§5. Comparison of Two Strategies for Cooperative Internet Searching System of Multiple Scientific Archives – Open Source Software and Cloud-type Server

Takaiwa, Y. (Tsukuba Univ. of Tech.)[1],
Gotoh, H. (Kyoto Univ. Museum),
Namba, C., Iguchi, H., Matsuoka, K., Hemuki, S.,
Kikutani, E., Sekimoto, M., Nakamura, Y. (KEK),
Kugo, T. (Kyoto Sangyo Univ.),
Tanabashi, M. (Nagoya Univ.),
Kanaya, K., Ukegawa, F. (Univ. Tsukuba),
Yoshikawa, T. (Nagoya Women's Univ.),
Kimura, K., Ohshima, Y., Suzuki, S. (IMS),
Murakami, M. (NIPS),
Yagyu, S. (Grad. Univ. Advanced Studies),
Yamada, Y. (NAO)

i) Goals For scientific archives such as the Fusion Science Archives (FSA) at NIFS, maintaining their catalog databases available to the users is the key task. Considering that the studies using such archives may not be confined in a single archives, it would be useful if one can find related documents of different archives by a unified platform of database services, which we call as *Archives Information Sharing* (AIS). Projects of promoting the AIS for scientific archives have been performed under the NIFS collaboration program, and the final common goal of them is a well-managed system to provide such services via internet. Two different candidate systems have been studied; one is the system based on an open-source software system and the other is the system of a cloud type database server.

Previously two projects, though closely related and cooperating, were launched: "Development of a Cooperative Internet Searching System of Multiple Scientific Archives" (project ***a***) and "Construction of Archival Information Database Utilizing Cloud-type Server"(project ***b***), which were reported in a previous NIFS Annual Reports. In this year, entering into the next stage, they are combined to one for the purpose of evaluation. This article, thus, summarizes the procedures for realization of AIS and comparison of two.

ii) Procedures for AIS AIS requires that archival data of different archives are imported to a single unified system of databases in which documents searches for multiple archives are made possible.

1. Databases at Individual Archives Individual archives maintains its own database of documents (*"Finding Aids"*) which may use any softwares. In this project Filemaker Pro and EXCEL are used at the participating archives. In order to unify the data from different archives into a unified database they must comply with standards of archival data descriptions. We adopt the EAD as the standard in the first stage because it is recognized as a *defacto* standard, however, the full-flexibilities of the standards have to be reduced due to the limitation of target unified systems.

The point at this stage was the notion of hierarchy structure in archival documents arrangements originating from well-known archives principles of provenance, original orders, respect of fonds, *etc.* By this, the documents description may be proceeded from global structure (higher hierarchy) down to detailed documents (lower hierachy), and intermediate hierarhy levels which may be physical units of storage or conceptual classifications of the materials, are processed as needed for convenience. The databases have to be compliant with the structure, and if the proper standards are adopted the procedures can be standardized, too.

2. The Unified Systems The system for the project ***a*** is Archon, which is constructed with PHP and MySQL for web pages and database engine, Apache as the web server, on the Linux OS, which is located at the YITP, Kyoto University, and which accomodates Yukawa, Tomonaga and Sakata memorial archives database data. For project ***b***, the Sokendai, together with NIFS and other institutes, made a rental contract of Application Provide Service (APS) site of Infolib with Infocom Company. The Infolib is the general purpose archival database management system.

3. Data Transfer to the Unified Systems Two methods of data transfer from individual archives to the unified systems are supported; manual input to the web form boxes and bulk data transfer using specially formatted data files. For the latter method CSV can be used in any systems. for systems. The Archon accepts data by the EAD (XML) tag structured simple text files. On the other hand, the Infolib may accept EXCEL files as well as CSV. In order to prepare these specially formatted files it is required to write scripts of Filemaker and EXCEL at individual archives databases.

iii) Summary Above procedures are summarized in the table below for comparison of two systems.

	platform	**system**	**Transfer**
a	Linux PC	Archon open software:	CSV, EAD, manual (web)
b	Infocom: cloud data server	Infolib ASP	CSV, EXCEL, manual (web)

In both systems unified archives databases may be constructed in a similar manner if individual databases were made compatible with the standards. Apart from the technical details, the difference is found in distribution of cost for various tasks in managing systems; open softwares require more time and man-power for designing, setting-up, and customization, on the other hand for commercial rental servers customization option cost may be charged as addition to the contract fee.

[1] present affiliation, KEK.