

The Dublin Core Metadata Registry: Requirements, Implementation, and Experience

Harry Wagner (wagnerh@oclc.org)
and
Stuart Weibel (weibel@oclc.org)

OCLC Research
6565 Frantz Road
Dublin, Ohio 43017 USA

Contents

1. Introduction
2. Architecture of the DCMI Registry
3. Functional Characteristics and Benefits of the DCMI Registry
4. Relationship of the DCMI Registry to ISO/IEC 11179
5. Related Registry Activities
6. DCMI Registry Deployment Experience
7. Barriers to Installation and Adoption
8. Prospects for Further Development
9. Conclusions

Abstract

Metadata registries are an important digital library research area with the promise of satisfying the needs of metadata designers, practitioners, and users. This paper describes the deployment experience involving the Dublin Core Metadata Initiative (DCMI) metadata registry [1] and discusses the opportunities and prospects for metadata registries as part of the evolving Web-based metadata infrastructure. The motivation and architecture of the DCMI registry are discussed. Benefits and beneficiaries are described, as well as barriers to installation and adoption of metadata registry technology. In addition, prospects for further development are discussed.

1. Introduction

The emergence of global networking has made it possible to use the Internet as a global file system managed with a globally distributed ‘operating system’ – the Web. The foundation protocols of the Internet include the Domain Name System, perhaps the world’s most ubiquitous and successful registry system. As Internet and Web protocols have evolved, the need for registries for a variety of purposes has become evident, including the management of metadata terms.

The Dublin Core Metadata Registry is one approach to meeting such needs. Emerging from a common interest in the Dublin Core Metadata Initiative (DCMI) community, the DC-Registry working group [2] has, since December 1999, focused on identifying the functional characteristics for a metadata registry that meets the needs of our community, and providing a forum for open discussion of related issues.

These discussions have benefited from the perspectives of many researchers and practitioners in the DCMI community. The development of the software itself has been largely concentrated in OCLC Research, with significant input from UKOLN (the UK Office for Network Learning) at the University of Bath, and the Library and Information Science Department of the University of Tsukuba . As an Open Software initiative, OCLC makes the software available under the Dublin Core Public License Version 1.0 [3].

The Dublin Core Metadata Registry is designed and deployed as a multilingual registry of metadata terms. It is not configured as a schema registry, though as a general purpose application designed around Resource Description Framework (RDF) [4] entities, it could be used to manage RDF schemas. Other projects, for example IEMSR at UKOLN (Johnston, 2005, [5]) take this approach, and there are commercial applications designed for schema management as well (SchemaLogics, for example [6]).

The registry is intended to serve as a discovery mechanism and resolution service, with the goal of promoting the reuse of existing terminologies represented in multiple languages.

The registry is packaged with the Dublin Core metadata vocabulary, but is not limited to this, or any given vocabulary. Additional metadata terms can be included at the discretion of the implementer. The only limitation regarding what metadata can be imported into the registry is that it must be represented within the registry in RDF because of the reliance on RDF for internal registry data encoding.

2. Architecture of the DCMI Registry

The DCMI metadata registry is a distributed application in the sense that it is a collection of registries loosely federated with other metadata registries running at other locations. Each of the DCMI registries shares a common data format, user interface, and inter-registry communication features that enable them to cooperate and share metadata declarations in a loose federation.

The DCMI metadata registries are being deployed to serve the needs of communities of practice that differ according to domain, policy constraints, or natural language. The current DCMI registry design dates back to the publication of “*Plans for a distributed registry of Dublin Core in multiple languages*” (Baker, 1998). The motivation for distributing the registry is simply to support the needs of communities of practice in tailoring their metadata implementations to meet local needs and constraints while

sustaining high degrees of compliance with an international standard. Satisfying this need is a principal goal of the DCMI metadata registry.

DCMI metadata terms are intended to be broadly applicable across domains. Extensibility was among the original design criteria, supporting the need to define additional, more specific terms when necessary. The design of the registry reflects this extensibility requirement. The Dublin Core community classifies terms as elements, element-refinements, encoding schemes and controlled vocabulary terms. Dublin Core practitioners often conceptualize DCMI terms in these logical groupings. The registry supports this by providing pre-defined views of the data that match these conceptualizations.

DCMI registry implementations may differ from one another in a number of ways, depending on the communities they serve, the languages they support, and the number of local extensions they implement. However, from architectural and organizational perspectives they share important similarities. First, each registry runs the same base software and supports the same interfaces to its content. Additionally, each of the communities these registries serve is comprised of:

- Read-only users. These include both the humans and applications that are the primary consumers of the registry content. Additionally, the read-only users provide feedback to the registration authority regarding change requests to the registry content.
- A registration authority, responsible for approving registry content. For example, the DCMI Usage Board is the registration authority for the registry available at the Dublin Core Web site [1]. The Usage Board evaluates proposed new terms that are suggested by the larger Dublin Core community. Approved terms are then passed to the registry steward for inclusion in the metadata registry.
- Stewards are responsible for application support and maintenance. Their role is limited to the development, support and ongoing maintenance of the registry software. They rely on the registration authority for decisions regarding the actual registry content.

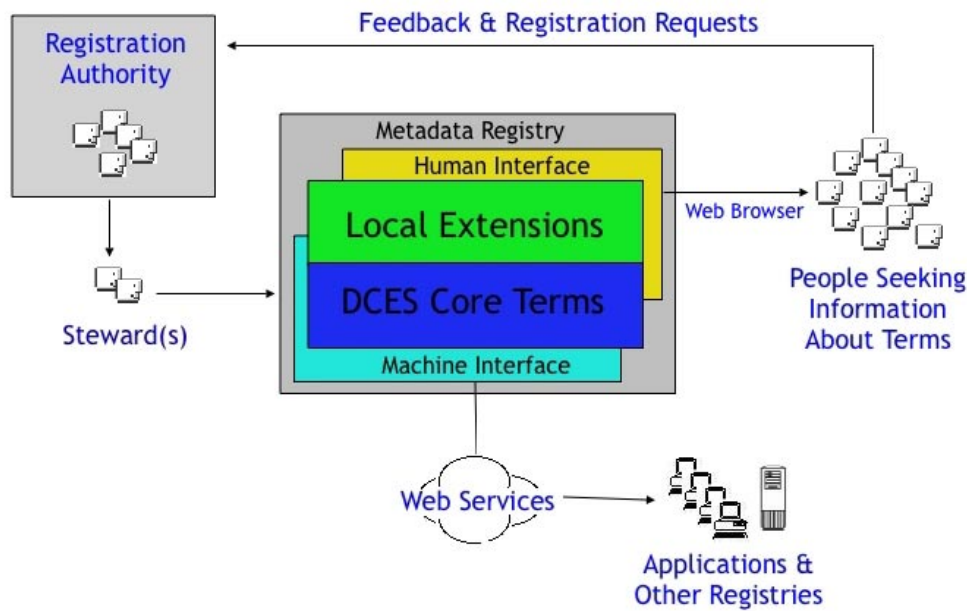


Figure 1: DCMI registry architecture overview

From a technical perspective, the registry is a server-side Java application. As such, it can run on any platform that supports Java. It is built entirely on open-standards and is distributed as open-source. A number of databases are supported, including PostgreSQL, MySQL, and Oracle [7].

The registry relies on open-source distributions and public standards for managing registry contents, including:

- Extensible Markup Language (XML) for parsing, serializing and exchanging registry content. All registry output is in XML format. Xerces [8] is used for processing XML.
- RDF and Resource Description Framework Schema (RDFS) [9] for defining metadata terms and their associated properties. Registry content is represented internally as RDF statements. Jena [10], an open source Java framework for Semantic Web applications, is used for reading, writing and parsing RDF and ontology languages.
- Extensible Stylesheet Language Transformations (XSLT) [11] for presenting search and browse results to users in a human-friendly (Web browser) format.
- The design of the registry is strongly oriented towards Web services, providing application access to registry content via both Simple Object Access Protocol (SOAP) and Representational State Transfer (REST) [12]. Axis, an Apache open-source project for creating SOAP-style Web services, is used for processing the

SOAP-style application interface[13]. The REST-style application interface simply relies on HTTP Post and Get.

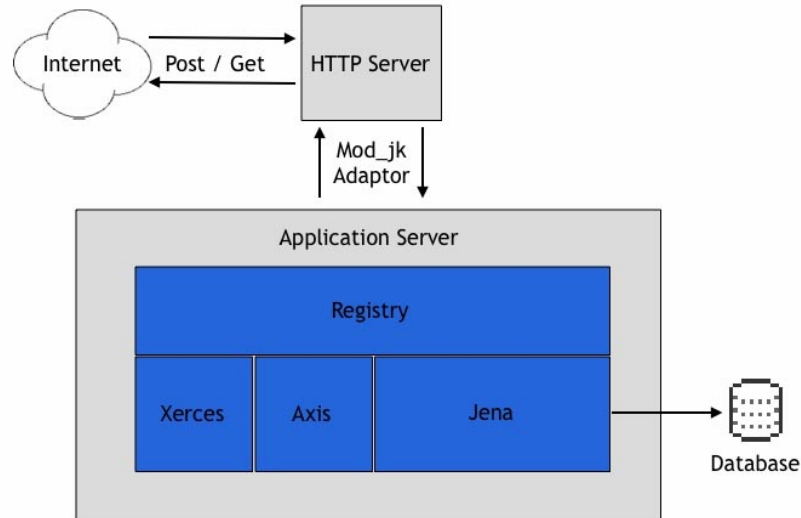


Figure 2: DCMI registry software components

Usability assessment has been limited to discussions within the Registry Working Group. These discussions were conducted on the (publicly available) working group mailing list and focused primarily on functional requirements and technical solutions. Various prototypes were developed, based on these discussions, to facilitate this effort (Heery and Wagner, 2002).

The registry currently supports 25 languages, including translations of both the user interface labels and the Dublin Core terms. The translations of Dublin Core terms have been done largely on a volunteer basis. These translations range in authority from those sponsored by national libraries or other national information agencies to the volunteer efforts of single individuals [14]. Much of this work has been contributed by active participants in the DC-Registry working group and the DC-Localization and Internationalization Working Group [15]. The translations, being voluntary, vary with regard to completeness and accuracy, and continue to evolve as new translations become available. Provenance information regarding the source of term definitions is provided within the registry for all terms. The English language rendition comes from the formal standard (ISO 15836:2003) [16].

The DCMI Abstract Model is “a reference model against which particular DC encoding guidelines can be compared” (Powell, et al., 2005). The Abstract Model was advanced in 2003, and became a formal DCMI Recommendation in March of 2005. The Abstract Model articulates the architecture of DC metadata and the DCMI Registry is intended to be the authoritative source of information about DC metadata terms. As such, these two

entities must be aligned. While design and implementation of the DCMI Registry predates the abstract model, no conflicts between them have been identified.

3. Functional Characteristics and Benefits of the DCMI Registry

The functionality of metadata registries is a subject of active research and can be expected to evolve over time. At present, registries are primarily resolution services, resolving identifiers to information about resources (either physical or conceptual). In the case of metadata registries, term identifiers are resolved to information about terms. It is useful to compare the benefits of such a system to more direct means of declaring element sets, such as schemas and static pages (either electronic or paper).

Schemas are static representations of terms, their properties, and their relations to one another. They are meant for machine interpretation, and are commonly written in either XML or RDF schema languages. They require an understanding of modeling techniques as well as experience with schema languages to maintain and interpret.

Static pages (HTML or PDF for example) require no specialized infrastructure or systems development. They are simpler to update and display and are easier to use for many simple look-up tasks.

Metadata registries must add value that substantially exceeds that provided by static representations (schemas or static pages) in order to be widely deployed and used. Metadata registries provide people and applications with an authoritative source of in-depth information about terms and their relationships to other terms within a given set of vocabularies. They are intended to serve the needs of more than one class of users, including:

- organizations that maintain large, and/or multilingual metadata vocabularies,
- metadata applications, communicating with the registry via Web services,
- schema and ontology designers, and
- creators of metadata instance data.

Metadata registries meet these needs in a number of ways:

Discovery: Discovery is a prerequisite to reuse. Designers of metadata applications need tools that enable them to easily discover terms that are already in use and that can simply be adopted. Besides reducing costs, high levels of re-use are an important ingredient for standardization and interoperability. While a static web page or schema can provide this information, the user must first find the document, and then determine whether or not it is the latest version (particularly difficult if the document is paper-based). Metadata registries, deployed within recognized communities of practice, provide known sources for up-to-date information about terms. This information is available in formats suitable for humans and applications, and can be provided in a number of natural languages.

Authority: Registries can be managed according to formal policies determined by responsible organizations, including access constraints and formal measures of term

currency or status. The existence and management of such policies support two key attributes necessary for the success of the Semantic Web; trust and provenance. The “Web of Trust” (Berners-Lee, 1998) recognizes that authority of information is an increasingly important issue. Metadata registries provide a useful tool for assuring unambiguous, up-to-date information about metadata terms, their provenance, and available translations.

Dynamic representation of information: Registries can offer customized representations of a given metadata set to satisfy different classes of users:

- Multiple languages – A registry can provide term definitions in any number of supported natural languages. This is particularly important in environments where metadata designers, creators, and users need to manage metadata in multiple languages.
- Multiple encoding – The expression of metadata is bound to a chosen encoding, and while in an ideal world there would be only one, the reality is that there are multiple encoding formats, including, HTML (and variants), XML, RDF and Notation 3 (N3) (Berners-Lee, 2001). The DCMI registry provides quick-access links to a variety of encoding formats for terms, including usage examples.
- Customized conceptual views – Organizing metadata terms into collections serves two useful purposes; it facilitates term management, and enables communities to view their data in ways that are more meaningful to them. Metadata registries include information about definitions and syntax of metadata terms but they must also be flexible enough to provide logical views of that information that promote discovery and comprehension.

Information suitable for both humans & machines: Documents formatted for humans, while being easier for people to use in some cases, cannot be easily processed by applications. The ability of a registry to provide applications with information about terms *via* Web services opens up opportunities to increase the effectiveness of metadata systems through automation.

For example, consider a hypothetical application that harvests metadata related to educational materials, and processes this information based on intended audience. An application such as this might be aware of specific terms, such as ‘*audience*’, and use these terms to aggregate instance data based on the value of this term. What happens though when ‘*audience*’ is further refined, as happened when the element-refinements ‘*educationLevel*’ and ‘*mediator*’ were later added? Such an application would very likely also be interested in these new terms. How would it discover they exist, and that they were related to ‘*audience*’? Several alternatives are possible:

- An application can be hard-coded, or parameter-driven, for the terms it uses. Such applications would require someone to discover the new terms, recognize their relationship to terms currently being used, and then modify the application to use the new terms. This scenario is prone to failure, and falls short of our expectations for the Semantic Web.
- An application can be schema-aware, and capable of parsing schemas that are likely candidates for new terms. However, this level of application sophistication

- requires additional resources and technical expertise to develop and support. Second, and more importantly, schemas are commonly versioned. It is fairly common practice to add new terms to a new version of a schema, rather than changing an existing schema. This is intended to reduce the potential for breaking existing applications, but also leaves applications unaware of new term additions.
- An application could use a Web services interface such as are provided by the DCMI registry to identify new terms and their relationship to other terms. This is more than sufficient for the hypothetical application used in this example to recognize that new terms were added, and to understand their relationship to existing terms. Such ‘registry-aware’ applications won’t require code changes or human intervention to evolve with the terminology they use. One example is the DCMI translation tool [17], which uses the registry’s application interface to collect and present information about the natural language properties (label, definition and description) associated with each DCMI term.

Illustration of complex formal relationships among terms: Most systems embody complicated metadata where there are relationships among terms. This is true for any DC metadata that has qualifiers (schemes for the constraint of term values, or element refinements that narrow the meaning or scope of a term). As the influence of ontology research increases, these relationships will become more complex and will require tools that enable practitioners to easily visualize the nature of these relationships. Registries are better suited to this task than static pages. They can display data in ways that promote ease of comprehension. For example, consider the relatively simple relationships enabled with RDF Schema (RDFS), which includes a ‘*subPropertyOf*’ assertion. Using RDFS, metadata practitioners can make explicit that ‘*abstract*’ and ‘*tableOfContents*’ are refinements of ‘*description*’. However, the opposite is not true. RDFS does not provide terminology that enables practitioners to explicitly assert that ‘*description*’ is further refined by ‘*abstract*’ and ‘*tableOfContents*’. This must be inferred. Such relationships can be made more explicit in registries. This will become a prominent benefit of registries as ontology languages (such as the Ontology Web Language, or OWL [18]) and reliance on logical inference using ontologies becomes more widespread.

4. Relationship of the DCMI Registry to ISO/IEC 11179

ISO/IEC 11179 [19] is a standard that describes a model for metadata registries. It is comprised of multiple parts, with varying levels of conformance. These include framework, classification, metamodel (a basic set of properties for describing registry content), data definition guidelines, naming and identification of data elements and content registration. The goal of these specifications is to provide a common framework for registries that will insure increased likelihood that:

- Registered items can be uniquely identified.
- Properties for describing registry content are unambiguous.
- Registry content can be unambiguously mapped between the metamodel and the implementation.

The overriding goals of the 11179 specification are to promote the understanding and sharing of metadata terms and definitions. This is also true for the DCMI metadata

registry. However, while the DCMI registry shares a similar mission with the 11179 standard, it differs somewhat in its approach, specifically with regards to technology. Typically, 11179-compliant implementations have been XML-centric in their approach. XML is pervasive, and widely successful as a data exchange encoding. However, XML does not lend itself well to two key aspects of 11179 compliance:

- 1) content properties that are unambiguous, and
- 2) insuring the uniqueness of identifiers. Part 3 (Registry Metamodel and Basic Attributes), part 4 (Formulation of Data Definitions) and part 5 (Naming and Identification Principles) address these issues and provide guidance for achieving compliance using an XML-centric approach.

The DCMI registry also relies on XML as a means of data exchange, but uses additional technologies (RDF and XML Namespaces) that are layered on top of XML. The issue of insuring unique identifiers for each registered item is addressed by using fully qualified term names for identifiers. Term names, such as:

`http://purl.org/dc/elements/1.1/creator`

consist of a namespace:

`http://purl.org/dc/elements/1.1/`

and a local term name:

`creator`

Properties described using RDF have three key advantages over simple XML tags:

- 1) they take advantage of XML namespaces,
- 2) they have attributes which describe their essential nature, and
- 3) these attributes can be processed by applications.

Together, these features insure that these properties are unambiguous and that their semantic meaning is clearer. For example, consider a geographic area described with the `extent` property. This property is identified with a URL that resolves to a series of attributes describing it and its relationship to other properties:

`http://purl.org/dc/terms/extent`

RDF-aware applications can process this information and identify a number of assertions about this property. The attributes describing the term in the schema are publicly declared assertions. Specifically:

The English rendition of the label is:	<i>Extent</i>
The English rendition of the description is:	<i>The size or duration of the resource.</i>
The source for this term definition is:	http://purl.org/dc/terms/
It has an RDF Type of:	<i>Property</i>
It has a dc term type of:	<i>element-refinement</i>
It has an associated version:	<i>extent-002</i>
It has a date issued of:	<i>2000-07-11</i>
It was last modified on:	<i>2002-06-15</i>
It is related to and refines:	<i>format</i>

However, without the explicit declaration of attributes as is provided with RDF, applications must rely on a pre-arranged understanding of, or assumptions about, the meaning and nature of simple XML tags, such as '*<extent>*' or '*<geographicExtent>*'.

The ISO 11179 community is currently exploring ways to expand the specification to include greater utilization of XML namespaces and to provide additional support for the semantic understanding of registered items and their properties. Part of this effort is being undertaken by the Extended Metadata Registry Project (XMDR) [20]. One of the key issues being considered is how to incorporate extensions to the specification to include support for technologies, such as RDF and OWL, which promote unambiguous semantic specification. This increased focus on declarations of semantics is reflected in the 2005 Open Forum on Metadata Registries conference [21], an annual international conference, the theme of which, was "Semantic Interoperability: Where Meaning Meets Metadata"

The DC-Registry working group recognizes the importance of a standard approach for metadata registries. They have closely followed the work being done on the ISO/IEC 11179 specification, and where possible, have sought opportunities for mutual cooperation. It is expected that the extensions being considered by the XMDR project will help bring the two communities closer together.

5. Related Registry Activities

In an era of increasing importance of machine-to-machine transactions, registries will be important for exchanging well-structured information and schemas. Kotok (2003) describes several examples, including Universal Description, Discovery and Integration (UDDI) and electronic business using eXtensible Markup Language (eXML), registries that promote discovery of Web services and product information [22].

Metadata registries, in turn, provide resolution services for terms or abstract concepts, resolving identifiers to information about metadata terms used to describe products and services or intellectual assets. Their goal is not to provide access to products or services, but rather to promote understanding of metadata terms used to describe products and services. The DCMI Metadata Registry falls into this category.

Application profiles are an important part of the changing landscape of metadata standards, affording a necessary mechanism for combining terms from several schemas into a composite schema that is tailored to local functional requirements (though an application profile may be comprised of terms from a single schema as well). An application profile is generally rendered in the form of a compound schema (as described in Heery and Patel, 2000). Operationally, application profiles require both documentation and standards of community practice to become useful and effective. The development of application profiles and their ancillary support is still as much a topic of research as practice, but as they proliferate, they will be prominent entities in metadata registries.

The variety of metadata registries is substantial, reflecting the research interests and operational needs of the many organizations that have undertaken them. They differ according to the underlying technology strategies, desired functionality, and accessibility by humans, applications, or both. Prominent examples of metadata registry applications span a variety of domains.

The Environmental Data Registry provides information used to describe environmental data used by local, state and U.S. government agencies. It is intended to serve as an authoritative and comprehensive source of information about terminology used by these agencies. This registry is part of a larger System of Registries (SOR) implemented by the U.S. Environmental Protection Agency [23].

The U.S. Department of Defense Metadata Registry and Clearinghouse [24] is intended to provide a common source of information about metadata terms and related technologies that are used within the defense industry. The registry is designed to promote interoperability and reuse of metadata and related software within authorized agencies. Details of the contents of this registry are no longer accessible to the general public.

The METeOR registry serves the Australian Institute of Health and Welfare (as described by Braddock, 2005) This registry is a replacement for the Knowledgebase Registry Like its predecessor, it registers metadata related to Australian social services, specifically: health, community services and housing assistance. It includes browse and search interfaces and is based on the ISO/IEC 11179 specification [25].

The European Library (TEL) Registry [26] is a collection of terms and properties designated for use in TEL application profiles (van Veen and Oldroyd, 2004). TEL is a cooperative effort involving 8 European national libraries and the Italian Central Cataloging Institution (ICCU). Their goal is to provide an advanced resource discovery service for researchers that goes beyond enumeration of available terms, and includes information about application profiles, schemas, and support for data entry forms and proposing of new elements.

The CETIS registry, developed and maintained by the Centre for Educational Technology Interoperability Standards, is intended to “*help people in the UK HE (higher education) & FE (further education) community to find out about terms used in the field of learning technology standards*” [27]. This registry serves as a reference tool, providing browse and search access to a variety of definitions related to technology. Definitions are provided collaboratively by members of the CETIS special interest groups (SIGs).

The German Metadata Registry provides “*an overview of the metadata efforts and implementations within Germany and German-language areas*” [28]. This registry includes term definitions, documentation, links and other material related to a number of subject specific domains, including education, medicine, physics and mathematics. Both German and English languages are supported.

The Development of a European Service for Information on Research and Education (DESIRE) Metadata Registry [29] was among the first registry metadata projects. This registry was developed by UKOLN and is intended to serve as a discovery and navigation tool for a variety of metadata resources, including namespaces, registration authorities, application profiles and cross-vocabulary mappings between terms. It relies on XML as the underlying technology and is the basis for a number of evolving registry activities described below. While each of these projects continue to have a web presence and a functional registry, not all are currently active.

- The SCHEMAS Registry [30] includes terms and metadata activity reports related to projects funded by the IST Programme and other European national initiatives. It is implemented as a human-readable registry designed primarily to inform metadata designers.
- The CORES registry (Heery, et al., 2003), is a registry of metadata vocabularies that focuses on the sharing and reuse of metadata terms, and the creation of application profiles [31]. It builds on work done in the SCHEMAS project. Among the practical outcomes of this activity is a resolution among several metadata organizations to identify their metadata terms with Uniform Resource Identifiers (Baker & Dekkers, 2003).
- The Metadata for Education Group (MEG) Registry [32] is both an extension, and a rewrite, of the DESIRE registry. Unlike DESIRE, the MEG registry is RDF based. It is intended to serve the needs of the MEG group by providing a known source of information related to education-specific semantics, application profiles and metadata specifications. The registry also includes a Java client designed to facilitate application profile creation (Heery *et al.*, 2002).
- The Information Environment Metadata Schema Registry (IEMSR) [33] is a recent registry project that builds on the MEG registry. It is expected to include both the Dublin Core and IEEE LOM [23] metadata vocabularies and serve as a known source of information for education-related application profiles (Johnston, 2005).

6. DCMI Registry Deployment Experience

At present, five DCMI metadata registries have been deployed, at the following locations:

- OCLC Online Computer Library Center, Dublin, Ohio, USA
- The University of Tsukuba, Japan
- The University of Goettingen, Germany
- The Library of the Chinese Academy of Science, (Beijing) China
- The National Library of New Zealand, New Zealand

The registry deployed at the Library of the Chinese Academy of Science in Beijing is a recent addition. It is expected to play a role in a project to link all of the Chinese digital libraries. Each of the remaining registries is currently being used primarily for research.

The registry application statistics offer a valuable source of insight regarding registry functionality, and its usefulness. Table 1 provides a summary of pages served by the Web (human) interface of the Dublin, Ohio registry. These statistics cover a period of

activity spanning November 1, 2004 through April 30, 2005. A total of 137,581 pages were accessed during this period. Registry users do not identify themselves and hence cannot be questioned about their reasons for using the registry, however logs reveal certain usage patterns. For instance:

- The registry provides two Web interfaces, one for browsing the registry content, and one for text searching. The browse feature was used over 10,000 times, compared with only 272 search requests, indicating a clear preference for browsing over searching.
- There were 31,930 requests for term-level detail. This includes term-level information in all of the supported languages.
- There were 47,613 accesses for alternate encoding views (term information encoded in RDF, N3, or N-triple). These data encodings are accessible as links from the *term detail* views formatted for human reading, but can also be accessed directly via URL. The large number of page views for these encodings is a good indication of the interest that registry users have in Semantic Web technologies.
- There were 1,262 requests to set or change language preferences. The default language for this registry is English. The large number of calls to change language preferences is indicative of a significant number of international users whose first language choice is not English.
- The term detail view provides quick links to usage examples (7,669). The large number of access requests indicates the importance of this feature.

Request Type	Page Count	Percentage
Browse registry content	10,415	7.5%
Search registry content	272	0.1%
View item detail	31,930	23.2%
View alternate encoding (RDF, N3, etc.)	47,613	34.6%
Set or change language preferences	1,262	0.9%
View usage examples for a term	7,669	5.5%
Canonical view of term	4,444	3.2%
Other (i.e., Provenance information)	33,976	24.6%

Table 1: Page summary statistics: DCMI Registry 2004.11.01 – 2005.04.30

Registries are expected to serve applications as well as humans. The DCMI Registry supports two application interfaces, both based on Web services. One is SOAP-based and the other is REST-based. The registry statistics provide insight into these two interfaces and how they are being used. The overwhelming majority of application use comes from the REST-style services. During the period from November 1, 2004 through April 30, 2005 there were 4,841 calls to the REST services, versus 1,727 SOAP-style calls, possibly reflecting the greater ease of implementation of REST services.

The application interfaces provided are designed to satisfy a number of different information request types. This was considered an important functional requirement, and according to an informal survey recently conducted within the DC-Registry Working

Group, is still perceived as one of the most important features of metadata registries. Based on this information one might expect the usage statistics to show patterns indicative of use by applications. The statistics show interest in the application interfaces, but the use-patterns currently reflect experimentation, rather than actual use.

7. Barriers to Installation and Adoption

Organizations adopt technology because it promises to add value: to improve delivery of goods and services or reduce the cost of providing them. Identifying this value at the present time is a largely qualitative assessment. Partly this is due to the fact that the metadata milieu is evolving. The value of metadata registries will become more evident as the difficulties associated with large deployments of heterogeneous metadata systems emerge. There is growing evidence of the problems of uncoordinated metadata assignment and design in systems which attempt to integrate metadata from many sources. For example, studies of the NSF-funded National Science Digital Library program have identified inconsistent metadata as a major impediment to effectiveness (Dushay and Hillmann, 2003). Metadata registries can help address such deficiencies.

Organizational commitment is another important factor that has to be considered. Deploying and sustaining software systems entails substantial costs; organizations must make a significant and persistent commitment to deploy them. There must therefore be a strong business case to support their adoption.

Deploying a metadata registry requires a thorough understanding of a variety of factors:

- What is the scope of the registry?
- Who is it intended to serve?
- Is there an organization to establish and administer registry policies?
- Are there technical resources available to manage the day-to-day technical aspects of running the application?
- How should the data be organized to best promote its comprehension? Metadata should be organized so that it can be viewed in ways meaningful to the registry users.

The degree of technical expertise required to install, customize, and maintain a registry, and the amount of planning required to make it productive, can be substantial.

8. Prospects for Further Development

Development of metadata registry infrastructure remains at an early stage, meeting primarily administrative needs, and with human users as the primary consumer. Today, metadata registries are essentially a resolution service, resolving term identifiers to information about terms. However, one can imagine a metadata environment that benefits from a wider variety of services and a greater degree of application-to-application communication. Possible scenarios include:

- Automated crosswalks and mappings. The Dublin Core metadata registry is already capable of registering metadata terms that can be encoded in RDF, including terms from other metadata standards. However, simply including terms from more than one metadata standard within the same registry does not solve the interoperability problems that exist between standards. A means to automate the mapping of terms between standards is needed and is a good candidate for future development.
- Querying and use of ‘arbitrary’ schemas (application profiles) by applications. Application profiles, and how to generate them in a machine-readable format, is an ongoing research activity. When a standard for generating machine-readable application profiles emerges and is adopted it is conceivable that the Dublin Core registry will include support for application profiles.
- Automated invocation of authority management, harvesting opportunities, and metadata triggers. The cost of creating instance metadata is a large impediment to the effectiveness of metadata systems and resource discovery and management. Efforts to reduce that cost must rely largely on industrializing the creation of metadata through more effective tools and automation. As such techniques emerge, they will benefit from gathering authority-controlled data from a variety of sources. This will only become possible when applications can discover both the content and structure of such data without human intervention, and metadata registries will likely be an important part of this infrastructure.
- Automated harvesting of metadata terms. Registering terms in the current DCMI registry involves importing terms using one of the import tools provided with the software. Terms must be identified and then selected for importing. In the future it will be desirable to automate harvesting of metadata repositories, extracting and automatically registering relevant terms.

9. Conclusions

The DCMI Registry has evolved over a period of years to meet the needs of the DCMI community, which have ranged from authoritative management of the DCMI vocabulary, to public browse and search capability, to application interfaces. Feedback and participation from the community have helped steer the design and implementation, and the architecture and functionality have evolved to a model very much in keeping with the distributed idiom that currently dominates Internet developments. Distribution of management and function incurs some management costs, but confers technological resilience as well as meeting important policy requirements that are essential components of global information systems.

Nonetheless, metadata registries have not yet become an integral part of the metadata infrastructure of the Web. We expect this integration will increase as multi-lingual or cross-domain metadata applications are deployed on a large scale, and the benefits for automated management support becomes more evident.

References

- Baker, Thomas (1998). "Plans for a distributed registry of Dublin Core in multiple languages." *Dublin Core Working Draft*, 1998-10-28.
<http://dublincore.org/documents/1998/10/28/distributed-registry/>.
- Baker, Thomas and Dekkers, Makx (2003),
"Identifying Metadata Elements with URIs: The CORES Resolution"
D-Lib Magazine, Volume 9 Number 7/8.
<http://www.dlib.org/dlib/july03/baker/07baker.html>.
- Baker, Thomas; Dekkers, Makx; Heery, Rachel; Patel, Manjula, and Salokhe, Gauri (2001) "What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives". *Journal of Digital Information*, Volume 2 Issue 2
Article No. 65, 2001-11-06
<http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/>
- Berners-Lee, Tim (2001). "Notation 3: An RDF language for the Semantic Web" 2001-11-27.
<http://www.w3.org/DesignIssues/Notation3.html>
- Berners-Lee, Tim (1998). "Realising the Full Potential of the Web." W3C Website.
<http://www.w3.org/1998/02/Potential.html>.
- Braddock, David (2005). "METeOR: a metadata registry implementation experience"
Open Forum 2005 on Metadata Registries.
http://www.berlinopenforum.de/download/Braddock_David.zip
- Dushay, Naomi. & Hillmann, Diane (2003). "Analyzing Metadata for Effective Use and Re-Use." *DC-2003 Dublin Core Conference: Supporting Communities of Discourse and Practice – Metadata Research and Applications. September 28 -October 3, 2003*.
http://www.siderean.com/dc2003/501_Paper24.pdf
- Heery, Rachel, Johnston, Pete, Beckett, Dave, and Steer, Damien (2002). "The MEG Registry and SCART: Complementary Tools for Creation, Discovery and Re-use of Metadata Schemas" *DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence*. Florence, Italy. October, 13-17, 2002.
<http://www.bncf.net/dc2002/program/ft/paper14.pdf>
- Heery, Rachel; Johnston, Pete; Fülöp, Csaba and Micsik, Andras (2003). "Metadata schema registries in the partially Semantic Web : The CORES experience." *DC-2003 Dublin Core Conference: Supporting Communities of Discourse and Practice – Metadata Research and Applications. September 28 -October 3, 2003*.
http://www.siderean.com/dc2003/102_Paper29.pdf

Heery, Rachel and Patel, Manjula (2000). "Application profiles: mixing and matching metadata schemas" *Ariadne*, Issue 25, 24-Sep-2000.
<http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>.

Heery, Rachel and Wagner, Harry (2002). "A Metadata Registry for the Semantic Web" *DLib Magazine*. Volume 8, #5 May, 2002.
<http://www.dlib.org/dlib/may02/wagner/05wagner.html>.

Johnston, P. (2005). "What Are Your Terms?" *Ariadne*, Issue 43, April 2005.
<http://www.ariadne.ac.uk/issue43/johnston/>.

Kotok, Alan (2003). "Metadata Rules' - a report from the Open Forum on Metadata Registries." *WebServices.org* 2003-02-24.
<http://www.webservices.org/index.php/ws/content/view/full/2873>.

Powell, Andy, Mikael Nilsson, Ambjörn Naeve, Pete Johnston (2005). "DCMI Abstract Model." *DCMI Recommendation*. 2005-03-07.
<http://dublincore.org/documents/abstract-model/>.

van Veen, Theo and Oldroyd, Bill (2004).
"Search and Retrieval in The European Library: A New Approach"
D-Lib Magazine, Volume 10 Number 2.
<http://www.dlib.org/dlib/february04/vanveen/02vanveen.html>.

NOTES

[1] The Dublin Core Metadata Registry is available at:
<http://dublincore.org/dcregistry/>.

[2] The DC-Registry working group page is available at:
<http://dublincore.org/groups/registry/>.

[3] DCMI public license is available at:
<http://dublincore.org/dcpl/>.

[4] The Resource Description Framework (RDF) is a family of W3C specifications defining encoding standards to support Semantic Web applications.
<http://www.w3.org/RDF/>.

[5] The JISC IEMSR project is available at:
<http://www.ukoln.ac.uk/projects/iemsr/>.

[6] SchemaLogic Corporate homepage:
<http://www.schemalogic.com/page>.

[7] PostgreSQL is available at: <http://www.postgresql.org/>.

MySQL is available at: <http://www.mysql.com/>.

Oracle is available at: <http://www.oracle.com/index.html>.

[8] Xerces is available at: <http://xml.apache.org/xerces2-j/>.

[9] RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004.

<http://www.w3.org/TR/rdf-schema/>.

[10] Jena is an open source project originating in the HP Labs Semantic Web Programme, available at: <http://jena.sourceforge.net/>.

[11] XSL Transformations (XSLT). Version 1.0. W3C Recommendation 16 November 1999. <http://www.w3.org/TR/xslt>.

[12] SOAP Version 1.2. W3C Recommendation 24 June 2003.

<http://www.w3.org/TR/soap/>.

The REST protocol is described in chapter 5 of the doctoral dissertation by Roy Fielding: "Representational State Transfer (Rest)", in *Architectural Styles and the design of Network-based Software Architectures*. Univeristy of California, Irvine , 2000.

http://www1.ics.uci.edu/%7Efielding/pubs/dissertation/rest_arch_style.htm.

[13] Axis is an implementation of the SOAP protocol, available at:

<http://ws.apache.org/axis/index.html>.

[14] Sources and acknowledgements for Dublin Core registry term translations are available at:

http://dublincore.org/dcregistry/pageDisplayServlet?page=help_en-US.xsl#H7.

[15] Dublin Core Localization and Internationalization Working Group.

<http://dublincore.org/groups/languages/>.

[16] Information and documentation - The Dublin Core metadata element set (ISO 15836:2003

<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=37629&scopelist=PROGRAMME>.

[17] The DCMI translation tool is available at: <http://wip.dublincore.org/translate/>.

[18] OWL Web Ontology Language. W3C Recommendation. 2004-02-10

<http://www.w3.org/TR/2004/REC-owl-features-20040210/>.

[19] ISO/IEC 11179 standard for metadata registries.

<http://metadata-standards.org/11179/>.

[20] Extended Metadata Registry Project (2005-05-19) is available at: <http://xmdr.org/>.

[21] Eighth International Open Forum on Metadata Registries. April 6-8, 2005.
<http://www.berlinopenforum.de/>.

[22] Universal Description, Discovery and Integration (UDDI) is available at:
<http://www.uddi.org/>.

Electronic Business using eXtensible Markup Language (ebXML) is available at:
http://www.service-architecture.com/web-services/articles/ebxml_registry.html.

[23] Environmental Data Registry is available at <http://www.epa.gov/edr/> and the System of Registries (SOR) is available at: <http://www.epa.gov/sor/>.

[24] The U.S. Department of Defense Metadata Registry and Clearinghouse is available at: <http://diides.ncr.disa.mil/mdregHomePage/mdregHome.portal>.

[25] The METeOR registry is available at:
<http://meteor.aihw.gov.au/content/index.phtml/itemId/181414>.
The Knowledgebase Registry is available at:
<http://www.aihw.gov.au/knowledgebase/index.html>.

[26] The European Library (TEL) Registry is available at:
http://krait.kb.nl/coop/tel/handbook/tel_reg_v1.3.html.

[27] The CETIS registry is available at. <http://www.cetis.ac.uk/encyclopedia>.

[28] German Metadata Registry Project is available at:
<http://www.mpib-berlin.mpg.de/dok/metadata/gmr/gmr1e.htm>.

[29] Development of a European Service for Information on Research and Education (DESIRE) Metadata Registry is available at:
<http://desire.ukoln.ac.uk/registry/index.php3>.

[30] The SCHEMAS Registry is available at:
<http://www.schemas-forum.org/registry/>.

[31] The CORES registry is available at:
<http://www.cores-eu.net/registry/>.

[32] The Metadata for Education Group (MEG) Registry is available at:
<http://www.ukoln.ac.uk/metadata/education/regproj/>.

[33] The Information Environment Metadata Schema Registry (IEMSR) is available at:
<http://www.ukoln.ac.uk/projects/iemsr/>.

[34] IEEE WG12: Learning Object Metadata (LOM) is a working group focused on the standardization of learning object metadata.: <http://ltsc.ieee.org/wg12/>.

Author Details

Harry Wagner is a consulting software engineer in OCLC Research. He has been an active participant in the Dublin Core Metadata Initiative (DCMI) since 2001. He serves, with Rachel Heery, as co-chair of the DCMI Registry Working Group and is a member of the DCMI Advisory Committee. In addition to Metadata registries his research interests include open-source, Web services, and the Semantic Web activity.

Stuart Weibel, Ph.D., is a Senior Research Scientist in OCLC Research, and past Director of the Dublin Core Metadata Initiative. In his two decades in OCLC Research he has directed research in automated cataloging, document structure analysis, and electronic publishing. He has been an active participant in the Internet Engineering Task Force activities involving persistent identifiers, and was part of the team that implemented PURLs (<http:purl.org>).