§6. SGML-based Archival Finding Aid for Fusion Science Archives

Namba, C.,
Gotoh, H. (Kyoto Univ.),
Takaiwa, Y. (Tsukuba Univ. Technology),
Hanaoka, S., Iguchi, H.,
Sekimoto, M. (KEK)

A new Japanese act "Public Records and Archives Management Act" was approved in 2009, and came into force on April 1, 2011. Thus importance of archives is gradually recognized in Japanese societies. In NIFS a formal organization for archival activities "Fusion Science Archives (FSA)" was established in 2005. Since then historical materials on fusion research and/or organizations related to fusion research have been collected and preserved at FSA. They are expected to serve as the evidence of various facts or actions taken by fusion research community in the past. However, in order to access to these historical materials an appropriate catalogue of registered materials and a convenient electronic finding aid available through Internet are required.

In general, archival material is originally produced as a record of a certain action taken by the organization. And after completion of this action, these materials are preserved as archival materials. Thus these materials essentially include circumstances of the action or processes, in other words, archival materials have hierarchy structure reflecting the structure of the action and/or that of the organizations. This means that archival materials should be understood in the context of original actions. Finding aid for the historical materials should reflect the specific features of these materials.

Standard Generalized Markup Language (SGML) is recommended for description of archival database, because it allows to describe hierarchy structure of each record in the database. We use so-called XML (Extensible Markup Language), based on SGML and available for web utilization. XML requires DTD (Document Type Definition), as well as SGML, which provides a formal set of rules for the ordering and frequency of the elements in XML.

Encoded Archival Description (EAD) is a de-facto standard for data of archival finding aid and is accepted as an international standard as DTD for XML. EAD DTD defines the markup and rules for encoding an archival finding aid. For example, element declarations, attribute declarations, parameter entity declarations and so on. In Fig. 1 we show a hierarchical tree structure of EAD.

EAD makes possible to describe the hierarchy structure of archival data: they are 1) information on the original organization, which produced the document, 2) the history of the documents, 3) person who preserved the documents, 4) place, where the documents are stored and so on. When the database is based on EAD, the retrieval results are shown in the hierarchy order and help us to understand the structure of archival data as is shown in Fig. 2.

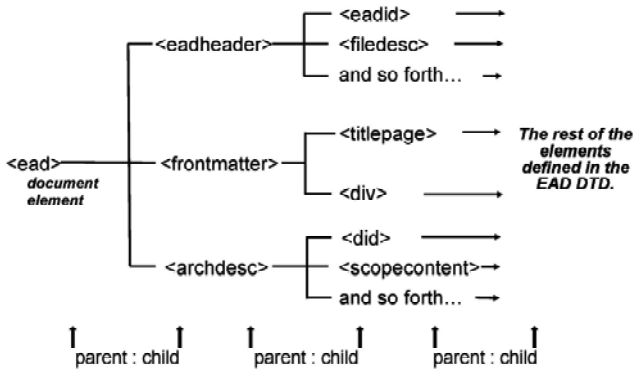

Fig. 1 Hierarchical tree structure of EAD[1)]

Fig. 2 An example of retrieval results showing the hierarchy structure

Here, we introduced following hierarchy structure in our EAD-based database.

Repository level : Fusion Science Archives
Collection level : a set of materials given by a certain individual or organization.
Series level : a set of materials in one box (ID = B301a, for example)
File level : a set of materials in one file (ID = B301-01, for example)
Item level : each individual material (ID = B301-01-01, for example)

Note that, in some cases we have no file level description.

Our archival database was constructed utilizing the tool developed at National Institute of Japanese Literature (NIJL), so called "Archival materials information sharing Database (AMISDB)" under the collaboration with Sokendai, NIJL, High Energy Accelerator Research Organization (KEK) and other researchers from universities, and now available through homepage of Fusion Science Archives at URL= http://www.nifs.ac.jp/archives/index.html.

1) EAD Application Guidelines for Version 1.0 : SGML and XML Concept http://www.loc.gov/ead/ag/agconc.html#sec2)

This work was conducted under NIFS Collaborative Research Program (NIFS07KVXJ010).