# The Digitalization and Resource Sharing of Archives in Taiwan

*Li-Kuei Hsueh; Graduate Institute of Library, Information and Archival Studies, National Chengchi University; Taipei, Taiwan; I-Mei Hung; Institute of History & Philology, Academia Sinica; Taipei, Taiwan; Li-Chiao Wang; National Cheng Chi University Libraries; Taipei, Taiwan; Chun-Ya Wen; Institute of History & Philology, Academia Sinica; Taipei, Taiwan; I-Chia Chang; Institute of Information Science, Academia Sinica; Taipei, Taiwan*

## Abstract

*The National Digital Archives Program, directed by the National Science Council, includes the following participating institutions which hold archives: National Palace Museum, Academia Historica, Taiwan Historica, Institute of History & Philology, Academia Sinica, Institute of Modern History, Academia Sinica, Institute of Taiwan History, Academia Sinica, National Taiwan University, Aletheia University, etc. Under the umbrella of the overall plan, these institutions work separately but share results; the main purpose of the program is "Establishing the special nature of archive contents and developing the unity of digital resource sharing." This paper mainly discusses the current status of the archive digitalization each above mentioned institution is carrying on, including the extent of digitalized archives of each institution, method of digitalization, standards for archive description, and online indexing method, etc. Also, this paper will discuss the development of archive resource sharing in Taiwan, including: metadata, name authority control, as well as construction of the union catalog.*

## Keywords

*Archives; Digitalization; Resources sharing; Taiwan*

## I. Introduction

The National Digital Archives Program (NDAP) being carried on by the Republic of China Executive Yuan National Science Council (abbreviated) has its origins in shared concern felt by both the government and the public about digitalized resource archives in the information age. The "National Digital Archives Plan" was approved by the Executive Yuan and established in July of 2000. From January 1, 2003, the National Science Council reconfigured the National Digital Archives Plan into a national effort, renaming it NDAP and formally beginning various tasks related to digital archiving of valuable resources in the Taiwan region [1].

Important archival institutions in Taiwan are: National Palace Museum, Academia Historica, Taiwan Historica, Institute of History & Philology, Academia Sinica, Institute of Modern History, Academia Sinica, Institute of Taiwan History, Academia Sinica, National Taiwan University, National Tsing-Hua University, Taipei National University of the Arts, Aletheia University, Diwan College of Management, etc. Archives undergoing digitalization in the institutions include: Ming-Qing Archives, China Archives in Modern History, Tan-Hsin-Tang-An, government documents and archives, the Taiwan Soutokufu Archives, Japanese Occupation and Postwar Archives, the Taiwan Provincial Governor's Office Archives, Taiwanese old title deeds, photographs, and a number of private archives, etc. In characterizing the extent of the digitalized archives, the time scope of the archives spans the hundreds of years that have passed from the Ming and Qing dynasties, the ROC period, the Japanese occupation to present day. The archives include economic, societal, political, figure, etc. They can be said to be all-inclusive.

The task of archive digitalization is a time-consuming, labor-intensive, and cost-intensive process, especially when the fact that the archives are unique and irreplaceable is taken into consideration. The model for archive digitalization in Taiwan is one in which all funding is controlled at the national level. The National Science Council, a government agency, organizes and manages the funds, in a top-down model. The advantage of this method is that, due to the inclusion of the program in the government budget, funding is comparatively adequate; add to this the presence of a specialist assigned to organization and management, quarterly, semi-annual, and annual evaluations, as well as the technical support and promotional activities needed by specialist teams in the digitalization process, the task of archive digitalization is made more complete.

The national archive plan is a cross-field digital resource creation and sharing plan. In the execution of archive resource digitalization, the model of "separate construction and resources sharing" model is used:

### 1. Separate construction

The NDAP sets institutions as operating units; each archiving institution separately creates its own digital databases for its archives, following international standards organized by the archives group. The metadata, cataloging standards, digital creation standards system structure, etc. reflecting the distinctiveness of each institution's archives are then developed.

### 2. Resources sharing

NDAP provides support in programs. Support programs highly relevant to digital resources construction, besides management, maintenance, education, and promotion, are the content and technical divisions. The Content Development Divisions supports the various units in cooperation and sharing in the process of execution. Specialized groups are established for each field, such as the archives group. This group plans the direction of development for archive digitalization, while cooperating and sharing relevant experiences and results. The most direct developmental goal is the establishment of a Taiwan area archival unified retrieval system, which is based on a "cross-institution, half-integrated" method. This system would provide a unified interface for users to retrieve information from

relevant databases. Through the interface, users can connect to the separate databases and gain comprehensive archival information. Aside from this, the archive group also plays the role of coordinator in resource organization between the various institutions. The group takes the digital archive resources of the various institutions, organizes them according to a plan and standards established by the group (such as the archival core elements XML format), then uses the union archives finding aid to enter the information into the cross-field combined catalog developed by the national plan. The special nature and diverse cross-field sharing nature of the archives is thus emphasized. The Research and Development of Technology division provides support for the necessary technical development during this process.

Using the "separate construction, and resources sharing" model described above, the Taiwan region digital archiving program has shown substantial results since its beginning in 2002. This paper will primarily discuss the current situation of digital archiving content and digitalization construction in various institutions of Taiwan; this paper will also analyze the development of archive resource sharing mechanisms in the Taiwan region, and will finally make suggestions for archive resource sharing in Taiwan.

## II.  Situation of Archival Digitalization in Taiwan

In Taiwan the digital archives can be divided into three different categories based on their time of origin: 1. Min-Qing era archives, 2. Japanese occupation era archives, and 3. Republic of China period archives. With regard to archives, digitalization, categories, and construction conditions, a summary follows.

### 1.  Ming-Qing Era Archives

The two main institutions holding Ming-Qing archives are the National Palace Museum and the Institute of History & Philology, Academia Sinica. Archives held in the National Palace Museum include the Grand Council Archives. The Institute of History & Philology, Academia Sinica holds the Grand Secretariat Archives. Also, the National Taiwan University Library holds the Tan-Hsin-Tang-An and the manuscripts of Ino Kanori. Aletheia University holds the personal files of Dr. Mackay. All of these are archives dating from the Ming-Qing era; their digital archiving situation is as follows:

### (1)  National Palace Museum

Qing-era central archives held by the National Palace Museum include files from the Qing inner court, manuscript collections and historical documents collected by the Qing history institute. 204 boxes of archives were brought to Taiwan by the Nationalist government in 1949, comprising roughly 400,000 manuscripts. These, and those from other sources include: Palace Memorials, the Grand Secretariat Archives (including national Credentials, imperial edicts, and imperial ordinances, etc.), copies of memorials collected by the Grand Council, archives of the Grand Council (such as record books of edicts, position papers, monthly memorials, court letters archives, record books of military

campaigns against riots, record books of telegraphed edicts, etc.), and archives of the Historiography Bureau, primarily files from the Qing State History Office and the Qing Historiography Bureau, including imperial annals, gazetteers, charts, draft biographies, packets of biographical materials and general chronological biography archives; a number of documents have also been donated by various parties in recent years.

Out of these, digitalization work on archives from the Grand Council began in 1997; a total of 190,837 manuscripts are to be digitalized. In the process of digital archiving, in order to retain the original appearances of the files, image joining is used. In cataloging, the original information on registration cards of the files are used, including manuscript number, submitting officer, official position, incident, time, imperial comment, and box number (7 items total). Currently the institution has joined NDAP as well as other archiving organizations, together researching and carrying on the work of archive digitalization.

### (2)  Institute of History & Philology, Academia Sinica

Based on a 1952 estimate, the Grand Secretariat Archives held by the Institute of History & Philology, Academia Sinica are composed of about 310,000 manuscripts, over 4000 of which date from the Ming period; most of these were collected by the emperor Kangxi in his efforts to record Ming history. Qing dynasty documents make up the majority of the manuscripts held by the Institute; they include (1) various files formerly stored by the Grand Secretariat, including edicts, court letters, source materials collected for compiling historical writings, etc., (2) various files of the Secretariat itself, (3) personal documents, (4) test forms of civil service examinations and related files, and (5) Shenyang archives[2].

Since digitalization work began in 1996 (initially using black/white and grayscale scanning, switching to image files saved on disks in 2003) [3], over 200,000 documents had been digitally scanned as of 2005. With the construction of metadata, around 170,000 documents are now available in their fullness for use online.

### (3)  National Taiwan University Library

The National Taiwan University Library holds the Tan-Hsin-Tang-An, the Ino Kanori manuscripts, An-Li-Da-Sha documents, as well as old photographs of Taiwan, among other valuable archives. Out of these, the Tan-Hsin-Tang-An are administrative and judicial files dating from 1776-1895 in the Tanshui Ting, Taipei Fu, and Hsinchu County. These documents are invaluable in studying the administration, judiciary, economics, society, and agriculture, etc. of Qing-era Taiwan; these are also important historical sources in studying traditional Chinese law and judicial rulings. These are also comparatively well-preserved traditional Chinese County-level documents.

### (4)  Aletheia University

The Dr. Mackay archives digitalization plan pursued by Aletheia University allows researchers to study the history of the Presbyterian Church in northern Taiwan; the Northern Presbyterian Church is a part of the larger family of the World Council of

Churches and the Reformed Churches. Its historical origins can be traced back to two preachers: English Presbyterian Church preacher Dr. Maxwell in southern Taiwan in 1865 and Canadian Presbyterian Church preacher Dr. Mackay in northern Taiwan in 1872[4]. The contents of the plan include: "30 diaries kept by Mackay" (with around 10,000 sets of information, this plan provides the English electronic text format and Chinese translation records), "correspondence between Mackay and other countries and friends," "notes kept by Mackay during his education," "notes kept by Mackay during his geographic, geologic, and botanical studies in Taiwan," "Pinpu and Fulau vocabulary recorded by Mackay," "records related to observations of the Pinpu," as well as "over 400 photographs related to Mackay's work" kept by the historical institute of Aletheia University, among other important files [5].

### 2. Japanese occupation-era archives

The primary institution that holds Japanese occupation-era archives is the Taiwan Historica. Taiwan Soutufuku Archives, divided by year of effectiveness, include permanent public records, 15-year public records, the public record overall catalog of the above records, a record of correspondences, categorization catalogs, as well as Taiwan Soutufuku special permanent documents. Documents range from 1895 to 1945[6].

The Taiwan Soutufuku archive digitalization project joined the Digital Archive Plan in 2002, beginning the "Digital Japanese Ruling and Postwar Archive Project." Important archives are the "Taiwan Soutufuku archives" and the "Taiwan Soutokufu Senbaikyoku Archives;" the archives include blueprints for the governor's mansion, Taipei's Chung-Shan Bridge, administrative region diagrams, mining region maps. There are over 20,000 map archives approximating A3 size in total [7].

### 3. ROC-era archives

ROC-era archives refer to files produced by various government institutions after the establishment of the Republic of China and subsequent reestablishment of operations in Taiwan. Primary institutions holding archives are the Academia Historica and the Institute of Modern History, Academia Sinica. Digitalized content are described below:

#### (1) Academia Historica

The "Digital Archives of Government Documents and Presidential Records Plan," contain the following five main categories of archives:

a.Democratic government archives: these primarily contain primary records of important national decisions and directives, beginning from July of 1925 until about June of 1949. There are 7,086 compilations, totaling 1,134,746 pages.

b.President Chiang Kai-Shek documents: these primarily contain: writings, revolutionary writings, Chiang family genealogical records, leader family records, photographic records, literature and books, special documents, special telegrams, special files and 10 other separate categories. The digital archiving plan involve special documents, including personally approved documents and drafts, 72 compilations in total with 6,716

manuscripts; special files, including categorized information and normal information with 1,077 compilations, 73,609 manuscripts in total; special telegrams, with 38 cases, 436 compilations, and 141,756 manuscripts in total; photographic records, including photographs and photographic film, with 196 compilations, 50,798 photographs, and 223,844 pieces of film.

c.Resource Council archives: these archives are primarily divided into the two categories of Chinese-language files and foreign-language files. Chinese-language files are composed of files from both the Council itself and those of organizations in other countries. Foreign-language files are those documents pertaining to relations between the Council and its subsidiaries with American organizations or other foreign entities, including correspondence, telegrams, and various files; Chinese, English, and Japanese files are all archived. Archives start from 1939 and end in 1952. There are 20,840 Chinese-language compilations and 8,243 foreign-language compilations, 29,083 compilations in total; 4,322,850 pages, and 3,875 maps.

d.Taiwan Provincial Government Department of Land: these archives can be divided into: laws and regulations, land reports, meeting records, human organization, receiving and processing of daily production, land reform information, redistricting of agricultural fields, the abolishment and use of national forests, regulations for use of non-urban land, deed organization, use and disuse of agricultural land, etc.; these are important for the study of Taiwanese land reform. They range from 1944 to 1992 and 2,404 compilations total, with 360,600 pages.

e.President Chiang Ching-Kuo documents: the documents archived by this institution include: the loyal industriousness archives, government and army files, records of meetings with foreign emissaries, compilations of his words and actions, personal notes, speech transcripts, photographs and writings, as well as audiovisual sources, etc. There are over 2,000 Chinese-language file compilations, over 280,000 pages.

#### (2) Institute of Modern History, Academia Sinica

The archives held by the Institute of Modern History, Academia Sinica (hereafter referred to as the Institute of Modern History) can be divided into the two categories of economic and diplomatic. The diplomatic archives include records from 1860 to 1928, and are the oldest and of the greatest historical value among the archives held by the Institute. The economic department archives range from the 1950s to the 1970s, and are highly important files in studying the postwar economic development of Taiwan [8].

The completed portion of the "Postwar Taiwan Economic Development Archives Digitalization Plan" includes archives from the Executive Yuan Economic Stabilization Council, the Executive Yuan American Aid Utilization Council, the Taiwan Region Production Affairs Management Council, International Trade Consideration Council and other institutions, as well as the "Ministry of Industry, Ministry of Economics Business Files Digital Archives." The "Modern Important Economic and Diplomatic Files Digital Archive," the "Mr. Kuo-Ding Lee Multimedia Database, " and the "Taiwan Soutufuku, Senbaikyoku File Digital Archives" are still in the works[9][10].

### (3) Institute of Taiwan History, Academia Sinica

The Institute of Taiwan History, Academia Sinica did not join the NDAP until 2004, and primarily deals with historical sources related to the study of Taiwanese history; it previously cooperated with the Taiwan Provincial Consultative Council in archive digitalization. The two primary parts of the archives are:

a. Government historical sources held by the Taiwan Provincial Consultative Council (from 1946 to 1998): this can be further divided into provincial council files from 1946 to 1951, special session provincial council files and provincial council files from 1952 to 1998; provincial council reports and records from 1946 to 1998; videotapes of councilmen in session. All are valuable resources in studying the development of Taiwanese democratic government.

b. Archives held by the Institute of Taiwan History, Academia Sinica: personal writings, Japanese occupation-era Taiwanese studies of ancient literature and local history, Japanese occupation court documents, etc. The documents range from the Qing dynasty, Japanese occupation, and the postwar period. Organization and construction of digital resources faces presents significant challenges. The primary goal of development is organizing and developing a foundation for a historical source database, to increase the ease of use of historical documents [11].

## III. Analysis of Archival Digitalization in Taiwan

Substantial results and research have been produced as a result of the digital archiving work performed by the various institutions in Taiwan, from the individual efforts beginning in the 1990s to the NDAP project organized and promoted by the National Science Council in 2002. With regard to the metadata standards, digital image format, authority controls, archive indexing system, etc. of each institution's efforts, we analyze the construction of digital archiving in Taiwan; they are displayed in Table 1 and are described below:

### 1. Metadata

Metadata is the basis for construction of archive information databases, and is an important element in presentation of digital archive content. In performing the work of digitalization, every institute in Taiwan views it as a critical task and actively views and conforms to related international standards in order to facilitate integration of the archives into the international level. In construction of metadata, the majority of institutions refer to the two standards of Encoded Archival Description (EAD) and MARC21, while some chose Dublin Core (DC). The choice mainly depends on the degree of expertise in using the documents. The institutions that deepen archive contents choose EAD and MARC21 in order to facilitate professional archiving and academic use; institutions that emphasize accessibility and usability choose DC, in order to accelerate data registration and mutual accessibility. Besides following international standards, each archiving institute takes into account its own qualities and those of its archives to create metadata structure appropriate for use; this is done while maintaining the flexibility needed for integration into the international level.

### 2. Digital Image Formats

Before the execution of NDAP, many archiving institutions were already investing resources in digital image creation. The primary aim was to preserve the original appearance of the archived files, avoiding damage to cultural heritage resulting from time and storage environment. The other aim was to apply technology to facilitate ease and openness of use. But as a result of quickly improving technology, many early-adopting institutions have produced digital images of differing formats. But in the interest of protecting original files, the institutions have not reprocessed completed files, instead focusing their attentions on creating digital images for unprocessed archives. Studies found that, because of the example set by the NDAP technical development multi-level plan, all archiving institutions make use of three levels of digital image formats: archive level, business level, and browsing level, primarily focusing on purposes of archiving, business cooperation value-adding, and use in open public information. In using current technology, most institutions use the highest quality of scanning, subsequently converting files into different levels of image quality. In terms of formats, institutions of RGB 24-bit color, Tiff and JPG formats, as well as the JTiff format developed for special browsing methods. Resolutions are generally between 72 to 300 dpi. In regulation of formats, institutions were given flexibility in taking into account the circumstances, special qualities, and demands on the archives. Given the diverse types of file formats, this is an appropriate method for digital archiving.

### 3. Authority Controls and Authority Files

Authority controls were a concept introduced in planning construction of databases to aid in deep analysis and to provide effective indexing. Most archiving institutions chose to develop "name authority files," "geographic authority files," and "organization alias authority files." The ISAAR (CPF) standards created by ICA were universally consulted, and the entries being constructed primarily centered on the contents of each institution's own archives. There were two methods in development of authority files, the first being independent construction (as in the Institute of Modern History) and the second being cooperative construction (as in the National Palace Museum and the Institute of History and Philology). But whether independent or cooperative construction, there was adequate discussion of both authority entries and cataloging rules. In cooperative construction, similar archives (in time and archives type) were used as the optimal foundation, becoming a good example for the name authority databases planned for construction under NDAP.

### 4. Finding aids

Archives indexing systems are the primary tool for file use, and are also the primary means for responding to "purpose of use" in the process of digitalization. Various archiving institutions use metadata as a foundation, setting needs based on such, organizing function design, and developing and constructing the archives finding aids. There are three primary types of development in the various institutions: independent development, outsourcing development, and cooperative development. The technical abilities

of each archiving institution were the primary principle in weighing and selection. In the four years since its advent, the results of the plans have gradually materialized, and various archiving institutions have begun opening digital archiving results based on the file indexing systems. There are both catalogs and full text image systems. The majority of catalogs are free, while full text imaging systems are inconsistent in this aspect; there are inevitably mechanisms for restrictions of use. These archives indexing systems all represent vast archives databases. In facing the needs for file use of the external world, the method of organizing and presenting these databases for use will is a goal for the work of archive digitalization in Taiwan.

**Table 1-1. Construction Situations of Digital Archiving Institutions**

| Institutions | Academia Historica | Taiwan Historica |
|---|---|---|
| Metadata Standards | EAD、MARC21 | EAD、MARC21 |
| Digital Image Formats | ▫Three-tiered formats：archive level/business level/browsing level<br>▫Digital Formats：Tiff, Jpg<br>▫Resolution：72/150/200-300 dpi<br>▫Color Mode：RGB<br>▫Color bits：24bits<br>▫Compression Quality：Uncompressed/75%<br>▫Construction Method：Direct scanning/degrading conversion<br>▫Average File Size：53kb/260kb/12-23mb | ▫Three-tiered formats：archive level/business level/browsing level<br>▫Digital Formats：Tiff, Jpg<br>▫Resolution：300/100dpi<br>▫Color Mode：RGB<br>▫Color Bits：24bits<br>▫Compression Quality: Uncompressed/90%/80%<br>▫Construction method: direct scanning/ degrading conversion<br>▫Average file size: under 1mb/10-50mb/100-700mb |
| Authority Files | Name authority files | Japanese-occupation era name authority files Japanese-occupation era and early post-war period institution authority files |
| Finding Aids | "Democratic | "Taiwan Soutufuku |

| Institutions | Academia Historica | Taiwan Historica |
|---|---|---|
| | Government Archives:" indexing "Chiang Kai-Shek Documents and Photographs:" Index column thematic usage pulldown menus. | Document Retrieval System:" |

**Table 1-2. Construction Situations of Digital Archiving Institutions**

| Institutions | Institute of Modern History, Academia Sinica | Institute of History and Philology, Academia Sinica |
|---|---|---|
| Metadata Standards | EAD | EAD |
| Digital Image Formats | ▫Three-tiered formats: archive level/business level/browsing level<br>▫Digital Formats: Tiff, Jpg<br>▫Resolution: 200/300dpi<br>▫Color Mode: RGB<br>▫Color Bits: 24bits<br>▫Compression Quality: Uncompressed/24%<br>▫Construction method: direct scanning/degrading conversion<br>▫Average file size: 200kb/1.5mb/35mb | ▫Three-tiered formats: archive level/business level/browsing level<br>▫Digital formats: Tiff, Jpg, Jtif<br>▫Resolution: 72/150/300dpi<br>▫Color Mode: RGB<br>▫Color Bits: 24bits<br>▫Compression Quality: Uncompressed/86%<br>▫Construction method: direct scanning/degrading conversion<br>▫Average file size: 100kb/1mb/3mb/35mb |
| Authority Files | Name authority controls Location name authority controls Authority alias authority controls | Name authority files Location name authority files |
| Finding Aids | "Archive Catalog | Free online catalog use |

| Institutions | Institute of Modern History, Academia Sinica | Institute of History and Philology, Academia Sinica |
|---|---|---|
| | Indexing system:"The retrieval system possesses both browsing and indexing capabilities. | Pay-to-view full-text image usage. |

**Table 1-3. Construction Situations of Digital Archiving Institutions**

| Institutions | Institute of Taiwan History, Academia Sinica | National Palace Museum | National Taiwan University Library |
|---|---|---|---|
| Metadata Standards | Dublin Core | EAD | Dublin Core |
| Digital Image Formats | • Three-tiered formats: archive level/business level/browsing level • Digital formats: Tiff, Jpg • Resolution: 72/300dpi • Color mode: RGB • Color bits: 24bits • Compression Quality: Uncompressed/72% • Construction method: direct scanning/degrading conversion • Average file size: 15kb/7mb/40mb | • Three-tiered formats: archive level/business level/browsing level • Digital formats: Tiff, Jpg • Resolution 200/350dpi • Color Mode: RGB • Color bits: 24bits • Compression Quality: Uncompressed/30‰ • Construction method: direct scanning • Average file size: 5/30MB | • Three-tiered formats: archive level/business level/browsing level • Digital formats: Tiff, Jpg • Resolution: 150/300dpi • Color mode: RGB • Color bits: 8bits • Compression Quality: Uncompressed/85%/75% • Construction method: direct scanning/degrading conversion • Average file size: 520kb/11mb/63mb |
| Authority Files | | Qing-era name authority files | |
| Finding Aids | Free online catalog use Full-text image use subject to approval | "Military affairs document compilations," Qing-dynasty document archive overall index," and "Qing-era archive name authority data" all make use of three retrieval methods: browsing index, normal index, simple index | "Tan-Hsin-Tang-An," "manuscripts of Ino Kanori," "ancient Taiwanese tablets:" searching in the columns of title, theme, keywords, author, names, location names, year, media form, and language,etc. |

## IV. Development of Archival Resources Sharing Mechanisms in Taiwan

Based on the archive digitalization work being carried out by various institutions across Taiwan, the "separate construction" aspect of their efforts can be known. In the NDAP's "separate construction and resources sharing" model, with regard to the Taiwan region archive resource digital sharing mechanisms, archive core metadata standards, organization and development of authority controls, archive unified indexing system, union archives finding aids construction, etc. can be used to characterize the development of "sharing."

### 1. Construction of Core Elements in Archives Metadata

With regard to the basic task of digital archiving – metadata construction, out of the Academia Historica, Taiwan Historica,

National Palace Museum, Institute of Modern History, Academia Sinica, Institute of History & Philology, Academia Sinica, National Taiwan University library, and other important archiving institutes, the first five institutions follow EAD international standards, while only the National Taiwan University Library and Institute of Taiwan History, Academia Sinica use DC. From this can be known that, in the digital archiving plan sponsored by the National Science Council, the primary domestic archiving institutions have reached a consensus on following international standards. Also, the NDAP archives group gathered the various archiving institutions starting from 2004 and opened conference discussions; based on the EAD structure, a shared digital archive core metadata standard plan was to be developed, including core archive position, XML DTD and file cataloging principles. Archive description standardization is one of the basic tasks in archive management. Through a shared cataloging format, communication and reading errors can be reduced when there is information exchange between different institutions. The establishment of shared description principles can help in establishing consistency in archive catalogs, protecting and raising the quality of archive description, accelerating archive digitalization, as well as providing modernized retrieval methods and services [12]; such is also of great help to future members' development of metadata.

In discussing "archive core elements," the archives group collected the metadata structure of six plan members; the information was given to the technical development group for comparison. 19 core elements were decided upon in conference, five of which were set as necessary elements: "archiving unit," "building number," "description level," "title," and "time." The deciding principle for necessary elements was the simultaneous fulfillment of: 1. having at least four of the above institutions agreeing to join the core elements group, and 2. following the suggestions of EAD (Encoded Archival Description)(Library of Congress, 2005) and ISAD (G) (General International Standard Archival Description)[13]. At the same time in order to take into account the requirements of later joining the union catalog, the metadata working group made correlations between the 19 core elements and the DC elements, and drafted exported XML document examples and the DTD format for the reference of archive group members.

## 2. Organization and Development of Authority Controls

Besides archives description and database construction, there are also plans and development of organization in authority controls, which serve to raise the precision of indexing. Among archiving institutions, currently the Institute of History & Philology, Academia Sinica, the Institute of Modern History, Academia Sinica, and the National Palace Museum make use of a shared cataloging format for archive authority files. Using the Institute of Modern History as an example, the aspects included by "name authority control standards" are: 1. name, 2. tenure, 3. job title, 4. nation of residence, 5. Chinese calendar, 6. Western calendar, 7. page beginning [14]. The file creation aspects used by the National Palace Museum and the Institute of History & Philology are as described above, including: year of birth and death, aliases, citizenship, brief biography, talents, resume, writings, family relations, and social relations. From this can be seen that the authority file aspects used by the National Palace Museum and the Institute of History & Philology are more detailed than those used by the Institute of Modern History. In the national plan, the Content Development Division archives group will begin organization and development work on the name authority files of each archiving institution. The name information of each institution will be compared, and then through discussion the core elements of names will be decided upon, an organized archive name information database will be created, the cooperation between each institution will be shared construction and shared usage; however, such will not be a centralized system – through a copying method, each institution will be able upload constructed name core elements information to the system, where detailed information will be accessible by each institution through URL linkage. As a result, institutions that have not yet constructed their own name authority files can find the necessary information from the organized name information database and, using the copying method, choose and download the needed information to form their own name authority files. Consequently, because of the temporal and special qualities of the files, the organized archive name information database is expected to become the most comprehensive modern Chinese (including Taiwan) archive name information database.

## 3. Union Archives Finding Aids Construction

Digital archive contents under NDAP include: the six institution plans - Grand Secretariat files, recent key diplomatic and economic files, Qing-dynasty military operations files, democratic government files, Tan-Hsin-Tang-An, and Taiwan Soutufuku files, as well as open selection plan documents such as personal documents, Tainan war damage documents, Dung family acupuncture. To facilitate domestic archive digital information sharing, the Archive Group first designed the union archives finding aids. In developing the union archives finding aids, organizing preparatory tasks were performed; these can be divided into "setting core elements" and "archive browsing structure analysis." Technical personnel then assisted in writing the DTD format, exporting XML as a means of file exchange. The various members then decided whether to provide the URLs of the original catalogs, in order to present the complete database description of archives.

In the "archive browsing structure analysis," the control level principle in archiving studies was followed, using the concept of multi-levels: "full set – series – compilation – individual document." The Archiving Group interviewed each institution and recorded the browsing level of each institution's database, DTD-related information, etc. In writing the XML conversion program, the method of "exporting XML of all column positions" was used, because this method can provide an alternative format backup for the database; at the same time, in the face of increasing mutual linkage between digital databases, exporting the XML of all the column elements in the database can help in exporting different elements based on the needs of different shared catalogs. Exporting different column elements can easily attain the purpose of

exchange with different parts of the world.

Because Archive Group members will, in the future, import the entire union catalog in addition to providing the union finding aid, in order to prevent weariness among archive group members resulting from responding to the needs of different union catalogs, the union archives finding aids plays the role of uniting actor. The information of Archive Group members is imported to the union catalog through a single channel in order to prevent the challenges of Archive Group members facing different needs alone. As a result, the connections between Archive Group member databases, the archive union finding aids, and the union catalog can be shown as in the Figure, below:
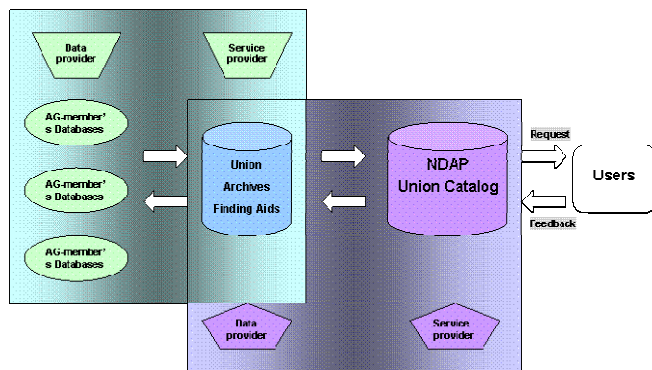


**Figure 1.** Relationship between Union Archives Finding Aids and the NDAP Union Catalog

In terms of the OAI (Open Archives Initiative) mechanism, the union catalog is a type of data service. The union archive finding aid is a data service to the Archives Group; it is a tool specifically for retrieving archives. To the union catalog of digital collections, it is a data provider, as it is the provider of the content of the union catalog. Finally, the users can access to archives collections through the NDAP Union Catalog.

At the same time as the completion of the important preparatory tasks, the planning and construction of the union archive finding aids is progressing. Such includes the construction of hardware and the system, design of the interface, etc; the functions desired in presentation of the union archives finding aids are repeatedly communicated and correlated with the technical personnel. The union archives finding aids is to become officially operational online in 2006.

## 4. Construction of the NDAP Union Catalog

In the beginning of planning for NDAP, a cross-field union catalog was expected to be constructed in order to provide for the opening and use of all forms of digital results; as a result the "Union Catalog System Construction Plan" was created. The union catalog of the National Digital Archives Program (hereafter referred to as union catalog), in order to unify the above mentioned databases from different academic fields and of different content forms, makes use of the Dublin Core (DC) in metadata standards [15]; this is widely recognized as a simple but effective core element group, and includes 15 basic elements in describing one archive. It possesses four important qualities: "easy to produce or maintain, widespread and easy to understand language, global usage, high flexibility" [16]. The union catalog compares metadata from about 88 different databases in the NDAP and DC, extracts important data from the correlation and inserts it into the union catalog to provide users with catalog-style browsing functions. In creating the union catalog, in order to respect the differing archives of the various digital archiving plans and the different foundations of related international standards, forced regulation was not used in construction of the union catalog. Rather, through detailed content analysis as well as communication and cooperation-heavy methodology, it was only executed after reaching a unified consensus. The most important construction method is cooperating with the content group of the digital archive technology plan, and through creating consensus within the group, reaching the goal of organizing the field. The creation and work of the archive group is one example.

Since development of the union catalog began in May 2002, after two years of public explanation, comprehensive research and system testing, the official version of the union catalog was introduced in August 2004. In the current form of the union catalog, there are three primary functions: categorical browsing, index searching, and related information. The union catalog currently provides four types of categorical browsing: thematic categories, participating organizations, temporal domains and spatial regions. The thematic categories browsing can be used with "simple list" or "list with thumbnail," and can also be used with "ontology" searching, as shown in Figure 2. Based on NDAP participating organizations, the major content holders and non-solicited projects are used as the viewing categories for information navigation.



**Figure 2.** Simple List and Thumbnail Style Browsing

The temporal domains browsing allows the user to either select or enter the desired time, then information retrieval to find archives belonging to the year of origin. The temporal domains browsing primarily adopted the "two thousand-year Western-Chinese calendar conversion" of the Computing Centre, Academia Sinica[17] to create a unified day-order (i.e., days away from A.D. 1) from Chinese calendar years in the metadata and a user-selected Western year of origin; The unified day-order can

then be compared to attain the goal of temporal domains browsing, as shown in Figure 3. With continually growing digital objects, a different method of temporal domains search must be developed, such as archaeological year and geological year. At the same time, a retrieval method making use of selected Chinese year of origin must be developed. The spatial regions browsing allows a user to select a location on a map to search for digital objects; at the same time, the concept of image levels is used to reveal greatest area of distribution for digital objects. The primary search interface is provided by GIS group of the Computing Centre, Academia Sinica [18]; the screen is as shown in Figure 3. With WebGIS, users can specify a geographic area on a world map, which is then transformed into sets of coordinates to retrieve the matched digital objects. On the technical level, the latitude/longitude information in the metadata is used to attain the purpose of information retrieval.



**Figure 3.** Time Axis and Geographical Distribution Browsing

In terms of the function of information retrieval, the union catalog provides full-text search and metadata search. Full-text search uses the search values inputted by the user in order to perform a text search on all elements and categorical catalogs, then divides the search results by categorical catalog, as in Figure 4. Metadata search is a type of advanced search; users can use complex commands on the 15 elements based on concept of archive content to restrict the area of search and raise the precision of search.



**Figure 4.** Full-text Search and Metadata Search

Finally, in related information, an introduction to the union catalog and requirement overview, answers to frequently asked questions, and historical accomplishments, etc. are provided. The primary topics include a requirement overview, union catalog working group introduction, related website links, communication information, intellectual property disclaimer, answers to frequently asked questions, and the historical accomplishments of the union catalog.

## V. Conclusion

The task of archive digitalization requires spending large amounts of labor and financial resources, and requires the participation of many organizations and personnel if it is to attain expected results. The Taiwan region has been reliant on the enthusiastic efforts of the National Science Council. In developing the NDAP, domestic institutions possessing large archives participated, encompassing almost all important domestic archiving institutions. Through the adoption of a common archive format, based on existing digitalized archives, the goal of resource sharing can eventually be reached. The digitalization model used by Taiwan can be said to be a top-down model; this model possesses the benefits of having a single concentrated management unit, a single authority in charge of finances, shared development of needed technology, etc. The foundations of digital archiving have already been created in Taiwanese archiving institutions, and archives creation of a union catalog specifically for archives are all the common efforts of the domestic archiving community. Finally, the results of archives digitalization can be benefit to the users.

## References

[1] C. K. Yang, "Current Situation of Taiwan Soutufuku Archives and Future Plans." Taiwanese Documents, 52(3), 463(2001).

[2] Institute of History & Philology, Grand Secretariat Archives at the Institute of History & Philology http://saturn.ihp.sinica.edu.tw/%7Edahcr/project90.html(2004/10/7). (2004).

[3] see[2].

[4] Aletheia University, National Digital Archives Plan Content Development Topic: Mackay and Oxford College http://www.au.edu.tw/ox_view/mackay/introduction.htm(2004/10/7). (2004).

[5] see[4].

[6] see[1].

[7] M. C. Hsiao, Description of Taiwan Historica Japanese Occupation-era Archive Large Image Scanning Efforts http://www.ndap.org.tw/1_newsletter/content.php?uid=539 (2004/10/15). (2004).

[8] Institute of Modern History, Important Recent Files on Diplomacy and Economics Digital Archives Plan http://dipeco.sinica.edu.tw/(2004/10/14). (2004).

[9] S. H. Chuang, "Summary of Recent Archive Digitalization Efforts by the Institute of Modern History, Academia Sinica" Newsletter for Modern Chinese History Issue., 32,108-114 (2001).

[10] see[8].

[11] Institute of Taiwan History, Institute of Taiwan History Digital Archiving Plan http://ithda.sinica.edu.tw/index.html(2004/02/10). (2004).

[12] National Digital Archives Program, Taiwan–Content Development Division– Archives group, Archive Description Organization

Principles (Draft) (Unpublished). (2003).

[13] International Council on Archives [ICA], (ISAD (G))：General International Standard Archival Description,2nd ed. (Ottawa：ICA, 2000).

[14] see[8].

[15] Metadata Group, Dublin Core Elements List http://www.sinica.edu.tw/~metadata/standard/dublincore-chi.htm (2005/04). (2004).

[16] Metadata Group, Dublin Core Description http://www.sinica.edu.tw/~metadata/standard/standard-frame.html (2005/04). (2005).

[17] Computing Centre, 2000 Year Chinese-Western Calendar Conversion http://www.sinica.edu.tw/~tdbproj/sinocal/luso.html (2005/04). (2005).

[18] Computing Centre, GIS Group, Space Information Technology Processing and Application http://gis.ascc.net/NDASupport/(2005/04). (2005).

## Author Biography

*Li-Kuei Hsueh received his MA in library and information studies from Leeds Polytechnic (1988) and his PhD in library and information studies from Loughborough Univ. (1991). Since then he has taught in the National Open Univ. of Taiwan (1991-1996). He is a professor of the Graduate Institute of Library, Information and Archives of Chen-chi Univ. of Taiwan (1996-present).*