

OPEN ARCHIVES FORUM: OAI ARCHIVES – INVENTORY OF SERVICE PROVIDERS

| | |
|------------------------|---------------------|
| Project Number: | IST-2001-32015 |
| Project Title: | Open Archives Forum |

| | |
|--|---|
| Date of Delivery: | 30-09-2003 |
| Title: | Open Archives – Inventory of Service Providers |
| Workpackage contributing to the Document: | WP2 |
| Total Number of Pages: | 8 |
| URL: | http://www.oaforum.org/otherfiles/tv-services.pdf |
| Author: | Birgit Matthaei |
| Contact Details: | Birgit Matthaei Humboldt-University Berlin, CMS (Computing Centre) Rudower Chaussee 26, 12489 Berlin, Germany |

1 OAI ARCHIVES

1.1 Inventory of Service Providers

This inventory schedules 37 service providers. For each the aspects name, URL, contact, short description and harvested data are registered as displayed below. This inventory is an extension of the OAI and OA-Forum database resources¹. on the state of August 2003.

¹ Open Archives Initiative: <http://www.openarchives.org/service/listproviders.html> •
Open Archives Forum: http://www.oaforum.org/oaf_db/list_db/list_services.php

| | |
|-------------------|---|
| Service | <i>AmericanSouth.org</i> |
| URL | http://www.americansouth.org/ |
| Contact | |
| Service & content | Different disciplines, such as history, sociology, archeology, geography, anthropology, American studies, linguistics, art, religion, and/or music. Southern history and culture. Contextual and interpretive tools. |
| Harvested data | List at: http://www.americansouth.org/modules.php?op=modload&name=NS-Browse&file=index , list “Results from other archives” |
| Service | <i>ARC: A Cross-Archive Service</i> |
| URL | http://arc.cs.odu.edu/ |
| Contact | dlib@list.odu.edu |
| Service & content | first end-user service provider • source available through SourceForge • hierarchical harvesting Arc is an experimental research service of Digital Library Research group at Old Dominion University. Arc is used to investigate issues in harvesting OAI compliant repositories and making them accessible through a unified search interface. It is not a production service and may be subject to unscheduled service interruptions and anomalies. |
| Harvested data | Harvests all known archives |
| Service | <i>ARCHON</i> |
| URL | http://archon.cs.odu.edu/ |
| Contact | dlib@list.odu.edu |
| Service & content | physics metadata • based on Arc• features: citation indexing, equation-based searching A Digital Library that federates Physics collections with varying degrees of metadata richness |
| Harvested data | APS, CERN, Emilio, NTRS-Physics, arXiv_ List at: http://mercury.seven.research.odu.edu/archon , link: “Browse”, list: “Archive Groups” |
| Service | <i>arxiv.org</i> |
| URL | http://arxiv.org/ |
| Contact | www-admin@arXiv.org |
| Service & content | ArXiv is an e-print service in the fields of physics, mathematics, non-linear science and computer science. |
| Harvested data | |
| Service | <i>BICTEL/e</i> |
| URL | http://edoc.bib.ucl.ac.be/ |
| Contact | dardenne@sceb.ucl.ac.be |
| Service & content | E-prints and dissertations of nine universities of the French Community of Belgium. |
| Harvested data | List at: http://edoc.bib.ucl.ac.be/participant_bictel.html |
| Service | <i>Celestial – Open Archives Gateway</i> |
| URL | http://celestial.eprints.org/ |
| Contact | |
| Service & content | OAI “cache”, also 1.1 -> 2.0 converter Celestial is a service that harvests metadata from repositories that support the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), and caches that data for other services to harvest. Celestial can harvest from repositories that support the OAI-PMH 1.0, 1.1, and 2.0 versions. Metadata can be harvested from Celestial using version 2.0. By utilising static metadata records, and permanent resumption tokens Celestial aims to provide a high-performance and reliable mirror for existing OAI-PMH compatible resources. Celestial implements the OAI providence schema (records the path a metadata record has taken). |
| Harvested data | List at: http://celestial.eprints.org/cgi-bin/status |

| | |
|-------------------|--|
| Service | CERN Document Server |
| URL | http://cdsweb.cern.ch/ |
| Contact | cds.support@cern.ch |
| Service & content | Over 550,000 bibliographic records, including 220,000 full text documents, of interest to people working in particle physics and related areas. Covers preprints, articles, books, journals, photographs, and much more. |
| Harvested data | List at: http://cdsweb.cern.ch/ (bottom of the page) |
| Service | citebaseSearch |
| URL | http://citebase.eprints.org/ |
| Contact | tdb01r@ecs.soton.ac.uk |
| Service & content | arXiv metadata • citation based indexing, reporting Citebase is currently only an experimental demonstration. It allows researchers to search across free, full-text research literature e-print archives, with results ranked according to many criteria (e.g. by <i>citation impact</i>), and then to navigate that literature using citation links and analysis. Citebase has articles from Physical, Mathematical, Computer Science, Psychology, Neuroscience, and Biomedical Sciences. |
| Harvested data | arXiv.org: http://arXiv.org/ cogprints: http://cogprints.soton.ac.uk/ Biomed Central: http://www.biomedcentral.com/ |
| Service | Cyberthèses – Publication et diffusion en ligne des thèses |
| URL | http://www.cybertheses.org/ |
| Contact | vbouletr@univ-lyon2.fr |
| Service & content | in development |
| Harvested data | |
| Service | CYCLADES |
| URL | http://www.ercim.org/cyclades/ |
| Contact | webmaster@ercim.org |
| Service & content | scientific metadata • features: personalization, recommendations, collaboration CYCLADES is a system, designed to provide an open collaborative virtual archive environment, which (among others) supports users, communities (and their members) with functionality for (i) advanced SEARCH in large, heterogeneous, multidisciplinary OAI compliant digital archives (ii) COLLABORATIVE work; and (iii) FILTERING and RECOMMENDATION of records, users, communities and collections. CYCLADES allows users to register any OAI archive to CYCLADES, which then will be automatically harvested and indexed. |
| Harvested data | Archives that adhere to the Open Archives Initiatives harvesting protocol specifications |
| Service | DINI Search Service |
| URL | http://edoc.hu-berlin.de/e_suche/oai.php |
| Contact | dobratz@rz.hu-berlin.de |
| Service & content | DINI- Search Service for OAI Data Provider of German University Libraries, which mostly contain digital dissertations, videos of lectures, master theses, public readings, university publications. |
| Harvested data | List at: http://edoc.hu-berlin.de/e_suche/oai.php , list at “Archiv” |
| Service | DiVA - Academic Archive On-line |
| URL | http://publications.uu.se/portal/index.xsql?lang=en |
| Contact | eva.muller@ub.uu.se |
| Service & content | In DiVA you can find theses, dissertations and other publications in full-text from the universities of Stockholm, Södertörn, Umeå, Uppsala and Örebro. |
| Harvested data | |

| | |
|-------------------|--|
| Service | <i>DP9- An OAI Gateway Service for Web Crawlers</i> |
| URL | http://arc.cs.odu.edu:8080/dp9/index.jsp |
| Contact | liu_x@cs.odu.edu |
| Service & content | gateway to expose repository contents in HTML suitable for web crawlers DP9 is an open source gateway service that allows general search engines, like Google, to index OAI-compliant archives. It stands between the crawler and the archive, intercepts the crawler's requests, forwards them to the archive, and translates the archive's output from XML into HTML. This allows OAI archives hidden in the deep Internet to be indexed by search engines that don't venture into the deep internet. DP9 also supports OAI name resolution and service linking. DP9 does this by providing consistent URLs for repository records, and converting them to OAI queries against the appropriate repository when the URL is requested. This allows search engines that do not support the OAI protocol to index the "deep Web" contained within OAI compliant repositories |
| Harvested data | All OAI-compliant data providers List at: http://arc.cs.odu.edu:8080/dp9/index.jsp |
| Service | <i>EEPI - Erasmus Electronic Publishing Initiative</i> |
| URL | http://web.eur.nl/ub/eepe |
| Contact | Ellermann@ubib.eur.nl |
| Service & content | in development |
| Harvested data | |
| Service | <i>iCite</i> |
| URL | http://icite.sissa.it/ |
| Contact | asnicar@sissa.it |
| Service & content | physics metadata • features: citation based access to arXiv metadata An automatic citation indexing system covering physics journals. |
| Harvested data | List of Journals: http://icite.sissa.it/ , list at "Journal" |
| Service | <i>MetaArchive.org</i> |
| URL | http://www.metaarchive.org |
| Contact | |
| Service & content | MetaArchive.org is a metadata harvesting and searching service that will offer multiple institutions a combination of new technologies for sharing information about locally maintained resources of interest to scholars, as well as a means of seeking and discovering complementary information held by other institutions. MetaArchive.org will provide a publicly searchable web interface to this metadata aggregation. The web interface will enable both comprehensive searches and searching strategies targeted to particular subject domains. |
| Harvested data | List at: http://callimachus.library.emory.edu:8180/metaarchive , link: "Browse" |
| Service | <i>my.OAI</i> |
| URL | http://www.myoai.com/ |
| Contact | webmaster@myoai.com |
| Service & content | covers all registered metadata • features: result sets, personalization, many other advanced features my.OAI is a full-featured search engine. |
| Harvested data | Selected list of metadata databases from the Open Archives Initiative project. |
| Service | <i>NDLTD Union Catalog</i> |
| URL | http://rocky.dlib.vt.edu/~etdunion/ |
| Contact | Iming@vt.edu |
| Service & content | Electronic Thesis/Dissertation OAI Union Catalog, experimental project. This is a service built by harvesting metadata from Open Archives of electronic theses and dissertations. The underlying technology is based on layered Open Archives with data being harvested from source archives and then stored in a Union Catalog. This Union Catalog is then front-ended with a search engine for demonstration purposes, but the data is just as easily accessible to other service providers, both local and remote. |
| Harvested data | List at: http://rocky.dlib.vt.edu/~etdunion/ |

| | |
|-------------------|--|
| Service | Networked Computer Science Technical Reference Library |
| URL | http://www.ncstrl.org |
| Contact | dlibug@cs.odu.edu |
| Service & content | metadata harvesting replacement for Dienst-based NCSTRL • based on Arc • computer science metadata NCSTRL provides unified access to technical reports and e-prints from computer science departments, institutes and laboratories. This is an OAI-based implementation of the NCSTRL project. This version replaces the Dienst architecture and protocol with the OAI metadata harvesting protocol. |
| Harvested data | List at: http://www.ncstrl.org/ , link: “Browse” |
| Service | NTRS – NASA Technical Reports Server |
| URL | http://ntrs.nasa.gov/ |
| Contact | help@sti.nasa.gov |
| Service & content | The NASA Scientific and Technical Information (STI) Program's mission is to collect, archive, and disseminate NASA aerospace information, and locate domestic and international STI pertinent to NASA's missions and Strategic Enterprises. Examples of NASA's STI include research reports, journal articles, conference and meeting papers, technical videos, mission-related operational documents, and preliminary data. NTRS also collects scientific and technical information from sites external to NASA to broaden the scope of information available to users. NTRS's Simple Search searches for NASA information only and its Advanced Search can search for NASA and non-NASA information. |
| Harvested data | List at: http://ntrs.nasa.gov/?method=about |
| Service | OAI Harvester Crawling Status |
| URL | http://www.mds.rmit.edu.au/~ajk/oai/interop/summary.htm |
| Contact | ajk@mds.rmit.edu.au |
| Service & content | Results from an automated crawler. View of a log of actions taken against the site of a repository (including error messages). |
| Harvested data | List at: http://www.mds.rmit.edu.au/~ajk/oai/interop/summary.htm |
| Service | OAIster |
| URL | http://oaister.umdl.umich.edu/ |
| Contact | oaister@umich.edu |
| Service & content | OAIster provides a search service on OAI archives. Goal is to create a collection of freely available, difficult-to-access, academically-oriented digital resources that are easily searchable by anyone. |
| Harvested data | Harvests all known archives List at: http://oaister.umdl.umich.edu/o/oaister/viewcolls.html |
| Service | OLAC: Open Language Archives |
| URL | http://www.language-archives.org |
| Contact | olac-admin@language-archives.org |
| Service & content | OLAC, the Open Language Archives Community, is an international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by: (i) developing consensus on best current practice for the digital archiving of language resources, and (ii) developing a network of interoperating repositories and services for housing and accessing such resources. |
| Harvested data | List of Archives: http://www.language-archives.org/archives.php4 List of Services: http://www.language-archives.org/services.php4 |

| | |
|-------------------|---|
| Service | <i>Perseus Digital Library</i> |
| URL | http://www.perseus.tufts.edu/cgi-bin/vor |
| Contact | dasmith@perseus.tufts.edu |
| Service & content | Digital library of resources for the study of the humanities. The Perseus system harvests registered OAI repositories and incorporates the information into its search interface. For documents that are web-accessible, Perseus creates two links: one directly to the outside page, and one with added Perseus links. As with texts in Perseus, these links point back to related resources in the digital library. |
| Harvested data | Perseus harvests all registered OAI data providers List at: http://www.perseus.tufts.edu/cgi-bin/collview |
| Service | <i>PhysNet</i> |
| URL | http://www.eps.org/PhysNet/ |
| Contact | PhysNet@ISN-Oldenburg.de |
| Service & content | PhysDoc - Physics Documents Worldwide - offers lists of links to document sources, such as preprints, research reports, annual reports, and list of publications of worldwide distributed physics institutions and individual physicists, ordered by continent, country and town. |
| Harvested data | |
| Service | <i>ProPrint</i> |
| URL | http://www.proprint-service.de/ |
| Contact | m.schulz@www.proprint-service.de , i.tappenbeck@www.proprint-service.de |
| Service & content | ProPrint aims to establish a distributed print-on-demand service based upon the two document servers of Humboldt-University and the State and University Library Goettingen. The OAI protocol has been extended with a new metadata set and an extended protocol request to reach those aims. State: Test implementation |
| Harvested data | Document servers of Humboldt University Berlin and the State and University Library Goettingen |
| Service | <i>Public Knowledge Project Open Archives Metadata Harvester</i> |
| URL | http://www.pkp.ubc.ca/harvester/ |
| Contact | harvester@pkp.ubc.ca |
| Service & content | The Public Knowledge Project has developed a number of discipline-specific Research Support Tools (RST), which accompany individual research studies indexed from e-journal and conference paper websites covering a wide range of disciplines. The RST utilizes the study's metadata to search relevant open-access databases for related studies, theory, news, policies, and other resources, as well as offering access to the study's metadata and citation, to a personal portfolio, and to email and comment options. |
| Harvested data | List at: http://www.pkp.ubc.ca/harvester/archives.php |
| Service | <i>RDN – Resource Discovery Network</i> |
| URL | http://www.rdn.ac.uk/ |
| Contact | rdn-support@rdn.ac.uk |
| Service & content | RDN bring together the existing data within the RDN and build a harvested database using and further developing the WSE (Web Search Environments) software. WSE is a search system which takes URIs, titles, description and record URIs fields from subject gateway records and then runs a webcrawl from there, tracking the routes and storing the results in RDF databases. This data then be made available both through the RDN ResourceFinder pages at www.rdn.ac.uk and through the hubs of the RDN. |
| Harvested data | There are currently seven hubs: ALTIS - Hospitality, Leisure, Sport and Tourism, BIOME - Health and Life Sciences, EEVL - Engineering, Mathematics and Computing, GEsource - Geography and Environment, Humbul – Humanities, PSIgate - Physical Sciences, SOSIG - Social Sciences, Business and Law |

| | |
|-------------------|---|
| Service | <i>Repository Explorer</i> |
| URL | http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai |
| Contact | hussein@vt.edu |
| Service & content | An interactive, web-based tool to test repositories for compliancy with various levels of the Open Archives Initiative Protocol. |
| Harvested data | |
| Service | <i>SAIL-eprints</i> |
| URL | http://eprints.bo.cnr.it/ |
| Contact | a.tugnoli@jolly.bo.cnr.it |
| Service & content | SAIL-eprints (Search, Alert, Impact and Link) is an electronic open access service provider for finding scientific or technical documents, published or unpublished, in Chemistry, Physics, Engineering, Materials Sciences, Nanotechnologies, Microelectronics, Computer Sciences, Astronomy, Astrophysics, Earth Sciences, Meteorology, Oceanography, Agriculture, and related application activities. |
| Harvested data | List at: http://eprints.bo.cnr.it/ , link: "Data Provider" |
| Service | <i>Scirus, the web search engine for scientific information</i> |
| URL | http://www.scirus.com |
| Contact | webmaster@scirus.com |
| Service & content | Scirus distinguishes itself from existing search engines by concentrating on scientific content only and by searching both web and (often proprietary) databases. Search engines generally index the web by letting a robot crawl websites. However, a lot of information on the Web is invisible to search engines, because they are either not allowed on the site or are technically not capable of crawling them. Sixty percent of the information on the Web is thought to be "invisible" in this way. Scirus' aim is to provide scientists with one comprehensive search platform covering both the web and the normally "invisible" databases. |
| Harvested data | List at: http://www.scirus.com/about/#sources |
| Service | <i>Sheetmusic Service [Sheet Music Consortium]</i> |
| URL | http://digital.library.ucla.edu/sheetmusic/ |
| Contact | shs@library.ucla.edu |
| Service & content | The Sheet Music Consortium is a group of music libraries working with digital library programs in their respective institutions toward the goal of building an open collection of digitized sheetmusic using the Open Archives Initiative: Protocol for Metadata Harvesting (OAI:PMH).Harvested metadata about sheet music in participating collections is hosted by UCLA Digital Library Program, which provides a service allowing access via this metadata to sheet music records at the host libraries. |
| Harvested data | List at: http://digital.library.ucla.edu/sheetmusic/librarian?SEARCHPAGE&Browse , list "Collection" |
| Service | <i>Socionet Personal Zone</i> |
| URL | http://spz.socionet.ru/index-en.shtml |
| Contact | admin@socionet.ru |
| Service & content | Workplace for Researchers, Tutors and Students in the Social Sciences Information Space. At the Socionet Personal Zone (SPZ) one can create personal collections of research materials (in Economics) with OAI-PMH 2.0 interface. To create collections one has to be registered at the SPZ. |
| Harvested data | linked RePEc sites |

| | |
|-------------------|--|
| Service | <i>tool2know</i> |
| URL | http://www.tool2know.com/ |
| Contact | info@tool2know.com |
| Service & content | In the world wide web, resources like e-magazines, data collections, e-books and PDF documents are difficult to trace and TOOL2KNOW offers a way to track them through a storyboard based on a dartboard approach. In this version of TOOL2KNOW, you can find two hyper guides: the guide to the Italian Museums websites and the one to the Italian scientific research websites. A third hyper guide about Italian newspapers websites is under construction. |
| Harvested data | List at: http://www.tool2know.com/default.asp?g=1 , link “List” |
| Service | <i>TORII, an open portal for open archives</i> |
| URL | http://torii.sissa.it/ |
| Contact | tips@medialab.sissa.it |
| Service & content | physics metadata • features: personalization, recommendations, WAP access Unified access to various open archives (Physics and Computer Science). Filtering and advanced searching. Open peer review (QCT). Personalization. State: Prototype |
| Harvested data | List at: http://torii.sissa.it |
| Service | <i>UIUC Digital Gateway to Cultural Heritage Materials</i> |
| URL | http://nergal.grainger.uiuc.edu/cgi/b/bib/bib-idx |
| Contact | t-cole3@uiuc.edu |
| Service & content | The Library of the University of Illinois at Urbana-Champaign (UIUC) is creating and implementing a suite of Open Archives Initiative (OAI)-based metadata harvesting services, search services, and tools designed to facilitate discovery and retrieval of certain classes of scholarly works, thereby making visible portions of the currently "hidden" Web of scholarly information resources. |
| Harvested data | List at: http://oai.grainger.uiuc.edu/AboutCollections.htm |
| Service | <i>Vac</i> |
| URL | http://archon.cs.odu.edu/VAC_Homepage.html |
| Contact | |
| Service & content | A gateway service to harvest non-OAI collection into OAI compliant repository Status : Work in progress |
| Harvested data | |