Enforcing Interoperability with the Open Archives Initiative Repository Explorer

Hussein Suleman
Department of Computer Science
Virginia Tech
Blacksburg, VA, USA
+1 540 231-3615
hussein@vt.edu

ABSTRACT
The Open Archives Initiative (OAI) is an organization dedicated to solving problems of digital library interoperability by defining simple protocols, most recently for the exchange of metadata. The success of such an activity requires vigilance in specification of the protocol as well as standardization of implementation. The lack of standardized implementation is a substantial barrier to interoperability in many existing client/server protocols. To avoid this pitfall we developed the Repository Explorer, a tool that supports manual and automated protocol testing. This tool has a significant impact on simplifying development of interoperability interfaces and increasing the level of confidence of early adopters of the technology, thus exemplifying the positive impact of exhaustive testing and quality assurance on interoperability ventures.

Categories and Subject Descriptors
D.2.12 [**Software Engineering**]: Interoperability – *Interface definition languages.*

C.2.2 [**Computer-Communication Networks**]: Network Protocols – *Protocol verification.*

General Terms
Reliability, Experimentation, Standardization, Verification.

Keywords
Interoperability, Protocol, Testing, Validation.

1. CONTEXT
The work of the Open Archives Initiative (OAI) was initiated in connection with a meeting of representatives of various electronic pre-print and related archives (e.g. NDLTD, arXiv, NCSTRL) in Santa Fe, USA in October 1999 [2]. From this meeting emanated an agreement among the archivists to support a common set of principles and a technical framework to achieve interoperability. This started the process of defining standards, broadened to include digital libraries other than pre-print archives.

A technical working group provided input into the process of creating and testing those standards under widely differing conditions. This process culminated in the announcement of a new protocol for interoperability, the Open Archives Initiative Protocol for Metadata Harvesting [3], in January 2001. These standards are now being disseminated to all interested parties who wish to adopt a low-cost approach to interoperability, with support from a growing number of members of the Open Archives community.

2. MOTIVATION
After the inaugural meeting of the OAI in 1999, a handful of archivists began to implement the agreed-upon interoperability protocol at their distributed sites. This effort was immediately hampered by a varying interpretation of the protocol specification. This was largely due to the difficulty of precisely specifying a protocol that would both be general and applicable to multiple domains. The client/server architecture chosen by the OAI led to a classic "chicken-and-egg" problem since client implementations would need to interface correctly with server implementations – there was subsequently a low degree of confidence in the correctness of early implementations in each category. Coupled with this, even when clients and servers subscribed to the same interpretation, there was not high confidence that other client/server pairs would interoperate successfully.

As one approach to address these concerns, we developed a protocol tester that would allow a user to perform low-level protocol tests on a server implementation without the need for a corresponding client implementation. This now-widely used software, the Repository Explorer, aids in standardizing the protocol understood by various different archives subscribing to the OAI model of interoperability.

3. DESIGN OF REPOSITORY EXPLORER
The Repository Explorer is implemented as a web-based application (see Figure 1) to take advantage of the ubiquitous nature of WWW clients, and to alleviate the need to install multiple components on client machines to support all the software components used during testing.

The software supports both manual and automatic testing, but with an emphasis on the former. In automatic testing mode, a series of protocol requests, with legal and illegal combinations of parameters, are issued to the archive being tested, and the responses are checked for compliance with the expected range of responses. In manual mode, the software allows a user to browse

through the contents of the archive using only the well-defined interface provided by the protocol – in this instance the user has full control over all parameters and can test individual features of the protocol.

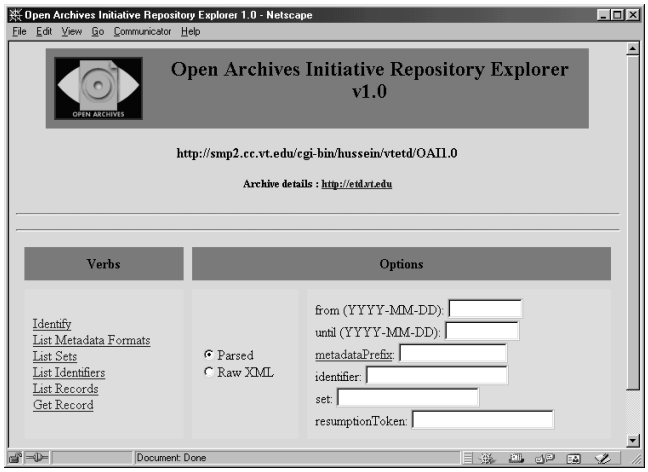

Figure 1. Repository Explorer basic interface

4. VALIDATION PROCEDURE

The OAI protocol is a request/response protocol that works as a layer over HTTP, with responses in XML. The Repository Explorer performs validation at multiple levels in an attempt to detect the widest range of possible errors. Figure 2 depicts the flow of response data received from a server during the validation process.

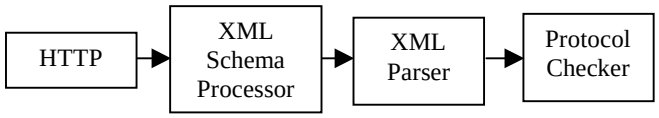

Figure 2. Outline of validation/testing process

Each step of this validation procedure performs incremental checking as described below:

1. When submitting an HTTP request, HTTP errors need to be detected and handled. The OAI protocol requires that explicitly illegal requests generate errors as HTTP status codes, and these are checked for.

2. Once the request is issued and the response is successful, the returned XML needs to be checked for validity. Since all XML responses are specified in the protocol specification using the XML Schema Language [1], this part of the validation is accomplished using an XML Schema processor, which attempts to match the schema with the generated XML data stream. Many structural and encoding errors in the XML are detected in this phase of testing.

3. Unfortunately, schema processing does not always work flawlessly because of its many external dependencies, e.g., schemata are downloaded from the WWW as required. As a redundant mechanism, XML errors are also detected during the parsing and tree generation phase.

4. Lastly, the tree representation is checked for semantic correctness. For example, where controlled vocabularies are used but not encoded into the XML Schema, these are checked at this stage (e.g., the standard list of metadata formats used by the OAI)

Once checks are performed, if the software is being used in manual mode, a new interface is generated to display the data received from the server and present the user with additional options to perform further operations.

5. CONCLUSIONS

Over the course of implementing the original and revised protocol, most if not all implementers used the Repository Explorer to test their implementations. Their positive feedback supported the original motivation that a compliance test would greatly ease the process of implementation.

Many lessons were learnt during the process of designing the test software, the most significant being that testing is a decidedly non-trivial problem if the target (namely, the protocol specification) is not stationary. However, designing a new version of the protocol in tandem with updating the test software ensured that the new specifications contain mechanisms (like self-identification) that can be exploited to perform more exhaustive automatic tests.

6. FUTURE WORK

Further standardization of the software libraries used in development can lead to a tool suite that will not only be adaptable to future versions of the OAI protocol, but also possibly to other client/server protocols. While XML tools are still not widely deployed and are notorious for high degrees of complexity, future combinations of XSD (Schema), XSLT (Transformation) and XPath (Path Specifications) could make protocol testing a fully automated process for emerging client/server protocols using XML as underlying technology.

The test suite is under constant development as best practices emerge among users of the community, thus ensuring that the Repository Explorer maintains its status as a compliance test, for which it was intended.

7. ACKNOWLEDGMENTS

Many thanks go to the group of OAI protocol authors and testers who used the Repository Explorer and provided useful feedback.

8. REFERENCES

[1] Fallside, David C. (editor). XML Schema. W3C, October 2000. http://www.w3.org/XML/Schema

[2] Van de Sompel, Herbert and Carl Lagoze. The Santa Fe Convention of the Open Archives Initiative. D-Lib Magazine, Volume 6, Number 2, February 2000. http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html

[3] Van de Sompel, Herbert and Carl Lagoze. The Open Archives Initiative Protocol for Metadata Harvesting. Open Archives Initiative, January 2001. http://www.openarchives.org/OAI/openarchivesprotocol.htm