

EAD files in CIC Metadata Repository

Online EAD finding aids have been created by a number of CIC libraries and are used to describe significant archival collections, other kinds of collections, and many important individual resources contained in those collections. While a relatively modest percentage of the individual objects mentioned in EAD finding aids have so far been digitized, much more digitization of such resources is planned. For the CIC metadata portal we currently harvest metadata records describing online EAD finding aids (or components of finding aids) available from:

- The Michigan State University Vincent Voice Library
- [University of Illinois at Urbana-Champaign Archives](#)

The University of Illinois at UC additionally provides metadata for harvesting describing non-EAD finding aids and other archival collection holdings. We also harvest additional metadata records describing individual objects mentioned in EAD finding aids (e.g., the Frank M. Hohenberger photograph collection at Indiana University, digitized images described in selected University of Illinois at UC finding aids).

EAD finding aids in their entirety are not being harvested. Generally, in cases where a metadata record for the EAD finding aid has been harvested, metadata about individual digital objects mentioned in that EAD finding aid are not being harvested (and vice versa), or at least there is no explicit link between object metadata and finding aid metadata provided. Moreover, the methods used to describe the EAD finding aids as information resources (e.g., cross-walks used from EAD to DC) are not consistent from institution to institution. Based on conversations with several participating CIC institutions, we'd like to explore better and more uniform ways to incorporate metadata about EAD finding aids and objects mentioned in EAD finding aids in the CIC metadata repository and possibly in a related portal or CIC EAD registry.

1 Remarks on the finding aid related metadata currently accessible through the CIC metadata portal

⇒ EAD-derived metadata record structure & overlap atypical

EAD files and other archival finding aids mentioned above are represented in the current CIC metadata portal by about 8,000 records. EAD finding aids are often being represented by metadata records that have a very unique structure. Due to their nature these records tend to contain a lot of fields and a very general description (up to 144 metadata elements per record, while the average number is 10). This might lead to an overwhelming list of results from finding aids, especially if individual records derived from finding aids are added and contain duplicate data from the same finding aid.

⇒ EAD do not fit in the distinction digital/analog resources

Most of the records describing EAD finding aids point to HTML versions of the finding aids or parts of the finding aids. Often these finding aid components describe only analog resources. This is not clear to the end-user when viewing the short result metadata records.¹ The finding aids records we're currently harvesting only contain a few records linked to digital resources. Given the long-term focus of the CIC metadata portal on digital content, descriptions of subordinate components in a finding aid that describe individual analog folders, items or series should not be included in the CIC metadata portal

¹ See the article by Chris Prom and Tom Habing "Using the Open Archives Initiative Protocols with EAD" in International Conference on Digital Libraries, Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries 2002

⇒ **No access to the “online item” when it exists**

As the metadata records we have are derived from only parts of the EAD finding aids, there is no indication in the metadata we have of the existence of an online version of any of the items described by the finding aid, even when such exists – e.g., selected of the Vincent Voice Library tapes have been digitized and are available online, selected of the photographs described by the Illinois EAD files are available online, but there is no way for the end-user to know this looking at the metadata we've harvested. A method should be developed to provide a link to digital objects referred to in EAD finding aids via the <dao> or <daogrp> elements.

2 Distinctive features of EAD finding aids

The “put-everything-together-and-see-what-happens” philosophy of the metadata portal leads to a mix of items of different nature. EAD files in particular are both metadata and a resource in and of themselves. Their hybrid nature leads to difficulties in presentation of results to end-users. Treating metadata about and derived from EAD finding aids in the same manner as other metadata creates a result that does not appear to be fully comprehensible to end-users.

Currently the CIC portal has the concept of an item, a collection of items, a record (description of either an item or of a collection of items). The items may or may not be accessible online.

A finding aid does not describe an item; a finding aid does not describe a collection of items; it describes collections, items, and the relationships and context between those items. A finding aid is a digital resource which describes other resources, whether analog or digital. Archival description proceeds from the general to the specific, using a series of hierarchical relationships represented in the <archdesc> and <dsc> (description of subordinate components) elements. The finding aid is neither an item, nor a collection, but it is a description of a collection and the subordinate parts of that collection, including possibly series, subseries, folders, and items. Therefore descriptive information at the top-levels of the finding aid is theoretically inherited by the lower levels. However, including inherited information in an OAI record complicates retrieval by introducing duplicate metadata into many files.

A finding aid is typically represented as an xml document, which may or may not be publicly exposed. Markup practices vary extensively at many institutions, making the application of a common stylesheet impractical. Furthermore, the current DTD definitions allow no way for institutions to point to their preferred HTML representation of the content in the entire EAD/XML file. However, it may be possible to extract urls or other pointers to digital objects linked from a finding aid using the <dao> and <daogrp> elements.

3 Integrating EAD files in the CIC metadata portal: a strategy

The Vincent Voice Library from the Michigan State University is composed of 357 EAD files, providing 1239 OAI records with a dc:type=text. A record describing the entire EAD file would be of type text. Another record describing the collection represented would mention the fact that this is a collection of audio files. Individual records describing the digital items mentioned in the EAD files (if available) would be of type audio. A collection record describing the collection of all EAD files from the Vincent Voice Library would also be desirable.

- A. **Each EAD finding aid as a whole is an online resource** and should be represented in the CIC metadata aggregation by its own individual descriptive metadata record. The metadata record describing the EAD as a whole should be derived from the EAD content above the <DSC> element in accord with an appropriate, generic cross-walk following the crosswalk defined in the EAD Application guidelines² and existing initiatives developed in the similar contexts.
- B. **Collection level descriptions** are used to provide context for items contained in the CIC metadata repository. Metadata records that describe complete EAD finding aids contained in a larger collection should be tied in an appropriate manner to collection level descriptions of the larger EAD collection. If an EAD describes individual items for which a digital surrogate is known to exist, then the EAD itself represents a collection, and an appropriate collection level description will be derived from the top-level EAD elements (and possibly other sources as well).

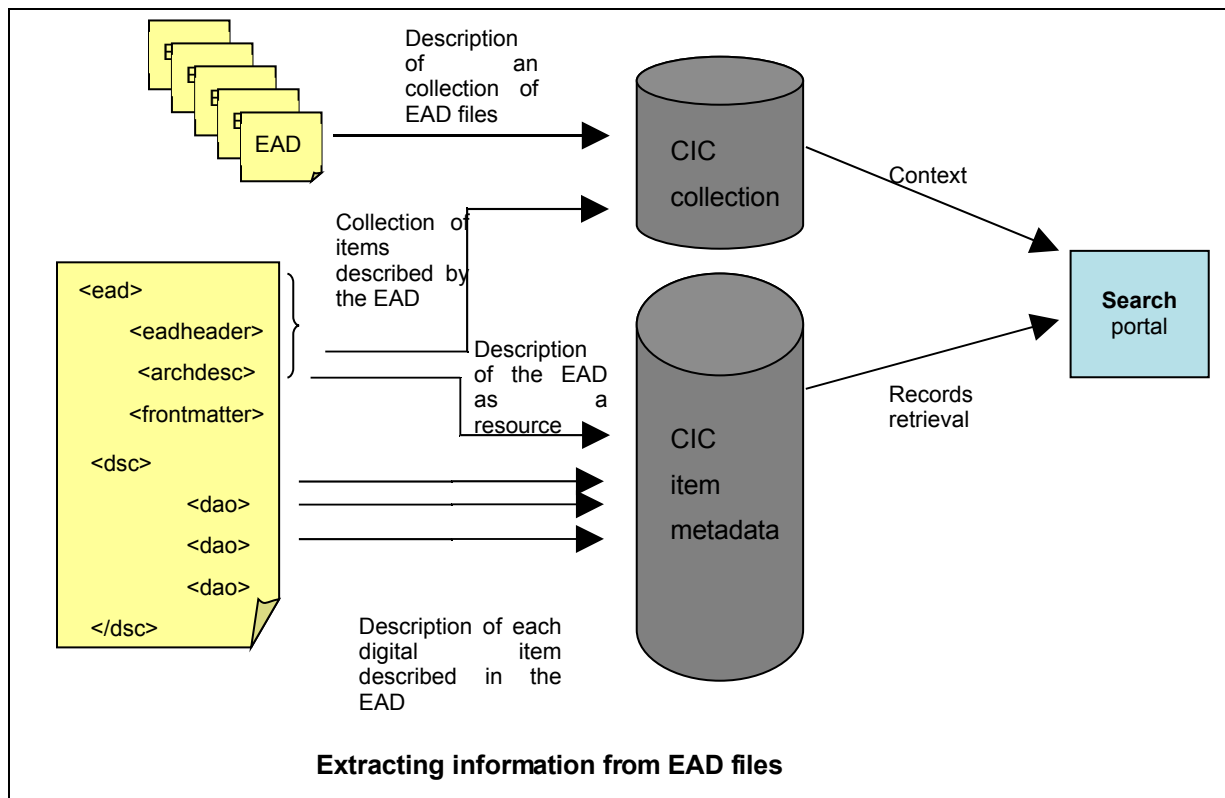
² <http://lcweb.loc.gov/ead/ag/agappb.html#sec3>

- C. **Objects mentioned within an EAD for which a digital surrogate** is known to exist should have their own descriptive metadata record included in the CIC metadata aggregation. These records should include a relationship element tying the object description to a collection-level description derived from the parent EAD finding aid.

Example of search result in the CIC portal on the term “war”

Title	War chief and family, with horses and sled, Acoma
Author/Creator	Hohenberger, Frank Michael, 1876-
Contributor	Indiana University. Digital Library Program
	Lilly Library (Indiana University, Bloomington)
Type	image
URL	http://purl.dlib.indiana.edu/iudl/lilly/hohenberger/Hoh037.000.0027
See also	IsPartOf http://www.dlib.indiana.edu/collections/lilly/hohenberger/index.html
Collection	Frank M. Hohenberger PhotographCollection

The “collection” is added for each result in order to provide a context to the hit.



4 A separate EAD portal

While the above approach for integrating EAD information into the CIC metadata portal should help make EADs and digital information resources they describe visible within the larger aggregation of CIC metadata, the approach does not fully exploit all the information contained in the EAD files. We would propose implementing for experimental purposes a separate portal to EAD finding aids only.

A. An EAD registry

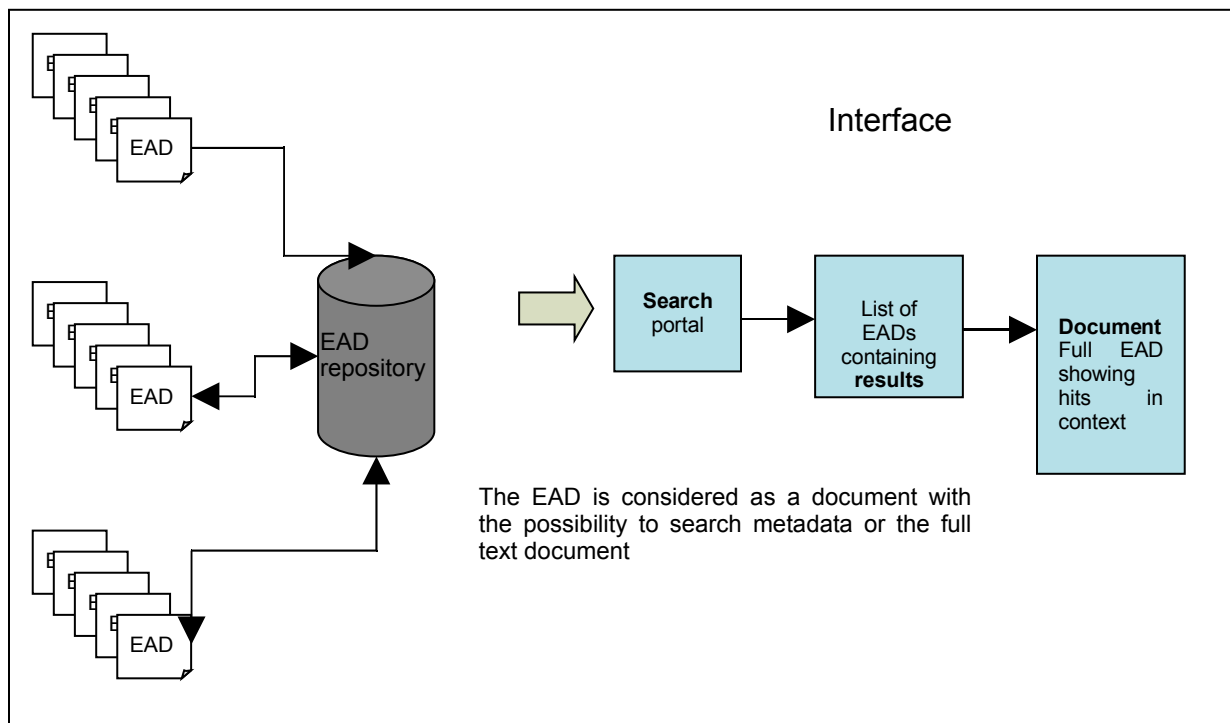
All records describing EAD files as a whole would be made available through a registry. Digital objects could be linked as “illustrations” and all those EAD records could lead to the original EAD file (HTML version).

B. A common interface to EAD files

The objective would be to provide a search interface which would lead to the relevant part of the finding aid, represented in its context, thanks to a general description.

The DLXS software from the University of Michigan manages EAD files and offers an interface to it. It has allowed to display the University of Illinois at Urbana-Champaign finding aids <http://nergal.grainger.uiuc.edu/cgi/f/findaid/findaid-idx>.

This could be used to test an interface for aggregated EAD files (not necessarily public) if several CIC institutions agree to contribute material for the test.



5 Contributing to the CIC-EAD service

In order to implement this strategy, the CIC service would collect the EADs.

⇒ A proof-of-concept on finding aids information aggregation

A XML copy of the complete finding aid would be made available on the Web or sent by email to the CIC service. (If complete finding aid in XML is available via the Web, the metadata record describing EAD finding aid should include the URL for the XML file, e.g.,)

The CIC service would

- Develop a standard process to create the collection level description and the EAD description;
- Develop an interface for a registry of finding aids in CIC institutions;
- Integrate the EAD resources in the CIC portal;

- Optionally aggregate EAD files into a specific portal.

⇒ **Long-term strategy**

The longer term objective would be that interested data providers contribute EAD files in OAI repositories. Clay Redding at Princeton has built an OAI repository with EAD files and is creating a toolkit to include EAD files into static repositories. This work could be usefully re-used by the CIC institution in order to ensure a very low technical barrier to EAD files sharing.

The CIC service would harvest EAD finding aids directly using OAI-PMH. Clay Redding has provided an EAD XML Schema appropriate for this purpose. This schema could be used until such point as the EAD Schema working group provides an official version of the schema (expected within 6 months).