A framework for basic administrative metadata in digital libraries

ZHENG Qiaoying & WANG Shaoping

Information Research Institute, Shanghai Jiaotong University Libraries,
Shanghai 200240, China

Received Feb. 22, 2008
Revised May 20, 2008
Accepted May 29, 2008

Abstract Administrative metadata means the expansion of the metadata research to the administrative level of resource development. Based on the basic administrative sections in the information resource lifecycle (IRL), the framework for basic administrative metadata (FBAM) is helpful in constructing open interoperable platforms for acquisition, processing and services of information resources in digital libraries. It facilitates the seamless communication, the cooperative construction and management, and the sharing of digital resources. The formulation of FBAM follows the principles of modularity and openness that promote interoperability in resource management. It also adopts the structured methodology of information system design, with which the FBAM data model is developed in conformity with <indecs> and PREMIS. The capabilities of FBAM are driven by a metadata repository with administrative information that is contained in FBAM records.

Keywords Administrative metadata, Digital library, Digital resources, Interoperability, Openness, Information resource lifecycle

1 Introduction

Broad attention has been paid to and in-depth researches have been conducted by information professionals on metadata that describe the content of resource, i.e. descriptive metadata (DM), such as Dublin Core (DC). Meanwhile, people came to realize that metadata are able to describe not only various kinds of content, but also "data" in other different aspects of the processes of information development, utilization, and management[1].

The description of management policies and mechanisms of resource in digital libraries by administrative metadata (AM) is an extension of metadata to the administrative level of the resource. Normalized and open disclosure of administrative data in acquisition, collection, and utilization processes of the resource by AM is helpful for the user or the third party system to share local resources conveniently and smoothly. In addition, it facilitates various information systems to exchange management policies and mechanisms effectively and to promote interoperability

Correspondence could be addressed to Zheng Qiaoying (E-mail: zheng@mail.lib.sjtu.edu.cn) or Wang Shaoping (E-mail: spwang@mail.lib.sjtu.edu.cn)

CJLIS
Vol. 1 No. 1, 2008
pp 27–37
National Science Library,
Chinese Academy of
Sciences

between different systems through AM with appropriate security measures. The actual representation mechanisms for the resource or the collection of resources in an information system can be formed with the help of the binding of AM with the said system. It is easily customized or modified through changes made to the AM. Such mechanisms can also support automatic searching and matching of the resource or the collection of resources and parsing of resource structures by intelligent agents, thus integrating, on top of that, multiple collections of resources. Similarly, the actual operational mechanisms for the management policies of an information system can be formed with the binding of AM with the said system, and likewise, can also be easily customized or modified through changes made to the AM. Such mechanisms can also support automatic searching, parsing, and proper processing of the management policies and regulations by intelligent agents.

2 Different kinds of administrative metadata (AM)

It is necessary for the information professionals to have a prior knowledge about the administrative functions of metadata while studying and developing DM. Administrative elements or modules such as rights, date valid and date available in DC, the application template such as service metadata in ROADS[2], meta-metadata in IEEE LOM[3], etc. are stipulated in those DM schemes.

It was the administrative components[4] (AC) of Dublin Core metadata initiative (DCMI) that first clearly put forward an AM scheme. The scheme worked out a metadata set that aimed at the effective management of DC metadata.

The Digital Library Federation's Metadata Encoding and Transmission Standard (METS)[5] have divided the metadata family into three categories: Namely, 1) The descriptive category, 2) the administrative category, and 3) the structural category. The AM category is again subdivided into four main forms: a) Technical metadata, b) intellectual property rights metadata, c) source metadata, and d) digital provenance metadata. Although not actually being a metadata scheme, METS has built up a broader framework of AM in coverage than AC.

Cornell University has sponsored a website for AM studies[6]. This website assembled AM projects of electronic resource management, e.g. DLF ERMI[7], but with a focus on acquisition of digital resources, especially those of electronic journals.

The OAIS reference model[8] formulated by the Consultative Commission of Space Data Systems (CCSDS) has highlighted the concept of preservation metadata that describe information related to long-term preservation of digital resources such as content, tagging, metadata, context, digital resource provenance, etc., among them many are related to AM.

In IT community, several projects with the scheme of AM, such as P3P[9], PICS[10], ODRL[11], and XACL[12], have currently been developed. These projects describe specific management mechanisms of information systems in resource processing and utilization.

All the above-mentioned AM schemes and projects, each of which generally has its own vocabulary, syntax, encoding schemas, etc., were designed or formulated separately for specific application situations. Some of them are metadata that manage metadata (DM), some are those managing acquisition of resource, some are the ones that manage technical processing of resource, and some manage service procedures of information systems, and so on. All those metadata are, if viewed horizontally, connected with the administrative tiers of resources. However, if viewed vertically, they are only concerned with the description of management mechanisms or policies of a certain management sector in the whole information resource lifecycle. If we could not coordinate the development of different kinds of AM, it would be difficult to develop open and interoperable digital library systems, at least at the administrative level for information resource. We could not avoid reinventing the wheel every time we should have an AM set to describe a specific management situation, and we could not ensure such AM would promote interoperability of information systems without certain coordinating mechanisms or principles. In order to facilitate interoperability, it is necessary to have the following prerequisites established:

- A unified "core" element set is required for AM with the same or similar functions; and
- Existing AM elements can be reused as much as possible for AM with different functions.

3 A unified framework of AM

A unified framework of AM in line with the information resource lifecycle (IRL) will help set up a core element set, and a mechanism favorable for reuse of existing AM elements.

IRL gives us an all-round view of information management in libraries and information service centers with the dynamic process of creation, transfer, transformation, and utilization of information resource. We can roughly describe the resource movement in libraries and information service centers as follows:

- Resource is acquired and aggregated into digital or non-digital objects.
- Non-digital objects are processed and transformed to digital ones. The digital objects are combined to form compound ones; and digital and compound objects can be further organized into a collection with the help of a knowledge organization system (KOS). All the resources, whether they be digital or non-digital objects, whether they be information content or information objects or resource collections, can be described by DM.
- Resource collections and metadata that describe resource can be registered, archived, and delivered to users through service sectors.

The resource, as described above, has actually gone through the following three functional sections: Namely, acquisitions, processing, and services. That is to say, a

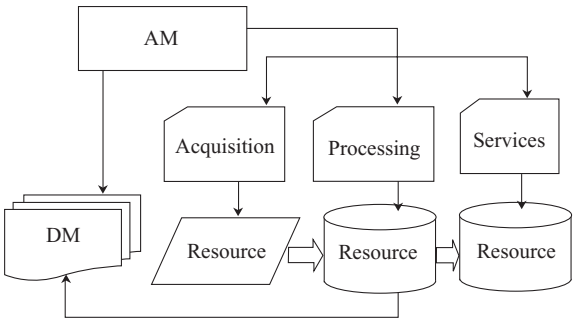

Fig. 1 A simplified diagram of information resource lifecycle with AM and DM.

process encompassing from resource creation to its utilization by users. All these three sections are controlled by management mechanisms and policies that can be described by AM, as shown in Fig. 1, a simplified diagram of IRL. Rather than restricted by any given management situation, IRL provides the possibilities of analyzing AM with an all-round point of view of information management that is desirable for capturing, on the whole, the basic characteristics of AM. The three sections in the IRL, reduced in size, have narrowed the scope of objects studied on the one hand, and increased commonality among different objects within the given scope on the other hand. It is easier to "round up" in each management section a Basic Administrative Metadata that means a sort of "basic" metadata built on a core AM element set for a given management stage of resource, reflecting commonality of management mechanisms on that stage, and extending as required by actual management situations. A framework for Basic Administrative Metadata will provide a context in which all "special" AM are no longer isolated and unrelated ones, but an organic integration that can be easily merged. In line with IRL, we can arrange such isolated "special" AM schemes and projects as described in Section 2 in the way shown in Table 1.

We can see from the above analysis that the IRL-based AM framework provides us with a core element set that can be readily extended to meet the needs of specific resource management situations, and that extant AM schemes or projects can be reused because they are all covered and linked by such a framework.

Table 1 AM arranged in line with IRL

IRL administrative sections	Acquisition	Processing	Services
Some administrative functions	Resource evaluation Rights management Resource registry	Metadata administration Digitization Collection organization	Resource evaluation Rights management Access control and privacy protection
Some relevant AM	PICS ODRL DLF ERMI –	AC METS AM – –	PICS ODRL XACL P3P

4 FBAM data model

4.1 Investigations and requirements analysis

The first step toward the establishment of an FBAM is the in-depth investigation and analysis of requirements for resource management data in library and information services, especially in digital libraries. The authors selected 23 libraries and information service centers for questionnaire investigations, and followed up with a field investigation at one of the sampled institutions, Shanghai Library which plays a role both as a public library and as an information service center.

In formulating the pre-selected administrative data elements for the questionnaire, we investigated and analyzed the resource management stages in those libraries that we visited, and factored the international standard series of ISO 8459 *Information and Documentation — Bibliographic Data Element Directory*[13] and their Chinese equivalents of GB/T19688[14] into our consideration. Both of them cover the following five essential parts of library operation, namely, 1) Interlibrary loan applications, 2) acquisitions applications, 3) information retrieval applications, 4) circulation applications, and 5) data elements for the exchange of cataloging and metadata. The standard series lists respectively directories data elements that support systems of interlibrary loan, acquisitions, information retrieval, circulation and cataloging. It discusses relationships among the data elements, and provides examples of usage.

Statistics showed that most of the pre-selected data elements were qualified by the respondents, and the elements with higher percentage of choice were generally referenced for future AM elements.

On the basis of our questionnaire findings, the authors also took a field investigation at Shanghai Library, an organization featuring both libraries and information institutes. Our purpose was to delve more deeply into their practices of information acquisition, processing, and services in a real situation of a library in operation. First-hand materials concerning policies/requirements of administration and management information in relevant sections of the workflow have been obtained so as to make up for the lack of the authors' direct communication and personal contacts with those pertinent library staff during the questionnaire investigation period. The result of the investigation is a summary of the working processes and relevant basic administrative data.

In addition, we created E-R (entity-relationship) models of management information requirements with the help of the regular structured methodology that is being practiced in the design of information systems.

4.2 Establishment of an FBAM

The FRAM data model is based on <indecs>[15] and PREMIS[16] models. Entities and relationships of the E-R analysis are mapped into the space of AM framework to constitute the data model of the framework. There are basically three types of mapping situations:

- Entities and relationships mapped to the space of AM framework with relevant entities and relationships in <indecs> or PREMIS;
- Entities and relationships mapped to the space of AM framework without relevant entities and relationships in <indecs> or PREMIS; and
- Attributes of entities in E-R models mapped to entities in AM framework.

For example, there are following entities and relationships in our E-R model for the services section of IRL:

Entities: user, service provider, third party, infoResource, service policy, contract;

Relationships: asks for service, provides service, signs, acquires, executes, searches, etc.

If mapping is represented by "→", we have

user, service provider, third party → agent (PREMIS), or party, person (<indecs>);

InfoResource → intellectual entity, object (PREMIS), or intellectual property, creation <indecs>;

Asks for service, provides service, signs, acquires, provides, executes, searches, etc. → event (PREMIS or <indecs>).

There are no objects of mapping for *service policy*, and *contract*.

Attribute "rights" of *service policy* represents rights of users to utilize information resources, which can be mapped to "rights" of <indecs> or PREMIS.

The entities and relationships without objects of mapping in our E-R models were further abstracted on the basis of their characteristics and attributes. For example, *resource acquisition policy* and *order* in the E-R model for the acquisition section, and *processing regulations and contract* in the E-R model for the processing section were all unable to be mapped to proper objects. *Service policy* in the service section together with *resource acquisition policy* and *processing regulations* reflects action criteria regarding specific tasks to realize information resource management, which can be combined and reduced to "*policy*".

Finally, in the space of the AM framework, following entities and relationships were established:

Agent that represents persons or corporate bodies such as users, service providers, collectors, etc. that play a role in IRL;

Event that is a result of abstraction of relationships operating between *agent* and *resource*, or between *agent* and other entities (e.g. users search resource, which describes any meaningful action or occurrence in IRL);

InfoResource that includes information resources, whether digital or not, with various representation formats;

Rights that were attributes of entities in the E-R models, and have been enhanced to an entity on account of its important connections with such entities as *agent*, *event*, and *infoResource*, resembling *rights* in <indecs> and PREMIS;

Policy that is an incorporation of *acquisition policy*, *processing regulations*, and *service policy*;

Contract that reflects agreement on rights relationship between relevant agents, a new entity without object of mapping;

OrderBills that are bills in acquiring or collecting resource, a new entity without object of mapping;

AcquisitionSource that is any source from which resource can be acquired, a new entity without object of mapping;

AcquisitionDecision that is a decision whether a specific object of resource should be acquired, a new entity without object of mapping;

EvaluationLabel that represents a result of evaluation of the given resource by a certain resource evaluation system, a new entity without object of mapping;

ServiceRequest that is a request delivered to any information service provider for a desired service, a new entity without object of mapping; and

ServiceMode that is the approach in which an information service institute provides a service, a new entity without object of mapping.

Agent has relationships with *InfoResource*, *Policy*, *Contract*, *OderBills*, *AcquisitionSource*, *AcquisitionDecision*, *EvaluationLabel*, *ServiceRequest*, and *ServiceMode* through *Event*, and is entitled to related *Rights*. *Agent* may also have relationships with *Rights* through *Event* (e.g. agents "discuss" permissions for utilization of given resources). Processing of resource and its metadata is reflected by the chain "*Agent—Event—InfoResource—Rights*".

On the other hand, *Policy*, *Contract*, *OrderBills*, *AcquisitionSource*, *Acquisition-Decision*, *EvaluationLabel*, *ServiceRequest*, and *ServiceMode* are all related to specific resource. Fig. 2 shows the AM framework data model with its main entities and relationships.

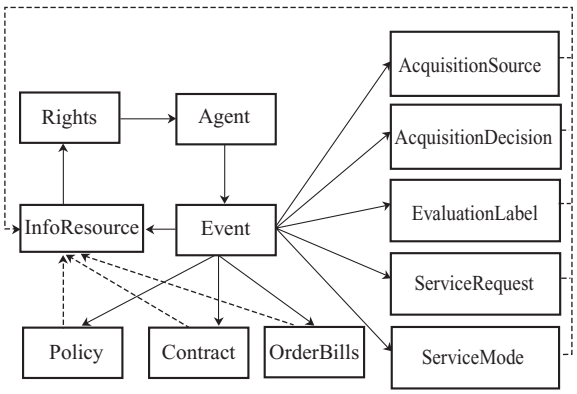

Fig. 2 AM framework data model with its main entities and relationships.

The design of an FBAM has followed the principles of modularity and openness, favoring interoperability in resource management.

- AM in an FBAM manages the processes of resource acquisition, processing and services, and is divided into three relatively independent, but closely interrelated modules of acquisition, processing and services that are in line with the workflows of resource management in a digital library. Elements in every module are grouped in three classes, i.e. common elements, special elements, and local elements. The common elements are those common in all three modules, e.g. *Agent* (including users and staff of organizations concerned), *InfoResource*, *Rights*, etc. The special elements are those in specific modules (e.g. *AcquisitionSource* in the *acquisition module*, and *ServiceRequest* in information service module). Local elements are those elements extended by specific applications to meet the needs of the local application environment.
- FBAM is open to other well accepted or applied standards and schemes with the aid of reusing and wrapping. For example, the attributes of *Agent* have reused vCard elements, other commonly used metadata such as DC and MARC can be either wrapped or linked for the description of *InfoResource*, and the metadata language of PICS can be embedded for evaluation of digital resource. Futhermore, FBAM is also open to extensibility mechanism, with which any application can extend elements in the original framework vertically or horizontally. The vertical extension is a subdivision of the existent elements, i.e. qualification of the elements in accordance with the specific application context to derive from more special sub elements or qualifiers. And by horizontal extension we mean establishment of new elements or element sets semantically not coinciding with the existing element set to produce new functions.
- The design of an FBAM is not limited to the internal functional requirements of a specific application, but gives precedence and support to its interoperability. The interoperability is based first and foremost on XML coding system, which is well accepted in the digital library community. Secondly, the basic logical framework suitable for resource management, which has established on the basis on the extant standards and systems, supports the interoperability. It facilitates seamless interchange of AM between different applications and systems.

5 Capabilities of FBAM

Generally speaking, FBAM can be used as platforms for management of resource acquisition, content creation and processing, resource integration and publication, and can describe specific management mechanisms and policies related to these management situations.

To specify FBAM capabilities, we consider a metadata repository that acts as a central storage of metadata for an information service organization. In addition to DM describing content of information resources, AM are needed in the repository to contain information about DM (e.g. the source of DM, the date and responsible party of record creation, modification or deletion, ingest-specific information, etc).

National Science Library, Chinese Academy of Sciences

In FBAM configuration, metadata management in a repository belongs to the processing module that deals with processing of resource and its metadata. There are those core elements such as Agent, InfoResource, Event, Rights, etc. and their qualifiers in the processing module that need a few words of explanation:

- *Agent* can reflect any party responsible for record creation, modification or deletion.
- *InfoResource* can wrap a DM record or be linked with an external DM record with the help of the common qualifiers "wrap", that is the same as MDWrap of METS, and "link".
- *Event* can represent actions in record management such as creation, update, batch processing, etc.
- *Rights* can specify qualification for an *agent* to take a given action.

Table 2 An example of FBAM records

Element	Element qualifier	Element qualifier	Encoding scheme	Value
Identifier	–	–	–	ADM20080205
InfoResource	InfoResourceIdentifier	–	–	METI000001
	Wrap	–	DC	
	–	xmlData	–	<dc:title>Investment and fund-raising issues in China's large environmental protection infrastructure projects
	–	–	–	</dc:title>
	–	–	–	<dc:creator>Liu Xue'an
	–	–	–	</dc:creator>
				…
	InfoResourceCategory	–	–	Record
	Source	–	–	SJT
	InfoResourceType	–	–	metadata.bibliographic
	TechnicalInfo	Format	–	DC
		CharacterSet	–	UTF-8
Rights	RightsIdentifier	–	–	METR000001
	AccessRights	Condition	–	Cataloger
		ActionRights	–	Creat;download;modify;upload
Agent	AgentIdentifier	–	–	META000001
	AgentName	–	–	Huang Min
	AgentType	–	–	Person
	AgentRole	–	–	Cataloger
	LinkingEventIdentifier	–	–	METE000004
	LinkingRightsIdentifier	–	–	METR000001
Event	EventIdentifier	–	–	METE000004
	EventType	–	–	Metadata processing
	Action	–	–	Create
	ActionDataTime	–	ISO8601	2007-08-09T01:30:00+08:00
	Action	–	–	Modify
	ActionDataTime	–	ISO8601	2007-08-11T01:35:00+08:00
	LinkingInfoResourceIdentifier	–	–	METI000001

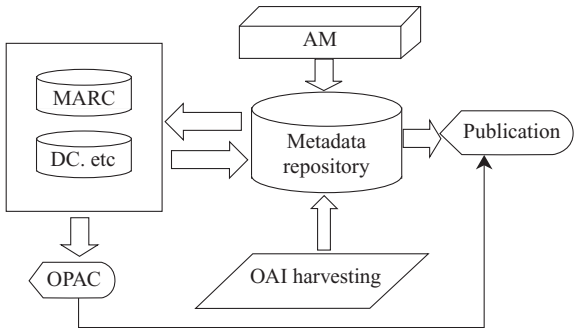

Fig. 3 AM used in a metadata repository.

- *ActionDateTime*, a qualifier of *event*, can describe the date or time when an action takes place.
- *LinkingInfoResourceIdentifier*, a qualifier of *event*, can link an event to the relevant InfoResouce.
- *LinkingEventIdentifier*, a qualifier of *agent*, can link an agent to the relevant event;
- *LinkingRightsIdentifier***,** another qualifier of *agent*, can link an agent to the relevant rights.
- *Source*, a qualifier of InfoResource, can describe the source of DM.

Table 2 is an example of FBAM records that delivers administrative data of a DC record with its creation and modification information. The qualifier "wraps" acts as a wrapper that embeds an XML-encoded DC metadata record.

The elements and qualifiers can be properly encoded for processing in the local system or exchange with other information systems. Fig. 3 describes how AM is used in a metadata repository that stores metadata from local DM systems of MARC, DC, etc., and those harvested from external metadata sources.

6 Conclusions

AM is the expansion of metadata to the tier of resource administration and management. FBAM has provided a logical framework of AM that can unite existing and future AM under the same roof of interoperability with coordinating mechanism as displayed by IRL.

AM has come out of the requirements of open management of resource in digital libraries. It is unimaginable to promote resource sharing without the very sharing of policies and pertinent mechanisms governing resource management. The studies of AM in the basic sections of IRL will help enhance standardization and normalization of resource management data, and promote communication and sharing of administration and management information, which will lead to resource sharing at a still higher level of fulfillment.

National Science Library, Chinese Academy of Sciences

References

1 Zhang, X.L. Studies and Applications of Metadata (in Chinese). Beijing: Beijing Library Press, 2002.
2 Resource Organization and Discovery in Subject-based Services: ROADS. Retrieved on April 12, 2008, from http://www.ukoln.ac.uk/metadata/roads/.
3 IEEE Standard for Learning Object Metadata. Retrieved on April 12, 2008, from http://ltsc.ieee.org/wg12/par1484-12-1.html.
4 AC-Administrative Components: Dublin Core DCMI Administrative Metadata. Retrieved on April 12, 2008, from http://www.bs.dk/standards/AdministrativeComponents.htm.
5 METS: Metadata Encoding & Transmission Standard. Retrieved on April 12, 2008, from http://www.loc.gov/standards/mets/.
6 A Web Hub for Developing Administrative Metadata for Electronic Resource Management. Retrieved on April 12, 2008, from http://www.library.cornell.edu/cts/elicensestudy/home.html (no longer maintained).
7 Jewell, T. D., Anderson, I., & Chandler, A., et al. Electronic Resource Management: The Report of the DLF ERM Initiative. Retrieved on April 12, 2008, from http://www.diglib.org/pubs/dlf102/.
8 Reference Model for an Open Archival Information System (OAIS). January 2002. Retrieved on April 12, 2008, from http://public.ccsds.org/publications/archive/650x0b1.pdf.
9 The Platform for Privacy Preferences 1.1 (P3P 1.1) Specification: W3C Working Group Note 13 November 2006. Retrieved on April 12, 2008, from http://www.w3.org/TR/P3P11/.
10 Platform for Internet Content Selection (PICS). Retrieved on April 12, 2008, from http://www.w3.org/PICS/.
11 Open Digital Rights Language (ODRL): Version 1.1. Retrieved on April 12, 2008, from http://www.w3.org/TR/2002/NOTE-odrl-20020919/.
12 XML Access Control Language. Retrieved on April 12, 2008, from http://www.trl.ibm.com/projects/xml/xacl/xacl-spec.html.
13 International Organization for Standardization. Information and Documentation — Bibliographic Data Element Directory, ISO 8459. Geneva: ISO, 1998–2002.
14 General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Information and Documentation — Bibliographic Data Element Directory, GB/T19688 (in Chinese). Beijing: China Standard Press, 2005.
15 Rust, G., & Bide, M. The <indecs> Metadata Framework: Principles, Model and Data Dictionary. 2000: 13–15.
16 PREMIS Working Group. Data Dictionary of Preservation Metadata. Dublin, Ohio: OCLC, 2005:1-1. Retrieved on April 12, 2008, from http://www.oclc.org/research/projects/pmwg/premis-final.pdf.