Leveraging OAI Harvesting to Disseminate Theses

Hussein Suleman, University of Cape Town (hussein@cs.uct.ac.za)
Edward A. Fox, Virginia Tech (fox@vt.edu)

ABSTRACT
NDLTD, the Networked Digital Library of Theses and Dissertations, supports and encourages the production and archiving of Electronic Theses and Dissertations (ETDs). While many current NDLTD member institutions and consortia have individual collections accessible online, there has until recently been no single mechanism to aggregate all ETDs to provide NDLTD-wide services (e.g., searching). With the emergence of the Open Archives Initiative (OAI), that has changed. The OAI's Protocol for Metadata Harvesting is a robust interoperability solution that defines a standard method of exchanging metadata. While working with the OAI to develop and test the metadata harvesting standard, we have set up and actively maintain a central NDLTD metadata collection and multiple user portals. Member sites are encouraged to contribute to this central archive by supporting the OAI protocol, along with particular standards and conventions that have been specifically devised to support ETDs. We discuss in this article our experiences in building this distributed digital library based upon the work of the OAI.

Keywords
NDLTD, ETD, interoperability, standards, protocols, theses and dissertations

Author Information

Hussein Suleman
Department of Computer Science
University of Cape Town
Cape Town
South Africa
+27 21 650 2663
hussein@cs.uct.ac.za

Dr. Hussein Suleman is a Senior Lecturer in Computer Science at the University of Cape Town and a past member of the Digital Library Research Laboratory at Virginia Tech. His research relates to aspects of digital library architecture and interoperability. He works closely with the Open Archives Initiative in developing and supporting standards and with NDLTD in implementing those standards.

Edward A. Fox
Department of Computer Science
Virginia Tech
Blacksburg, VA, USA
+1 540 231 5113
fox@vt.edu

Dr. Edward A. Fox, director of NDLTD, is a Professor of Computer Science at Virginia Tech. There, he also directs the Internet Technology Innovation Center and Digital Library Research Laboratory. He is currently chairman of the NSDL Policy Committee and serves on the OAI Steering Committee.

Leveraging OAI Harvesting to Disseminate Theses

ABSTRACT

NDLTD, the Networked Digital Library of Theses and Dissertations, supports and encourages the production and archiving of Electronic Theses and Dissertations (ETDs). While many current NDLTD member institutions and consortia have individual collections accessible online, there has until recently been no single mechanism to aggregate all ETDs to provide NDLTD-wide services (e.g., searching). With the emergence of the Open Archives Initiative (OAI), that has changed. The OAI's Protocol for Metadata Harvesting is a robust interoperability solution that defines a standard method of exchanging metadata. While working with the OAI to develop and test the metadata harvesting standard, we have set up and actively maintain a central NDLTD metadata collection and multiple user portals. Member sites are encouraged to contribute to this central archive by supporting the OAI protocol, along with particular standards and conventions that have been specifically devised to support ETDs. We discuss in this article our experiences in building this distributed digital library based on the work of the OAI.

INTRODUCTION TO NDLTD

The Networked Digital Library of Theses and Dissertations (NDLTD) is a loose federation of member institutions and organizations that publish Electronic Theses and Dissertations (ETDs). Since its 1996 inception, NDLTD has grown to include over 100 individual member institutions and many consortia from countries around the world (Fox et al., 1996; Fox et al., 1997; Suleman et al., 2001).

The mission of NDLTD includes the development of graduate education; preservation and dissemination of knowledge; and the advancement of current technology to support publishing, including multimedia tools and digital libraries. Central to its mission is the dissemination component and, to this end, various efforts have centered on enhancing access to ETDs. These include the distribution of software toolkits to manage the ETD workflow as well as easily share ETDs with remote sites, the development of standards for ETD metadata, and central portals that users may use to easily locate ETDs at institutions scattered around the globe. The last of these, based on a Union Catalog, has benefited greatly from current developments in digital library interoperability, especially the standards developed by the Open Archives Initiative (OAI).

EARLY EFFORTS IN INTEROPERABILITY

While not a primary goal, it has long been a desire of NDLTD to unite the various member sites into a single collection to support the researcher or student seeking ETDs (Fox et al., 1996; Fox et al., 1997).

The first attempt to provide services across multiple NDLTD members was a federated search system (Powell & Fox, 1998). This system allowed users to go to a central site and perform a meta-search across multiple ETD collections. The search language and syntax at each site was known and described internally in a precise specification language. This then was used to reformulate the users' queries appropriately for each site. The search results were cached as they arrived but the results were not merged. This federated search system was popular as a proof-of-concept of the capabilities of federated searching. There were, however, many shortcomings with this initial system, including:

- The reliance on consistent operations at every participating member site resulted in more points of failure – every time a member site changed any aspect of its search engine interface, the federated search system would need reconfiguration.

- Searches were as slow as the slowest member site.

- If a remote site was momentarily unreachable, the federated search results would be incomplete.

- There was no simple way to merge result sets since this would require parsing of the result set pages (which were in HTML) as well as retrieving each individual document or metadata record listed prior to ranking.

- It was not always simple to add new remote sites to the system since each new site could require the specification of a new search language and/or syntax. Furthermore, when remote search engines were not based on single-request stateless operations, additional complexity could result from the client-server interaction.

Similar observations were made by other projects that relied on federated searching. A notable example is the original Networked Computer Science Technical Reference Library (NCSTRL) project (Leiner, 1998), which used the Dienst protocol and software (Lagoze & Davis, 1995) to create a distributed digital library of technical reports in the field of computer science. Reliability and speed problems were addressed in NCSTRL by replication and caching. The complexity of mapping search queries and results was avoided by establishing a standard protocol and an associated software toolkit. In spite of this, adoption was considered a technical hurdle by many implementers. In addition, each remote site was a potential point of failure and this prompted further discussion of approaches to interoperability.

In order to address all of these problems, representatives of NCSTRL, NDLTD, and various other organizations that had an interest in digital library interoperability, met in Santa Fe in October 1999 to develop a simpler solution for large-scale production-quality interoperability. The result of this meeting was the Santa Fe Agreement (Van de Sompel & Lagoze, 2000), a set of guiding principles for interoperability, and a protocol for transferring metadata among sites. The organization that grew out of that meeting, the Open Archives Initiative [1], has further developed the original agreement and actively advocates and supports it in its current form – the Open Archives Initiative Protocol for Metadata Harvesting (Lagoze et al., 2002).

THE OAI PROTOCOL FOR METADATA HARVESTING (OAI-PMH)

The OAI-PMH is a client-server protocol based on HTTP, which facilitates the incremental transfer of metadata among networked systems. The protocol is designed to be simple and general, thus suitable for use in various different contexts and communities. NDLTD, as one such community, has adopted this protocol to develop an alternative to the federated search system used in the past.

The OAI-PMH differs from previous approaches to interoperability in that there is more limited interaction among remote sites and the central site. The primary purpose of the protocol is incremental bulk transfer of metadata (harvesting) – there is no remote search facility. Instead, a provider of services acquires data from a data provider, stores and processes it locally, and then supplies services to users

based on that data. A brief summary of the requests that make up the core of the protocol is provided in Table 1.

Service Request	Expected Response
Identify	Description of archive - standards and protocols implemented
ListMetadataFormats	List of supported metadata formats
ListSets	List of archive sets and subsets
ListIdentifiers	List of record identifiers, optionally corresponding to a specified set and/or date range
GetRecord	Single metadata record corresponding to a specified identifier and in a specified metadata format
ListRecords	List of metadata records corresponding to a specified metadata format and, optionally, a set and/or date range

Table 1. OAI-PMH service requests and expected responses

The OAI-PMH uses current standards wherever applicable. All data that is transferred in response to a request is encoded in an XML format defined using XML Schema (Fallside, 2001). The requirement of conformance to an XML Schema means that OAI-based tools such as the Repository Explorer (Suleman, 2001) can perform machine validation of responses. In current OAI-PMH implementations, responses are generated using the HTTP CGI mechanism because of its stability as a Web technology – when Web Services standards such as SOAP (Box et al., 2000) are eventually finalized, the OAI-PMH can easily be amended to use those mechanisms.

NDLTD has adopted use of the OAI protocol for metadata transfer because of the ease with which the protocol can be customized to handle the specific needs of NDLTD. Included among the advantages of OAI-PMH are:

- Simplicity: various toolkits can be (and have been) developed to ease adoption among ETD sites.
- Support for arbitrary metadata formats: NDLTD has developed a new standard for metadata specific to ETDs (as discussed later) and this is trivially supported by the OAI protocol.
- Sets within the archive: in the context of NDLTD these can correspond to geography, subject areas, and organizational structures.

THE NDLTD UNION ARCHIVE

As NDLTD members implement the OAI protocol on their individual archives, work has begun on developing mechanisms and policies for collecting metadata into a central site for the provision of user-level services across all participating NDLTD members.

In order to support multiple service providers, the Union Archive was created with the purpose of collecting metadata from remote sites and republishing it as a single collection. Sites that provide user-

level services such as searching and browsing then can harvest from this single collection. Some of the advantages of a central merged collection of metadata are:

- Multiple services can be developed at a central site, without having to harvest metadata multiple times – this is analogous to caching or replication.
- The development of experimental services does not negatively impact the production servers of individual members by repeatedly requesting transfers of the same metadata.
- The system is robust – all metadata may be accessed locally irrespective of the current status of remote sites.
- The merged collection can easily be replicated using mechanisms such as those developed by the Internet2 Distributed Storage Initiative (Pande & et al., 2002).

The Union Archive is designed to function as both a provider of services (harvester) and a provider of data. It harvests metadata from the remote sites, stores it in an internal database and then republishes this metadata through its own OAI data provider interface. This is illustrated in Figure 1. The depicted Union Catalog components refer to systems that process the metadata and provide services to users while the Union Archive simply accumulates the metadata.

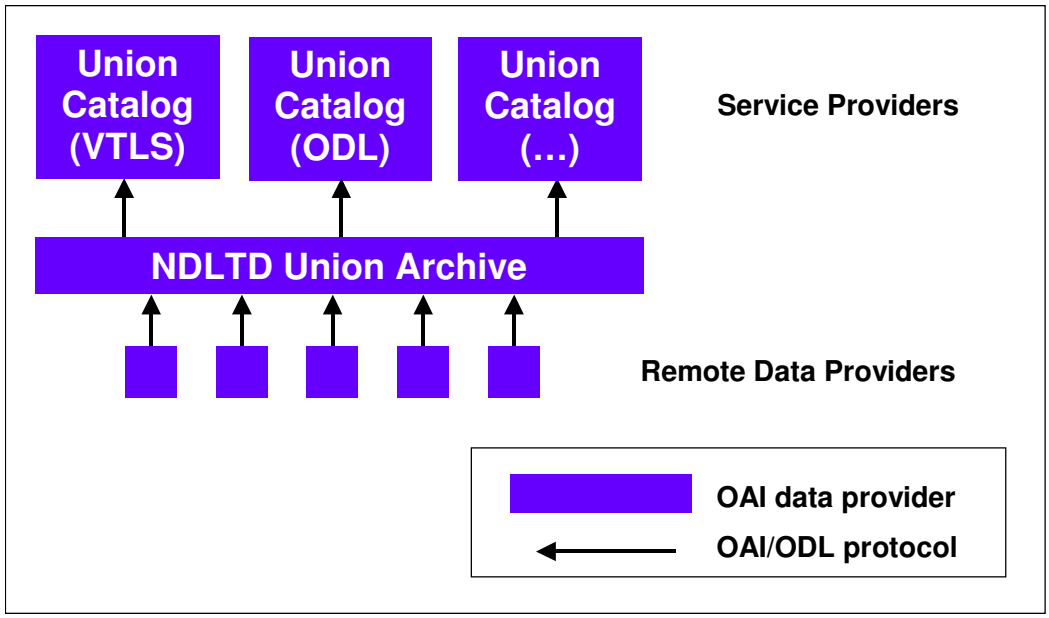

Figure 1. Architecture of Union Archive and Catalogs

There are multiple service providers that harvest from the Union Archive, including the official VTLS Union Catalog for NDLTD [2] and the experimental Open Digital Library (ODL) Union Catalog. Each of these systems provides a public user interface with search and browse capabilities. As of the fall of 2002 there are 13 ETD collections that participate in this project, and more are in the process of making their collections OAI-accessible. The collections and their respective document sizes are listed in Table 2.

Some additional sites contributed metadata by means other than OAI directly to the Union Catalog maintained by VTLS. These records are in the process of being incorporated into the Union Archive.

Archive	# Records
Virginia Tech	3562
Humboldt University of Berlin	301
University of Duisburg	389
Technical University of Dresden	18
PhysNet	185
MIT	72
CalTech	102
Uppsala University	1700
University of British Columbia	2
Louisiana State University	351
North Carolina State University	178
University of South Florida	39
ETD Individuals	3
Total	6902

Table 2. Contents of Union Archive (as of 18 September 2002)

Harvesting Schedule

Version 1.1 of the OAI protocol supports incremental harvesting whereby service providers may request records by specifying only a date range. Harvesting algorithms designed to deal with the absence of time information have to contend with either marginally outdated data or duplicate records (Suleman & Fox, 2002a). In practice, the Union Archive harvests data from remote locations once daily since ETD metadata is relatively static and it is not critical to incorporate new entries immediately. There is still some duplication of harvested records but this will be eliminated in the future when all parties switch to using OAI-PMH v2.0 since the new version of the OAI protocol supports times in addition to dates.

Configuration of Union Archive

The configurable information for the Union Archive is stored in an XML file, which defines the archives to harvest metadata from as well as parameters to control the harvesting algorithm and the Union Archive in general. In the latter category, a database in which the harvested data is stored is defined in terms of the parameters necessary to make connections to it. Then, parameters are defined for each archive as listed in Table 3. Harvesting is attempted at most every “interval” days and with “interrequestgap” seconds of delay between requests to avoid flooding data providers with requests. An external scheduler (cron) periodically executes a program that tracks the elapsed time since the last harvest and initiates a new harvesting operation if necessary.

Field	Semantics
identifier	Unique label for the data provider
url	*baseURL* of the data provider
metadataPrefix	*metadataPrefix* for metadata format to harvest
interval	Periodicity of harvesting (in days)
interrequestgap	Number of seconds to delay between consecutive requests to the archive
set	*setSpec* of set to harvest (if omitted, harvest entire archive)

Table 3. Semantics of fields in configuration file

OPEN DIGITAL LIBRARY COMPONENTS

The Union Archive was developed in conjunction with and as a test case for the Open Digital Library (ODL) methodology (Suleman & Fox, 2001; Suleman & Fox, 2002b). ODL is a set of guiding principles and protocols for building digital libraries as networks of communicating components, where each component is an extended Open Archive.

All ODL protocols are based on an Extended OAI-PMH v1.1 (XOAI-PMH), which supports inter-component interaction by introducing general data containers into all responses, a finer granularity for timestamps, and a service request to submit records to an archive. Some of these are already incorporated into the latest OAI protocol (v2.0) so a future ODL framework will require fewer extensions.

Based on the XOAI-PMH, a set of specialized protocols was devised to serve as the interfaces to components that provided basic digital library functionality. For the ETD Union Catalog prototype, this included components to browse, search, track new items, filter archives that are not fully OAI-compliant, and merge together OAI data sources. The last of these serves as the core of the Union Archive, together with filters for some of the archives which are not fully compliant with the OAI protocol. The other components support a rudimentary user interface for resource discovery, as shown in Figure 2.

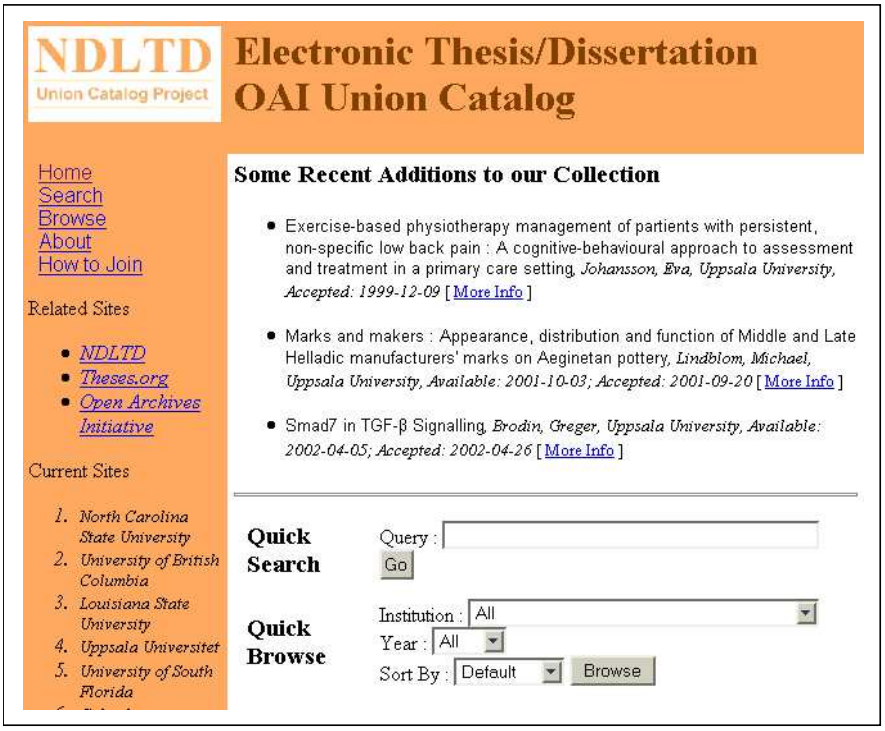

Figure 2. ETD Open Digital Library

Each of the user interface components harvests metadata on a periodic basis from the Union Archive. The harvested data is then pre-processed appropriately depending on the functionality of the component – for example, the search engine component builds inverted files. When a user interacts with the system, the user interface layer communicates with each component using specialized versions of the XOAI-PMH. In the case of the search engine, query terms are encoded into the OAI-PMH set parameter, as specified by the ODLSearch protocol (Suleman & Fox, 2002b). The response from the component is a list of identifiers or records (depending on whether ListIdentifiers or ListRecords was issued) that are then transformed, using XSLT, into HTML and combined with other fragments of HTML before being sent back to the user's browser.

This componentized approach was arguably effective as a strategy for developing digital library services for NDLTD because:

- Components were replaced/upgraded without any effect on the whole system.
- Several components were reused in other projects.
- Service providers such as VTLS were easily able to harvest metadata directly from the Union Archive in order to build a separate production-quality user portal.
- Research projects such as VIDI (Wang, 2002) were able to integrate with the ODL system at the higher level of the search engine.

DATA ENCODING ISSUES

With many NDLTD members being from non-English-speaking countries, languages and character sets are an important problem that needs to be solved in any distributed digital library solution. Of the 13 current participants in the project, 5 use metadata that cannot be encoded using standard ASCII. While extended ASCII may work for now, it is anticipated that other members from Asian countries will provide data that cannot be encoded in any variation of ASCII.

As a general solution to this problem, the OAI requires that all data transmitted by the OAI protocol be encoded in UTF-8 with numerical Unicode entities wherever necessary. This ensures that there is no loss of data and that validation of XML is always possible without the need for external entity files. However, it raises the bar on the development of tools to process this data. All tools must support UTF-8 – this is usually not much of a problem as long as data is converted to a human-readable format when needed. But since the XML standard requires that parsers translate all entities, any intermediate XML parser or XSLT transformation converts entities to an appropriate human-readable form. To work around these conversions of the intermediate data, Unicode entities are double-escaped in the Union Archive before any XML fragments are passed to XML parsers. For example, to prevent an intermediate XML parser from converting "<" into "<", the "<" can be escaped into "<" before parsing and that will get safely converted into "<".

Dynamic WWW-based user interfaces (such as that shown in Figure 2) introduce additional encoding complexity since the set of special characters in the CGI specification (Gundavaram, 1996) does not correspond to the set of special characters in the XML specification. Also, browsers will automatically escape some characters, such as spaces, to comply with the HTML specification. This resulted in additional escaping and unescaping of data before and after communication between server and client.

ETD METADATA SET (ETDMS)

ETDMS is a standard developed for expressing metadata related to theses and dissertations (Atkins, 2001). The Union Archive harvests records in ETDMS from all archives that support it, in addition to the default Dublin Core format (Dublin Core Metadata Initiative, 1997). Currently, only 8 of the 13 archives support ETDMS, but all new sites are encouraged to support this richer standard.
ETDMS builds on the Dublin Core (DC) standard by:

- Adding in an element for an alternative title.
- Specifying tightly constrained semantics for the existing DC fields, e.g., dates may only be in certain ISO8601 formats.
- Adding in elements specific to the degree rather than the ETD: name of the degree, level of the degree, area of study, and institution that granted the degree.
- Adding optional attributes to free-text fields to specify language of metadata, translation status, vocabulary scheme, and a name authority link.

TOOLS AND ETD-DB EXTENSIONS

To support participation in the Union Archive, an extension to the ETD-db ETD management software (Atkins et al., 2001) was created. This extension is a drop-in module to retrospectively add OAI-PMH v1.1 support to an archive that uses any version of the ETD-db software - the current version of ETD-db comes prepackaged with it. The ETD-db extensions are used by 7 of the 13 sites participating in the Union Archive.

The software is written in Perl and has the same system requirements as ETD-db – thus, there are no additional system requirements for an existing site running ETD-db to become an OAI data provider. The software need only be copied and configured, where the configuration script confirms that the databases are accessible and allows the installer to enter OAI-specific parameters such as the "repository name" and "administrator's email". Once all parameters are specified, the ETD collection is accessible as an OAI data provider through the web server's CGI mechanism.

The OAI extension to ETD-db will export metadata in 4 different formats: Dublin Core, ETDMS, MARC, and RFC1807. For each record, the Dublin Core version is equivalent to the ETDMS version, with additional tags removed. The MARC version of the record follows the crosswalk recommended by the ETDMS specification, while the RFC1807 version is simply a best-effort mapping since no official crosswalk exists.

Similar tools to these extensions are available for the E-Prints software [3], used by at least one of the 13 participating sites. Various toolkits are made available at the OAI website to ease the development of OAI interfaces to custom-built digital libraries as well as any other digital libraries that lack OAI support.

RIGHTS MANAGEMENT

Currently, only freely accessible metadata is harvested from ETD archives. Thus, when the unmodified metadata is republished, no intellectual property rights are violated. Also, the Union Archive retains all original record identifiers so that they may be traced to their respective sources, and to enable duplicate detection.

In terms of digital objects, the OAI can allow linking from metadata to HTML pages within a digital library, rather than directly to the data. The net effect of this is that the source archives can enforce rights management for their resources. The Union Archive uses this mechanism to avoid dealing with the myriad of rights policies in use among NDLTD members.

FUTURE WORK

Research is currently underway to build more reliable archives with better performance by replicating metadata over high-speed network connections; this is related to the Internet2 Distributed Storage Initiative (Pande et al., 2002) and the LOCKSS project (Reich & Rosenthal, 2001). Within the ODL project, higher-level digital library services are being developed to facilitate alternative forms of discovery (e.g., by recommendation).

ETDMS provides the potential for interesting services in the future. For example:

- Constrained vocabularies may be used to build subject indices for ETDs.

- Name authority links may be used to link ETDs to other publications from a single person.

As the user-level services become more popular, the Union Catalogs and Archive are becoming important components in ETD discovery solutions. Ultimately, using OAI-PMH is part of the larger initiative to provide high quality and high availability digital library services to students and researchers, with minimum effort from contributing sites. To get closer to this goal, all NDLTD members are encouraged to contribute to the Union Archive. More information can be found on the NDLTD website [4].

ACKNOWLEDGEMENTS
Thanks are given for the support of NSF through its grants: IIS-9986089, IIS-0002935, IIS-0080748, IIS-0086227, DUE-0121679, DUE-0121741, and DUE-0136690.

NOTES

1. http://www.openarchives.org/ (Open Archive Initiative).

2. http://www.vtls.com/ndltd/ (VTLS NTLTD Union Catalog).

3. http://www.eprints.org/ (EPrints.org, Open Citation Project).

4. http://www.ndltd.org/union.html (NDLTD)

REFERENCES

Atkins, A., (2001). Resources for Developers of ETD databases. Website http://scholar.lib.vt.edu/ETD-db/developer/.

Atkins, A., E. A. Fox, R. France, and H. Suleman (editors), (2001). *ETD-ms: an Interoperability Metadata Standard for Electronic Theses and Dissertations -- version 1.00,* NDLTD. Available http://www.ndltd.org/standards/metadata/ETD-ms-v1.00.html.

Box, D., D. Ehnebuske, G. Kakivaya, A. Layman, N. Mendelsohn, H. Frystyk Nielsen, S. Thatte, and D. Winer, (2000). *Simple Object Access Protocol (SOAP) 1.1*, W3C Note, 8 May 2000. Available http://www.w3.org/TR/SOAP/.

Dublin Core Metadata Initiative, (1997). *Dublin Core Metadata Element Set Version 1.1: Reference Description*. Available http://www.dublincore.org/documents/dces/.

Fallside, D. C. (editor), (2001). *XML Schema, Part 0, Part 1, and Part 2*. W3C Recommendation, 2 May 2001. Available http://www.w3.org/TR/xmlschema-0/, http://www.w3.org/TR/xmlschema-1/, and http://www.w3.org/TR/xmlschema-2/.

Fox, E. A., J. L. Eaton, G. McMillan, N. A. Kipp, L. Weiss, E. Arce, and S. Guyer, (1996). "National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources", *D-Lib Magazine*, Volume 2, Number 8, September 1996. Available http://www.dlib.org/dlib/september96/theses/09fox.html.

Fox, E. A., B. DeVane, J. L. Eaton, N. A. Kipp, P. Mather, T. McGonigle, G. McMillan, and W. Schweiker, (1997). "Networked Digital Library of Theses and Dissertations: An International Effort Unlocking University Resources", *D-Lib Magazine*, Volume 3, Number 9 September 1997. Available http://www.dlib.org/dlib/september97/theses/09fox.html.

Gundavaram, S., (1996). *CGI Programming on the World Wide Web*, O'Reilly and Associates, March 1996. Available http://www.oreilly.com/openbook/cgi/.

Lagoze, C. and J. R. Davis, (1995). "Dienst – An Architecture for Distributed Document Libraries", *Communications of the ACM*, Volume 38, Number 4, April 1995, p. 47.

Lagoze, C. and H. Van de Sompel, (2001). "The Open Archives Initiative: Building a low-barrier interoperability framework", *Proceedings of JCDL 2001*, Roanoke, VA, USA, June 2001, ACM Press, pp. 54-62.

Lagoze, C., H. Van de Sompel, M. Nelson, and S. Warner, (2002). *The Open Archives Initiative Protocol for Metadata Harvesting*, Open Archives Initiative, June 2002. Available http://www.openarchives.org/OAI/openarchivesprotocol.html.

Leiner, B. M., (1998). "The NCSTRL Approach to Open Architecture", *D-Lib Magazine,* Volume 4, Number 12, December 1998. Available http://www.dlib.org/dlib/december98/leiner/12leiner.html.

Pande, A., M. Kothapalli, R. Richardson, and E. A. Fox, (2002). "Mirroring an OAI Archive on the I2-DSI channel", *Proceedings of JCDL 2002*, Portland, Oregon, USA, July 2002, ACM Press, pp. 293-294.

Powell, J. and E. A. Fox, (1998). "Multilingual Federated Searching Across Heterogeneous Collections", *D-Lib Magazine*, Volume 4, Number 8, September 1998. Available http://www.dlib.org/dlib/september98/powell/09powell.html.

Reich, V. and D. S. H. Rosenthal, (2001). "LOCKSS: A Permanent Web Publishing and Access System", *D-Lib Magazine*, Volume 7, Number 6, June 2001. Available http://www.dlib.org/dlib/june01/reich/06reich.html.

Suleman, H., (2001). "Enforcing Interoperability with the Open Archives Initiative Repository Explorer", *Proceedings of JCDL 2001*, Roanoke, VA, USA, June 2001, ACM Press, pp. 63-64.

Suleman, H., A. Atkins, M. A. Gonçalves, R. K. France, E. A. Fox, V. Chachra, M. Crowder, and J. Young, (2001). "Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access – Part 1: Mission and Progress", and "Part 2: Services and Research", *D-Lib Magazine,* Volume 7, Number 9, September 2001. Available http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html and http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html.

Suleman, H. and E. A. Fox, (2001). "A Framework for Building Open Digital Libraries", *D-Lib Magazine,* Volume 7, Number 12, December 2001. Available http://www.dlib.org/dlib/december01/suleman/12suleman.html.

Suleman, H. and E. A. Fox, (2002a). *Beyond Harvesting: Digital Library Components as OAI Extensions*, Technical Report, Virginia Tech NCSTRL Collection, January 2002.

Suleman, H. and E. A. Fox, (2002b). "Designing Protocols in Support of Digital Library Componentization", *Research and Advanced Technology for Digital Technology: Proceedings of the 6th European Conference*, ECDL 2002, Rome, Italy, September 16-18 2002, pp. 568-582.

Van de Sompel, H. and C. Lagoze, (2000). "The Santa Fe Convention of the Open Archives Initiative", *D-Lib Magazine*, Volume 6, Number 2, February 2000. Available http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html.

Wang, J., (2002). *A Lightweight Protocol Between Visualization Tools and Digital Libraries*, Master's Thesis, Virginia Tech, Blacksburg.