The OAI-PMH NASA Technical Report Server

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

JoAnne R. Calhoun, Calvin E. Mackey
NASA Langley Research Center
Hampton VA 23681 USA
{JoAnne.R.Calhoun,Calvin.E.Mackey}@nasa.gov

ABSTRACT
The NASA Technical Report Server (NTRS) is now based on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). This metadata harvesting version of NTRS represents a significant improvement over the previous distributed searching implementation of NTRS. In addition to being an OAI-PMH service provider, the new version of NTRS is also an OAI-PMH aggregator. This allows NTRS to serve as a one-stop shop for harvesting of NASA metadata.

Categories and Subject Descriptors
H.3.7 [Information Storage and Retrieval]: Digital Libraries.

1. INTRODUCTION
NASA’s history with web-based DLs dates back to 1993, when a WWW interface was provided for the Langley Technical Report Server (LTRS). However, it was not until 1995 that the NASA Technical Report Server (NTRS) was setup to provide integrated searching between the various NASA web-based DLs. NTRS proved extremely successful with the public; a detailed history of NASA’s WWW DLs can be found in [1].

2. NASA OAI Data Providers
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 [2], 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. 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.

3. NASA OAI Service Provider
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 (only 6 of 20) because of many nodes having highly variable availability.

However, this is no longer an issue 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 and an advanced search interface 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.

New in the OAI-PMH version of NTRS is the inclusion of repositories that are not in the nasa.gov domain. At the moment, we include repositories from the Physics eprint Server (arXiv), Biomecnetral, Aeronautical Research Council (the UK-equivalent of NASA) and the Department of Energy. The simple search interface searches only the NASA repositories by default. The advanced search interface (which features fielded searching) offers the possibility of including non-NASA repositories. Several other interfaces are provided as well, including: browsing, weekly updates, and administration.

4. Conclusions
NTRS is implemented as a specialized bucket [3], and uses a variety of technologies to provide its services, including the Virginia Tech OAI-PMH harvester, an OAI-PMH repository (thus making NTRS an aggregator), a MySQL database, the awstats http log analysis facility, and a variety of support scripts to integrate the various aspects. Both the user interface and baseURL for harvesters is http://ntrs.nasa.gov/.

This model for implementing DLs appears to be very successful. We have been able to quickly (i.e. 1 day) replicate this model for a version of the NACA Technical Report Server, currently under testing by NASA staff.

5. REFERENCES