

Towards Cooperative Authority Control

MIYAZAWA Akira
National Institute of Informatics

1. Introduction
2. Operation model
3. Database configuration model
4. Database format model

1. Introduction

Here I am proposing some models for authority control work and its implementation from technical viewpoint. The purpose of these modeling is for the consideration of our effort to materialize international cooperation on authority control. At the previous meeting, we have recognized we have still long way to go. Without establishment of authority control in the library side, cooperation is just a dream. And none of three countries have achieved the nation wide authority control system. But, I believe it is still worth to figure out concrete design of our goal, because, without it, our discussion might be quite uncertain and vague.

In the following sections, I present models in three aspects. The first one is operation model, which is to identify what authority control is. The second one is database configuration model, which is for consideration of physical sharing method of cooperative data. The last one is database format model, which discusses the contents of our shared authority database, though it is not fully developed in this stage.

2. Operation model

The first phase of the authority control work is “Search phase” which is started by a cataloger who has a book to be cataloged and has found a name to be recorded as a heading. He/she, then, searches the authority file (the file can be either machine readable, book form or card form), if the name is already in the file.

When the name is found in the file, the second phase “Identification phase” is operated. He/she must make sure the name at hand is the same person/organization in the authority file. If the name at hand is identified as the name found in the file, just use it following relevant procedure of the cataloging environment.

When the name is *not* found in the file, the third phase “Confirmation phase” starts. This is to check whether the name is really new one’s name or just variant names of already listed one. Of course, if the name is found in *see also* reference, this is just the “Identification phase”. But, still the new name can be a new variation of existing entry, and if so, he/she will register the name at hand as *see also* reference of the entry.

When the name is really new one, he/she goes into the last phase “Establishment phase”. Following an appropriate cataloging rule, he/she establishes a new unique heading form of the name, and then, registers the name to the authority file. In the case of book form, the new heading may be kept at a temporary file and waits update in the next version.

3. Database configuration model

Centralized model maintains one database at a certain site. Operation can be carried out at more than one site, as far as the search and registration turn around time is short enough for operation (online operation may be essential). But, search phase, and registration phase must be operated against this centralized file (though, they can be operated from distant site). For the operation of centralized works, especially for identification phase and confirmation phase, well-organized coordination will be necessary to keep consistency of the authority file.

Distributed model maintains more than one database at more than one (usually distant) site. Each database contains a part of the whole database exclusively. Therefore the search phase must access to the all sites, and registration phase must be operated at a certain site designated for the particular record. How to divide whole database into these sites depends on the design. For example, nationality of the person may be used to decide the site. (Though, it may be problematic in some cases). Identification phase and Confirmation phase are, naturally, done in corresponding site.

Shared distributed model maintains, again, more than one database. But, in this case, each database in the site is identical and contains whole entries of the database. Proper protocol, or method of synchronization, is required to keep these databases identical. Addition to the database or change of data must be sent to other sites, and must be processed in each site. This has advantage for search phase, because you need not to access to distant site. It will shorten the turn around time for search phase. And also registration phase can be done on the site. But, there may be a problem that concurrent update can cause inconsistency in this configuration. We might need some additional rule or conventions for operation.

4. Database format model

Usually, an author name authority record consists of an established heading, *see*

from references, *see also* references, links to other entries and notes, together with some control fields, such as identifier, record creation date and record update date. But, in our environment, where multiple cataloging rules are in effect and multiple control operations are carried out, we may have to extend this format.

One possible solution is a simple merged format. This is just a merger of formats used in each participant. If the participants are just Japan, China and Korea, the format will have an identifier, all the fields of Japanese authority record, all the fields of Chinese authority record and all the fields of Korean authority record. When these participants have individual records for same person, simply merge these records to incorporate one CJK authority record. Advantage of this format is its simplicity and is easy to start, but you have to note that the processing other participant's portion of a record is not easy work.

Of course, we can make new format fitting for our requirements. But, we don't have our concrete requirement yet, and we need further investigation of this area.

On the other side, there may be a solution to maintain individual format within each site. This solution premises distributed or shared distributed mode in configuration. In this case, each site needs to receive transaction from other sites. If the sites send transaction in their native format, receiver must be prepared to process all these formats. It may not be big problem if there are just two outside formats. But, if we have three, four or more formats to handle, it will be too much complicated to maintain. Therefore it is preferable to have common interchange format here. In this case, each site has to convert their transaction format to this common format. This approach has advantage that each site can keep their individual control system without change, and they can simply add systems for cooperation.