

Constructing Metadata Schema of Scientific and Technical Report Based on FRBR

Xiaozhu Zou¹, Siyi Xiong¹, Zhi Li¹ & Ping Jiang²

¹ Institute of Scientific and Technical Information, Nanjing University of Aeronautics and Astronautics, Nanjing, China

² Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Correspondence: Siyi Xiong, Library, Nanjing University of Aeronautics and Astronautics, Nanjing, 211110, No. 169 Sheng Tai West Road, China. Tel: 025-84892935, 86-182-5198-0566. E-mail: yiyi50211@163.com

Received: February 28, 2018

Accepted: February 12, 2018

Online Published: March 27, 2018

doi:10.5539/cis.v11n2p34

URL: <http://dx.doi.org/10.5539/cis.v11n2p34>

The research is financed by the Library of Nanjing University of Aeronautics and Astronautics in 2014 service innovation work project "Research on Helicopter science and technology report depth integration and characteristic database construction based on FRBR, project number: TSG201401" and Graduates Scientific Research Innovation Project of Jiangsu Province "Research on characteristic database Construction of science and technology report in helicopter topic based on FRBR, project number: kfj20170912".

Abstract

Scientific and technical report is an important document type with high intelligence value. But the resource distribution of different carrier forms of scientific and technical report is not integrated and the resource description is not deeply and specific enough to the report document type, which influences the information searching accuracy and efficiency for users. Functional Requirements for Bibliographic Records (FRBR), an emerging model in the bibliographic domain, provides interesting possibilities in terms of cataloguing, representation and semantic enrichment of bibliographic data. This study employs the FRBR conceptual model and entity-relationship analysis method to design in-depth descriptive metadata schema of scientific and technical report by analyzing the entities and mapping the bibliographic attributes corresponding to the characteristics of report, which can help to integrating and disclosing scientific and technical report resources.

Keywords: metadata schema, FRBR, scientific and technical report, information integration

1. Introduction

Scientific and technical report (hereafter referred to as "report") collections are important intelligent resources to the research users. Through investigation of the resource developing situation of report (Yantao,2016), there are mainly three problems:

- 1)The report collections have several carrier forms such as microform, prints and electronical form, which increases the inconvenience for users.
- 2)There are no uniform resource discover platform of report collections to reveal the access for users.
- 3)The description of the report collections resources is not complete and deep, especially on the subject index.

So there are needs to develop a database and an searching platform of report collections to provide users the comprehensive and convenient access to the resources, the basis of which is to design a descriptive metadata schema. Metadata schema of scientific and technical report is structural data describing report collections to help users retrieve and identify the required resources, to carry out detailed and comprehensive cataloging of data units and to help resource management and long-term preservation.

FRBR, short for Functional Requirements for Bibliographic Records, is a report released in 1998 by International Federation of Library Associations (IFLA). Since then the effort to develop and apply FRBR has been extended in many innovative and experimental directions. The aim of the FRBR is to determine the needs of users of library systems, and based on their needs, to form functional requirements for these systems (Rudic,2011). The entities as well as relationships between them, required for a library system/catalog to meet

the needs of users, are suggested.

FRBR offers a flexible framework to represent any cultural content in library, new benefits for improving searching and visualization (Decourselle, Duchateau, & Lumineau, 2015). The last decade has seen the emergence of application studies of FRBR. One of the earliest survey was conducted at the Online Computer Library Center (OCLC) focused on the benefits of FRBR cataloguing, it also reviews the major solutions which implemented this model (Hickey,2005). The representative FRBR projects include the Music Australia project adopting the FRBR information model (Kerry, 2005), Red Light Green Union Catalogue in FRBR display, and the FRBR work-set algorithm develop project conducted by OCLC (Thomas, 2018). The FRBR report notes that the study endeavors to be comprehensive in terms of the variety of materials that are covered. FRBR proposed a flexible framework for the description and identifying of various resources, which is applicable to design metadata schema of report collections.

2. The Characteristics of Science and Technical Report

Scientific and technical report, as one of top ten document types, is born in the early 20th century and develops rapidly after the Second World War. According to America National Information Standards Organization, Scientific and technical report collections convey the results of basic or applied research and support decisions based on those results. A report includes the ancillary information necessary for interpreting, applying, and replicating the results or techniques of an investigation. The primary purposes of such a report are to disseminate the results of scientific and technical research and to recommend action.

Unlike other document types, scientific and technical report collections have some unique characteristics:

- 1) It has a unique number, known as the report number which is usually composed of the initials of the performing organization or the sponsor followed by serial numbers. A report may have a contract or grant number and an accession or acquisition number.
- 2) It is not usually published or made available through commercial publishing which often takes independent and complex preview process, so it doesn't need to wait for a long publishing process, which makes scientific and technical report collections have a strong timeliness.
- 3) Its contents usually focus on frontier projects, revealing the latest research results. Many exploratory researches often appeal first in scientific and technical report collections.
- 4) It may be written for an individual or organization as a contractual requirement to recount a total research process, usually attaching with detail data, graphics and facts and including full discussions of unsuccessful approaches.
- 5) Its distribution may be limited or restricted, its readership may be limited, and its contents may include classified, proprietary, or copyrighted information.

3. The Concept of FRBR

FRBR defines a structured framework based on the Entity-Relationship (E-R) model which is commonly used in relational database. Referring to the E-R model, FRBR abstracts the key objects that users are interested in the bibliographic record from a universal perspective to define the entities, sums up the various characteristics that are needed to identify the entities to define the attributes and tease out the main link between entity and attribute to define the relationships.

As shown in Figure 1, the FRBR E-R model includes three classes of entities, labeled Group1, Group2, and Group3. Group 1 comprises the products of intellectual or artistic endeavor that are named or described in bibliographic records: Work, Expression, Manifestation, and Item. According to the FRBR report released by IFLA, the entity Work is defined as "A distinct intellectual or artistic creation," Expression as "the intellectual or artistic realization of a work in the form of alphanumeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms," Manifestation as "the physical embodiment of an expression of a work" and Item as "a single exemplar of a manifestation" (IFLA,2018). Group2 comprises those entities responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of such products: Person and Corporate body. Group3 comprises an additional set of entities that serve as the subjects of intellectual or artistic endeavor: Concept, Object, Event, and Place. The three groups of entities connect with each other through the attributes, which reveals the hierarchical relationships and reflects the transverse connection between entities, thus constitutes a three-dimensional network model.

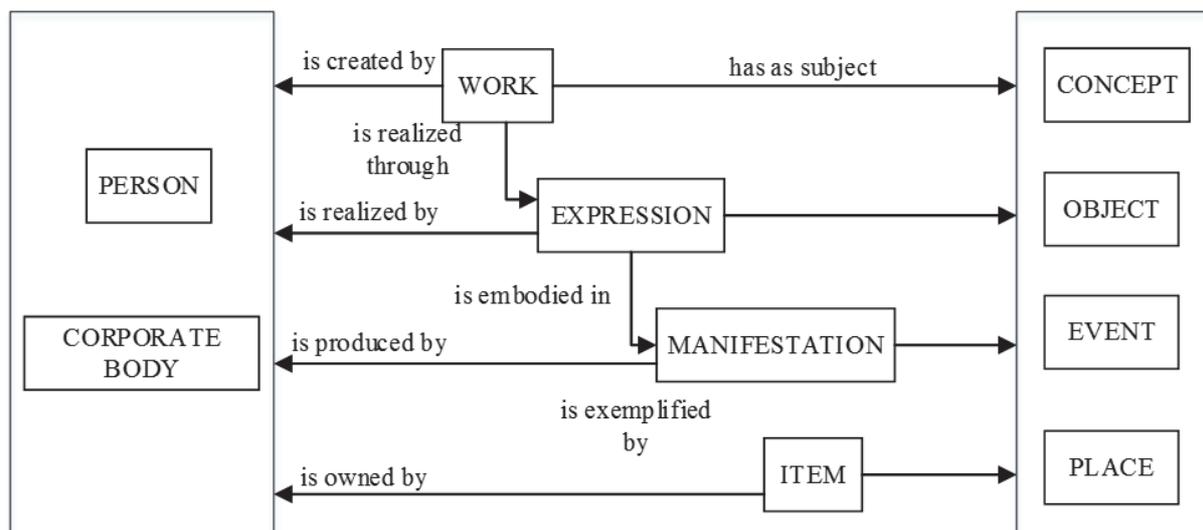


Figure 1. The FRBR entity-relationship model

4. Metadata Schema Design of Scientific and Technical Report Based on FRBR

4.1 The Entity Analysis based on FRBR

Metadata is the basis of information structure and the core data to reveal the information both in traditional literature and digital resources. The organization and the manifestation of metadata directly influence the user's understanding of the resources and the retrieval efficiency. Designing the metadata based on FRBR needs to identify the entities of scientific and technical report collections firstly. The first group focuses on a report literature itself consisting of four entity-Work, Expression, Manifestation and Item. The second group is about those responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of the entities in the first group including person (an individual) and corporate body (an organization or group of individuals and/or organizations), for example the authors sponsor and accession organizations of a report. The third group is about the subjects of a report, divided into concept (an abstract notion or idea), object (a material thing), event (an action or occurrence), and place (a location).

Take the report "Pilot-in-the-loop Method Development" as an example to implement the entity analysis based on FRBR (W/E/M/I stand for Work/Expression/Manifestation/Item):

W Pilot-in-the-loop Method Development

E1 Pilot-in-the-loop Method Development (report number: ADA609969)

MI Printed version

I1 A copy preserved in the library

The Work entity "Pilot-in-the-loop Method Development" may have several Expression which are report with the same report name but different report number. One Expression may have several Manifestation, like printed or electronic version. One may have several Item, like a copy preserved in the library or an electronic text offered by a database.

In this research, the scope of data analysis focuses on scientific and technical report collections, of which the work entity may have several expression entities (for example several report versions in different language), the expression entity may have several manifestation entities such as printed, microform and digital version, and the manifestation entity may have several items.

4.2 Mapping the Metadata Elements

Each entity type is assigned a set of attributes. Work has attributes such as title and form, Expression has a language attribute (translations of the same work are different Expressions), Manifestations has attributes like typeface and Items has attributes such as condition and location. Users can search the wanted entity through the attributes. The attributes are defined through logical analysis of the bibliographic records data to extract terms which can reflect the characteristics of the entities from user's perspectives. With the entities identified, the metadata elements can be selected after analyzing the correspondence with the attributes of different entities.

The attributes defined for each of the entities in the model will not necessarily be exhibited by all instances of that particular entity type. In some case the logical attribute parallels an individual metadata element, but in most cases the logical attribute represents an aggregate of discrete data elements. The attributes of entities have been investigated in order to map the metadata elements to the attributes applicable to scientific and technical report collections.

4.2.1 Metadata Elements Mapping to the Attributes of Group 1

As defined in FRBR, there are 84 logical attributes of the first group entities, including 12 of a work, 25 of an expression, 38 of a manifestation, 9 of an item. Excluding the attributes for musical work and cartographic work, there are 7 logical attributes of an expression left as the following: title of the work, form of work, date of the work, other distinguishing characteristic, intended termination, intended audience, context for the work. Table 1 shows the metadata elements mapping to the logical attributes of a work corresponding to the characteristics of scientific and technical report collections.

Table 1. Metadata Elements Matched with The Attributes of Work

Logical attributes	Metadata elements	Description
title of the work	title	the word, phrase, or group of characters naming the work
date of the work	year	the year that the work was originally created
other distinguishing characteristic	local doi	the object unique identifier in local

Excluding the attributes for serial, musical notation, recorded sound, cartographic image/object, remote sensing image and graphic or projected image, there are 12 logical attributes of an expression left as the following: title of the expression, form of expression, date of expression, language of expression, other distinguishing characteristic, extensibility of expression, revisability of expression, extent of the expression, summarization of content, context for the expression, critical response to the expression, use restrictions on the expression. Table 2 shows the metadata elements mapping to the logical attributes of an expression corresponding to the characteristics of scientific and technical report collections.

Table 2. Metadata Elements Matched with The Attributes of Expression

Logical attributes	Metadata elements	Description
form of expression	report type	such as prepare report collections, intermediate report and summary report
language of expression	language	the language of the context
other distinguishing characteristic	report no	the specific report number
summarization of content	abstract	the summarization of the content

Excluding the attributes for serial, sound recording, image, visual projection and remote access electronic resource, there are 22 logical attributes of a manifestation left as the following: title of the manifestation, statement of responsibility, edition/issue designation, place of publication/distribution, publisher/distributor, date of publication/distribution, fabricator/manufacturer, series statement, form of carrier, extent of the carrier, physical medium, capture mode, dimensions of the carrier, manifestation identifier, source for acquisition/access authorization, terms of availability, access restrictions on the manifestation, reduction ratio (microform), polarity (microform or visual projection), generation (microform or visual projection), system requirements (electronic resource), file characteristics (electronic resource). For user's task, some specific external attributes can be omitted, such as reduction ratio, polarity and generation. Table 3 shows the metadata elements mapping to the logical attributes of a manifestation corresponding to the characteristics of scientific and technical report

collections.

Table 3. Metadata Elements Matched with The Attributes of Manifestation

Logical attributes	Metadata elements	Description
date of Publication/distribution	production date	date of Publication/distribution
fabricator/manufacture	manufacturer	fabricator/manufacture
form of carrier	resource type	electronical/printed/microform
manifestation identifier	manifestation id	manifestation id in local
terms of availability	availability	full text or abstract
mode of access	access mode	download directly, document delivery or borrowing
access address	urls	access address in database

The logical attributes of an item defined in FRBR are the following: item identifier, fingerprint, provenance of the item, marks/inscriptions, exhibition history, condition of the item, treatment history, scheduled treatment, access restrictions on the item. Table 4 shows the metadata elements mapping to the logical attributes of an item corresponding to the characteristics of scientific and technical report collections.

Table 4. Metadata Elements Matched with The Attributes of Item

Logical attributes	Metadata elements	Description
item identifier	item id	the item identifying number
provenance of the item	source db	source database
scheduled treatment	create date	create date
treatment history	update date	update date
access restrictions on the item	item url	access address of the item

4.2.2 Metadata Elements Mapping to the Attributes of Group 2

As defined in FRBR, the logical attributes of a person are name of person, dates of person, title of person and other designation associated with the person. The logical attributes of a corporate body are name of the corporate body, number associated with the corporate body, place associated with the corporate body, date associated with the corporate body and other designation associated with the corporate body. For the Group 2 entities, we focus on the responsibility relationship of the work entity, Table 5 shows the metadata elements mapping to the logical attributes of Group 2 corresponding to the characteristics of scientific and technical report collections.

Table 5. Metadata Elements Matched with The Attributes of Group 2

Logical attributes	Metadata elements	Description
name of person	author	the name of author
other designation associated with the person	address, email	post/email address
name of the corporate body	organization	the responsible corporate body

4.2.3 Metadata Elements Mapping to the Attributes of Group 3

Scientific and technical report collections are generally the results of certain research tasks and projects, so a report has both the characteristics of literature and project. We need to build a corresponding relationship between the report and the project, which is concerned by the users. The Group 3 entities is about the subject of the content, the logical attributes of which include term for the concept, term for the object, term for the event and term for the place. Table 6 shows the metadata elements mapping to the logical attributes of Group 2 corresponding to the characteristics of scientific and technical report collections.

Table 6. Metadata Elements Matched with The Attributes of Group 3

Logical attributes	Metadata elements	Description
term for the concept	subject classification	the classification subject or number
term for the object	model alias	the research object, for example the plane model
term for the event	project name/no	project name/no
term for the place	nationality	research nation

5. Conclusion

In-depth disclosure is the premise of resources polymerization. The FRBR model were derived from a logical analysis of the data that are typically reflected in bibliographic records. FRBR supports a more systematic and meticulous analysis of the attributes of report collections and inspire the way that metadata might be aggregated to combine different element sets relating to one resource.

This study designs descriptive metadata applicable for report collections based on FRBR model through the E-R method and selecting 27 attributes mapping to the metadata elements of report, which is the basis to develop the uniform resources integrating database of report collections.

Acknowledgments

This paper is a revised version of a keynote lecture presented at the 5th International Academic Conference for Graduates at Nanjing University of Aeronautics and Astronautics, 2017. The financial support of Library of Nanjing University of Aeronautics and Astronautics and Graduates Scientific Research Innovation Project of Jiangsu Province is appreciated.

References

- Decourselle, J., Duchateau, F., & Lumineau, N. (2015). A Survey of FRBRization Techniques. *19th International Conference on Theory and Practice of Digital Libraries, 9316*, 185-196. https://doi.org/10.1007/978-3-319-24592-8_14
- Hickey, T. B., & O'Neill, E. T. (2005). Frbrizing OCLC's WorldCat. *Cataloging & Classification Quarterly, 39*(3-4), 239-251. https://doi.org/10.1300/J104v39n03_15
- IFLA. (2018). Functional Requirements for Bibliographic Records. Retrieved from http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm
- Kerry, K. (2005). The Austlit Gateway and scholarly bibliography: a specialist implementation of the FRBR. *Cataloging & Classification Quarterly, 39*(3-4), 87-102. https://doi.org/10.1300/J104v39n03_07
- Rudic, G. (2011). Constraints modeling in FRBR data model using OCL. *AIP Conference Proceedings, 1389*(1), 813-816. <https://doi.org/10.1063/1.3636857>
- Thomas, B. H., Jenny, T. (2018). *FRBR Work-Set Algorithm Version2.0*. Retrieved from <http://www.oclc.org/content/dam/research/activities/frbralgorithm/2009-08.pdf>.
- Yantao Q. (2016). *Study on aggregation patterns of foreign scientific and technical reports of helicopter based on FRBR* (Master dissertation). Nanjing University of Aeronautics and Astronautics, Nanjing, China.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).