Registry and Repository Face to Face
Sun Microsystems in Burlington, MA
January 4, 2001

Participants

Scott Nieman, Norstan Consulting	Chris Ferris, Sun
Farrukh Najmi, Sun	Len Gallagher, NIST
Yutaka Yoshida, Sun	Mike Kass, NIST
Mike Rowley, Excelon	Sally Fuger, AIAG

Attending by teleconference:

Nikola Stojanovic, Columbine JDS	David Webber, XML Global;
Joe Baran, Extol	

Regrets:

Marianne Hondo	Joel Munter

The major requirement is that there be a canonical query syntax, preferably an existing standard.   After considering various alternatives, the two main contenders were XPATH and OQL. The group agreed to compare XPATH and OQL using the most complicated query situations identified in the Registry Services document.

Query Syntax Selection Table

The group identified a set of requirements and compared the two options. Following is a table showing the results of the comparison.

Priority	Requirement	OQL	XPATH
High	Ease of use by RegistryClients	High	Medium More cryptic syntax
Medium	Ease of use by RegistryServices	Low	High
High	Independent of Implementation (RDBMS, ODB, XML)	High	High
High	Ability to support content based queries on XML managed object content	Low Requires mixed syntax	High
Medium	Ability to support content based queries on non-XML managed object content	Medium May be applied to any content type that has an object model	No support
Low	Ability to call methods defined in the RIM	High	No support
Low	Query processing performance	High Assuming ODB or RDB	Medium Assumes file system + XPATH processor
High	Complex query features	High Superset of SQL, includes AND, OR, NOT, LIKE, Order by, Group by, etc.	Medium AND, OR, NOT, LIKE
High	Ease of acceptance by ebXML membership	Low	Medium

XPATH Mapping Exercise

To evaluate the feasibility of XPATH as the query syntax, we were concerned that the syntax would require storage of metadata as XML. At this time, all known registry/repository implementations use RDB.

Mike Rowley led the exercise to map the OQL classification query to its XPATH equivalent. This proved to be difficult and he asked if we could change our DTDs to make the XPATH query mapping simpler. After a brief discussion, the group determined that changing the RIM to make XPATH implementation easier was not an acceptable solution.

The Breakthrough: A Virtual Document Approach

At this point, Mike suggested a virtual document approach. In this approach, the XPATH queries are based on virtual documents that are convenient views of the actual metadata submitted. The schemas for the virtual documents are algorithmically derived from the RIM. There is one virtual document for each type of query generated.

Use of virtual documents
?? Makes query syntax simpler;
?? Prevents users from creating excessively complex and computationally intensive queries.

Ad Hoc Queries: Do we need them?

Len Gallagher raised a concern about ad hoc queries, and whether there is a need to have them at all. He stated that the objective of a query is to get all managed objects that satisfy certain conditions and unrestricted ad hoc queries can interfere with system efficiency.

The group then discussed whether there is a need for ad hoc queries. It was noted that it is possible for people to, either accidentally or intentionally, submit an over-complicated query that results in excessive use of resources. It was decided that the RA may limit ad hoc queries by defining a time-out which, if the query cannot be executed in that time, will throw a "query too complex" exception.

The group spent some time defining what constitutes an ad hoc query against an ebXML registry. It was agreed that ad hoc queries have the following characteristics:
?? Ability to give the power to define the search criteria to the client
?? Current scope is limited to XML content and metadata
?? Not known in advance

As an alternative to true ad hoc queries, Len Gallagher proposed an alternative method: highly structured non-ad hoc managed object queries. This method limits the query to known joins, restricting the extent to which a query can become over complicated.
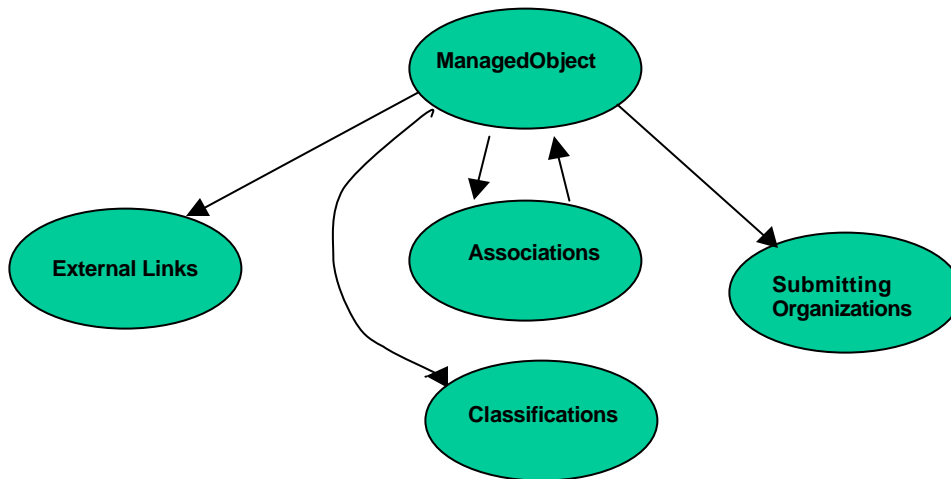
Survey of Group's opinions on Query Syntax

After a review of query approaches including ad hoc using XPATH, ad hoc using OQL, and non-ad hoc but highly structured managed object queries, the majority of the group believed that the preferred method is ad hoc queries using XPATH syntax.

XPATH Exercise Details

Mike Rowley developed several examples of queries based on the structures in the RIM as the group discussed and reviewed them. We created virtual XML document instances, which will need to be converted to DTD or Schema. Based on these example instances, we walked through all the known OQL queries in the RS document and mapped them to XPATH.

The details of the exercise will be sent in a separate document.

End of Summary of Day One, which was a very productive day.

<Notes for Farrukh>

Bug in spec: root classification nodes must have a unique name to distinguish from names of other root classification nodes. Farrukh to investigate this problem.

Inheritance diagram in RIM – make classification be a sibling of association instead of a child.

In RIM, association-related methods need to be added to the interfaces (Information Model Public View – every blue line needs methods added)

Fix relationship between ManagedObject and Package – the same ManagedObject can occur in many packages. Also need to deal with package containing other packages.

Move URI back to ManagedObject.

FPI – investigate – issues log?

</Notes for Farrukh>