A bibliographic logistics for processing mega-medium collections: a practical system of the NCL to solve the current barriers in cataloging and its services

Diing-jong Yan and Ching-fen Hsu
Bureau of International Exchange of Publications, Taipei, Taiwan, China

Abstract:

The NCL—like all Asian libraries—will face, in the area of cataloging, barriers in terms of language, types of data types, and services. The various MARC (or metadata) conversion of titles will become the norm in cataloging. Our Library is constantly striving to establish a pan-bibliographic system to break through these barriers.

In the past three decades, in order to manage traditional collections and digital collections, we have built more than 60 databases or systems. Each of these systems separately manages a different format of bibliographic record in order to highlight the data’s language, bibliographic type, theme, or content characteristics. For different purposes, some of the data will duplicate the information in currently existing databases. These databases were originally intended to enhance the effectiveness of bibliographic control. Nevertheless, they not only were costly in terms of library maintenance resources, but also caused problems in accessing the collected information.

With a view toward having good bibliographic control in order to break down existing barriers and to face future challenges, the NCL started a new project in 2010: the development of a MARC-compatible system that is sufficiently compatible with the Library's current bibliographic records. This system's data structure is based on MARC21 and also utilizes special techniques to integrate a variety of metadata formats beyond MARC. The system is designed not only to integrate various formats of bibliographic records, but also to preserve the original bibliographic records. In addition, it is designed to analyze in every way all the relevant bibliographic information in order to enhance standard bibliographic control and every kind of information and content service.
We believe that library bibliographic control in both the present and future are the foundation of information access. Hence this article will present a brand-new bibliographic system, which became fully functional in its first phase in 2010. This system from a rather broad perspective offers the design and development of bibliographic management processes in order to maintain and accommodate the compatibility and expanded flexibility of bibliographic records from a variety of sources. We hope that by trying the same platform to manage metadata from a variety of traditional and digital collections, we will eliminate barriers to bibliographic control that are encountered by libraries.

We firmly believe that this new complex library bibliographic system, with its comprehensive functionality, will certainly be able to resolve or reduce current difficulties in library management and that it will also deliver Library 2.0 services for our patrons in the new century.

1. The NCL's system: Development and difficulty:

1.1. Overview of development:

Founded in Nanjing in 1933, the library moved to Taipei and commenced operations in 1954. A computer room was established in 1982 and in 1984 a "bibliographic information service system - NCLAIS (NCL Automated Information Service). This service began using computers to process bibliographic data. Among its paper publications were the following: "Catalog of Books Published in the Republic of China;" "Index of Periodical Articles from the Republic of China;" "Index to the Official Gazette of the Republic of China;" "Union Catalog of Chinese Periodicals from of the Republic of China;" and "Union Catalog of Western-language Periodicals in the Humanities and Social Sciences Union Catalog." The NCLAIS has also issued a computer card catalog. In October 1994, the NCL's information network officially began operation and established seven large databases: NCL Online Catalog; the Index to Periodical Literature System; the NCL [ROC] Gazette Online; the [ROC] Government Publication Catalog; the Official Publications Echo Network; (Contemporary Literature; and the NBINet Union Catalog. They contain nearly 1.7 million entries on a variety of Chinese books catalog and can be searched in the library, over the telephone, or online.1

By 1993, the bibliographic system established by the NCL had come to include 41 comprehensive or specialized bibliographic (or full-text) databases. At the end of 2002, the main bibliographic system had 8,411,850 entries. While building the "Comprehensive Library Bibliographic System" in 2010, we looked at 67 bibliographic record systems. We selected the bibliographic data from 58 of those to be the principal source of content for the "integrated bibliographic system." These systems held 18,845,376 bibliographic and authority records at the end of 2010.

Looking back at the establishment of these systems, we see that each had its own background and bibliographic control requirements at different times. Their contents can be divided into five categories:
1) The library catalog system; includes modules for interviews, cataloging, periodicals, circulation, OPAC, WEBPAC, etc.

2) The national catalog system. This consists of the following systems: nationwide bibliographic information system; union catalogs for Chinese rare books and genealogy; a national catalog of government publications; official overseas reports; periodical guides; newspapers; information on new books; ISBN’s; doctoral and master's theses; and full-text images of Chinese rare books.

3) Document information systems. They consist of images of indexes to periodical articles; table of contents services for periodicals; electronic images and full text of the [ROC] Gazette; the government statistical survey; stone and metal rubbings; posters from the world of arts in Taiwan; etc.

4) Specialized catalog systems. They include catalogs of treatises on the Confucian classics, on Dunhuang, and on Han dynasty philosophers; a catalog of research papers on Chinese culture; local Taiwan documents; Taiwan memoirs; overviews of Taiwan; etc.

5) Information systems on persons, i.e., contemporary literary materials; contemporary artists; who's who in government agencies; etc.

Through the establishment of these systems, we have made the NCL’s resources fully available for our users to search and utilize. This is not a national bibliographic information system established by a commercial database company to create. Instead, it is a product of a library, the likes of which libraries around the world seldom develop.

1.2. Difficulties in the development of the NCL’s system:

Too many queries, too much browsing leave system at a loss:

When the NCL’s information system was established, it did not have an integrated query interface. Readers needed to search each query one by one from each system, and then compiled and compared the results themselves. In 2004, the Library established a decentralized integrated distributed query interface known as "Chinese Knowledge Portal" to provide readers with a single query interface. Even so, at the time of the query, they had to wait for each system to return the results separately. In 2007, therefore, the NCL installed a centralized integrated query interface called the "ISSR system," which through the OAI mechanism collected data and set up a centralized database. Because it still could not cover all the systems and all the purchased electronic resources, users accepted it less well than was expected. Hence readers were hesitating at the Library’s home page.

Many systems grow freely out of various parts of the Library:

At present there are 58 databases included in broad bibliographic control system: 40 are still maintained and operated; 14 add no new data; and 4 have been shut down. 12 units established these 58 systems: 22 by the Reading Room; 7 each by Special
Collections, Reference, and Sinology Liaison; 4 by Acquisitions; 3 by Sinological Materials; 2 each by Cataloging and Book Registration; and 1 each by 4 other units. Approximately 42 persons perform system development and maintenance, not including those who archive data. Given "human" differences and the lack of horizontal organization, this type of bibliographic system cannot be efficiently maintained, operated, and developed on a sustained basis.

**Bibliographic clutter due to old and new systems coexisting:**

The NCL’s holdings are recorded and arranged by different systems. They also lack macro-control. Some of the holdings have duplicate bibliographic records; others have none at all. Established bibliographic systems have different archiving formats, working platforms, recording rules, and workflow. All of this causes conflicts in the searching, comparison, and exchange of data.

From bibliographic records to digital text, from series to single books, from books to tables of contents, we are still waiting for innovations in order to provide correct bibliographic information, given the absence of effective authority record control of bibliographic information to provide the correct remains to be innovative. Uneven bibliographic quality and the lack of a sophisticated quality control mechanism will cause the reader trouble in identifying and selecting data after searching.

**Facing the digital era calmly:**

Without consolidation and simplification of its systems, the NCL will get half the result with twice the effort and be unable to improve daily operations in the areas of acquisitions, cataloging, collections, reading, and reference. In addition, the systems lack human resources. In fact, most are one-person systems, with each person having special skills. Unless there is a plan for comprehensive development, there will inevitably be disorder among the existing systems, and the Library will not be able to keep up with the times.

Bibliographic resources provided by the publishing industry are more "timely, accurate, rich, and diverse" compared to those coming from libraries. Hence they naturally win more favor from readers than library catalogs do. Given the exponential growth of knowledge production today, libraries have no time to deal effectively with network and digital resources. Therefore they will gradually neglect the basic functions of preserving human knowledge and disseminating information.

**Review and assessment of the NCL’s systems:**

By law, the NCL is repository for our country’s various publications, academic dissertations, and government publications. Its Cataloging Unit performs the cataloging, and its International Standard Book Number Center is responsible for assigning domestic ISBN's and for CIP preparation. The Bibliographic Information Center maintains and operates the union catalog system that includes 75 domestic libraries. In addition, large amounts of data from the Index to Periodical Literature (established in 1983) is already available in the form of bibliographic resources that are "complete in variety and rich in content." Then, in response to the Library’s requirements in terms of general operations and special projects, data is retrieved
from the aforementioned databases or there are established separate databases with either special subjects (like a bibliography of treatises on Han dynasty philosophers) or special contents (like stone rubbings). These primary and derived databases on the one hand strengthen the Library’s bibliographic resources. On the other hand, with its archiving formats and record standards each going its own way, these databases also create difficulties in management, maintenance, and operation, as well as for the readers using them.

**Evolution in catalog creation and use:**

Starting in the late 1990’s, search engines became users’ first choice in finding information. As a result, the library community has also begun to rethink existing library system functions, bibliographic content, and service modes. Library automation has been growing, indeed, for more than 30 years. Even though the completed Library Information System (LIS) has been constantly developing new features as it follows the transformations in information technology, it still cannot satisfy librarians’ needs in building and servicing library catalogs. Library departments have their own demands of the system, as do library information specialists. Their expectations of the LIS include the following:

1) more direct access to information;
2) ability to communicate through the API interface;
3) stronger interflow with systems inside and outside the NCL;
4) improved interface and function of external bibliographic services. 2

A library’s operational efficiency is of course closely related to the function of its automation system, but the function of each system must protect bibliographic and authority data. Only then can the system be effective.

**2.1. Librarians’ requirements in catalog-building:**

Most libraries at present still use systems of resource organization dating from the 19th century. As for cataloging, they still follow ISBD and individual countries’ cataloging rules, and are very strict in their bibliographic content records. The use of MARC in automated system also works in concert with their cataloging standards. ISBD, AACR2, MARC, classification and heading systems, and other related standards are commonly used worldwide in the bibliographic control of library resources. The resulting output of these operations provides librarians and users with the ability to search, manage, identify, and access all types of information.

At this late date, libraries with their unique MARC formats are carrying large amounts of bibliographic and authority records. Nevertheless, the popularity of the full-text search, and even the rise of search engines, are changing the users’ searching habits. This means a divergence in the ways the library community looks at the formats, content, and presentation of existing bibliographic records. It is certain that MARC—because of its strict, thin, and flat structure—can only with difficulty be used to describe dynamic digital resources. Therefore it cannot appropriately display, in the Internet environment, complex semantics or multi-level relationships among bibliographic data. The content of a MARC bibliographic record aims primarily at the description of a single item, not of the levels of the organization of knowledge. 3 Even
so, the value of MARC records has not yet disappeared completely because of all this criticism. The library community is trying hard to extend the use of MARC. Hence MARC21 is constantly being revised, while MARC and XML are correspondingly increasing the scope of disseminating MARC data in the Internet. Also, the experiment of mechanically applying the FRBR structure to MARC bibliographic data is making catalogers discover the necessity of in-depth analysis of the specific bibliographic characteristics of various academic disciplines. In fact, only by coordinating the revision of cataloging rules and MARC can they utilize MARC structure in showing links between works.4

Because creating MARC records is rather cumbersome, many libraries use non-MARC metadata to build correlated catalogs of digital resources. Cataloging librarians are trying to simplify their bibliographic fields by using Dublin Core fields to create different types of resources, but there is insufficient use of this. MARC and non-MARC metadata have their advantages and disadvantages. Besides training the technical services librarians of the digital era in traditional cataloging, libraries also should seek ways to coexist with other metadata.

2.2. User requirements of cataloging services:

Because users’ expectations of a catalog vary according to the information and the content they require, traditional library catalogs should be able to achieve the users’ single, clear search goals within their existing collections. Nevertheless, given the information-gathering habits of users in the digital age, the “inventory” list type of cataloging service is declining. Driven by search engines, modern users hope to have a single, simple, and easy portal for inquiry but also a simple and easy interface. In this way, they can be stimulated and guided in shuttling around in a broad range of bibliographic knowledge resources. Optimally, if possible, users would enjoy immediate access to online full-text data. By happily utilizing comment, tagging, and other functions, they would share their mutual information needs with other users. Hence users hope that even if they cannot get to a library, they could get these services anywhere, anytime.5 In short, the users need a cataloging service interface that is simple but sophisticated, available everywhere, and fast.

Libraries spend plenty on their cataloging operations, while their catalogers grow tired of processing the endless supply of materials entering the libraries. The use of their public catalogs, however, actually yields less benefit than before. In terms of stability of their business operations, libraries need to seek a balance between their product and their users’ requirements. Library and information science experts suggest the following strategies for making catalogs more flexible:

1) Search for new catalog users;
2) Development of new levels of catalog use;
3) Extension of the catalog’s life cycle through innovation and cost reduction;
4) Development of new ways for current users to use catalogs.6

There are a number of possible innovative practices in cataloging services. They include the following: integrating existing library catalog interface and resource discovery tools so as to improve the query function; and--as for research needs--creating a single item within which there is seamless contact between title, table of
contents, cover, full text, quoted passage, and index. Different types of libraries have different combinations of service projects and staff. Hence each library must determine the orientation of its own internal organization and discuss related issues before it can set appropriate strategic direction for change.7

2.3. The reuse of bibliographic resources:

While browsing documents that discuss the applicability of traditional catalogs, we find titles with terms like "rethink," "restrategize," "repurpose," "revitalize," etc. This indicates that the purpose of reviewing current cataloging operations is not to put an end to MARC. Instead, the purpose is how to put the significant number of MARC records through an infrastructure that reconstructs metadata, thereby reusing MARC as part of a broader, more diverse set of tools, standards, or protocols.8

Cataloging of library resources in the future should be a matter of cooperation between the library and information science communities. As their budgets grow ever tighter, libraries should increase community cooperation and the sharing of bibliographic records through the entire supply chain for information resources, thereby maximizing the use of data.9 The most practical approach is to integrate the metadata created separately by libraries and publishers. During the different stages of the publishing process, the library and information science communities will assign descriptive content for each publication. By cooperating in the mutual utilization of this content, the library and publishing communities will help to improve the inquiry service for readers and sales of publications,10 and they will save the time now spent on the duplication of bibliographic description.

3. Thought and concept behind the NCL’s systems:

3.1. Development strategy for building bibliographic systems with bibliographic structure:

The "bibliographic field area" formed by the generation and dissemination of bibliographies can be divided into three sections: upstream, midstream, downstream. The upstream section is linked with the producers of knowledge and the stage of the production of knowledge. The downstream section is linked with the consumers of knowledge and the stage of the consumption of knowledge. The midstream section is bibliographic management’s core area of technical processing. For the most part, libraries and information agencies are individually or collectively responsible for this last section.

The key point of the bibliographic area is the integrated management of the upstream, midstream, and downstream sections. This causes the bibliographic workflow to link the two stages of knowledge production and knowledge consumption, thereby forming the comprehensive management of every area and section in the realm of bibliography. The NCL calls this management--integrated from top to bottom and from beginning to end--the Bibliographic Logistics System Plan. It has divided the various bibliographic or cataloging management areas into 10 core tasks and brought them into a common system for management. The 10 core tasks of comprehensive bibliographic management are as follows:
Bibliographic workflow management is the basis for bibliographic operational management. In an environment of bibliographic areas or integrated sections, bibliographic records keep pace with books in their functional activities upstream, midstream, and downstream. Thus there is an increased need for bibliographic record projects, including a number of record projects that are extended and related to bibliography. In other words, in every stage of bibliographic workflow, each component of core management and its bibliographic record projects are not quite alike. Regardless of whether such changes increase or decrease the projects, they have formed the phenomenon of the growth of bibliographic records.11 We therefore know that our Library’s bibliographic management has entered the era of "great caution."

2.3. Exploring the relationship between “bibliography” and “full text” in the digital age:

Interlocking unity of bibliography and full text:

In the era of paper resources, "bibliographic titles" and "items" were always separate entities. Usually the "book" came first. Only then was there created a “bibliographic title” based on the book. Later, "bibliographic control" moved toward the mechanism of cataloging-in-publication (CIP), whereby bibliographic titles began to appear earlier than books. With the arrival of the digital age, electronic resources (such as e-books) are contained in one file from the cover, the copyright page, and the table of contents to the full text. As an entity, the "item" almost places "bibliographic data" in that file. As for the differences between them, the principal one is that "classification numbers," “subject terms,” and “keywords”—as used by professional catalogers—turn into "note" data that provide explanations and do textual research. At bottom, they are actually "physical description" data that simply and easily distinguish one book from another. This phenomenon of interlocking unity of bibliography and full text is gradually rendering the contributions and accomplishments of professional catalogers non-distinctive. Since users often end up with full text in the course of searching for information, they either feel that “bibliography” is not that helpful or that like air, sunlight, and water, “bibliography” is important but they don’t notice its existence.
The vertical relationships between the whole and its parts:

In the past, librarians cataloged their collections by emphasizing the “bibliographic unit (book title or named title).” With the publication of large quantities of information...
resources, users’ demand for bibliographic content analysis became more and more ardent. As a result, the analysis in a bibliographic became more thorough and detailed, like going from a tree’s roots to its branches and leaves. In general, the vertical level can be divided into four parts. A book, for example, might contain a series title, a sub-series title, chapter and section titles, and full text. Other traditional and digital documents can also be divided into these four levels, which are built like a tree and can guide searchers to further broaden or narrow the search results.

After bibliographic information on a variety of documents is gathered in one database, the levels of analysis in a bibliographic record are a relationship that cannot be ignored by an entire bibliographic system with organization and structure. This type of corresponding relationship is one of the items needed for a vehicle that finds, identifies, selects, and acquires knowledge. More importantly, it is one of the important bibliographic relationships that must deal with processing when general and thematic bibliographic information is integrated.

<table>
<thead>
<tr>
<th>Type</th>
<th>Series</th>
<th>General Books</th>
<th>Periodicals</th>
<th>Research Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Title</td>
<td>Title</td>
<td>Title</td>
<td>Report Title</td>
</tr>
<tr>
<td>Level 2</td>
<td>Sub-series</td>
<td>Title</td>
<td>Issue Number &amp; Title</td>
<td>Chapter &amp; Section</td>
</tr>
<tr>
<td>Level 3</td>
<td>Chapter &amp; Section Title</td>
<td>Chapter &amp; Article Titles</td>
<td>Title</td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td>Full Text</td>
<td>Full Text</td>
<td>Full Text</td>
<td>Full Text</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Conference Papers</th>
<th>CD’s &amp; DVD’s</th>
<th>Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Conference Title</td>
<td>Collective Title(?)</td>
<td>Title</td>
</tr>
<tr>
<td>Level 2</td>
<td>Proceedings Title</td>
<td>Disc Title</td>
<td>Book &amp; Periodical Titles</td>
</tr>
<tr>
<td>Level 3</td>
<td>Paper Titles</td>
<td>List of Contents</td>
<td>Chapter &amp; Section Titles</td>
</tr>
<tr>
<td>Level 4</td>
<td>Full Text</td>
<td>Contents</td>
<td>Full Text</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Web Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Website Name</td>
</tr>
<tr>
<td>Level 2</td>
<td>Web Page Name</td>
</tr>
<tr>
<td>Level 3</td>
<td>Item (?) / Stage</td>
</tr>
<tr>
<td>Level 4</td>
<td>Web Page Contents</td>
</tr>
</tbody>
</table>

Figure 2: Vertical relationships between whole and parts of different works

Mutually derivative horizontal relationship among bibliographic information:

The first group of FRBR entities includes four basic relationships: works, content expression, vehicle manifestation, and items. While moving forward with the
establishment of our new bibliographic system, we will break it down into the following three aspects of the horizontal relationship.

1) Related works, i.e., each type of item with content expression or vehicle manifestation derived from a single work. Such types include vehicles (like e-books, all kinds of files, etc.), genres (comics, movies, theater, etc.), objects suitable for reading (for children, the visually impaired, et al.), editions (with revised or altered content), and language (translations into a variety of different languages).

2) Adding value to content, i.e., items produced by deducing and explaining the contents of specific works, or items that entail recommendation or analysis of those specific works. Such items include vernacular, annotations, recommendation and appreciation, guided reading, selected works, vocabulary, book reviews, criticism, marketing literature, and special television programs.

3) Influence of the works, i.e., the connection between specific works and other works, or the influence exerted by the recommendation and approval of others. This category includes specific works that are quoted, cited, placed on a reading list, or entered into competitions; or that win awards or to the reader’s attention in the course of the latter’s work or study.

Analysis and exploration of the horizontal relationship between titles provide important guidelines for gathering these work items. In addition, in terms of establishing new bibliographic systems, they will further satisfy the user’s higher functional demands.

3.3 Moving from comprehensive bibliographic queries to integrated queries for bibliographic records:

A comprehensive query system that straddles multiple bibliographic databases provides a model for searching for information, a model that can simultaneously search many databases, has a simple and easy user interface, and offers a variety of options for sorting query results. As such it serves the needs of persons who are trapped in every manner of database and unfortunately search in individual databases. Nevertheless, it has its inconvenient aspects. For example, search results are not all together, and different databases will appear simultaneously, one by one, on the same entity’s same record. Users thus have to browse and filter the results one by one, without getting any content analysis. The comprehensive mechanism is used only for personal inquiry service, and there is a lack of feedback on the work done by librarians.

A comprehensive system treats each bibliographic record equally, ignoring multiple correlations between bibliographic records. As a result, it is increasingly unable to meet the needs of most users. Therefore, we need to construct a new bibliographic system that moves from "comprehensiveness" to “integration.”

The "integrated type" of bibliographic system must squarely face the trend toward the merger of bibliographic records. In addition, it must be able to handle both the vertical relationship between bibliographic whole and part and the horizontal mutually derivative bibliographic relationship. When a user’s search results appear, it can
provide immediate correlation between works and items. By means of the three-dimensional correlational model provided by the system, users can quickly grasp the strategies and techniques for further action. The current comprehensive query system lacks the capability of determining and handling these bibliographic correlations.

### 3.4. Attaining bibliographic record functionality through bibliographic authority correlation:

With regard to individual persons and groups from the second group of FRBR entities, it is demonstrable that a single bibliographic record still cannot make for good bibliographic control or provide search functionality. Bibliographic records principally convey works and their creators; in some cases, the creators are either the theme of the knowledge represented or the represented works themselves. In the digital era, the bibliographic records do not entirely “originate” with a library, nor do they all come from the hands of librarians. In this kind of environment, in which many persons work on bibliographic records with multiple sources of content, implementing the concept of "bibliographic control" requires different thinking or it will not happen.

A bibliographic record is a piece of work held in a cataloger’s hands, and it is inspected and recorded according to any number of cataloging and recording standards. When the system begins to bring together everybody’s bibliographic records, then we find the discrepancies among them, and serious discrepancies mean "bibliography out of control ". At this point we think of display shelves in a store, with only a bar code, i.e. no price tag, on any commodity, because the prices are recorded in an exclusive field in a database. So if one wants to raise price on any commodity, or lower it for a sales promotion, one can only make that revision in that database. Changes and revisions of the content of bibliographic records should be done the same way.

An authority record contains all sorts of different data, such as the correct form and the incorrect forms of a title or an author’s name. All authority records are important criteria for a system as it compares every kind of data in order to differentiate between originals and duplicates. They will also be important efficient instruments in days to come for users as far as "relevancy" in searching is concerned. Because we cannot know in advance exactly how anyone will produce any given bibliographic record, we can only establish an “authority record,” i.e., a mechanism for collecting every sort of result. Such records consist of titles, authors (individuals or groups), publishers, standard library control numbers and variety of topics needed to build the control table, such as time, dynasties, imperial reign titles, place names, languages, and other vocabulary. We by necessity utilize this set of authority records to improve the quality and maintain the accuracy of bibliographic records.

### 3.5 Constructing a new bibliographic control system based on literary authors:

The NCL began collecting and arranging literary materials in January 1984, when it organized the "Documentary Exhibition of Contemporary Women Writers." In March 1994, it established and developed the "System of Images and Full Text for Materials
on Contemporary Literary History," and it also began trying to manage "tacit knowledge" derived from literary authors. In regard to this, we propose the following:

Relationships among writers: writers' personal family background, growth, learning and life experiences all appear in or affect their writings to one degree or another. The principal relationships include time period; regional culture; life experience; blood relatives; discipleship and literary school membership; interpersonal association; and the influence of their writings.

Links among literary works themselves: in the collected works of any given major author, one can usually discover similarities and differences in his or her treatment of specific problems. The principal links among the work themselves are as follows: subject matter; chronological setting; plot evolution; role of interpretation; writing techniques; and works and derivative works or products.

The relationship between works and criticism: this relationship is actually between author and readers. It is even more a matter of multiple meaning and spatial relationships across generations. The principal relationships between works and criticism include: identification, critiques, questioning, supplementing, and textual research.

In addition, there are other things associated with literary authors: literary societies, literary awards, literary chronicles, financial aid for authors, literature forums, and laws and regulations on literature.12

4. The NCL’s system in actual practice:

4.1. Objectives in establishing the system:

The NCL is both the principal supplier of bibliographic copy for libraries and information organizations and the basic source of bibliographic materials for which the general public or researchers are searching. Given its dual task of collecting and supplying bibliographic items, it must come up with a new comprehensive plan to meet present and future requirements for the establishment, management, preservation, and utilization of its bibliographic resources. At the same time, it has to keep in cost-effectiveness and fulfilling the catalog’s potential. Based on the NCL’s internal situation, bibliographic service technology, and environmental changes, the establishment of its wide-area bibliographic management system has the following explicit objectives:

1) the merger of the databases established over the years by the NCL, so as to create a bibliographic system that contains every type of publication and then to gradually move towards a representative and integrated "National Bibliography."

2) the integration of the union catalog holdings of National Bibliographic Information Network’s (NBINet) 75 cooperative libraries, in order to make this bibliographic system a "National General Bibliography."

3) the integration of MARC and non-MARC bibliographic records; the establishment of one format (field) for both types of bibliographic records; and, given the trend
towards the FRBR framework for cataloging, the establishment of the link between bibliography and authority.

4) on the basis of the bibliographic quality control center’s many years of experience, we must verify bibliographic records through comparison, get rid of duplicates, consolidate, and upgrade, so that the different records for the same book will be integrated into a single most complete and richest bibliographic record.

5) in the initial stage of using the wide-area bibliographic management system as the basis for the union catalog, each system’s special features will still revert to the operation of that system. In the future, we will examine the server and reflect on whether or not it can integrate the features of all the systems.

6) the re-designing of workflow process, so as to avoid the duplication of a bibliographic or authority record for the same entity by different librarians in different systems. We will also establish consensus among librarians concerning the re-use of bibliographic records, and through the process of enhancing the content, we will add value to bibliographic and authority records more efficiently.

4.2. Re-designing the workflow in catalog preparation:

Analysis of catalog-building:

In her study of the feasibility of implementing the FRBR framework in NBINet, Professor Zhang Huizhu found that bibliographic quality is relevant to whether or not an ideal FRBR data association framework can be established. Problems such as gaps in content and different ways of recording data will affect the handling of the data relationships. 13 We want to integrate the data with different recording standards and formats into one entity, as well as to establish relationships among related bibliographic or authority records. These are the biggest challenges in building the database.

Generally speaking, the NCL has the resources to cover the bibliography and contents of each level, but only by using different formats and standards to build databases. The two main data formats are MARC and XML. Under MARC there are Chinese MARC (CMARC) and MARC21, and the definitions of the record fields in every XML database have their differences. In terms of record content, each database follows different standards for classification numbers, titles, subject terms, or keywords. For the most part, MARC records employ a national or international standard, while XML databases employ their own. Every database also differs as to the degree of control over the entries in a record. Some records in different databases, in fact, describe the same work or entity. Different departments within the NCL are responsible for these databases, which, moreover, were developed by different vendors. Therefore, database management, maintenance, and development are completely independent. This means that even in both the re-use of data inside the Library and moving out and shifting of that data, we have to go through the data conversion process at different levels.
Fusing the catalog and re-designing standards:

The plan for establishing the wide-area bibliographic management system not only involves the integration and fusion of the catalog. It must also provide for the resolution of the various standards of description or analysis followed by each database. Only then can there be re-designing from the sources of data in order to improve the efficiency of existing data integration and facilitate maintenance of, or addition of value to, follow-up data. Therefore, it is equally important to re-design standards and to fuse catalog content.

Whether a database is primary or derived, each system has its own special consideration at the time of its establishment. Catalog fusion emphasizes higher efficiency in the use of data, integration of human and financial resources, and improved system management. The disabling of an individual database might result from the phenomenon of natural selection during the fusion process. The retention or abolition of existing databases will depend on evaluation of their purposes and system functions. Of course, the decision to retain or abolish might take more than a short time. So in order to preserve the original special characteristics of each database, a large bibliography will have the function of managing the "inventory" of that database. In other words, after the fusion of the catalog, the original records from every database will still be retained in the system.

If a library builds a catalog by following consistent standards, it will be able to improve catalog management and information retrieval. Although the library and information science community has criticized the rigidity of traditional cataloging, and even though the subject analysis standards used by libraries have their shortcomings, it is still true that in the course of promoting an information correlation framework, libraries obtain a certain affirmation in assembling bibliographic records. Differences between independently established standards for using resources mean more than just labor-intensive maintenance for each standard. There is also another problem. Given the lack of correspondence between the various standards, it will be difficult—when searching for information—to attain efficiency in gathering related information through the subject analysis content in bibliographic records.

Transforming processes and reconstructing systems:

There are a number of possible reasons for a clustered database with different standards for record creation: incomplete coordination between plans for special projects and a library’s general direction of development; the lack of horizontal communication between departments; and performance of similar work by different departments. The core tasks of librarians are every type of resource management, organization, and service. Although these tasks vary, librarians cannot provide a final product to the user without inter-departmental coordination. In order for a library to cope with the many changes in the dynamic environment of information supply and demand, it must have inter-departmental cooperation, a more meticulous and rigorous planning process, and flexibility. Only then can it implement such a large-scale project.

When an organization has the same workforce, the need to cope with multi information requests is increasing and not abating. A single department is unable to
handle all business similar to the same nature and other departments must lend a helping hand. Due to the diverse expertise of staff in different departments, the criteria of building and circulating data vary. As a result, consistent workflow procedures must go beyond the department pattern to scrutinize every facet of process for want of special skills and manpower. Such a change in NCL has not yet been finalized, but preliminary plans to build resources management tasks are summarized as: system development and application, records of content description, subject classification standards, authority control, archival work, digital object management, reference services, advice to readers’ queries, advice to staff’s queries, promotion of educational training and management of activities, environments and so on. These will serve as a basis for follow-up workflow planning in detail.

4.3. System Structural Planning:

In order to satisfy diverse information requests by general users, researchers, and library staff, the flexibility of data import and export related to all types of information is our important need, required to build this system. Because the range of this plan covers a whole slew of data format types including MARC and XML, all information sources must be kept with their contents intact. A framework for multiple data formats must be developed so that follow-up data may easily be integrated and correlated. Furthermore, different user groups will be provided with the correct format of information at their request.

Multiple formats taking care of all sides:

As to the library community in Taiwan, CMARC and MARC21 (or its predecessor USMARC) have widely been used in machine-readable format. Some libraries have adopted either one for creating their bibliographic and authority records; others selected CMARC for Chinese bibliographic records whereas bibliographic records for Western languages are most often in the MARC21 format. The union catalog system of NCL is the main source of two kinds of MARC bibliographies. Therefore, two kinds of two-way comparison standard for MARC and its conversion process have become the basis of operations for many years. In fact, different MARC format conversions may have difficulty in corresponding to each other. In addition, CMARC has rarely been revised and the media reports on updated information about many revised MARC21 fields; they can hardly correspond to CMARC. On the other hand, CMARC codes for Chinese resources (such as Chinese musical instrument, musical form) or field (such as Chinese rare book and related fields) will not correspond to those of MARC21.

Since MARC format conversion is doubtful, XML database with dissimilar definitions of each field will make the process more complicated. As systems aside from MARC format database are often created for specific propriety databases, many fields in XML records will not talk to their counterparts of other databases. In order to compare, analyze follow-up data and establish relationship between information queries, MARC and non-MARC bibliographic output services and the forefront foundation works include: preparation of field comparison tables for various data formats, analysis of corresponding field results, evaluation of format types in the system internal storage, and modification of corresponding fields based on formats in the system internal storage. After a long and cumbersome process analysis between
Library staff and manufacturers, a final decision has been made by manufacturers to accept the expended MARC21 structure for the solution of library’s internal storage format. In order to facilitate follow-up notes, below is a concise description of the X-MARC Structure:

1) Based on the existing fields of MARC21, X-MARC has expanded its field structures, including all CMARC and other metadata, which can not be mapped into MARC21 fields for keeping all contents of the original records intact. The big challenge of this work is to consider the data fusion during expansion of new framework, taking into account the need to restore integrated data back to various MARC and XML formats.

2) Although MARC has 886 fields, they are only suitable for other MARC format records. The magnitude of future revision of MARC21 is hard to predict. Therefore, expansion of MARC21 fields does not include any field between Leader and 9XX. The X-MARC must go beyond the existing MARC21/ISO2709 format restrictions to retain the extant structure.

3) At present X-MARC has adopted a 4-digit field expansion. The segment of field numbers is roughly being divided as follows: fields 1000 – 1999 for relevant system control data; fields 2000 – 2999 for independent database and basically they contain fields to be decided what their proper usages will be; fields 3000 – 3999 for expansion and common use of the original design of 9 XX of MARC21; 4000 – 4999 for use of CMARC specific data.

4) When bibliographical or authority records are in the repeated comparison, the fusion process does not appear in the query result field and a negative field will be assigned. For example, certain record has been integrated to a duplicate record and the number 500 of repeated fields will not show in the reader query results and then a “-500” will be stored in the database for reference to library staff performing maintenance works. (Figure 3) This kind of design will be discussed in the next paragraph.
Elements of Data:

Library bibliographical records are from time to time being criticized too cumbersome to be fitted for practical use. Among thousands of MARC fields or sub-fields, part of human cost of the project time has not been computed or properly recorded for users’ inquiries. For example, coded data contain detailed information on specifications for audio-visual data carrier, color, sound and characteristics, but on the public bibliographic records only shows the name of “Data type marked” (GMD). The development of FRBR, FRAD, and RDA, however, will push the future bibliographies up to the level of granularity. So far as the cataloging work is concerned, which parts should use the library human resources? Martha M. Yee has pointed out the dilemma of cataloging in her article: “Data with less structure and less granularity could be easier and cheaper to apply and might have the potential to be adopted in a more standard fashion across all communities, but that data would limit the degree to which powerful indexing and sophisticated display would be possible.” If we look from the angle of data construction, granularity is indeed necessary. But the human resources and costs for overall refinement are beyond the library capabilities. As a result, designing cataloging granularity based on the core purpose of bibliography use will meet our present needs.

From our long experience of processing and using data of union catalogs, we found that lack of fine structure of bibliographical records in the beginning will constraint data screening, analysis, addition or re-use at a later stage. A saying goes “The path from simplicity to complexity is tortuous while the other way round straight.” Under the premise of integration of diverse data, a broad bibliographic database records are seeking principle of granularity and structure. Under our limited resources, this principle sets the basis for the establishment of related information elements. After initial analysis, NCL has various metadata records of different categories, but all of them are nothing more than the work itself and people who work around the work, time, places, objects, themes and so on. These things are a
concern to entities of FRBR. Each entity has different properties and attributes. Inter-
relationships among properties and attributes are formed a foundation of
bibliography or bibliographic family. Under limited manpower, to produce the inter-
relationship among records, we must proceed to analyze different metadata in the
component parts. Starting from the smallest element of data “points,” we connect
them to be a data string and then produce a meaningful knowledge face.

Modification of Authority Records and Experiments:

Another issue of component data processing is how to set the scope and practice of
data components. After many discussions, we found that library authority control is
essentially the data element management. Therefore, wide-area components of
bibliographical control system may refer to the principles of traditional authority
control. As far as the scope is concerned, individuals, groups, uniform titles, themes
are types of authority records and future planning should expanded to include
standard numbers, names of place, dynasties, imperial regal titles, all kinds of codes
(such as region, language code) so as to lay a foundation of data associations in
term of time and space. In practice, we should enhance the refinement of traditional
authority records and find out from existing data the properties of meaningful
contents to store them in structural format.

As components are much more, frameworks of past records are not completely
suitable for use of future systems, but authority records established under many
years of authority control may have already covered the basic information
components. Therefore, the modification of non-structural contents of MARC
authority records by adding other metadata elements is prevalent at present time.
For various component records refer to FRAD attributes. To establish the field of
component records, there is no instance of FRAD structure to follow and we hope
the system will lead the way toward the development of FRAD architecture. In order
to establish a practical foundation to find the necessary items for component records,
we have proceeded to do experiments with personal name authority records which
are numerous with intact contents. For this experiment, 100 Chinese personal
names with copious reference sources have first been selected. MARC authority
records bearing those names are being transferred to the systematic individual name
records which are manually modified by adding, deleting or editing contents. During
the experimental period, the framework of the name of the component records and
the feedbacks of the record workers are completed as initial steps. Like the
previously mentioned field expansion program of bibliographical records, component
records under the strict MARC21 authority format will be expanded using a 4-digit
field structure. (Figure 4).
Components Code and Association:

In the process of cataloging and authority control, text description for the same entity or attribute may have variations due to the cataloging policy, library authority record rules, and the languages for cataloging. If the uniqueness of physical components could not clearly be described by language, it would be very difficult to completely gather the contents of numerous amounts of data. In order to use these components for data comparison and related works, all components of the system will be categorized and numerically coded. The system will automatically assign a unique number for each component entity. These groups of numbers will be used for follow-up programs and related data processing. The protective action of descriptive language in component records will hardly affect the previously established association. For example, the description of certain name in a bibliographic record comes from the component number of that name which then elicits various complex properties related to that component record. Through the program’s operation process, the horizontal and vertical relationships between data and attributes are expected to reason out the relationship between data and information. (Figure 5). Therefore, besides the component codes being used to match and recognize the real entity, the most important thing is to improve the efficiency to the system front-end users’ query, identification, selection and guidance.

Backbone (Core) Architecture of Storage and Sub-storage:

The practical system structure design of broad bibliographical management involves very complex technology. Instead of minute descriptions, we will give in a nutshell the concept of the database structure design. There are three targets of designing overall system architecture -- to safeguard the original data, to integrate common data, and to link related data. The database structure is different from that of the general system; especially the goal is to integrate data. To safeguard raw data under the same system architecture is superfluous. Except inventories the main purpose is archiving various types of records, and recording contents of assessment reference.
After the analysis of all types of data fields, initially fields fall into three categories: 1. Objective description of physical fields (such as name of work, author, etc.) 2. Subjective analysis of entity or value-added field (such as keyword, class number, summary, etc.) 3. Management information fields (such as maintenance personnel, date, etc.) The structure of data storage is divided into backbone (core) storage and sub-storage. The former stores integrated backbone records, including retained fields of the first category which have been repeatedly compared and backbone fields of the third category; the latter stores original sources of category 2 and the negative field numbers for all records which have not been admitted in the fields of backbone records. For example, if new records are duplicates of certain records in the database after comparison, they will be calculated and integrated. The field of new records (MARC21 field 500) which has not been included in the backbone records will be stored in the sub-bibliographic database with a “-500.”

Although the record field storage is divided into “backbone record” and “sub-storage record,” the method of display varies according to different users and their purposes. Library staff maintenance interface according to different purposes provides service: 1. according to the order of field number; 2. according to backbone concentration; 3. according to three sources of independent display. Library staff may switch the operations as need arises. User query interface also provides two different ways of display -- general and authorized use. The results of general search only display the backbone of fused bibliographical contents. After authorized users log on their accounts, they may be able to see the backbone of fused bibliography and original records before integration.

4.4. Data processing and Building characteristics:

Conversion of various XML data to a unique format (X-MARC) is to open a door convenient to the subsequent data management. Because of the system with a huge amount of data and diverse formats and each process not affecting the overall framework, it is necessary to set up parameter mechanism dealing with different information that needs adjustments.

Data Check:

In order to take into account follow-up data processing and output, all kinds of data must go through various level of evaluation to ensure the usefulness of records and their fields in the database. Data check can be divided into data format and contents. The system of MARC format data will be evaluated according to MARC21 and principles of cooperative cataloging by library communities in Taiwan. As to the XML records with inconsistent field definitions, even though they have been converted to MARC-based records prior to evaluation, they are hardly suitable for use under the same evaluation standard for MARC format records. But they could only be used to evaluate whether or not XML format records have core fields.

Data Classification and Rating:

In the face of continued growth in numbers of a large database with various levels of complexity in contents, the entity of data integration must be handled in each individual process. At proper time flag each record based on the degree of
completeness of contents and specific forms for data matching, refinement, and service operation. System parameters are currently configured to provide various types of information classification and scoring rules. Classifications are mainly based on records that indicate the level of completeness of first- or second-priority fields. However, during the process of integration of MARC and non-MARC bibliographic records, the amount of fields in bibliographic records could not completely reflect the importance of various elements of bibliographic records. Some non-essential but necessary description for contents or specific nature of works may have been ignored during the evaluation of completeness of fields. Thus after analysis of record contents from various sources, the system accounting will give non-essential fields which provide important information a weighted score of different ways to highlight the important messages of bibliographic records. Since the grade or score is the result of program operations, the generated numerals should not be construed as absolute criteria to judge the merits of bibliographies. This practice only provides a basis of reference to the management or filter of various data at a later date.

5. Barrier Breakthrough and Innovation of Wide-range Bibliographic Management System:

The wide-range bibliographic management system established by NCL is the results of many years' experience through hard work trying to construct the best bibliographic system dealing simultaneously with traditional and digital media. We believe we have broken through barriers and created innovation in cataloging in the following five items.

5.1. A Complete Huge National Bibliography:

The purpose of the development of a bibliographic system should not only limit itself to the management (control) of every single piece of all known works, but also discover more single pieces of “existing works not under control” or “existing works which should not be overlooked.”

An ideal comprehensive national bibliography = controlled bibliography + bibliography to be controlled.

Controlled bibliography is clearly known to everyone and not much should be said. The major sources of bibliography to be controlled come from bibliographic citations, references, known literature, book reviews and so on. The bibliography of single piece of main works being processed may cast light on unpublished, less circulated works and works not intended to be made public and lost works, etc. In so doing, the national bibliography will become nearly comprehensive.

5.2. The Comparison of Intelligent Data:

During the process of union catalogs, we have accumulated experiences to find a considerable number of duplicate bibliographical records. But the scope of this project for data expansion comes from record format, data sources of different principles. The design of matching program must take into account various data formats and special features of contents to produce more flexible matching mechanism. Initially, workers tend to take their jobs seriously. But the difference from
the past is the expectation of the matching program to provide the features for learning and memory. The process of manpower carried out logical operations to sum up similar situations as the basis for people to follow in future. The final goal is to achieve the result of interaction among manual operations in terms of automatic matching through programs. To improve the needed vocabularies and rules for data matching will raise efficiency for the job and gradually reduce the degree of human intervention.

5.3. Interrelationships of Bibliographic Data:

Interrelationships of data should be considered as the biggest challenge facing us. The difficulty of program design is only secondary. The biggest problem is that when records are first created for each database, catalogers' attention is being directed to single piece description and they leave no clues to the entities and relationships between them. Under the very limited conditions, programming could not directly establish interrelationships of data. Such work must be done indirectly by analyzing various entities of authority and bibliographic records. Through relationships of various entities and bibliographies in addition to human recognition and assistance, we may gradually establish a foundation for the association of bibliographic families and entities. In the long run, a design suitable for NCL’s information system framework should be to rethink the policies of cataloging operation planning, guidelines, workflow, special professional training and other aspects to consolidate the associated information infrastructure.

5.4. Sustained Growth of Bibliographic Records:

In the past the preparation of bibliography was conducted in such a way that the job be completed at the first try. A bibliography too simple and crude is not a good bibliography. Basic bibliographic data for the national bibliographic records have been proposed for discussion of requirements. However, because of continued exponential growth of information resources, the compilation of bibliographies should not be limited to a few people and everyone has a chance to make a contribution, especially experience of a single piece of works, evaluation, and added values and so on. As a result, bibliographies should not be the achievements of a single person, but are the efforts of entire staff members combined together. Bibliographies are no longer to be completed at one try, but are the results of cumulated values added constantly through continuous collective contributions of entire staffers.

5.5. Dealing with the Era of Multiple Bibliographies:

Publishers, libraries, scholars and experts can prepare bibliographies. Besides basic bibliographic data, bibliographic contents may include many subject analyses, notes, textual research, and others. Thus a single piece of work from different sources has become a common phenomenon. Bibliographical system should also be led toward the direction of integration of different sources of bibliographies for comparison, consolidation of operation so that bibliographies will become more comprehensive. To jump over the single-bibliography hurdle, multiple bibliographies as the core practices for our future bibliographic system will break through the dilemma facing us. The preparation of bibliographies may also resort to mathematics.
6. Conclusion:

Library catalog by subject cataloging and the function of indexing system has been successfully based on users’ search terms displaying consistent clustering bibliographic query word list. In recent years system vendors have introduced multi-faceted classification functions to the library system. Bibliographic queries have been enhanced after the introduction of language, regions, years and other services to filter query results. The users could accurately locate the required information. Otherwise, the query results which have come from either clustering or multi-faceted classification will be confined to the same plane switching ranges of the various query results. If in the future we gradually establish associations of various entities and we do expect that bibliographic query service will guide users in the three-dimensional presentation accessing to all levels of information needs.

Today, search engines are fast and large and nothing is lost in the search results. Readers are immediately satisfied with their timely needs and other features. [In the past], readers faced complex interface in terms of library collection systems, inconvenience of maneuver, cumbersome operational procedures and they did not know how to take advantage of those library services. Whenever readers could not help using library system to search information, they would produce many expectations which would be increasingly cumulated. Perhaps when readers gradually walk away from the bibliographic system, it would be the time for the library to change the status quo beyond the previous time.

This is the second phase of system development. Based on the traditional norms, the new procedures for managing library data and services include: (1) retention of the existing data in full and integration of record contents of MARC and non-MARC to facilitate the extension of usefulness in future; (2) extension of MARC structural design, expansion of MARC fields so that under the same record structure, the valued contents of non-MARC fields will be fused; (3) analysis of detailed bibliographic data on components, establishment of independent records of individual component entities through the identification of individual components and contents of records, establishment of the association between various entities; (4) analysis of special features of each record field, retention of important descriptive record contents through scoring fusion process.

In 2002 National Science Council (NSC) established the Office of Digital Archives Program (NDAP). The development of digital resources has officially been accepted as priority items of government policies. NCL was the first institution to plan digital archives. During the past 10 years we have been committing ourselves to transposing our library collections to digital archives. More than 40 millions digital files and diverse collections of digital resources have accelerated numerous digital publications. As a result, the new improvements have greatly influenced the structures of NCL’s collections and readers’ behaviors of using library materials. Therefore, under the environment of management and control of multi bibliographies, collection of digital resources, management, integration, dissemination and readers’ services, the use of new framework of services and concepts of information expedite adaptation and transformation to meet the challenge of mega media waves of the library. Our library community’s reasonable solution plan in the face of network and non-network digital documents, digital publications, digital newsletters and
commentaries, and reading habits, the access of information and so on will challenge those rapid changes in current trends.

In 2010 while visiting Taipei, Professor Paula Kaufman, librarian, the University of Illinois at Urbana-Champaign read our statement of the concept and prototype of this system. She specifically said: “Suppose we get into a physical or virtual library, regardless of form or location of the data collection, we use the same system only to find the contents needed. Our long-time dream of a “one-stop shopping” service has been almost achieved.”

We firmly believe this new consolidated function with a comprehensive library bibliographic system will be able to solve or abate the current library management difficulties and will also provide our patron friends with a new-era library 2.0 service.

**A Word of Thanks:**

NCL’s incumbent curator Zeng Shuxian has continued to promote the project of management system of comprehensive bibliographies, a plan originally initiated by the 12th curator Gu Min. Under his leadership, former curator Gu Min with all staff members began to explore the theory of management of comprehensive bibliographies. Wu Meiying, the deputy curator, Wang Peiyiing, chief of Information Office, Mr. Chen Liyuan and personnel of various units jointly performed system development. Wang Zhenan, general manager of Tudor Information Company, Ltd. with all his team members shouldered the responsibility for the development of information technology. These people were the best contributors to presenting the comprehensive bibliographic system in front of every one of you.