both communities, and with those of museums as well. Many of the bodies and persons that generate records also write books or create art or collect objects found in museums (painters, sculptors, archaeologists, botanists, and geologists for example). The person who authors a book also generates records that document the writing of the book (manuscripts, correspondence with publishers and colleagues, and so on), among other records that document other life and work activities. She may also be an artist or perhaps an archeologist. People and organizations associated with allied cultural heritage institutions overlap.

Archival records, however, document the lives and work of people and organizations in *all walks of life* and not just the fraction of people who write books or have books written about them, or create or collect objects found in museums. Farmers, business people, health professionals, teachers, social workers, community organizers, civil rights attorneys and works, religious leaders, and many, many others are represented in archival holdings. Birth certificates, marriage certificates, military records, property records, and other types of

⁶ ISAAR second edition: <http://www.ica.org/download.php?id=1648>

⁷ EAC-CPF Official Web site: <http://eac.staatsbibliothek-berlin.de/>

records ensure that just about everyone is represented. Of course any of these people can write a book or have a book written about them. The vast majority of the people documented in records, however, are simply not present or associated with the holdings of libraries and museums.⁸ Archivists certainly do not and will not formally describe the majority of the people documented in their holdings, but many of them will be described, and many of these descriptions will be unique to archives, the *sole responsibility* of archivists.

The centrality of providing context leads archival archivists to describe individuals, families, and organizations in a manner different than in libraries. Common to both is authority control proper: recording the names used by and for the entity, and designating (based on rules) which among them is to be preferred and which not for facilitating access to resources, displayed in descriptions of the resources, and in retrieval lists. As important is name control is, the archival imperative to provide context involves documenting the lives of individuals and history of families and organizations, quite frequently in great detail.

As libraries are increasingly working in a global, shared environment, the nature of library authority control has begun to extend beyond name control towards *identity control*. VIAF aggregates authority records from different authority files from around the world. Authoritative names for the same person frequently differ from one authority file to another, depending on, among other factors, the evidence available when the record is created and the cataloging rules used in creating it. VIAF groups the records for one person or organization together. The objective is that each group contains all records for the same identity, and that records for different identities are not grouped together. While the name strings remain of paramount importance, contextual information becomes important in distinguishing when persons or organizations with the same name are in fact different entities, and when those with similar though different names are the same entity. Biographical and historical information becomes important in determining identity, as names alone are weak identifiers. This shift towards identity control is reflected in changes in the MARC authorities format.

The MARC authorities format, maintained by the Library of Congress Network Development and MARC Standards Office, has added several fields that may be broadly associated with establishing the identity of the described entity.⁹ While it has been the practice to add birth and death dates to personal names to distinguish one person from another, the format now has a special field for precisely encoding life dates and dates of existence for corporate bodies and families. Among other fields associated with identity are associated places; addresses; fields of activity; associated groups; occupations; gender; associated languages; and biographical or historical description. The data represented by these data assist in reliably and confidently determining identity. With the addition of these MARC data fields, library authority control more closely resembles archival context description.

Despite these recent changes that bring library authority description closer to archival description practices, the changes do not fully capture the range or the detail required in archival context description as codified in ISAAR (CPF). In addition to the MARC authorities fields, ISAAR(CPF) calls for legal status; mandates or sources of authority; and functions and activities (which is not the same as field of activity). The archival communication standard

⁸ On December 17, 2014, 26% of the nearly 3M EAC-CPF in SNAC matched VIAF cluster records.

⁹ *MARC 21 Format for Authority data* Library of Congress Network Development and MARC Standards Office, <http://www.loc.gov/marc/authority/>

accommodates prose description of the descriptive data *and* use of formal links among the CPF entities, the descriptions of archival records, and the description of functions and activities.

The archivists' objective is to not only identify the people documented in archival records, but to describe them in sufficient detail that users will be able to understand the social context in which the related historical records were created and used. Thus the description commonly is far more elaborate than is accommodated by the brevity assumed in the MARC authorities format. Biographies and histories quite frequently comprise several paragraphs or chronologies with multiple date-place-event entries. Such complex, detailed description in library authority control is not necessary, but in archival description it is considered essential for establishing context. Though it is conceivable that library authority control could be enhanced to accommodate the archival requirements, the library cooperatives exist first and foremost to serve the needs of the library community—as they should. There is no compelling motivation for the library community to enhance these systems to accommodate the rather complex descriptive needs of the archivists.

Archivists need a cooperative program that ensures that its primary professional curatorial responsibilities are fulfilled. Such a program needs to be fully integrated into the management and processing of archival resources. While distinct from existing library cooperatives, the archival cooperative will necessarily collaborate with the library and museum communities, to align identities across the curatorial domains, to share data with one another, to provide common pathways to resources across the cultural heritage spectrum.

G. Cooperative Linked Archival Description

A primary objective of the SNAC Research and Demonstration project was to demonstrate that the names of corporate bodies, and persons, and information about them embedded in archival record descriptions could be extracted into archival authority records linked to the archival record descriptions from which the information was extracted. The intention was to take the first step towards realizing in archival systems the economic benefits of linked authority systems found in libraries, though with a significant innovation.

The SNAC Cooperative envisions a more efficient cooperative model than what is currently employed in the library community. Library linked authority systems are local, that is, authority records, even if derived from a cooperatively maintained source such as the NACO Name Authority File, are maintained locally. While there are definite catalog maintenance efficiencies to be realized in the linked local system, the disconnect between the locally maintained authority records and the cooperatively maintained authority records introduces inefficiency. When authority records are revised in the cooperatively maintained authority file, the local authority records copies of those records must be re-derived, or the revision duplicated. This must be repeated in each local authority file derived from the cooperative authority file. SNAC envisions the cooperatively maintained authority file being the local authority file, which is to say, the cooperatively maintained authority records would be linked to the local archival record descriptions, and no local authority file would be maintained. The cooperative authority file would function as the authority file, eliminating the last major inefficiency found in current library systems.

Such an approach is not without significant technological challenges. But recent developments that were motivated in large part by and designed to support the Semantic Web and Linked Open Data provide workable methods for addressing the challenges. In

particular, the growing use of *persistent identifiers* for Internet accessible resources, including finding aids and catalog records. In essence, the archival cooperative being planned, envisions a global archival description system that is, in essence, a managed, controlled subset of the Internet. Further, the name access points in EAD-encoded finding aids and MARC21 records can be revised to reference directly authority records in the shared cooperative authority file.¹⁰ By linking the local descriptions to the shared authority file, revisions in the latter can be directly promulgated in the former.¹¹

Realizing the envisioned global linked archival description system will be a socially and technologically complex undertaking that will take time to unfold. Working cooperatively is compelling because of the economies that can be realized through cooperatively sharing the work describing the social and historical context of archival records. But the benefits of a global archival description system extend far beyond the economic, as an international archival description cooperative will also transform the nature of historical research in fundamental and powerful ways.

H. More than Economic Benefits

The shared context of the archival records and the imperative to describe that context provides the archival community with an economic rationale for working cooperatively and sharing data. Established network and computer technologies provide the technological means to work cooperatively in an efficient, effective manner. The imperative to preserve context through description, based on longstanding professional principles, distinguishes archival descriptive practices from that of allied cultural heritage communities, and these distinctive practices need a dedicated community-based standards and technological infrastructure.

The benefits of a dedicated archival cooperative are more than economic, and extend beyond the archival community to the users of archives: historians of all kinds, educators, students, genealogists, documentary editors, and many others.

Among the benefits for the users of archives are the following:

- Integrated access to millions of historical resources located in hundreds of archives, libraries, and museums around the world. Such access is of enormous economic benefit to researchers.
- Access to the professional, social, and intellectual networks within which the people documented in historical resources lived and worked, revealing relations and connections among people and documents.
- Access to emerging national and international integrated cultural heritage such as DPLA and Europeana.
- Access to the information needed to plan research travel.

¹⁰ MARC subfield for "Authority record control number or standard number" and EAD attribute "source".

¹¹ There is one major exception: when two or more identities are mistakenly combined in one authority description, the description must be split into two or more descriptions. Such revisions always require manual review to determine which of the two or more identities is to be associated with the local description.

- A biographical and historical reference resource for use by teachers and students in the teaching of history and the teaching of research methods, and pathway to digitized resources that can be used in the classroom.
- A means for those interested in family history to connect family members to the professional and social networks within which they lived and worked.

In addition to the processing efficiencies described above, the following are benefits for archival institutions and archivists:

- Benefiting the users of archives in all of the ways described above.
- The effectiveness and impact of archival description is enhanced through use of the data in new and innovative ways.
- Description is exposed and made reusable both within and outside the archival community.
- Raises the profile of archival holdings.
- Increases the discovery of and use of archival holdings.
- Local archival holdings linked into a vast network of related holdings, providing global context and enhancing the value of the holdings.
- Integration of cooperative data into the local archival discovery system, providing pathways from local holdings into the global network, and providing high quality contextual data.
- Enhance the value of the holdings by exposing the interrelations among holdings in multiple repositories.
- A reference resource for both processing and reference staff.