Lifecycle and Requirements for Digital Collection Management of Thai Theses and Dissertations

Yuttana Jareonruen

Doctor of Philosophy Program in Information Studies, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand
E-mail: jyuttana@wu.ac.th

Kulthida Tuamsuk*

Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand
E-mail: kultua@kku.ac.th

ABSTRACT

This research was aimed at studying the situation, problems, and requirements for digital collection lifecycle management of Thai theses and dissertations. The mixed research method used was composed of: (1) Study of the problem and situation in which the qualitative method was applied. The research site covered 10 higher education institutions where the Thailand Digital Collection (TDC) project is operated. The informants were key administrative officers of the TDC project of each institution. In-depth and structured interviews were conducted on an individual basis to obtain the most accurate answers. (2) Study of requirements based on the quantitative research method to survey the requirements for the digital collection management system for Thai theses and dissertations from 84 purposively-selected TDC project officers and 527 end users selected by accidental sampling, totaling 611 samples. Research findings are as follow: (1) The study of the situation and problems of digital collection lifecycle management shows that Thai higher institutions systematically manage their digital collection. The management lifecycle is consistent with the *Guidance documents for lifecycle management of ETDs*, which included seven steps: program planning, creation, submission, and ingestion, access and retrieval of digital objects, archiving and preservation, evaluation and assessment, inter-operation (creation of institutional collaboration), and development of link data. (2) The study of requirements for digital collection management of Thai theses and dissertations shows five system requirements: acquisition and gathering, digitization, metadata standards, management of rights, and storage and retrieval, all of which are at M (mandatory) and D (desirable) levels.

Keywords: electronics theses and dissertations, digital collection lifecycle management, academic libraries, Thai higher institutions, Thailand Digital Collection

1. INTRODUCTION

A thesis or dissertation is mandatory for higher education studies. It demonstrates achievement of each program offered by an institution and is deemed one of the products from the institution's mission to prepare graduates for society. Additionally, theses are a primary resource for students and researchers in different disciplines (Rasuli, Alipour-Hafezi, & Solaimani, 2015), which, when completing all procedures required, are kept in the academic library in the form of hard copies and electronics files (Wang, 2014). A thesis carried out at a higher institution is generally kept in a library collection resource after the student has submitted the final report to his/her faculty, which in turn sends it to the library for storing. The steps include binding, cataloging, and shelving before each thesis is ready for service. However, circulation of theses is confined within the library or the affiliation network. Almost all libraries do not allow individuals to borrow a thesis from the library. Since the 1970s, when electronic publishing technology and word processor programs became available and more widespread, theses in the form of digital copies have increasingly become popular, especially in PDF files, and are under each library's retrieval system (Massicotte & Botter, 2017).

The concept of electronic theses emerged from an agreement made at a conference in Ann Arbor, Michigan organized in 1987 by the University Microfilms International. VirginiaTech was the first company to develop the system's platform with the aim to promote in electronic form to publicize theses and encourage exchange of the technology among higher institutions (Suleman, Atkins, Goncalves, France, & Fox, 2001). Nevertheless, higher education institutions (HEIs) were not fully prepared for electronic thesis production due to a number of limitations, e.g., budget, regulations, readiness of officers, and supporting infrastructures. Thus, collection development in the form of the 'digital library' was initiated for users to gain access to theses. Although this approach does not cover the entire lifecycle of electronics theses, its advantage in administrative effectiveness is realized. Moreover, users are able to gain access to this resource from anywhere and at any time just like they are for electronic theses (Shrivastava & Gupta, 2016).

In regard to the collaboration of HEIs, the Networked Digital Library of Theses and Dissertations (NDLTD) is an international organization dedicated to promoting the adoption, creation, use, dissemination, and preservation of electronic theses and dissertations (ETDs) (http://www.ndltd.org). NDLTD supports electronic publishing and open access to scholarship in order to enhance the sharing of knowledge worldwide. There are members in more than 50 universities, mainly from the USA, Europe, and South America. The concept of ETDs lifecycle management developed by NDLTD has become a guideline for ETDs management processes in several HEIs to transform from printed material into digital formats (Cayabyab, 2015). However, the reports of NTLTD's annual conferences revealed that most HEIs see ETDs program as just a shift in the format process of information storage and service. The ETDs project in most institutions was aimed at facilitating the accessing, storing, downloading, and printing of documents. In fact, there are more considering factors when implementing the ETDs project, such as copyright management, appropriate technology for the system, roles of stakeholders, and other standards and technical aspects (Donovan, 2014). Do and Gewissler (2017) also report that most successful ETDs projects are found in developed countries. The ETDs projects in developing countries are mostly in the infancy stage and seem to be unsuccessful in the long term.

In Thailand, the Office of Higher Education Commission (OHEC) initiated the digital collection management project within the Thai Library Integrated System (ThaiLIS) under the name "Thailand Digital Collection (TDC) in 2000 to 2001. The project has been ongoing until now with the main objective to promote Thai university libraries to develop their own digital collections of theses, research reports, academic articles, rare books, instructional media, and institutional archives. As such, all network universities are able to have access to the digital information. TDC architecture is a central unit with OHEC as the center and each university library is a branch that builds its own digital collection, which is in turn uploaded onto the central unit by means of the Digital Collection Management Software. At the beginning, OHEC supported 24 universities by providing necessary equipment such as computer nodes, system software, and scanners (Office of the Higher Education Commission, 2015). Now there are 171 main and associate members of TDC, with over 405,000 digital theses and other collections in the system. Each year, more data are added for common sharing among the network institutions and interested people. There are over 2,100,000 users and over 20 million downloads per year. In 2013 alone, downloads accounted for 32,863,268 times (TDC Use Statistics, 2013).

Based on the TDC annual reports and literature reviews, while the number of digital theses and dissertation in TDC and the number of users are increasing every year, there are significant problems in terms of management and services. For the management, because each university has different practices in managing theses and dissertations, some universities are responsible by the graduate schools, while some universities

are responsible by the libraries. There are also aspects such as the capacity of network technology, metadata standards, copyright, and rights management. Moreover, as the theses and dissertations are collected in PDF format, accessing to the content of documents is not possible. The system is not in a linked open data platform that would allow better accessibility and wider use (Loipha, 2001; Tinamas, 2012; Walailak University, 2014). The researchers saw that TDC had a need to improve the system by following the guidelines for lifecycle management of ETDs developed by NDLTD (Schultz & Skinner, 2014; Donovan, 2014). Also, TDC service on ETDs for the Thai academic community can be a part of NDLTD for international use in the future.

This paper is a part of the research that aimed at development of the ETDs management and service system for the TDC network. The study of the digital collection lifecycle management of theses and dissertations in Thai higher institutions served on the TDC network, based on the concept of NDLTD's ETD lifecycle management, was conducted. This comprises ETD program planning; creation, submission, and ingestion of digital materials; access and retrieval; archiving and preservation; and evaluation and assessment (Schultz & Skinner, 2014; Halbert, 2014; Donovan, 2014). The *DeLOs digital library reference model* (Candela et al., 2007) and the TDC framework developed by Walailak University (2016), both of which were used in this study, comprise five system requirements: acquisition and gathering, digitization, metadata standards, rights management, and storage and retrieval.

2. RESEARCH OBJECTIVES

The objectives of this research were to: (1) Study the situation and problems of ETDs digital collection management of the Thai HEIs providing their ETDs collections to the TDC network, based on the *Guidance documents for lifecycle management of ETDs* (Schultz & Skinner, 2014; Halbert, 2014; Donovan, 2014). (2) Study the requirements for ETDs digital collection management for Thai HEIs under the TDC network from the TDC project staff/practitioners and end-users' perspectives.

3. RESEARCH METHODOLOGY

This study was a mixed-method research, performed in two stages in response to the research objectives as follows.

3.1. Stage 1

Study of the situation and problems of ETDs digital collection management of the Thai HEIs, in which the qualitative method was applied. The research site covered ten institutions including nine universities providing their DTDs in the TDC collection, and one administrative office of TDC. The informants comprised twelve key administrative officers of the TDC project of each institution, selected by purposive sampling method. Data collection was done using the structured interview questions based on *Guidance documents for lifecycle management of ETDs* (Schultz & Skinner, 2014; Halbert, 2014; Donovan, 2014), comprised of 5 processes: ETD program planning; creation, submission, and ingestion of digital materials; access and retrieval; archiving and preservation; and evaluation and assessment. In-depth interviews were conducted on an individual basis from November 2017 to April 2018 in order to obtain the most accurate answers from the administrative perspectives. Lists of the ten institutions in this study were: Chiangmai University, Khon Kaen University, King Mongkut's University of Technology Ladkrabang, King Mongkut's University of Technology North Bangkok, King Mongkut's University of Technology Thonburi, Mahasarakham University, Ramkhamhaeng University, Taksin University, Thammasat University, and Office of Information Technology Administration for Educational Development at OHEC.

3.2. Stage 2

Study of the requirements for ETDs digital collection management for Thai HEIs, using the quantitative research method to survey the opinions of 84 purposively-selected TDC project staff/practitioners and 527 end-users of TDC's theses and dissertations database, selected by accidental sampling, totaling 611 samples. The questionnaire was done based on the *DeLOs digital library reference model* (Candela et al., 2007), a framework of guidance for building good digital collections (NISO, 2007) and the TDC framework developed by Walailak University (2016), comprised of five functions of digital collection management: acquisition and gathering, digitization, metadata standards, right management, and storage and retrieval. Data collection was done from May to August 2018. The data was analyzed by using percentage to identify the respondents' opinions on requirement levels.

Assessment criteria for requirement levels were based on M, O, D, E, and N according to the *User requirements documentation* (Robertson & Robertson, 2012), with the following details: M (mandatory requirement), level 5—the basic characteristic required in the system, D (desirable requirement), level 4—satisfactory characteristic that should be in the system except

in case of limitations in development (e.g., time frame, budget), thus if there is no development in this version, the characteristic should appear in the next version; O (optional requirement, level 3—optional basic characteristic that can be replaced by another characteristic such as the set characteristic or the only one standard to be selected; E (possible future enhancement), level 2—the characteristic not necessary for usage that, if existing, will increase system value depending on the system developer; and N (not necessary), level 1—the characteristic that is not necessary for the system and is not required by the user.

4. RESEARCH FINDINGS

4.1. Situation and Problems of ETDs Digital Collection Management of Thai HEIs

The findings were obtained from the interview and divided into five main categories based on the ETDs lifecycle management: Program planning; creation, submission, and ingestion; access and retrieval; archiving and preservation; and evaluation and assessment.

4.1.1. Program Planning

Written indication of background and importance of the digital collection project: All universities were found to have no program planning, but the service has been set to respond to the institution's key strategies.

Planning for request for support: All universities were found to plan budget requests and receive library budget allocations.

Proposing of implementation plans: Clarified planning was found with accurate annual time frames for production or import of digital collection, partly as a result of the agreement each institution made with the ThaiLIS network.

4.1.2. Creation, Submission, and Ingestion

Explanation of the policy and regulations for proposing electronics theses: The general policy and regulations of the graduate school are used, which cover only the printed version and the printed thesis supplementary file. There is no clear or written policy directly related to implementation of production and import of digital objects.

Building of primary supporting tools: Manuals and electronics data files are publicized, which the students can download from the Web page, but these have not been made as Web services, checklists, or assisting tools that are integrated with the institution's information system.

Procedures for checking completeness of thesis outlines: The procedures were found to be stipulated as the workload of faculties and the graduate school. The library is responsible for accepting the book and file of the final version of theses to proceed to the next step.

Data file production process for the system: It was found that such production was carried out in two major systems including library automation and institutional repositories, the two of which use different metadata, i.e., MARC21 and Dublin Core metadata where no integration of data files of the two systems has been carried out.

The digital data file format stipulated and supported by the original affiliation: Storage of data files was found to be in Microsoft Word and PDF only. There is no officially announced stipulation or standard in terms of digital data files. However, besides the data file of a thesis book, some universities have converted the CD/DVD-ROM containing the file of the applied program and the tool used by the student for doing the thesis into ISO format as a backup purpose.

The process for checking patents and intellectual property: There is no concrete policy or process, nor any approach for this process except announcement of approval for general users to click the accept button of the thesis patent by the original affiliation and inventor.

Thesis registration process: It was found that universities accept data from the faculty and program of study for officers from the graduate school to enter the data into the graduate school's system. No channel was found where students enter the data themselves like the overseas ETD management system, except only in the case of institutions that use Chulalongkorn University's iThesis. Therefore, the data in the information system of the graduate school is updated only once a year, or not updated at all until a student makes a request for a thesis defending examination.

The process for making an approval page for usage based on patent law: No universities were found to have this process. University libraries observe that their graduate students have signed their names as acknowledgement of the patent issue when they applied for the program, or when they propose the thesis topic. This document is used as certification. As for dissemination of work by the Internet network, most universities use an approval contract in the form of Creative Commons License for publicizing and access to the digital collection.

Cataloging process and importing digital collections for user service: The findings show that 1) the thesis book file is principally entered through Dublin Core metadata before being placed in the TDC system to serve users. There is no international metadata standard for theses (ETD-MS); only unqualified Dublin Core metadata are used as stipulated by the working group of the ThaiLIS network, in which the details

are nearly similar and interchangeable. 2) Setting topics and keywords for retrieval is done by means of the manual for topics of LCSH (Library of Congress Subject Headings) and the Thai topic manual. Entering of details in general has as yet not been based on the standard that supports meaningful retrieval such as Resource Description and Access (RDA) or Bibliographic Ontology.

4.1.3. Access and Retrieval

Access to digital information resources: Publicizing theses and dissertations in digital format was found to be performed in two ways. 1) Publicizing digital collection in the form of born-digital, which provides service via a digital collection information system or institutional information bank and licensing of access, via central data of TDC and institutional information bank using IP address group and license registration. 2) Publicizing of digital collection in retrospectively reformatted form: It was found that this is not carried out now because the project used data conversion from Microsoft Word files from the start. There is no policy to scan old thesis books stored in the form of printed media due to limitations of manpower and budget.

Policy on access and retrieval services: Several link sources have been built to enable users to have access to digital collections from the Web OPAC of university libraries. Users have an access channel via a direct retrieval menu as well as by a link from Web OPAC to the searched digital object. However, presently users are not able to have access to a university's digital material via its library's Web page, nor are they able to gain access through a general mechanism like Google.

Policy for publicizing and allowing digital collections to be discovered over a wider circle: There is no clear policy found in this respect except for inviting users by public relations, which is carried out whenever the semester starts but not on a regular basis all through each fiscal year.

Producing of retrieval tools: In general, the universities were found to be using the existing tools. No additional retrieving tools have been developed such as Facet or Linked Data, which are specific tools for retrieving digital collections of theses.

4.1.4. Archiving and Preservation

Development of archiving and preservation system: No policy related to or development was found for archiving and preservation of digital collection in full-preservation form. What is done at present includes reserving backup data in indexes and digital objects under the presently used information system only (e.g., automated library, institutional information bank, and e-Thesis system). There is no policy for selecting reliable recorded media (e.g., static storage, cloud) for transferring and

preserving digital collections.

Analytical and systematizing processes for digital collection: Systematization of importing online digital collection was found where documents are transferred via a webpage to the central database of TDC using Dublin Core metadata to produce indexes. Thai universities under the ThaiLIS network have entered an agreement to construct a model for naming files, directories, and systematic ordering of data files based on manuals and clear inspection. However, there is no control of the import document. The new files are still recorded on top of former ones as the present system does not enable version control, and the existing theses have already been finalized.

Recording and storage of digital collection: The original copies were found to be stored in the computer node, which is the file server of each university. Then the data is transformed into PDF files, most of which have not undergone embedding of metadata directly onto the digital object. There is instead description of details of the digital object from tabulated metadata for retrieval index. Besides, in the compressing or de-sizing of the digital materials for serving users, all universities were found to have no policy to produce master files and delivery files that match their context and users' accessing tools. The digital objects served are the same as the original version in the institution's storage.

Formatting and preservation of digital objects: No formatting or preservation was found for digital objects, nor were there agreement or approaches for this process.

Preservation of metadata and URIs of digital objects: No preservation of metadata and URIs of digital objects was found, nor agreement or approaches for this process.

4.1.5. Evaluation and Assessment

Determining stakeholders of the universities' digital collection project: This has not been done in writing. In addition, the overall management of theses in the universities is separately performed at the graduate school, the library, and the information technology center. DTDs collection is under the responsibility of the library.

Determining the method, tool, and indicator for project evaluation: Nothing has been concretely done regarding the determination of method, tool, and indicator for project evaluation, except evaluation of routine work and the overall efficiency of library service. Moreover, the analysis shows that the universities do not realize the existence of standard criteria, methods, tools, and indicators for evaluating the project. If there are reliable and standard criteria at an international level, they will be used systematically in due course.

The process for constructing a systematic tool: There is a need for university libraries to develop an evaluation tool in

the form of database on the Web in addition to evaluating the digital collection. This will support implementation of the library in all aspects. However, the universities still do not know how to start.

Evaluation of infrastructure, storage, and capacity in the use of digital collection: No evaluation of infrastructure, storage, and capacity has been performed since the existing system cannot accommodate such work. Statistics related to access of use from general users cannot be a direct indicator for the evaluation.

The overall research findings on the situation and problems of ETDs digital collection management of Thai HEIs can be concluded as illustrated in Fig. 1.

In addition, the researcher collected more data on the building of collaboration among Thai HEIs in the management of DTDs collection, from which key elements can be drawn as follows: Determining stakeholders of the universities' digital collection project: There is no written determination on this. In addition, the overall management of theses and dissertations in the universities is separated, i.e., under the graduate school, the library, and the information technology center. DTDs collection is under the library's responsibility.

Infrastructures, the basic component of the management and storage of the present digital collection in Thai universities for storage systems, digital collection management, digital object delivery, and information discovery: The sub-components of these include cataloging and searching systems. In addition, the data analysis discloses that all HEIs need

support of infrastructure from the center in order to assist their implementation of the digital collection project. These infrastructural components include 1) storage and delivery of digital collections, namely, the storage site, computer node, database, and platform for an institution, and the service system for retrieval and transformation of digital documents; 2) metadata: the system for managing the metadata format, exchange of schema, and mechanism for harvesting metadata; 3) retrieval: the database system, the retrieval system that is compatible with the searching mechanism; and 4) the basic network that supports the increase or development of applied programs that differ from one institution to another.

Software for system management: It was found that each university uses different software for operating the digital collection project, e.g., software from different distributors or developers. This leads to a problem in terms of practice, and there is still no concrete solution or agreement made. Besides, the HEIs need software that can be commonly used at the consortium and the information network levels as well as at the communication network or computer network level.

Patents and organization culture related to patents: This was found to be one of the bigger problems when working together, especially the former concept of the information resource owner. All universities share the same opinion that solutions in terms of patents and organizational patent culture depend on understanding of each institution's policy towards
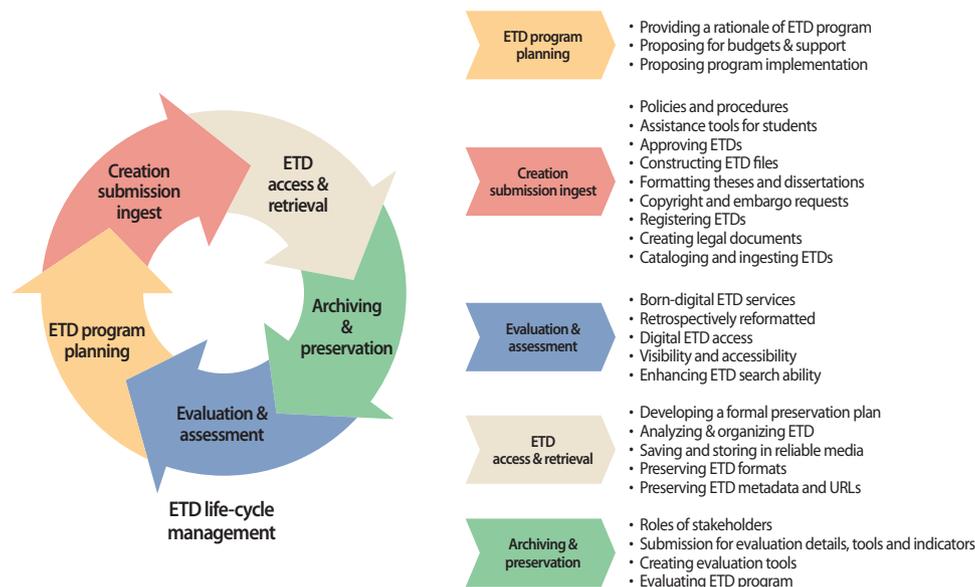

Fig. 1. Process of electronic thesis and dissertation (ETD) lifecycle management of digital collection in Thai universities.

implementers.

Administration and advisory of the collaborating network: Problems were found in implementation because the project has been initiated from the center, i.e., OHEC. The advisory board or the committee has been appointed as a working group with no power for decisions or order. The project structure itself makes it difficult to appoint advisors or experts for management and advice.

4.2. Requirements for DTDs Collection Management for Thai HEIs

Informants included 184 males (30.1%) and 427 females (69.9%) whose highest education level ranges from lower than bachelor's degree (26.7%) to bachelor's degree (66.3%), master's degree (4.9%), and doctorate's degree (2.1%). Their experience in the use of TDC service ranges from less than 1 year (42.2%) to 1-3 years (32.7%), 4-6 years (14.2%), 7-9 years (7.4%), and over 10 years (3.4%). The objectives of TDC use included: to follow up academic advances (24.2%), for doing research, thesis or independent study (53.7%), and supplementing academic work (21.9%). The frequency of use ranged from more than once a year (8.8%) to more than once a month (20.0%), more than once a week (38.5), and more than once a day (32.7%).

Five categories of requirements for DTDs collection management were investigated: Acquisition and gathering, digitization, metadata standards, rights management, and storage and retrieval. The requirements of two groups of users were analyzed: staff/practitioners and end-users based on M, O, D, E, and N under the *User requirements documentation* (Robertson & Robertson, 2012). The results can be concluded as shown in Table 1.

4.2.1. Acquisition and Gathering

The highest requirement of both staff/practitioners and end users at the M level is to have a web application interface between the system and users that can work on general browsers (66.7% and 48.9%, respectively). The lowest requirements at D level are: a system to accommodate file import, which is an RIS or BibTex file from commercial Google Scholar (41.70%), and a system to accommodate exchange of data with the graduate school or an external system using the scholar metadata standard (41.70%).

4.2.2. Digitization
4.2.2.1. Data Objects

Staff/practitioners equally require three types of data objects at the highest level (M), namely, for the system to accommodate

a single digital object, the system to accommodate statement files (such as Microsoft Word .doc and .docx), and the system to accommodate import and display of digital objects in the form of PNG pictures (50.0% for all of the three categories). The highest requirement at M level of end users is for the system to accommodate a single digital object (44.0%). The least requirement of practitioners at D level is for the system to accommodate digital objects in the form of TIFF pictures (38.1%), which is also similar to the least requirement of end users (34.0%).

4.2.2.2. Multimedia Objects

The highest requirement of practitioners at M level is for the system to accommodate migration of digital objects in the form of format migration (52.4%), whereas the end users require the system to accommodate digital objects in the form of MPEG-4 video file (.mp4) moving pictures the most (M level) at 39.3%. The least requirement of practitioners at D level is for the system to accommodate Flash video files (.flv) (41.7%). The least requirement of end users is for the system to accommodate Audio Video Interleave Files (.avi) (33.1%).

4.2.2.3. Digital Object Identifiers

The highest requirement at M level of staff/practitioners is for the system to accommodate digital object identifiers (53.6%), while the highest requirement at M level of end users is for the system to accommodate digital objects based on the Handle System for use with the identifier for referencing (37.8%). The lowest requirement at D level of staff/practitioners is for the system to identify digital objects using the Universally Unique Identifier (41.7%), which is similar to the least requirement of end users (35.2%)

4.2.3. Metadata standards
4.2.3.1. Structural Metadata

Staff/practitioners have the highest requirement at M level for a system of structural metadata that assists in managing patent data, allowance of access, and technical details of theses of users (64.3%). The highest requirement of end users at D level is for the system to accommodate the standard data structure machine-readable cataloging format (MARC) (39.4%). The staff's requirement at D level is for the system to accommodate standard of specific data such as VRA Core at 41.7%, which is similar to the least requirement of end users (33.9%).

4.2.3.2. Descriptive Metadata

The highest requirement at M level of staff and practitioners is data value standards in MeSH form (Medical Subject Headings)

(48.8%). The highest requirement of end users at D level is that the metadata in the system should be compatible with the markup language (39.1%). The lowest requirement of staff/practitioners at M level is data content standards in the form of FRBR (Functional Requirements for Bibliographic Records) (39.3%), while the least requirement of end users at D level is data content standards in the form of RDA (35.5%).

4.2.4. Rights Management

Staff/practitioners show the highest requirements at M level equally in two items: a system for permitting or denying access from outside to collections with set access rights and for the system to accommodate the use of metadata for giving details in patents of digital objects (67.9% for the two items). The highest requirement of end users at D level is a system for permitting or denying access from outside to collections with set access right (38.5%). The lowest requirements of staff/practitioners at M level are a system to accommodate levels of authorization and definitions of work for each group of users and for the system to accommodate password administration for users in the form of system-assigning passwords (60.7% for both items). The lowest requirement of end users at D level is for the system to accommodate use of metadata for giving details in patents of digital objects (35.4%).

4.2.5. Storage and Retrieval
4.2.5.1. Storage

Staff/practitioners require all of the items under this category at M level; the highest one being ability to browse, search, and retrieve efficiently (70.2%). Their least requirement at M level is the system capacity to distinguish special collections of searching tools in the digital thesis collection system (48.0%). This is similar to the least requirement of end users, which is, however, at D level (34.0%).

4.2.5.2. Retrieval

Staff/practitioners show the highest requirement at M level for retrieval capacity under basic search and advanced search (Boolean logic, Truncation/wild card, and limit search, 72.6%), while the highest requirement of end users at D level is for the system to display results of metadata to explain the digital object directory in the forms of MARC21, MARCXML, and Dublin Core metadata (37.5%). The least requirement of staff/practitioners is outlines in the forms of DTD, XML schemas, and technical descriptions for display of records for benefit of use with other databases (45.2%). The lowest requirement at D level of end users is for the system to have a Breadbox tool that allows users to review, refine, and limit the retrieving screener and various facets existing in the searching tools of the system (32.7%).

Table 1. Requirements for ETDs digital collection management

Requirements for ETDs digital collection management	Practitioners (N=84)			End-users (N=527)		
	No.	%	Level	No.	%	Level
1. Acquisition and gathering						
1.1 Users can access the system from the IP address groups of the institution that is affiliated.	46	54.8	M	192	31.4	M
1.2 The user interface must have a Web application interface between the system and users that can work on general browsers.	56	66.7	M	299	48.9	M
1.3 The system can export data records to exchange with external systems or use for other purposes of the institution.	45	53.6	M	218	35.7	D
1.4 The system to accommodate the increasing of ETDs with no restrictions on the number of records in the database.	55	65.5	M	233	38.1	D
1.5 The system to accommodate storing digital objects at both the collection level and the digital object level.	42	50.0	M	225	36.8	D
1.6 The system to accommodate the creation of transaction records by importing MARC21 files from LMS systems of institutions that are members of ThaiLIS and TDC networks.	49	58.3	M	230	37.6	D
1.7 The system to accommodate an import, which is an RIS or BibTex file from commercial Google Scholar.	35	41.7	D	211	34.5	D
1.8 The system to accommodate exchange of data with the graduate school or an external system using the scholar metadata standard.	35	41.7	D	235	38.5	D
2. Digitization						
2.1 Digitization of data objects						
2.1.1 The system to accommodate a single digital object.	42	50.0	D	269	44.0	D
2.1.2 The system to accommodate the management of digital object with multiple formats embedded.	34	40.5	M	252	41.2	D

(Continuing)

Table 1. Continued

Requirements for ETDs digital collection management	Practitioners (N=84)			End-users (N=527)		
	No.	%	Level	No.	%	Level
2.1.3 The system to accommodate the management of digital objects in which one object can be displayed as multiple formats (such as TIFF, JPEG, and PDF/A).	37	44.0	M	222	36.3	M
2.1.4 The system to accommodate statement files (such as Microsoft Word .doc and .docx).	42	50.0	M	213	34.9	M
2.1.5 The system to accommodate image digital objects in TIFF format.	32	38.1	M	208	34.0	D
2.1.6 The system to accommodate image digital objects in JPEG format.	36	42.9	D	220	36.0	M
2.1.7 The system to accommodate image digital objects in PNG format.	42	50.0	D	219	358	D
2.1.8 The system to accommodate image digital objects that are computer graphic images.	34	40.5	D	212	34.7	D
2.1.9 The system to accommodate image digital objects stored in a dataset.	40	47.6	D	218	35.7	D
2.2 Digitization of multimedia objects						
2.2.1 The system to accommodate digital objects in the form of Flash video files (flv).	35	41.7	D	234	38.3	D
2.2.2 The system to accommodate digital objects in the form of MPEG-4 video file (mp4).	40	47.6	D	240	39.3	D
2.2.3 The system to accommodate digital objects in the form of MPEG video file (mpg) movies.	38	45.2	D	219	35.8	D
2.2.4 The system to accommodate digital objects in the form of audio video interactive file (AVI).	39	46.4	D	202	33.1	D
2.2.5 The system to accommodate digital objects in the form of QuickTime (MOV) movies.	39	46.4	D	218	35.7	D
2.2.6 The system to accommodate migration of digital objects across media (media migration).	40	47.6	D	231	37.8	D
2.2.7 The system to accommodate migration of digital objects across format (format migration).	44	52.4	D	222	36.3	D
2.3 Digital object identifiers						
2.3.1 The system to accommodate description of digital objects with Dublin Core metadata (qualified and unqualified).	44	52.4	M	215	35.2	D
2.3.2 The system to accommodate description of digital objects with RIS metadata.	35	41.7	D	225	36.8	D
2.3.3 The system to accommodate handle system (HS) for use with the identifier for referencing.	40	47.6	D	231	37.8	D
2.3.4 The system to accommodate to identify digital objects using the University Unique Identifier.	35	41.7	D	215	35.2	D
2.3.5 The system to accommodate a mechanism to prevent errors in file naming, or when accessing digital objects through the Internet.	44	52.4	M	225	36.8	D
3. Metadata						
3.1 Structural metadata						
3.1.1 The system to accommodate description of terms in digital collections with tools that are international standards (such as ISBD, subject headings, taxonomy).	44	52.4	M	236	38.6	D
3.1.2 The system of structural metadata that assists in managing patent data, allowance of access, and technical details of theses of users.	54	64.3	M	226	37.0	D
3.1.3 The system of structural and relationship of digital objects in digital collections with structural metadata.	48	57.1	M	207	33.9	D
3.1.4 The system to accommodate format.	47	56.0	M	241	39.4	D
3.1.5 The system to accommodate data structure standard in DCMES.	47	56.0	M	211	34.5	D
3.1.6 The system to accommodate data structure standard of research information system.	37	44.0	D	220	36.0	D
3.1.7 The system to accommodate standard of specific data such as VRA Core.	35	41.7	D	223	36.5	D
3.2 Descriptive metadata						
3.2.1 The system to accommodate data value standards in LCSH.	40	47.6	M	225	36.8	D
3.2.2 The system to accommodate MeSH.	41	48.8	M	220	36.0	D
3.2.3 The system to accommodate data value standards in TGN.	34	40.5	D	222	36.3	D
3.2.4 The system to data value in AACR.	38	45.2	M	227	37.2	D
3.2.5 The system to accommodate RDA.	39	46.4	M	217	35.5	D
3.2.6 Data content standards must be in the form of RDA.	39	46.4	M	223	36.5	D
3.2.7 Data content standards in the form of FRBR.	33	39.3	M	226	37.0	D

(Continuing)

Table 1. Continued

Requirements for ETDs digital collection management	Practitioners (N=84)			End-users (N=527)		
	No.	%	Level	No.	%	Level
3.2.8 The system can use metadata to manage digital collections consisting of various information resources (Application profiles).	36	42.9	M	224	36.7	D
3.2.9 System metadata must be compatible with the Markup language.	37	44.0	M	239	39.1	D
4. Rights management						
4.1 The system should display a license agreement or license for every collection in the system.	52	61.9	M	223	36.5	D
4.2 The system to accommodate permitting or denying access from outside to collections with set access rights and the system to accommodate use of metadata for giving details in patents of digital objects.	54	67.9	M	235	38.5	D
4.3 The system to accommodate the use of metadata to provide copyright details of digital objects in the system.	57	67.9	M	216	35.4	D
4.4 The system to accommodate determining the level of use (authorization) and a detailed description of the work (definition) for each group of users according to usage level.	51	60.7	M	223	36.5	D
4.5 The system to accommodate the use of password administration for users' system-assign password.	51	60.7	M	228	37.3	D
4.6 The system to accommodate the use of password administration for users' selected password.	52	61.9	M	225	36.8	D
4.7 The system must restrict the access rights to full documents of users/user groups, both at the object level and the collection level.	51	64.0	M	230	37.6	D
5. Storage and retrieval						
5.1 Storage requirements						
5.1.1 The system to accommodate the distributing of born-digital ETDs.	52	61.9	M	229	37.5	D
5.1.2 The system to accommodate the distributing of retrospectively reformatted ETDs.	50	59.5	M	218	35.7	D
5.1.3 Users can browse, search, and retrieve digital collections efficiently.	59	70.2	M	211	34.5	D
5.1.4 The system to accommodate system capacity to distinguish special collections of searching tools in the digital thesis collection system.	41	48.8	M	208	34.0	D
5.1.5 The system to accommodate a search mechanism that can be searched from a single search together with other digital information resources of the library.	48	57.1	M	231	37.8	D
5.1.6 The system to accommodate indexing for searching digital collections from search engines that are commonly used (such as Google).	57	67.9	M	215	35.2	D
5.1.7 The system to accommodate a registration mechanism that allows digital objects in the collection to be used in conjunction with open institutional repositories.	42	50.0	M	219	35.8	D
5.1.8 The system to accommodate a mechanism for extracting metadata from digital objects that users upload to the system, as well as to be able to use metadata with information systems that use the OAI-PMH protocol.	43	51.2	M	238	39.0	D
5.2 Retrieval requirements						
5.2.1 The system to accommodate search data using basic search and advanced search methods (Boolean logic, truncation/wild card, and limit search).	61	72.6	M	228	37.3	D
5.2.2 The system has a Breadbox type tool that users can review, refine, limit, search filter, and has various facets available in the system's search tools.	46	54.8	M	200	32.7	D
5.2.3 The system to accommodate taxonomy browsing which administrators can group taxonomy on the display screen by setting it as a parameter.	40	47.6	M	205	33.6	D
5.2.4 The system can display the metadata used to describe the list of digital objects in MARC21, MARCXML, and Dublin Core metadata formats.	44	52.4	M	229	37.5	D
5.2.5 The system to accommodate icons which represent the type of digital information resources in the system and the format type of digital objects in the search results screen.	45	53.6	M	205	33.6	D
5.2.6 The system to accommodate searching from mobile, smartphone, and tablet devices in the form of Web applications.	52	61.9	M	203	33.2	D
5.2.7 The system to accommodate DTD schemes, XML schemas, and technical documentation for records output for the benefit of using with other database systems.	38	45.2	M	203	33.2	D
5.2.8 The system to accommodate A-to-Z list creation by name, author name, institution name, collection name, and degree name.	60	71.4	M	217	35.5	D

ETD, electronic thesis and dissertation; MARC, machine-readable cataloging; LMS, learning management system; ThaiLIS, Thai Library Integrated System; TDC, Thailand Digital Collection; ISBD, International Standard Bibliographic Description; DCMES, Dublin Core Metadata Element Set; LCSH, Library of Congress Subject Headings; MeSH, Medical Subject Headings Data Value Standards; TGN, Getty Thesaurus of Geographic Names; AACR, Anglo-American Cataloging Rules; RDA, Resource Description and Access Cataloging; FRBR, Functional Requirements for Bibliographic Records.

5. DISCUSSION AND CONCLUSION

The results of this study on the lifecycle and requirements for DTDs collection management for Thai HEIs can be summarized and discussed in two aspects as follows. The study of the situation and problems of DTDs collection management for Thai HEIs in the TDC network – The findings demonstrate that Thai HEIs manage their digital collection systematically, with a lifecycle of management that is consistent with the *Guidance documents for lifecycle management of ETDs* (Schultz & Skinner, 2014; Wang, 2014). However, the findings show that Thai HEIs want to establish academic collaboration among other universities in order to share digital collection services. In this regard, the requirements fall into four categories, namely, infrastructures, system management software, support for patents and organizational culture related to patents, and administration and advising. These four requirements are the contexts added under the digital collection collaboration of Thai higher institutions.

The study of the requirements for DTDs collection management of Thai HEIs show that staff/practitioners responsible for digital collection of theses and electronic theses and end users have the following five requirements related to the system as follows.

Acquisition and gathering: Staff/practitioners and end users require a production system and acquisition of theses and electronic theses which are applied programs workable via browsers, followed by system functions that can accommodate exchange and import of external files and import of theses and electronic theses references from commercial data and Google Scholar.

Digitization: Staff/practitioners require a system that supports conversion of multiple digital files, while end users only want simple data files that are compatible with smartphones and tablets. Both groups of users do not emphasize the importance of big data file storage (such as TIFF or AVI), even though these data files allow production of master files and long-term digital preservation. As for digital identifiers, both groups prefer
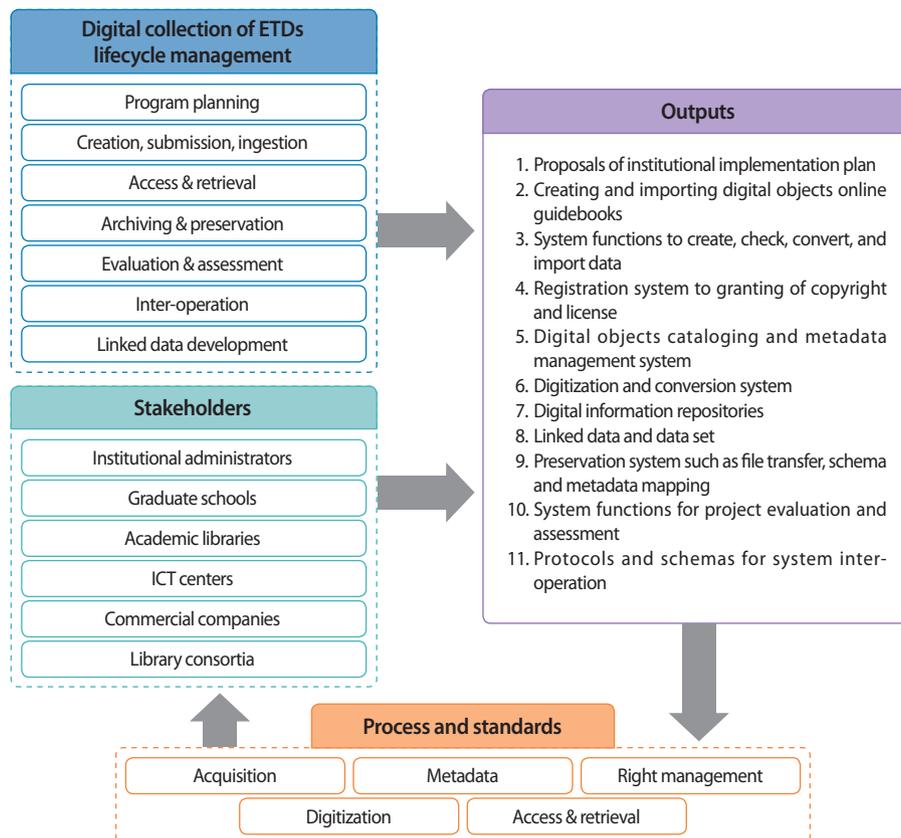

Fig. 2. Requirements for lifecycle management of digital collection of Thai higher institutions. ETD, electronic thesis and dissertation; ICT, information and communication technology.

external and international identifiers such as digital object identifiers rather than internally created identifiers, even though the two types of identifiers can work interchangeably.

Metadata standards: Staff/practitioners require a system that accommodates explanation of digital collection directories by means of international tools (such as international book standards, topics, and glossaries). End users want directory descriptions for patent management, access authorization, and the technical details of each thesis. Besides, both users require metadata that is compatible with a record structure similar to MARC, which is normally used in general libraries. The findings reflect familiarity with metadata that have been used for a long time in university libraries. What staff/practitioners and end users agree on is that there is no necessity for the use of metadata to describe specialized collections such as VRA core, TGN, or RAD.

Rights management: Staff/practitioners see the importance of a system mechanism that allows access and want to offer collection service in a closed system, while end users see the importance of disclosure of term agreements and mechanisms for licensing.

Storage and retrieval: Staff/practitioners and end users require a system with storage and retrieval mechanisms similar to the existing library system. The key functions required comprise Basic Search and Browse, followed by other retrievals such as Refine search, Relevance search, and Facet search. The least requirements of the two groups are a mechanism that links digital collection of theses with an open information bank of an external system. Their least requirements do not mean that they do not require it, but after prioritization, both users see the importance of retrieval mechanisms the most.

From the analyses of the status and problems of and the requirements for DTDs collection management of Thai HEIs in this study, the following four categories of TDC system specification can be drawn (Fig. 2).

First, a lifecycle of ETDs digital collection management, which include: Program planning; creation, submission, and ingestion; access and retrieval of digital objects; archiving and preservation of digital objects; evaluation and assessment; inter-operation (creation of institutional collaboration); and development of link data.

Second, stakeholders comprise: Institutional administrators, graduate schools, higher institutions' libraries, institutional information technology centers, commercial database service providers, and collaborative networks of higher institutions.

Third, system requirements comprise: Standards for acquisition and provision of digital objects; metadata management standards; management standards for patents and intellectual properties, including access licenses for digital objects; digitization standards; and storage and retrieval standards.

Fourth, required outcomes from the system, which can be a sub-system, module, or system function, comprise: A system or website that provides details of the project and the policy for implementation at the institutional level; manuals for creating and importing digital objects as online guidebooks; functions for creating, checking, converting, and importing data; a registration system for granting copyrights and license; digital object cataloging and metadata management systems; digitization and conversion a system; digital information repositories; linked data and data sets; preservation systems such as file transfer, schema, and metadata mapping; project evaluation systems; and protocols and schema for system inter-operation.

REFERENCES

Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C., ... Schuldt, H. (2007). *The DeLOs digital library reference model foundations for digital libraries*. Pisa: A Network of Excellence on Digital Libraries.

Cayabyab, A. T. (2015). A review of emerging ETD initiatives, challenges and future developments. *International Journal of Information and Education*, 5(10), 772-777.

Do, D. T., & Gewissler, L. (2017). *Managing ETDs: The good, the bad, and the ugly*. Retrieved June 10, 2019 from https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1949&context=charleston.

Donovan, B. (2014). Managing the lifecycle of ETDs: Curatorial decisions and practices. In M. Schultz, N. Krabbenhoeft, & K. Skinner (Eds.), *Guidance documents for lifecycle management of ETDs* (pp. 100-117). Atlanta: Educopia Institute.

Halbert, M. (2014). Guide to options for ETD programs. In M. Schultz, N. Krabbenhoeft, & K. Skinner (Eds.), *Guidance documents for lifecycle management of ETDs* (pp. 172-182). Atlanta: Educopia Institute.

Loipha, S. (2001). Electronics theses and dissertation (ETDs) searching. *Journal of Information Science*, 19(3), 19-29.

Massicotte, M., & Botter, K. (2017). Reference rot in the repository: A case study of electronic theses and dissertations (ETDs) in an academic library. *Information Technology and Libraries (ITAL)*, 36(1), 11-28.

NISO. (2007). A *framework of guidance for building good digital collections*. Retrieved June 10, 2019 from http://framework.niso.org/.

Office of the Higher Education Commission. (2015). *Summary of the satisfaction evaluation form for using the full database of electronic documents in Thailand*. Bangkok: Office of the Higher Education Commission.

Rasuli, B., Alipour-Hafezi, M., & Solaimani, S. (2015). *Understanding electronic theses and dissertations through a business model perspective: The case of Irandoc ETDs*. Retrieved June 10, 2019 from https://www.researchgate.net/publication/284715166.

Robertson, S., & Robertson, J. (2012). *Mastering the requirements process: Getting requirements right*. New Jersey: Addison-Wesley.

Schultz, M., & Skinner, K. (2014). Introduction. In M. Schultz, N. Krabbenhoeft, & K. Skinner (Eds.), *Guidance documents for lifecycle management of ETDs* (pp. i-viii). Atlanta: Educopia Institute.

Shrivastava, P., & Gupta, D. K. (2016). *National level ETD efforts: A comparative analysis*. Gujarat: INFLIBNET Centre.

Suleman, H., Atkins, A., Goncalves, M. A., France, R. K., & Fox, E. A. (2001). Networked digital library of theses and dissertations. *D-Lib Magazine*, 7(9). Retrieved June 10, 2019 from http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html.

TDC Use Statistics. (2013). Retrieved January 20, 2019 from https://tdc.thailis.or.th/tdc/.

Tinamas, S. (2012). *Management of digital copyrighted materials in Thai higher education institutes*. Retrieved June 10, 2019 from https://tise2015.kku.ac.th/drupal/?q=node/19.

Walailak University. (2014). *Meeting minutes of Thai Academic Digital Collection for electronic full-text theses and dissertations 2014*. Bangkok: Office of the Higher Education Commission.

Walailak University. (2016). *Meeting report of the Committee for System Development of Full-Text Documents in Electronic Form, 28 January 2016*. Bangkok: Office of the Higher Education Commission.

Wang, X. (2014). Guidelines for implementing ETD programs: Roles and responsibilities. In M. Schultz, N. Krabbenhoeft, & K. Skinner (Eds.), *Guidance documents for lifecycle management of ETDs* (pp. 14-44). Atlanta: Educopia Institute.