Which Metadata for Ancient Arabic Manuscripts Cataloguing?

Mohammed Ourabah Soualah
Université Lumière Lyon 2
ELICO – ENSSIB
med_soualah@yahoo.fr

Mohamed Hassoun
ELICO – ENSSIB
mohamed.hassoun@enssib.fr

Abstract

Three million ancient Arabic manuscripts are jealously preserved in libraries, in conservation institutions and in private homes. A non-consulted document is like a dead document—its fate is related to its consultation by users. The access to this kind of document is very difficult because of two main reasons: first, the fragile state of the manuscripts that make them easily degradable, necessarily limiting handling. Second, the dispersion of the manuscripts in different locations in the world makes access difficult. Thus, the digitization and the online availability of the ancient Arabic manuscripts are sought as a solution. The question remains: how to access the digitized images of the manuscripts? Despite the difficulty of interpreting the images, building the cataloguing is an imperative. Our paper deals with the Arabic manuscript cataloguing problems. It reviews the various aspects of the Arabic manuscripts cataloguing metadata. We propose a new cataloguing model and a potential interoperability project between catalogs encoded with different formats.

Keywords: Ancient Arabic manuscripts; cataloguing; digitization; Dublin Core; metadata.

1. Introduction

Ancient manuscripts are living witnesses to human civilization. They represent a real knowledge medium of a specific era. The access to these works is a real problem because of their fragility which limits handling; and because distant countries conserving them require travel for access. So digitizing the Arabic manuscripts and spreading them through the web seems to be the safest solution.

However, finding search tools that allow fast access to the digitized resources remains a serious problem. Access to the content of the digitized manuscript image is a real challenge. The cataloguing solution to this problem seems indirect and a tiresome task but it is precise and realistic. Cataloguing digital manuscripts is different from cataloguing originals. The former includes specifics of the digital document (e.g. format, links, storage). Therefore, the goal is to find metadata that allows both to be described. Our work aims to highlight the difficulty of finding metadata for the Arabic manuscripts. Thus, after a short presentation of the various modes of manuscripts cataloguing, we describe the different types of metadata that are used. This allows us to present our Arabic manuscripts cataloguing model. Then, we focus on catalog interoperability and present a Dublin Core use in order to reach this goal.

2. The Ancient Arabic manuscript

The ancient Arabic manuscript is a handwritten document created before the emergence of printing in Arabic. The manuscript is a witness to an era, containing the knowledge of that time. The manuscript is made in an artisanal way and with a rare and expensive material (papyrus, animal skins, etc.). Thus, the manuscript is considered both as an archeological (IRHT, 2006) and a scientific object. We can come closer to it by describing its physical characteristics, contents and history.

2.1. Ancient Arabic Manuscripts Characteristics

The ancient Arabic manuscript structure shows no well-defined shape. They are usually of an unequal size and contain a various number of layers. The appearance of the book in Arabic script is historically confused with the beginnings of Islam (Humbert, 2002). Since that time, the Arabic language has undergone an extraordinary expansion and allows Arab graphemes to be used in about 130 languages (Humbert, 2002).. There are two categories of manuscripts: those that have been written in Arabic language, as well as the *a'jami* ones (*a'jami* is the Arabic term for foreigner) (Humbert, 2002). The *a'jami* manuscripts use Arabic graphemes, but their contents are written in other languages, such as Persian, Turkish, Berber, Urdu, etc.

Arabic manuscripts are highly sought after by scholars and historians. Their contents are only partially explored. Arabic manuscripts are considered as priceless treasures for codicologists, paleographers and historians, for whom every aspect of the manuscript contains considerable important information (Delamarre, 2004).

2.2. Access to Ancient Arabic Manuscripts

The consultation of Arabic manuscripts poses technical and geographical problems. Digitization and publication on the web provides some solutions—allowing many to consult these manuscripts and avoid the handling of the originals.

Unfortunately, digitization in image mode doesn't allow access to the full text of the manuscript contents, so alternatives must be considered. The first principle is to describe the manuscript to include the codicological aspect, the paleographical aspect and the manuscript history. In addition to the description the catalog description will contain the link to the digitized form (the URI of the manuscript image).

3. The Ancient Arabic Manuscript Digitization

The Manuscript digitization procedure aims to reproduce the analog form of the manuscript into its digital form. The digitization output generates a manuscript image, called the manuscript image mode. This image mode lacks support for Information retrieval. OCR (Optical Character Recognition) programs are used to provide a first pass at a text searchable version. The professionals proceed to check the transcription in order to make an improved textual version called the manuscript textual mode.

3.1. Digitization Objectives

The digitization of manuscripts is motivated by the objective of preservation, dissemination and re-use of the manuscripts. Nowadays, it is accepted that digitization is a means of preservation rather than conservation itself (Cédelle-Joubert & Buressi, 2002). Therefore, the digital form of the documents revives the forgotten ones and renews the possibility of reading and interpreting the document.

3.2. Digitization Solution Structure

Previously, we described the two modes of manuscript digitization—the text mode and the image mode. The former implementation takes too much time, while the latter presents difficulties for full-text research. An intelligent digitization solution is not to choose only one of the two modes, rather combining them to obtain best results. The transition from one to another becomes possible thanks to the linking technology.

The manuscript image obtained from the digitization is not structured; it can only be visualized. Therefore, access to the manuscript images requires an association between the images and structured or unstructured data in order to support the possibility of information retrieval.

So, the solution is to design a reliable cataloguing system that will finely describe the digitized manuscript. The catalog will be made up of a set of descriptive records, which will provide links to the corresponding manuscript image. The solution can be enriched by annotation and browsing—services in the image database that are not the object of this paper.

Access to the manuscript image requires a search in the catalog. Once the information is found, the system displays the corresponding records. Then, thanks to a link, the manuscript images will be displayed with a simple click. Figure 1 below shows both the solution for the Arabic manuscripts digitization and their online access.

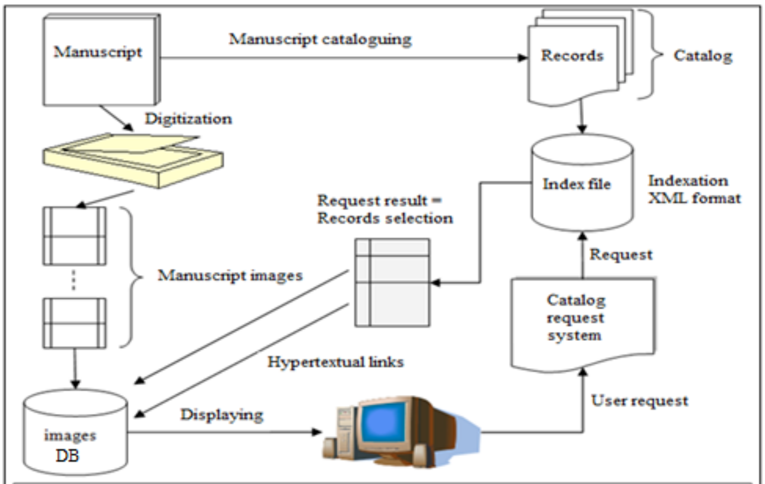

Fig.1: Arabic manuscripts digitization system and manuscripts images access

4. Ancient Arabic Manuscript Cataloguing

4.1. What's the Manuscript Cataloguing?

The description of the document identifies its availability for the reader (Ben Lagha, 2002). Two aspects of the manuscript description face each other:

Aspect 1: The manuscript is regarded as a single and isolated document.

Aspect 2: The manuscript is associated with the archive collection.

In the first case, information about the manuscript contents, its nature and its origin, is provided. After the cataloguing procedure, each record describes precisely its associated manuscript. In the second case, the manuscript is considered as a part of a collection. The manuscript is not described as a single piece, but its description is related to the manuscript collection description (Guesdon & Rodriguez, 2005).

The classification of a manuscript collection is a real problem because of its heterogeneity. Indeed, the manuscript collection is generally formed by a various kind of documents, which makes the catalog preparation difficult.

4.2. The Traditional Arabic Manuscript Cataloguing

Arabic manuscript cataloguing is a major effort because of the lack of a standard cataloguing protocol. Several attempts have been made without producing a unified model. Different catalogs have been built in the past, based on a formal description of the documents. The principle is to provide brief information about the manuscript such as the title, the subject, the author, a short summary, etc. However, the cataloguing was done differently from one cataloguer to another. This method results primarily in a list of records in a particular manuscript collection (Kaileh, 2004).

Arabic manuscript cataloguing is a product of two different schools: the first is the Arabic school and the second the Orientalist school, consisting primarily of western cataloguers who took the initiative to list Arabic manuscripts.

The two schools made a good effort. Despite some variations, they've used the same class of elements to describe the Arabic manuscripts: These consist of some 'standard' elements, which occur in most of catalogs; and some optional elements, which were used only by some cataloguers. Among the standard elements, we find:

> *The manuscript title*, *the manuscript author*, *the copyist*, *the manuscript completion date*, *the writing style*, *the physical description of the manuscript* and *the number of lines per folio*.

The optional elements most often used can be listed as below:

> *The place of origin of the manuscript*, *the incipit, the explicit*, *the ink used* and *the condition of the manuscript's base material*.

Considering these common practices, the objective to look for a representative set of descriptive metadata for ancient Arabic manuscripts.

5. Descriptive Metadata of Arabic Manuscripts

The objective of our work is to enable users to access the digitized Arabic manuscripts. At first glance, access to the images using their content seems more attractive. However, the Frank Lebourgeois team's work, at the LIRIS-RFV laboratory – INSA (Institut des Sciences Appliquées – Lyon) showed the difficulty of indexing the content ancient Arabic manuscripts, primarily because of the variability of use of Arabic characters, the low quality of the manuscript support and the low resolution of images (Kaileh, 2004). Therefore, access to the manuscript images using the catalog is an obvious solution. For relevance and representativeness, it is important to conduct a thorough cataloguing of the manuscript, providing researchers both keyword access and controlled access using standard entries.

The reader starts his research using a particular request. It can be the manuscript title, its author or any other manuscript feature described in the metadata. But, what is metadata in our case? Metadata is sometimes defined as data about data. It is a structured data set describing any resource (Peccate, 2003). Metadata allows the management and supports the accessibility of the described resource. Therefore, the reliability of the Arabic manuscript catalog is based on the relevance of the selected metadata.

5.1. Ancient Arabic Manuscripts Cataloguing Protocols

In our study, two detailed and comprehensive protocols retained our attention. The first one is defined by the IRHT (Institut de Recherche et d'Histoire du Texte – Paris), the second one is defined by the scientific committee for the Ancient Arabic manuscript record establishment. The committee members met in June 8^{th} and 9^{th} 1989 in the headquarters of the Prince Abdul-Aziz Al-Saûd foundation in Casablanca.

Both protocols view the manuscript in its various aspects: codicology, paleography and history. But, they propose two different cataloguing approaches. The IRHT proposes the specimen cataloguing mode while the El-Saud foundation proposes the volume cataloguing mode.

In the specimen manuscript cataloguing mode, the manuscript is regarded as an indivisible work although it may have a heterogeneous structure. Thus, the descriptive record is associated with the whole manuscript even when that whole is composed of several volumes. When the manuscript consists of several heterogeneous parts (also called volumes), the cataloguer, using the volume cataloguing mode, will make a descriptive record for each volume. Each volume is considered as a separate manuscript.

Despite the difference of the cataloguing modes, the descriptive metadata remains the same in both cataloguing modes.

5.2. The Ancient Arabic Manuscript Cataloguing Elements

Other issues at this level are often characterized as differences of opinion between cataloguing purists and Internet cataloguers. The first advocate for cataloguing based on the principles of cataloging regarding manuscripts, while the second are mainly interested in increasing the access points to the manuscript images and by the navigational capabilities of the image database.

a. Manuscript bibliographic cataloguing:

The catalog is composed of a set of files, sorted alphabetically, known as bibliographic records. Access to this catalog is done by searching the title, author, subject, and location, among others. Authority procedures define the catalog headings and are usually used as bibliographic record entries.

In this traditional cataloguing method, the cataloguer is mainly interested in the manuscript description, according to the cataloguing protocol that has been defined. In this case the access to the manuscript is via the authority list which the cataloguer implements.

We can summarize the ancient Arabic manuscript cataloguing metadata following the two schools' protocols as follows:

Manuscript identification:

- *Identification number, volume number* (when the manuscript is composed of more than one volume)*, microfilm copy number* (when the manuscript has been microfilmed).
- *Foliation or pagination* (which must indicate the numbering type used).

Manuscript Contents:

- *Manuscript title* (a manuscript can have several titles: authors' title, cataloguers' title, scribers' title, etc.). Sometimes, this information is missed, so it must be provided by the cataloguer.
- *Authors' names* (first name, surname, nickname, celebrity name, *kunya*, *ism*, *laqab*, etc.).
- *Death date* (must be cited in hegira calendar and in its Gregorian calendar equivalence).
- *Birth date*, *existence period, manuscript subject.*
- *Incipit* (first phrases of the manuscript), *Explicit* (last phrases of the manuscript), *Colophon.*
- *Copyist names* (He is a person responsible for the manuscript transcription), *annotations*.
- *Transcription place, transcription date, manuscript summary, contents table, index.*

Physical manuscript description:

Manufacturing material, *number of folios*, *number of lines per page*, *Palimpsest or not*, *signature*, *craftsman identification*, *manuscript condition* (quality), *coverage size*, *conservation state*, *size of folios*, *the writing style*, *drawings*, *illumination*, *ink* (color), *binding* (technology, décor, date), *drawdown*, *loose sheets*.

Additives:

Manuscript reading (*qira'at* in Arabic), *listening* (*sma'at* in Arabic), *Visa* (*ijja'zat* in Arabic: license issued by the master to his disciple to authorize him teach his module), *corrections*.

Copies references:

Copies of the manuscript available in other libraries, *printed copies*, *copy printer*, *publication references*, *translation references*, *sources*, *observation*.

Manuscript history:

Manuscript possession (Persons' names) *and acquisition type* (purchase, loan and donation).

b. Manuscript cataloguing used for the online edition:

In addition to the descriptive elements, 'Internet purists' add to the bibliographic catalog record some information which supports free text information retrieval enriching the record and providing access to the images. We can add, for example:

- *The table of contents*: each line contains a link pointing to the corresponding manuscript image.
- *The index*: It can be made in different ways (automatic indexing, manual indexing).
- *Annotations*: The information reported by the researcher on the image can be added to the catalog in order to improve the manuscript access.

5.3. Discussion

The choice of one cataloguing method rather than another is not always a useful approach—the pragmatic solution in this case is to use both cataloguing modes. Indeed, the capability provided by the computerization of the catalog for information retrieval, inspires us to use the two cataloguing modes at the same time. But, we recommend adding the specific metadata to identify the cataloguing mode as well.

Our strategy is to find a general way to enable the ancient Arabic manuscript encoding using both the volume cataloguing mode and the specimen cataloguing mode. The strategy is to use the bibliographic manuscript cataloguing metadata to which metadata for online indexing will be added. Our approach will obviously generate additional work for the cataloguer. But we expect that implementation of automatic tools will assist in building image annotation and the navigation in the image database.

The process is to select a word or a region on the manuscript image. Then the researcher will undertake the image annotation in an appropriate text field. The latter will be integrated into the index. Thus, the catalog will not include only the conventional metadata entries, but will also use rich indexing of annotations.

6. Ancient Arabic Manuscript Cataloguing Encoding

Encoding of catalog data aims to correspond with the encoding format of the manuscript metadata, whether based on a database management system (DBMS) or an open encoding format like XML. The DBMS option provides powerful access and information retrieval methods. However, the rapidly changing proprietary model of many databases leads to questions concerning the portability and the interoperability of the cataloguing system and its data. As portability and interoperability are primary objectives in the digitization projects, the XML format is more recommended.

In our recent conference paper (Soualah & Hassoun, 2010b) we have described the different encoding formats that are used by major libraries and institutions, in their Arabic manuscripts cataloguing. Below, we describe the most used formats for the Arabic manuscripts cataloguing.

6.1. The Use of EAD (Encoded Archival Description)

EAD was developed in 1993 at the University of Berkeley. The EAD is designed to be a shared tool used by all the archival community to allow diffusion of research data via the Web. The EAD data is a file that describes the contents of documents package held by an archive. The EAD is a DTD created for encoding archival inventories and manuscripts catalogs (Queyroux, 2003). EAD contains 146 elements, divided as shown below:

- *Generic elements* (41): They are common to most texts. They are used for the layout. Example: <p> (paragraph), <table>, <num> (number), <abbr> (abbreviation), etc.
- *Metadata elements* (23): They provide information on the record, or on the inventory. Example: <author> (author), etc.
- *Structural elements* (18): They are contained in segments providing information about the research instrument and describing the corpus of documents. Example: <frontmatter>, <archdesc> (Archive description), etc.
- *Specific information elements* (36): <origination> (producer name), <physdesc> (physical description), <accessrestrict> (access restriction), etc.
- *Entering elements* (12) : <persname> (person name), <corpname> (corporation name), <subject>, <geogname> (geographic name), etc.
- *Localization elements* (16): They are used to describe relationship different parts of the inventory and links to digital documents.

Among those elements, only eight are mandatory: <head>, <eadheader>, <eadid>, <filedesc>, <titlestmt>, <titleproper>, <archdes> and <did>. EAD doesn't impose any representation constraint, but, the encoder has to respect the element names and their hierarchy.

6.2. The use of TEI (Text Encoding Initiative) Manuscript Description

The TEI Manuscript Description (TEI-ms) is TEI application specialized in the manuscript description. The TEI P5 version includes both metadata of both MASTER project and EAMMS project (Centre de Ressources Numériques TELMA, 2008).

The TEI-ms offers two methods to describe the manuscripts. The first method, called the simple method, consists of series of nested paragraphs, composed of short phrases describing the manuscript. The second method, called the complex method, defines the record with a detailed markup structure, which describes finely the manuscript. The first method is intuitive and simple to use. It offers the cataloguer familiar surroundings for manuscript description, but, it also presents serious difficulties for a conceptual indexing. The second method requires the cataloguer to respect TEI-ms structure, but it facilitates the indexing step and the information retrieval.

<msDesc> is the root element of all the manuscripts descriptive elements. It must be placed after <sourceDesc> TEI element. It describes only one manuscript at a time. It contains seven sub-elements described as bellow:

- <msIdentifier>: It defines the needed information to identify a manuscript.
- <head>: It's the header.
- <msContents>: It describes the intellectual content of the manuscript.
- <physDesc>: It contains the physical description of the manuscript.
- <history>: It contains elements describing the history of the manuscript.
- <additional>: It contains additional information about the manuscript (e.g. bibliography).
- <msPart>: This element describes other manuscripts assembled into a single one.

6.3. EAD and TEI-ms insufficiencies

As we have reported in Soualah & Hassoun (2010a, 2010b) the TEI-ms lacks the capacity to describe exhaustively the ancient Arabic manuscripts. We have proposed some adjustments of the TEI in order to overcome the various cataloguing problems. The TEI adjustment concerned the introduction of the Arabic transliteration into TEI, the integration of the different cataloguing modes and a solution to the problem of the old Arabic names description.

The various inadequacies of the TEI-ms in describing the Arabic manuscripts are still valid for the EAD standard. In addition EAD is more suitable for archive fond description. For example, it has no element to describe incipit, explicit, colophon, contents table, index, etc. Therefore, an effort to adjust the two standards is required to adequately describe the Arabic manuscripts.

7. Encoding formats interoperability: An ambitious project

Because our objective is to improve the availability of manuscript images corresponding to a specific user query, there is little interest in providing cataloguing for resources that remain unavailable online. To reach this objective, several solutions are possible to overcome differences in encoding format: the standardization of the encoding format, the use of a mediator format and building domain ontologies to bridge the differences.

7.1. Standardization

The manuscript encoding standardization strategy proposes the establishment of a common cataloguing protocol for Arabic manuscripts, using the same encoding tools used by various conservation institutions. This strategy gives researchers and users access to effectively indexed, standardized data. (Westeel, 2004).

In real life, this solution seems impossible to implement, because of the variability of the current encoding requirements and working methods of the different conservation institutions. Standardization requires deep changes in institutions' habits. So this solution has to be rejected.

7.2. Using Dublin Core as a Mediator Format

The Dublin Core defines a basic set of fifteen metadata. The metadata concerns the digital document content (Title, Description, Subject, Source, Coverage, Type and Relation), intellectual property (Creator, Contributor, Publisher, Rights) and version (Date, Format, Identifier and Language).

We are inspired by the OAI-PMH protocol model and the role of Dublin Core in providing interoperability for that protocol. The idea is that Dublin Core takes the role of a minimum set, which is used as a bridge between the different catalogs in various formats.

The strategy is to generate bibliographic records using the Simple Dublin Core terms, starting from the existing basic catalogs, which may be encoded in different formats. A link is automatically established between the record using the Dublin Core format and the original record via the <dc: relation> element.

Utilizing the Dublin Core structure, the access to the bibliographic records is done in a controlled way by name, by title and by subject. Moreover, keyword access is provided using Dublin Core catalog by using the <dc:description> element. This element is designed to carry the intellectual contents of the digital resource and provides the best candidate for the implementation of the free text indexing system.

Figure 2 describes the structure of the use of Dublin Core to support the interoperability needs of the catalogs of digitized Arabic manuscripts:

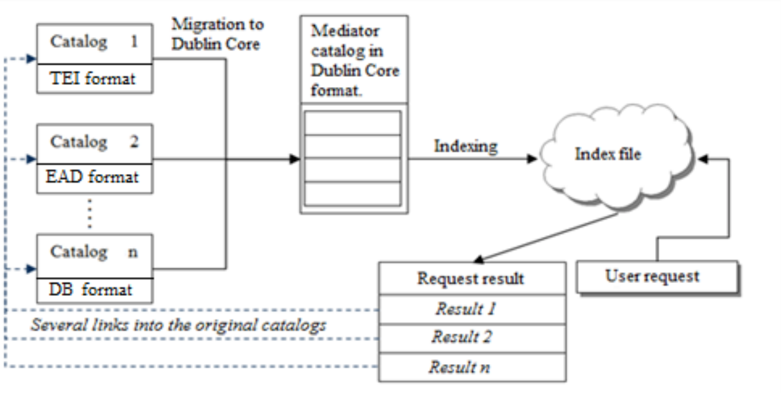

FIG. 2. *Solution structure of the catalogs interoperability using Dublin Core*

In the literature, criticism of Dublin Core lies in its descriptive weakness in relation to some particular aspects of the original documents. Considering the manuscripts description needs, fir instance, Dublin Core has no element to describe the colophon, the incipit, the explicit, or other specialized data.

Nevertheless, the simplicity of its elements allows an easy adaptability to the other formats. The element <dc: description> is a powerful concept, which makes possible to relate the metadata of the original catalogs, which don't correspond to the Dublin Core standard.

The metadata migration starting from the different catalogs towards the Dublin Core mediator catalog will be done thanks to a mapping program. This process requires knowledge of the source catalog structures. Consequently, collaboration with the manuscript conservation institutions is imperative.

Table 1 summarizes the manuscript metadata categories and their correspondence in Dublin Core standard under its current form:

TABLE 1: Ancient Arabic Manuscript description using the Dublin Core standard

Cataloguing metadata of the ancient Arabic manuscript	Status	Corresponding elements (Dublin Core)
Manuscript title	Mandatory	<dc :title>
Manuscript identification	Mandatory	<dc: identifier>
Manuscript manufacture date	Optional	<dc: date>
Language or dialect of the manuscript.	Mandatory	<dc: language>
Author: The responsible of the manuscript intellectual contents	Mandatory	<dc: creator>
Manuscript type (textual, image, etc.)	Optional	<dc: type>
Manuscript subject	Mandatory	<dc: subject>
Copyist: The person who wrote the manuscript	Optional	<dc: contributor>
Artisan, manufacturer, town, etc.	Optional	<dc: publisher>
Temporal and geographical manuscript coverage	Optional	<dc: coverage>
The source format, sheets number, manuscript dimensions, etc.	Optional	<dc: format>
Manuscript image location uniform resource locator (URL).	Optional	<dc: relation>
Manuscript information (copies number, copies localisation, etc..).	Optional	<dc: source>
Public or private rights	Optional	<dc: right>
The manuscript description (homogeneity, etc.), support, ink, the condition of the manuscript, writings, layout, decoration, colophon, binding, the manuscript history (owners, sellers, origin, etc.), readers, listeners, text analysis, summary.	Mandatory	<dc: description>

Access description to the mediator catalog (Dublin Core)

The description of the entire project is not the objective of this paper. The project is based on a multilingual catalog, where the user may make his requests through different languages, mainly English, French or Arabic.

Two access modes for the mediator catalog under the Dublin Core metadata format are envisioned. The first one is the controlled access and the second one is the free access mode:

- **Controlled access**: An implicit indexing involving the author, the subject, the title and the copyist. The index entries will be done by using <dc: title>, <dc: creator>, <dc: subject> and <dc: contributor>.
- **Free access**: In addition to the preceding elements, the index will be enriched by the <dc:description> content elements. The information retrieval system will be based on the Boolean model, described in a previous paper (Soualah & Hassoun, 2010).

The proposed solution ensures adaptability between the original catalogs and the mediator catalog under the Dublin Core format. The considered system does not require any extra effort on behalf of the existing systems. Consequently, all the institutions will continue to produce bibliographic records of digitized Arabic manuscripts in their own formats, and these records will be automatically converted into Dublin Core for the purpose of interoperability.

8. Conclusion

The Ancient Arabic manuscript cataloguing project is a necessary for a better management of manuscripts. It is the key that ensures good visibility of the manuscripts. Indeed, the online availability of the digitized manuscripts could not be useful without an effective search strategy, possible with an efficient catalog implementation.

The Arabic digitized manuscripts cataloguing challenge lies in determining the appropriate metadata protocol. Several prior projects proposed their own approaches. Our work aims to = federate these cataloguing systems, integrating all available metadata for Arabic manuscripts. We further propose to integrate the manuscript structure into the catalog using table of contents, indexes and annotations. It is a careful effort that will allow the cataloguer to increase the accessibility of digitized Arabic manuscripts.

Our work proposes an interoperable catalog solution, making a collective virtual catalog of available digitized Arabic manuscripts using Dublin Core as a base. The virtual collective catalog will allow access to all digitized Arabic manuscripts that satisfy a specific user request, regardless of their physical or storage locations.

The contextual indexing of the catalogs will be based on the building of a domain ontology, used to connect the user to the manuscript. All the requests will be initially be submitted through the ontology, bridging all the catalogs in the different formats. The catalogs will return the result, using the ontology as a basis to translate the results into a form the user can comprehend.

References

Ben Lagha, Sassa. Inforge, Ecole des HEC, Université de Lausanne – Document numérique. Volume 6 – n°1-2/2002 - "Les dossiers numériques" - Publications Hermes sciences – October 2002.

Cédelle-Joubert, Laure and Buressi, Charlette – To tally a project – Review: "Conduire un projet de numérisation" – Editions TEC&DOC – June 2002.

Centre de Ressources Numériques TELMA –TEI presentation – available online at http://www.cn-telma.fr/. Consulted on the 23/05/2008.

Delamarre, Aurélie. Which cataloguing for the contemporaneous manuscripts? Mémoire d'étude – DCB 2004.

Guesdon, Marie-Geneviève and Rodriguez, Nathalie. Les manuscrits arabes, turcs et persans à la bibliothèque interuniversitaire des langues orientales – MELCOM 27, Alexandrie – May 2005.

Humbert, Genviève – Arabic writing tradition – Revue des mondes musulmans et de la méditerranée – November 2002.

IRHT – Institut de Recherche et d'Histoire des Texte – Introductory training to the medieval manuscripts –Booklet, irht – 2006, available online at http://aedilis.irht.cnrs.fr/stage/index

Kaileh, Hala – Digitized Arabic manuscripts online access – doctorate thesis in ICOM – Université Lumière – Lyon 2 – January 2004.

Queyroux, Fabienne. EAD, la description archivistique encodée - Bibliothèque de l'Institut de France. La numérisation des textes et des images : Techniques et réalisations – travaux de recherche – Juillet 2003.

Peccate, Patrick. Roundtable of XML Campus - 28/02/2003 – available on http://peccate.karefil.com/software/metadataCampusXML.pdf. Consulted on the 04, 15th 2011.

Soualah, Mohammed Ourabah and Hassoun, Mohamed. (2010a). The TEI P5 Manuscript Description Adaptation for cataloging digitized Arabic manuscripts. TEI Conference and Members' Meeting 2010. Thu 11 Nov to Sat 13 Nov 2010, Zadar, Croatia.

Soualah, Mohammed Ourabah and Hassoun, Mohamed. (2010b) A multilingual online access to the digitized Arabic manuscripts – 13ème Colloque International sur le Document Electronique : Document électronique entre permanence et mutation – 16/17 November 2010 – INHA, Paris.

Westeel, Isabelle. Colloque EBSI/ENSSIB. Patrimoine et numérisation : la mise en contexte du document. Bibliothèque municipale de Lille. Montréal. 13-15 october 2004.