FRBRisation of Koha in the Context of CMARC (a UNIMARC-derived format)

Naicheng Chang
General Education Center, Tatung University
Taiwan, China
E-mail: ncchang@ttu.edu.tw

Yuchin Tsai
National Center for High-performance Computing
National Applied Research Laboratories
Taiwan, China
E-mail: thomas@nchc.org.tw

Meeting:

187 — *Advancing UNIMARC: alignment and innovation* **— IFLA UNIMARC Programme (UNIMARC)**

Abstract:

Functional Requirements of Bibliographical Records (FRBR) is a conceptual model developed by IFLA Study Group on the Functional Requirements for Bibliographic Records in 1998 (IFLA, 1998). The new Resource Description and Access (RDA) rules which are expected to replace AACR as the world's leading cataloguing code are being based on these, and this will have serious implications for MARC formats. Implementing the model in rules and then adapting MARC formats accordingly have substantial help in the estimate of the cost of cataloguing work and the estimate of the amount of work to be done to convert the current records to FRBRized structures. More importantly, the FRBR model provides a new direction on developing a new generation Online Public Access Catalogue (OPAC) display pattern and on new methodology of searching/retrieving bibliographic records. This research project incorporates the FRBR model into the Koha Library Management Systems (LMS), and develops a new generation FRBR OPAC in Chinese. The research achievements include 1.developing FRBR application function library called LibFRBR which is used to convert existing records of different types of material into FRBRized structures in Koha, 2.designing a feasible Chinese FRBR OPAC user interface, 3.designing a feasible Chinese FRBR managing

1. Background

Functional Requirements of Bibliographical Records (FRBR) is a conceptual model developed by the IFLA Study Group on the Functional Requirements for Bibliographic Records in 1998 (IFLA, 1998). The aim of the FRBR framework is to produce bibliographic records which provide precise information to answer user needs. FRBR has been implemented by very few of the internationally available library systems and even then only as a pilot. However, the new Resource Description and Access (RDA) rules which are expected to replace AACR as the world's leading cataloguing code are being based on FRBR and this will have serious implications for MARC formats.

MARC is an exchange format and there is no standard way of exchanging work level data because it was embedded in different ways in records formulated according to Anglo-American Cataloguing Rules. In new rules formulated according to RDA, it will also be necessary to see how best user tasks' needs are to be fulfilled.

Presently, the issue of how to implement FRBR in a real IT system remains unclear. This research project incorporates the FRBR model into Chinese MARC (CMARC) implemented in the Open Source Software (OSS) Library Management Systems (LMS) Koha System. It includes the methodology for inputting bibliographic records according to FRBR. This will yield estimates of the amount of work to be done to convert the current records and the amount of manual (cataloguer) intervention needed. This in turn can be converted into an estimate of the costs of this in catalogues of different sizes.

Furthermore, the FRBR model provides a new direction on developing a new generation Online Public Access Catalogue (OPAC) display pattern and on new methodologies of searching/retrieving bibliographic records, in which this research project applies technology from the concept of Web2.0 such as TimeSheet Style.

FRBR Model

FRBR is based on the entity-relationship model and is defined by four entities, *work*, *expression*, *manifestation*, and *item*, with three kinds of relationships, primary, responsibility, and subject relationships associated with the four entities as shown in Figure 1. *Work* and *expression* are defined to reflect intellectual or artistic content; *manifestation* and *item* are defined to reflect physical form. The first of primary relationships indicates that a *work* is "realized through" *expression*. Similarly, the relationship connecting *expression* with *manifestation*, indicating that an *expression* is "embodied in" a *manifestation*; the last relationship identifies the *manifestation* "exemplified by" an individual *item* for ensuring that all copies (i.e., *items*) of the same *manifestation* are linked to that *manifestation*. The responsibility relationship indicates that the entities in the second group (*person* and *corporate body*) are linked to the first group by four relationship types; the "created by"

relationship that links both *person* and *corporate body* to *work*; the "realized by" relationship that links the same two entities to *expression*; the "produced by" relationship that links them to *manifestation*; and the "owned by" relationship that links them to *item*. In the case of subject relationships, the FRBR model represents a set of entities of which each may serve as the subject of a work and may include concept, object, event, and place.
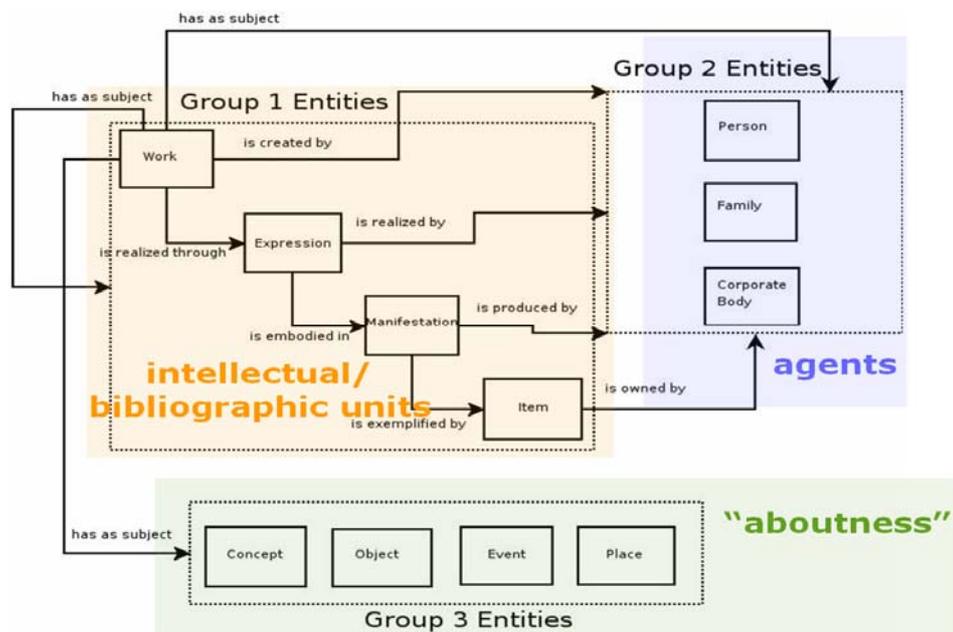

Figure 1: FRBR entity-relationship model (Childress, 2006)

There are additional relationships between group 1 entities which are not shown in the entity-relationship model, and these are to show how the relationships operate in the context of the four primary entities in the model (i.e., *work, expression, manifestation,* and *item*). For instance, in the work-to-work relationships, the adaptation relationship type involves *works*: the intellectual or artistic content of one *work* has been judged sufficiently different from the other to constitute a separate *work*.

Koha and Koha-Taiwan

The significant cost savings that can be achieved through taking the open source route are increasingly important given the library financial climate. Koha has been one of the longer-established OSS Library Management Systems in the US (Breeding, 2009). And it has started to attract attention in the UK (Bissels and Chandler, 2010).

Koha-Taiwan (2011) serves Koha Chinese users around the world, and is the only Koha

community that currently has been developed in Asian countries. It has its own website (http://koha-tw.org) (2011) maintained by the affiliation of one of our authors , National Center for High-performance Computing, National Science Council, Taiwan. Starting in 2006, Koha-Taiwan continues to release Chinese Koha initially with version 2.2.9 and updates.

Koha-Taiwan is a successful model around the world in terms of its keeping up with cutting edge technologies especially with trends of LMS. It looks at issues in several aspects related to CMARC, for example, issues of different scripts in the same record (in MARC 21 and CMARC and Chinese internal codes (i.e. double byte character set)) when implementing Koha (Chang, Tsai, and Hopkinson, 2009). Also, it looks at issue of implementing UNIMARC, CMARC, and MARC 21 on Koha (Chang and Tsai, 2010). A further example is to develop cloud cataloguing technology based on the concept of Software as a Service (SaaS) and Service-Oriented Architecture (SOA) to create a user-friendly and flexible cataloging interface. This cloud cataloguing infrastructure will apply results from this FRBR research.

This research is a Chinese FRBRrized Koha prototype project with Web 2.0 concept. This includes an entity-relationship model based FRBR application function library called LibFRBR which is used to convert existing records of different types of material into FRBRized structures in Koha, and a new generation Online Public Access Catalogue (OPAC) display pattern and a new methodology of searching/retrieving bibliographic records. The FRBR-based systems are well suited to Taiwan's Chinese Machine-Readable Cataloguing format and Chinese Cataloguing rules. The results of this research will be realised through Koha-Taiwan as a General Public License (GPL) for further application within Koha community.

2. Koha_ LibFRBR testbed

The platform of the testbed is Koha 3.02.03.000 released at the end of 2010 running under the Linux Ubuntu operating system. In the testbed, 500 CMARC bibliographic records with subjects mainly on history and information provided by the National Center for High-performance Computing Library were tested.

The testbed was built to examine mainly four levels 1) building a FRBR application function library called LibFRBR to convert existing records into FRBRized structures in Koha; 2) practicing mapping algorithm between CMARC/MARC 21 and FRBR in the Koha environment (Tsai, 2006; Delsey, 2002); 3) Designing a new generation Chinese FRBR OPAC user interface in Koha environment; 4) developing a feasible Chinese FRBR working interface for cataloguers.
We applied LibFRBR to FRBRize in practice bibliographic records in several aspects described below.

(1) Edit, add, and delete data of entities of 3 groups

(2) Edit, add, and delete data of attributes. This includes 12 attributes from *work*; 25 attributes from *expression*; 38 attributes from *manifestation*; 9 from *item*; 5 attributes from *corporate body*; and each 1 attribute from *concept object, event,* and *place*.

(3) Edit, add, and delete data of relationships of "realized through","embodied in", and "exemplified by" in group 1; "created by","realized by", "produced by", and "owned by" relationship in group 2; and subject relationships in group 3.

(4) cross groups/entities relationships with relation types of reproduction, successor, adaptation, whole/part, revision, and so on.

(5) Each entity has its own data sheet which records information of ID numbers and attributes. Each relationship is recorded in the relationship data sheet which includes ID numbers, relation types, and time.

(6) Web-based viewer template to precede all the functions in LibFRBR.

The system structure is outlined in Figure 2.
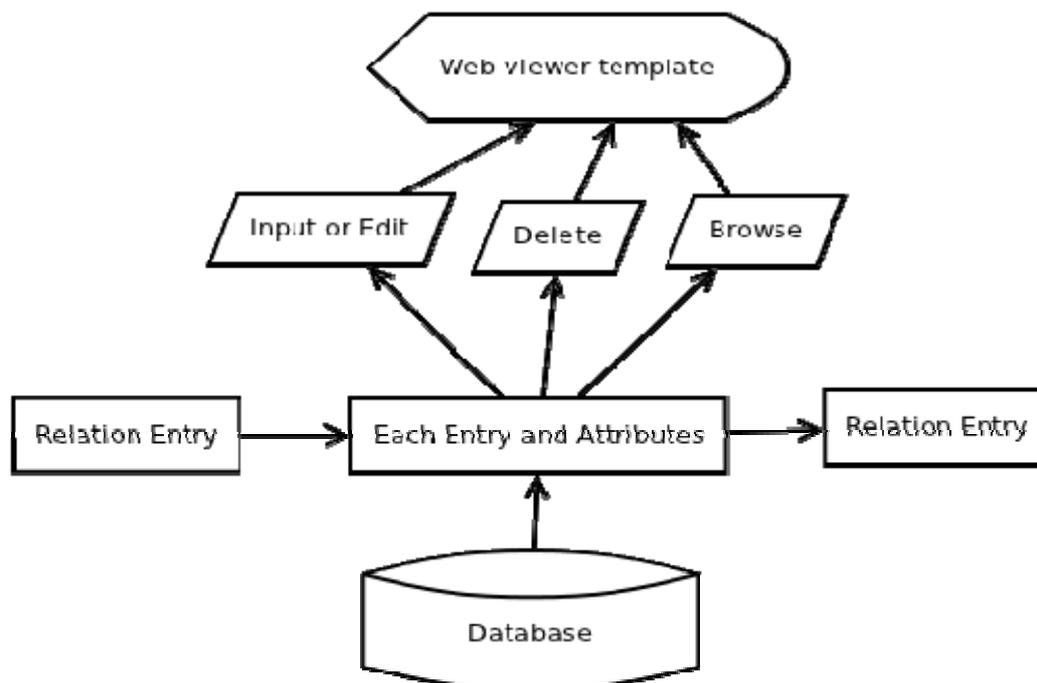
Figure 2: System outline of Koha_LibFRBR

3. Examples of bibliographic relationship types in the FRBR model and mapping of CMARC to FRBR (in Figure 6)

1. Example of creating *expression* "realized through" *work*

      step1. 'create' *work*

      step2. Input data for *work* attribute and click Insert to save *work*

      step3. Select 'add new relation type' in top level and choose "realized through"

      step4. Input data for *expression* attribute and save

      step5. Back to *work* and rerun step4 and step5 for other *expression*

   The result is shown in Figure 3 below.

Figure 3: Creating *expression* "realized through" *work*

2. Example of relationship between *person* and *work* is shown in Figure 4.

Figure 4: Relationship type between *person* and *work*

3. Example of adaptation relationship between two *work*s as shown in Figure 5:
Example

    w1 司馬光-資治通鑑

    has an adaptation ←

    → is an adaptation of

    w2 柏楊版資治通鑑

step1. Add *work* (W1)

step2. Input data for *work* (W1) attribute and save

step3. Select 'add new relation type' and choose "has adaptation (*work*)"

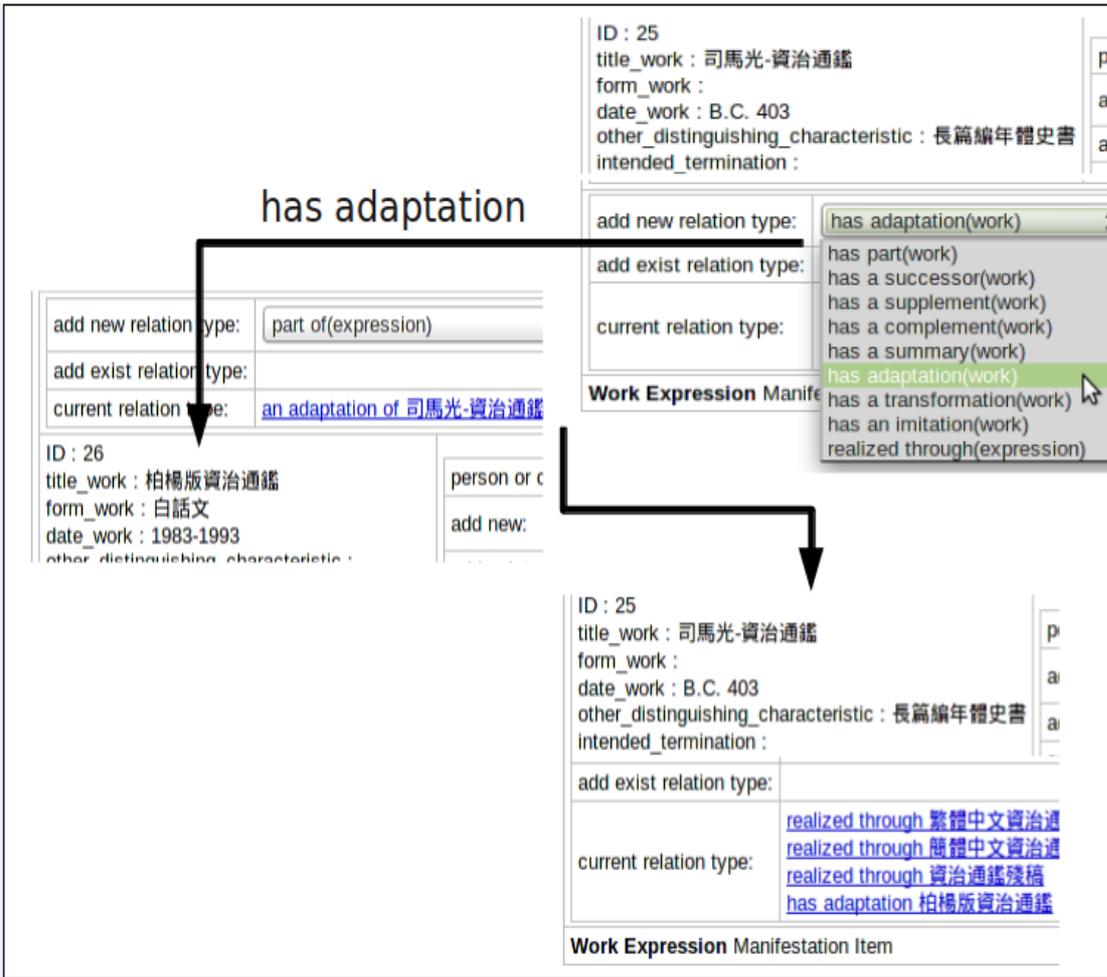step4. Input data for *work* (W2) attribute and save

Figure 5: Adaptation relationship between two *work*s

001　　　[skip]	607　　2 -> csh　　　item-item_identifier
010 1 a -> 9573208601	a -> 中國　　x ->歷史
manifestation-manifestation_identifier	700 1 s -> 宋　　　　a -> 司馬
b ->平裝 manifestation-terms_availability	b -> 光　　　　4 -> 原著
d -> NT $ 180	manifestation-statement_responsibility
manifestation-terms_availability	702 1 a -> 柏楊
100　　a -> 19950612d1992xxxxxxxm	4 -> 譯
y0chiy09xxxxx	manifestation-statement_responsibility
expression-language_expression	801 0 a -> cw
manifestation-date_publication_distrib	manifestation-place_publication_distribution
105　　a -> z　000yy　work-form_work	b -> 國家高速電腦中心
expression-form_expression	item-item_identifier
200 1 a -> 蕭鸞眼淚	item-provenance_item
manifestation-title_manifestation	c -> 19950612
f -> 司馬光原著	item-treatment_history
manifestation-statement_responsibility	809　　p -> BOOK
g -> 柏楊編譯	d -> 610.23
manifestation-statement_responsibility	e -> 4646
205　　a -> 八版	l -> v.34
manifestation-edition_issue_designation	856　　u -> http://www.ylib.com.tw/home.asp
210　　a -> 台北市	manifestation-manifestation_identifier
manifestation-place_publication_distribution	z -> YLib 遠流博識網
c -> 遠流	2 -> http
manifestation-publisher_distributor	manifestation-mode_access
d -> 民 81	805　　a -> 國家高速電腦中心
manifestation-date_publication_distribution	b -> LB
215 1 a -> 269 p.	c -> 000984　　　item-item_identifier
manifestation-extent_carrier	d -> 610.23　　　item-item_identifier
d -> 21 cm. manifestation-dimensions_carrier	e -> 4646　　　　item-item_identifier
225 2 a -> 柏楊版資治通鑑	p -> BOOK
manifestation-series_statement	l -> v.34　　　　item-item_identifier
v -> 34 manifestation-series_statement	f -> D00CYJ00
461 0 1 -> 2001　　a -> 柏楊版資治通鑑	item-marks_inscription
v -> 34	z -> CR
605　　2 -> csh　　　item-item_identifier	m -> 0
a ->資治通鑑　　　　x->譯文	t -> CCL

Figure 6: Mapping of CMARC to FRBR

Koha FRBR librarian's interface and FRBR OPAC

Instead of the single flat record concept underlying current cataloging standards, FRBR uses an entity-relationship model of metadata for information objects. This difference is also reflected in the FRBR working interface for cataloguers. This research develops a feasible Chinese FRBR working interface for cataloguing librarians. Cataloguers can input/output FRBR bibliographic records in an efficient working model showed in Figure 7. Furthermore, the Koha OPAC as shown in Figure 8 which displays pattern with technology from the concept of Web2.0 called TimeSheet Style. The timelines shows the publication years of *work*s.

Figure 7: Koha Chinese FRBR working interface for cataloguers

Figure 8: Chinese FRBR based OPAC with Timelines

5. Evaluation and discussion

A. Datafields that fail to be mapped to FRBR

There are cases either in CMARC3 or MARC 21 of datafields that can not be mapped to FRBR. In the case of CMARC3, our research shows that there are total of 133 fields with 199 subfields that fail to be mapped to FRBR. We must consider solutions to minimize the loss of data in bibliographic records due to a low mapping rate. We propose below four solutions based on practices.

a. Skip, i.e. lose the data

b. alternative mapping rules with looser definition

c. customized mapping rules and create home fields/subfields

d. building tables for data in bibliographic records where it goes to, if institutions are not able to do a, b or c

B. Transforming field codes to datafields

FRBR format provides detailed descriptions of datafields and relationships; nevertheless, traditional MARC formats apply codes for their bibliographic data. Our practice shows that the system presents codes instead of datafield names/value after mapping. We suggest that this needs programming to transform field codes to datafields so that the system interface is readable for cataloguing librarians when working with a cataloguing interface and for end users using an OPAC.

C. MARCXML in LibFRBR

The National Central Library in Taiwan has developed and maintained a framework for working with CMARC3 data in an XML environment. The CMARC3XML framework is intended to be more flexible and extensible than original CMARC3 (ISO 2709) to allow users to work with MARC data in ways specific to their needs. In the FRBR environment, it contains not only bibliographic record data, but also numbers of relationships of entities and attributes of bibliographic record data. Our research shows that it could be a critical issue when a bibliographic record database becomes substantially large in terms of representing fully the complex relationships in FRBR environment. It is our view that a native XML-based database instead of a relational database could be a solution for FRBR_based libraries.

D. Strengthen searching function with FRBR in OPACs

Web-based OPACs demonstrate advances on traditional OPACs, especially in terms of their potential to integrate many document types and sources via a single interface. The OPACs offer the function of keyword searching by author, title, subject, keyword, call number, and others. Also, OPACs offer function of natural language searching which is improved by combining with other techniques such as boolean searching and proximity searching. With the feature of 'relationship' in FRBR, FRBR-based OPACs could strengthen their searching ability to help readers to find, identify, select, and obtain materials.

E. LibFRBR system performance evaluation

This project is a prototype research project with the object of locating obstacles and difficulties when building prototype services utilizing the FRBRized database. We focus on input/output FRBR bibliographic records data, but not LibFRBR system evaluation. Nevertheless, we have proceeded with an in-depth user survey to find out what librarians and readers think about the Koha FRBR model.

In the continuation of this research project, we plan to combine AACR3 and RDA to re-define the database fields in order to optimize the system performance, and will integrate FRBR to more aspects within Koha system such as data input/output, keyword searching

features, and extend to circulation module.

6. Conclusion

At the time of writing there is no real FRBR practice in any LMS in Taiwan nor in the Koha official website;; therefore, we expect this research project could assist the worldwide library and information community and integrated library system community to understand the concept of implementing the FRBR in terms of feasibility and future developments; help to plan actively FRBR-based systems well-suited to Chinese Machine-Readable Cataloguing format and Chinese Cataloguing rules; provide the Koha user community with a first generation FRBR-based Koha system to cope well with current trends in integrated library system development; and stimulate more worldwide FRBR-related research and implementations to accelerate the development of FRBR-based integrated library systems.

References:
Bissels, G. and Chandler, A., (2010).Two years on: Koha 3.0 in use at the CAMLIS Library, Royal London Homoeopathic Hospita, *Program: electronic library and information system*, 44(3): 283-290.

Breeding, Marshall, (2009). Investing in the future: automation marketplace 2009. *Library Journal*, 134(6):26-39. Available at: http://www.libraryjournal.com/article/CA6645868.html

Chang, Naicheng and Tsai, Yuchin, (2009). *An evaluation of multi-language/multi-script functions in Koha*. Paper presented at World Library and Information Congress: 75th IFLA General Conference and Council, August 23-27, 2009, Milan, Italy. Available at: http://www.ifla.org/files/hq/papers/ifla75/135-chang-en.pdf

Chang, Naicheng Tsai, Yuchin, and Hopkinson, Alan, (2010).An evaluation of implementing Koha in a Chinese language environment, *Program: ElectronicLibrary and Information Systems*, 44(4): 342-356.

Childress, Eric, (2006). *FRBR and OCLC research.* Available at: http://www.oclc.org/research/presentations/childress/20060410-uncch-sils.ppt

Delsey, T., (2002). *Functional analysis of the MARC 21 bibliographic and holdings formats*. Available at: www.loc.gov/marc/marc-functional-analysis/analysis.pdf

IFLA Study Group on the Functional Requirements for Bibliographic Record,

(1998). *Functional Requirements for Bibliographic Records: Final Report.* The Hague, Netherlands : International Federation of Library Associations and Institution. Available at: http:// www.ifla.org/VII/s13/frbr/frbr.htm

Koha-Taiwan, (2011). *Koha-Taiwan at Google Groups*. Available at: http://groups.google.com/group/kohataiwan

Koha-tw.org, (2011). Available at: http://trac.koha-tw.org/#KOHA 台灣

Tsai, Chiachi,(2006). *A study of implementation of FRBR_based OPAC*. Unpublished dissertation of Graduate Institute of Library and Information Studies, National Taiwan Normal University.