E. Fox
Virginia Tech University
Department of Computer Science

International Interdisciplinary Open Archives and ETDs

In October 1999, a group met in Santa Fe, New Mexico (USA), to discuss the "*Universal Preprint Service*". We developed the Santa Fe Convention, agreeing to work together, and later renamed our effort as the **Open Archives Initiative**. Today, OAI is being deployed to create international interdisciplinary open archives. What is crucial is that OAI is flexible - able to support different approaches to such aggregation efforts.

At that very first meeting, I argued that while many groups were developing vertical or subject-focused archives (e.g., on math, physics, computing, economics, ...), there also was need to support horizontal aggregation efforts, leading to, for example, all the electronic theses and dissertations in Ohio, or Portugal, or Australia. In other words, it was important to suit the thematic, social, organizational, and political realities inherent in organizing distributed groups of scholars to share their publications with others. Fortunately, OAI is amenable to such tailoring, as long as needed work is undertaken. In this presentation we explore the multiple dimensions relating to these matters.

First, there is the matter of what authors will do. I argue that authors should be willing and able to not only write documents, but also to help suitable others to locate their works. This calls not only for submitting works, as into OAI repositories, but also for helping describe those works so others can find them. While these skills are not common today, they should be - that is one of the arguments for requiring graduate students to submit their own theses and dissertations, i.e., to make sure they can do this for at least one work in which they have strong interest. Yet, even if a student is required to submit and create a metadata record (e.g., using a tool like **My Meta Maker**, which may make the process easier, thus helping the author to do more), it is not clear how well that will be done, and what knowledge the student will have to classify the work using a subject-specific thesaurus or category system. Cataloging is often a difficult process, but something that graduate students could master if given suitable training (especially if given a tool like in OCLC's CORC, to aid semi-authomatic cataloging). I argue that such training makes sense for two reasons - to support their future work as an author (to catalog this and future works), and to make it easier for them to browse and search for other works using categories (once learned). Why not give authors tools so they can identify the best subject descriptors for their work, so that others can find their submission based on such fine grained cataloging? While this may be feasible and make sense, there is yet one other crucial answer regarding what authors will do. The answer is simply that students will undertake efforts that are commensurate with the motivation / pressure / loyalty that they feel to their department, institution, or other repository-building group.

Second, there is the question of how to deploy OAI for a given community. On the one hand, from its earliest days, OAI has been designed to support the universal or world community. If a work is placed in an OAI-compliant repository, its metadata can be harvested as a Dublin Core compliant XML record, and be a part of the collection of all such metadata worldwide. Yet, is that all that is needed to support physicists, or mathematicians, or students writing their thesis? Not really! It would be better if key members of the community were to plan how to adapt OAI to community needs. In particular, that means identifying one or more other metadata schemes to use, so that authors or others can fill in descriptions according to those schemes (including entering in subject descriptors according to a suitable classification system). Thus, the **Networked Digital Library of Theses and Dissertations**[1] has encouraged efforts, just completed, to release *ETDMS*, the electronic thesis and dissertation metadata standard. A metadata record that follows the ETDMS

[1] http://www.ndltd.org

guidelines should convey crucial information about the work, the degree, those who helped supervise its development and testing, and crucial classification data.

Third, there is the question of what classification system(s) to use. In the case of a particular discipline-oriented archive, the most popular classification system(s) for that discipline should be used. In the case of a thesis or dissertation, based on the discipline involved, such a classification system certainly should be applied. However, in addition, it makes sense to support the needs as well of those aiming to discover works who are not experts in a specific classification scheme. For such cases, it makes sense to consider three schemes. One is to have the author assign keywords and key phrases, preferably in large numbers, that are uncontrolled but still fit well as descriptors of the content. Another scheme is to use the Library of Congress Subject Headings (LCSH) scheme, to give a general or high level cataloging, preferably following a system like the Anglo-American Cataloging Rules (e.g., AACR2). A third scheme is to use something like the Dewey Decimal Classification (DDC) or LCSH. In sum, then, for the example of a dissertation in computing, one should add descriptors based on either or both of LCSH and DDC, as well as the ACM classification system.

Fourth, there is the matter of what scheme to put in place for supporting authors. We believe the short answer is: "whatever works the best!" The point is that someone or some group should run the archive, and should be able to sustain operations for a long time. Part of sustaining operations involves being able to get authors to contribute. Another part involves running the archive, and preserving content over long periods of operation. All these requirements can be met well by libraries (e.g., University Libraries at Virginia Tech) or library consortia (e.g., OhioLink for the state of Ohio in USA), or national libraries (e.g., the National Library of Portugal, selected by the leading universities to assume this responsibility).

Regarding working well, however, if something other than a thesis or dissertation is involved, e.g., a paper for a conference or journal, it may be most appropriate for submission to be to a discipline focused archive rather than an institution-oriented. Yet, to ensure sustainability, it may be best for the archive to also be supported by some institution, such as a university or publisher (e.g., Oldenburg's support of PhysDoc and ACM support of CoRR). While in the discussion above we sketch some of the concerns and solutions related to adapting OAI to needs such as are emphasized in this workshop, we can in the space available provide only an overview. We encourage study of the other presentations made in this workshop for further details and insights.

International Interdisciplinary Open Archives and ETDs

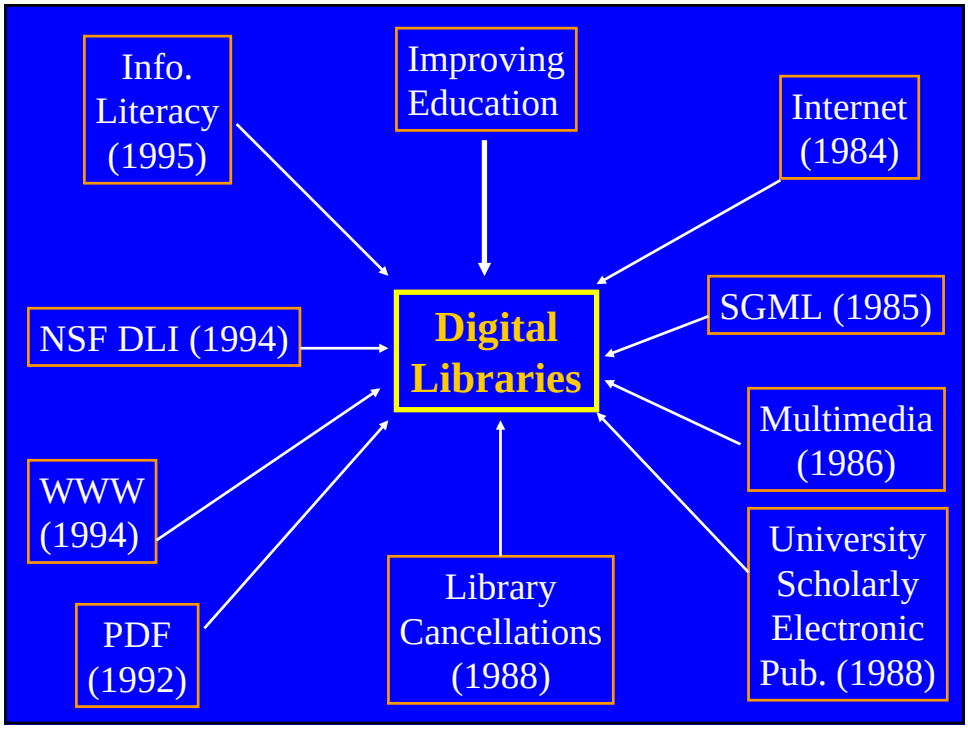
Edward A. Fox (Virginia Tech)
http://fox.cs.vt.edu – fox@vt.edu

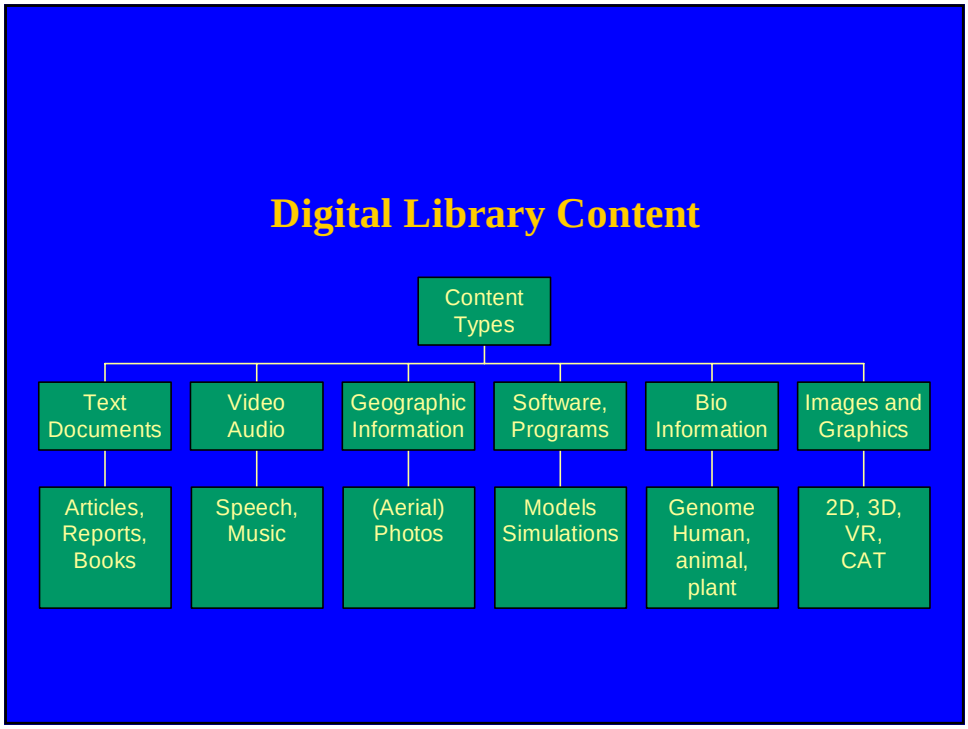
Workshop on International Interdisciplinary Open Archives and Subject Specific Services in Mathematics and Physics

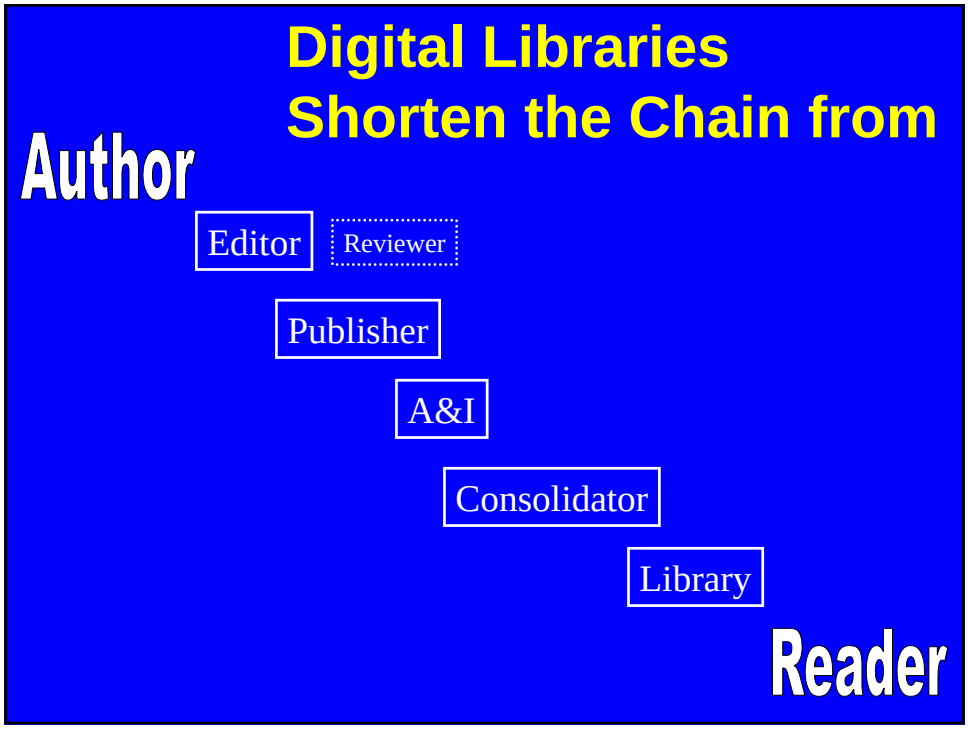
Duisburg, Germany – 3 Sept. 2001

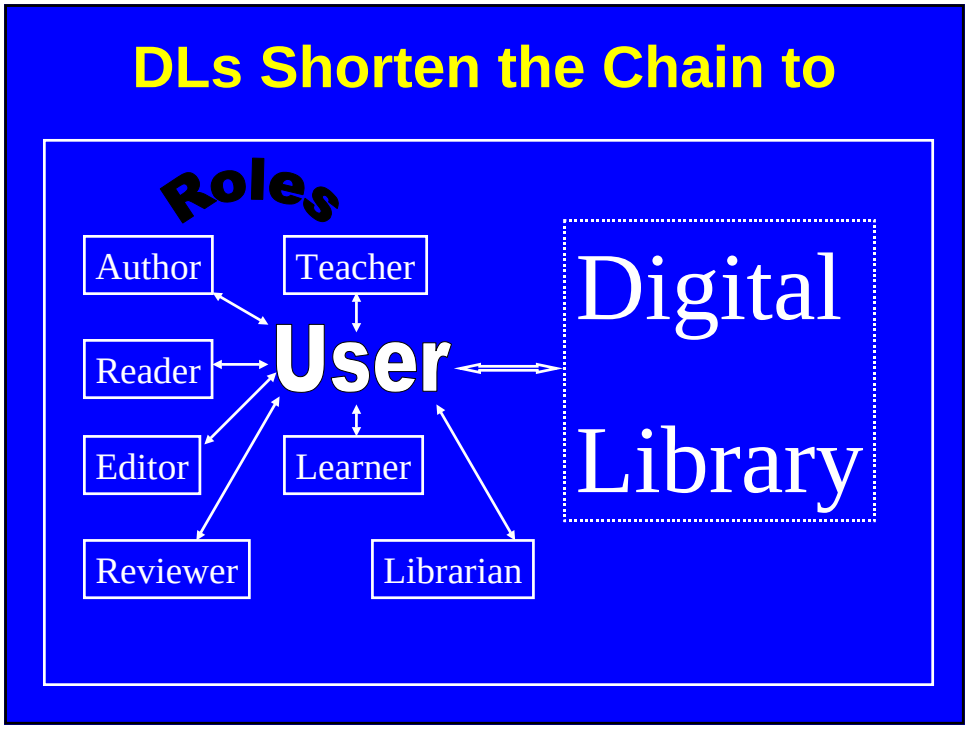
Acknowledgements (Selected)

- **Sponsors:** ACM, Adobe, IBM, Microsoft, NSF, OCLC, SOLINET, SURA, US Dept. of Ed. (FIPSE), VTLS, …
- **Faculty/Staff:** Tony Atkins, John Eaton, Carl Lagoze (Cornell), Gail McMillan, James Powell, Shalini Urs (Mysore), …
- **VT Students:** Fernando Das Neves, Robert France, Marcos Goncalves, Hussein Suleman, …


Info. Literacy (1995)
Improving Education
Internet (1984)
NSF DLI (1994)
Digital Libraries
SGML (1985)
Multimedia (1986)
WWW (1994)
University Scholarly Electronic Pub. (1988)
PDF (1992)
Library Cancellations (1988)


Digital Library Content
Content Types
Text Documents
Video Audio
Geographic Information
Software, Programs
Bio Information
Images and Graphics
Articles, Reports, Books
Speech, Music
(Aerial) Photos
Models Simulations
Genome Human, animal, plant
2D, 3D, VR, CAT


Digital Libraries
Shorten the Chain from
Author
Editor
Reviewer
Publisher
A&I
Consolidator
Library
Reader


DLs Shorten the Chain to
Roles
Author
Teacher
Reader
User
Editor
Learner
Reviewer
Librarian
Digital
Library

A Digital Library Case Study

- Domain: graduate education, research
- Genre:ETDs=electronic theses & dissertations
- Submission: http://etd.vt.edu
- Collection: http://www.theses.org

Project:

Networked Digital Library of Theses & Dissertations

(NDLTD)

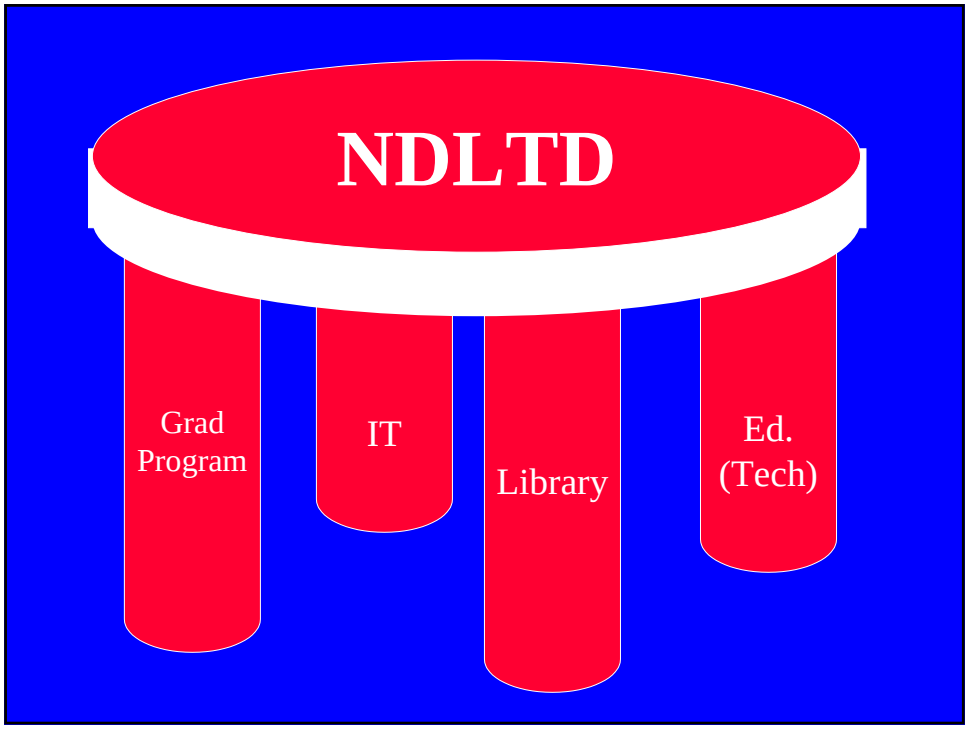
http://www.ndltd.org

The Networked Digital Library of Theses and Dissertations

www.NDLTD.org

Training Authors
Expanding Access
Preserving Knowledge
Improving Graduate Education
Enhancing Scholarly Communication
Empowering Students & Universities

Leader of the Worldwide ETD (Electronic Thesis and Dissertation) Initiative



What led to today's meeting?

- 1987 mtg in Ann Arbor: UMI, VT, …
- 1992 mtg in Washington: CNI, CGS, UMI, VT and 10 universities with 3 reps each
- 1993 mtg in Atlanta to start Monticello Electronic Library (regional, US Southeast): SURA, SOLINET
- 1994 mtg at VT: std: PDF + SGML + multimedia objects
- 1996 funding by SURA, US Dept. of Education (FIPSE)
- 1997 meetings in UK, Germany (and many follow on), ...
- 1998 – 1st symposium – Memphis (20)
- 1999 – 2nd symposium – Blacksburg (70)
- 2000 – 3rd symposium – St. Petersburg (225)
- 2001 – 4th symposium – Caltech (200)
- 2002 - May 30 – June 1, BYU; 2003 – Spring, in Berlin

What are the long term goals?

- Millions of students / year getting grad degrees are exposed / involved
- 200K/yr rich hypermedia ETDs that may turn into electronic portfolios (images, video, audio, …)
- Dramatic increase in knowledge sharing: literature reviews, bibliographies, …
- Services providing lifelong access for students: browse, search, prior searches, citation links
- Hundreds/thousands of downloads / year / work

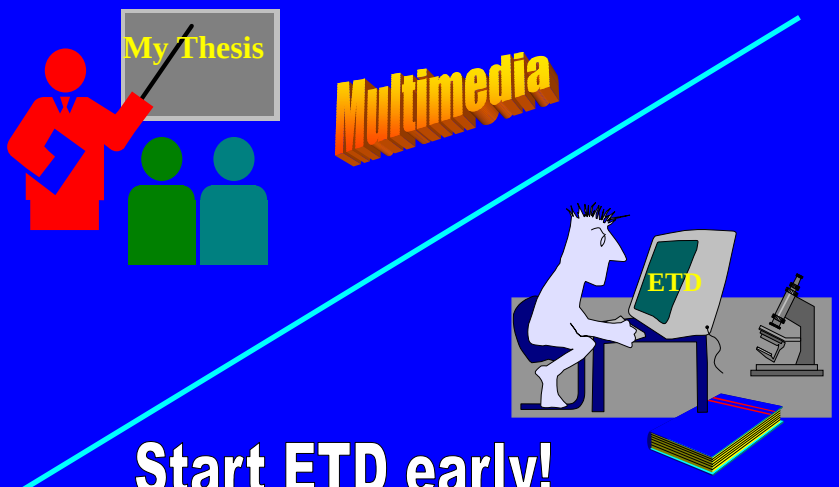
ETDs: Library Goals

- Improve library services
  - Better turn-around time
  - Always available
- Reduce work
  - catalog from e-text
  - eliminate handling: mailing to UMI, bindery prep, check-out, check-in, reshelving, etc.
- Save space

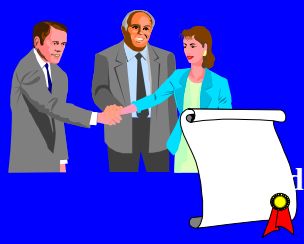
What are we doing?

- Aiding universities to enhance graduate education, publishing and IPR efforts
- Helping improve the availability and content of theses and dissertations
- Educating ALL future scholars so they can publish electronically and effectively use digital libraries (i.e., are Information Literate and can be more expressive)

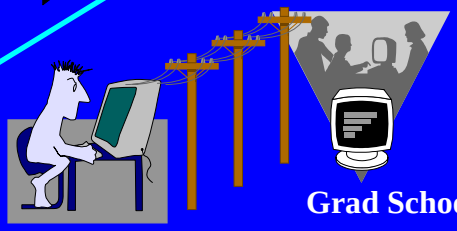
Student Defends & Finalizes ETD



Student Gets Committee Signatures and Submits ETD



Approval form



Grad School

Graduate School Approves ETD, Student is Graduated

Quality control

Library Catalogs ETD, Access is Opened to the New Research

Status of the Local Project

- Approved by university governance Spring 1996; required starting 1/1/97
- Submission & access software in place
- Submission workshops for students (and faculty) occur often: beginner/adv.
- Faculty training as part of Faculty Development Initiative
- Over 3000 ETDs in collection – some have audio, video, large images, software, …

Archiving ETDs

- Every 15 minutes back-ups made of not-yet-approved submissions
- Hourly back-ups of newly approved ETDs
- Weekly back-ups of entire ETD collection
- Copies stored on-site and off-site

Costs/Savings at VT

- Graduate School stopped shipping to the library 3000 copies of paper TDs/year
- Library stopped binding, shelving, and circulating 3000 copies of TDs/year
- 166 ft of shelf space saved/year by the library
- VT used existing equipment in Library (vs. start-up costs for staff, hardware and software from from a zero-base estimate: $65,000 – see http://scholar.lib.vt.edu/theses/)

Institutional Members

- **Cinemedia**
- **Coalition for Networked Information (CNI)**
- **Committee on Institutional Cooperation (CIC)**
- **Consorci de Biblioteques Universitàries de Catalunya**
- **Diplomica.com**
- **Dissertation.com**
- **Dissertationen Online (Germany)**
- **ETDweb, a Division of Answer4.com**
- **Ibero-American Science & Technology Education Consortium (ISTEC)**
- **National Documentation Centre (NDC), Greece**
- **National Library of Portugal (for all universities)**
- **OCLC Online Computer Library Center**
- **OhioLINK**
- **Organization of American States (SEDI/OAS)**
- **Southeastern Library Network (SOLINET)**
- **UNESCO (www.unesco.org/webworld/etd)**

National / Regional Projects

- **Australia**
  - U. New South Wales (lead)
  - U. of Melbourne
  - U. of Queensland
  - U. of Sydney
  - Australian National U.
  - Curtin U. of Technology
  - Griffith U.
- **Germany**
  - Humboldt University (lead)
  - 3 other universities
  - 5 learned societies: Math, Physics, Chemistry, Sociology, Education
  - 1 computing center
  - 2 major libraries
- OhioLINK: 79 colleges/univs
- Consorci de Biblioteques Universitàries de **Catalunya**, as group, www.cbuc.es:
  - Universitat de Barcelona
  - Universitat Autonòma de Barcelona
  - Universitat Politècnica de Catalunya
  - Universitat Pompeu Fabra
  - Universitat de Girona
  - Universitat de Lleida
  - Universitat Rovira i Virgili
  - Universitat Oberta de Catalunya
  - Biblioteca de Catalunya

US University Members (52)

- Air University (Alabama)
- Baylor University
- Brigham Young University (part, whole)
- Caltech
- Clemson University
- College of William & Mary
- Concordia University (Illinois)
- East Carolina University
- East Tenn. State U. – require fall 2000
- Florida Institute of Technology
- Florida International University
- George Washington University
- Louisiana State University
- Marshall University (W. Va.)
- Miami University of Ohio
- Michigan Tech
- Mississippi State University
- MIT
- Montana State University
- Naval Postgraduate School (CA)
- New Jersey Inst. of Technology
- New Mexico Tech
- North Carolina State University
- Northwestern University
- Penn. State University
- Regis University
- Rochester Institute of Tech.
- Texas A&M
- U. of Colorado Health Science Center
- U. of Florida
- U. of Georgia
- University of Hawaii, Manoa
- U. of Iowa
- U. of Kentucky
- U. of Maine
- U. of North Texas – required since 8/99
- U. of Oklahoma
- U. of Pittsburgh
- U. of Rochester
- U. of South Florida
- U. of Tennessee, Knoxville
- U. of Tennessee, Memphis
- U. of Texas at Austin – required in 2001
- U. of Virginia
- U. of West Florida
- U. of Wisconsin - Madison
- Vanderbilt U.
- Virginia Commonwealth U.
- Virginia Tech - required since 1/97
- West Virginia U. - required fall 1998
- Western Michigan U.
- Worcester Polytechnic Inst.

Other Countries - 52 Members

- Australia
- Belgium
- Brazil
- Canada
- China, Hong Kong
- Columbia
- Germany
- India
- Italy
- Korea
- Mexico
- Netherland
- Norway
- Russia
- Singapore
- S. Africa
- S. Korea
- Spain
- Sudan
- Sweden
- Taiwan
- UK

Type 1 Members

University Requires ETDs

- Adobe Acrobat and/or XML/SGML tools
- Automated submission & processing
- Archive/access through UMI, (OCLC,) Virginia Tech, ...
- (Local) WWW site, publicity
- (Local) Assistance provided as requested: email, phone, listserv(s)

Type 2 Members

University Agrees to Require ETDs

- Like Type 1 but set date not reached
- Usually has an option or pilot
- May: wait for new AY; start with all who enter after; …
- Build grass roots support
  - **Advisory committee: representative? expert?**
  - **Champions to spread by word of mouth**
  - **Approval: Senates, Commissions, Deans, Students**
  - **Publicity to reach community**

NDLTD Members, Types 3-7

- 3. Part of university requires ETDs
- 4. University allows ETDs
- 5. University investigating, has pilot
- 6. University consortium joins:
  - CIC (Big 10 coordinating body)
- 7. Non-university organization joins
  - CNI (Coalition for Networked Info.)

Counts of ETDs at Selected U's

University/Institution	ETD Collection size
ADT: Australian Digital Thesis Program (Australia)	238
University of Bergen (Norway)	45
California Institute of Technology	2
Consorci de Biblioteques Universitaries de Catalunya (Spain)	151
East Tennessee State University	106
Humboldt-University (Germany)	430
Louisiana State University	3
Mississippi State University	33
MIT	62
North Carolina State University	301
Pennsylvania State University	83
Pontifical Catholic University (PUC) (Brazil)	90
Gerhard Mercator Universitat Duisburg (Germany)	126
Universitat Politecnica de Valencia (Spain)	189
University of Florida	174
(continued)	

Counts of ETDs at Selected U's (cont'd)

University/Institution	ETD Collection size
University of Georgia	121
University of Iowa	6
University of Kentucky	19
University of Maine	27
University of North Texas	337
University of South Florida	25
University of Tennessee	12
University of Tennessee, Knoxville	28
Uppsala University (Sweden)	178
Virginia Tech	3393
West Virginia University	1006
Worcester Polytechnic Institute	83
TOTAL	**7268**

Counts of University Scanned ETD Collections

University/Institution	ETD Collection Size
MIT	5,581
National Documentation Center, Greece	12,000
New Jersey Institute of Technology	26
University of South Florida	150
TOTAL	**17,763**

VT ETD Access Logs

	1997/98	1998/99	Increase 1997/98 -1998/99	1999/00	Increase 1998/99 -1999/00
Requests for PDF files (mostly full ETDs)	221,679	481,038	117.0%	578,152	20.2%
Requests for HTML files (mostly tables of contents and abstracts)	165,710	215,539	30.1%	260,699	21.0%
Requests for multimedia	1,714	4,468	160.7%	12,633	182.7%
Distinct files requested	6,419	21,451	234.2%	16,409	-23.5%
Distinct hosts served	29,816	57,901	94.2%	87,804	51.6%
Average data transferred daily	156 MB	219 MB	40.4%	382 MB	74.4%
Data transferred	55 GB	78 GB	40.4%	137 GB	75.6%

VT ETD Access by Int'l Sites

International Domain	1997/98	1997/98 rank	1998/99	1998/99 rank	Increase 1997/98 -1998/99	1999/00	1999/00 rank	Increase 1998/99 -1999/00
United Kingdom	6,735	1	11,347	1	68.5%	25,583	1	125.5%
Malaysia	876	16	4,190	6	378.3%	16,147	2	285.4%
France	2,138	7	4,797	5	124.4%	14,960	3	211.9%
Germany	6,727	2	3,374	9	-49.8%	14,384	4	326.3%
Canada	3,413	4	9,632	3	182.2%	13,543	5	40.6%
Spain	590	18	3,647	8	518.1%	9,918	6	171.9%
Italy	1,430	12	3,095	10	116.4%	9,300	7	200.5%

Multimedia Use in ETD Collection

File type	Examples	Count
Still image	BMP, DXF, GIF, JPG, TIFF	328
Video	AVI, MOV, MPG, QT	58
Audio	AIFF, WAV	18
Text	PDF, HTML, TXT, DOC, XLS	7601
Other	Macromedia, SGML, XML	51

For professional societies

- Like “writing across the curriculum”, e.g., Chemical Markup Language, MathML, …
- Besides writing: computing/communications, information literacy, personal digital library management, tool use, research methods, collaboration, archiving/preservation
- Data sets, communities of users of them
- Classification systems / browsing / searching
- NRC’s “Issues for Science and Engineering Researchers in the Digital Age”, 57 pages

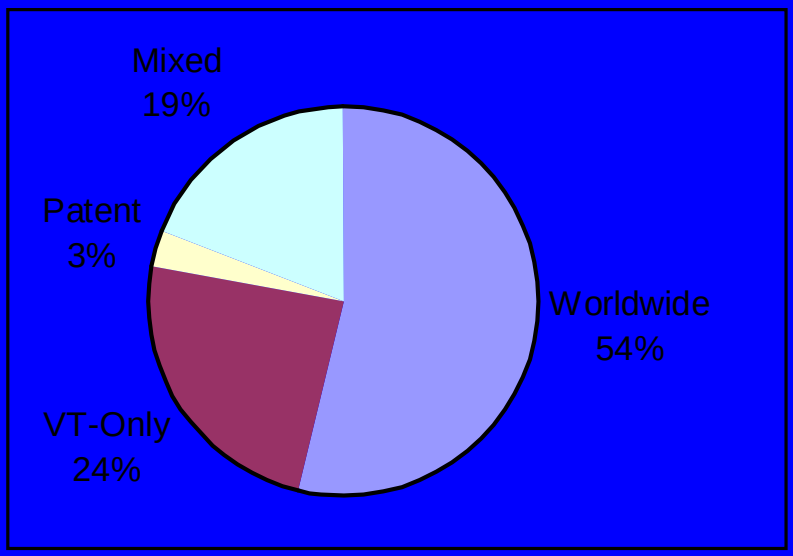
Relationship with publishers

- **Concern** of faculty and students that still wish to publish books or journal articles, voiced: campus, Chronicle, NPR, Times
- **Solution**: Approval Form gives students, faculty choices on access, when to change access condition; use IPR controls in DL
- **Solution**: by case, work with publishers and publisher associations to increase access
  - **AAP, AAUP**
  - **AAAS, ACM, ACS, Elsevier, ...**

Some responses from publishers

- **ACM**: need to acknowledge copyright
- **Elsevier**: need to acknowledge copyright
- **IEEE-CS**: endorse initiative
- **ACS**: After first publication, can release
- **Textbook publishers**: different market, manuscript significantly reworked
- **General**: restricting access to local campus will not cause any problems

Access Choices at VT (7/2000)



www.theses.org

- James Powell student project, D-Lib Magazine description in Sept. 1998
- XML description of each site
  - type of search engine / service
  - language
  - coverage (for resource discovery)
- Adding Z39.50 gateway capability and integrating with MARIAN, along with Harvest and Open Archives protocols

Access Approaches

- Goal: Maximize access and services, e.g., by encouraging:
- UMI centralized services
- VTLS: planned free union collection of metadata
- Distributed service: Dienst, Z39.50
- Regional services (e.g., OhioLink)
- Local servers with browse, search
  - From local catalogs to local archives
- WWW robot indexing and search services

Access Possibilities



Web search engines	www.theses.org	www.openarchives.org	library catalog clients	3rd Party Services (e.g., UMI)

Virginia Tech	MIT	National Library of Portugal	CBUC (Spain)	Ohio Link	National Projects: AU, GE, …

Why might a university want to be involved?

- To improve graduate education / better prepare your students / increase their knowledge and visibility
- To unlock university information
- To save money for students and for the university / improve workflow
- To build an important digital library

Multiple objectives

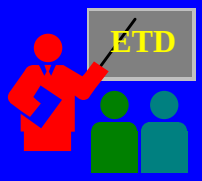
- **Sharing research results**
  - **Decrease costs, increase services**
  - **Increase knowledge of users**
- **Adding to author knowledge/skills**
  - **Epub, DL, IPR**
- **Enhancing organization's infrastructure**
  - **CS department, library**
  - **University, Laboratory**

How can a university get involved?

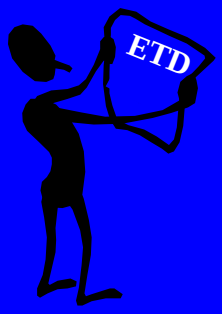
- Select planning/implementation team
  - **Graduate School**
  - **Library**
  - **Computing / Information Technology**
  - **Institutional Research / Educ. Tech.**
- Send us letter, give us contact names
  - **www.ndltd.org/join**
- Adapt Virginia Tech solution
  - **Build interest and consensus**
  - **Start trial / allow optional submission**

Convene Local Planning Group

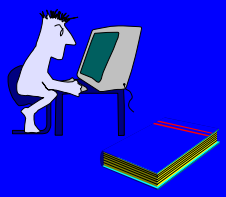
Build Local ETD Site



Workshop/Training



Policies



Digital Library

Inspection/Approval

NUDL

- 1/15/99 NUDL proposal to NSF under DLI2 international program, later redone as separate bilateral projects
  - Partners: Germany, Mexico (Puebla and Monterrey), Brazil
  - Problems: Multilingual search, multimedia submissions, requirements/usability, …
- Start with ETDs, then expand to other student works, portfolios, data sets, (CS) courseware, ...

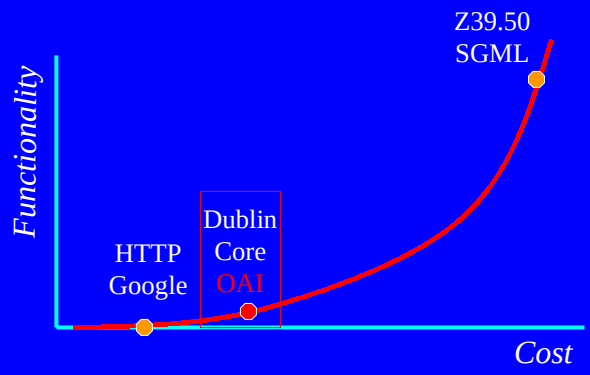
Open Archives Initiative

OAI

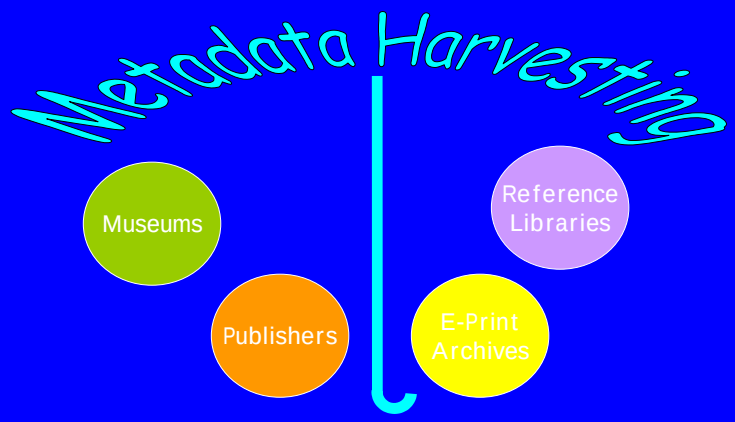
www.openarchives.org

openarchives@openarchives.org

Interoperability Trade-offs



Technical Umbrella for Practical Interoperability…



…that can be exploited by different communities

Open Archives Initiative (OAI)

- xxx@LANL, high-energy physics (Ginsparg, 1991)
- CSTR + WATERS = NCSTRL (Lagoze,1994)
- xxx + NCSTRL = CoRR collaboration (1998)
- Universal Preprint Service protoproto, Oct. 21-22, 1999, Santa Fe – led by LANL, CNI, DLF, Mellon --> OAi
- Santa Fe Convention (see Feb. D-Lib Magazine article)
- Follow-on mtgs: 6/3@San Antonio, 9/21@Lisbon (ECDL)
- Archives -> Open Archives
  - Support unique archive identifiers
  - Implement Open Archives metadata set (DC, using XML)
  - Implement OA harvesting protocol (derived from Dienst protocol)
  - Register the archive
- Build tools, layer other services: linking, searching, …

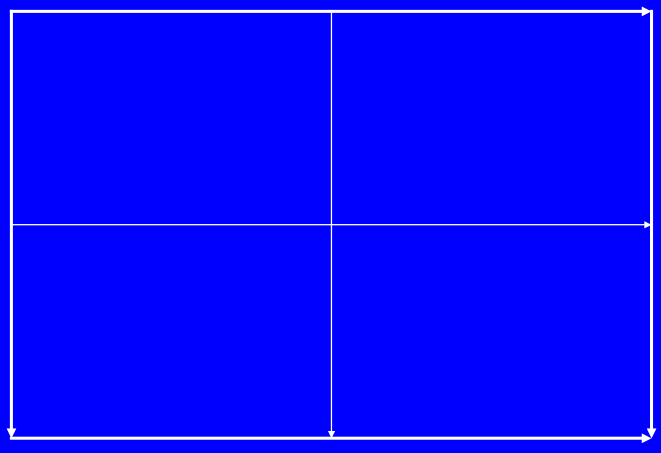
OAi Philosophy

- Self-archiving = submission mechanism
- Long-term storage system = archive
- Open interface = harvesting mechanism
- Data provider + service provider
- Start with "gray literature"
  - e-prints/pre-prints, reports, dissertations, …

Approaches to Open Archives

Build By Institution

Build By Discipline



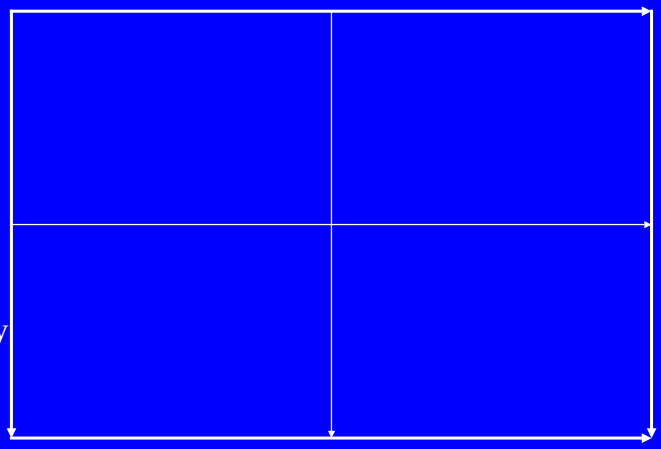
Approaches to Open Archives

Build By Institution

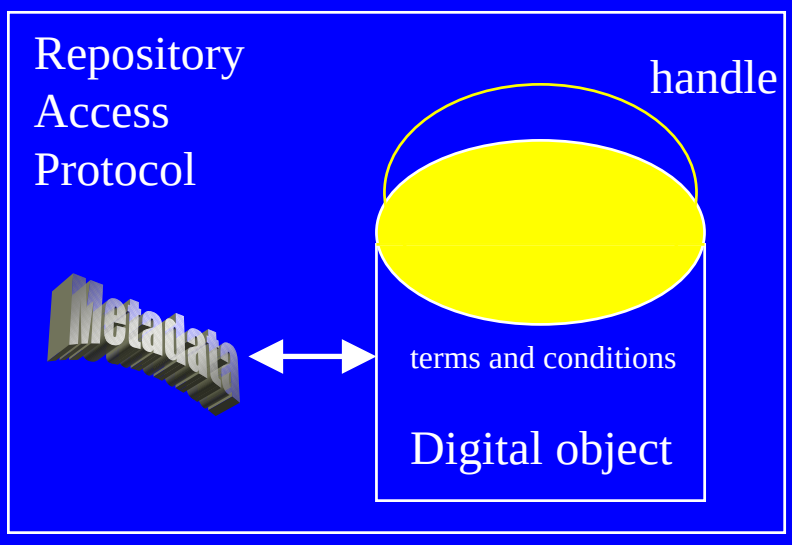
Build By Discipline

Access by

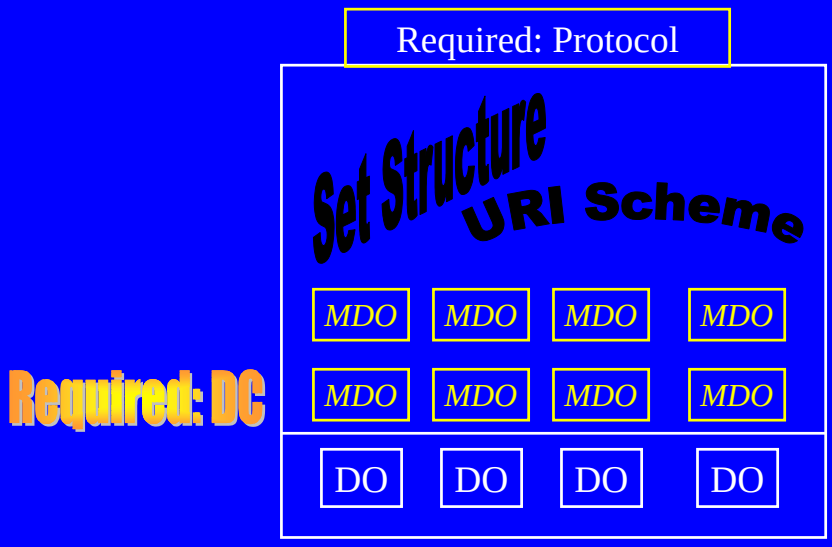
Author
Category
Interdisciplinary
Year
Language
Query …



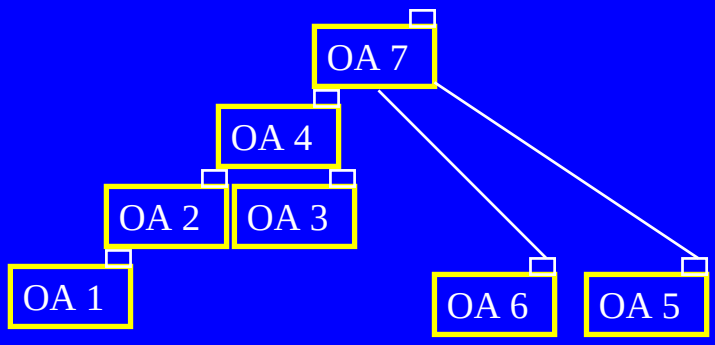
Repository of Digital Objects



OAI – Repository Perspective



OAI – Black Box Perspective



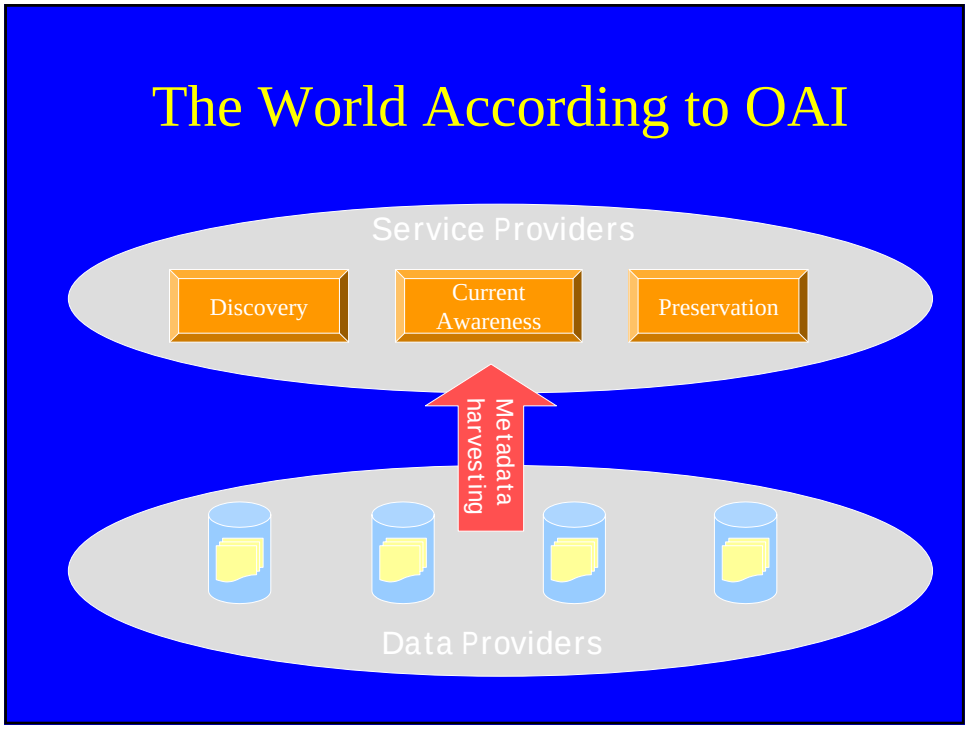


Figure 1. Layers Related to Open Archives Initiative

Services

Citation / Linking
SFX
CiteSeer

Authoring
Submission
Metadata Creation
Citation Checking

Editorial: Reviewing, Certification

Copy-Edit / Add Value
Text/MM Editing
Citation DB Updating
Authority Control
Preservation Conversion
Gazetteer
Cataloging

Summarization
Search/Browse
Annotation
Collaboration
• • •

Registry
Archives: Name, ID, Description, Terms and Conditions, …
Archive Formats: Name, Standard, Preservation Process, …
Metadata Formats: Name, XML DTD, …
Protocols
Tools
Services
• • •

Repository

Repository for NDLTD
Metadata Formats: OA Metadata Set, NDLTD Standard (DC-based) Set
Transaction Log
Training Resources

Open Archives Harvesting Protocol

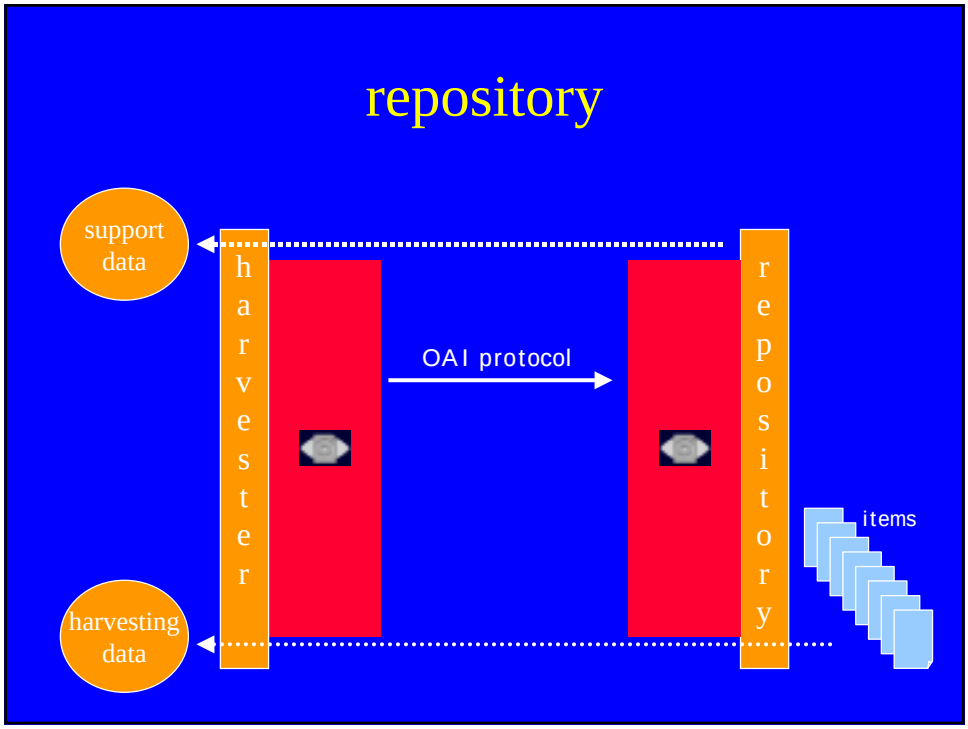
VT Partition
Record (Metadata)
Record (Full Content)

UVA Partition
Metadata | Content
• • •
• • •
Caltech Partition
Metadata | Content

NCSTRL Repository
EconWPA Repository
• • •
RePEc Repository

Key Features of the OAI Metadata Harvesting Protocol

- definitions & concepts
  - repository
  - record
  - identifier
  - datestamp
  - set
- protocol features
  - HTTP encoding
  - metadata prefix & schema
  - flow control
- protocol requests
  - supporting requests
  - harvesting requests



identifiers

locally unique key for extracting a record from a repository

oai-identifier = oai:archive-identifier:record-identifier

Registered URI Scheme

Archive Identifier: Registered within OAI

Unique ID within archive: (syntax is archive-specific)

example = oai:ncstrl:ncstrl.cornellcs/TR94-1418

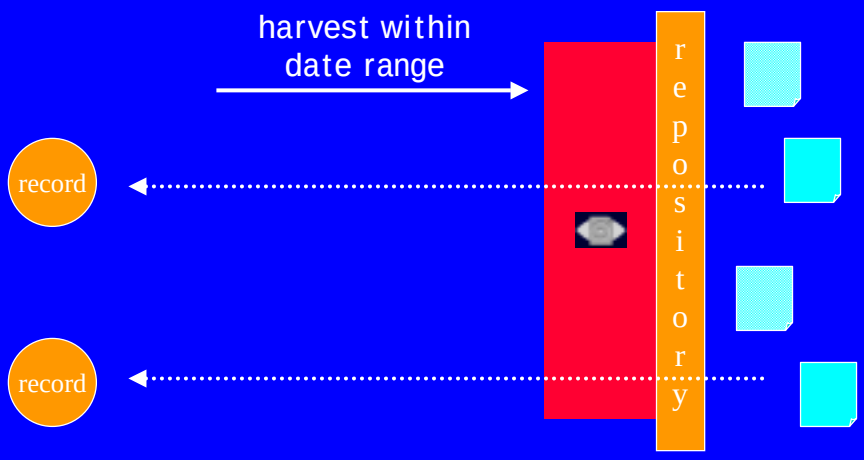
Request, Response – OAI, VT ETDs

Request

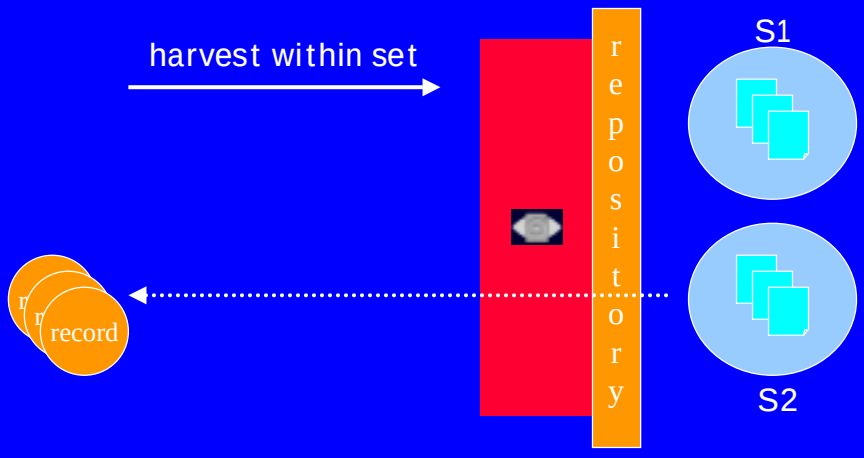
http://scholar.lib.vt.edu/theses/OAI/cgi-bin/index.pl?
verb=GetRecord&metadataPrefix=oai_etdms&identifier=oai:VTETD:etd-520112859651791

Response

selective harvesting - datestamps



selective harvesting - sets




Registering as a Data Provider
Consequences of Registration
Protocol Testing


Registered Data Providers

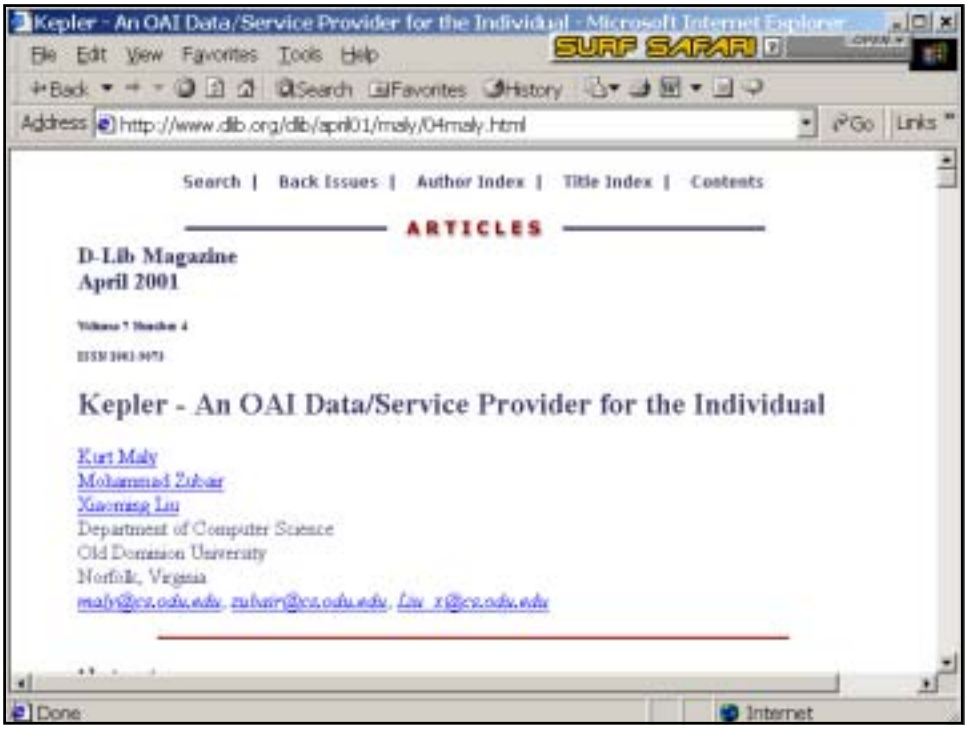
OAI Tools

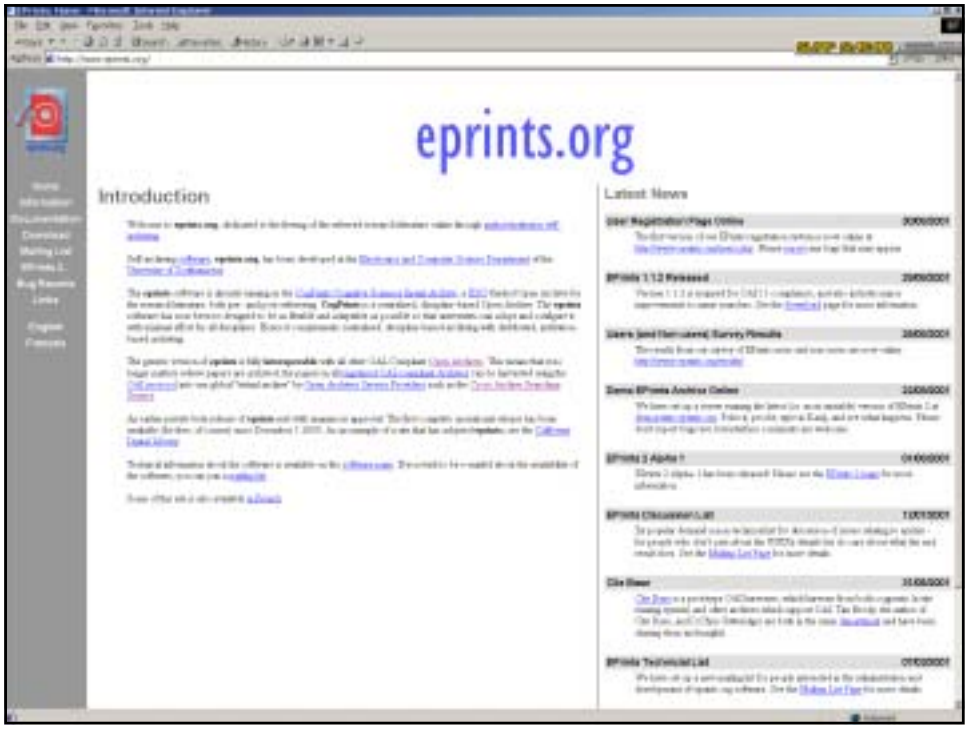
- Related resources, e.g., XML, Unicode
- Servers and utilities, e.g., ARC, Kepler, EPrints
- XML Schema Validator
- Repository Explorer
  - Interactive Browsing
  - Testing of parameters
  - Multiple views of data
  - Multilingual support
  - Automatic test suite

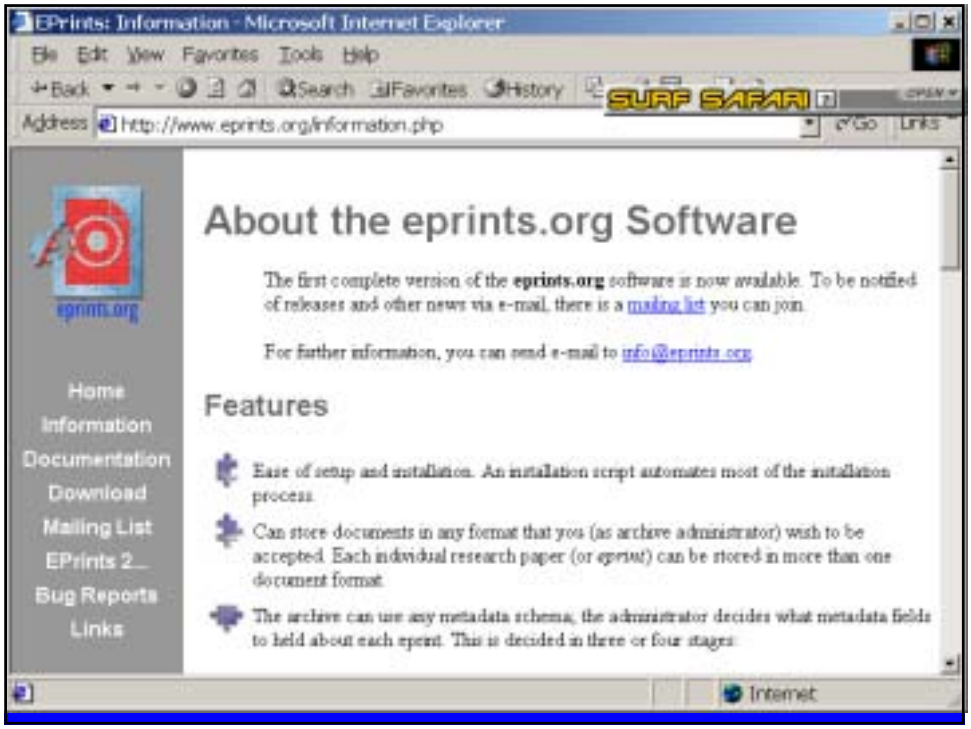
ARC (arc.cs.odu.edu)

ARC

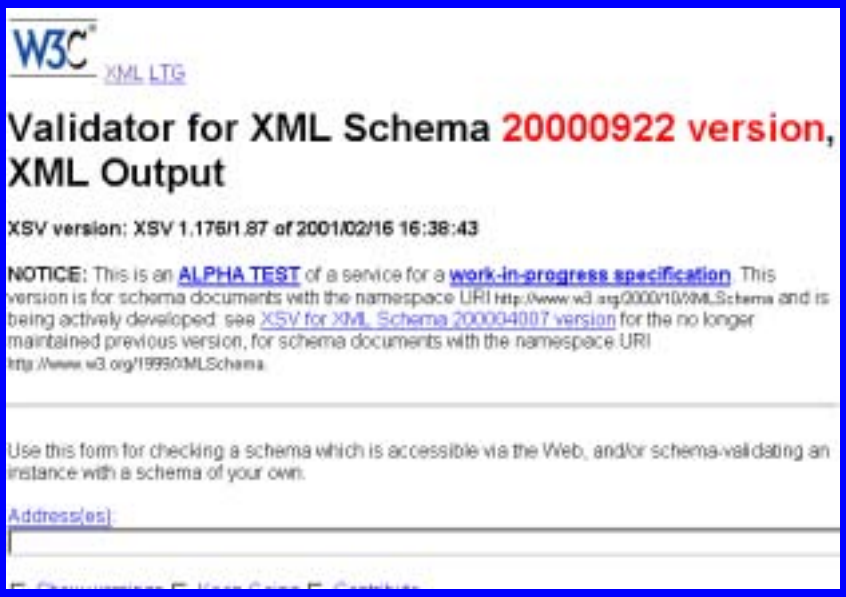

arc
Cross Archive Searching Service
Simple search Advanced Search Help
This is page 1, hits (1--10) of total 66 hits.
Result Pages: 1 2 3 4 5 6 7
SEARCH RESULTS
Title
Creators
Description
Archive
Document ID


Kepler - An OAI Data/Service Provider for the Individual - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.dlib.org/dlib/april01/maly/04maly.html
Search | Back Issues | Author Index | Title Index | Contents
ARTICLES
D-Lib Magazine
April 2001
Kepler - An OAI Data/Service Provider for the Individual
Kurt Maly
Mohammad Zubair
Xiaoming Liu
Department of Computer Science
Old Dominion University
Norfolk, Virginia
Done
Internet


eprints.org
Introduction
Latest News


EPrints: Information - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
SURF SAFARI
Address http://www.eprints.org/information.php
Go Links
eprints.org
Home
Information
Documentation
Download
Mailing List
EPrints 2...
Bug Reports
Links
About the eprints.org Software
The first complete version of the eprints.org software is now available. To be notified of releases and other news via e-mail, there is a mailing list you can join.
For further information, you can send e-mail to info@eprints.org
Features
Ease of setup and installation. An installation script automates most of the installation process.
Can store documents in any format that you (as archive administrator) wish to be accepted. Each individual research paper (or eprint) can be stored in more than one document format.
The archive can use any metadata schema; the administrator decides what metadata fields to hold about each eprint. This is decided in three or four stages.
Internet

XSV Schema Validator



OAI Repository Explorer

- Serves as a compliancy test
- Allows browsing of open archives using only OAI protocol
- Sends requests on behalf of user, parses and checks responses and displays browsable interface
- Will detect most discrepancies in protocol
- http://purl.org/net/explorer

RE 1.3

Participating in the OAI Community

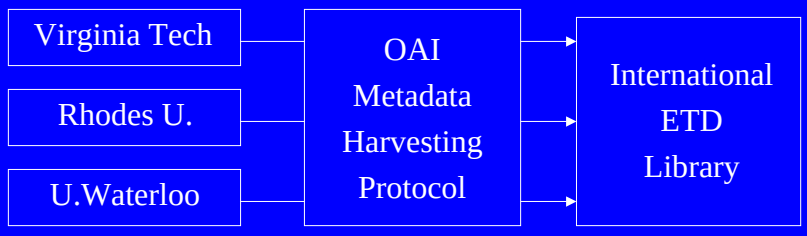
- Listservs
  - oai-general – discussion of OAI related issues
  - oai-implementers – sharing technical questions and agendas
- OAI website (www.openarchives.org)
  - Post news and links to OAI related activities
- Community-specific
  - How does OAI apply to your community?

Detailed Case Study: NDLTD

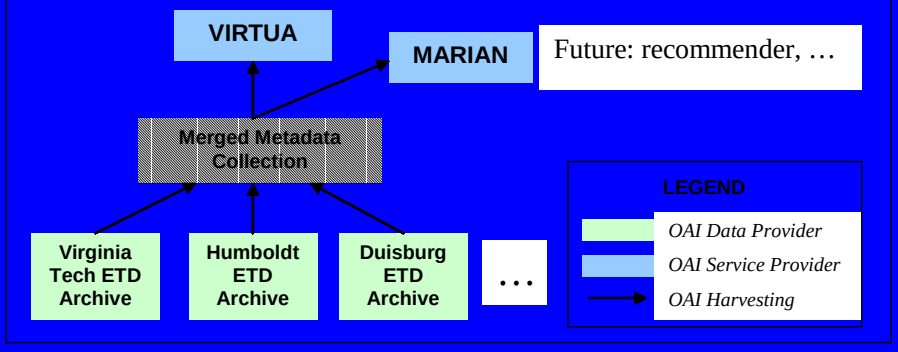
- Metadata: MARC21 (coded in XML), ETDMS (see www.ndltd.org/standards)
- Protocols in use: Z39.50, Harvest, Dienst, OAI, as well as http (web sites)
- OCLC's LAF (authority control) to work with RDF implementation of ETDMS
- Union collection -> VTLS's Virtua, Virginia Tech's MARIAN
- Phased efforts for development and testing over more than a year

Case Study: NDLTD

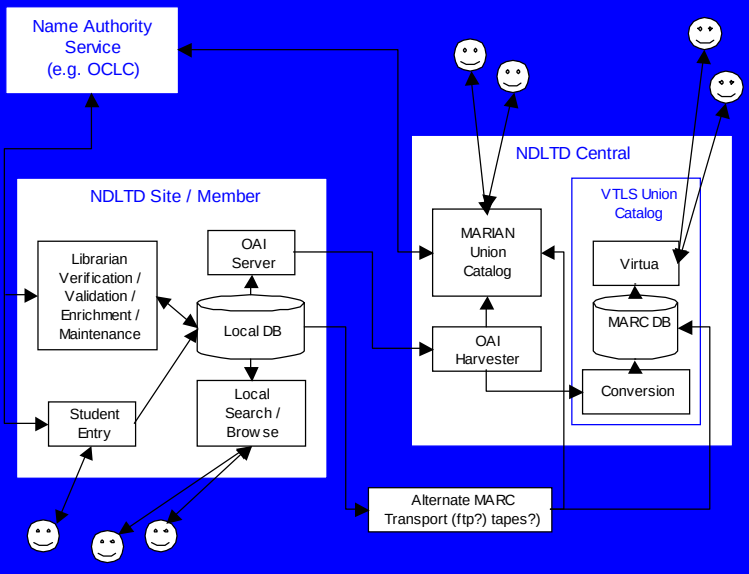
- Multiple independent university/library/ center-based collections of electronic documents



ETD Union Collection (OAI)



Union Catalog Creation



VTLS

- **VTLS will**
  - **Support NDLTD through a union catalog service implemented with Virtua**
  - **Accept metadata in MARC-21 or UNIMARC, and help identify other converters for other types**
  - **Accept metadata in one other format, namely ETD-XMS**
  - **Accept data in various character sets, with UNICODE preferred, but in some cases the submitter may be required to convert**

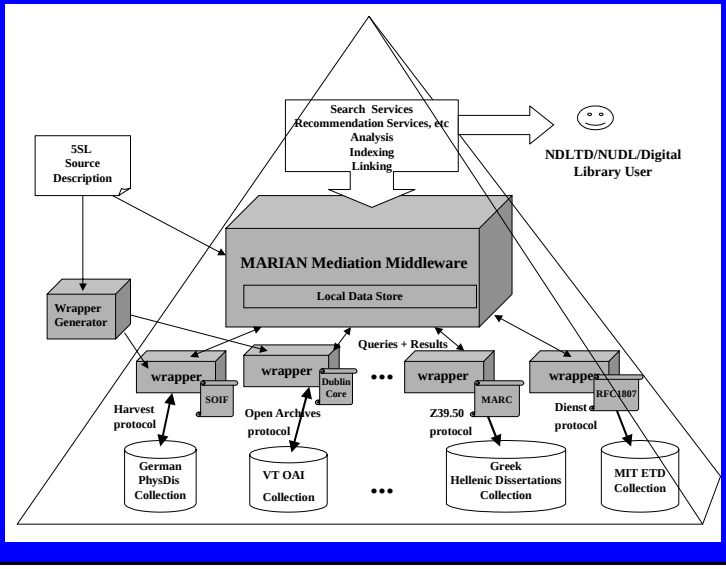
MARC XML-DTD

- XML Transport format for US-MARC records
- Standardized metadata exchange format for traditional library services joining OAI

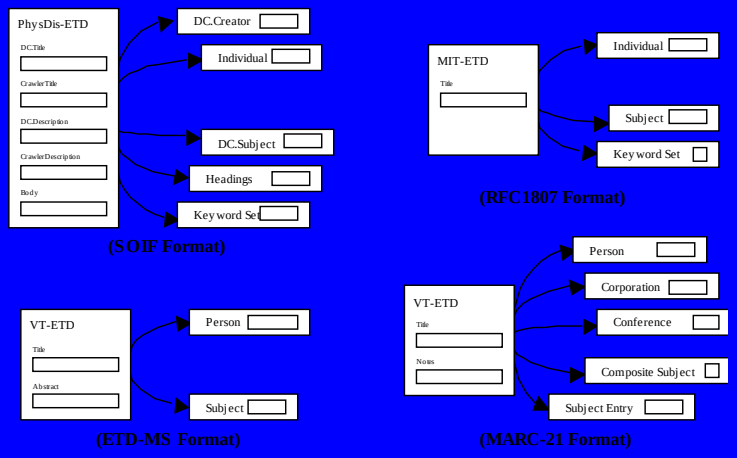
MARIAN

- Digital Library Search & Retrieval System
- Principles
  - Network representation
  - Class-based retrieval
  - Weight-valued functions and weighted sets
- Interoperability
  - System: wrappers and harvesting
  - Syntax: OAI standards (XML, Unicode, …)
  - Structure: information networks
  - Semantics: class-based retrieval
    : collection views

System & Syntactic Interoperability



Structural Interoperability through Information Networks



Community Assistance by OAI?

- Awareness
- Training
- Tools
- Test and validation
- Sharing experiences and solutions

Summary

- NDLTD
  - Status, JOIN!
  - ETDMS, Union catalog
- OAI
  - Philosophical and technical aspects
  - Community building / support