

OASIS Indexing Decisions

Non-public indexes

ID Number

<eadid> (used for updating finding aids by OIS)

Public indexes

1. **Anywhere**

entire finding aid except <eadheader>

2. **Names (organizations) and Names (people and organizations):**

<name> anywhere

<persname> anywhere

<corpname> anywhere

<famname> anywhere

NOTE: Dates and relation information found in name elements are indexed as keywords together with the names. Names should be marked up to allow for optimal retrieval (i.e. normalized as much as possible, especially punctuation).

3. **Places**

<geogname> anywhere

4. **Titles (books, poems, songs, etc.)**

<title> anywhere

NOTE: These are titles of *works* (as opposed to cataloger-created descriptions)

5. **Subjects and Genres**

<ead><archdesc><scopecontent> anywhere

<ead><archdesc><<bioghist> anywhere

<occupation> anywhere

<genreform> anywhere

<subject> anywhere

6. **Repository Name**

<ead><archdesc><did><repository>

7. **Call number/accession number**

<ead><archdesc><did><unitid>

<ead><archdesc><<admininfo><acqinfo><p><num>

<num> anywhere

8. **Dates**

<date> within <dsc>

<unitdate> anywhere

NOTE: All dates need to be four-digit numbers. That is, if a date element contains the phrase "1920-1930," a search for 1920 will retrieve that finding aid, but if the phrase is "1920-30," it won't.

March 30, 2011

© 2011 President and Fellows of Harvard College
<http://hul.harvard.edu/ois/systems/mat/indexing-decisions.pdf>

Additionally, only 1920 and 1930 are actually indexed. That is, if a date element contains the phrase "1920-1930," a search for 1920 will retrieve that finding aid, a search for 1930 will retrieve the finding aid, and a search for 1920-1930 will retrieve that finding aid. A search for 1925 will not.

9. **Container Listing:**

anything with a <c>

Notes

Punctuation is ignored in indexing: () . , [] -

Diacritics and special characters are normalized for indexing (ISO Latin-1).

The names of processors are discoverable via keyword search. Only the encoder's name, as tagged in the <eadheader>, is completely blocked from the public.

Names of donors can be identified via keyword search, but are not indexed.

The user is presented with a list of finding aids that met her/his search criteria, with each finding aid only appearing once on the list. The user selects a finding aid and is given a "key word in context" view of the document, with each occurrence of the search term highlighted. On selecting the "full text view" of the finding aid the user enters the document at its top, and is directed to do a key word search using the browser's "find" feature within that finding aid.

At this point, all of the indexes are envisioned as being key word indices (as opposed to exact string indexes) because of the lack of standardization in the forms of these elements and the lack of subelements (such as surname and forename) to help.