

[www.openarchives.org](http://www.openarchives.org)

Open Archives Initiative

OAI

openarchives@  
openarchives.org



“Opening Remarks & Historical  
Overview” - ACM SIGIR’2001

Ed Fox (w. Lagoze & Suleman):B

# 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 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History SURF SAFARI ? OPEN

Address <http://arc.cs.odu.edu/> Go Links

**arc** **Cross Archive Searching Service** 

Simple search Advanced Search Help

**Search all bibliographic fields**

Search for

Sort results by

Group results by

**search**

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.

*This prototype is based on the [UPS](#) project and the [NCSTRL+](#) based digital library developed by Old Dominion University*

Internet

# ARC

The screenshot shows a Netscape browser window titled "ARC - Netscape". The address bar is empty. The main content area displays the ARC logo (the letters "arc" in a stylized font) and the text "Cross Archive Searching Service". To the right is a logo for "OPEN ARCHIVES" featuring a stylized eye and a camera lens. Below this is a navigation bar with links for "Simple search", "Advanced Search", and "Help". A message states "This is page 1, hits (1--10) of total 66 hits." Below this are links for "Results Pages: 1 2 3 4 5 6 7". A section titled "SEARCH RESULTS" contains a table with the following information:

<b>Title</b>	<i>Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size to Convey Nominal and Quantitative Data</i>
<b>Creators</b>	Nowell, Lucille Terry
<b>Description</b>	centerH2Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size To Convey Nominal and Quantitative Data/h2/center centerH3Lucille Terry Nowell/H3/centerBCenterABSTRACT/Center/BpIn producing a user interface design to visualize search results for a digital library called En
<b>Archive</b>	NDLTD
<b>Document ID</b>	oai:VTETD:etd-111897-163723

At the bottom of the page, a footer note reads: "This prototype is based on the UPS project and the NCSTRL+ based digital library developed by Old Dominion University". The browser's status bar at the bottom shows "Document: Done" and various navigation icons.

[Search](#) | [Back Issues](#) | [Author Index](#) | [Title Index](#) | [Contents](#)

---

## ARTICLES

---

### D-Lib Magazine April 2001

Volume 7 Number 4

ISSN 1082-9873

## Kepler - An OAI Data/Service Provider for the Individual

[Kurt Maly](#)

[Mohammad Zubair](#)

[Xiaoming Liu](#)

Department of Computer Science  
Old Dominion University  
Norfolk, Virginia

[maly@cs.odu.edu](mailto:maly@cs.odu.edu), [zubair@cs.odu.edu](mailto:zubair@cs.odu.edu), [Liu\\_x@cs.odu.edu](mailto:Liu_x@cs.odu.edu)

---

# eprints.org



- Home
- Information
- Documentation
- Download
- Mailing List
- EPrints 2...
- Bug Reports
- Links
- English
- Français

## Introduction

Welcome to **eprints.org**, dedicated to the freeing of the refereed research literature online through [author/institution self-archiving](#).

Self-archiving software, **eprints.org**, has been developed at the [Electronics and Computer Science Department](#) of the [University of Southampton](#).

The **eprints** software is already running as the [CogPrints Cognitive Sciences Eprint Archive](#), a JISC-funded Open Archive for the research literature, both pre- and post-refereeing. **CogPrints** is a centralised, discipline-based Open Archive. The **eprints** software has now been re-designed to be as flexible and adaptable as possible so that universities can adopt and configure it with minimal effort for all disciplines. Hence it complements centralised, discipline-based archiving with distributed, institution-based archiving.

The generic version of **eprints** is fully **interoperable** with all other OAI-Compliant Open Archives. This means that it no longer matters *where* papers are archived, the papers in all registered OAI-compliant Archives can be harvested using the OAI protocol into one global "virtual archive" by [Open Archives Service Providers](#) such as the [Cross Archive Searching Service](#).

An earlier private beta release of **eprints** met with unanimous approval. The first complete operational release has been available (for free, of course) since December 1 2000. As an example of a site that has adopted **eprints**, see the [California Digital Library](#).

Technical information about the software is available on the [software page](#). If you wish to be e-mailed about the availability of the software, you can join a [mailing list](#).

Some of this site is also available in [French](#).

## Latest News

### User Registration Page Online 30/08/2001

The first version of our EPrints registration system is now online at <http://www.eprints.org/users.php>. Please [report any bugs](#) that may appear.

### EPrints 1.1.2 Released 29/08/2001

Version 1.1.2 is required for OAI 1.1 compliance, and also includes minor improvements to name searches. See the [download page](#) for more information.

### Users (and Non-users) Survey Results 29/08/2001

The results from our survey of EPrints users and non-users are now online: <http://www.eprints.org/results/>

### Demo EPrints Archive Online 22/08/2001

We have set up a server running the latest (i.e. most unstable) version of EPrints 2 at [demo.eprints.org](http://demo.eprints.org). Poke it, prod it, type in Kanji, and see what happens. Please don't report bugs yet, but interface comments are welcome.

### EPrints 2 Alpha 1 01/08/2001

EPrints 2 Alpha-1 has been released! Please see the [EPrints 2 page](#) for more information.

### EPrints Discussion List 12/07/2001

By popular demand a non-technical list for discussion of issues relating to eprints - for people who don't care about the UNIXy details but do care about what the end result does. See the [Mailing List Page](#) for more details.

### Cite Base 31/05/2001

Cite Base is a prototype OAI harvester, which harvests from both [cogprints](#) (a site running eprints) and other archives which support OAI. Tim Brody, the author of Cite Base, and I (Chris Gutteridge) are both in the same department and have been sharing ideas and insights.

### EPrints Technical List 07/03/2001

We have set up a new mailing list for people interested in the administration and development of eprints.org software. See the [Mailing List Page](#) for more details.



Welcome to **CogPrints**, an electronic archive for papers in any area of **Psychology**, **Neuroscience**, and **Linguistics**, and many areas of **Computer Science** (e.g., artificial intelligence, robotics, vision, learning, speech, neural networks), **Philosophy** (e.g., mind, language, knowledge, science, logic), **Biology** (e.g., ethology, behavioral ecology, sociobiology, behaviour genetics, evolutionary theory), **Medicine** (e.g., Psychiatry, Neurology, human genetics, Imaging), **Anthropology** (e.g., primateology, cognitive ethnology, archeology, paleontology), as well as any other portions of the physical, social and mathematical sciences that are pertinent to the study of cognition.






- 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](mailto: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:



English  
Français

1. Decide a maximal set of metadata fields that should be stored (for example, "authors", "title", "journal", "journal volume", etc.)
2. Decide what types of eprint should be stored (for example, refereed journal article, thesis, technical report, unpublished preprint)
3. For each eprint type, decide which metadata fields should be stored for eprints of that type, and which of those fields are mandatory.
4. Decide how these metadata fields should be projected into the Open Archives world. (If necessary, interoperability can be switched off, but this is strongly discouraged.)



Eprints can be placed in a configurable, extendible subject hierarchy. This hierarchy can be used to view and search the archive.



Submission of all eprints is via a simple but extremely powerful WWW-based interface. Papers can be uploaded just as files, in a compressed bundled file (such as a .zip file), or automatically mirrored from an existing website by specifying a URL.





Data integrity checks are performed automatically without the need for administrator intervention. Some "stub" routines allow individual sites to add their own integrity checks if they desire. These checks can greatly reduce the workload of the site administrator.



Authors can have associated metadata. Again, the site administrator can decide what metadata fields will be stored with each author's record.

 Users can subscribe either as authors or readers, via a web form or an automatically processed e-mail account.


 Submitted papers (if administrators desire) go through a moderation process. Submitted papers are placed in a buffer, where they can be approved by a moderator, rejected outright, or returned to the author for amendment. (The extent of this moderation is, of course, up to the individual institution running the archive: at [CogPrints](#) we use it to make sure the subject matter is suitable, and that all figures appear correctly.)

 Moderation process is also performed using a WWW-based interface. In fact, once the site is up and running, it will require very little maintenance that cannot be performed using the WWW-based maintenance interface.

## Requirements

A major objective in this work has been to make the system usable at the minimum possible expense. The only cost of running the archive need be the basic cost of the hardware. The hardware and software requirements are:

 Any computer capable of running UNIX. Obviously, the faster, the better, but any Intel Pentium II processor will give good performance.

 A TINY operating system. *Linux* (a most advanced and free TINY implementation)

# XSV Schema Validator



## Validator for XML Schema 20000922 version, XML Output

XSV version: XSV 1.176/1.87 of 2001/02/16 16:38:43

**NOTICE:** This is an **ALPHA TEST** of a service for a **work-in-progress specification**. This version is for schema documents with the namespace URI <http://www.w3.org/2000/10/XMLSchema> and is being actively developed: see [XSV for XML Schema 200004007 version](#) for the no longer maintained previous version, for schema documents with the namespace URI <http://www.w3.org/1999/XMLSchema>.


---

Use this form for checking a schema which is accessible via the Web, and/or schema-validating an instance with a schema of your own.

Address(es):

Show warnings  Keep Going  Contribute

# RE 1.3



## Open Archives Initiative - Repository Explorer

*explorer version - 1.3 : protocol version - 1.0/1.1 : August 2001*

This site presents an interface to interactively test archives for compliance with the OAI Protocol for Metadata Harvesting [ [Click here for details](#) ]

JavaScript is required

Note: To avoid HTTP errors, please wait for each page to finish loading before clicking on any link.

Please enter the URL to the OAI interface (everything before the ?) or choose a predefined archive from the table

American Memory [LoC]
arXiv
JCDL Pictures
LTRS [NASA]

[ [View Archive Website](#) ] [ [Test and Add an archive to this list](#) ]

Verbs	Parameters
<a href="#">Identify</a>	from (YYYY-MM-DD) : <input type="text"/>
<a href="#">List Metadata Formats</a>	until (YYYY-MM-DD) : <input type="text"/>
<a href="#">List Sets</a>	metadataPrefix : <input type="text"/>
<a href="#">List Identifiers</a>	identifier : <input type="text"/>
<a href="#">List Records</a>	set : <input type="text"/>
<a href="#">Get Record</a>	resumptionToken : <input type="text"/>

# RE Interactive Browsing



## Open Archives Initiative - Repository Explorer

*explorer version - 1.1 : protocol version - 1.0 : April 2001*

This site presents an interface to interactively test archives for compliance with the OAI Protocol for Metadata Harvesting [ [Click here for details](#) ]

JavaScript is required and Netscape Navigator is recommended

Note: To avoid HTTP errors, please wait for each page to finish loading before clicking on any link.

---

Please enter the URL to the OAI interface (everything before the ?) or choose a predefined archive from the table

- American Memory [LoC]
- Digital Cultural Heritage Community [UIUC]
- Perseus [Open Language Archives]
- RDN [Resource Discovery Network]

[ [View Archive Website](#) ] [ [Test and Add an archive to this list](#) ]

# RE Parameter Testing

Verbs	Parameters	
<a href="#">Identify</a>	from (YYYY-MM-DD) :	<input type="text"/>
<a href="#">List Metadata Formats</a>	until (YYYY-MM-DD) :	<input type="text"/>
<a href="#">List Sets</a>	<u>metadataPrefix</u> :	<input type="text"/>
<a href="#">List Identifiers</a>	identifier :	<input type="text"/>
<a href="#">List Records</a>	set :	<input type="text"/>
<a href="#">Get Record</a>	resumptionToken :	<input type="text"/>
Language	Display	Schema Validation
<input type="text" value="English"/>	<input checked="" type="radio"/> Parsed	<input type="radio"/> None
	<input type="radio"/> Raw XML	<input checked="" type="radio"/> Local mirror of schemata
	<input type="radio"/> Both	<input type="radio"/> Online schemata
<a href="#">home</a>	Send all comments to <a href="mailto:hussein@vt.edu">hussein@vt.edu</a> --- <a href="#">Digital Libraries Research Laboratory@Virginia Tech</a>	

# RE Browsing

## Archive Self-Description

<b>Repository Name</b>	Virginia Tech Electronic Thesis and Dissertation Collection
<b>Base URL</b>	<a href="http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTETD.pl">http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTETD.pl</a>
<b>Protocol Version</b>	1.0
<b>Admin Email</b>	<a href="mailto:webmaster@scholar.lib.vt.edu">mailto:webmaster@scholar.lib.vt.edu</a>
<b>Other Information</b>	<pre>description:   oai-identifier:     scheme: oai     repositoryIdentifier: VTETD     delimiter: :     sampleIdentifier: oai:VTETD:etd-171110282975860 description:   eprints:     content:       text: Theses and Dissertations produced by students at Virginia metadataPolicy:   text: Metadata may be used by commercial and non-commercial user dataPolicy:   text: Full texts are individually tagged and the rights statemer</pre>

# RE Browsing



## Open Archives Initiative - Repository Explorer

*explorer version - 1.1 : protocol version - 1.0 : April 2001*

<http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTETD.pl?verb=ListSets>

---

### List of Sets

*Click on the link to list the contents*

[All theses and dissertations](#)

---

**Request URL :** <http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTETD.pl?verb=ListSets>

**Response Date :** 2001-06-14T20:27:00-05:00

---

Verbs	Parameters
-------	------------

from (YYYY-MM-DD) :



# RE Browsing



## Open Archives Initiative - Repository Explorer

*explorer version - 1.1 : protocol version - 1.0 : April 2001*

<http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTETD.pl?verb=ListIdentifiers&set=All>

### List of Record Identifiers

*Select a link to view more information*

[oai:VTETD:etd-3345131939761081](#) [[display record in Dublin Core](#)] [[display metadata format](#)]  
[oai:VTETD:etd-171110282975860](#) [[display record in Dublin Core](#)] [[display metadata format](#)]  
[oai:VTETD:etd-05012000-14030054](#) [[display record in Dublin Core](#)] [[display metadata form](#)]  
[oai:VTETD:etd-3621112139711101](#) [[display record in Dublin Core](#)] [[display metadata forme](#)]  
[oai:VTETD:etd-133422039701091](#) [[display record in Dublin Core](#)] [[display metadata format](#)]  
[oai:VTETD:etd-23281533974920](#) [[display record in Dublin Core](#)] [[display metadata formats](#)]  
[oai:VTETD:etd-123322282975860](#) [[display record in Dublin Core](#)] [[display metadata format](#)]  
[oai:VTETD:etd-255314202974780](#) [[display record in Dublin Core](#)] [[display metadata format](#)]  
[oai:VTETD:etd-335713312971890](#) [[display record in Dublin Core](#)] [[display metadata format](#)]  
[oai:VTETD:etd-104722369631841](#) [[display record in Dublin Core](#)] [[display metadata format](#)]

# RE Browsing



## Open Archives Initiative - Repository Explorer

*explorer version - 1.1 : protocol version - 1.0 : April 2001*

<http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTETD.pl?verb=ListMetadataFormats&identifier=oai:VTE>

### List of Metadata Formats

*Click on the link to view schema*

Prefix=[[oai\\_rfc1807](#)]

Namespace=[<http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1807.txt>]

Schema=[<http://oai.dlib.vt.edu/OAI/rfc1807.xsd>]

[[display record](#)]

Prefix=[[oai\\_marc](#)]

Namespace=[[http://www.openarchives.org/OAI/oai\\_marc](http://www.openarchives.org/OAI/oai_marc)]

Schema=[[http://oai.dlib.vt.edu/OAI/oai\\_marc.xsd](http://oai.dlib.vt.edu/OAI/oai_marc.xsd)]

# RE Browsing

[http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTETD.pl?verb=GetRecord&metadataPrefix=oai\\_etdms&i](http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTETD.pl?verb=GetRecord&metadataPrefix=oai_etdms&i)

## List of Fields

### header:

identifier : oai:VTETD:etd-3345131939761081  
datestamp : 1997-03-31

### metadata:

#### thesis:

title: Conceptual Development and Empirical Testing of an Outdoor Recreation Experience Preference Scale  
creator: Walker, Gordon James  
publisher: Virginia Polytechnic Institute and State University  
subject: outdoor recreation  
subject: recreation experience preference scales  
subject: recreation experience matrix  
subject: recreation opportunity spectrum  
description: This dissertation examines four issues, including: (a) whether outdoor recreation experience preference (REP) scales exist; (b) whether these experiences can be measured using a framework called the Recreation Experience Matrix (REM); (c) how well the Recreation Opportunity Spectrum (ROS) variables of activity, setting, and expertise explain the types of experiences outdoor recreationists receive; and (d) how well two new variables--primary mode and mode dependence--explain the types of experiences outdoor recreationists receive. In order to address these issues, a questionnaire was distributed at Mount Rogers National Recreation Area in Virginia during October and November 1996. A total of 410 people completed this questionnaire. Of these, 336 provided useable data.

# RE Multiple views of data

1-3345131939761081

## Raw XML Output

```
<?xml version="1.0" encoding="UTF-8"?>

<ListSets xmlns="http://www.openarchives.org/OAI/1.0/OAI_ListSets" xmlns:

<responseDate>2001-06-14T20:32:33-05:00</responseDate>
<requestURL>http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTETD.pl?ver

<set>
  <setSpec>All</setSpec>
  <setName>All theses and dissertations</setName>
</set>

</ListSets>
```

31939761081

# RE Multilingual Support



Open Archives Initiative - 揣繼坪霜  
擬

霜擬 曉掛 - 1.1 : ?? 曉掛 - 1.0 : 2001 爛 血坡

掛桴莢咯玲蹀蝠誑宅聆松郗志Harvesting 坳鉤OAI ??紫偶腔賜醞 [ 割剝載嗣跨考 奠僻瀧爵 ]

割剝猛厥JavaScript 標祐松斬Netscape 霜擬

蛸碑: 咯旌轄 HTTP 渣响, 脹善藩玲瑤狙婁賦吁繼姪奠僻蟀諉

怀 善OAI賜醞腔URL(? 坪衄轉 ) 麼植蹈桶篋恁憲玲蹀悖珂隅砵腔紫偶

- American Memory [LoC]
- Digital Cultural Heritage Community [UIUC]
- Perseus [Open Language Archives]
- RDN [Resource Discovery Network]

[ 脹艘紫偶厘桴 ][ 聆松甜樓 玲蹀紫偶善蹈桶篋 ]

雄裸

統杆

# RE Automatic Test Suite



## Open Archives Initiative - Repository Explorer

*explorer version 1.1 : protocol version 1.0 : April 2001*

Open Archives Initiative :: Metadata Harvesting Protocolv1.0  
Protocol Tester v1.1 :: Virginia Tech DLRL :: April 2001

Testing : Identify

URL : <http://oai.dlib.vt.edu/~hussein/cgi-bin/NDLTD/VTETD.pl?verb=Identify>

Test Result : OK

---- [ Repository Name = Virginia Tech Electronic Thesis and Dissertation Collection ]

---- [ Protocol Version = 1.0 ]

---- [ Base URL = <http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl> ]

---- [ Admin Email = <mailto:hussein@vt.edu> ]

Testing : ListMetadataFormats

URL : <http://oai.dlib.vt.edu/~hussein/cgi-bin/NDLTD/VTETD.pl?verb=ListMetadataFormats>

Test Result : OK

---- [ Sample Metadata Format = oai\_rfc1807 ]

Testing : ListSets

URL : <http://oai.dlib.vt.edu/~hussein/cgi-bin/NDLTD/VTETD.pl?verb=ListSets>

Test Result : OK

---- [ Sample Set Name = All ]

# RE Error in Response

## Archive Self-Description

<b>Repository Name</b>	Virginia Tech Electronic Thesis and Dissertation Collection
<b>Base URL</b>	http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTDErr1/VTETD.pl
<b>Protocol Version</b>	1.0
Error: Missing field : <Identify>/<adminEmail>	
<b>Other Information</b>	<pre>description:   oai-identifier:     scheme: oai     repositoryIdentifier: VTETD     delimiter: :     sampleIdentifier: oai:VTETD:etd-171110282975860 description:   eprints:     content:       text: Theses and Dissertations produced by students at Virginia metadataPolicy:   text: Metadata may be used by commercial and non-commercial user dataPolicy:   text: Full texts are individually tagged and the rights statemer</pre>

# RE Error in XML



explorer version - 1.1 : protocol version - 1.0 : April 2001

<http://oai.dlib.vt.edu/~hussein/cgi-bin/NDLTDerr1/VTETD.pl?verb=Identify>

## XSD Schema/Instance Validation Error !

Errors in XML instance

```
<?xml version='1.0'?>
<xsv docElt='{http://www.openarchives.org/OAI/1.0/OAI_Identify}Identify' instanceAsses
<importAttempt URI='http://oai.dlib.vt.edu/OAI/1.0/OAI_Identify.xsd' namespace='http://
<importAttempt URI='http://oai.dlib.vt.edu/OAI/oai-identifier.xsd' namespace='http://w
<importAttempt URI='http://oai.dlib.vt.edu/OAI/eprints.xsd' namespace='http://www.oper
<invalid char='4' code='cvc-complex-type.1.2.4' line='11' resource='file:///tmp/file2V
<fsm>
<node id='1'>
<edge dest='2' label='{http://www.openarchives.org/OAI/1.0/OAI_Identify}:responseDate'
</node>
<node id='2'>
<edge dest='3' label='{http://www.openarchives.org/OAI/1.0/OAI_Identify}:requestURL' />
</node>
<node id='3'>
```



# Implementation Utilities

- Protocol handlers
  - OCLC
  - Virginia Tech
  - UIUC
- Metadata Utilities
  - MARC to DC (OCLC, Virginia Tech, ...)
- [eprints.org](http://eprints.org)

# 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](http://www.openarchives.org))
  - Post news and links to OAI related activities
- Community-specific
  - How does OAI apply to your community?

# Externally funded initiatives

- European Community
  - Open Archives Forum
  - Cyclades Project
- Andrew W. Mellon Foundation
  - Funding for 7 service providers
- Digital Library Federation
  - Gateways for access to member's digital collections
- National Science Foundation
  - NSDL ([www.nsd1.nsf.gov](http://www.nsd1.nsf.gov)) Core Infrastructure
  - Virginia Tech awards IIS-9986089, 0086227, 0080748 with joint funding by DFG (Germany), CONACyT (Mexico)

# Where do we go from here 2001-2002

- Controlling the stampede
- Technical re-evaluation leading to "final" 2.0 specification
  - OAI Technical Committee
- Strategy for standardization
- Community building focused on verification and validation

# Case Study: NSDL

- National Science, mathematics, engineering and technology education Digital Library (NSF)
- Urgent need: “doors open” Oct. 2002
- Core integration track: building on collections
- Collections track: building portals and centralized repositories (metadata, learning objects / educational resources)
- Metadata: DC, LOMS/IMS
- Problems: will publishers share metadata? Will those with small repositories adopt OAI?

# Case Study: NCSTRL

- CSTR and WATERS -> NCSTRL
  - Federated search of regular sites, harvesting of lite sites
- Changes: disinterest in central service, decline in interest in dept report series, increase in interest in personal web pages (ACM allowance)
- Kepler to support personal Open Archives
- Shift from Dienst-based service to OAI-based service underway in Fall 2001 (aided by Virginia's Internet Technology Innovation Center, through ODU, UVA, and Virginia Tech – along with others)

# Case Study: SOLINET

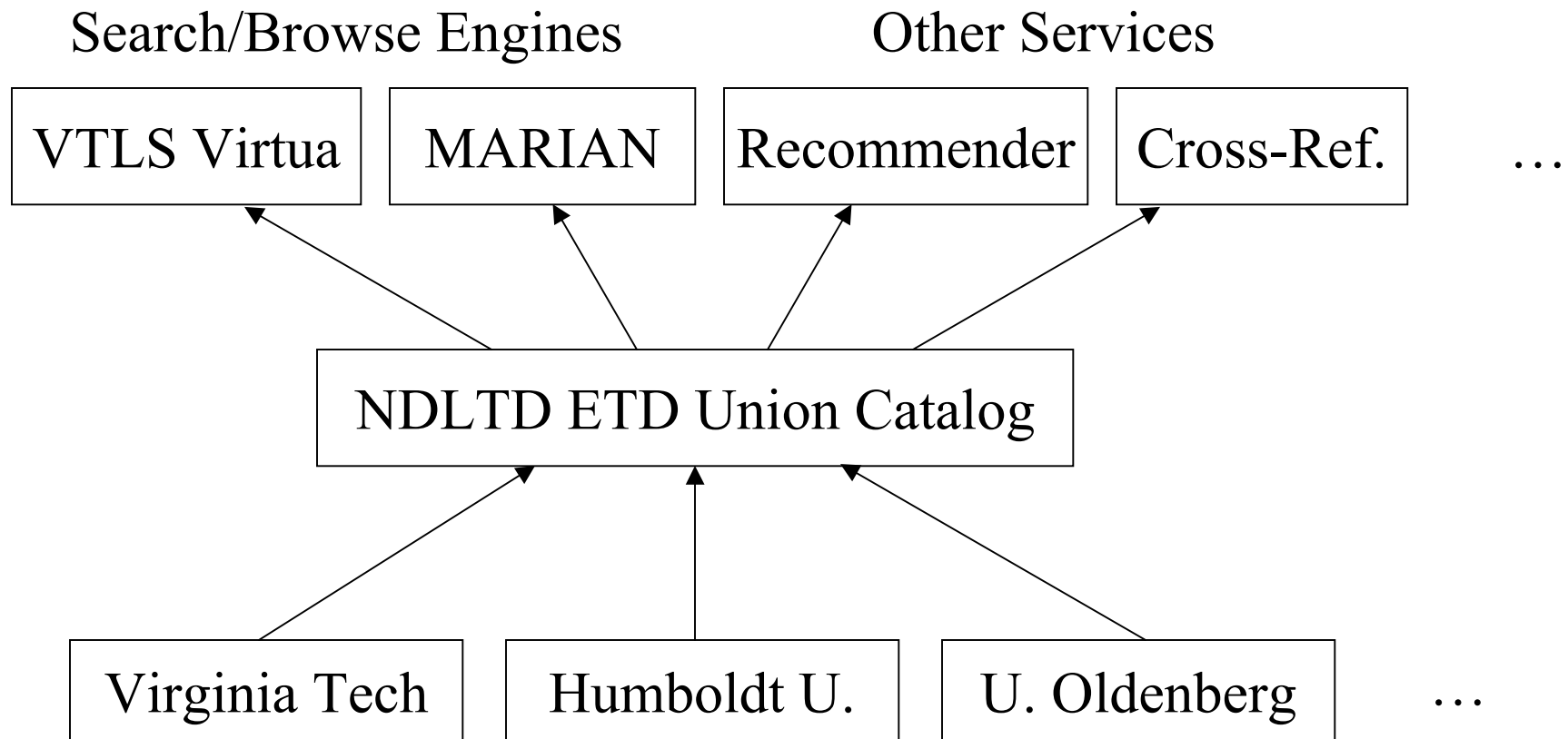
- Mellon Foundation
- SOuttheastern Library NETwork (Atlanta)
- Deadline: February 2003
- 10+ univ. collections about American South
- Scholars to learn about OAI, decide how to apply, work toward controlled vocabulary
- Harvesting to central site
- New central DL services (to be developed)

# Detailed Case Study: NDLTD

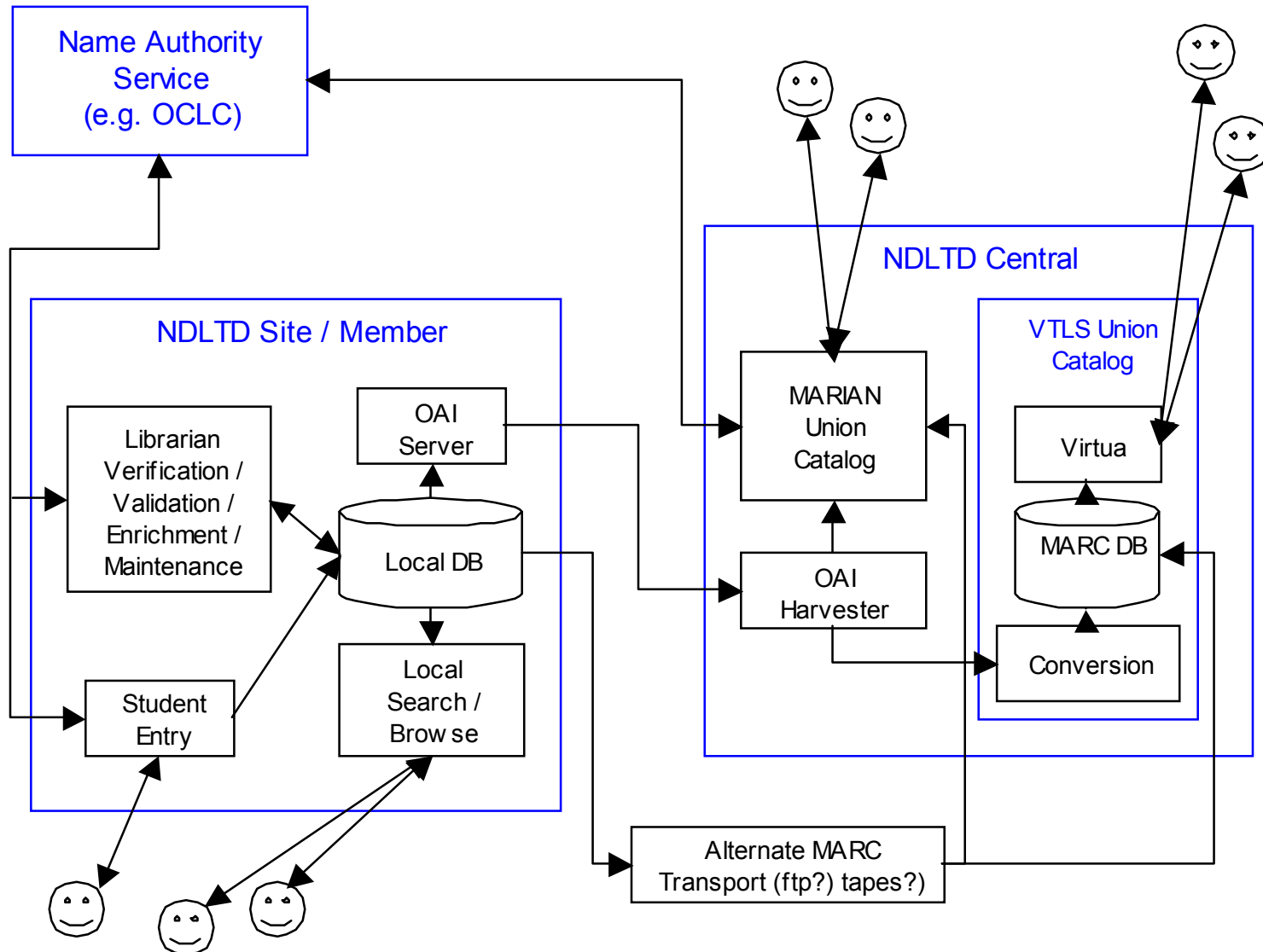
- Metadata: MARC21 (coded in XML), ETDMS (see [www.ndltd.org/standards](http://www.ndltd.org/standards))
- Protocols in use: Z39.50, Harvest, Dienst, OAI, as well as http (web sites)
- OCLC's LAF (authority control) to work with RDF implementation of ETDMS
- Union collection -> VTLS's Virtua, Virginia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



# Case Study: NDLTD Services



# 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**

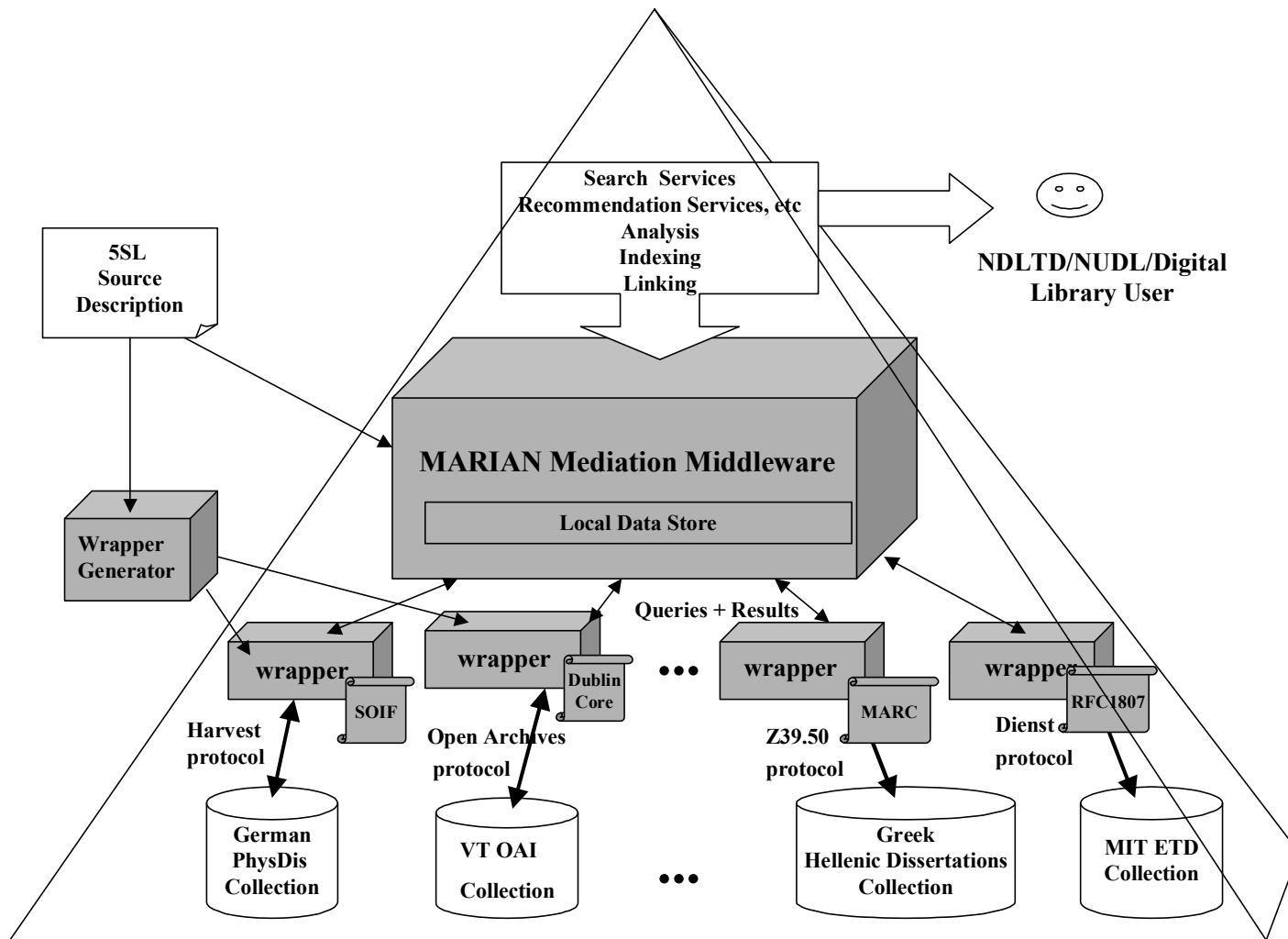
# MARIAN

- Digital Library Search & Retrieval System
- Principles
  - Network representation
  - Class-based retrieval
  - Weight-valued functions and weighted sets

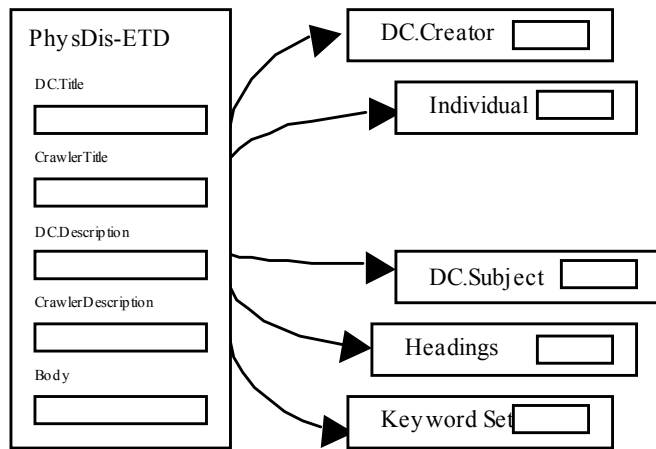
# MARIAN and Interoperability

- System: wrappers and harvesting
- Syntax: OAI standards (XML, Unicode, ...)
- Structure: information networks
- Semantics: class-based retrieval  
: collection views

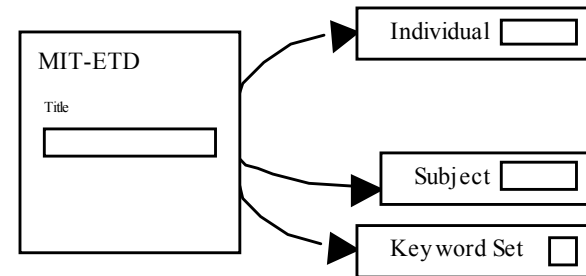
# System & Syntactic Interoperability



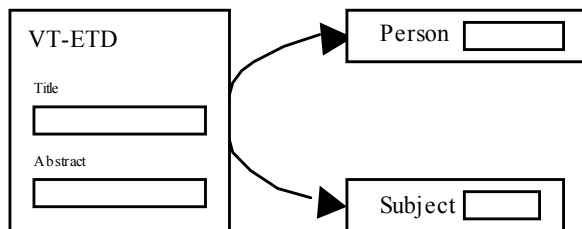
# Structural Interoperability through Information Networks



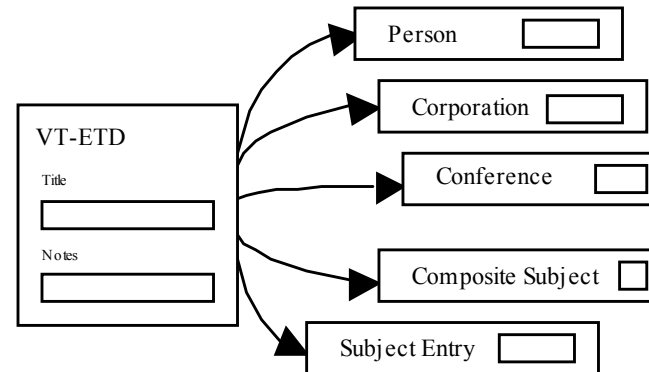
**(SOIF Format)**



**(RFC1807 Format)**



**(ETD-MS Format)**



**(MARC-21 Format)**

# Community Assistance by OAI?

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



# Questions to Address

- Is DC sufficient, or is there a list of one or more metadata standards existing or that can be developed to suite community needs?
- Is there a natural set structure, or several?
  - Year? Topical areas? Location / institution?
- What are the social, economic, political issues regarding who will run an Open Archive?
- Will all share metadata or must there be federated search as well?
- Protocol – Community Issues?

# Protocol – Community Issues

- Record – about – optional: rights? ...
- Deleted records: how long to remember?
- Identifiers: URI scheme
- Link between record and digital object: can use DC identifier for the digital object
- Set scheme
- Metadata: prefix (e.g., oai\_dc) and schema

# Protocol – Community Issues

- Resumption token: time-to-live
- Acceptable use policies: for harvesting interface, for metadata harvesting
- Identify: repositoryName, baseURL, protocolVersion, adminEmail

# Conclusion

- Interoperability
- History / evolution of OAI
- Protocol for metadata harvesting
- Implementations and support
- Current situation / progress
- Community building and support by OAI