Research on Metadata Schema Registry System Oriented Knowledge Service

Peng Wu ,Beibei Zhang
Department of Information Management
School of Economics and Management
Nanjing University of Science &Technology
Shaohua Qiang
School of Management Science and Engenieering
Nanjing University of Technology

School of Economics and Management, Nanjing University of Science &Technology
No 200,Xiao ling wei Street, Nanjing ,China

Abstract: -With the arrival of knowledge service age, the integration of information resources of Legacy System will turn from system level to standard level. Metadata schema registry will be the research hotspot of library science and intelligence science, and it can solve the heterogeneous problem between information systems. This paper is faced to the requirements which knowledge service have of integration of heterogeneous information resources in legacy information system, and proposed the application model and system framework of metadata schema registry system.

Key-Words: - Metadata schema registry ,Heterogeneous information resources, Knowledge service ,Legacy System

1   Introduction

Establishment and use of metadata schema are important tache to solve the heterogeneous problem and realize the sharing of information between information systems.Now the problems those metadata field faces are not the lack of projects. To the contrary, there are too many competing projects existed, and each project needs to be standardized. The chaotic situation promotes the production of metadata schema registry. Resources describer, network search developer and interoperable system designer need to know the correlative resource description format and the exact metadata project definition. Metadata schema registry makes for improving the expansion and interoperable capabilities of discovery of network information resources.It can help people to acquire the authoritative version information about a metadata project and the exact definition of project elements timely.

As Rachel Heery pointed at the DC-2005 Registry global meeting of the Working Group, integration of information resources is not the simple conversion of data level any more, but the value-added sharing and compatibility based on the metadata schema. Because of the varieties of application demands and the diversification of development main bodies, multiple metadata models will continue to exist in the real environment which aims at different description purposes. Even in the same area they also can have different requirements of metadata due to different purposes, and this can lead to interoperable difficulties between metadata models. In this way, realizing the inter-compatibility between different metadata schemas, and then realizing the inter-access and inter-search between recorded data that according to metadata schema become the problems to be solved currently.

2   Research Status

Metadata schema registry is proposed by DCMI Dublin Core Metadata Initiative , and the objective is to provide new standard for registration and search service of existed standard for users. In the fields of library science and intelligence science, the research on metadata schema registry has been a hotspot. In the foreign countries, the application field of metadata schema registry is wide, such as the environment field, education field and government field. But in the domestic country, the research on the metadata schema registry is mainly confined to a certain field, such as digital library field and geographic space field. At present, many metadata schema registry system based on Internet have set up in the foreign countries, most of which are built based on ISO/IEC11179. They are all systems which can

issue, register and search the definition information of metadata schema and the criterion of coding, conversion and application, and some of systems support automatic registration and publication, which can facilitate the sharing and management of information resources in fields. There are the brief introductions about the foreign metadata schema registry system in Table 1. There is little domestic research on metadata schema system. There are only seven papers about it in CNKI database search (search title is metadata schema registry or register and system), most of which are introductory ones. Metadata schema registry system which developed in the domestic country is still not mature. Metadata schema registry systems which are using include the Chinese mirror of DC metadata schema registry system and Chinese ecological metadata management system which developed by Chinese Academy of Sciences.

From the foreign and domestic research, because metadata schema registry system involves library science, intelligence science and computer science and so on, the difficulty of integrated use is higher. The foreign researches are mainly concentrated in the application of metadata schema registry system in the fields of information resources management and information service and so on. The domestic researches are on the beginning stage of introduction of foreign advanced theories and simple explorations. Both in the foreign and domestic countries, heterogeneous information resources development faced to legacy system and research on metadata schema registry system of knowledge service lack, and theoretical and technical embedded research more lack.

Some researchers maybe thought about the demands ,concept or implementation of knowledge service ,or solution of heterogeneous information resources of the view of software system isolated. In the dynamic and distributing network information environments, we need the new solution based of the data schema, so we integration the demands of knowledge service for heterogeneous information resources and the theory of metadata schema registry based technology of software engineering to construct the framework oriented service.

Table 1 introduction of foreign metadata schema registry system

name	System sorts	Registered modes	Application fields	Governing unit
ROADS Metadata Registry	Single naming domain	Artificial	Interoperability of inter-gateway	UKLON ILRT
DESIRE Metadata Registry	Inter-naming domain	Artificial	Description of information resources pool	JISC UKLON
CORES Metadata Schema Registry	Inter-naming domain	Artificial	Description of information resources pool	UKLON
DCMI Metadata Registry	Single naming domain	Artificial	Network Information Resources	DCMI
ebXML Registry services	Distributing	Automatic	Commerce Fields	OASIS
UDDI Registry	Distributing	Automatic	Application of industry groups	CNNIC
GoXML Registry	Distributing	Automatic	Commerce Fields	XML Global Technologies, Inc.
German Metadata Registry [10]	Inter-naming domain	Artificial	Natural Science	
Schemas Registry	Inter-naming domain	Artificial	Network Information Resources	UKOLN
Environmental Data Registry [12][13]	Single naming domain	Artificial	Environmental Science	EPA
GILS	Single naming domain	Automatic	E-government	U.S. government

In the process of government informatization and enterprise informatization in domestic country, many legacy information systems are brought, plentiful information resources are still stored in the relatively isolated department information database and risk of "Information Island" is prominent. How to perform the function of legacy information systems economically and scientifically and promote the inner

information resources to keep and add value have been problems to be solved. And metadata schema registry system can solve the problem from the angle of compatibility of data standard. Comparing to other integrated projects from data and system level, it has features of low-input and better sustainable development.

3. Metadata Schema Registry Model

Metadata schema registry system is a system which is used to store, organize, manage and share metadata. It does not include the real data, and just provides definition information, origin and location about metadata. This system provides the link of relative support about metadata. When it is used with the idiographic information database, the system can enable users to better understand and access data, and make information more meaningful.

3.1 Process of metadata schema registry model

At present, there are many types of metadata schema registry system. According to different classifications, metadata schema systems can be divided into different sorts. According to the naming domain of system management, systems can be divided into single naming domain metadata schema registry system, inter-naming domain metadata schema registry system and distributing metadata schema registry system. Also, according to automatic and artificial upload of uploaded templates, systems can be divided into automatically registered metadata schema registry system and artificially registered metadata schema registry system. In this paper, the research model is mainly automatically registered metadata schema registry system of inter-naming domain.

A standard and canonical metadata schema registry system can increase the usefulness and sharing of metadata. The conceptual model of metadata schema registry system is as the figure below:
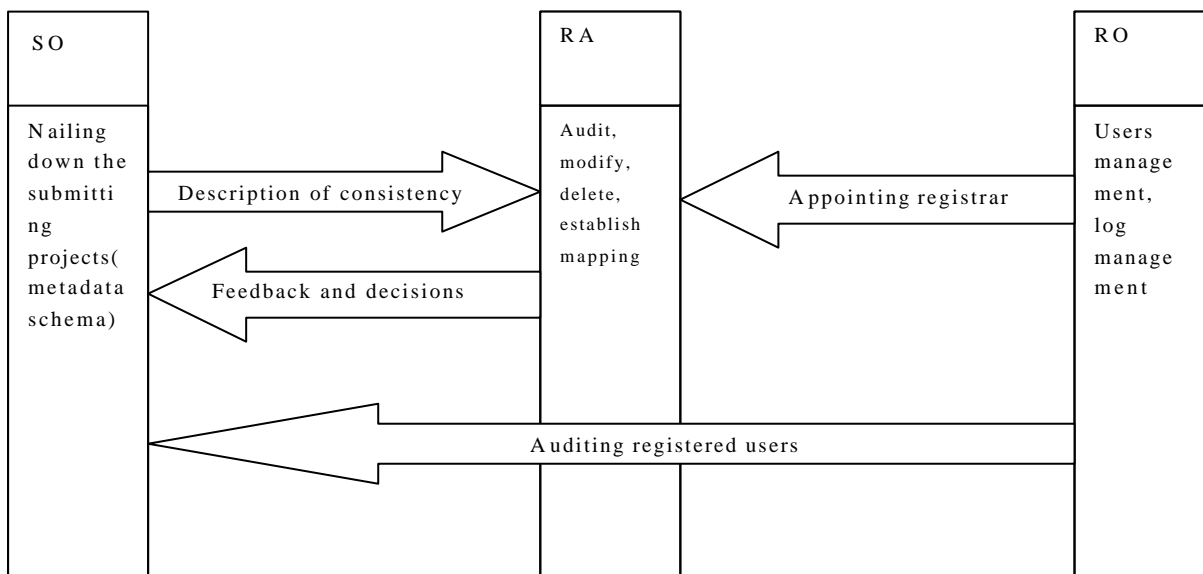

Fig 1 conceptual model of metadata schema registry system

(1)Registry Agency (RA)

Registry agency (mainly are the registrars, perhaps include a series of rights organizations such as control committee) is mainly responsible to audit, modify and delete the metadata schema which submitted by the submitting organization, and establish mapping of same semantic elements between different metadata schemas. The registrars of some large metadata schema registry systems are generally appointed by the control committee.

(2) Responsibility Organization (RO)

Responsibility organization is an authoritative organization of thematic fields of relative registered items which is appointed by a certain organization. It

can ensure the consistency and coherence of relative registered items, coordinate the essential attribute of relative content submitted by submitting organization, and be responsible for the integrity and accuracy of registered items' attribute values, such as semantics and value. For example, the administrators of metadata schema registry system are responsible to maintain the management of registry system users and log and appoint the registrars.

(3) Submitting Organization (SO).Submitting organization should accord to the membership of registry system, follow the registry process, and put forward registry requirements to the registry system. It should submit the necessary, coincident and

correlative information to the registry system, and ensure its quality. In this paper, submitting organization is registered users. According to a series of proves submitted by users, administrators judge if they accord with the terms of submitting organization.

(4) Users. Registry system needs to establish the rules that other users use system, and provides the use of search of public open content. In the antitype system , the users can only inquire of elements or browse, inquire of mapping relation between different standards, and can not submit between different standards, and can not submit, modify and delete system content.

3.2 Function of metadata schema registry model

Here we show the functions of metadata schema registry system:

(1) Supporting metadata interoperability. Metadata interoperability is mainly finished by mapping and conversion between metadata schemas.

(2) Helpful to the management of metadata schema. Through the register of metadata models and correlative information of elements, MR system hackles and records the evolution of metadata schema, and supports the management of metadata schema.

(3) Promoting the reuse of metadata schema and facilitating the establishment of metadata application projects. MR system is open, through the functions of inquiry and issue of MR system, the visibility and accessibility of registered metadata models are improved. In short, MR system supports the share of metadata modes, and reduces unnecessary redesign and rework of metadata modes developers.

(4) Supporting metadata mining. Data mining can find potential and possible data model and internal relations, rules, development, trends from a lot of structured data, and these data is often stored in the form of structured static database. Similarly, metadata mining can also find information that users demand from resource contents. But the tool of mining here is metadata, and it include the mining based on metadata reuse and XML/DTD, mining based on ontology, mining based on words and expressions analysis. At this time, detailed information about varieties of metadata of metadata registry system will provide help to metadata mining.

4 Metadata Schema Registry model oriented knowledge of heterogeneous information resource

Here we show the basic Metadata Schema Registry Model of Knowledge Service Metadata in table2. This model is of an interactive web site/service which provides all the necessary information required for all units and suppliers to understand the metadata and vocabulary standards for special applications. in this model, there are   five layers, they are user layer, service layer, bindin g layer ,profile layer ,and schema layer, each block represents a distinct kind of information; and the arrows represent flows of information at design time.

4.1 User layer and Service layer

In the user layer, this model offer the interface of knowledge   services   such   as   heterogeneous information index or aggregation ,content mining. In the service layer, the model transfer the demands of user to specific servers such as feeds and channel services, schema/services registries, link resolution terminology services and so on ,these service offer the elaboration model of heterogeneous information.

4.2 Binding layer

The binding layer will document the relevant binding types for all the elements in the MDM standard enabling them to be referenced and reused.

The demonstrator will document html and RDF bindings types. There are several ways in which metadata can be encoded and represented. For example, Dublin Core embedded into HTML meta tags. Each of these is termed a "binding model".

A "binding model" specifies the syntax used to encode the metadata by "Application Binding Interface" . Some application profiles can be represented usefully with more than one binding type; and most binding models will be applicable to several application profiles. The resulting binding specifications are represented in the model by the "Application Bindings" block. This combination of application profiles and binding types will mean that software developers will be able to build software that complies with a wide range of metadata standards, some of which are conceptually incompatible, while retaining system compliance.

 The metadata records associated with the resources held by an organization, users does not define how resources   or   metadata   are   stored.   However, interoperable interfaces to these resources require the exposure of records that conform to specific MDR bindings.

4.3 Profile layer

In profile layer, there will be several metadata application profiles for a resource, so a one-to-many relationship will exist between "resources and their metadata" and "domain application profiles". They will overlap - for example, the element "Title" is likely to appear in all or most application profiles. Clarification of the content of each application profile will allow the different user communities to take into account their specific needs without needing to be

concerned about other communities' metadata. Some domain application profiles may apply to more than one service.

Application profiles will be defined in human-readable form and will have specific bindings that allow them to be machine operable and the metadata registry will capture all of the Application Profiles used for special services and content resources allowing them to be interrogated and reused as templates for new applications.

4.4 Schema layer

In Schema layer, to construct standard metadata schema namespace ,a metadata model defines the underlying structure of metadata records used within a particular business, and the relationship of the constituent parts of a description to the resource(s) being described. The model determines the metadata

terms required to make up a valid description. For example, the Dublin Core Abstract Model defines the use of "elements", "element refinements", "encoding schemes", etc.

The encoding schemes will comprise a series of bindings and potentially lists of controlled vocabulary, and the terms used to construct metadata descriptions within a particular application profile must be declared before they can be used, such declarations can take the form of human-readable documents or machine-readable 'schemas' or both.

Typically, a core set of terms is made available by a global standards body for widespread use (e.g. the DCMI or IEEE). Additional terms need to be declared separately by the specific domain users. These will all be documented in the model registry.
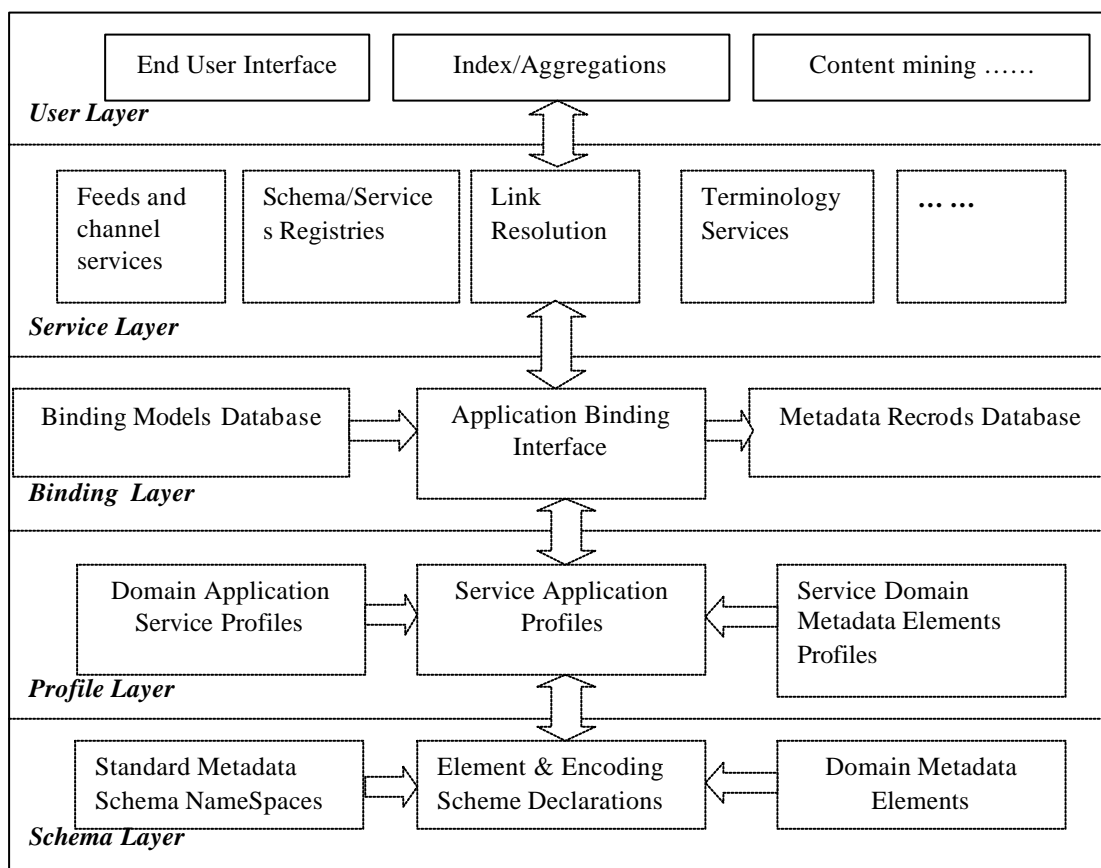

Fig2 Metadata Schema Registry Model oriented knowledge Service

4   Conclusion

Integration .of heterogeneous information resources of Legacy System is always of a difficult problem in the time of 'Information Service'. In the time of 'knowledge service', how can we resolve the problem of information isolation based on the layer of data schema, promote the value of organization

information resource while we must keep the system running, and offer the know service, it's a very challenge also a chance for the research and application of metadata. In this thesis, we introduced the theory of metadata schema registry form the library science, and explored the application model for the integration of heterogeneous information

resource oriented knowledge service. We can use the technology of Uml, Xml, Ontology to construct the application domain based on this model, this show the valuable future of this research.

References:

[1]Heery R. Draft DCMI open Metadata Registry FunctionalRequirements, *Available from <http://dublin core.org/groups/registry/fun_req_ph1.shtml>*,(Last visited: May.2007)

[2]Heery R. Wagner H.A Metadata Registry for the Semantic Web. *DLib Magazine*, 2002, 8 5 : 15-17

[3] Dongwon Jeong *, Doo-Kwon Baik, Incremental data integration basedon hierarchical metadata registrywith data visibility, *Information Sciences* 162 (2004) 147– 181

[4] Rachel Heery, Pete Johnston, The MEG Registry and SCART: Complementary Tools for Creation, Discovery and Re-use of Metadata Schemas, *Proc. Int. Conf. on Dublin Core and Metadata for e-Communities* 2002: 125-132,Firenze University Press

[5]ROADS Metadata Registry, *Available from <http://www.ukoln.ac.uk/roads/templates/>*,(last visited at Apr.2007)

[6]DESIRE Metadata Registry, *Available from . <http://desire.ukoln.ac.uk/registry/>,* (last visited at Apr.2007)

[7]Dublin Core Metadata Registry, *Available from . <http://dublincore.org/dcregistry/>*, (last visited at Apr.2007)

[8]DESIRE, Desire Metadata Schema Registry,2000.*Available from: <http://desire.ukoln.ac.uk/registry/>*(Last visited June 2007).

[9]MEG, MEG registry project, 2002. *Available from:<http://www.ukoln.ac.uk/metadata/educati on/reg proj/>* (Last visited: June 2007).

[10]DCMI Registry Working Group, The Dublin Core Metadata Registry. Prototype, 2002. *Availablefrom:<http://wip.dublincore.org:8080/ dcregistry/index.html>* (Last visited: June 2007).

[11]CORES: a forum on shared metadata vocabularies(2003),*Available from:<http://www.coreseu.net/>*(Last visited: June 2007)

[12]Baker, T., Blanchi, C. et al. (2003). Principles ofMetadata Registries. White Paper of DELOS Registries Working Group, *Available from:<http://delosnoe.iei.pi.cnr.it/activities/stan dardizationforum/Registries.pdf>*,(Last visited: June 2007)

[13]Heery, R. and Patel, M. (2000, September). Application Profiles: mixing and matching metadata schemas. Ariadne 25,*Available from:<http://www.ariadne.ac.uk/issue25/app-pr ofiles/>*,(Last visited: May ,2007)

[15]Manola, F., and Miller, E., Editors.(2003). RDFPrimer. *World Wide Web consortium W3C Working Draft,23 January 2003. Availablefrom<http://www.w3.org/TR/2003/WD -rdf-primer-20030123/>*,(Last visited: May ,2007)

[16]Brickley, D. and Guha, R.V., Editors, (2003). RDF Vocabulary Description Language 1.0: RDF Schema.*World Wide Web Consortium W3C Working Draft, 23 January 2003, Availablefrom:<http://www.w3.org/TR/2003/W D-rdf-schema-20030123/http://www.w3.org/200 1/Annotea/>*,(Last visited: May ,2007)

[17]Hewlett-Packard Labs. Jena Semantic Web Toolkit, *Available from:<http://www.hpl.hp.com/semweb/jena.htm >*, (Last visited: May , 2007)

[18]CORES Schema Creation & Registration Workshop SZTAKI, Budapest, 6-7 March 2003 (2003). *Available from:< http://www.cores-eu.net/>*,(Last visited: May , 2007)

[19]Rachel Heery, Registry for Educational Metadata Schemas Final Project Report and Recommendations, *final report*, 7 April 2003

[20]Abdourahamane Baldé1 and Marie-Aude Aufaure, How can metadata contribute to add semantic information to clusters? *The Electronic Journal of Symbolic Data Analysis -* Vol.3, N.1 (2005), ISSN 1723-5081