<body>

Web Content Management Using the OASIS ebXML Registry Standard

Keywords: Registry, Repository, Web Services, Content Management, Semantic Web, Meta data, Distributed Systems, E-business, eCommerce, E-Government

Farrukh Najmi
XML Standards Architect
Sun Microsystems
USA
farrukh.najmi@sun.com
http://ebxmlrr.sourceforge.com

Biography

> Farrukh Najmi is an XML Standards Architect at Sun Microsystems, Inc. His focus in recent years has been on Registries, Repositories, Enterprise Content Management, and eGovernment. He is an active contributor to the ebXML standard and is a principal author and editor of the **[OASIS ebXML Registry Specifications]** . Farrukh is the specification lead for the Java API for XML Registries **[JAXR]** . He is a co-founder of **[freebXML.org]** where he currently leads **[freebXML Registry]** , an open source project that provides a royalty-free implementation of the ebXML Registry and JAXR standards.
>
> Previously, Farrukh helped develop Sun's Java Messaging Service Reference Implementation. Prior to joining Sun, Farrukh was at Hewlett Packard where he led the development of a large-scale enterprise Java based Clinical Information Systems.

Table of Contents

Introduction and Brief History

Under the sponsorship of UN/CEFACT and OASIS, the ebXML initiative was launched in 1999 to develop a modular suite of specifications that enables enterprises of any size and in any geographical location to conduct business over the Internet. Within the ebXML activity, the ebXML Registry working group was originally chartered to define a specification for a registry and repository that would meet the focused requirements of publishing and discovery of artifacts defined by other ebXML specifications in order to facilitate ebXML-based business integration.

Despite the narrow focus of its original charter, the ebXML Registry specification was designed from its inception to manage arbitrary content and standardized yet extensible meta data. With version 3, the ebXML Registry specification has evolved into a standard that defines an interoperable Enterprise Content Management API (ECM) for web services.

Overview of Enterprise Content Management

We begin with a definition for ECM. Enterprise Content Management provides the platform and tools for scalable and secure production and delivery of information assets within and across enterprise boundaries.

An ECM product typically has a wide array of features. Some examples of ECM features for the production of information assets are:

- Managing users, user roles and user groups
- Content authoring and transformation
- Aggregation from diverse info sources
- Content versioning
- Describing content with standard metadata
- Content workflow: edit, update, approve
- Localization to multiple languages

Some examples of ECM features for the delivery of information assets are:

- Web page Assembly
- Site index and search
- Personalization
- Privileges Management
- Caching and Replication
- Syndication and delivery to subscribers
- Vending or selling of online content

When considering ebXML Registry as an ECM standard it is important to maintain a clear distinction between ECM products and the role of ebXML Registry as an ECM standard. While an ECM product may provide many of the ECM features listed above, ebXML Registry standard is only focused on those aspects of ECM that are significant for interoperability between ECM systems. Specifically, the ebXML Registry standard does not compete with ECM products. Instead it provides ECM products with a standard API and information model that enables and ECM product to interoperate with other ECM products that support the ebXML Registry standard.

While there are many applications of ECM, Web Content Management (WCM) stands out as a prevalent ECM application.

Overview of Web Content Management

WCM allows non-technical business people to manage web sites. Specifically it enables geographically distributed people to build and manage web sites collaboratively.

The resulting web site content can be associated together in meaningful ways to reflect natural relationship between content. It can also be cataloged and categorized to facilitate rapid discovery. It can be validated to improve the integrity of content.

Collaboration on the web site content is supported via workflow capabilities such as the ability to manage the life cycle of content from the time it is submitted until the time it is deleted. In the interim period it may undergo several edits and approval cycles.

Collaboration on the web site content requires that all actions are restricted to authorized individuals and their agents.

The task of creating web sites is further made simple by providing a templates driven approach to web content creation.

Introduction to ebXML Registry

Thus far we have described Enterprise Content Management (ECM) and a specific feature of ECM, namely Web Content Management (WCM). We will now introduce the ebXML Registry standard, its features and capabilities and how it can be used for WCM.

What is ebXML Registry

We introduce the ebXML Registry with an informal definition that uses a familiar analogy:

```
ebXML Registry is to the web what relational databases were to
enterprise applications
```

In traditional enterprise applications a relational database provides the means for applications to persists and share data and metadata. Today enterprise applications are evolving into web applications or web services that also have the need to persists and share data and metadata. The data and metadata is often in an XML format and may also consist of a variety of content types such as text, images, sound and video. An ebXML Registry may be used by web services and application to store and share content and metadata.

A more formal definition of ebXML Registry follows:

```
An ebXML registry is an information system that securely manages any
content type and the standardized metadata that describes it.
```

What ebXML Registry is Not

It is common misconception that ebXML Registry is useful only within the larger ebXML architecture to facilitate B2B collaboration as a means to publish and discover ebXML Collaboration Protocol Profiles, Business Process Specifications and Core Components. Within our database analogy this would be tantamount to saying that databases are only useful for keeping Employee records (or some other early database application). A related misconception that can be dispelled similarly is that ebXML Registry is only useful as a web services registry.

Another common misconception is that ebXML Registry is useful to applications only at design time and not at run time. This misconception has its roots in the first misconception since publish and discovery of Collaboration Protocol Profiles, Business Process Specifications and Core Components is primary a design time activity. Within our database analogy this misconception would be tantamount to saying that databases are only useful at application design time.

Finally another misconception is that ebXML Registry is tightly coupled with the rest of the ebXML stack and that using ebXML Registry requires using the rest of the ebXML stack. A related misconception that ebXML Registry is not consistent with a web services framework. In fact ebXML Registry has no required dependencies on any other ebXML Registry specification and it is itself a web service with a SOAP interface that is described within a WSDL description.

Features of ebXML Registry

The ebXML Registry standard is defined by two specifications. The ebXML Registry Information Model (ebRIM) specification defines the standard metadata that can be stored in the registry. The ebXML Registry Services (ebRS) specification defines the API for services provided by the registry.

In essence, ebRIM defines what gets stored in the registry while ebRS defines how it gets stored and managed within the registry.

Figure 1 highlights some of the main features of the ebXML Registry standard.

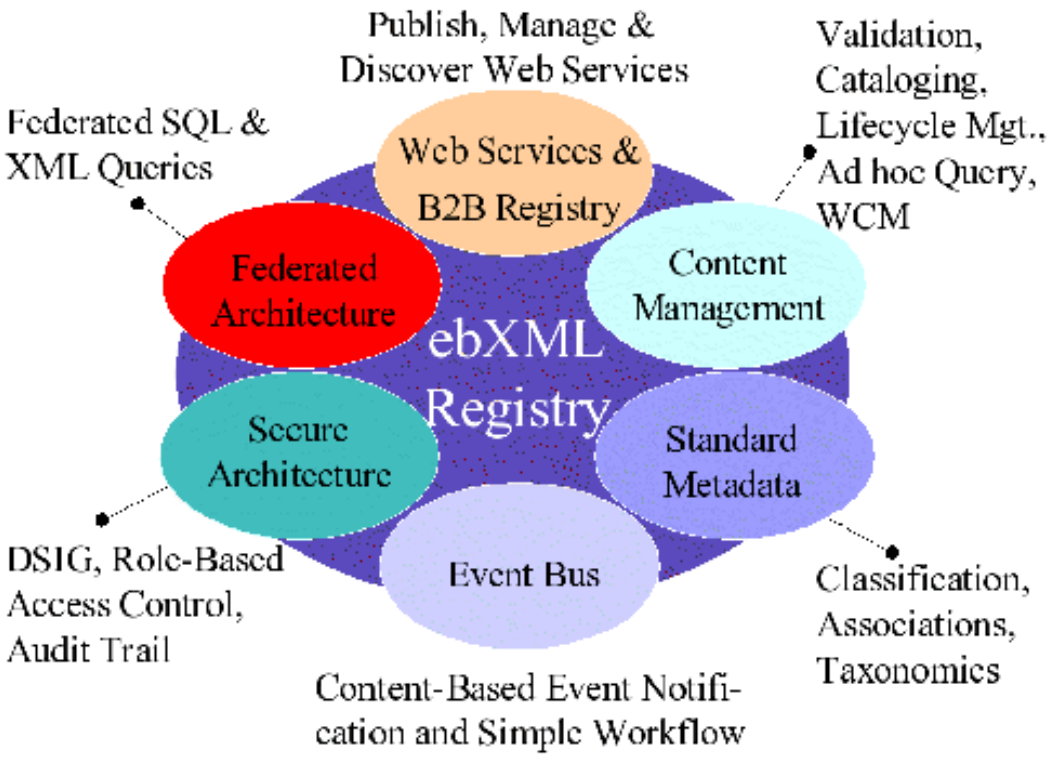

Figure 1: Main Features of ebXML Registry At a Glance

Web Service and B2B Artifacts Registry

The support for publishing, management and discovery of Web Service and artifacts describing B2B collaborations is central to ebXML Registry. Instead of providing special case treatment for these artifacts, the ebXML Registry defines a general purpose repository which is capable of storing any type of content including but not limited to service descriptions in WSDL and ebXML CPP/A. The content in the repository is described or cataloged using metadata that is stored in the registry. The ebRIM does define standard metadata for description of service as shown in **Figure 2** .

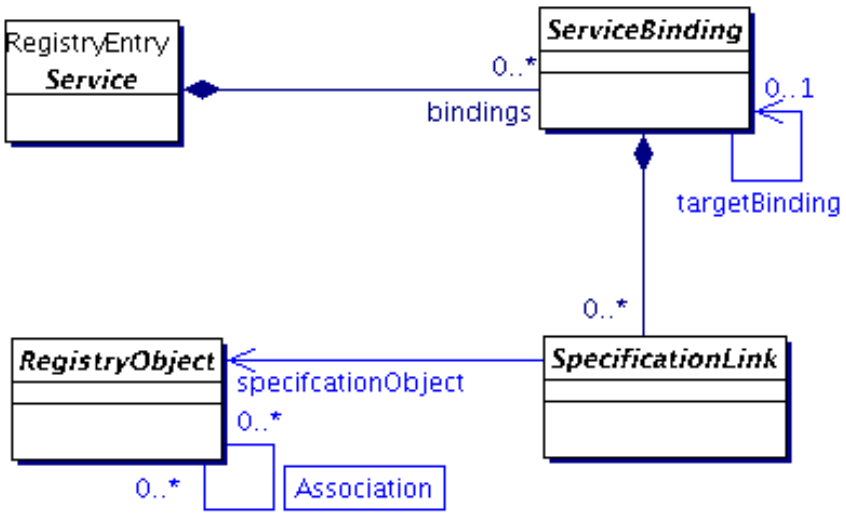

Figure 2: Information Model Classes for Service Description

Content Management

The ebXML Registry provides a variety of content management features to manage the arbitrary content in its repository.

Content Validation

When content is submitted to the registry it may be automatically validated using a content specific validation service that applies business rules to determine whether the content is accepted or rejected. For example, it is possible to enforce that all PortTypes in a WSDL description must have a SOAP binding in order to be accepted.

Content Cataloging

When content is submitted to the registry it may also be automatically cataloged using a content specific cataloging service that converts content into metadata selectively. For example, it is possible to configure a cataloging service for GIF images that automatically categorizes each GIF image as either Blank and White, Greyscale or Color. The content validation and cataloging services operate in a content specific manner even though the registry has no intrinsic knowledge about specific content types. This is possible because the ebXML Registry standard defines a standard SOAP API for cataloging and validation web services. A web services designed specifically to validate or catalog a specific content type may be implemented to support the standard SOAP API and be registered with the ebXML Registry for that content type.

Life Cycle Management

Once content is accepted by the registry it manages all life cycle aspects for it as shown in **Figure 3** . This includes updates and approvals as well as deprecation (marking as obsolete) and eventual

deletion.

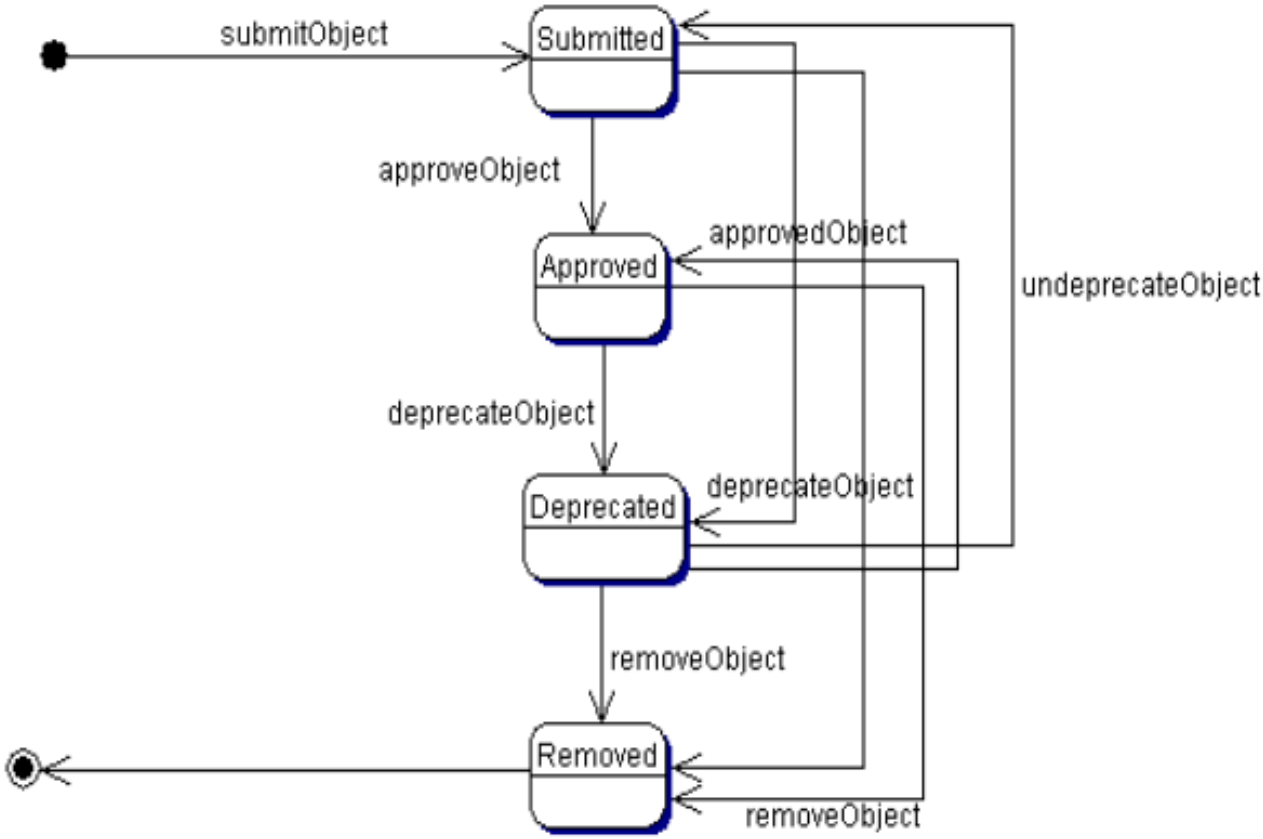

Figure 3: Life Cycle Management of a RegistryObject

Content Discovery

Content discovery is an essential aspect of Content Management. An ebXML Registry may be searched with ad hoc queries using SQL 92 or XML filter query syntax. The ad hoc nature allow the query to be as specific or as general as the situation demands. This has the potential for making queries quite complex. This complexity is easily managed using parameterized registry resident queries where the gory details of the query are not exposed to the client. Instead the client is only responsible for providing the parameters in order to execute the query. The listing below shows an ad hoc query that retrieves the collection of Organizations that are classified by the Automotive Industry and the Japan Geography.

```
SELECT id FROM OrganizationWHERE
   id IN (SELECT classifiedObject FROM Classification
          WHERE
               classificationNode IN (SELECT id FROM ClassificationNode
                WHERE path = "/Geography/Asia/Japan'))
  AND
   id IN (SELECT classifiedObject FROM Classification
          WHERE
               classificationNode IN (SELECT id FROM ClassificationNode
```

```
                WHERE path = '/Industry/Automotive'))
```

Web Content Management

The core Web Content Management support within ebXML Registry is provided by the LifeCycleManager and QueryManager services defined by ebRS. A standard SOAP based API to the LifeCycleManager service allows submission of content and metadata over the web. Any submitted content or metadata may be assigned an arbitrary URL relative to the base URL of the registry. The HTTP based API to the QueryManager interface to ebXML Registry acts like a web server and serves the submitted content and metadata at the URL defined for it.

Standardized Metadata

ebXML Registry allows web content to be described or cataloged by standardized Metadata. The ebRIM specification defines a standard set of classes that may be used as metadata to describe arbitrary content. The main metadata classes and their relationships are shown in the UML class diagram in **Figure 4** .

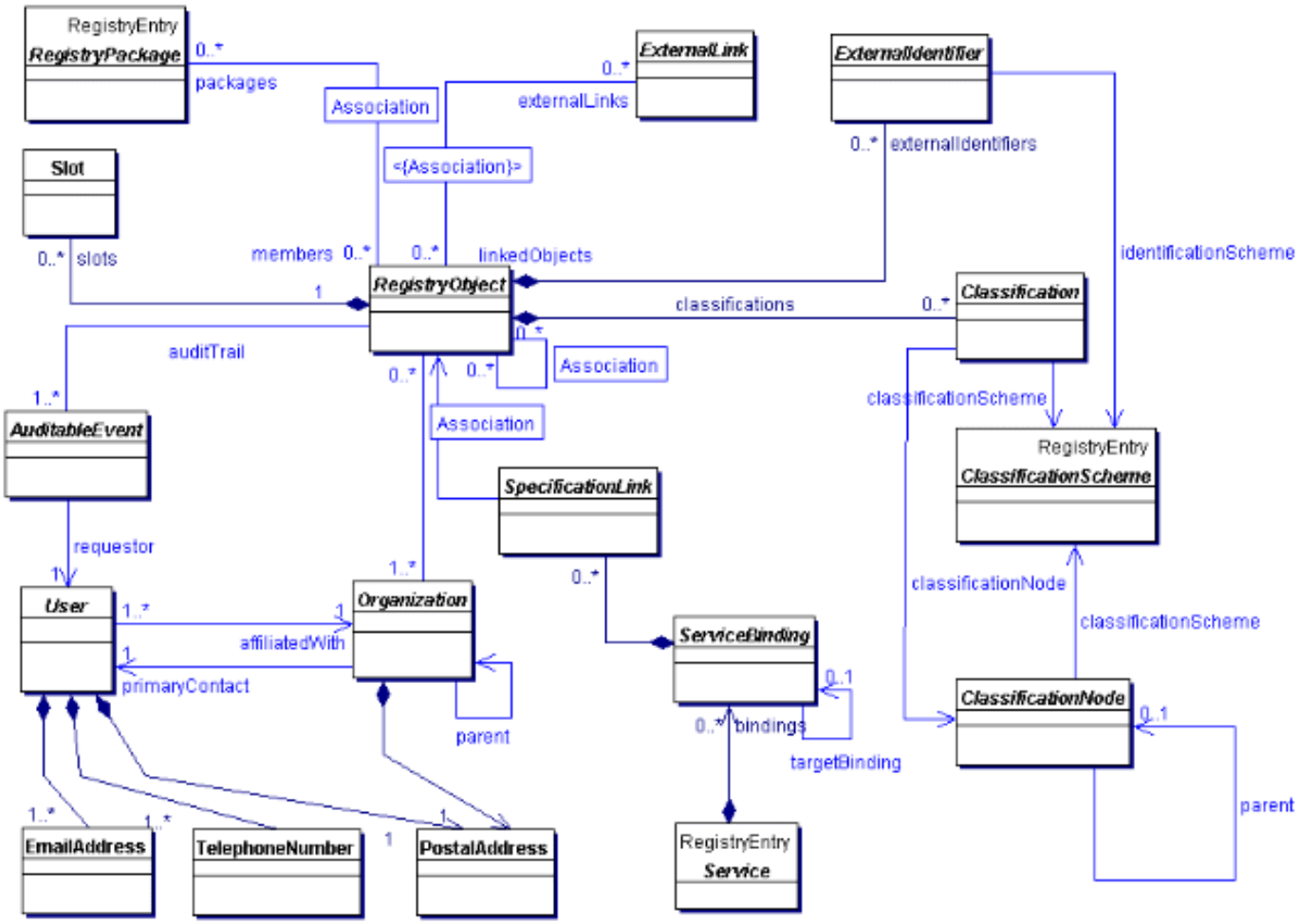

Figure 4: ebXML Registry Information Model: Public View"

At the center of the model is a RegistryObject which is a common base class for all metadata classes. The information model is object oriented as shown in **Figure 5**

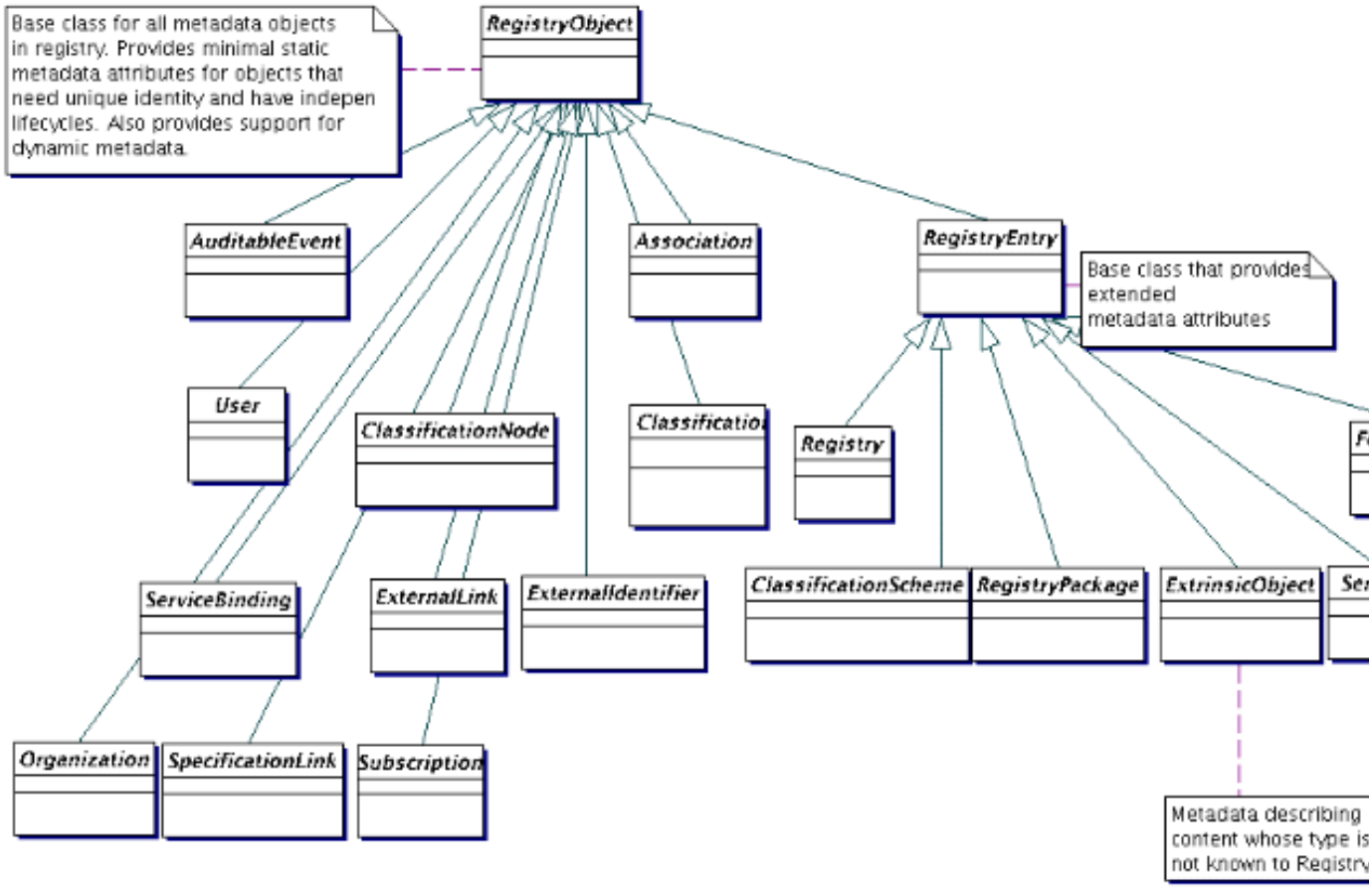

Figure 5: ebXML Registry Information Model: Inheritance View"

The ebRIM allows content to be classified using taxonomies (called ClassificationSchemes). These taxonomies may be entirely user-defined or defined by domain specific vertical standards such as RosettaNet and HL7. The ebXML Registry allows these taxonomies to include complete information about taxonomy values and their structure. This allows the registry clients to browse taxonomy structure and allows the registry to validate taxonomy values when Classification metadata is submitted to the registry.

The ebRIM allows content to be related with other content using relationships (called Associations). These relationships may be entirely user-defined or defined by domain specific vertical standards such as RosettaNet and HL7.

The ebRIM also allows content to be grouped together using the familiar file and folder metaphor using a RegistryPackage which is like a folder and RegistryObjects which are like files. Analogous to folders, a RegistryPackage may contain other RegistryPackages and RegistryObjects. This file and folder metaphor together with user-defined URLs plays a special role in Web Content Management capabilities of ebXML Registry. A user may create new folders structures using RegistryPackage

objects within the registry. They can submit metadata and content as files within specified folders. Each file or folder can have a URL that represents its place within the content hierarchy within the registry.

Content-based Event Notification

ebXML Registry supports a flexible content-based event notification capability which can be used to enable simple workflow by notifying interested subscribers of events related to content managed by the registry. A subscription may be created for receiving a specific type of event. The event is selected using the ad hoc SQL and XML filter query syntax supported by the registry for content discovery. The subscription includes information on how and where to deliver event notifications. Subscribers may be notified by sending the a Notification about the change to a registered a web service or an email address.

Secure Architecture

To ensure content integrity ebXML Registry requires that all publish requests must be signed using XML digital signatures signed by a public key that matches a registered user. In addition all submitted content must also be signed to ensure that it has not been tampered with. The registry is required to verify the signatures within each request that it receives. Similarly all registry responses sent to client are also signed and may be verified by clients to ensure integrity of the response.

During the processing of a client request, the registry uses the digital signature to establish the identity of the user associated with the client. This identity is then used to determine what roles are associated with the user and what groups the user belongs to. Roles and groups may be defined by Registry Administrators according to the need of the community or vertical the registry caters to. It should be noted that groups and roles are managed as special canonical ClassificationSchemes.

The ebXML Registry supports custom fine grained Access Control Policies (ACP) which may be associated with specific object to define who can do what actions to that object. This is based upon the XACML 1.0 standard and may be used to enable Group and Role Based Access Control (RBAC). In the absence of a custom ACP, access to an object is governed by a default ACP that allows anyone to read the object but only the owner or Registry Administrator to modify or delete the object.

All actions within an ebXML Registry that change the state of an object are logged as an AuditableEvent instance within the registry. A complete audit trail is available for all changes that an object has undergone. The audit trail includes who made what change and when.

Federated Architecture

Distributed content management is supported by the ebXML Registry using a federated architecture. Two or more registries may form a loosely coupled federation and appear to be a single logical registry. A client may submit a federated query to any member registry within a federation and the registry must return a unified result set after applying the query to each member of the federation in parallel.

Metadata within an ebXML Registry may reference metadata within any other ebXML Registry. For example an object in one ebXML Registry may be associated with an object in any other ebXML Registry using a remote object reference. Unlike some other registry standards, the ebXML Registry distributed capabilities do not require replicating every objects, all the time to all other registries in the distributed registry topology. The ebXML Registry does however support selective replication of objects from one registry to another for performance and fault tolerance reasons. The registry event notification capability is used to keep replicas in sync across different registries.

As an application of the ebXML Registry federated architecture consider the following scenario. What follows next is a vision of what is achievable with the ebXML Registry federated architecture.

Imagine that The Netherlands government operates a small scale ebXML Registry for each department in the federal government. All these departmental registries are federated together to form The Netherlands Government registry federation as shown in **Figure 6** .

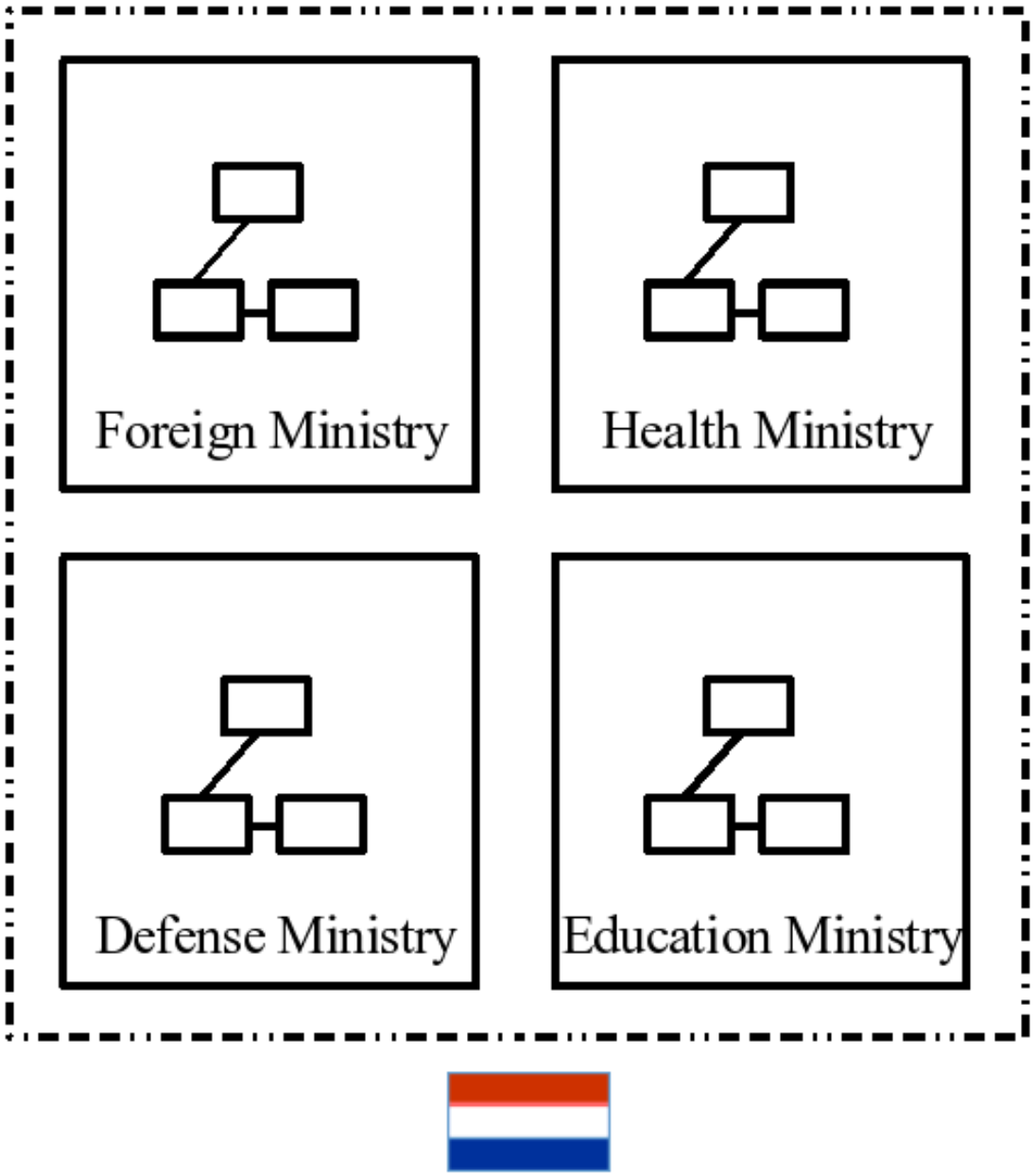

Figure 6: The Netherlands Government Registry Federation

Now imagine that other governments in the EU have similar Federal Government registry federations as shown in **Figure 7** .

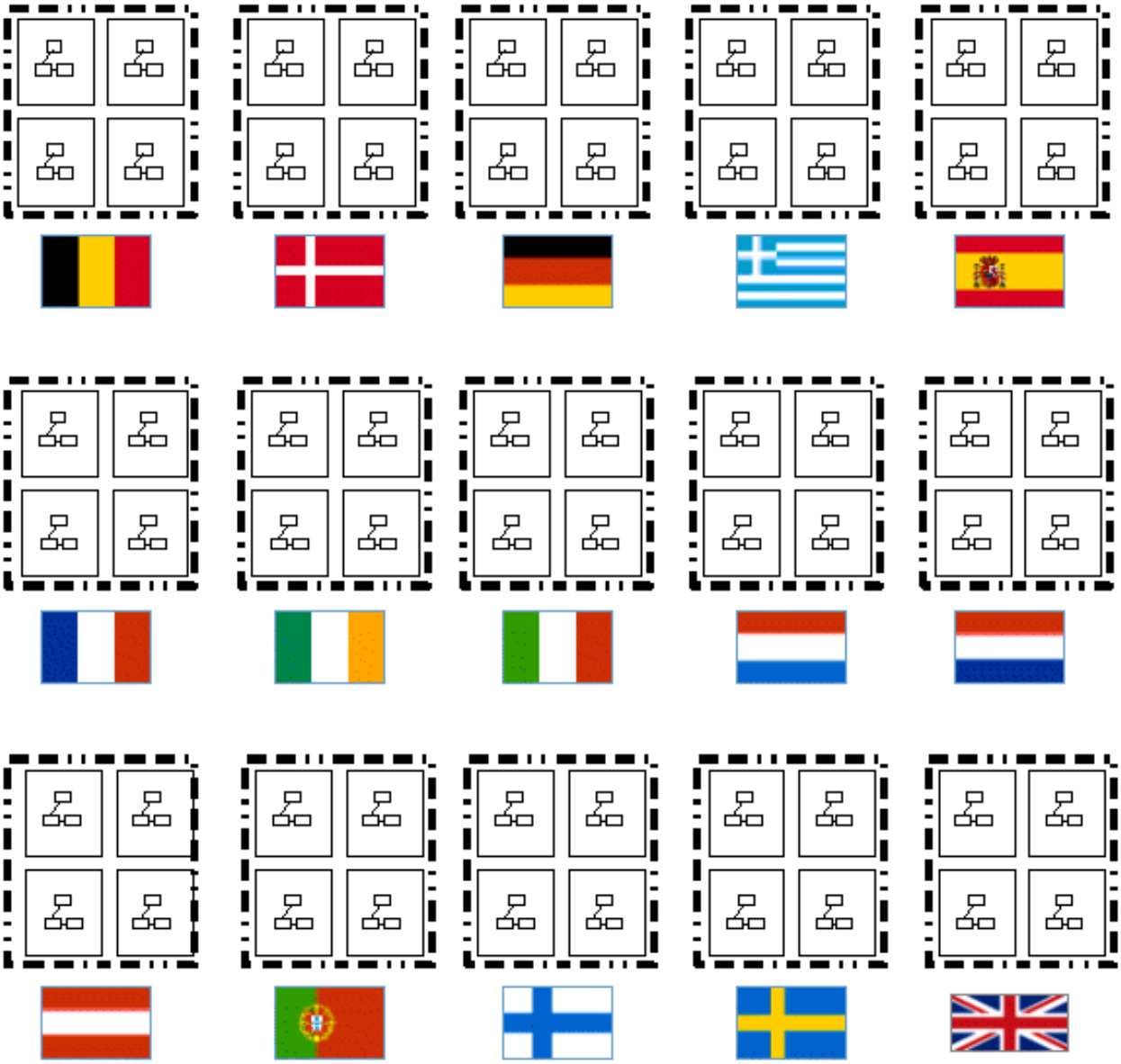

Figure 7: National Government Registry Federations

Now imagine that each nations federation is federated with other EU national federations to form the EU registry federation as shown in **Figure 8** . Note that it is possible for an object in any physical registry to reference objects in any other registry.

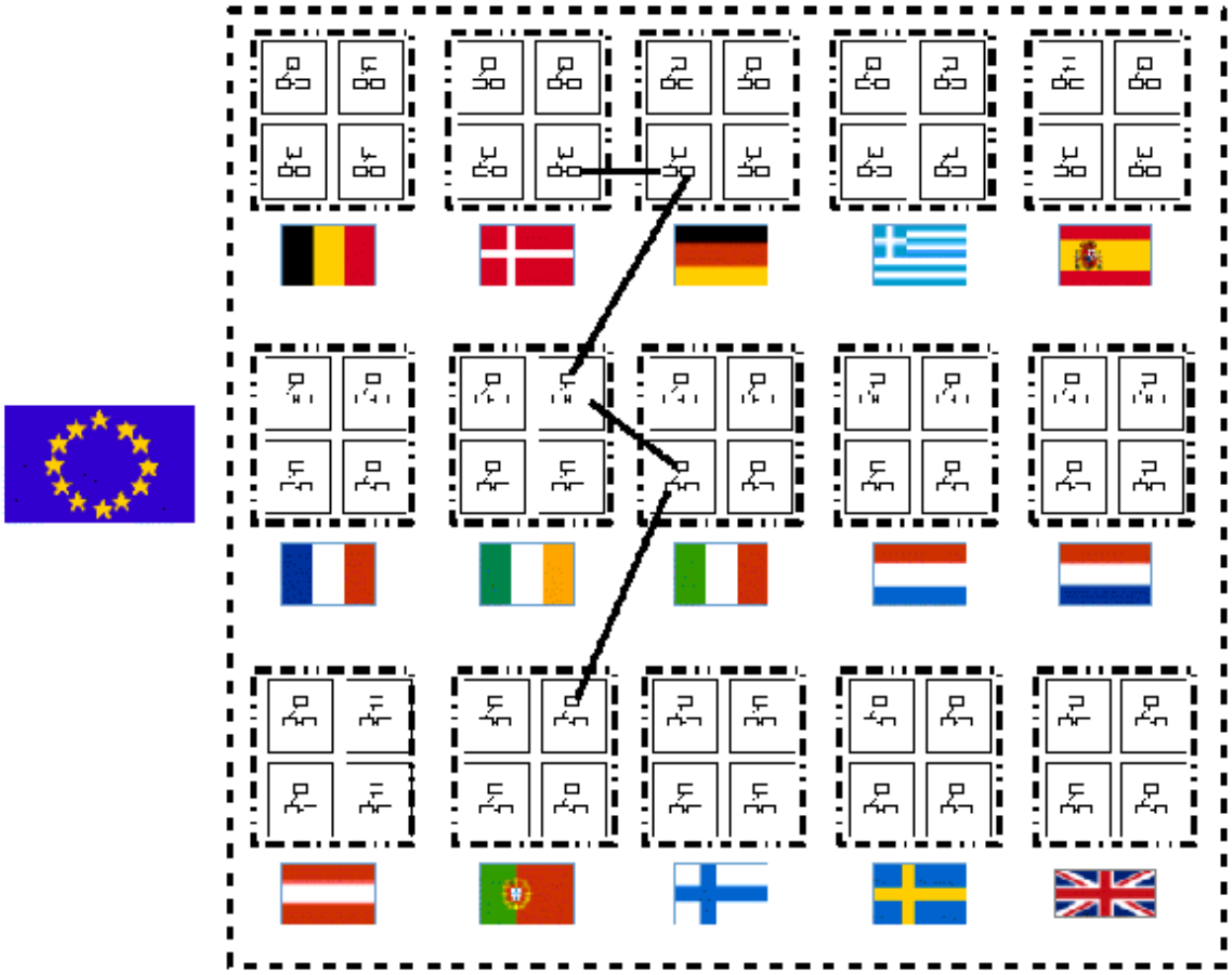

Figure 8: European Union Registry Federation

Because a physical registry may belong to multiple federations, imagine that the Health Ministry registry in each nations is federated with the Health Ministry registry in all other EU nations to form an EU Health Ministry registry federation. This could be useful in Epidemic Management scenarios as described in **[EpidemicMgt]** .

Thus ebXML Registry enables seamless federated content management using a scalable federated architecture that is loosely coupled and does not require tight agreements between participating registries. The EU brings about European unification. ebXML Registry brings about seamless content unification.

ebXML Registry in the Real World

Having described the major features of the ebXML Registry we now describe some actual examples of how ebXML Registry is being used in the real world. Most of these examples are based upon the **[freebXML Registry]** which provides a royalty free open source implementation of ebXML Registry standard and is the de facto reference implementation for the standard.

Adobe, a leading document technology vendor is using the ebXML Registry as a content repository

within it eForms product named Adobe® Intelligent Document Platform product. Users fill out PDF based eForms online. When the form is submitted the form data is stored within the ebXML Registry.

Apelon, a health care information system provider is using the ebXML Registry as a clinical guideline registry to provide decision support for doctors and nurses.

General Motors and Volkswagon, within the automotive industry are using ebXML Registry as a B2B artifacts registry to manage business processes and partner profiles.

Government of Canada is using ebXML Registry in an eGovernment pilot project to standardize processes, roles and business data across Government of Canada within the Government Strategic Reference Model. Republica, a Finnish software vendor provides leading edge B2B collaboration software and services for government agencies. Republica is using ebXML Registry to manage ebXML Core Components and business vocabulary. Republica's ebXML Registry based product is deployed at Finnish Ministry of Finance.

HL7, a health care vertical standard is using ebXML Registry to manage HL7 Conformance Profiles in a an registry operated by NIST in the US Government.

RosettaNet, a leading supply-chain consortium that defines eBusiness Process Standards is using ebXML Registry to manage RosettaNet technical dictionary including business processes and related artifacts.

Sabre, a leading travel commerce vendor uses ebXML Registry to as a B2B artifacts registry to manage business processes, partner profiles, XML schemas and related artifacts.

Statistical Data and Metadata Exchange (SDMX) initiative is sponsored by BIS, ECB, IMF, OECD, EUROSTAT, and UN and World Bank to address standardization of the exchange of statistical information. SDMX.org is using ebXML Registry in a case study involving a virtual hub for external debt data and metadata for various governments.

Sun Microsystems, a leading hardware and software vendor uses ebXML Registry to manage internal and external facing web services. The Sun web services registry is available as a resource to employees and partners.

Future Road Map

Currently the ebXML Registry TC is in the final stages of approving version 3 of the ebXML Registry specifications. All functionality described in this paper corresponds to this version. Work has begin on defining version 4 of the ebXML Registry standard. Next we give a brief glimpse of some important features planned for version 4.

Currently the ebXML Registry supports client managed version control which allows clients to explicitly create new version of objects when they feel it is appropriate. Registry managed version control (RMVC) is planned for version 4. RMVC will perform automatic version control on all metadata and content managed within the registry in a manner similar to version control systems

such as CVS.

Currently the ebXML Registry Information Model (ebRIM) defines a standard set of classes. It also allows these classes to support dynamic attribute extensibility using the Slot class of ebRIM. A Slot instance represents a user-defined (extensible) attribute for an object. The TC is considering adding type extensibility in ebRIM in version 4. Type extensibility would allow verticals and standards to extend ebRIM in a domain specific manner and add new classes to the model.

Version 3 of ebXML Registry focused on Content Management. In January 2004, ebXML Registry TC embarked on a strategic new direction with the establishment of the Semantic Content Management Sub-Committee (SCM SC). The Semantic Content Management SC is chartered to extend ebXML Registry to add direct support for publish, discovery and usage of ontologies and knowledge bases. The goal is to enable collaborative building of distributed knowledge bases and using these knowledge bases as metadata to describe arbitrary content. For example ontology classes may be used to classify content more precisely than current taxonomy based classifications. Knowledge bases may be used by the registry to support inference capabilities when processing semantically aware queries resulting in much more efficient discovery.

Conclusions

The original vision behind ebXML Registry standard was to specify a registry for B2B artifacts and web services. However, it was designed from the ground up as a general purpose content repository. With version 3 it has evolved into an interoperable standard for secure, distributed, web-based content management. It is being deployed in a diverse range of use cases by organizations, governments **[egovRegistry]** , verticals and standards **[WSRP-ebRR]** world wide. It is gaining traction within emergency management use cases as shown in an Epidemic Management demonstration at XML 2003 **[EpidemicMgt]** . ebXML Registry does not replace or compete with Content Management products. It provides them with a standards based API as a path to open, interoperability with other content management products. Recently, ebXML Registry has set a trajectory to add Semantic Content Management capabilities in its next version. The vision is that ebXML Registry will be for the semantic web what web servers are to the web. ebXML Registry aims to become the semantic web server for the semantic web.

Bibliography

[OASIS ebXML Registry Specifications]
The ebXML Registry Specifications are available at (http://www.oasis-open.org/committees/regrep/documents/2.5/specs).

[JAXR]
The Java API for XML Registries 1.0 specification is available at (http://www.jcp.org/en/jsr/detail?id=093).

[freebXML.org]
freebXML.org is an initiative that aims to foster the development and adoption of ebXML and related technology through software and experience sharing. (http://www.freebXML.org).

[freebXML Registry]

The freebXML Registry project provides a royalty free, open source implementation of ebXML Registry and JAXR standards. It may be downloaded from (http://ebxmlrr.sourceforge.com). A brochure on the project is available at (http://ebxmlrr.sourceforge.net/presentations/freebXMLRegistryBrochure.pdf).

[egovRegistry]

An ebxml Registry Repository Open Source Implementation and Its Application to eGovernment. Open Office format (http://ebxmlrr.sourceforge.net/presentations/oasis-demo-xml2003.sxi). Power Point format (http://ebxmlrr.sourceforge.net/presentations/oasis-demo-xml2003.ppt).

[EpidemicMgt]

Epidemic Management Using OASIS ebXML Registry. Open Office format (http://ebxmlrr.sourceforge.net/presentations/ebxmlrr-egov.sxi). Power Point format (http://ebxmlrr.sourceforge.net/presentations/ebxmlrr-egov.ppt).

[WSRP-ebRR]

Using ebXML Registry to Publish, Manage and Discover WSRP Artifacts (Draft) may be downloaded from (http://www.oasis-open.org/committees/download.php/5639/wsrp-pfb-ebxml-tn-draft-03.doc).

</body>