

DIGITAL RESOURCES AND REQUIREMENTS FOR THE DEPLOYMENT OF DIGITAL LIBRARIES AT THE SOCIAL SCIENCES LIBRARY

PHAN TÂN^(*)

NGUYỄN DUY THỎA^(**)

and NGUYỄN VĂN HỘI^(***)

The Social Sciences Library, which is under the management and administration of the Institute of Social Sciences Information within Vietnam Academy of Social Sciences (VASS), has long been well-known by domestic and foreign researchers as a valuable archive. The construction of the digital collection of the Social Sciences Library and the website into electronic portal is very essential to meet the provision of information, to enhance the quality of servicing information, promote, introduce data resources, as well as to facilitate domestic and foreign scholars to approach copious amount of sources in a faster, easier, and more comprehensive way.

This article will briefly introduce selected sources of electronic information and digital collections of the Social Sciences Library, and at the same time, specify the requirements for expanding digital library in today's context.

I. Foundation of electronic information and digital collection

1. Directory information source

Nowadays, information source of the Social Sciences Library is quite diversity. Most documents of the Library are included in database and readers can lookup online on the Internet. The system of 16 database directories is considered to be the biggest and most diverse one in Vietnam and it is arranged into forms of documents and subjects by these following groups:

Group of database directory of handwriting documents includes:

- Database directory of Rural Conventions in Vietnam includes handwritings in Sino and Romanization Vietnamese characters, translations, which relate closely to Vietnam feudal

^(*) Dr., Vice Director of Institute of Social Sciences Information.

^(**) and ^(***) Institute of Social Sciences Information.

village culture with 6,960 folder recording over 5,000 villages all over Vietnam.

- Database directory of *Hagiography, Canonized Conferment* (Thần tích thần sắc) with 13,109 records of ancient villages in Vietnam, listing name of human spirits, history, legends of those divinities. In addition, there also have ordinance records of villages and the Hagiographies in Sino and Romanization Vietnamese characters with 4,221 records. Today, Hagiography, Canonized Conferment records in Sino and Romanization Vietnamese characters have been digitized and put out for service.

Database directories of books includes: 1/ Database directory of books with 132,928 records, consisting books in Latin, especially some ancient documents in Russian added after being transliterated from Latin and some in ethnic minority languages. The oldest document added to this database was published in the eleventh century. Old newspapers, magazines gathered in volumes (mostly left by the École française d'Extrême-Orient library) are added in the form of book record. 2/ Database directory consists of books on Hanoi with 1,636 records. 3/ Database directory of books belongs to the library system of VASS with 110,313 records, consisting books imported to specialized libraries of VASS since 1988. 4/ Database directory consists of Russian books with 69,650 records. 5/ Database directory consists of Japanese books with 9,649 records, including a lot of

books in 17th century. 6/ Database directory of Chinese books with 41,349 records, mostly are ancient Chinese documents including the oldest Chinese documents from 11th century.

Images database with 57,860 records. This is a priceless collection of photographs and postcards preserved in library repository, including photographs and postcards collected and taken by École française d'Extrême-Orient and the Institution of Social Sciences Information before 1990.

Database directory of disks that includes records of 910 old disks is also a precious directory. However, readers can only find general information about this database, and cannot exploit further due to the lack of necessary equipment.

Writings database deducted from journals with 182,192 excerpt records from over 100 domestic and foreign journals on social sciences, including 40 domestic up-to-date journals of social sciences.

Database listing title of newspapers, magazines in CDS/ISIS has only 839 records. It is expected that in 2014, this database will be completed in new software (Millenium).

Gazettes database with 6,738 records, including circulars, decrees, etc. from 1974 to 2000.

Database of thesis, essays major on social sciences conducted by scientists working at VASS contains 393 records.

So, until today, the database directory of the Social Sciences Library has about 600,000 records. Compared to other

leading information library centers in the country, this database directory system is an extremely large, strong on research ability and diverse in types of document. At the end of 2013 and the beginning of 2014, this database system has been updating in new software (Millenium) and readers are able to search online directly at opac.issi.vass.gov.vn.

2. Digitized information resource

Library has been constructing, implementing a very valuable digital collection. Documents added into digital include: handwritten documents (Rural conventions, Hagiography, Canonized Conferment, Logs of the state-established documents and Hagiography, Canonized Conferment in Sino and Romanization Vietnamese characters); books in ancient languages consisting of ancient Chinese, ancient Japanese; modern document includes Newsletter for the purpose of studying.

Until now, the Library has a digital collection up to thousands pages in general.

3. Website of the Institute of Social Sciences Information

Today, the Institute of Social Sciences Information has an official website in home page. In there, users can search all information relating to the Institute; find News serving for research by database directory. However, it is simply a static website which cannot be used as an electronic portal due to the lack of software to operate searching and exploiting digital resources and Database directories as well as applying services related to digital collections in the Social Sciences Library .

II. Practical requirements for the implementation of digital library.

1. Organizing and policies on exploiting digital collections

About organizing digital collections:

On the basis of electronic information resource as well as information infrastructure in the Institute, digital documents can be organized in three following collections:

- Collection of digital documents accessible through the Internet: contains digitized documents that can be accessed for free by other libraries and downloadable via LAN and the Institute of Social Sciences Information's website. Digital books and journals, articles that can be read online should only be downloaded when necessary; otherwise, links to documents' home pages are provided.

- Collection of digital documents processed by the Institute of Social Sciences Information from its information resource. These type of digital documents can be accessed through online service system.

- Collection of digital documents from endogenous documents, including researched results of Vietnam academy of social sciences. This collection is expected to access through online service and can become one of the most frequently accessed digital resources in the future.

About policies on exploiting:

In terms of digitized documents which are expected to access through online service with maintenance fee, serving

needs to be divided into two main types, which are: documents related to copyrights and documents that unrelated to copyrights. Ancient books and documents stored in EFEO are objects that do not have copyrights. On the other hand, projects, scientific researches whose copyrights belong to VASS can be posted online to use with charge per page. To other digitized documents such as newspapers, articles, journals; considering copyrights is essential. In case of unable to determine copyright, the solution is temporary posting online and serving with charge and limited times. Particularly, to documents such as Rural conventions, *Thần tích thần sắc* or other handwriting documents, the policy only permits the access to database directory and not full version of digitized documents.

2. Audience

The audience of the library in online electronic portal can temporarily be divided into 2 following groups:

Group 1 is readers who have library cards. These are true type of audience traditionally served by the library, thus, charging fee through online service is quite easy, by depositing and or additional charging when making cards or giving accounts to access.

Group 2 is readers that do not have library cards, existing in virtual form and only can be identified through accounts. Regarding this audience, the Social Sciences Library usually has to charge fee over bank accounts or SMS.

3. Forms of digitized information and digital resource

- Online information directory: all types of readers can use this online service for free.

- Directory of selected or periodical online information: This form of information is applied to each specific group of readers, possibly with fee. This service is carried out by library officers and supplied periodically according to demand of readers in the form of directory; specific fees can be set to different fields of documents demanded by readers.

- Digitized information online consists three main types: 1/ free digital resource, readers can search and download by themselves; 2/ Digital resource with charging fee periodically through accounts or other forms of prepayment; 3/ Digital resource that accessed one time only after charging fee through bank accounts or SMS.

- Information in hard copies for readers purchasing online, through bank accounts or SMS: As soon as money arrived at the library accounts, readers can receive hard copies by post or e-mail.

It is noticeable that the website of the Institute of Social Sciences Information also needs to point out access level, help readers to know their level to access. Besides, library needs to increase cooperation with other libraries to help readers to access official documents through borrowing service between libraries.

4. Constructing digital resource and digitized documents

Constructing Database directories:

To ensure the consistency and international standard of information directory of document put into database for online service, data processing has to be carried out by specialized officers, departments following due process and function. In this process, it is necessary to compose professional guideline that bears common regulations to ensure the consistency of document input and cooperation between departments, especially when dealing with old database.

Constructing digitized databases:

Constructing comprehensive database also needs to follow standardized steps to ensure accuracy and adequacy for digitizing documents. The process consists of following steps:

Step 1: Sketch a plan for digitizing documents, choose document, software, and determine data capacity after digitizing to prepare technical equipment.

Step 2: based on plan sketched, the department of preserve and department of customers are responsible for extracting documents out of storage, checking document condition in accordance to regulation.

Step 3: The digitizing team receives documents from screening team and carries out digitizing document.

Step 4: documents after being digitized are transferred to inspecting team to check the adequacy and ensure technical

standard of digitized documents (to each page).

Step 5: inspecting team transfers documents to electronic team to put into storage and update through the Internet system.

Every step needs signatures to determine responsibility of each department. This procedure will ensure that digitizing is implemented correctly in every document page and there are no mistakes in arranging digitized document.

5. Requirements for technical construction of electronic portal

a. Integrating electronic portal, exploiting and managing digitized documents

Software for managing digitized document is the key to construct and manage digitized databases of a library, which is integrated in library software system as a consistent and united body, ensuring consistency and fluent cataloging data and digitized data (full texting documents, movies, sounds, etc.) This software needs to meet all the following demands:

- It is a content management system or CMS, allows the management of content, construction of digitized data including: electronic books, E-magazines, digital data like images, drafts, maps, video, sounds, and courseware for distant learning.

- It is possible to manage a large database; at least 10 million pages (about 100GB) and has fast searching speed. It can be accessed in multi-facet,

multi-format form through Internet connection. It also has ability to integrate with all types of digitized data including Word, Excel, PDF, HTML, XML, TXT, PostScript, scanned images, sounds, video... It has ability to integrate with many data sources; it has ability to exchange, import/export/exchange data to other software based on common interfaced standards: METS, Dublin core, etc.

- Single sign-on, allows users to login once and access to all services. Establishes and installs all levels of privacy security to digitized documents. Providing access to documents and network security mechanism in three types of objects: General database available to access, apply to service, objects of stored digitized documents. It has mechanism of logging administration and recording all movements in system.

- It has data transfer and data converter to allow transferring texts into different forms to be exploited directly online. Each type of content needs to follow metadata or follow automatic cataloging standard Dublin core (creator, subject, summary, key words...).

- It allows during publishing process, specific information of texts can be retrieved from the database (contains comprehensive information and metadata), then associate with presentation form (in Page Templates) to create a complete document.

- It supports dynamic report samples. Statistic reporting system is diverse in

each module. Modules need to have consistent and sensible layout, have metadata searching mechanism. It also supports simple search, advanced search, full-text search and via Google fast and accurately.

- It uses Unicode encoding multi-language TCVN-6909 standard and simultaneously supports TCVN-5712.

- It allows the management in three service conditions: free, charging fee and restricted data. It manages accounts of readers who purchase data. To satisfy and expand the purchase of digitized documents through the Internet, the system also needs to support data purchasing service to help readers in the library or any guests search, choose, and order publications online...

- It allows merging multiple files into a one that its capacity may reach 100 GB; it establishes customized cataloging frame for each different data.

- It integrates system of automatic digitizing: automatic digitizing, identify image file, such as JPG, GIF, TIF, etc., to other forms, for example, PDF, Word, Excel, etc. Its ability to digitize identification has over 90% accuracy. It supports fluently further configuration for storing digitized documents.

- It has ability to adjust automatically the interface language.

b. Modules of digitized document management software

Modules of digitized document management software need to satisfy following requirements:

Module of digital resource:

- It allow users to manage digitized documents with files in different formats (pdf, doc, .docx, .xls, .xlsx, ppt...); construct automatic cataloging frames in Dublin core standard; merge many files into one and each file has capacity of 100 GB; divide data according to specialized multi-level subjects in the system.
- It manages digitized documents according to three conditions: free, charging fee and restricted data.
- It allows attaching cataloging record and copying cataloging information from document of system, reducing data entering task for library officials; creating documents of file collection and connecting various collections together. It can establish trailer file for documents and allows retrieving automatically cover page. It can show list of digitized documents according to topics, forms and conditions of data. It can transform data readable in computer \into formats supported by handset equipment, such as mobile phone, table, etc.
- The system also allows automatic general data indexing with a great number of documents or be done by library officials. The administrator can manage content of comments made by readers about the data. He also can manage the demand of readers to access restricted data and files; establish flexible security levels to digitized documents and files.
- The admin can establish only one digital folder contained all document files of system; import/export cataloging

information and digital files in cataloging standard: Dublin core, METS, Excel; create statistic record sample according to different criteria: digitized documents, most purchased documents...

Module of online searching (OPAC):

- Readers can access to data by unit and confidential level. It supports simple search, advanced search, full-text search fast and accurately. It allows readers to view trailer file made by creator while processing documents (create trailer files); it can manage list of data recently viewed in their home page. Support functions in data store for readers to manage on their own list of datas they want to purchase.
- Support management of allowing or not allowing readers to download data although it is for free; show digitized documents according to topics, collections and relations to each other.
- Readers can post comments, choose data in order of their concern; they can purchase all documents in the same collection, file or send request to purchase documents with restricted access.
- It supports the function of identifying password to check each action made by readers.

Module of list:

- It supports users to create multi-level specialized system. It is customized to use various mechanisms for different materials.
- It supports the managing of users, readers and data according to different units in the system.

- It supports dictionary of authors, publishers, subjects. To ensure the consistency in editing and increase searching data capability; a program allows the construction of data dictionary system (converted file) for files that needs to be consistently controlled when entering data or taken out for searching.
- Full-text indexing is integrated in the software system of digital library; it can be indexed by hands or automatically different levels associated with data management model.
- It supports users to establish dynamic report templates; allows them to construct specific system of report for their need without depending on software supplier.

Module of system management:

- It creates, provides access and usage rights to users in each unit and each function of each module in the system; it can change users' rights; delete accounts of users.
- It can create specific list of circulation points that an account user can manage; establish system parameter for program to use in working process; establish parameters of email system of library; establish parameters of IP address, IP range workstation access to the operator distribution system of digital documents.
- The system log has to record all operations of different functions of program, showing who is doing what in which moment. Manager can turn off/ on logging mode to each specific function in each specific module; log accessing by user, moment and working content.

- It allows the creation of statistic maps of system operation monthly, weekly, daily, and in each module.

- Users can adjust default language in the system and change theme on language. Managers can copy data automatically and semi-automatically to database.

Module of readers:

- It allows managing readers' account information, from pictures, birthday, ethnic, education level, occupation, and work, address... to card number, date of issue, expiration date. Based on that, it has mechanism of gathering identical groups when necessary.

- It can manage information of how readers use the library: list of documents viewed the most, account history of readers, history of purchasing digital document, etc.

- It can manage the permission of readers. Establish the range of access (unit, data security level, confidential level of readers...) data circulation policies for readers.

- It supports dynamic statistic reports in different criteria such as reader list, favorite document list.

Electronic source of information and digital collection of the Social Sciences Library is said to be quite big and diversity. Hopefully, with plan of implementing digital library carried out by the Social Sciences Library, in the future, the Social Sciences Library will become a copious and useful information resource for both domestic and foreign users.