Cataloguing CERN's archival records - principles and practice

Introduction

What are archival records?

Archival records are original records (in any medium) that are produced in the normal course of business or other activities and are judged worthy of permanent preservation. For most organisations this requires selection of around 5% of the total records created. They provide information about, and evidence for, these activities; but the context in which they were created must be understood before their content can be interpreted. It is important to know who created the record, when, why, how it relates to other records, and so forth. One of the aims of good archival management and cataloguing is to preserve as much as possible of this contextual information, without which the records lack informational and evidential value.

How should archival records be catalogued?

The main requirement for archival finding aids is that they should be compatible with archival description standards (summarised below). This demands a hierarchical cataloguing structure, which should be visually clear to users so it is easy for them to "drill down" through the layers. (It is not usually sufficient, therefore, just to include links of the "broader/narrower" type in an otherwise flat database.) Specialist archival cataloguing software is commercially available,[1] and these packages often include other useful modules such as conservation management and accessions registers. However, Encoded Archival Description (see below) is also widely used, and is sufficient for the needs of most Archives.

This report summarises the main principles and standards relating to archival cataloguing, and describes the current practice in the CERN Archive. It is intended as background for discussion of possible improvements.

Archival cataloguing – principles and standards

Some key standards

General International Standard Archival Description - ISAD(G)

Published by the International Council on Archives (ICA),[2] this provides guidance for the preparation of archival descriptions. It identifies 26 required elements, and stresses that "Archival descriptive standards are based on accepted theoretical principles. For

---

[1] e.g. DS Calm http://www.crxnet.com/page.asp?id=57 (accessed April 2008)

[2] http://www.ica.org/ (accessed April 2008). Two accompanying standards also exist:
International Standard Archival Authority Record for Corporate Bodies, Persons and Families - ISAAR (CPF) . Produced by the ICA to complement the first edition of ISAD(G), it is a way of producing standardised descriptions of the creators of archival documents, or archival authority records. These can then be shared between different repositories holding archival material from the same source.
International Standard for Describing Functions - ICA-ISDF. The first version of was released by the ICA in April 2008, and allows for standardised descriptions of functions.

example, the principle that archival description proceeds from the general to the specific is the practical consequence of the principle of respect des fonds."

Encoded Archival Description - EAD[3]
The EAD DTD, a non-proprietary encoding standard for machine-readable finding aids, was first developed at the University of California, Berkeley, in 1993. It is maintained in the Network Development and MARC Standards Office of the Library of Congress (LC) in partnership with the Society of American Archivists, and is compatible with ISAD(G).

Some key principles

Provenance and original order (*Respect des fonds*)
The main principle of archive management is *respect des fonds,* which comprises respect for provenance and respect for the original order of the records.

- Respect for provenance: Provenance is the relationship between records and the organizations or individuals that created, accumulated and/or maintained them. Records of different provenance must always kept separate (not, for example, combined to produce subject-based collections, or to fill in "gaps").
- Respect for original order: The organization and arrangement of records reveal much about the people and activities that produced them. If such an original order is discernible it must be preserved (i.e. the records should not be reorganised into more convenient or logical-seeming arrangements).

Of course, any sort of virtual arrangements can be made *in addition* to this, as is currently being done by supplying the Library with full texts of items missing from its own series.

Cataloguing should proceed from the general to the specific
Different levels of arrangement can be identified in archival collections,[4] and these are in turn represented in the levels of description. Archival description should proceed from the general to the specific, beginning with the highest level description, the fonds (e.g. CERN), then working down through sub-fonds (e.g. the IT Department), series, sub-series, etc., to item-level descriptions, as required. (The "item" is often a file rather than an individual document.)

Archival cataloguing - practice at CERN

CERN Archive holdings
The CERN Archive includes files of letters, memos, reports, rough notes and manuscripts, and other records produced by the CERN Council and its subordinate Committees, by previous Directors-General and other senior staff, and by CERN Divisions, and selected Experiments and Committees. It also contains the Pauli Archive: a collection of correspondence, manuscripts and other material representing the scientific legacy of Wolfgang Pauli (Nobel Laureate, 1945). These records are of interest to the Organization itself, and to the wider community for scholarly research. Some records held in the Archive are similar to Library holdings - CERN scientific documents,

---

[3] http://lcweb.loc.gov/ead/ (accessed April 2008)
[4] The word "collection" is used for convenience, but it is generally avoided in the archival profession as it implies the artificial creation of groups rather than the natural accumulation of archival fonds.

including divisional reports, for example; but the majority are not, which is why a different approach to their cataloguing is needed.

Web pages
The CDS/Aleph database is the main finding aid for the CERN Archive. However, because it does not provide all the functionality required for archival cataloguing it has been supplemented by Web pages (linked to the corresponding CDS records), including:

- A collection guide, which gives an overview of the Archive and acts as a gateway to other finding aids. This work is ongoing, but eventually each section of the guide will link to a collection-level description and to the lower level descriptions in CDS/Aleph.
- Fonds-level and collection-level descriptions based on ISAD(G). There are links between these pages and the relevant database entries.
- An authority record based on ISAAR (CPF).

CDS/Aleph database
The archival entries have been made to conform reasonably well to those of the Library in order to present a uniform information resource for researchers. Sections from the Archive collection guide are used as "narrow search" categories so users can limit their search if they wish. Apart from the fundamental requirements outlined above (and partially compensated by the provision of other Web-based finding aids), some specific areas for discussion concerning the cataloguing of archival records in CDS include:

- Should the Archive and Library bases continue to be searched together?
- If so, how should we deal with records that appear in both bases? Multiple hits are confusing and unhelpful for the user, especially since the records are not usually comparable (most Library entries are at the level of the document, whereas most Archive entries are at the level of the file or box).
- Links to full text (even for items provided from the Archive) are currently made only from the Library base (except for the Pauli fonds). This is done to minimise confusion to the user (if s/he wants a CERN scientific document, the first source is the Library, with the Archive supplying any missing documents where possible). Should this be changed?
- How can the display of "related numbers" be improved? To facilitate the retrieval of individual documents, "related numbers" are now included in the Archive catalogue entries (this was introduced relatively recently, so coverage is incomplete). This means that the description of a file will include, in addition to the file's own reference number, a list of the reference numbers of any CERN scientific documents contained in it. This improves searching, but produces over-crowded catalogue records.
- How could the date-range search interface be improved? A researcher interested in events that happened at CERN between 1960 and 1970 might find relevant items in a file ranging from 1969-2001, or 1954-1962, or 1964-1965, etc. It would be good if the user could just specify the period that is of interest, and let the system sort out the first date/last date search combinations in the background.

<div align="right">
Anita Hollier, DSU/SI
May 2008
</div>