Technical Report Interchange Through Synchronized OAI Caches

Xiaoming Liu[1], Kurt Maly[1], Mohammad Zubair[1], Rong Tang[1], Mohammed Imran Padshah[1], George Roncaglia[2], JoAnne Rocker[2], Michael Nelson[2], William von Ofenheim[2], Richard Luce[3], Jacqueline Stack[3], Frances Knudson[3], Beth Goldsmith[3], Irma Holtkamp[3], Miriam Blake[3], Jack Carter[3], Mariella Di Giacomo[3], Major Jerome Nutter[4], Susan Brown[4], Ron Montbrand[4], Sally Landenberger[5], Kathy Pierson[5], Vince Duran[5], and Beth Moser[5]

[1] Old Dominion University.
Norfolk, Virginia, USA
[2] NASA Langley Research Center,
Hampton, Virginia, USA
[3] Los Alamos National Laboratory,
Los Alamos, New Mexico, USA
[4] Air Force Research Laboratory / Phillips Research Site, Kirtland AFB,
New Mexico, USA
[5] Sandia National Laboratory,
Albuquerque, New Mexico, USA

Abstract. The Technical Report Interchange project is a cooperative experimental effort between NASA Langley Research Center, Los Alamos National Laboratory, Air Force Research Laboratory, Sandia National Laboratory and Old Dominion University to allow for the integration of technical reports. This is accomplished using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and having each site cache the metadata from the other participating sites. Each site also implements additional software to ingest the OAI-PMH harvested metadata into their native digital library (DL). This allows the users at each site to see an increased technical report collection through the familiar DL interfaces and take advantage of whatever valued added services are provided by the native DL.

1 Introduction

We present the Technical Report Interchange (TRI) project, which allows integration of technical report digital libraries at NASA Langley Research Center (LaRC), Los Alamos National Laboratory (LANL), Air Force Research Laboratory (AFRL), and Sandia National Laboratory. LaRC, LANL, Sandia and AFRL all have thousands of "unclassified, unlimited" technical reports that have been scanned from paper documents or "born digital". Although these reports frequently cover complementary or collaborative research areas, it has not always been easy for one laboratory to have full access to another laboratory's reports. The laboratories would like to share access to metadata with

links to full text document initially, and eventually replicate the document collections. Each laboratory has its own report publication tracking, management and search/retrieval systems, with varying levels of interoperability with each other. Since the libraries at these laboratories have evolved independently, they differ in the syntax and semantics of the metadata they use. In addition, the database management systems used to implement these libraries are different (Table 1).

Table 1. Native Metadata Formats and Library Systems

Laboratory	Native Metadata Format	Native Library System - Source	Native Library System - Destination
LaRC	MARC	BASIS+	TBD
LANL	USMARC+ Local Fields	Geac ADVANCE	Science Server
AFRL	COSATI	Sirsi STILAS	Sirsi STILAS
Sandia	MARC	Horizon	Verity

One major effort that addresses interoperability started with the Santa Fe Convention [11]. The objective of the Santa Fe Convention, now the Open Archive Initiative (OAI) [4] is to develop a framework to facilitate the discovery of content stored in distributed archives. OAI is becoming widely accepted and many archives are currently or soon-to-be OAI-compliant. While DL interoperability has been well studied in NCSTRL [1], STARTS [2] and other systems, OAI is significantly different in several aspects. Most significantly, OAI promotes interoperability through the concept of metadata harvesting. The OAI framework supports Data Providers (archives or repositories) and Service Providers (harvesters). A typical data provider would be a digital library without any constraints on how it implemented its services with its own set of publishing tools and policies. However, to be part of the OAI framework, a data provider needs to be 'open' in as far as it needs to support the OAI protocol for metadata harvesting (OAI-PMH). Service providers develop value-added services based on the information collected from cooperating archives. These value-added services can take the form of cross-archive search engines, linking systems, and peer-review systems.

OAI-PMH provides a very powerful framework for building union-catalog-type databases for collections of resources by automating and standardizing the collection of contributions from the participating sites, which has traditionally been an operational headache in building and managing union catalogs [7]. By implementing the OAI-PMH, the TRI system enables the sharing of documents housed in disparate digital libraries that have unique interfaces and search capabilities designed for their user communities. This allows a native digital library to export and ingest information from other digital libraries in a manner transparent to its user community. That is, the users access information from other digital libraries through the same native library interface the users are accustomed to
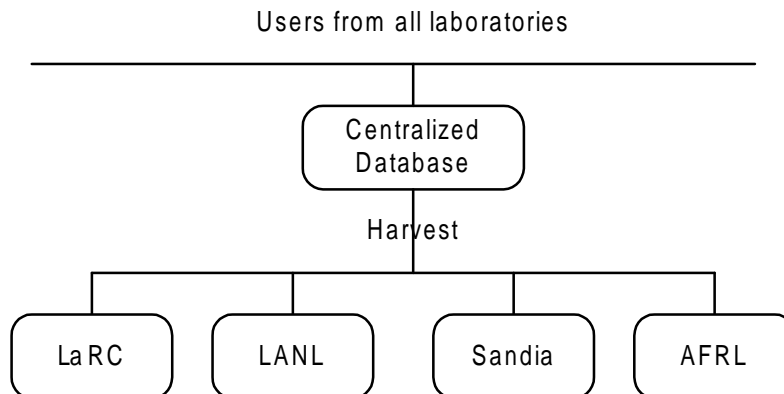
Users from all laboratories

```
                    Centralized
                     Database

                      Harvest

    LaRC        LANL        Sandia        AFRL
```

Fig. 1. Centralized Approach

using. The importance of this approach is that it not only allows for one-time historical sharing of a corpus amongst participating libraries, it also provides for continuous updating of a native library's collection with new documents when other OAI-compliant repositories add to their collections. Additionally, all libraries will always (with some tunable time delay) be consistent in having the totality of all holdings available within their own library.

Based on OAI-PMH, there are two approaches to build a federated digital library that allow users to access reports in all the libraries through a single interface: centralized and replicated. We had to determine which of these approaches would work better for the TRI project. In the centralized approach (Fig. 1), a federation service harvests metadata from the four OAI enabled libraries and provides a unified interface to search all the collections. This approach has been adopted by Arc [5], the first OAI service provider prototype, and other OAI service providers [8] [3] [10]. However, a centralized search service is not a suitable approach for the TRI project given that the primary object of the project is for participating laboratories to provide access to technical reports using their existing library interfaces. Besides this limitation, the centralized approach suffers from the organizational logistics of maintaining a centralized federation service, and having a single point of failure. The TRI system is based on a replicated approach, which addresses these problems (Fig. 2). This approach can be viewed as mirrored OAI repositories, where every laboratory has its own federation service. The consistency between these services is maintained using OAI-PMH. As a federation service is locally available, it becomes easy to push other laboratory's metadata into the native library. In addition, this approach supports several levels of redundancy, thereby improving the availability of the whole system. For example, a failure of a TRI system at one laboratory would not severely impact users at other laboratories. In fact, users at the affected laboratory will continue
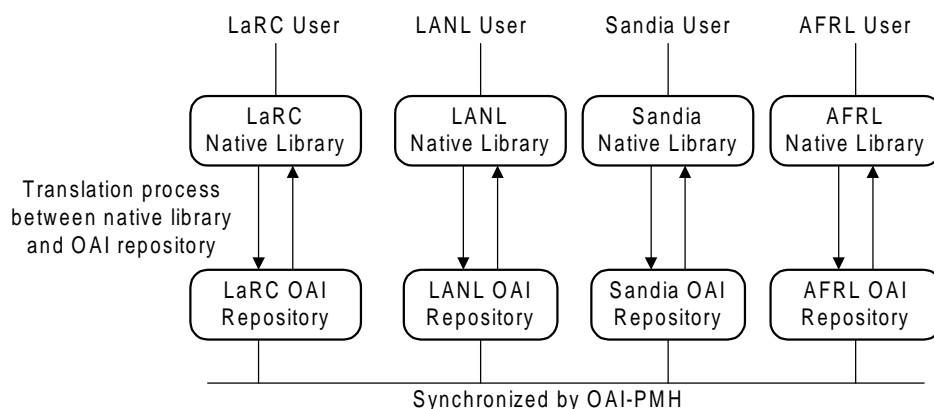
Fig. 2. Replicated Approach

to search and discover reports from other laboratories, though they may not be able to see reports that are added to the system at other laboratories during the down time.

A single node in the TRI system is based on Arc (http://arc.cs.odu.edu) [5], the first OAI service provider that has been in use for nearly two years. While Arc has a built-in infrastructure for OAI harvesting, there are many new challenges in TRI:

Integration with native DL: Since each laboratory has its own DL management system and native search interface, the TRI must be seamlessly integrated into native DL system.

Metadata translation: Because each DL uses different native metadata format, to enable interoperability, we need use a standard metadata format and there must be translation between the native and standard metadata formats.

Seamlessly support new participants: The system must support new participants with limited effort, and any new participant should not adversely impact the existing installations.

Changes progagation: Metadata is duplicated in each DL, so when add, update and delete operations occur in one native library, the changes must be propagated to other libraries.

The rest of the paper is organized as follows. Section 2 presents the architecture of TRI system. In Section 3 we discuss the OAI implementation and common modules across all participating laboratories. Section 4 discusses the issues of integrating TRI system with native library. In Section 5 we discuss records update, deletion and duplicate detection. In Section 6 we analyze the experiences to date and outline future work.
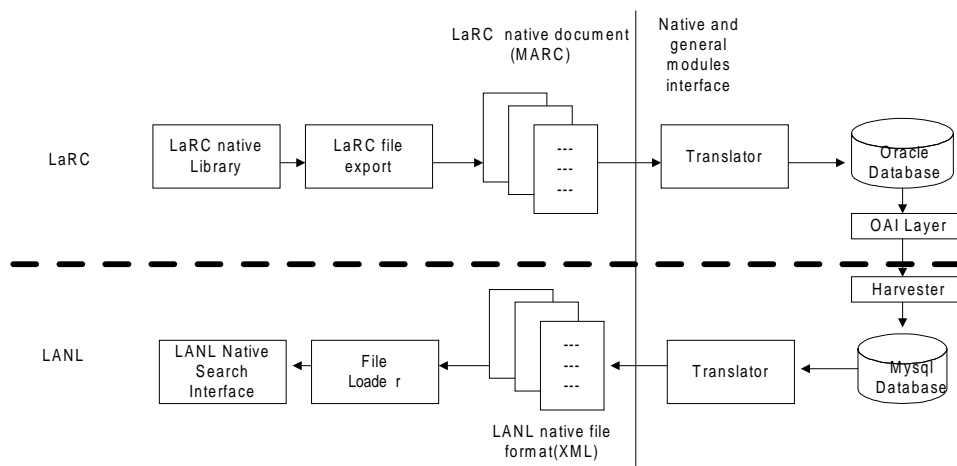
Fig. 3. A Typical Workflow- LANL shares documents from LaRC

2 System Architecture

In the TRI system, each participant has its own user community and a local search interface allowing users to retrieve data from other library systems. A translation process in each DL is responsible for translating native metadata format to a standard metadata format and vice versa, i.e., MARC tags are converted into Dublin Core (DC) [12] and DC into MARC. The standard metadata format is saved in an OAI compliant repository, which can selectively serve metadata when an external OAI harvesting request arrives. A harvester located at each DL periodically harvests metadata from other DLs (Fig. 3).

Since each library has its data format and management system that is maintained by local librarians/information specialists, a file-system based solution is a simple and flexible way for each library to import/export native metadata. The last modification time of records provides a basic mechanism to detect newly added or changed metadata. The exported native metadata is translated into unqualified DC format, which is the default used by OAI to support minimal interoperability. Although richer metadata formats such as MARC or Qualified DC would provide richer semantics and support greater "precision" in search results, the variation in technical report metadata formats (including many unique to a given laboratory) suggested that unqualified DC would be the best metadata format for the initial phase of TRI. As Figure 3 illustrates, the native metadata is converted into OAI-compliant DC, and DC metadata is harvested by other libraries. Once harvested, metadata is converted from DC into local metadata format and stored in an import directory. The local libraries then integrate the newly harvested metadata into their local systems.
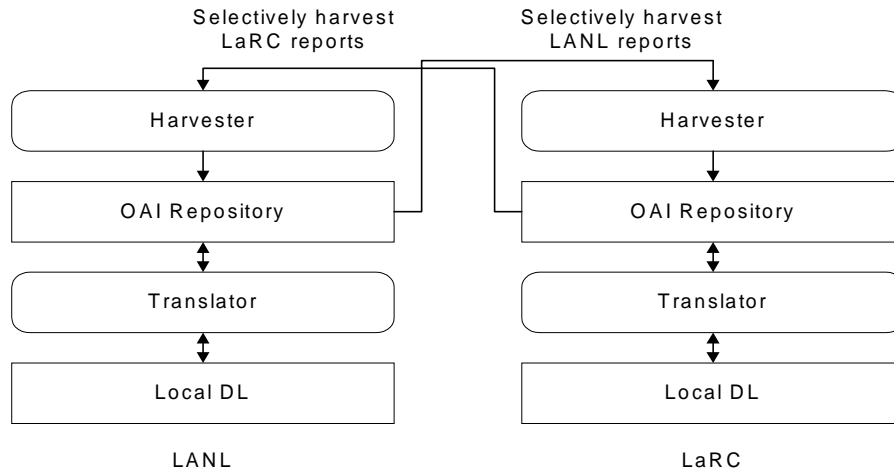
Fig. 4. OAI Repository and Harvester

The developed software is highly modularized and can easily support new participants with minimal effort. The software modules are:

Scheduler: A tool manages and schedules various tasks in TRI system.
OAI repository: A database-based system makes each library OAI-compliant.
Harvester: An application issues OAI request and collects metadata.
Translation tool: Translates native metadata format in each library to a standard metadata and vice versa.

These modules are the same for all repositories. The translation tool requires some customization for a particular library because its local metadata format will need to be mapped into a standard format. This can be accomplished by creating a mapping table between the metadata and the standard.

3 Harvester and OAI Repository

The harvester and OAI repository designs and configurations are based on Arc's implementation design. Arc uses an OAI layer over harvested metadata, making hierarchical harvesting possible. Figure 4 outlines the major components of the system and how they interact with each other.

3.1 Harvester

Similar to a Web crawler, the TRI harvester traverses the data providers automatically and extracts metadata, but it exploits the incremental, selective harvesting defined by the OAI-PMH. Historical and newly published data harvesting have different requirements. When a service provider harvests a data

provider for the first time, all past data (historical data) needs to be harvested, followed by periodic harvesting to keep the data current. To harvest newly published data, data size is not the major problem but the scheduler must be able to harvest new data as soon as possible and guarantee completeness - even if data providers provide incomplete data for the current date, this is implemented by a small overlap between each harvest.

The hierarchical harvesting concept introduced in Arc has a great deal of flexibility in how information is filtered and interconnected between data providers and service providers. In TRI, each repository harvests from other participants and is harvested by other participants. In the case of LaRC, there is a centralized repository harvesting from other NASA OAI-compliant repositories to build up its collection for the TRI project. The structure is also fault-tolerant because complete metadata sets are cached in each library, thereby duplicating data from the original source. If a library system crashes and is no longer accessible, its metadata records reside in other library repositories, thereby ensuring that the records are still available for search, retrieval and serving OAI requests.

3.2 Scheduler and Task Management

The scheduler manages various tasks in the TRI repositories. In each library, there are several typical tasks:

Local read: It makes native DL OAI-compliant and harvestable by other partners;

Remote harvest: It issues requests to OAI compliant repositories;

Local write: It writes harvested records into its local library system.

The scheduler's functions include: automatically launching these tasks, monitoring current status, and addressing network and other system errors. If the harvesting is successful, the scheduler tracks the last harvest time so that the next harvest will start from the most recent harvest.

Each task has its configurable parameters so that the participating laboratories have the flexibility in controlling the system. Tasks can be set up as a historical or fresh process and it allows combining multiple repositories to one single virtual repository (in the case of LaRC). The interval between harvesting is also configurable allowing system administrators to customize how often the data will be harvested: more frequent harvests require additional system resources but provide more current data. However, the whole system works in a coordinated way. For example, a typical working sequence is local read, remote harvest and local write.

The TRI scheduler can be configured as a daemon with its own timer or be controlled by a system timer (e.g. crontab files in Unix). At the initialization stage, it reads the system configuration file, which includes properties such as user-agent name, interval between harvests, data provider URL, and harvesting method. The scheduler periodically checks and starts the appropriate task based on configuration file.

4   Local Repository

While each site shares similar repository and harvester modules, they also have specific DL management systems and native metadata formats. We follow several guidelines in designing the local repository management in TRI system: Each library should maintain its own management system, an identical one is not feasible or possible; Considering the different software/hardware environment in each library, the interface between the native library and TRI system should be portable across platforms and should be simple; The effort to add a new participant should be minimal.

Based on these requirements we defined a file system based interface between native library and TRI general modules (Fig. 3). Each library exports its native format to a configurable directory, and the changed/added document is automatically marked by last modified time. The TRI local reader periodically polls this directory and any file whose modified date is newer than last harvesting time is translated into unqualified DC format and inserted into the OAI repository. Additionally, there is also an import directory in each library; the TRI local writer periodically checks whether any new/changed metadata is harvested from remote repository, translates it into local format and writes it to import directory. Each site may have its own program that exports metadata from local library system, and a loader that reads the import directory. Such a mechanism is highly integrated with a given local repository so its implementation is out of the control of the TRI common modules.

For historical reasons, each digital library may use different metadata formats. While it is possible to implement a one-to-one mapping for each metadata pair, the mapping complexity dramatically increases with the number of participants (n laboratories would require $n(n-1)$ mappings). With a common intermediate metadata format, only $2n$ mappings are necessary. So we chose unqualified DC as the common intermediate metadata format, and mapped each native metadata format to unqualified DC. However, with a common metadata format, the rich metadata element in each library may be lost as the common metadata format is the minimal subset of all libraries. This problem can be alleviated if we adopt a richer common metadata format in the future.

4.1   Mapping Metadata Formats

LANL, LaRC and Sandia use MARC in their local libraries, but each library has its own extensions or profiles. AFRL supports its own metadata format. Each library exports its metadata in its convenient way and also defines a bi-directional mapping table (See samples in Table 2 & 3).

In Table 2, the mapping table follows the structure of Library of Congress's MARC to DC crosswalk [6] with additional features from LaRC. In the MARC to DC mapping, the MARC file is parsed and corresponding fields are mapped to DC; some information may be lost, for example, the identifier field may be an ISSN number, technical report number or URL. Information like ISSN and URL

Table 2. LaRC MARC to DC Mapping(excerpt)

LaRC MARC Metadata Set	Dublin Core
D245a, D245d, D245e, D245n, D245p, D245s	title
D513a, D513b	coverage
D520b	description
D072a,D072b(001), D650a,D659a	subject
D090a(000), D013a, D020a, D088a, D856q, 856w	identifier

Table 3. DC to Sandia Mapping

Dublin Core element	Sandia Metadata Field
identifier	report numbers
identifier – URI	URL
subject	subject category codes
title	title
subject	keywords
creator	personal names
creator	corporate names
date	date
format – extent	extent
description	notes
rights	classification & dissemination

is clearly defined in MARC, but it will map to the undistinguished "identifier" field in unqualified DC, losing the distinctions between metadata fields.

4.2   Subject Mapping

Each library may use a different subject thesaurus and/or classification scheme. For example, LANL uses a combination of Library of Congress Subject Headings (LCSH) and subject terms from other relevant thesauri (including *International Energy: Subject Thesaurus (ETDE/PUB-2)* and its revisions). The metadata for a given LANL technical report may also include numerical subject categories or alpha-numerical report distribution codes representing a broad subject concept. Subject category code sources used by LANL include: *Energy Data Base: Subject Categories and Scope (DOE/TIC-4584-R#)* and its succeeding publication and revisions, *International Energy: Subject Categories and Scope (ETDE/PUB-1)*. Report distribution category code sources include various revisions of *Program Distribution for Unclassified Scientific and Technical Reports: Instructions and Category Scope Notes (DOE/OSTI-4500)*.

LaRC uses its own subject thesaurus and the NASA-SCAN system. The local library may organize the information by subject classification and it is necessary to do a subject classification mapping, for example, mapping NASA subject code (77 Physics of Elementary Particles) to LANL report distribution code (UC-414) (Table 4). Subject metadata is an area where the generically grouping
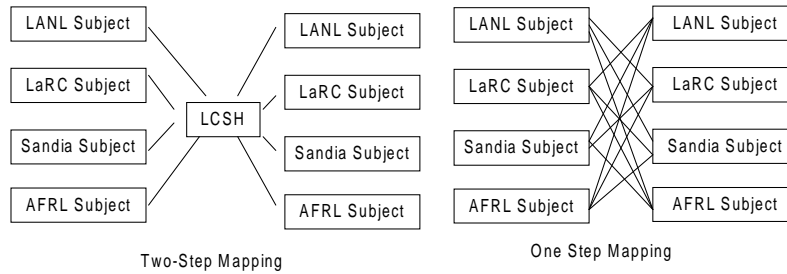
Fig. 5. Subject Mapping (Assume the unique subject schema is LCSH)

Table 4. Subject Mapping: LANL UC-414 maps to NASA SCAN 77

Digital Library	Subject Schema	Sample Subject Format
LANL	UC Report Distro Category ETDE Subject Category INIS Subject Category (old) INIS Subject Category (new) Text (LCSH) Text (other thesauri) Text (local subject heading)	UC-414 sddoeur 430100 edbsc E1610 inissc S43 inissc Controlled formatted text Controlled formatted text Locally controlled text
NASA	SCAN Text	77 PHYSICS ELEMENTARY PARTICLES AND FIELDS

the various subject related metadata into a single unqualified DC data element results in loss of the source information for a given thesaurus or classification scheme thereby complicating the subject metadata mapping.

There are several approaches to address the lack of unified subject access. One way is to use a standard terminology and map each library's controlled metadata to the standard [9]. However, the granularity of subjects/keywords is significantly different among participating libraries; a unified standard is difficult to define and two-step mapping may cause more inconsistencies. Another way is to perform an individual mapping for each subject category pair. This alternative approach is more accurate because only one-step mapping is used. However, both approaches may introduce significant human effort to maintain the relationships (Fig. 5). A third approach is to use an automatic classification algorithm, however, the precision of this mapping is low as we are dealing with limited metadata. The easiest approach is to map all numeric subject codes into text strings using the mapping provided by the contributing organization. We have implemented all the methods except the unified subjects, and we are currently evaluating the different approaches in terms of validity and cost.

4.3   Integration with the local library

The procedure of integrating with the local library is highly dependent on the existing library system. Here we describe the experience in LANL. LANL discussed various options for making TRI metadata available to local library users. One of the first suggestions, importing TRI metadata records from other institutions into the library's online catalog (the original source of exported LANL technical reports metadata) was ultimately rejected due to concerns about data mapping from the "lowest common denominator" Dublin Core format of TRI records to the MARC format required for the online catalog. It was decided to make TRI metadata records available through the library's Science Server software as a proof-of-concept test.

Science Server, a locally modified version of software provided by Science Server LLC, enables simple content management while delivering electronic journals and IEEE Conference and Standards records directly to the desktop. At LANL, Science Server was ultimately selected for integration of and access to TRI records for the following reasons:

1. Provides a unified, familiar search interface to library users;
2. Offers robust indexing and searching capabilities with support for full text links (hyperlinks to technical reports);
3. Permits the definition of "collections" for each harvested site, with appropriate access restrictions for the collections as needed. Since the Science Server product was originally designed for access to journal literature, the "journal paradigm" was adapted for technical reports - with the TRI database becoming one collection within Science Server, each TRI archive institution treated as a "title", individual report years handled as volumes/issues, and the individual reports handled as "articles".

With the above paradigm in mind, it was a simple matter to design a loader for Science Server that mapped the TRI Dublin Core fields into Science Server fields. TRI's configuration tables were updated to perform "local writes", exporting the records from each archive to Dublin Core XML flat-file format. These records were then copied to test version of the Science Server system, converted from DC (loaded) and indexed. At this point, approximately 72000 TRI metadata records are locally searchable through the test Science Server system.

4.4   Security

There are four types of interactions in an OAI based data/service provider framework.

User - Search Service: a user interacts with a service provider, for example an interaction of a search user with a cross-archive search service.

Data Provider - Service Provider: a service provider interacts with a data provider using the OAI-PMH, for example, when a service provider harvests metadata from a data provider.

Publisher - Data Provider: an author publishes a digital object in a data provider, for example, when a researcher submits her pre-print in a pre-print collection.

User- Data Provider: a search user has found metadata record and wants to retrieve an associated object.

One approach to make these interactions secure is the use of Secure Socket Layer (SSL), and the other approach is based on IP address-based restrictions. In the current TRI system, we take the latter approach since it is simpler and is sufficient for the security needs of all partners. Thus, clear text is used for all four types of operations and authentication is provided by checking that a user (or program) comes only from a pre-defined set of acceptable machines. SSL can be adopted in the future if the TRI members wish to exchange more sensitive metadata.

5   Deletion, Update and Duplication Detection

The TRI system is a fully distributed system with redundant data in each participating library; thus changes in one library need be propagated to other libraries. Furthermore, each library integrates data from many different sources, inside or outside of TRI project, which sometimes may lead to the existence of more than one legitimate copies of an article. Therefore we need to consider the duplication detection problem.

5.1   Deletion

Since the local library repository is not controlled by TRI system, the deletion is done in an advisory way. The deletion is initiated by the originating DL, the target TRI database deletes the records during the propagation of the information of action taken, finally an alert mechanism is provided to libraries that have imported the data to their local databases, and the deletion in a local DL is dealt with by its own management system.

The OAI-PMH defines a basic mechanism in dealing with deleted records: a record that is deleted can be indicated by a status of "deleted" in its header. This status means that an item has been deleted and therefore no record can be disseminated from it. This mechanism is integrated with local database management in our implementation. To initiate the document deletion, the local administrators mark a record as "deleted" in their administrative page. This information is kept in local TRI repository and when a remote site starts to harvest from this repository, it notices the "deleted" status based on the mechanism defined by OAI-PMH, and delete the record from its local TRI repository. At the same time the deleted records is marked in its local admin system. The system administrator can find deleted records in local admin page and apply the appropriate operations.

5.2   New and Updated Records

In OAI-PMH, updated or newly added records are identified by a "datestamp", which is defined as the date of creation, deletion, or latest date of modification. Similar to deleted records, updated records need to be integrated with local database management. When a file is changed or added to local export directory, the last modification date of this file is changed too. During each operation, the date of the last harvesting is saved, and it is compared with the date of each file under local export directory. Any file whose last modification time is newer than the last harvested time is imported into the local OAI repository and its datestamp is also changed. Later when a remote repository issues a fresh harvesting request, only the updated and new metadata is returned. This data is written into the import directory and later could be integrated into local search interface.

5.3   Duplication Detection

There are many cases in which duplication may occur. For example, one paper may be co-authored by authors at multiple TRI sites and the report indexed by the respective DLs. Especially in LaRC, there are multiple OAI repositories with overlapping collections. To accommodate each library's policy about duplicate records, the TRI system provides a mechanism that detects possible duplicates by similarity of key metadata fields like title and author. It then alerts the local system administrator of possible duplicate records to verify and delete.

6   Conclusions

In the first stage of the TRI project, LaRC and LANL installed TRI systems and each site has shared approximately 30K technical reports with each other. Both were able to automatically harvest newly published metadata from other site on a daily basis. LANL also loaded the harvested records into its native library, the Science Server, a system external to the TRI project repositories. ODU has finished the AFRL and Sandia translation modules and they will be deployed soon. We are also working on implementation of a user-friendly administrator page for deletion and other system management work.

During the implementation, one of the most significant problems is that unqualified DC does not match well with the sophisticated metadata formats used by the participants. The mappings, especially the subject mapping, is also difficult, and in many circumstances the semantics of original data is lost. This could be partially solved by defining a qualified DC profile for technical reports; however, the standard definition itself is time-consuming and is outside the scope of TRI. We intend to solicit additional participants for TRI after the current round of testing concludes. The initial result of using OAI-PMH as a mechanism for sharing data indicates that OAI-PMH is a flexible and powerful way to automate and standardize metadata exchange.

References

1. Davis, R., and Lagoze, C. (2000) "NCSTRL: Design and Deployment of a Globally Distributed Digital Library". Journal of American Society for Information Science, 51, 273-280.
2. Gravano, L., Chang, K., Garcia-Molina, H., Lagoze, C. and Paepcke, A. (1997) "STARTS:Stanford proposal for internet meta-searching". In Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 207-218
3. Harnad, S. and Carr, L. (2000) "Integrating, Navigating and Analyzing Eprint Archives Through Open Citation Linking (the OpCit Project)". Current Science (special issue honour of Eugene Garfield) 79, 629-638. http://cogprints.soton.ac.uk/documents/disk0/00/00/16/97/
4. Lagoze, C. and Van de Sompel, H. (2001) "The Open Archives Initiative: Building a low-barrier interoperability framework". In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Roanoke VA, June 24-28, 2001, pp. 54-62.
5. Liu, X., Maly, K., Zubair, M. and Nelson, M. L. (2001) "Arc - An OAI Service Provider for Digital Library Federation". D-Lib Magazine, 7(4). http://www.dlib.org/dlib/april01/liu/04liu.html
6. LOC (2001) "MARC to Dublin Core Crosswalk". Network Development and MARC Standards Office, Library of Congress. http://www.loc.gov/marc/marc2dc.html
7. Lynch, C. (2001) "Metadata Harvesting and the Open Archives Initiative". ARL Monthly Report 217, August 2001. http://www.arl.org/newsltr/217/mhp.html
8. McClelland, M., McArthur, D., Giersch, S. and Geisler G. (2002) "Challenges for Service Providers When Importing Metadata in Digital Libraries". D-Lib Magazine, 8(4). http://www.dlib.org/dlib/april02/mcclelland/04mcclelland.html
9. Koch, T., Neuroth, H. and Day, M. (2001) "Renardus: Cross-browsing European subject gateways via a common classification system (DDC)". IFLA satellite meeting: Subject Retrieval in a Networked Environment, OCLC, Dublin, Ohio, USA
10. Suleman, H. and Fox, E. A. (2001) "A Framework for Building Open Digital Libraries". D-Lib Magazine, 7(12). http://www.dlib.org/dlib/december01/suleman/12suleman.html
11. Van de Sompel, H. and Lagoze, C. (2000) "The Santa Fe Convention of the Open Archives Initiative". D-Lib Magazine, 6(2). http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html
12. Weibel, S., Kunze, J., Lagoze, C. and Wolfe, M. (1998) "Dublin Core metadata for resource discovery". Internet RFC-2413, ftp://ftp.isi.edu/in-notes/rfc2413.txt