

General Description of All Projects, Contd.

APPENDIX B GENERAL DESCRIPTION OF ALL PROJECTS

Project/Operating Agent	URL/Contact	Degree of Implementation	% STI	Data Type(s)	Key Archive Characteristics
ADONIS	www.adonis.nl/	operational	100%	e-journals	Creates e-journal archive as part of its CD-ROM product. Soon to go online.
American Psychological Association	wpickren@apa.org	prototype	mix	word processed business documents	Professional society creating a digital institutional archive.
* American Astrophysical Society	www.aas.org	operational	100%	e-journals and supplementary material	Learned society archive linked to archives in astrophysical sciences worldwide. Collaboration to maintain links. Escrow fund for migrating system.
* American Institute of Physics	www.aip.org/journals/archive/index.html	operational	100%	e-journals, supplementary data	Learned society publisher archiving its journals and those of its member societies. Well established licensing policy.
American Chemical Society	pubs.acs.org/journals/faqgen.html	operational	100%	e-journals, supplementary data	Learned society publisher archiving its journals. Public preservation statement. Moving to Oracle.
* Atmospheric Radiation Program	rqm@ornl.gov	operational	100%	data	US Department of Energy program to monitor radiation, etc. Large volume of incoming information and large quantity ftp'd per day.
* Carbon Dioxide Information Analysis Center (CDIAC)	cdiac.esd.ornl.gov/cdiac/	operational	100%	data	A World Data Center for Atmospheric Trace Gases. Primary global-change data and information analysis center of the U.S. Department of Energy (DOE). Unlike many traditional data centers that are discipline-based (for example, meteorology or oceanography), CDIAC's scope includes everything that could be of value to users concerned with the greenhouse effect and global climate change. Responsible for collection, documentation, value-added products, dissemination, customer support, and quality control.

* Indicates highlighted project.

General Description of All Projects, Contd.

	CEDARS (CURL Exemplars in Digital Archives) Project	www.leeds.ac.uk/c edars/	design	n/a	mixed; varies by demonstra- or project	Higher Education initiative funded by the Joint Information Systems Committee through the eLib programme. Funding was awarded to the Consortium of University Research Libraries (CURL) and the work is performed by three libraries – Leeds, Oxford and Cambridge. Main objective is to provide guidance in best practice for digital preservation, based on demonstrator projects and strategic working group discussions in broad areas of concern. Particularly interested in intellectual property issues. Main deliverables will be recommendations and scalable models for establishing distributed digital archives. Cedars is expected to influence national legislation on legal deposit and provide input to the national strategy for digital archives being developed through the British Library's National Preservation Office.
*	Defense Information Technology Testbed/US DoD (US)	tborkows@dtic.mil	prototype	100%	multimedia	Archiving multimedia objects including video from Unmanned Aerial Vehicles, still imagery, transcribed text, and audio voice-overs by pilots. Original material received from Joint Analysis Center in the UK and information is stored and preserved at Ft. Leavenworth, Kansas, US in a multimedia data warehouse system.
*	Distributed Object Computation Testbed (DOCT) San Diego Supercomputer / Center US PTO	www.sdsc.edu/DO CT	prototype	mix	patents	The Distributed Object Computation Testbed (DOCT) has created a testbed system for handling complex documents on geographically distributed data archives and computing platforms. The technologies should apply to the information needs the US PTO and other U.S. federal agencies. Technologies include replicated archives, redundant communication paths and advanced database technologies to access heterogeneous databases, in a secure environment.

* Indicates highlighted project.

General Description of All Projects, Contd.

*	Electronic Publications Preservation Project/NLC Electronic Collection (Canada)	collection.nlc-bnc.ca/e-coll-e/index-e.htm	operational	mix	e-journals and books	EPPP was a project of the NLC to pilot the acquisition, cataloging, preservation and access to electronic journals. This was completed and now the lessons learned have been incorporated into the NLC Electronic Collection. Electronic journals and books are accessible via the regular Web-based online public access catalog.
*	EVA -Finnish National Library/Helsinki University Library	renki.lib.helsinki.fi/eva/	late prototype	mix	e-publications	An 18-month project begun in June 1997. Emphasis is on development of a model for use at national libraries and local university libraries. Project has established guidelines for automatic harvesting of sites using a robot. Using a variety of tools available from EU and Nordic funded projects. Benefits of hierarchical storage versus sequential are outlined. Access is via relatively simple HTTP protocols. Metadata is extracted from the header of the collected site. Unique identifiers are added. Scheduler considers impact on Web servers and networks being hit by the robots.
	GenBank/National Center for Biotechnology Information	www.ncbi.nlm.nih.gov/Web/Genbank/index.html	operational	100%	biological sequence data	An active archive of DNA sequence data. Deposition and access is free. A GenBank number obtained upon deposition is required by biomedical publications. ASCII is the archival file. This is also used for downloading. Sybase version is used for searching, reporting and database management. Publically available resource, but changes and corrections to sequence data for particular species must be made by staff and approved by the researcher who "owns" the sequence. Commercial firms may download a copy of the entire database in order to ensure privacy of their search queries and results. Links are provided to MEDLINE file through the GenBank Number. Links may also be made to other databases that are linked from the NCBI Web site. Taxonomic authority file is also maintained.

* Indicates highlighted project.

General Description of All Projects, Contd.

*	HighWire Press (International)	intl.highwire.org/	operational	100%	e-journal	Department of Stanford University that supports small learned society publishers in electronic publishing. Includes archiving and additional added value linking.
*	JSTOR (International)	www.jstor.org/about	operational	10% in math; new Science Clusters	e-journal and digitized hard copy back issues	Nonprofit organization created out of the organization originally funded by the Mellon Foundation. Extensive work on operational cost models for institutional access to the archives. Includes both electronic and digitizing of paper back issues.
*	Kulturaw3/Royal Library, National Library of Sweden (KB) (Sweden)	kulturaw3.kb.se/html/projectdescription.html	prototype	unknown	all within domain regardless of format or data type; initial harvesting included images, sound, and text of various kinds	Aim of the project which began in September 1996, is to test methods for collecting, preserving and accessing Swedish electronic documents. Project includes periodicals, static documents, and dynamic documents with links (e.g., HTML pages). Less priority is given to conference, usenet groups, listservs, ftp archives and databases. The approach is to make the capturing as automatic as possible. Robots are being tested which collect publications based on the .se extension, location of the Web server in Sweden even though it has a .com extension, and foreign produced pages with a Swedish connection, e.g., translations of Swedish literature. Material would be collected and downloaded to the KB server routinely without review or selection. For electronic periodicals, a method similar to that of PANDORA is used to monitor and harvest every issue. KB will also test electronic delivery from producers. Other aspects of digital archiving that will be explored include hardware and software migration. Storage in several formats will be tested. The project will devise methods for limiting access until a "waiting period" has passed. This is related to review of the Swedish Deposit Law which is underway.

* Indicates highlighted project.

General Description of All Projects, Contd.

*	Long Term Ecological Research Network (US)	(Internet.edu/DTOC).	operational/ on-going network improvements	100%	data; some documenta- tion	Network of ecological data centers. Working on an architecture for even more extensive distribution. Also automatic extraction of metadata.
*	NASA Distributed Active Archive Centers (DAACs) (US)	ivanova.gsfc.nasa.gov/daac/	operational	100%	data	Distributed active data centers that ingest and store data related to the Earth Observing System Data and Information System (EOSDIS)/Global Change. Each collects and archives in a specific subject area. Centralized user access. Federated support to value-added product development and customer service.
*	National Digital Archive of Datasets (UK)/ Univ. of London Computer Center and Univ. of London Library under contract to the Public Records Office	ndad.ulcc.ac.uk	operational	<25%	data	A centralized archive for UK government digital data and related documents available since March 1998. Metadata catalogs are provided for searching.
	National Technical Information Service (US)	www.fedworld.gov /	operational	>50%	primarily public and private reports and other documents	Public outlet for US government technical reports. FEDWORLD service provides access to its own database and others. Participating in pilot project with the U.S. Federal Depository Library Program for electronic access to technical report at these libraries. Reviewing the cost recovery and technical implications.

* Indicates highlighted project.

General Description of All Projects, Contd.

*	Natural Environment Research Council (UK)	www.nerc.ac.uk/environmental-data/	operational	100%	data	One of six research councils in the UK. It has seven data centers which are responsible for the collection and preservation of data representing numerical environmental measurements in various areas including hydrology, geosciences, and earth and space. The data is collected both from remotely sensed imagery from satellites and from NERC's own aircraft observations. The centers are distributed among organizations (primarily universities) that have expertise in these areas. Federated centers, with standardized data management practices. Long-term preservation and backup is required in order to be considered as a possible NERC center.
	Networked Digital Library of Theses and Dissertations	www.ndltd.org/	operational; limited number of universities	mix	theses and dissertations	International network of universities with theses and dissertation servers. Input is PDF. Some supplementary data available via ftp. No consistent approach to digital electronic archiving; white paper has been produced
	Networked European Depository Libraries (NEDLIB)/European Union	http://www.konbib.nl/nedlib/	design	mix	primary focus on published materials; e-journals	Collaborative project of European national libraries to construct the basic infrastructure for a networked European deposit library. "The objectives of NEDLIB concur with the mission of national deposit libraries to ensure that electronic publications of the present can be used now and in the future.

* Indicates highlighted project.

General Description of All Projects, Contd.

*	OCLC Electronic Journals Project (International)	www.oclc.org.oclc/menu/eco.html	operational	mix	e-journals	An international bibliographic utility that provides electronic journal archiving as a service to its member libraries. The libraries subscribe to the e-journal with the publisher or through a jobber (also through OCLC) and the ability to archive the journal in perpetuity is part of the agreement. Some journals are available only through links to publisher sites, others deposit the electronic copies immediately. Working on business and pricing model for perpetual archive.
*	OhioLINK Electronic Journal Center (US)	www.ohiolink.edu	operational	mix	e-journals	Consortia of public, academic and special libraries in Ohio, US. Requires local deposit of electronic journal when consortial subscription is made with the publisher.
	Open Archive Information System/CCSD S	ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html	conceptual design (white paper)	n/a	data but may be applicable to other types	ISO requested that the Consultative Committee for Space Data Systems develop a reference model for interoperable archives. This reference model provides common definitions and a conceptual functional and data model for such archives. Some experts believe this may have value as a reference model for types other than data. Being used as a basis for the conceptual model for CEDARS.
*	Preserving and Accessing Networked Documentary Resources in Australia (PANDORA) (Australia)	www.nla.gov.au/pandora/	operational	unknown	Web sites, e-journals, etc.	Extensive effort to capture Australiana from the Web. Automatic harvesting and gathering schedules, selection guidelines, effort to create a national network of archives with the Australian state libraries.
	Profiles in Science/National Library of Medicine	www.profiles.nlm.nih.gov	operational	100%	documents, audio, images	Mixed media product on the history of medicine. Based on paper archives provided by donors. First product from NLM's Digital Library initiative. Paper and standard (generally ASCII) electronic archival copies are retained. Web version is optimized for current browser technologies. Metadata compatible with Dublin Core set.

* Indicates highlighted project.

General Description of All Projects, Contd.

	Protein Databank /Research Collaboratory for Structural Bioinformatics	rcsb.nist.gov/	operational; in transition from one site to another	100%	biological sequence data	Protein sequence data bank previously maintained by Brookhaven National Laboratory. Moving to the Research Collaboratory for Structural Bioinformatics, a collaboration of Rutgers University, the National Institute for Standards and Technology and the San Diego Supercomputer Center. Will be maintained as a reference data set for biological research. Previous unstructured ASCII files are being converted to a database for easier searching. ASCII will also be maintained as archive.
	Royal Society of Chemistry (UK)	Robert Welham isdir@rsc.org	operational	100%	e-journals	Archiving electronic version of publication using SGML. No links are maintained. Fees and access rights as for document delivery.
	San Diego Supercomputer / Center NARA	www.sdsc.edu/NA RA/	prototype	adminis- trative docu- ments	word processing documents, electronic mail, databases, text collections	The NARA extension to the DOCT project is exploring the application of high performance supercomputing to the creation and maintenance of a persistent object archive. The project includes technology to ingest 1 million records scalable to 40 million per year, handling heterogeneous data objects, and managing the migration to new commercial hardware and software.
	Scientific and Technical Information Center/ Taiwan	192.83.176.196/ejo urnal/nsce-j.htm	operational; internal journals only	100%	e-journals	STIC is currently archiving electronic journals and digitizing paper copies for titles published by the research institute. Expect to extend project to include other Taiwanese electronic journals in science and technology.

* Indicates highlighted project.

General Description of All Projects, Contd.

*	U.S. Environmental Protection Agency/ Environmental Management Information System	shepanek.robert@epamail.epa.gov	design	100%	data	An interactive system used to support all aspects of the analytical process related to environmental health and ecological assessment. Houses metadata descriptions for a variety of objects including project information, data sets, databases, models, images and multimedia. EIMS with some modification will support the long-term archiving of data and other objects. Project is currently underway to integrate EIMS with agency-wide records management approach. Datasets selected for use are loaded into the analytical component of the Oracle database, which also houses the metadata component. By transforming the original data structure into a common one under Oracle disparate databases can be integrated and common analytical tools such as SAS can be used. Data, information and tools not in the database are available via ftp. EIMS is predicated on a partnership model. Responsibility for administration and stewardship is distributed to the organization that owns the object. Refresh and archival decisions are subject to partner policies. EIMS metadata is based on the "20 year rule".
	U.S. Federal Depository Library Program	www.access.gpo.gov/su_docs/dpos/ec/about-ec.html	design stage; prototype with NTIS	n/a	government documents; pilot with technical reports	Group of 1200 research libraries in the US that are designated to provide regional access to government publications received from the agencies through the Government Printing Office. Recently published a strategy for transitioning to a more electronic depository program. Has a pilot underway with the National Technical Information Service.

* Indicates highlighted project.

	U.S. Department of Energy/Office of Scientific and Technical Information	www.osti.gov/	operational	100%	technical reports and other government publication products	Over the last several years have worked with Scientific and Technical Information Committee of DOE to receive technical reports (primarily from the national laboratories) in electronic form. Have established a system for receiving them, along with basic metadata, indexing them and providing access through its own system (InformationBridge) for all document security and limited distribution levels and through GPO Access (U.S. Government Printing Office) for unclassified documents. Analyzing a modified Dublin Core metadata scheme for new bibliographic database.
'2 for 1'/ingenta, Ltd.	www.ingenta.com/Tfedocs/aboutingenta.html	operational	>50%	e-journals	Provides electronic publishing services for small commercial and learned social publishers. Agreement with OCLC to archive journals through its Electronic Journals Program.	

* Indicates highlighted project.