Metadata Standard of Theses and Dissertations according to the Entity-Relationship Model

Cíntia de Azevedo Lourenço* and Lídia Alvarenga**

Universidade Federal de Minas Gerais, Av. Presidente Antonio Carlos, 6627,
Pampulha 31.270-010, Belo Horizonte MG, Brasil,
*http://www.eci.ufmg.br/cintialourenco <cal@eci.ufmg.br>,
**<lidiaalvarenga@eci.ufmg.br>

Cíntia de Azevedo Lourenço graduated in librarianship (1991) and also earned a masters degree (1998) in librarianship at Pontifícia Universidade Católica de Campinas, Brazil, PUC, and a doctoral degree (2005) in information science, at Federal University of the State of Minas Gerais, Brazil, UFMG. She is a graduate professor in the Information Science School of Federal University of Minas Gerais State , Belo Horizonte , Brazil . Her current research interests include document description, documents representation, cataloging, metadata and FRBR. She is a member of a institutional research group on these themes. Her curriculum vitae can be found at http://buscatextual.cnpq.br/buscatextual/visualizacv.jsp?id=K4773000E8.

Lídia Alvarenga graduated in librarianship (1968) at Federal University of the State of Minas Gerais, Brazil, UFMG, earned a masters degree (1984) in information science at Communication Faculty, Federal University of Rio de Janeiro, Brazil, UFRJ, and a doctoral degree (1996) in education, at Education School, UFMG. She is a graduate and undergraduate professor in the Information Science School of Federal University of Minas Gerais State, Belo Horizonte, Brazil. Her current research interests include knowledge organization, document representation and description, and bibliometric studies. She is a member of a institutional research group on these themes. Her curriculum vitae can be found at http://lattes.cnpq.br/9205421345814071.

de Azevedo Lourenço, Cíntia, and Alvarenga, Lídia. **Metadata Standard of Theses and Dissertations according to the Entity-Relationship Model.** *Knowledge Organization, 36(1)*, 30-45. 17 references.

ABSTRACT: With the automation of information systems and with the advent of digital libraries, norms, standards and techniques of library studies have been widely discussed, analyzed, reevaluated and reorganized. In this article the results of doctoral research, in which the Brazilian Metadata Standard for Theses and Dissertations (MTD-BR) was analyzed, is presented. This standard has been utilized in the digital Library of Theses and Dissertations Project, of the Instituto Brasileiro de Informação em Ciência e Tecnologia, IBICT (Brazilian Institute for Scientific and Technological Information), with the methodology of data modeling, according to the *Functional Requirements for Bibliographic Records* (*FRBR*), which is based on the Entity-Relationship Model. It was concluded that new studies should be carried out applying this methodology to other metadata standards, even if they are analyzed with other data modeling tools, such as the object-oriented model, and considering its relationship with the guidelines, principles and instruments of library studies.

1.0 Introduction

According to Mey (1987, 46) descriptive representation has a very clear function of identifying bibliographic items and an equally clear characteristic of adjusting catalogues to the universe of the user. According to the information science authors consulted here, until the 1980s, the development of descriptive representation was linked to the practice of cataloguing and of bibliography. It was always concerned with the identification of the essential and complementary elements necessary for a satisfactory universal document representation. But, in the 1990s, most international agencies in the area of library studies started to

be concerned with the improvement of descriptive representation since new types of document support, especially in support of digital media, emerged exponentially. This reality caused the appearance of an increasing need to adapt the codes and the practices of cataloguing to the changes, the new forms of electronic publication and the advent of the web access to information resources (IFLA, 1998).

The existence of universal cataloguing standards and the Machine Readable Cataloguing (MARC) exchange format, facilitated automation and exchange between bibliographic databases. Using computers the standardization allowed by one format guarantees data importation and exportation between different computer systems; in addition, it facilitates information retrieval and interoperability between different commercial hardware and software, allowing cooperative cataloguing. The Seminar on Bibliographic Records, held in Stockholm in 1990, was an important hallmark of these changes that contributed to the realization in the area of international library studies that, among other things, the constant pressure to have a "minimal level" of cataloguing needed to be carefully rethought and reevaluated. And this would have as the basis, not only the relationship between the data element of a record, but also, and specially, the needs of the users (IFLA 2003, 2). In this seminar, nine resolutions were adopted, and one of them led to studies of a basic descriptive representation core, that originated a review group at IFLA to define the *Functional Requirements for Bibliographic Records* (*FRBR*). The *FRBR* studies started in September, 1992 and were concluded in September, 1997.

The *FRBR* study group had as objectives (IFLA 1998):

> To delineate in clearly defined terms the functions performed by the bibliographic record with respect to various media, various applications, and various user needs .... to cover the full range of functions for the bibliographic record in its widest sense- i.e., a record that encompasses not only descriptive elements, but access points (name, title, subject, etc.), other "organizing" elements (classification, etc.), and annotations.

The analysis of the descriptive elements identified by the FRBR was developed with the use of semantic data modeling, as utilized in computer science through a tool called the Entity-Relationship Model (ERM). It is known that modeling techniques such as the entity-relationship and the object-oriented analyses are used in projects of systems development as a way to understand, in clearly defined conditions, the entities or objects with which an organization needs to keep information and the logical relationships between these entities or objects. Based on this semantic model, the FRBR Review Group identified entities and basic attributes of descriptive representation, and the following descriptive representation standards were utilized: ISBD (*International Standard Bibliographic Description*); GARE (*Guidelines for Authority and Reference Entries*); GSARE (*Guidelines for Subject Authority and Reference Entries*); UNIMAR (Marc Americano); AITF (*Categories for the Description of Works of Arts*).

However, *FRBR* did not intend to present to the library community a definite model of descriptive representation. Its main objective was rather to present a basic level for bibliographic records, a semantic model for the study and further development of standards of descriptive representation and cataloguing. Based on Delsey's thought it is possible to affirm that the practices of cataloguing and descriptive representation can be seen as refined techniques of data management, that are related not only to descriptive representation, but rather to the detailing of contents of digitalized documents in complete texts present in digital libraries. IFLA maintain a Study Group on *FRBR* that produce reports posted on IFLANET. Among terms of reference adopted in December 2007 include tasks of review and maintain the *FRBR* conceptual model on ongoing basis, incorporating revisions, when needed, in both entity-relationship and oriented object models. Study Group webpage include a bibliography on *FRBR* connected with studies and researches on these proposed models and applications (IFLA 2007). IFLA webpage include a bibliography on *FRBR* where could be found studies and researches on proposed models and applications. Literature records some experiments using algorithms to group existing bibliographic records into *FRBR* categories. A example is the Online Computer Library Center (OCLC) work investigating the relationship that cluster bibliographic itens into manifestations, expressions and works, using as prototypical work Shakespeare's *Hamlet* from a WorldCat records sample (Hickey at al,. 2002).

1.1 Digital libraries and librarianship

Before discussing the importance of descriptive representation in this new context, it is necessary to highlight some characteristics and context of digital libraries according to Alvarenga's thought (1999/

2003). Digital libraries did not appear from nowhere. Although the formal recognition of digital libraries is relatively new it is known that some traditional libraries, especially in economically and technologically more developed countries, have been working with technologies of electronic data processing, by the use of the online reference databases and the online public access catalogs (OPACs) since the end of the 1960s and beginning of the 1970s. The creation of the digital library was possible with the advent of new electronic technologies of information and communication that altered the division of work of various professional segments of society, such as knowledge producers, users and information professionals, directly reflected in the production processes, storage, treatment and retrieval of documents and information, radically altering their work processes and final products. Different types of documents, new or not, now share the same space of materialization in the digital environment. Authors of texts, sounds and images share today the same production space. As products of this complex scenario the so-called digital objects exist in the web; they are an unlimited and changing universe constituted of new documents. The digital environment currently can be seen as the "without precedent" space for the record and retrieval of text, sound and image documents; this space, presented itself as a great possibility for storage, memory and formats, also started to require new facilitating elements for their retrieval. In this new context, objects and catalogues are not separate, but share the same space – the digital space (Alvarenga 2003, 2-17).

Considering all of these changes, experience based on the accumulatied knowledge in the field of library studies and information science is now useful. This moment requires integrated, professional and cooperative action. The studies and instruments in the field of descriptive representation are basic needs for quality projects. This representation comprises the identification and attribution of metadata that allow the organization and retrieval of items, according to their specificities.

This research had as the main objective to build a metadata standard for theses and dissertations having the Brazilian Standard of Metadata for Theses and Dissertations (MTD-BR) as the reference instrument. This system was created for the Digital Library of Theses and Dissertations (BDLTD), of the IBICT (Brazilian Institute for Scientific and Technological Information). This study, thus, aims at the descriptive representation and interchange of files related to theses and dissertations in digital libraries utilizing the Entity-Relationship Model (ERM), as observed in the final report of the project *Functional Requirement of Bibliographic Record* (*FRBR*), developed by the International Federation of Library Associations (IFLA).

2.0 The object of analysis: the Brazilian Metadata Standard of Thesis and Dissertations (MTD-BR)

In the present research a metadata standard was proposed which was not destined for documents in general, but rather, for a specific segment: theses and dissertations. Theses or dissertations can be defined as manuscript documents presented as a partial requirement for receiving an academic degree (Joint Steering Committee for Revision of *AACR* 2004, 1-42). Hence, they are primarily manuscript documents which will have their bibliographic description adapted to description guidelines for electronic resources when inserted into a digital library.

As reference standard for the proposed work, the Brazilian Metadata Standard of Theses and Dissertations (MTD-BR) was chosen and will be described below. In Brazil several initiatives exist in universities in an attempt to structure and manage digital libraries' collections of theses and dissertations. Among them, the Digital Library of Thesis and Dissertation, Brazilian (BDTD), of the IBICT is noteworthy and it has as the main objective (Southwick 2004, 1) the:

> Integration of the Brazilian initiatives of electronic publications and bibliographic record of theses and dissertations, providing the final users with an integrated view of these initiatives through information services and products with aggregate value. The main service ... is that of search and retrieval of theses and dissertations documents produced in Brazil or produced by Brazilians outside Brazil ... from a single gateway, to perform searches in the several initiatives without the need to visit each of them individually and perform searches in local repositories of those data provider institutions.

The success of this integration project between digital libraries of theses and dissertations of different universities depends, specially, on a unique standard of document description that encompasses documents both in printed and electronic format.

The MTD-BR standard was developed as part of the BDTD project to generate information products and services that could identify and localize electronic

theses and dissertations. (IBICT 2005). Currently there are 32 institutions, among research institutes and universities that act as data or service providers, and are integrated with BDTD. They are listed in Appendix 2. It is not necessary that all participating institutions utilize the same software for integration of collections to the IBICT BDTD; they only need to use a system compatible with the MTD-BR standard. However, the IBICT, which is linked to the Brazilian Ministry of Science and Technology, makes available the TEDE (Theses and Dissertations Electronic Publications System), at BDLTD. This is a system developed to provide the implementation of digital libraries in Higher Education Institutions (IES) and its integration with the IBICT BDLTD and with the *Networked digital Library of Theses and Dissertations*, NDLTD, of Virginia Tech University (IBICT, 2005).

The whole MTD-BR standard document can be found in Appendix 1. It is the official document of the standard, edited in 2005 by the IBICT, and is also available at the IBICT site. It has three parts:

– Part A: List of elements of the schema;
– Part B: Standards adopted for filling out specific elements such as canonical tables, proposed and used by national or international organizations (e.g., ISO guidelines, tables with IBGE Units of Federation, abbreviations of depository libraries used at the Brazilian bibliographic commutation system, etc.);
– Part C: Standards adopted for attributes for identification of some contents of the elements.

It should be noted the difference in the meaning of "attribute" in the MTD-BR and in the entity- relationship model. In the MTD-BR an "element" is simple or composed, while in the model it corresponds to "attribute" of the entities. Attributes in the MTD-BR are facultative elements adopted in the schema.

2.1 Analysis of the MTD-BR according to the ERM/FRBR

One of the problems found in this research results from the terminological inconsistency in the names of the elements from the MTD-BR reference standard, the cataloguing rules and the entity-relationship model. The discussion of these inconsistencies will be introduced later, when the specific elements are discussed and analyzed. However, for a deeper analysis of the MTD-BR standard through the FRBR model, it is necessary that:

– the entities of the *FRBR* model are defined in terms of the type of material described which in the specific case of this research are theses and dissertations;
– their relationships are presented, such that the relationship between the entities is better understood; and,
– the respective attributes of these entities are specified.

Based on these needs, the present analysis has identified the entities, relationships, attributes, sub-attributes and some qualifiers found in the reference standard. Not only the elements of the MTD-BR standard were taken into account but also other foreseen elements were considered in the cataloguing rules of manuscripts and electronic resources, in the attributes of the entities suggested by the *FRBR* study, in the NDTTD-ETD-MS standards, and the Dublin Core. Therefore the results discussed will are of two types:

– real metadata, i.e., the ones present in the MTD-BR model, identified with grey stripes; and,
– potential metadata, i.e., those that are not in the MTD-BR, identified with the word potential.

The applicability of both real and potential metadata will be discussed and analyzed based on the theoretical and technical study performed during the present research about the cataloguing rules related to treatment of theses and dissertations.

For the proposed analysis, the reference standard and classification of their elements were as predicted by the entities in the ERM. It should be noted that the elements of the reference standard correspond to the attributes of the ERM. Thus the difficulty for the terminological correspondence is those elements of the reference standard which correspond to attributes of the ERM. Attributes of the reference standard qualify the attributes of the ERM and will herein be called qualifiers.

The figures that follow contain, on the left side, the entities of the ERM inside rectangles; then, the attributes are shown inside ellipses; the elements corresponding to the reference standard are in grey rectangles. The elements represented in a rectangle with round borders, when present, correspond to attributes of the reference standard that were incorporated into the modeling in this research as qualifiers. These qualifiers are characterized by elements of the reference standard that cannot be considered attributes since they are repeated in several elements of the

MTD-BR. However, they represent qualifications of the data contained in the elements, which are essential for information retrieval in a digital library.

2.1.1 Entities for theses and dissertations

2.1.1.1 The entity WORK

This entity represents the abstract side of the idea that involves the academic work of a thesis or dissertation; its attributes will be represented by the following descriptive metadata, as can be observed in Figure 1. From the four attributes referring to the entity WORK for theses and dissertations only three are represented in the MTD-BR standard (real attributes). Each of these attributes has specific contents:

– DEGREE (Real) – Academic degree associated to the thesis or dissertation, according to the Table of Degrees that accompanies the MTD-BR standard document;
– COVERAGE (Real) – Spatial or temporal scope of thesis or dissertation, as well as the area of applicability;
– DEGREE TITLE (Real) – Name of the academic degree associated with the thesis or dissertation;
– FORM (Potential) – Specific form of the academic work; for example, whether it is a thesis, dissertation, graduation work, monograph (the last two in professional master studies).

For the potential attribute FORM, the following Table of Form is proposed: thesis, dissertation, project, graduation work, report and monograph.

2.1.1.2 The entity EXPRESSION

This entity represents the accomplishment of the academic work as for the specific form of its content which in the case of theses and dissertations will be "textual." So, its attributes will be represented by the following descriptive and administrative metadata - real and potential (see Figure 2).

According to the reference standard and the *FRBR* model, there are five attributes referring to the entity EXPRESSION for theses and dissertations and similar works. Each of these attributes has specific contents:

– UPDATED ON (Real) – Date that informs when the thesis or dissertation was inserted and/or modified in the database; or the date in which it was registered in the library, in case it exists only in printed form;
– LANGUAGE (Real) – Language in which the thesis or dissertation was written;

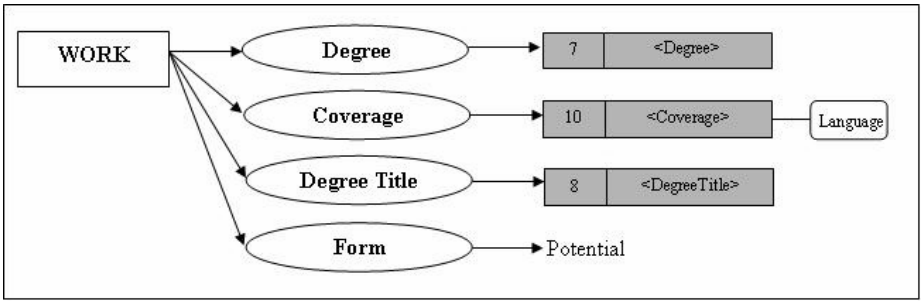

Figure 1. Entity WORK and its attributes

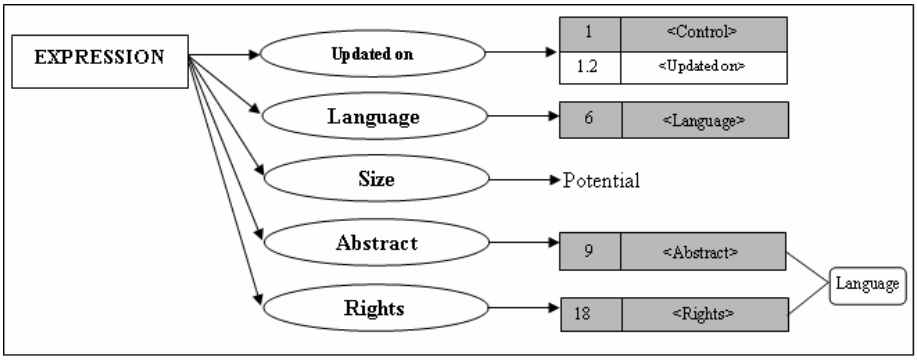

Figure 2. Entity EXPRESSION and its attributes

– SIZE (Potential) – Size of thesis or dissertation, that can represent both the number of pages of the manuscript, and the number of megabytes in the electronic resource;
– ABSTRACT (Real) – Abstract of the thesis or dissertation, both in the original and alternative language;
– COPYRIGHT (Real) – Informs about the conditions for distribution, reproduction and utilization of thesis or dissertation.

The potential attribute SIZE refers either to number of pages of a printed thesis or dissertation or to file size of a digital thesis or dissertation. This datum is relevant since with it the researcher can previously decide if the academic work is too long to require a copy or if the file can be downloaded during the research. In addition, this attribute is a metadatum that, although not contemplated in the reference standard, is present in other standards of digital libraries such as, for example, the Sistema Saber of USP. The field SIZE which is seen in the MTD-BR standard is more related to an Area of Notes or Observations; here it is more adequate as attribute of the entity MANIFESTATION as it can be verified in field 19 of Figure 3.

2.1.1.3 The entity MANIFESTATION

This entity is the physical incorporation of the academic work. This way, it will represent the major areas of description of the *AACR2*. According to the MTD-BR standard and the *FRBR* model, there are six attributes referring to the entity MANIFESTATION and all of them are represented the MTD-BR standard. Each of these attributes has specific contents:

– TITLE (Real) – Title of thesis or dissertation or similar works ;
  – TYPE OF MATERIAL (Potential) – Name of the type of physical support in which the thesis or dissertation was made on; according to the predicted standard it can be manuscript, electronic resource or both;
  – DATE OF DEFENSE (Real) – Date in which the thesis or dissertation was defended;
  – PLACE OF DEFENSE (Real) – Place where thesis or dissertation was defended;
  – NOTES (Real) – Reserved for the creation of sub-items (subfields) for specific use;
  – LEVEL OF ACCESS (Real) – Access restriction to the file.

19	<Size>	Namespace	Reserved for the creation of sub-items (subfields) for specific use	NR	F	

Figure 3. Element <Size> of the MTD-BR standard

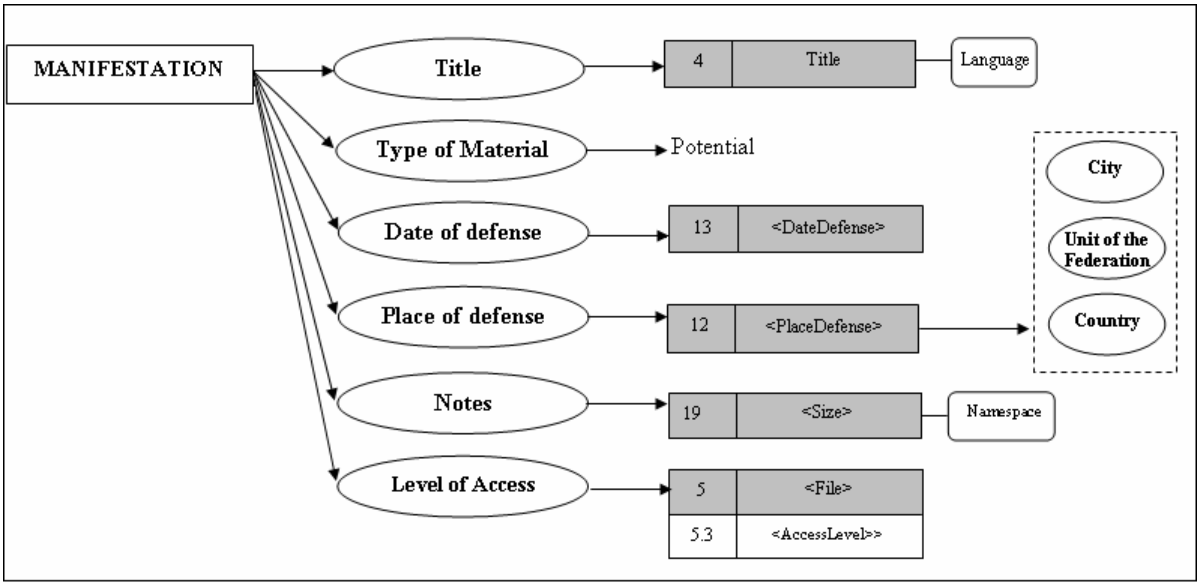

Figure 4. Entity MANIFESTATION and its attributes

In relation to the potential attribute TYPE OF MATERIAL; the lack of this attribute does not allow the user to know whether the thesis or dissertation exists as electronic resource or as manuscript or as both. It is known that the availability of the academic work in the electronic format is only possible with the author's authorization. Because of this, many theses or dissertations in a digital library have only their descriptive data recorded and online content is not accessible. At the BDTD for example, the only way to deduce that an academic work is not available in electronic format is when the URL of the file is accessed and the user is sent to the COMUT site for a copy request. However, the searches of many researchers could be faster if they could request the system to retrieve only the academic works that are either available in the electronic or printed format. Another suggestion is that in the proposed standard the name of the element <Size> is altered to NOTES, since the term "size" in the present proposal will be used for identifying the size of the of thesis or dissertation in pages or bytes, in the entity EXPRESSION.

2.1.1.4 The entity ITEM

This entity represents the concrete side of a work, a single copy of the thesis or dissertation. This way, it will specially represent the data of the record of the theses or dissertations in the digital or traditional library. Thus, its attributes will be represented by the following administrative metadata of the MTD-BR reference standard, such as seen in Figure 5.

According to the reference standard and the *FRBR* model, there are three attributes referring to the entity ITEM. Each of these attributes has specific contents:

– CALL NUMBER (Real) – Code that identifies the thesis or dissertation in the depository library collection.
– DOCUMENT IDENTIFICATION (Real) – Code that identifies the thesis or dissertation in both the digital library or depository library base.
– FILE URL (Real) – Electronic address of the thesis or dissertation file and/or of the origin digital library.

In relation to the attribute DOCUMENT IDENTIFICATION it is important to note that the URL of the file of a thesis or dissertation can represent this identification in a digital library. The record number is the most utilized in the case of online and traditional catalogues. This way, a greater reflection about the exact content of this attribute is necessary. It is important to define the exact character of this attribute so that this apparent ambiguity is better improved in the digital library. With this reasoning we get into three different types of document identification, administration number, record number in the database and URL of the file.

Another important point to be considered is that in the depository library it is more important that the call number is the registration number, which represents the single identification of the document in that library. It is believed that it is essential to have a clearer definition of what is the identification of a document in a digital library for the development and improvement of metadata standards for digital libraries of theses and dissertations and similar works.

2.1.1.5 The entity PERSON

This entity will specially represent the data referring to the indication of responsibility of physical persons of the theses or dissertations and similar works. So, its attributes will be represented by the following metadata structures of the MTD-BR standard, such as shown in Figure 6.

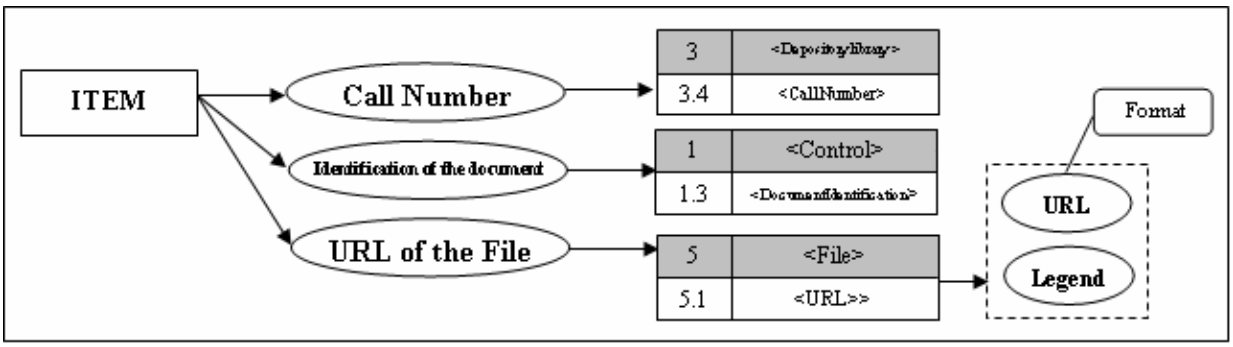

Figure 5. Entity ITEM and its attributes

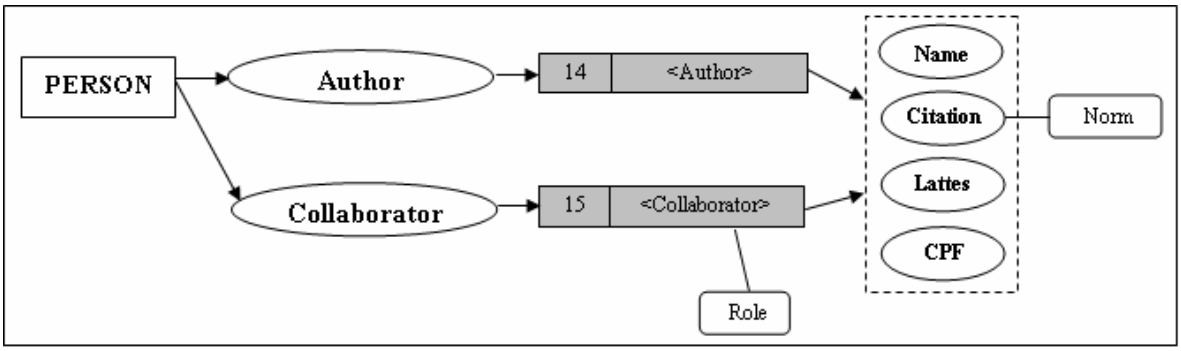

Figure 6. Entity PERSON and its attributes

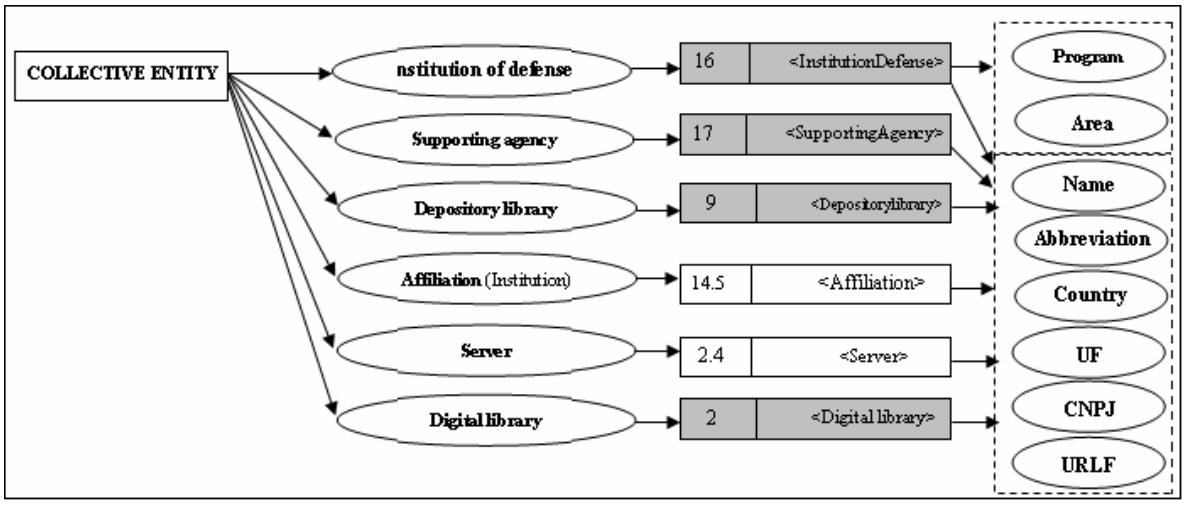

Figure 7. Entity COLLECTIVE ENTITY and its attributes

According to the mentioned standard and the *FRBR* model there are two attributes referring to the entity PERSON for theses or dissertations. Each of these attributes has specific contents:

– AUTHOR (Real) – Name of the author of thesis or dissertation;
– COLLABORATOR (Real) – Name of other persons that contributed in the thesis or dissertation and form of collaboration (supervisor, co-supervisor, committee member, etc).

In this entity, its two attributes will have specific sub-attributes with individual information about each person involved in the thesis or dissertation. In addition to these two levels of attributes, this entity also predict two qualifiers: the rule that creates a standard of citation input of the names of the people involved in a specific thesis or dissertation; and the role of each collaborator in a specific thesis or dissertation, which can be: supervisor, co-supervisor and committee member.

2.1.1.6 The entity COLLECTIVE ENTITY

This entity will specially represent the data referring to the indication of responsibility of collective entities of the theses or dissertations and similar works. So, its attributes will be represented by the following metadata in the MTD-BR standard, through which a thesis or dissertation will be related to the institutions involved in its production and dissemination, as seen in Figure 8. The importance of grouping these attributes, present also in other entities, has the same purpose indicated in the entity PERSON. Through the entity COLLECTIVE ENTITY it will be possible when retrieving information to group the theses or dissertations by the institutions where defense occurred, by the supporting agencies, by the libraries and so on.

There are six attributes referring to the entity COLLECTIVE ENTITY. Each of these attributes has specific contents:

– INSTITUTION OF DEFENSE (Real) – Name of institution where thesis was defended;

– SUPPORTING AGENCY (Real) – Name of the agency that financially supported the author;
– DEPOSITORY LIBRARY (Real) – Name of library where the thesis or dissertation is physically filed;
– AFFILIATION (Real) – Name of institution to which the person belongs;
– SERVER (Real) – Name of institution that houses the digital library;
– DIGITAL LIBRARY (Real) – Name of digital library responsible for the digital publication of thesis or dissertation.

This entity also contains three levels of attributes. All of the entity attributes COLLECTIVE ENTITY will have a basic core of six sub-attributes: Name, Abbreviation, Country, Unit of Federation, CNPJ, URL. However, two specific attributes will contain a specific core of sub-attributes: the attribute INSTITUTION OF DEFENSE will have, in addition to the mentioned sub-attributes, another two attributes: Program and Area.

2.1.1.7 The entity SUBJECT

This entity will represent specially the data referring to keywords or descriptors that will identify a thesis or dissertation and will relate it with other similar academic works. So, its attributes will be represented by the following metadata structures of the MTD-BR standard, such as shown in Figure 8.

According to the MTD-BR standard and the *FRBR* model, there are two attributes referring to the entity SUBJECT in a digital library of theses or dissertations. Each of these attributes has specific contents:

– DESCRIPTOR (Real) – Word that names the subject of thesis or dissertation.
– STANDARDIZED LIST (Real) – Source of the SUBJECT (thesaurus).

The reference standard MTD-BR identifies the entity SUBJECT through a descriptor accompanied of two attributes or qualifiers: the Language of the descriptor and the scheme from which the descriptor was extracted from (thesaurus). The qualifier "Scheme" will be considered in the modeling as an attribute of the entity SUBJECT: "Standardized List"

3.0 The entities and relationships in the proposed standard

With the identification and definition of the attributes of the reference standard, the next step is to identify, establish and define each relationship between these entities. Figure 9 should be consulted for the first instance of verification of the relationships of the proposed standard. The description of the relationships that includes only the basic entities of the proposed model appears below.

Thus, having in mind that the relationships are represented by expressions inserted between the entities, that the simple arrows represent the cardinality "1" and the double arrows represent the cardinality "plus 1," the proposed diagram is discussed. In the specific case of description of theses and dissertations, the entity WORK can be linked through an EXPRESSION, i.e., it is a textual production. When the research work is concluded each entity EXPRESSION is presented in the form of one or more MANIFESTATIONS, i.e., an EXPRESSION can be presented under more than one type of MANIFESTATION. The descriptive treatment given to this manifestation will be intimately related to the physical form of a WORK: manuscript or electronic resource.

If the WORK is done in the electronic format it will receive a descriptive treatment for electronic resources; but if it is done only as a manuscript, its descriptive treatment will follow the guidelines for manuscripts. However, if it exists in both formats the descriptive treatment should contain the information that represents both formats. Thus the need to specify in a metadata standard the information about physical support of the material, i.e., the type of the material.

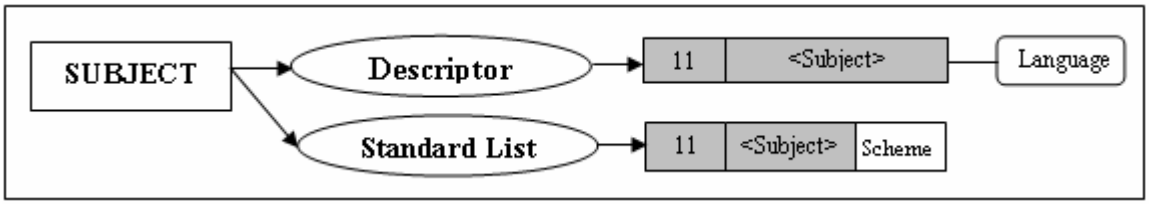

Figure 8. Entity SUBJECT and its attributes

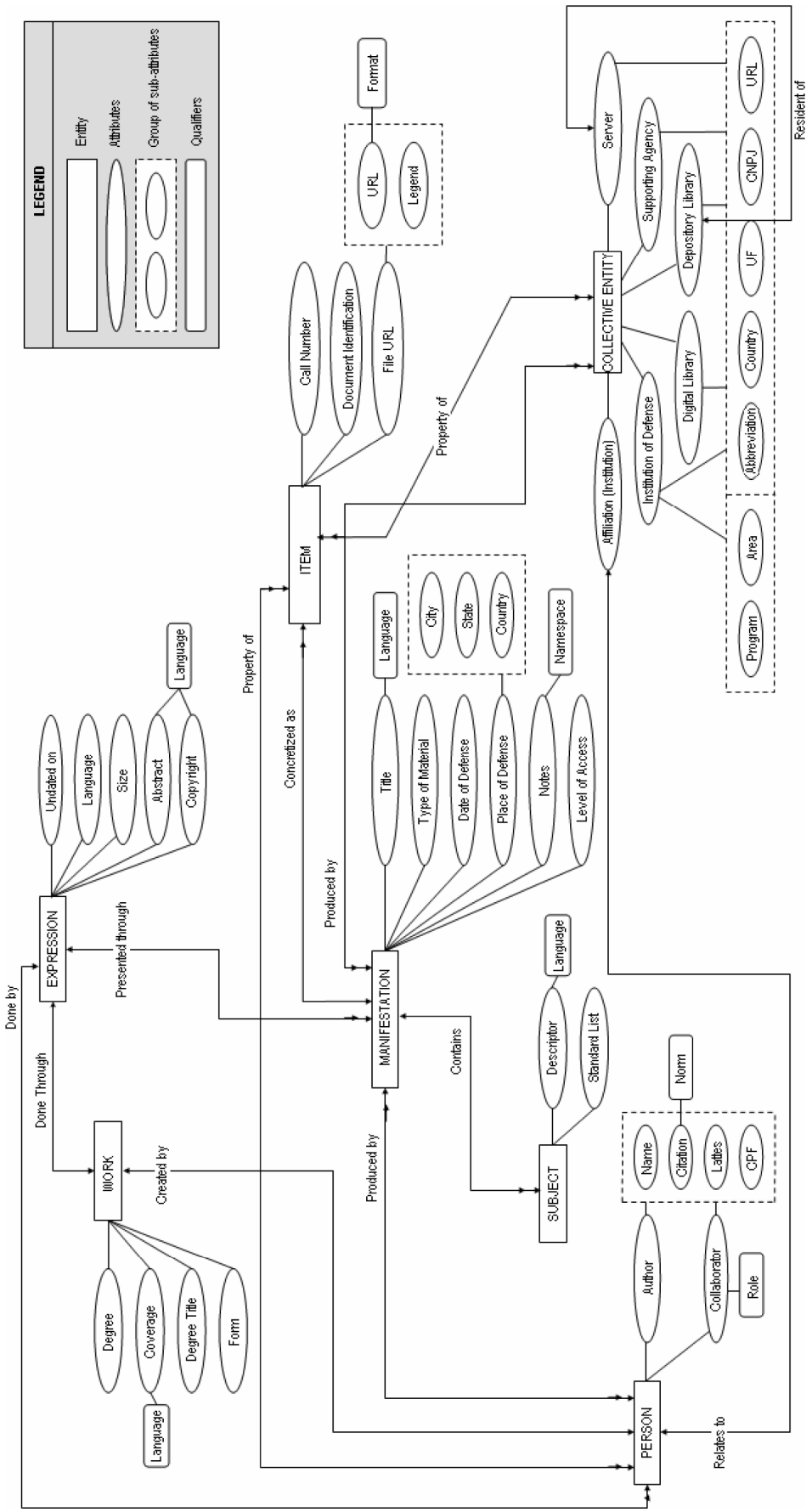

Figure 9. FRBR model Entities of the proposed standard

When a WORK becomes a MANIFESTATION, it is concretized as the entity ITEM which represents the original copy of this work; it may be the property of a library, or of a person. This way, a MANIFESTATION is concretized or exemplified through one or more ITEMS since it is possible to have more than one copy of each MANIFESTATION. In this sense, the entity ITEM will contain information about the records of control of this copy and its localization in the library.

Since theses and dissertations do not have standard numbers, such as ISBN, their unique identifiers will be their records in different libraries. In the specific case of a digital library this document identification may be represented also by the URL of the file if each thesis or dissertation has its own specific URL.

One or more ITEMS, in turn, are a property of one or more COLLECTIVE ENTITIES, i.e., both the digital and traditional libraries may have their copies of a thesis or dissertation, as well as the institutions where defense occurred, the supporting agencies, and the institutions to which the authors and collaborators belong.

As for the WORK as an abstract idea created by a PERSON, i.e., created by its author, since in its essence a WORK is initially idealized only by its author—when this work becomes an EXPRESSION, i.e., a text document, it is created by more than a PERSON, i.e., the document starts to be created both by its author and some collaborators such as the supervisor and co-supervisor.

When this WORK is finished it becomes a MANIFESTATION that will be produced by more than one PERSON and also by more than one COLLECTIVE ENTITY, i.e., it will be a production of the author and of several collaborators: supervisors, co-supervisors and committee members; it will also be a production of the institutions where defense occurred and of the supporting agencies.

When the MANIFESTATION of a WORK is ready it becomes a copy that can be the property of both the persons involved in its creation and of the institutions.

This way, a MANIFESTATION will be concretized and exemplified by an ITEM and may be the property of the author, collaborators, institutions where defense occurred, supporting agencies and digital and traditional libraries. So, a MANIFESTATION may be concretized in more than one ITEM; and one or more items may be property of more than one PERSON or COLLECTIVE ENTITY.

Finally, all basic entities, WORK, EXPRESSION, MANIFESTATION and ITEM, contain the entity SUBJECT since the group of these four basic entities represents the thesis or dissertation as a whole. This relationship is directly linked to the entity MANIFESTATION since this will represent the thesis or dissertation indexed in the libraries.

As observed in Figure 9, the entities PERSON, COLLECTIVE ENTITY and SUBJECT will contain elements that will link the thesis or dissertation to other scientific and academic works of their authors, collaborators or institutions where defense took place and the supporting agency; in addition, they will relate these theses and dissertations between themselves and with other documents external to the digital library that contain the same or similar subjects and vice-versa.

The entity PERSON will also be related with the entity COLLECTIVE ENTITY since authors and collaborators of a thesis or dissertation have an affiliation, i.e., they are individually linked to an Institution which is also a COLLECTIVE ENTITY; somehow the institution indirectly contributes through these persons in the production and dissemination of a thesis or dissertation.

The entity SUBJECT is an element that, in the areas of bibliographic description, represents one of the points of access of a thesis or dissertation, i.e., one of the structural metadata of a metadata standard was included in this modeling only with its main attributes in order to have a good descriptive representation.

4.0 Proposed metadata standard for theses and dissertations: Schema, standards for elements and standards for attributes

The proposal of a metadata standard will be presented in three parts: Part 1 – List of Elements of the Standard – Version 1; Part 2 – Standards for attributes; Part 3 – Standards for Qualifiers.

List of Elements of the Standard – Part 1 - Comprises the group of eight fields: Sequential number; Name of the attribute; Qualifier; Classification of the metadata; Content; Option (Repetitive and Non-repetitive); Option (Mandatory and Facultative); standard.

Standards for Attributes – Part 2 – Comprises the relationship of thirteen standards with their respective original sources or tables for application when using the proposed standard.

Standards for Qualifiers– Part 3 – Comprised of three standards with respective original sources and tables for application when using the proposed standard.

N.	Entity	Name of the Attribute	Qualific.	Classification of Metadata	Content	R / NR	M / F	Standard
1	Work	<Degree>		D / S	Academic degree associated to thesis or dissertation	NR	M	Table of Degree
2		<Coverage>	Language	D	Spatial or temporal scope of thesis or dissertation (Areas of the knowledge)	R	M	CNPq Table
3		<DegreeTitle>		D / S	Name of the academic degree associated to thesis or dissertation	NR	M	-
4		<Form>		D / S	Specific form of the academic work.	NR	M	Table of Form
5	Expression	<Datarecord>		A	Date of record of thesis or dissertation	NR	M	OAI Protocol
6		<Language>		D / S	Language of thesis or dissertation	NR	M	ISO 639
7		<Size>		D	Size of thesis or dissertation, that can represent both the number of pages of the manuscript and the number of megabytes in the electronic resource;	NR	F	-
8		<Abstract>	Language	D	Abstract of thesis or dissertation	R	M	-
9		<Copyrights>	Language	A	Informs about the conditions of distribution, reproduction and utilization of thesis or dissertation	R	M	-
10	Manifestation	<Title>	Language	D / S	Title of thesis or dissertation	R	M	-
11		<MaterialType>		A / S	Type of document of thesis or dissertation	R	M	Table of Type
12		<DateDefense>		D	Date in which thesis or dissertation was defended	NR	M	ISO 8601
13		<PlaceDefense>		D	Place of defense of thesis or dissertation	NR	M	-
		<UF>			Unit of Federation of city where thesis or dissertation was defended	NR	F	UF
		<Country>			Country where thesis or dissertation was defended	NR	M	ISO 3166
14		<Notes>	Name-space	D	Reserved for the creation of sub-items (subfields) for specific use	R	F	-
15		<AccessLevel>>		A	Restriction of access to File	NR	M	Table of Access
16	Item	<CallNumber>			Code that identifies the thesis or dissertation in the collection of depository library	NR	F	-
17		<DocumentIdentification>		A	Code that identifies the thesis or dissertation in the database of depository and/or digital library	NR	M	-
18		<URLFile>		A / S	Electronic address of the file of thesis or dissertation and/or presentation page of the metadata in the original digital library	NR	M	-
		<URL>	Format		Format of the File URL			URI Standard
		<Legend>	Language	A	Legend for File, description of its content	R	M	-
19	Person	<Author>		D / S	Name of the author of thesis or dissertation	R	M	-
		<Citation>	Norm		Form by which the person will be cited in some norm	R	M	-
		<Lattes>			Electronic access for CV of the Person at the Lattes Platform	NR	M	URI Standard
		<CPF>			CPF of Person	NR	F	Standard MF - 11 digits
		<Affiliation>			Name of institution to which the Person is affiliated to	R	F	-
		<Abbreviation>			Abbreviation of the Institution	NR	F	-
		<Country>			Country of the Institution	NR	F	ISO 3166
		<UF>			UF of Institution	NR	F	UF
		<CNPJ>			CNPJ of Institution	NR	F	MF Standard - 14 digits
		<URL>			Electronic address of the *site* of Institution	NR	F	URI Standard

Table 1. Proposed Standard – Part 1/3, List of standard elements – version 1

(to be continued on next page)

N.	Entity	Name of the Attribute	Qualific.	Classification of Metadata	Content	R / NR	M / F	Standard
20		<Collaborator>	Role	D / S	Name of collaborator of thesis or dissertation and form of participation.	R	M	-
		<Citation>			Form by which the Person wishes to be cited	NR	M	-
		<Lattes>			Electronic access for CV of the Person at the Lattes Platform	NR	M	URI Standard
		<CPF>			CPF of the Person	NR	F	MF Standard - 11 digits
	Person	<Affiliation>			Name of institution to which the Person is affiliated to	R	F	-
		<Abbreviation>			Abbreviation of the Institution	NR	F	-
		<Country>			Country of the Institution	NR	F	ISO-3166
		<UF>			Unit of Federation of Institution	NR	F	Unit of Federation
		<CNPJ>			CNPJ of Institution	NR	F	MF Standard - 14 digits
		<URL>			Electronic address of the *site* of Institution	NR	F	URI Standard
21		<InstitutionDefense>		D / S	Name of institution where thesis was defended	NR	M	-
		<Abbreviation>			Abbreviation of the Institution	NR	M	-
		<Country>			Country of the Institution	NR	F	ISO 3166
		<UF>			Unit of Federation of Institution	NR	F	Unit of Federation
		<CNPJ>			CNPJ of Institution	NR	F	MF Standard - 14 digits
		<URL>			Electronic address of the *site* of Institution	NR	F	URI Standard
		<Program>			Graduate Studies Program through which thesis or dissertation was defended	R	M	-
		<Name>			Name of the Graduate Studies Program	NR	M	-
		<Area>			Area of concentration of the Graduate Studies Program	NR	M	-
22		<SupportingAgency>		D / S	Name of agency that financially supported the author	R	F	-
		<Abbreviation>			Abbreviation of the Institution	NR	F	-
		<Country>			Country of the Institution	NR	F	ISO 3166
		<UF>			UF of Institution	NR	F	Unit of Federatio
		<CNPJ>			CNPJ of Institution	NR	F	MF Standard - 14
		<URL>			Electronic address of the *site* of Institutio	NR	F	URI Standard
23	Collective Entity	<Depositorylibrary>		A / S	Name of library where the physical copy is filed	NR	M	-
		<Abbreviation>			Abbreviation of depository library	NR	M	Registration CC
		<URL>			Electronic address of the *site* of depositor	NR	F	URI Standard
24		<Digital library>		A / S	Name of Institution responsible for digit dissertation	NR	M	-
		<Abbreviation>			Abbreviation of digital library	NR	M	--
		<URL>			Electronic address of the *site* of digital lib	NR	M	URI Standard
		<Server>			Name of institution that houses the digit stitution)	NR	M	-
		<Abbreviation>			Abbreviation of Covenant Institution	NR	M	-
		<Country>			Country of the Institution	NR	F	ISO 3166
	Subject	<UF>			UF of Institution	NR	F	Unit of Federatio
		<CNPJ>			CNPJ of Institution	NR	F	MF Standard - 14
		<URL>			Electronic address of the *site* of Institutio	NR	F	URI Standard
25		<Descriptor>	Language	D / S	Topics approached in the thesis or dissert which they were extracted from, when ap	R	M	-
26		<Standard List>		A	Thesaurus or list of headings of subject, s			-

Table 1. Proposed Standard – Part 1/3, List of standard elements – version 1

Legend: R= Repetitive NR= Non-repetitive M= Mandatory F= Facultative Metadata: A=Administrative D=Descriptive S=Structura

Some observations about the list of elements follows: the secondary and tertiary elements or sub-attributes were not numbered only because one could more clearly verify the final total of elements that the metadata standard would have in this proposal.

– Proposed standard - Part 1/3: List of elements of the standard – Version 1 (Table 1). For filling out the fields some standards of the reference standard were adopted and others were suggested. Thus, for a better understanding these standards will be exemplified below.
– Proposed standard - Part 2/3: standards for attributes

Registration CCN/COMUT	→	Table of Abbreviations of Institutions Source: < http://www.ibict.br >
ISO 3166	→	International Table of Countries Source: < http://www.iso.ch/iso/en/Standards >
ISO 639	→	Language: use two or three initial characters of the language. Ex.: POR, FRA, ALE, ING.????GER, ENG?? Source: < http://www.iso.ch/iso/en/Standards >
ISO 8601	→	YYYY-MM-DD or YYYY-MM or YYYY Source: < http://www.iso.ch/iso/en/Standards >
MF Standard - 11 digits	→	11 digits without hyphen and points. Ex: 99999999999
MF Standard - 14 digits	→	14 digits without hyphen and points. Ex: 99999999999999
URI Standard	→	URL Standard Source:< http://rfc.sun*site*.dk/rfc/rfc2396.html >
OAI Protocol	→	AAA-MM-DDThh:mm:ss **YYYY-MM-DD**[specification of hour, minute and second are facultative]. Source: <http://www.w3.org/TR/xmlschema-2/#dateTime >
Table of Access	→	Public Restricted
Table of Form	→	Thesis Dissertation Project TCC Monograph Report
Table of Degree	→	Doctor Master Professional Master
Table of Type	→	Manuscript Electronic Resource
CNPq Table	→	Table of Areas of the knowledge Source: < http://www.cnpq.br/areas/tabknowledge/index.htm >
Unit of Federation	→	Table of Units of Federation of IBGE

– Proposed standard - Part 3/3: Standards for Qualifiers:

ISO 639	→	Language: use two or three initial characters of the language. Ex.: POR, FRA, ALE, GER, ENG Source: < http://www.iso.ch/iso/en/Standards >
Table of Format	→	Table of Format of electronic file Source: < ftp://ftp.isi.edu/in-notes/iana/assignments/media-types/media-types >
Table of Role	→	Supervisor Co-supervisor Committee member

5.0 Final remarks

With the automation of information systems and the advent of digital libraries guidelines, standards and techniques of library studies have been widely discussed, analyzed, reevaluated and re-structured. In the last decades Brazilian library studies have attempted to be updated in relation to new methodologies on information treatment. So, it tries to apply tools that allow optimization of the use of the computer and Internet in relation to more precise information retrieval for the users both in projects of automation of information units and in projects of implementation of digital libraries. In this context, the guidelines, standards and rules utilized for information treatment concerning descriptive representation and thematic representation have been the focus of most researches on library studies in the world.

As for the study of guidelines of descriptive representation, the final report of the *FRBR* review group of IFLA is effectively a hallmark for a new era of research in the area of information treatment, offering librarians new tools of evaluation of standards of document description. Metadata has also been the focus of several studies in the area, such as in the research presented here, since they constitute effective resources for structuring standards of description of digital objects and for the development of architectures that allow a greater interoperability between digital libraries.

However, information treatment in a digital library environment goes beyond the simple descriptive rep-

resentation. And that is where the final report of the *FRBR* review group, despite its general character, is a great contribution to the theory of bibliographic description and to the way this description is applied to the "Internet Era" (AYRES et al., 2002), since it also emphasizes attention for the needs of the users. With the analysis done here it could be noted that the entity-relationship model proposed by the *FRBR* offers a vast field of study in the area of information representation—either descriptive or thematic—where Brazilian information professionals will have material for perfecting and refining their projects of library automation and development of digital libraries. As Delsey wrote (1997, 2-3):

> The principal value to be gained from modeling the logical structure of *AACR* is that it would assist us in shifting our focus from the process of cataloguing to the entities or objects that we are endeavouring to represent in our catalogues, from the specifics of individual rules to the operative assumptions and principles that inform the rules, and from the formal structure of the catalogue record to the logical structure underlying the data in the record. The discipline of the modeling exercise itself would serve to highlight anomalies within the rules and inconsistencies in the application of basic principles. It would also oblige us to clarify our thinking with regard to the concepts that are integral to the logical design of the code. Perhaps most important of all, the development of a model would provide us with a clear framework to be used in determining how to develop and extend the code to reflect newly emerging phenomena in the universe of information objects.

Thus, it is extremely important that these tools of data modeling are the focus of library study so that guidelines and standards are better adjusted to the current needs of the *on-line* and virtual catalogues.

Utilization of data models, such as the entity-relationship model-ERM, was shown to be very useful for the detection of possible application inconsistencies of the basic principles of cataloguing in descriptive metadata standards utilized for the information treatment in digital libraries. Another factor to be noticed is that this modeling allowed that a schema of digital information treatment was clearly understood and defined, contributing to the elaboration of proposals of metadata standards for digital libraries of theses, dissertations and similar works.

The present research allowed a view of how the data modeling tools can help in the improvement and adequacy of guidelines and standards of descriptive representation, both in traditional and digital libraries. With these tools the information professional can have a more administrative view in the analysis of metadata standards, and can better work the retrieval of informational resources in digital libraries, focusing specially in attending the needs of the users.

In this new era, where digital libraries are a concrete reality in world society, the information professionals increasingly need more tools that allow a better administration of these collections and a greater attention to the users. In traditional libraries this professional was present, which allowed the prompt attention to the doubts and difficulties of the users. Now, with the digital libraries this presence is not a reality anymore. So, the information professionals should foresee the doubts and difficulties as much as possible to facilitate an easier navigation in the digital libraries by the user. In this way these data modeling tools utilized by the system analysis may be of great aid for the information professional for overcoming this challenge. They would allow a deeper look into a specific reality, by collecting data and the information necessary for the development of systems that would better answer the demands of this reality.

References

Alvarenga, Lídia. 2001. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. *DataGrama Zero - Revista de Ciência da Informação* 2(6). Retrieved December 29, 2001, from http://www.dgz.org.br/Atual/Art_05.htm.

Alvarenga, Lídia. 2003. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. Encontros Bibli: *Revista Eletrônica de Ciência da Informação* 15. Retrieved October 01, 2003, from http://www.encontros-bibli.ufsc.br/Edicao_15/alvarenga_representacao.pdf.

Ayres, Marie-Louise et al. 1990. Informe sobre el éxito del projecto AustLit: Portal de bibliografía Australiana para la puesta em marcha de los modelos de eventos de FRBR e INDECS, así como otras implicaciones del FRBR. In *Proceedings of the 68th IFLA General Conference and Council. Glasgow, Scotland*. Retrieved March 01, 2004, from http://www.ifla.org/IV/ifla68/papers/054-133s.pdf.

Delsey, Tom. 1997. Modeling the logic of *AACR*. In *Proceedings of the International Conference on the*

Principles and Future Development of AACR. Toronto: American Library Association: Library Association, pp. 1-16.

Delsey, Tom. 2002. Functional analysis of the MARC 21 bibliographic and holdings formats. Washington: Library of Congress. Retrieved November 27, 2003, from http://www.loc.gov/marc/marc-functional-analysis/home.html.

Delsey, Tom. 1987. Developments and progress in cataloguing. In *Proceedings of the 58th IFLA General Conference and Council. Brighton.*

Delsey, Tom. 1998. The logical structure of the *Anglo-American Cataloguing Rules* - part I. Ottawa: National Library of Canada. Retrieved November 27, 2003, from http://www.nlc-bnc.ca/jsc/aacrint.pdf.

Delsey, Tom. 1999. The logical structure of the *Anglo-American Cataloguing Rules* - part II. Ottawa: National Library of Canada. Retrieved November 27, 2003, from http://www.nlc-bnc.ca/jsc/ aacrint2.pdf.

Delsey, Tom. 2000. The library catalogue in a networked environment. In *Proceedings of the Biocentennial Conference on Bibliographic Control in the New Millennium*. Washington: Library of Congress. Retrieved May 25, 2001, from http://lcweb.loc.gov/catdir/bibcontrol/delsey_paper.html.

Hickey, Thomas B., O'Neill, Edward T., Toves, Jenny. 2002. Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR). *D-Lib Magazine* 8n9: http://dlib.org/dlib/september02/hickey/09hickey.html.

International Federation of Library Association and Institutions; Cataloguing Section *FRBR* Review Group. 2007. Organization and reporting. http://www.ifla.org/VII/s13/wgfrbr/.

IBICT. 2005. BDTD. Retrieved February 15, 2005, from http://bdtd.ibict.br/bdtd.

IBICT. 2005. MTD-BR Padrão Brasileiro de Metadados de Teses e Dissertações – lista de elementos do schema (versão 2) Retrieved February 14, 2005, from http://www.ibict.br/schema.

IFLA Study Group on the Functional Requirements for Bibliographic Records. 1998. *Functional requirements for bibliographic records: final report*. Munich: K. G. Saur. Retrieved October 30, 2003, form http://www.ifla.org/VII/s13/frbr/frbr.pdf> ou <http://www.ifla.org/VII/s13/frbr/frbr.htm.

Joint Steering Committee for Revision of AACR. 2004. *Código de Catalogação Anglo-americano*. 2.ed.rev. São Paulo: FEBAB/IMESP.

Mey, Eliane Serrão Alves. 1987. *Catalogação e descrição bibliográfica: contribuições a uma teoria*. Brasília: Associação dos Bibliotecários do Distrito Federal.

Southwick, Silvia Barcellos et al. 2004. *Manual de instalação de ambiente computacional para implantação do sistema de publicação eletrônica de teses e dissertações (TEDE)*. Brasília: IBICT. Retrieved August 15, 2004, from http://tedesite.ibict.br/tde_manuais/arquivos/Bdtd%20Manual_instalacao_v2-1.doc.