OAI and NASA's
Scientific and Technical Information

Michael L. Nelson
Old Dominion University
Norfolk VA, 23529
mln@cs.odu.edu

JoAnne Rocker
NASA Langley Research Center
MS 157A
Hampton VA, 23681
j.rocker@larc.nasa.gov

Terry L. Harrison
Old Dominion University
Norfolk VA, 23529
tharriso@cs.odu.edu

Abstract

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is an evolving protocol and philosophy regarding interoperability for digital libraries (DLs). Previously, "distributed searching" models were popular for DL interoperability. However, experience has shown distributed searching systems across large numbers of DLs to be difficult to maintain in an Internet environment. The OAI-PMH is a move away from distributed searching, focusing on the arguably simpler model of "metadata harvesting". We detail NASA's involvement in defining and testing the OAI-PMH and experience to date with adapting existing NASA distributed searching DLs (such as the NASA Technical Report Server) to use the OAI-PMH and metadata harvesting. We discuss some of the entirely new DL projects that the OAI-PMH has made possible, such as the Technical Report Interchange project. We explain the strategic importance of the OAI-PMH to the mission of NASA's Scientific and Technical Information Program.

Key Words: OAI-PMH, Metadata Harvesting, NASA, Scientific and Technical Information

Author Biographies

Michael L. Nelson received his B.S. (computer science) from Virginia Tech (1991) and his M.S. and Ph.D. (computer science) from Old Dominion University (1997, 2000). He worked at NASA Langley Research Center from 1991 – 2002 and spent the academic year of 2000-2001 on a NASA fellowship at the University of North Carolina at Chapel Hill. He joined the computer science department of Old Dominion University during the summer of 2002.

JoAnne Rocker received her B.A. (international studies) (1990) and M.L.S. (1995) from the University of North Carolina at Chapel Hill. She worked at the NASA Langley Research Center Technical Library from 1993 – 2000, and from 2000 to the present she has worked in the NASA STI Program Office. Her tasks there include the coordination of NASA's digital library university research projects.

Terry L. Harrison received his B.A. (mass communications) from James Madison University (1992). Prior to becoming a graduate student in the computer science department of Old Dominion University, he worked as a technology instructor for XMLSolutions. He worked during the summer of 2002 for the NASA STI Program Office.

OAI and NASA's
Scientific and Technical Information

Abstract

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is an evolving protocol and philosophy regarding interoperability for digital libraries (DLs). Previously, "distributed searching" models were popular for DL interoperability. However, experience has shown distributed searching systems across large numbers of DLs to be difficult to maintain in an Internet environment. The OAI-PMH is a move away from distributed searching, focusing on the arguably simpler model of "metadata harvesting". We detail NASA's involvement in defining and testing the OAI-PMH and experience to date with adapting existing NASA distributed searching DLs (such as the NASA Technical Report Server) to use the OAI-PMH and metadata harvesting. We discuss some of the entirely new DL projects that the OAI-PMH has made possible, such as the Technical Report Interchange project. We explain the strategic importance of the OAI-PMH to the mission of NASA's Scientific and Technical Information Program.

Introduction

The Open Archives Initiative (OAI) is an international consortium focused on furthering the interoperability of digital libraries (DLs) through the use of "metadata harvesting". The metadata harvesting approach is defined in the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Lagoze, Van de Sompel, Nelson & Warner, 2002). Metadata harvesting is in contrast to the "distributed searching" approach of many previous DL interoperability projects for federating different DLs into a single service. While feasible for small numbers of nodes (e.g., < 20), large-scale distributed searching has proven difficult in an Internet environment for large numbers of nodes (e.g., > 100) (Nelson, 2001).

Key to the philosophy of the OAI is the separation of responsibilities into "service providers" and "data providers". In practice, a DL act as both a data provider and a service provider, so it is important to understand the distinction. A data provider is simply a repository, a collection of metadata records.   A service provider offers value added services (e.g., searching, browsing) on the metadata extracted from one or more data providers. More detailed reviews of the OAI-PMH, its origins, version history and applications can be found in (Van de Sompel & Lagoze, 2000; Lagoze & Van de Sompel, 2001; Van de Sompel & Lagoze, 2002; Lynch, 2001).

NASA was an original member of the OAI and has reengineered many of its DLs to capitalize on the popularity of the OAI-PMH. This includes the addition of data providers to existing DLs as well as the creation of new DLs as service providers that harvest from NASA and non-NASA data providers. The NASA Scientific and Technical Information Program supports NASA's active role in DL research and development.

NASA Scientific and Technical Information Program

The NASA Scientific and Technical Information (STI) Program has existed since the early days of the National Aeronautics and Space Administration. Its purpose, which derives from the National Aeronautics and Space Act of 1958, is "…to provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

The STI Program ensures that NASA remains at the leading edge of research and development (R&D) by quickly and efficiently capturing worldwide STI for use in problem solving, awareness, and knowledge transfer. Its data collection and dissemination supports NASA's mission to communicate scientific knowledge and understanding by collecting and transferring NASA's R&D to the aerospace and academic communities. The program collects the NASA-produced information from 10 NASA Centers and Headquarters, other sources in the U.S. (such as academia, professional societies and other government and military research facilities), and over 50 foreign countries, and maintains access to the largest collection of aerospace science and technical information in the world. An analysis and review of the STI Program's activities can be found in Pinelli (1990), but some STI Program activity highlights are to:

> Collect, announce, disseminate, and archive all STI resulting from NASA-funded and sponsored research to reduce duplication of effort and improve productivity and cost-effectiveness of the NASA research effort
> Acquire other domestic and international STI pertinent to NASA's missions
> Handle and publish all appropriately reviewed STI for NASA, thus requiring close coordination with export control, patent, copyright, and intellectual property organizations, and international partnerships
> Build and maintain the STI Database
> Coordinate the Agency's various Field Center STI programs
> Develop and implement all Agency policy and procedures for external release of STI
> Monitor a contractor facility, the NASA Center for AeroSpace Information (CASI) in Hanover, Maryland
> Negotiate and handle international STI agreements for NASA

Over the years, the STI Program Office has faced challenges in its task to acquire and promote information usage. Advances in information technology have fueled higher expectations and increased demand for easy and efficient access to scientific information. Internet savvy clientele are used to sophisticated search engines that retrieve documents, photos, music, and other kinds of data. These users now expect services from information providers such as:

> Desktop access to full-text documents and not just abstracts
> Rapid access to documents to meet customer requirements
> Wider access to varied information formats
> Preprints and other forms of gray literature not published in traditional forms

(e.g., photos, videos, and graphics)
High-speed Internet access and web-based architecture
Better data organization and user friendly interfaces

New challenges to information delivery are not the only concerns of the STI Program Office. The high cost of supplementing the NASA-produced information with commercial data in order to broaden the access to scientific and technical research has limited the kinds of information that the program can make available. Acquisition of commercial data is expensive and restrictive licensing hinders efforts by the STI Program Office to enhance its aerospace knowledge base. As the STI Program continues to face funding challenges, finding alternatives to commercially available information is necessary to ensure that NASA DL users have the information they need. The emergence of the OAI as a technology bridge connecting heterogeneous data sources offers the STI Program a way to build its collection circumventing the problems associated with commercial data. Further, as more information providers start using the OAI-PMH for data exchange, the breadth and scope of information available to the STI Program will grow.

NASA DL History

NASA's history with web-based DLs dates back to 1993, when a WWW interface was provided for the Langley Technical Report Server (LTRS) (Nelson & Gottlich, 1994; Nelson, Gottlich & Bianco, 1994; Roper et al., 1994) (1). Prior to this, LTRS was simply an anonymous FTP server that distributed technical reports authored and sponsored by NASA Langley Research Center. However, LTRS provided access to only reports from NASA Langley Research Center, and not other NASA centers and institutes. Beginning in 1994, software used to create LTRS was shared with other NASA installations and "LTRS-like" DLs were setup. However, it was not until 1995 that the NASA Technical Report Server (NTRS) (2) was setup to provide integrated searching between the various NASA web-based DLs (Nelson et al., 1995). Since then NTRS has proven extremely successful with the general public and averages approximately 30,000 search sessions per month. The total collection accessible through NTRS is approximately 4.5M abstracts and 300K full-text publications.

[TAKE IN FIGURE 1]

The architecture for this version of NTRS is shown in Figure 1, and a screen capture of the user interface is shown in Figure 2. Integration of the 20 separate nodes comprising NTRS was done through distributed searching. A single Perl script generates the NTRS interface, processes and dispatches queries to a daemon specifically assigned to each NTRS node. The main script gathers the results from each of the daemons and presents the results to the user. No attempt is made to integrate the results from each node in any meaningful manner; the architecture of the institutional nodes was exposed to the user. Although the searching is done in parallel, generating the results will not run faster than the slowest component. If for some reason a component is not reachable, the search will be held up waiting for that search request to timeout. Thus for performance reasons, user

queries are not broadcast to all 20 nodes by default, but rather only to six of the more popular nodes. 17 of the nodes support different versions of WAIS and the three nodes from the Astrophysics Data Service utilize a non-WAIS version of Z39.50 (Kurtz et al., 1999). Popular in the early days of the WWW, WAIS (Kahle et al., 1992) is an Internet based subset of the venerable Z39.50 standard for distributed information retrieval (Lynch, 1997).

[TAKE IN FIGURE 2]

While the current version of NTRS is popular and useful, its distributed searching architecture imposes many limitations. Since NTRS does not contain any holdings of it own, it is dependent on the quality and availability of its constituent nodes. As has been reported for similar distributed searching DLs, the availability of all nodes can be poor (Powell & French, 2000). Furthermore, syntactical differences between the various search engines limits the complexity of the searches that can be specified. While NTRS currently does some search syntax normalization, the mappings provided are less than complete.

Adopting the OAI-PMH to Existing NASA DLs

NASA was one of the original participants in the OAI. For the Universal Preprint Service (UPS) demonstration project that lead to the creation of the OAI (Van de Sompel et al., 2000), NASA contributed the metadata from the National Advisory Committee for Aeronautics (NACA) digital library. The NACA digital library is a retrospective digitization project that focuses on the capture and preservation of the reports authored by NASA's predecessor organization during 1917-1958 (Nelson, 1999) (3). In late 2001, both LTRS and NACA had OAI-PMH interfaces available for harvesting. This allowed the content in both collections to be indexed by a variety of OAI service providers While LTRS and NACA had previously been crawled by web crawlers and thus had their contents available from general search engines (e.g., Google, AltaVista, AskJeeves, etc.), the OAI-PMH interface was the first time that these DLs were truly interoperable with generalized services.

Adding OAI-PMH interfaces to the existing DLs took only a few days of programming for each interface. In fact, after a generalized approach was developed, several OAI-PMH interfaces for other nodes in NTRS were developed by the authors and distributed in less than a day. The approach used in these implementations was to add a modified "bucket" to the DLs. Buckets are object-oriented, intelligent web storage units that contain metadata, data and methods to operate on the metadata and date (Nelson & Maly, 2001). Buckets already implement approximately 30 methods (or "verbs" in OAI-PMH parlance), so adding the six OAI-PMH verbs as bucket methods was easily accomplished. Buckets were originally built to hold full-text content in DLs, but they have proved useful in other scenarios as well, including building OAI-PMH repositories. Figure 3 shows the architecture of the modified bucket. The six OAI-PMH verbs are added to the "package" that contains the other bucket methods. These OAI-PMH verbs are implemented generically; they depend on a library file that maps the general implementations of the

OAI-PMH verbs to the explicit DL. For example, it is in this file where native metadata formats are mapped to Dublin Core, rules for constructing identifiers are kept, and the appropriate filesystem or SQL calls are defined to extract the necessary metadata. Table 1 is a list of all the OAI-PMH repositories built using modified buckets. The Open Video Project (Slaughter, Marchionini & Geisler, 2000) is not a NASA project, but they have adopted a bucket-based OAI-PMH interface and it is included in the table for comparison purposes.

[TAKE IN FIGURE 3]

Name	URL	Notes
Langley Technical Report Server	http://techreports.larc.nasa.gov/ltrs/oai/ (1.1) http://techreports.larc.nasa.gov/ltrs/oai2.0/ (2.0)	Metadata in refer format; stored on a filesystem
NACA	http://naca.larc.nasa.gov/oai/ (1.1) http://naca.larc.nasa.gov/oai2.0/ (2.0)	Metadata in refer format; stored on a filesystem
Johnson Space Center Technical Report Server	http://ston.jsc.nasa.gov/JSCTRS/oai/ (1.1)	Metadata in a COSATI derivative; stored in a MS Access database dump
Open Video Project	http://www.open-video.org/oai/ (1.1) http://www.open-video.org/oai2.0/ (2.0)	Metadata in a locally defined format; stored in MySQL

Table 1. OAI-PMH Interfaces Built With Buckets

LTRS and NACA currently see a great deal of OAI-PMH harvesting activity. Some of this is surely due to spikes of harvesting and testing as new SPs become active. However, the breadth of activity (over 20 distinct harvesters have visited LTRS and NACA) indicates significant acceptance of OAI-PMH interfaces for these DLs. These DLs still see the same amount of traffic from interactive users and regular web crawlers, but the OAI-PMH capabilities result in even more exposure for their contents.

An OAI-PMH Version of NTRS

OAI-PMH use in NASA DLs has not been limited to retrofitting existing DLs. New services are being developed and existing services recreated using OAI-PMH as a core technology. One of the DLs being recreated using OAI-PMH is NTRS, which is in the final stages of internal testing at the time of this writing (4). Many of the limitations of the current distributed searching version of NTRS are being addressed by the OAI metadata harvesting version of NTRS. While the scale of the new NTRS is much smaller (currently only about 12K full text documents), it is expected to grow to be equivalent with the earlier, distributed searching version of NTRS as more NASA centers and institutes become OAI-PMH capable. For example, the NASA Jet Propulsion Laboratory (JPL) is not currently a participant in the OAI-PMH version of NTRS, but will be in the near future. At the moment, JPL is re-engineering its knowledge management processes and using OAI-PMH as a core technology for this purpose. When this project is complete, an OAI-PMH interface for its technical publications (suitable for public use in NTRS) will be a small subset of JPL's institutional OAI-PMH deployment. Other

centers, such as the NASA Marshall Space Flight Center (MSFC) are planning to replace their current DL infrastructure with off-the-shelf OAI-PMH capable DLs such as eprints.org (5).

The new NTRS offers many advantages that the earlier, distributed searching NTRS does not. For one, all the contents of NTRS are now searched by default. In the previous version of NTRS, not all nodes were searched by default because searching all nodes is "expensive" in terms of network resources. Furthermore, many of the nodes are highly specialized and are less likely to produce interesting results for the general query. In an attempt to limit the penalty of network access to nodes that are not likely to produce interesting results, only 6 of the 20 nodes selected by default. Of course, depending on knowledgeable users to correctly target their searches to the right nodes is an unrealistic assumption made by the previous NTRS version. However, this is not an issue in the new version of NTRS because metadata harvesting means maintaining a copy of metadata harvested in batch mode and not having to dynamically search each node for every query. This results in faster, more reliable searches for users. NTRS currently does not cache copies of full-text documents, so full-text document availability is still subject to transient network errors. The new version of NTRS provides both a simple interface (Figure 4) and an advanced search interface (Figure 5) that allows more targeted searching, including limiting the number of repositories to search. Syntactic differences between the 20 nodes of the previous version of NTRS made it infeasible to offer anything beyond just a simple search interface.

[TAKE IN FIGURE 4]

[TAKE IN FIGURE 5]

One interesting feature now offered by the OAI-PMH version of NTRS is that of "hierarchical harvesting" of the kind first introduced by Arc (Liu, Maly, Zubair & Nelson, 2001). Hierarchical harvesting is gaining popularity as OAI-PMH aggregators harvest metadata from data providers and re-expose the metadata to other harvesters, perhaps with some metadata normalization or cleansing. This means that NTRS can re-export its contents and act as an aggregator of public NASA STI. For example, the Elsevier sponsored OAI SP, Scirus, currently harvests both LTRS and NACA as separate repositories, even though they belong to a larger logical grouping. However, once the OAI-PMH version of NTRS is online, Scirus will be able to harvest a single NASA repository and have the entirety of the harvestable public NASA STI. Similarly, we are in discussions with the science.gov effort and hope to demonstrate the feasibility of an OAI-PMH hierarchical harvesting approach for their project. Hierarchical harvesting is a powerful capability that will make the OAI-PMH flexible enough to allow DLs to adapt to organizational structures. Guidelines for hierarchical harvest (known as "aggregators" in OAI-PMH parlance) can be found at (Lagoze, Van de Sompel, Nelson & Warner, 2002).

The new NTRS is itself a modified bucket, much like an extension of the OAI-PMH interfaces for LTRS and NACA presented above. The NTRS bucket has the standard

bucket methods, the OAI-PMH verbs (for hierarchical harvesting), and a set of NTRS specific verbs to correspond to simple and advanced search interfaces, the actual searching functions, etc. Figure 6 illustrates the architecture of the NTRS bucket. The metadata for NTRS is harvested using the harvester developed as part of the Open Digital Library Project at Virginia Tech (Suleman & Fox, 2001). The metadata is stored on a file system in its harvested form, then a locally developed Perl script loads the metadata into a MySQL database. Currently, only Dublin Core metadata (Weibel & Lagoze, 1997) is harvested from the participating nodes.

[TAKE IN FIGURE 6]

Technical Report Interchange (TRI) Project

An example of a new DL project made possible through the OAI-PMH is the Technical Report Interchange (TRI) project joint with NASA Langley Research Center, the Air Force Research Laboratory (AFRL), Sandia National Laboratory and Los Alamos National Laboratory (LANL). The purpose of the TRI project is to allow a targeted metadata harvesting between and only between known participants. The goal is to allow each site to maintain its existing DL infrastructure, but also allow each site to export its holdings in the standard OAI-PMH fashion as well as import from the other sites into the native DL. This is different from most OAI-PMH applications to date in that while existing data providers are given OAI-PMH interfaces, most of the SPs represent entirely new DLs or search interfaces. Table 2 (from Liu et al., 2002) shows the established systems and formats at each of the laboratories. COSATI is the metadata format defined by the Committee on Scientific and Technical Information.

Laboratory	Native Metadata Format	Native Source Commercial DL System	Native Destination Commercial DL System
LaRC	MARC	BASIS+	(TBD)
LANL	MARC + local fields	Geac ADVANCE	Science Server
AFRL	COSATI	Sirsi STILAS	Sirsi STILAS
Sandia	MARC	Horizon	Verity

Table 2. Native DL Systems and Formats

Given the heterogeneity, cost and inertia of the installed base(s), it is clear that each laboratory is not going to scrap their existing DL and standardize on a new DL system. Thus, the TRI participants setup cooperating OAI-PMH caches at each of the sites and do a complete harvest of all other TRI participants. A TRI package, which can work with either Oracle or MySQL, is distributed to each site and includes the following:

     An OAI-compliant repository (Java source code)
     Code to convert organizational metadata from its native format to Dublin Core
     Code to convert Dublin Core metadata back into native metadata format
     Harvesting code

Harvesting scheduler code
Log files to track when a site harvests or has been harvested

The TRI system is in the early stages of deployment at the time of this writing. However, it demonstrates the ability to use the OAI-PMH in environments other than the broad "communication with strangers" paradigm. The TRI project can be best described as "communication between friends." This is the first stage in community building, and the standards, formats and conventions that arise from community building are the keys to building high quality DLs.

OAI-PMH Enhanced Capabilities

OAI-PMH allows certain activities to be done more easily. For example, the NASA STI Program Office and the MAGiC grey literature project in the UK (Sidwell, Needham &Harrington, 2000) mirror each other's content (6). MAGiC makes a mirror of the metadata and full-text documents of NACA, and we do the same for their Aeronautic Research Committee (ARC) reports (ARC report series is UK equivalent to NACA report series). Although the OAI-PMH does not explicitly address the harvesting of full text documents, we learn of reports becoming available online (as each organization scans its back catalog) as they appear in their respective OAI-PMH interfaces. At that point, each organization simply caches a copy of the full-text document listed in the DC "identifier" field. For the ARC reports, the PDFs are exposed explicitly. For the NACA reports, the PDFs are not exposed explicitly, but there is a simple heuristic for extracting the direct URL of the NACA PDF. While this approach of using OAI-PMH has a chance of missing updates (updates to the PDFs not reflected in the metadata updates will be missed), this is considered to be a rare situation in practice and the associated risk to be worth the ease of mirroring with OAI-PMH.

Another interesting upgraded capability afforded through OAI-PMH deals with web crawlers. Although we are unaware of any general web crawlers that issue OAI-PMH harvesting requests, there is a gateway mechanism to assist general web crawlers in extracting the "deep" or "hidden" web residing in OAI repositories. DP9 is a gateway between general web crawlers and OAI repositories (Liu, Maly, Zubair & Nelson, 2002) (8). Crawlers starting at DP9 will see a succession of HTML pages (created with the `ListIdentifiers` verb) that point to individual records (through a `GetRecord` verb). The pages are rendered in minimal HTML from their XML source for the web crawler's benefit. `resumptionTokens` perform a type of pagination for the web crawler, allowing it to step through the repository n records at a time. People use general search engines such as Google for a variety of needs: one study of the popular computer science DL ResearchIndex found that only 6% of the sessions with ResearchIndex began from the ResearchIndex interface itself (Mahoui & Cunningham, 2001).

Future Work

Obviously, increasing the number of NASA OAI repositories as well as increasing the content available in the existing repositories is the first order of business. OAI-PMH is

driven by critical mass, and increased breadth and depth is necessary. However, beyond simply increasing participation, there are a number of interesting projects that the OAI-PMH is making possible that have not yet been undertaken. Specifically, the full power and expressiveness that OAI-PMH 2.0 offers through various optional features is not fully exploited in the current NASA DLs. Part of the issue is waiting for various harvester implementations to take advantage of these features as well. For example, although we do employ the "friends" capability to allow NASA repositories to "point" to other affiliated repositories, we are not aware of any harvesters that utilize this. As the number of data providers grows beyond what centralized registries can handle, we expect that the friends container to become increasingly useful in identifying pockets of affiliated DLs.

Some new to 2.0 capabilities that we do not yet utilize include response compression, which allows far more efficient communication between the harvester and repository by compressing the wordy and often redundant text of XML and bibliographic formats. Another area involves the optional "branding" containers, which allow a repository to recommend to harvesters how to render the harvested XML, and whether or not to attach a logo or text to the link to identify its "brand." We expect this to be useful; with Scirus (8) and OAI-PMH 1.1 we had to have these discussions via email. The branding containers will automate this negotiation.

The current NASA repositories only support DC. While unqualified DC is useful for baseline interoperability, it is not rich enough to express many of the concepts that are relevant to NASA publications. We did not initially focus on exporting MARC or COSATI because we did not believe there was enough interest in these formats to warrant a higher priority. However, this is no longer the case and we must address exporting semantically richer metadata formats.

Conclusions

The NASA STI Program has the responsibility for the collection, processing, dissemination, and stewardship of NASA scientific and technical information. Since 1993, Web-based digital libraries have been an integral part of this mission. However, as useful as these DLs are, the ability to interoperate with other U.S. governmental institutions, universities and other partners has always been limited to custom, one-off solutions. The OAI-PMH offers a simple, general interoperability framework for DLs and other Web-based services. The OAI-PMH has been added to existing DLs, such as LTRS and NACA, and has been the catalyst for re-implementing DLs such as NTRS. Not only does the new NTRS move away from the distributed searching model in favor of the metadata harvesting model, but NTRS creates new capabilities, such as providing a single point of harvesting NASA STI for other commercial and governmental projects.

Furthermore, we are just reaching the stage of implementing new projects with the capability of OAI-PMH. This includes the TRI project, which use OAI caches to synchronize the holdings of native DLs. Also, we are using the OAI-PMH/Web gateway,

DP9, to make our collections available to general web crawlers, and are beginning to use the OAI-PMH as the core technology for mirroring agreements.

The OAI-PMH represents a significant development in DL interoperability. The move from the spiraling complexity of distributed searching models to the simplicity of a metadata harvesting model has created a flurry of DL development. The number of OAI-PMH enabled DLs, harvesters, metadata converters and similar tools continues to grow at a monthly rate. The OAI-PMH is just reaching the point of critical mass, both in terms of contents and repositories, and also in OAI-PMH capable staff and tools. Soon, OAI-PMH will be just another core DL technology such as http or XML.

References

Kahle, B., Morris, H., Davis, F., Tienne, K., Hart, C. & Palmer, R. (1992). Wide Area Information Servers: An Executive Information System for Unstructured Files. Electronic Networking: Research, Applications and Policy, 2(1), 59-68.

Kurtz, M. J., Eichorn, G., Accomazzi, A., Grant, C. S., Demleitner, M. & Murray, S. S. (1999). The NASA ADS abstract service and the distributed astronomy digital library. D-Lib Magazine, 5(11). Available at http://www.dlib.org/dlib/november99/11kurtz.html.

Lagoze, C. L. & Van de Sompel, H. (2001). The Open Archives Initiative: Building a low barrier interoperability framework. Proceedings of the 1st ACM/IEEE Joint Conference on Digital Libraries, pp. 54-62. Available at http://www.openarchives.org/documents/oai.pdf

Lagoze, C. L., Van de Sompel, H., Nelson, M. L. & Warner, S. (2002). Open Archives Initiative Protocol for Metadata Harvesting – Version 2.0. Available at http://www.openarchives.org/OAI/openarchivesprotocol.html

Lagoze, C. L., Van de Sompel, H., Nelson, M. L. & Warner, S. (2002). Open Archives Initiative Protocol for Metadata Harvesting Implementation Guidelines. Available at http://www.openarchives.org/OAI/2.0/guidelines.htm

Liu, X., Maly, K., Zubair, M. & Nelson, M. L. (2001) Arc: an OAI service provider for digital library federation. D-Lib Magazine, 7(4). Available at http://www.dlib.org/dlib/april01/liu/04liu.html.

Liu, X. Maly, K., Zubair, M. Tang, R., Padshah, M., Roncaglia, R., Rocker, J., Nelson, M., von Ofenheim, W., Luce, R., Stack, J., Knudson, F., Goldsmith, B., Holtkamp, I., Blake, M., Carter, J., Di Giacomo, M., Nutter, J., Brown, S., Montbrand, R., Landenberger, S., Pierson, K., Duran, V. & Moser, B. (2002). Technical report interchange through synchronized OAI caches. Proceedings of the 6th European Conference on Digital Libraries, pp. 176-189.

Lynch, C. A. (2001). Metadata harvesting and the Open Archives Initiative. ARL Bimonthly Report 217, 2001. Available at http://www.arl.org/newsltr/217/mhp.html

Lynch, C. A. (1997). The Z39.50 Information Retrieval Standard. D-Lib Magazine, 3(4). Available at http://www.dlib.org/dlib/april97/04lynch.html

Mahoui, M. & Cunningham, S. (2001). Search behavior in a research-oriented digital library. Proceedings of the 5th European Conference on Digital Libraries, pp. 13-24.

Sidwell, C. A., Needham, P. A. S. & Harrington, J. D. (2000). Lightening grey literature: Making the invisible visible. New Review of Information Networking, 6, 121-136.

Nelson, M. L. & Gottlich, G. L. (1994). Electronic document distribution: Design of the anonymous FTP Langley Technical Report Server. NASA TM-4567. Available at http://techreports.larc.nasa.gov/ltrs/PDF/tm4567.pdf.

Nelson, M. L., Gottlich, G. L., & Bianco, D. J. (1994). World Wide Web implementation of the Langley technical report server. NASA TM-109162. Available at http://techreports.larc.nasa.gov/ltrs/PDF/tm109162.pdf.

Nelson, M. L., Gottlich, G. L., Bianco, D. J., Paulson, S. S., Binkley, R. L., Kellogg, Y. D., Beaumont, C. J., Schmunk, R. B., Kurtz, M. J., Accomazzi, A., & Syed, O. (1995). The NASA technical report server. Internet Research: Electronic Network Applications and Policy, 5(2), 25-36. Available at http://techreports.larc.nasa.gov/ltrs/papers/NASA-95-ir-p25/NASA-95-ir-p25.html.

Nelson, M. L. (1999). A digital library for the National Advisory Committee for Aeronautics. NASA/TM-1999-209127. Available at http://techreports.larc.nasa.gov/ltrs/PDF/1999/tm/NASA-99-tm209127.pdf.

Nelson, M. L. & Maly, K. (2001). Buckets: Smart objects for digital libraries. Communications of the ACM, 44(5), 60-62.

Nelson, M. L. (2001). Better interoperability through the Open Archives Initiative, New Review of Information Networking, 7, 133-145.

Pinelli, T. E. (1990). Introduction to National Aeronautics and Space Administration's scientific and technical information program. Government Information Quarterly 7(2), 123-126.

Powell, A. L. & French, J. C. (2000) Growth and server availability for the NCSTRL digital library. Proceedings of the 5th ACM Conference on Digital Libraries, pp. 264-265.

Roper, D. G., McCaskill, M. K., Holland, S. D., Walsh, J. L., Nelson, M. L., Adkins, S. L., Ambur, M. Y., & Campbell, B. A. (1994). A strategy for electronic dissemination of NASA Langley publications. NASA TM-109172. Available at http://techreports.larc.nasa.gov/ltrs/PDF/tm109172.pdf.

Slaughter, L., Marchionini, G. & Geisler, G. (2000). Open Video: A framework for a test collection. Journal of Network and Computer Applicatons, 23(3), 219-245.

Suleman, H. & Fox, E. A. (2001). A framework for building open digital libraries. D-Lib Magazine, 7(12). Available at http://www.dlib.org/dlib/december01/suleman/12suleman.html.

Van de Sompel, H., Krichel, T., Nelson, M. L., Hochstenbach, P., Lyapunov, V. M., Maly, K., Zubair, M., Kholief, M., Liu, X. & O' Connell, H. (2000). The UPS prototype: an experimental end-user service across e-print archives. D-Lib Magazine, 6(2). Available at http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html.

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

Van de Sompel, H. & Lagoze, C. (2002). Notes from the interoperability front: A progress report on the Open Archives Initiative. Proceedings of the 6th European Conference on Digital Libraries, pp. 144-157. Available at http://www.openarchives.org/documents/ecdl-oai.pdf

Weibel, S. & Lagoze, C. (1997). An Element Set to Support Resource Discovery - The State of the Dublin Core. International Journal on Digital Libraries. 1(2), 176-186.

Endnotes

1. Langley Technical Report Server, http://techreports.larc.nasa.gov/ltrs/

2. NASA Technical Report Server (old), http://techreports.larc.nasa.gov/cgi-bin/NTRS
3. NACA, http://naca.larc.nasa.gov/
4. NASA Technical Report Server (new), http://ntrs.nasa.gov/
5. DP9, http://dlib.cs.odu.edu/dp9/
6. Eprints.org, http://www.eprints.org
7. MAGiC, http://www.magic.ac.uk/
8. Scirus, http://www.scirus.com/

User

NTRS

broadcast query to all (selected) nodes

LTRS   ATRS   ....   NACA
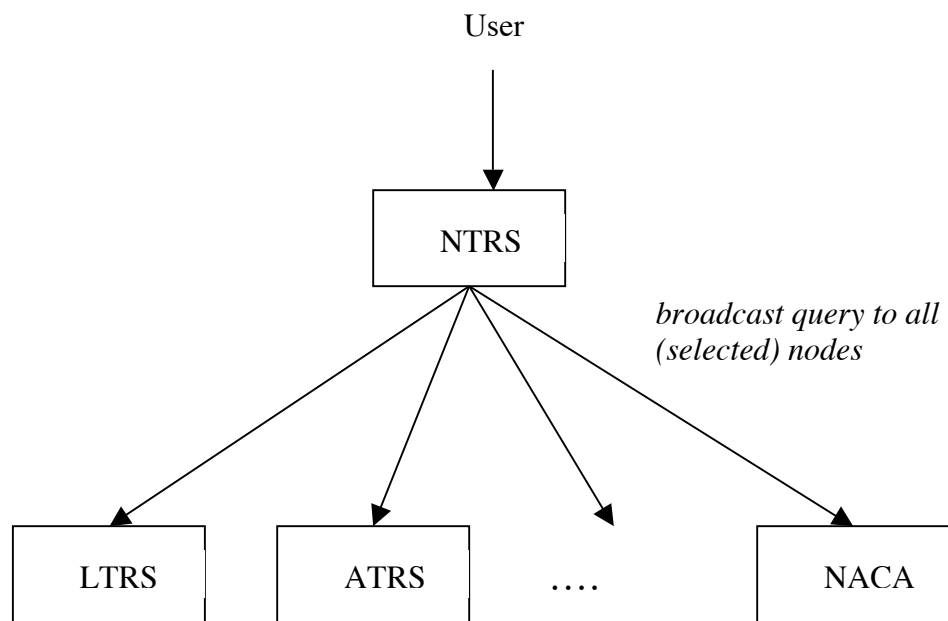
Figure 1. Architecture of the Distributed Searching NTRS
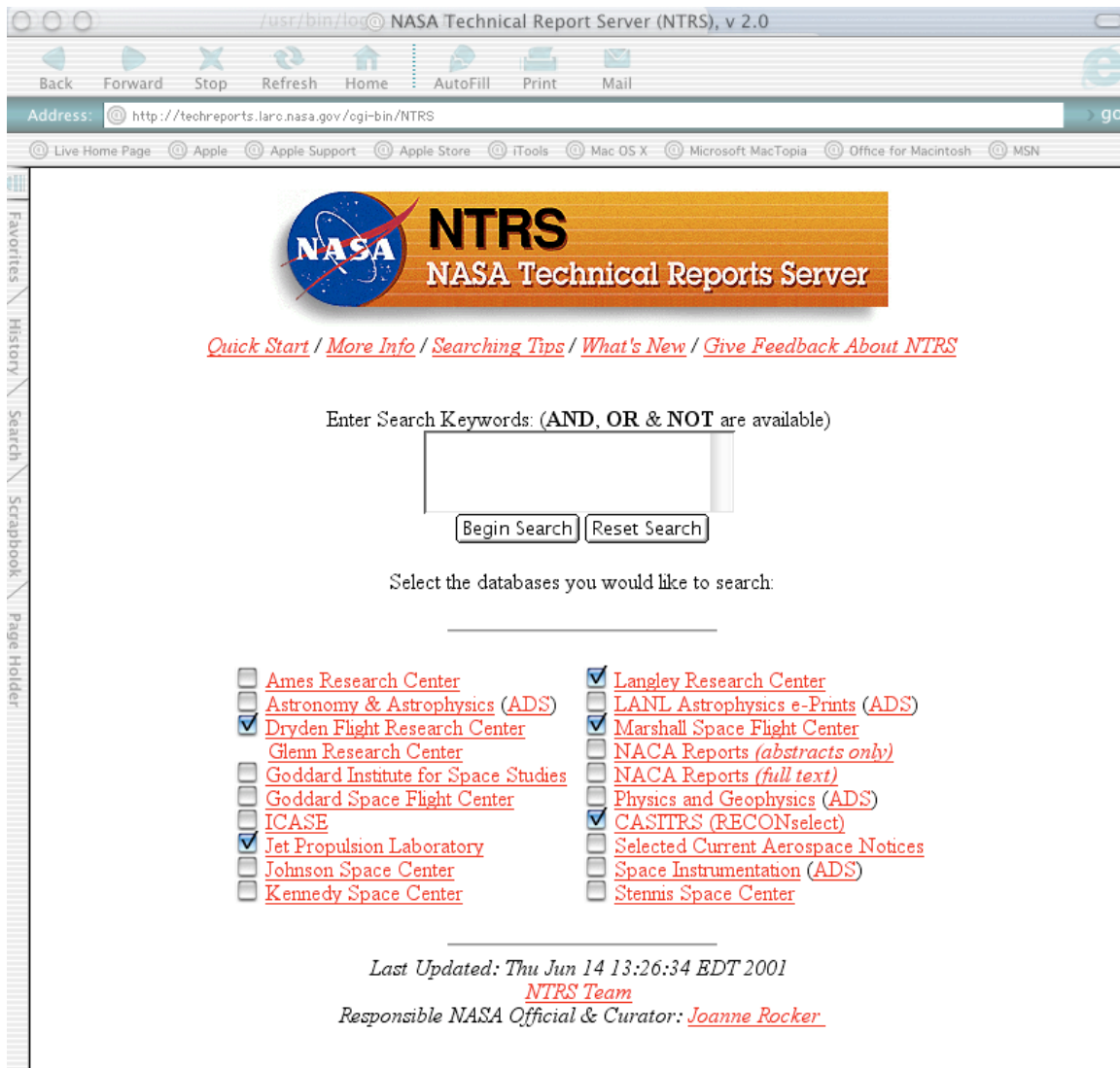
Figure 2. Screen Capture of the Distributed Searching NTRS

Bucket

index.cgi

_methods.pkg	_http.pkg	_log.pkg
source files for methods and OAI verbs	http dependency files	logs

_md.pkg	_state.pkg	_tc.pkg
metadata	bucket state	terms and conditions

default bucket packages

oai.pkg
oai.pl sets.pl

bucket payload
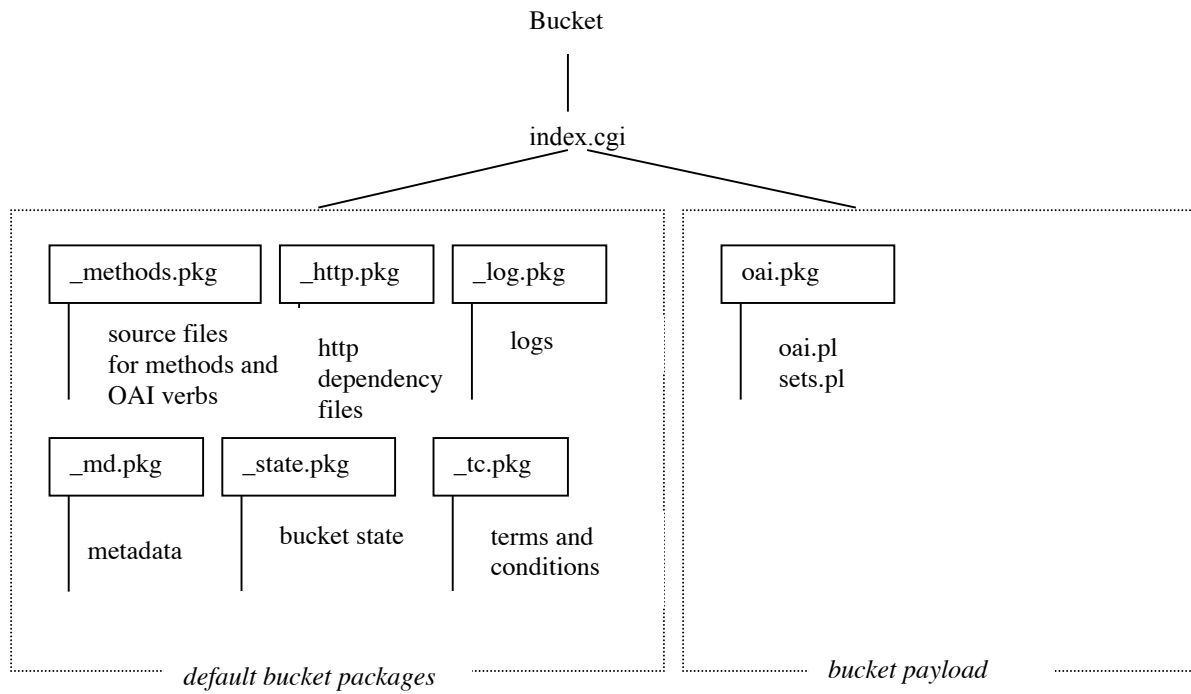
Figure 3. Architecture of an OAI-PMH Bucket

Figure 4. Simple Search Interface of the New OAI-PMH NTRS

Figure 5. Advanced Search Screen of the OAI-PMH NTRS

Bucket

|

index.cgi

_methods.pkg

source files
for methods, OAI
verbs and NTRS
searching

_http.pkg

http
dependency
files

_log.pkg

logs

_md.pkg

metadata

_state.pkg

bucket state

_tc.pkg

terms and
conditions

default bucket packages
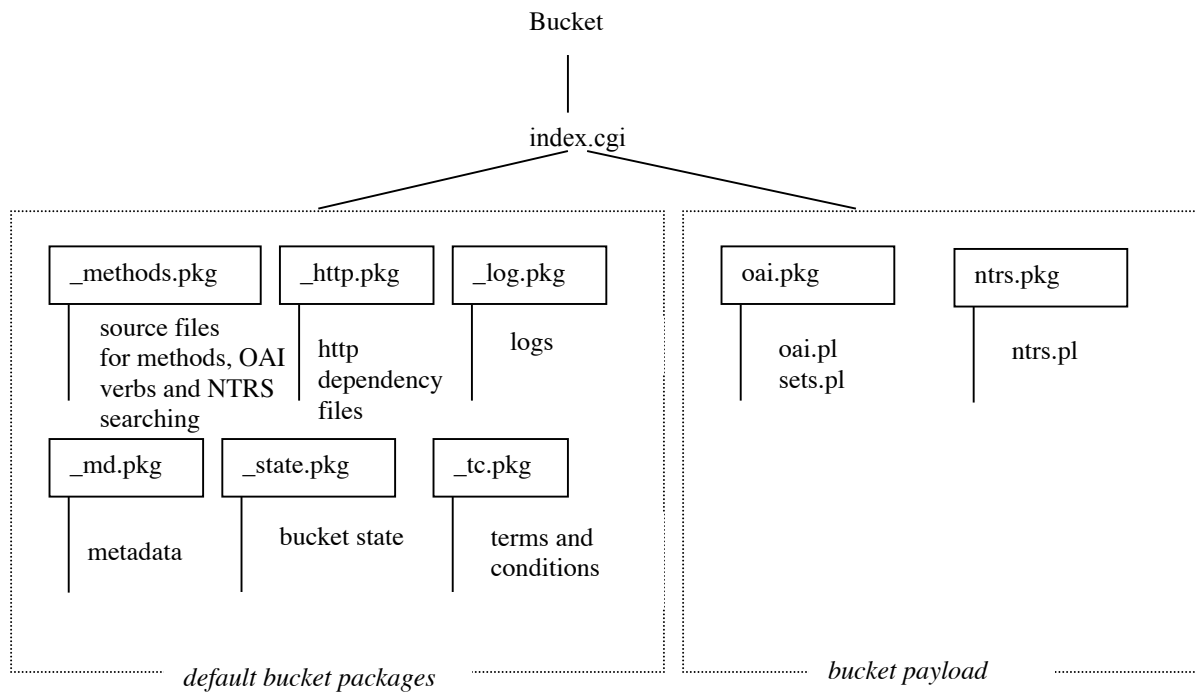
oai.pkg

oai.pl
sets.pl

ntrs.pkg

ntrs.pl

bucket payload

Figure 6. Architecture of the NTRS Bucket