Integration of Regular and Static OAI Repositories

OAI Metadata Harvesting Workshop
JCDL 2003 – May 31, 2003

Edward A. Fox

fox@vt.edu   http://fox.cs.vt.edu

CS   DLRL   Internet TIC

Virginia Tech, Blacksburg, VA, USA

Acknowledgements (Selected)

Announcement

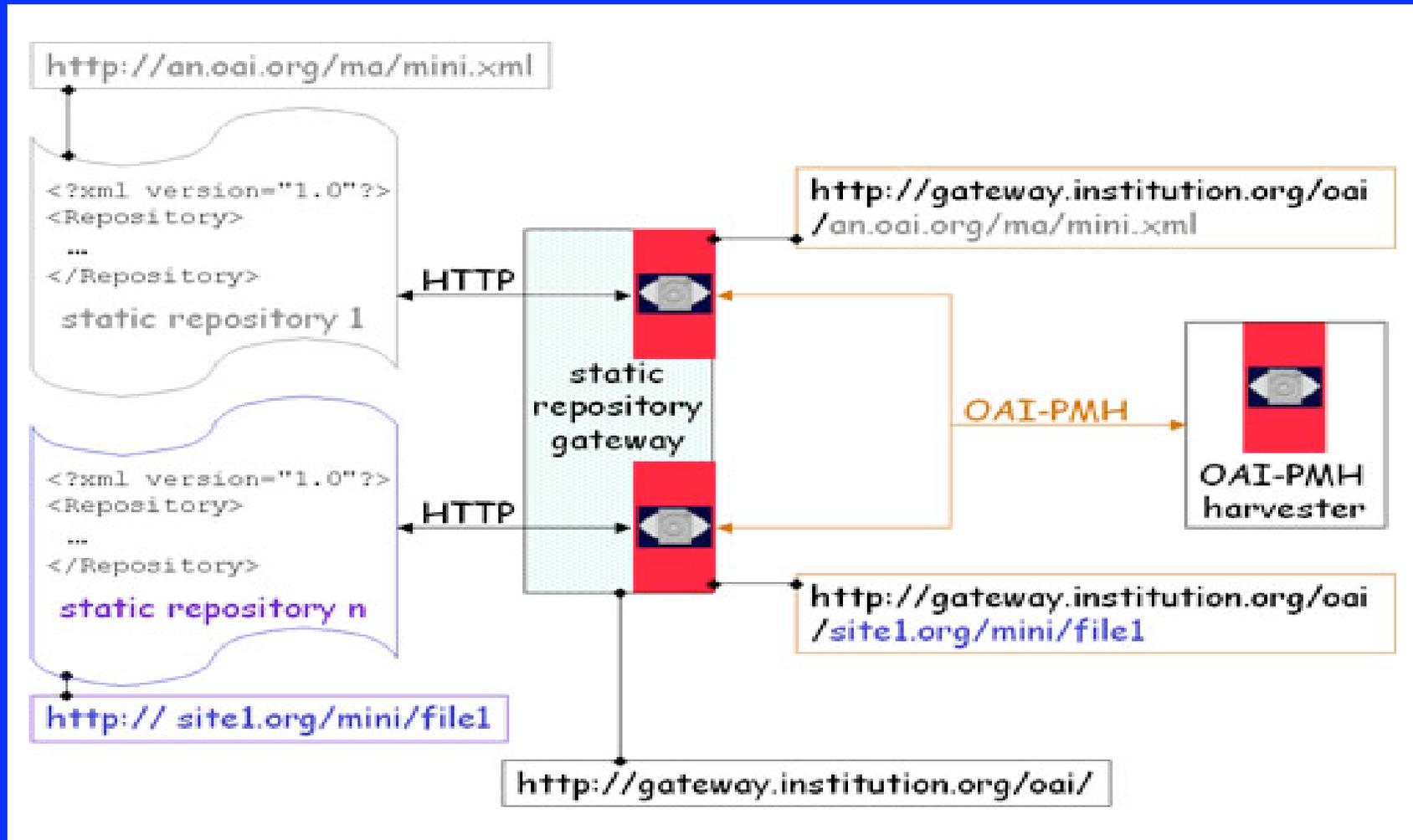- We can discuss this more broadly at US Workshop on Open Digital Libraries (at the Holiday Inn Ballston, Arlington, VA), on Monday, June 23rd through Wednesday, June 25th

Outline

- **OAI Static Repository Model (reminder)**
- **Focus on Education**
- **CITIDEL (including NCSTRL) and NSDL**
- **NDLTD (as complex case study)**
- **Automatic Classification from NDLTD to CITIDEL**
- **Selected Links**

The OAI Static Repository Model

Outline

- **OAI Static Repository Model (reminder)**
- **Focus on Education**
- **CITIDEL (including NCSTRL) and NSDL**
- **NDLTD (as complex case study)**
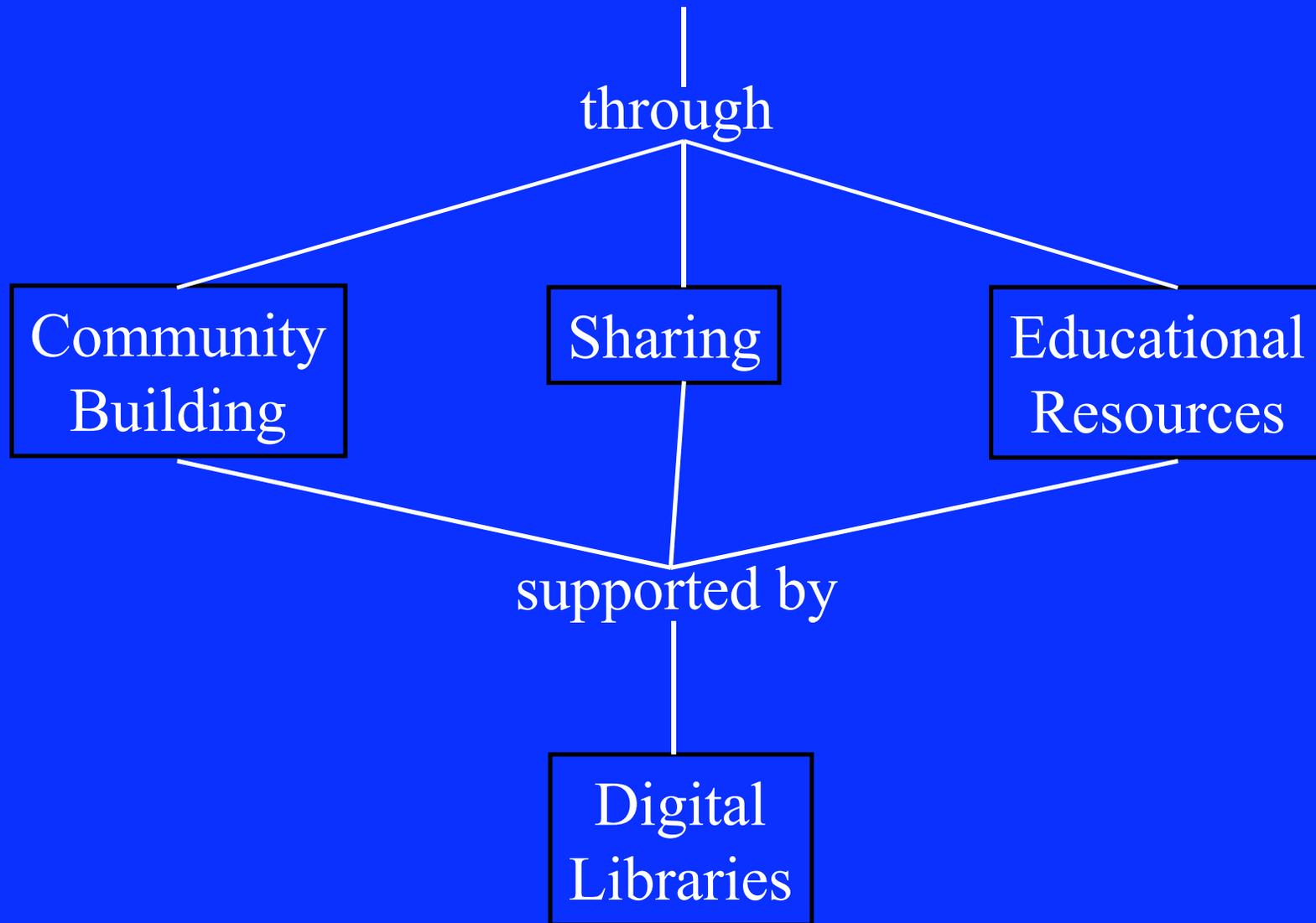- **Automatic Classification from NDLTD to CITIDEL**
- **Selected Links**

Advancing Education

through

Community
Building

Sharing

Educational
Resources

supported by

Digital
Libraries

CS -> CSTC -> CRIM

- NSF and ACM Education Committee are funding a 2 year project "A Computer Science Teaching Center" - CSTC - http://www.cstc.org/

- College of NJ, U. Ill. Springfield, Virginia Tech

- Focus initially on labs, visualization, multimedia

- Multimedia part is also supported by a 2nd grant to Virginia Tech and The George Washington University: http://www.cstc.org/~crim/ (with curricular guidelines also under development)

CS Teaching Center (CSTC)

- Instead of building large, expensive multimedia packages, that become obsolete and are difficult to re-use, concentrate on **small knowledge units**.

- Learners benefit from having well-crafted modules that have been **reviewed and tested**.

- Use digital libraries to build a **powerful base** of support for learners, upon which a variety of courses, self-study tutorials & reference resources can be built.

- ACM support led to Journal of Educational Resources in Computing (**JERIC**), accessible from **www.cstc.org**

COMPUTER SCIENCE TEACHING CENTER

a digital library of peer-reviewed teaching resources

home | about | contact us | submit a resource | search | browse resources | login

Welcome!

The CSTC is a digital library of reviewed resources for teaching computer science. We invite you to submit a resource and browse our collection.

Submit a Resource

- Step 1: Create a CSTC account
- Step 2: Login to CSTC
- Step 3: Complete Submission Application

Volunteer as a Reviewer

- Step 1: Create a CSTC account

CSTC Newsletter

Our monthly newsletter keeps you informed about the latest additions to the library. To subsribe to the newsletter, simply create an account and put a check in the checkbox under the CSTC newsletter section. If you want to unsubsribe at any time, just go to the "account info" page and deselect the checkbox.

ACM JERIC

Resources submitted to CSTC may be considered for publication in the ACM Journal on Educational Resources in Computing (JERIC).

Resource Summary

title	Applets for teaching computer graphics concepts
summary	**Bezier Applet** - Allows interactive creation of Bezier curves by adding and removing points from a canvas with the mouse. It also allows the user to display convex hulls and continuity hints for each curve. **2D Transforms Applet** - Demonstrates the effects of the order in which transformations are performed on an object in 2D. The user has control over three operations which may be scaling, rotation, and translation in any order. **3D Viewing Applet** - Demonstrates the basic concepts of 3D viewing. There are two views which allow the user to change the location of the camera and the point of interest and see the effects in the camera view. The user may choose between two different scenes to render and whether to use perspective or orthographic projection. **Lighting Applet** - This applet demonstrates the Phong shading model and allows the user to render one of four scenes and modify the parameters for ambient, diffuse, and specular light.
author	Patrick Min Princeton University
platform requirements	Java Virtual Machine 1.1 or higher

Resource Information

Required: The fields within this section are all required. You will not be able to submit this form without filling in all of the elements.

Title:

Interactive Multimedia Development

**Type of Resource:** Choose as many categories that are appropriate for the resource being submitted. Hold the shift or control key while clicking the mouse to make multiple selections.

Multimedia
Laboratories
Visualizations

ACM Subject Classification: Choose a list of subjects that describe the resource. Hold the shift or control key while clicking the mouse to make multiple selections.

Example Open Digital Library

USER INTERFACE

User Interface
OAI/ODL component
OAI/ODL protocol
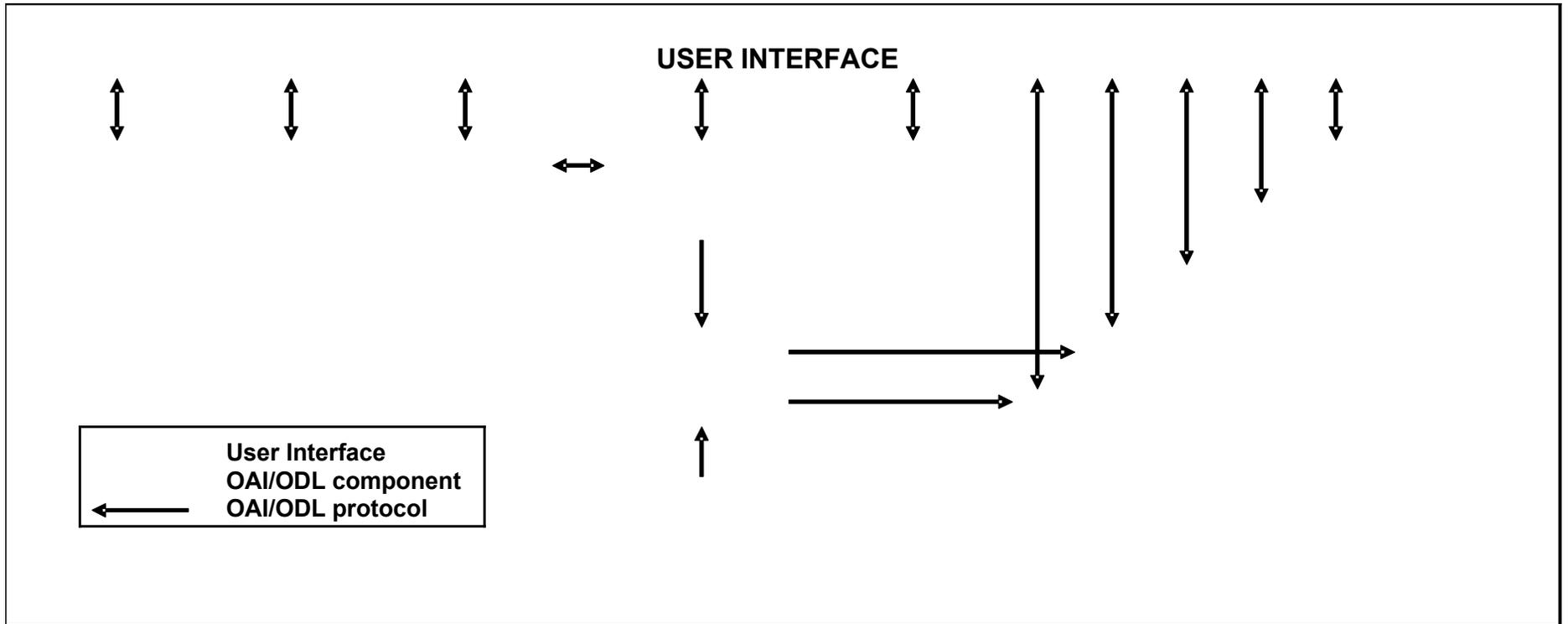
Digital Library for the
Computer Science Teaching Center (www.cstc.org)
(slide by Hussein Suleman)

Outline

- **OAI Static Repository Model (reminder)**
- **Focus on Education**
- **CITIDEL (including NCSTRL) and NSDL**
- **NDLTD (as complex case study)**
- **Automatic Classification from NDLTD to CITIDEL**
- **Selected Links**

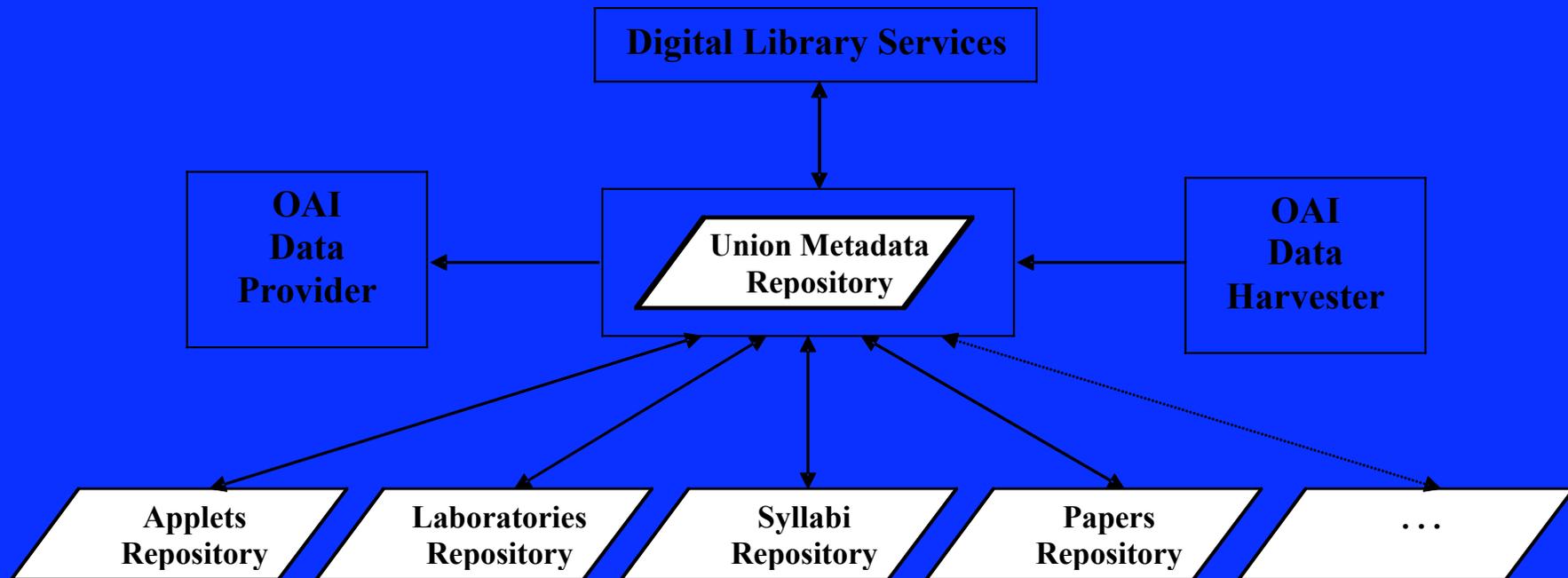# Computing and Information Technology Interactive Digital Educational Library (CITIDEL)

- Domain: computing / information technology

- Genre: one-stop-shopping for teachers & learners: courseware (CSTC, JERIC), leading DLs (ACM, IEEE-CS, DB&LP, CiteSeer), PlanetMath.org, NCSTRL (technical reports), Kepler?, …

- Submission & Collection: sub/partner collections ➔ www.citidel.org
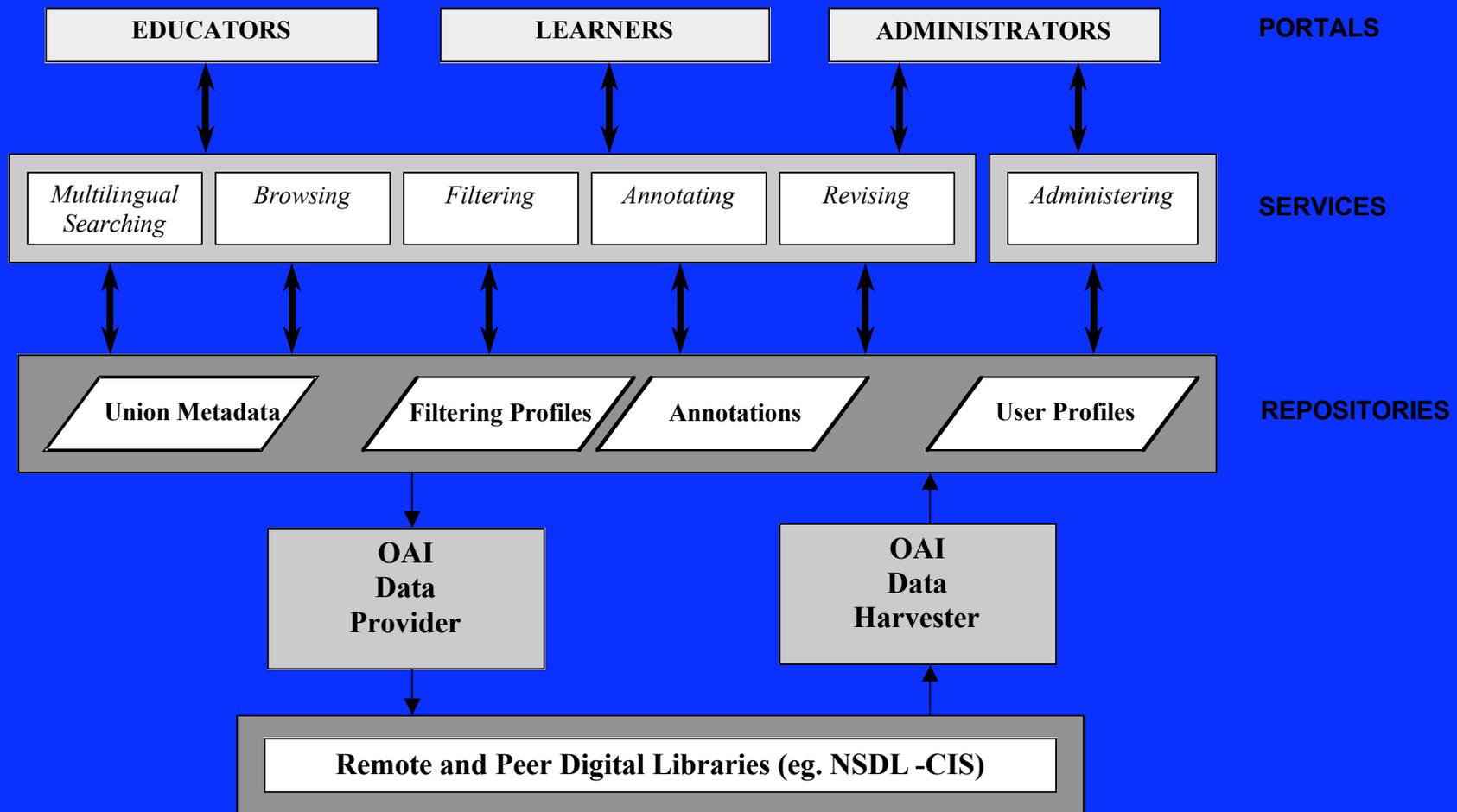
www.CITIDEL.org

- Led by Virginia Tech, with co-PIs:
    - Fox (director, DL systems)
    - Lee (history)
    - Perez (user interface, Spanish support)

- **Partners**
    - **College of New Jersey (Knox)**
    - **Hofstra (Impagliazzo)**
    - **Villanova (Cassel)**
    - **Penn State (Giles)**

Distributed repository structure

Digital library architecture for local and interoperable CITIDEL services

| EDUCATORS | LEARNERS | ADMINISTRATORS | **PORTALS** |

| *Multilingual Searching* | *Browsing* | *Filtering* | *Annotating* | *Revising* | *Administering* | **SERVICES** |

| Union Metadata | Filtering Profiles | Annotations | User Profiles | **REPOSITORIES** |

OAI
Data
Provider

OAI
Data
Harvester

Remote and Peer Digital Libraries (eg. NSDL -CIS)

EPrints for VT CS Technical Reports

Computer Science @VIRGINIA TECH

Welcome to VT CS Technical Reports

This is part of the NCSTRL collection, an international collection of computer science technical reports from CS departments and industrial and government resear laboratories. The collection is distributed among a set of interoperating servers operated by participating institutions. You may search for and access documents from different divisions within Virginia Tech and beyond.

Search the Titles, Abstracts and Keywords

[Search]

Browse

Browse the archive by Subject or Year.

Latest Additions

View items added to the archive in the past week.

Simple Fielded Search

Search the archive using the most common fields.

Advanced Search

Search the archive using a more complex range of fields.

Registered Users Area

Registered users may submit items to the archive.

Case Study: NCSTRL Costs/Benefits

| Stakeholders		Sample Potential Cost	Sample Potential Benefit
Providers	Faculty	Lower value for P&T	Faster publishing
	Students	Less recognition	Broader set of outlets
	Practitioners	Limited relevance	Ease of publishing, > quantity
Users	Faculty	Lower quality of work	Broader access to resources
	Students	Higher access costs (vs. department available material)	Lower access costs (vs. journal available material)
	Departments	New maintenance costs	Broader visibility
	University libraries	Additional access costs	Access to new resources
	Practitioners	More difficult access	Access to new resources

CITIDEL: Current Architecture
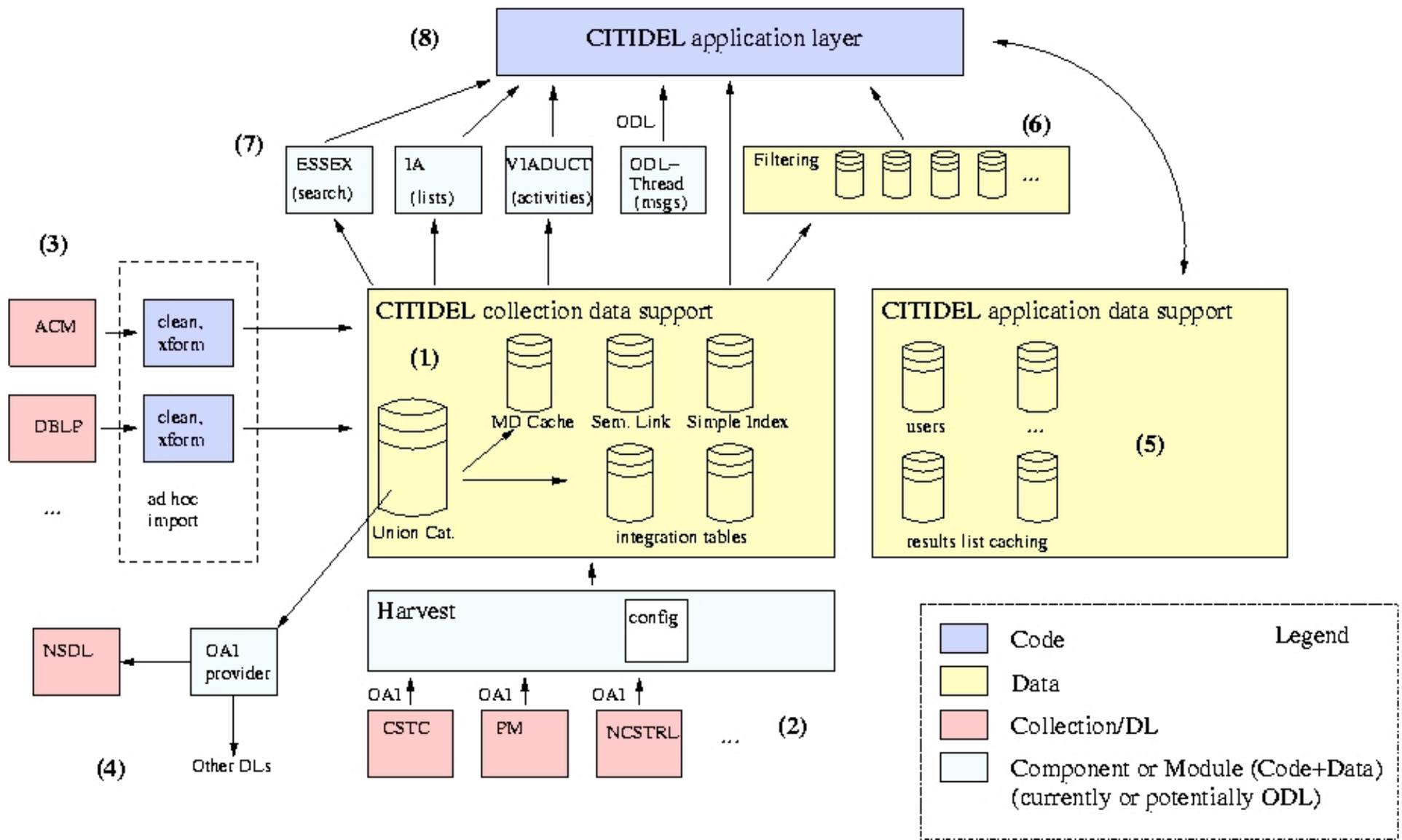
Slide from Aaron Krowne

CITIDEL -> NSDL

- A collection project in the

- National STEM (science, technolgy, engineering, and mathematics) education Digital Library – **NSDL**
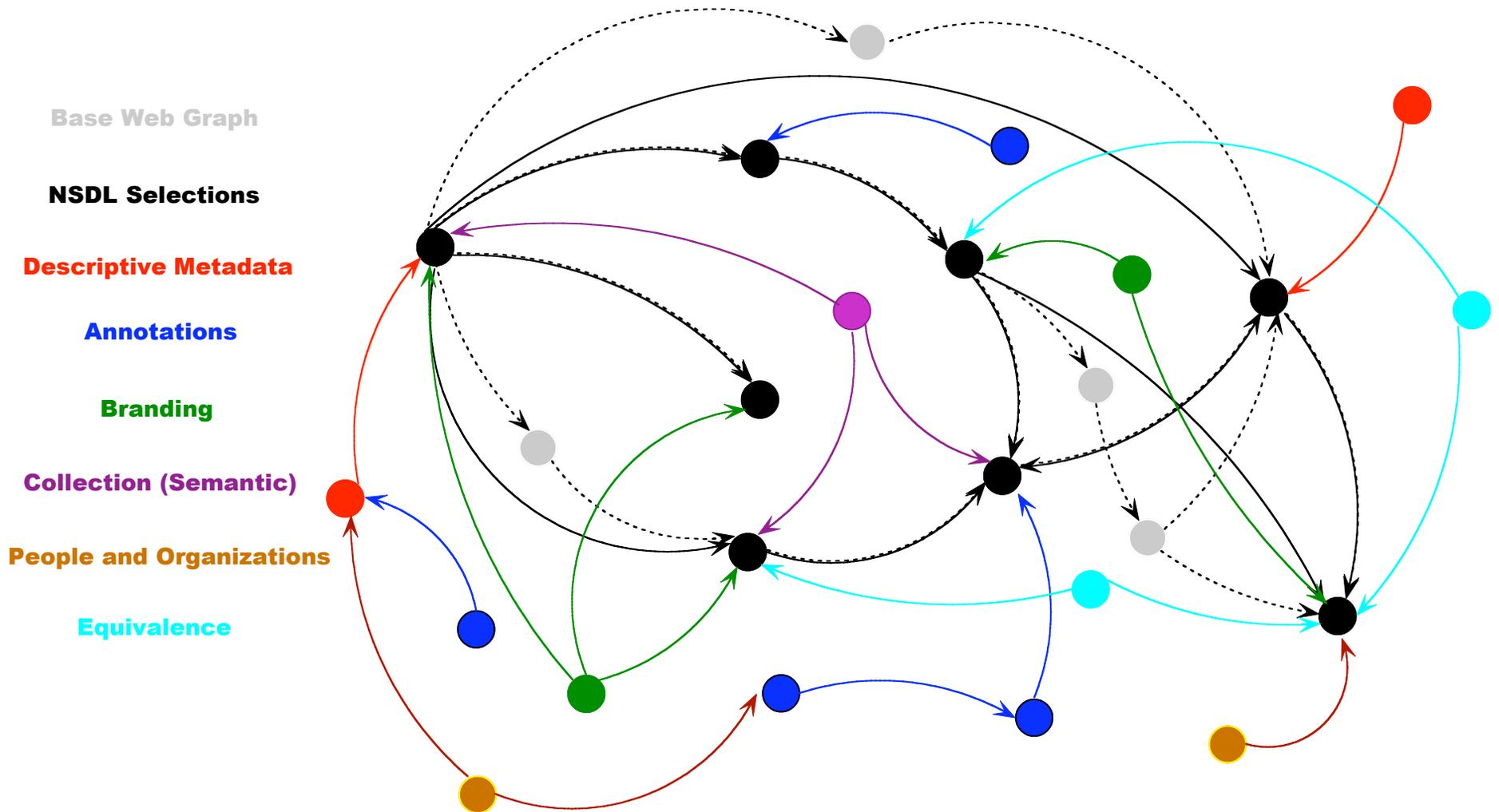
- **-> LEARNS**
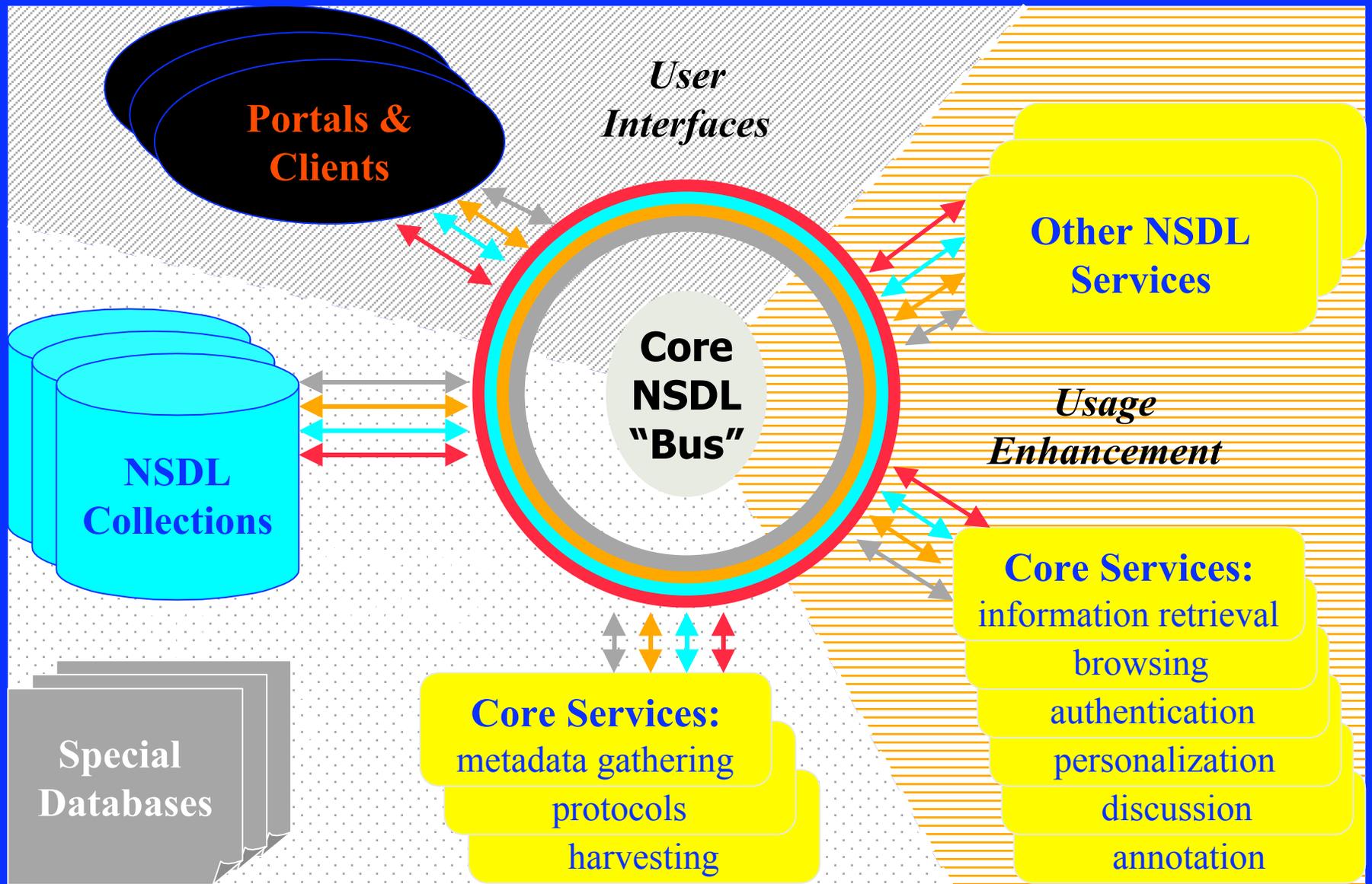
Collections

- Discovery of content

- Classification and cataloguing

- Acquisition and/or linking; referencing

- Disciplinary-based themes define a natural body of content, but other possibilities are also encouraged

- Access to massive real-time or archived datasets

- Software tool suites for analysis, modeling, simulation, or visualization

- Reviewed commentary on learning materials and pedagogy

Slide from Lee Zia

Proposed Basis for Adding Value to Interconnected DLs
A Data Warehouse, Specialized for Relationships

Base Web Graph

NSDL Selections

Descriptive Metadata

Annotations

Branding

Collection (Semantic)

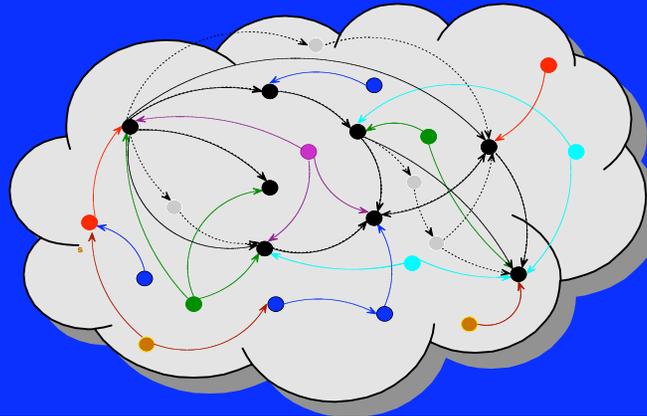People and Organizations

Equivalence

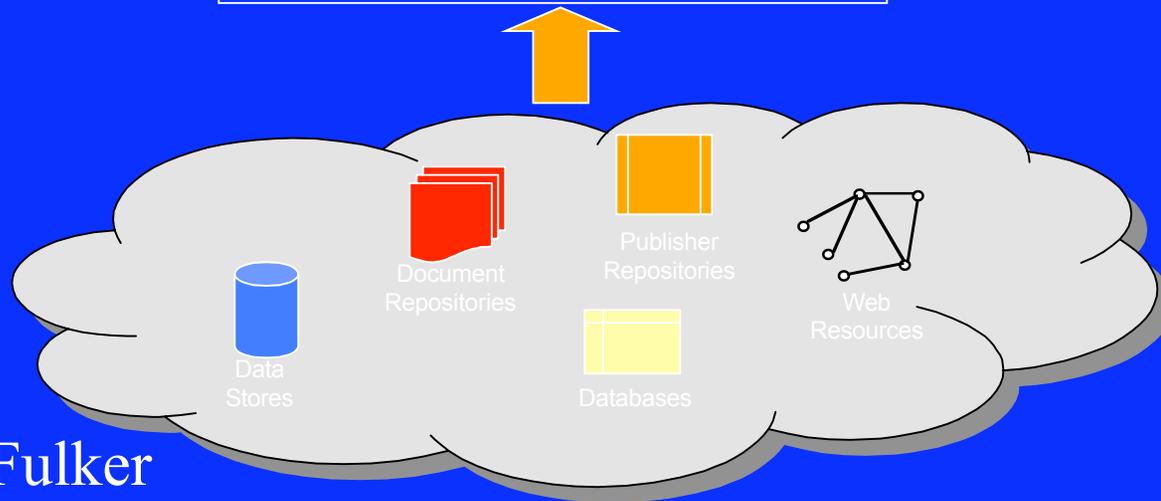**Diverse Network of Partner Libraries and Services (retail)**

Specialized Mining

Data Annotation

NSDL Data Warehouse: Entities and their Relationships (wholesale)

Harvesting, Gathering, Normalization

Digital Sources

Data Stores

Document Repositories

Publisher Repositories

Databases

Web Resources

Slide from Dave Fulker

CI and Central Search Engine

- Central portal as anachronism
- Interaction with other projects/portals
  - Publisher/society – Elsevier, AIP, ACM, EI
  - ARL Portal, DLF, OAIster
  - Institutional repositories
  - Course management systems
  - A & Is with full-text links
  - Integrated library systems (SFX, Encompass)
  - CrossRef
  - Biomed Central, Public Library of Science

Slide from Bill Mischo

Outline

- **OAI Static Repository Model (reminder)**
- **Focus on Education**
- **CITIDEL (including NCSTRL) and NSDL**
- **NDLTD (as complex case study)**
- **Automatic Classification from NDLTD to CITIDEL**
- **Selected Links**

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 are the long term goals?

- 400K US 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

Student Gets Committee Signatures and Submits ETD
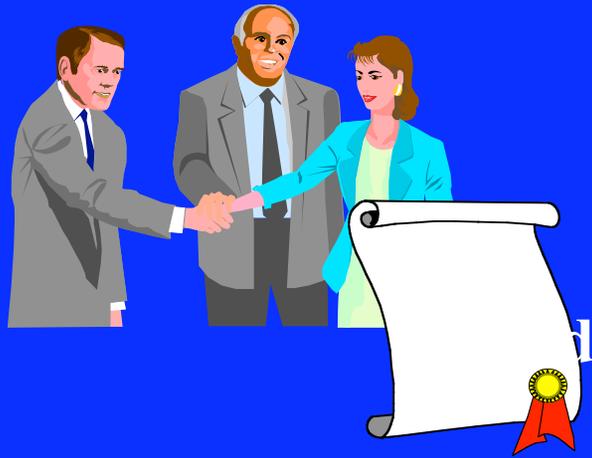
Approval form
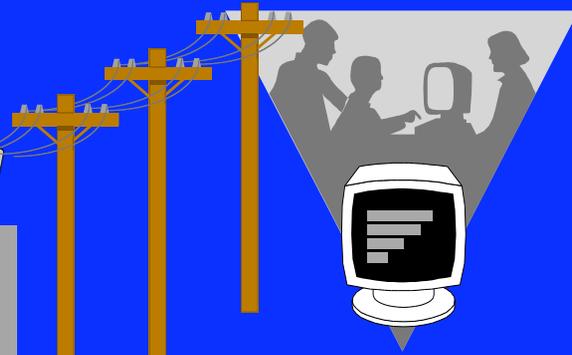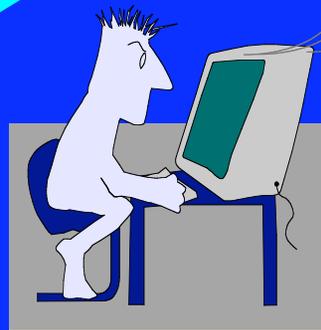
Grad School

Library Catalogs ETD, Access is Opened to the New Research

WWW

NDLTD

Digital library access control

Access to VT's ETDs

http://scholar.lib.vt.edu/theses/

	1997/98	1997/98	1999/00	2000/01	2001/02
ETD files requested	231,709	483,030	578,152	2,173,420	4,497,199
Abstracts requested	165,710	215,493	260,699	573,149	471,917
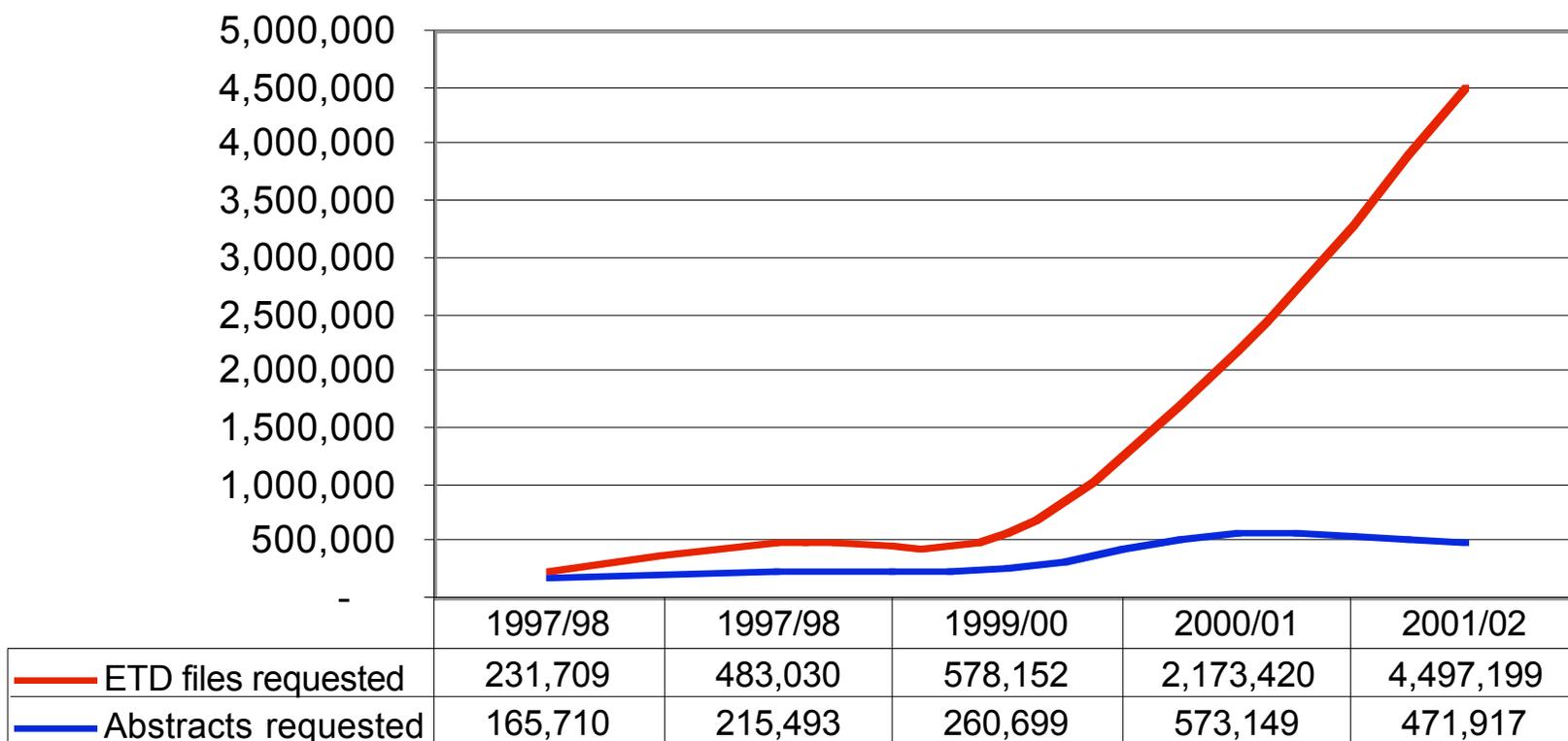
Brief History of ETD Meetings

- 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, ...
- 1998 – 1st symposium – Memphis (20)
- 1999 – 2nd symposium – Blacksburg (70)
- 2000 – 3rd symposium – St. Petersburg (225)
- 2001 – 4th symposium – Caltech (200)
- 2002 – 5th syposium – BYU, Provo, Utah
- 2003 – 6th syposium – Berlin (215)
- 2004 – 7th syposium – U. Kentucky
- 2005 – 8th syposium – Sydney, Australia

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.
- **Belgium**
- **Brazil**
- **Germany**
  - Humboldt University (lead)
  - 3 other universities
  - 5 learned societies: Math, Physics, Chemistry, Sociology, Education
  - 1 computing center
  - 2 major libraries

- **India**
- **Lithuania**
- **Spain:** Consorci de Biblioteques Universitàries de **Catalunya**, as group, www.cbuc.es: 9 sites
- **Sudan**
- **UK (British Library, JISC, Edinburgh)**
- **UNESCO (especially Latin America, Eastern Europe, Africa)**
- **USA:**
  - **CIC ("Big 10")**
  - **Ohio:** OhioLINK: 79 colleges/univs
  - SOLINET
- …

US University Members

- Air University (Alabama)
- Baylor University
- Boston University
- Brigham Young University
- Caltech
- Clemson University
- College of William & Mary
- Concordia University (Illinois)
- Drexel University – required 4/2002
- East Carolina University
- East Tenn. State U. – required 1/2001
- Florida Institute of Technology
- Florida International University
- Florida State University
- Florida Tech
- George Washington University
- Georgetown University
- Johns Hopkins University
- Louisiana State University – required 1/2002
- 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 – required 9/2002
- Northwestern University
- Penn. State University
- Regis University
- Rochester Institute of Tech.
- Texas A&M

- U. of Central Florida
- U. of Colorado Health Science Center
- U. of Florida – required 8/2001
- U. of Georgia – required 9/2001
- U. of Hawaii, Manoa
- U. of Illinois, Urbana-Champaign
- U. of Iowa
- U. of Kentucky – required in CS only
- U. of Maine – required in CS, Spatial Info Sci/Eng
- U. of Missouri-Columbia
- U. of North Texas – required since 8/99
- U. of Oklahoma
- U. of Nevada, Las Vegas
- U. of New Orleans
- U. of North Texas – required 8/1999
- U. of Oklahoma
- U. of Pittsburgh
- U. of Rochester
- U. of South Florida – required 8/2002
- U. of Tennessee, Knoxville
- U. of Tennessee, Memphis
- U. of Texas at Austin – required 6/2001
- U. of Virginia – required 1/2003
- U. of West Florida
- U. of Wisconsin - Madison – part reqt 12/1999
- Vanderbilt U.
- Virginia Commonwealth U.
- Virginia Tech - required 1/97
- Wake Forest U.
- West Virginia U. - required 8/1998
- Western Kentucky U. – required 9/2004
- Western Michigan U.
- Worcester Polytechnic Inst. – required 7/2002
- Yale U.

Other Countries (selected)

- Australia
- Belgium
- Brazil
- Canada
- Chile
- China, Hong Kong
- Columbia
- Finland
- France
- Germany
- Greece
- India
- Italy
- Jamaica
- Korea
- Lithuania
- Mexico

- Netherland
- Norway
- Poland
- Russia
- Singapore
- S. Africa
- S. Korea
- Spain
- Sudan
- Sweden
- Taiwan
- Thailand
- UK
- Venezuela

Institutional Members

- **Australian Digital Theses Program**
- **British Library**
- **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)**
- **MathDISS International**
- **National Documentation Centre (NDC), Greece**
- **National Library of Canada**
- **National Library of Portugal**
- **OCLC Online Computer Library Center**
- **Office of Scientific and Technical Info (US Dept of Energy)**
- **OhioLINK**
- **Organization of American States (SEDI/OAS)**
- **Southeastern Library Network (SOLINET)**
- **Sudanese National Electronic Library**
- **UNESCO (www.unesco.org/webworld/etd)**

Access Possibilities
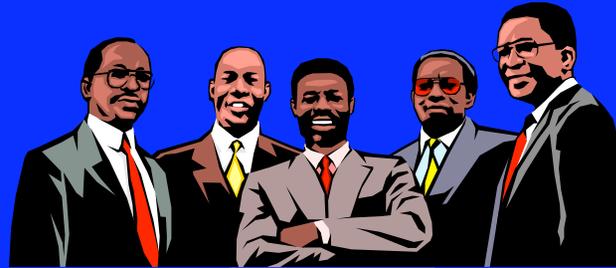
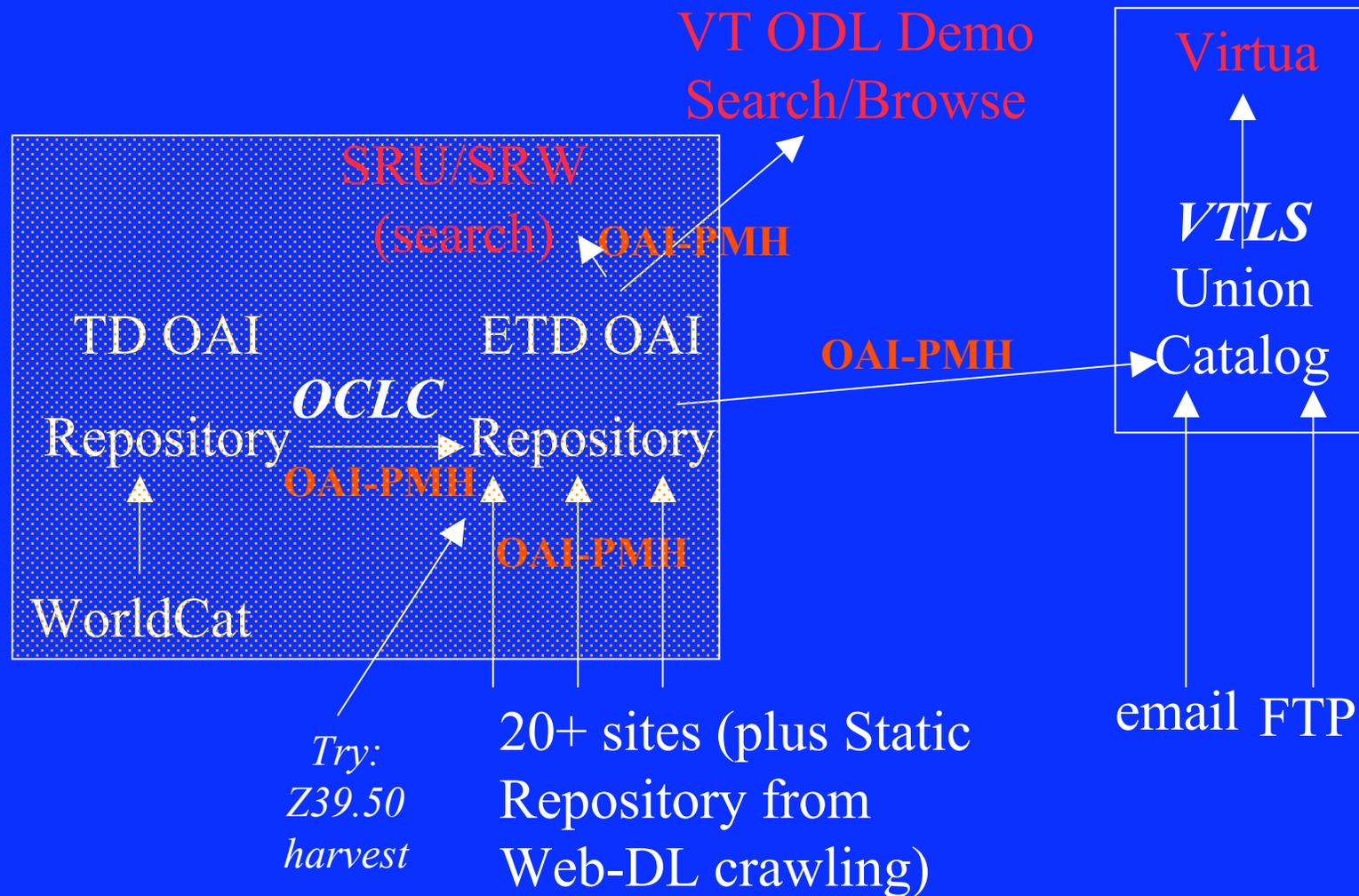| Web search engines	www. theses. org	www. openarchives. org	library catalog clients	3rd Party Services (e.g., UMI)
Virginia Tech MIT	National Library of Portugal	CBUC (Spain)	Ohio Link	National Projects: AU, GE, …

NDLTD Union Catalog Architecture

VT ODL Demo
Search/Browse

Virtua

SRU/SRW
(search) **OAI-PMH**

VTLS
Union
Catalog

TD OAI ETD OAI

OCLC

Repository Repository

OAI-PMH

WorldCat

OAI-PMH

OAI-PMH

*Try:
Z39.50
harvest*

20+ sites (plus Static
Repository from
Web-DL crawling)

email FTP

Outline

- **OAI Static Repository Model (reminder)**
- **Focus on Education**
- **CITIDEL (including NCSTRL) and NSDL**
- **NDLTD (as complex case study)**
- **Automatic Classification from NDLTD to CITIDEL**
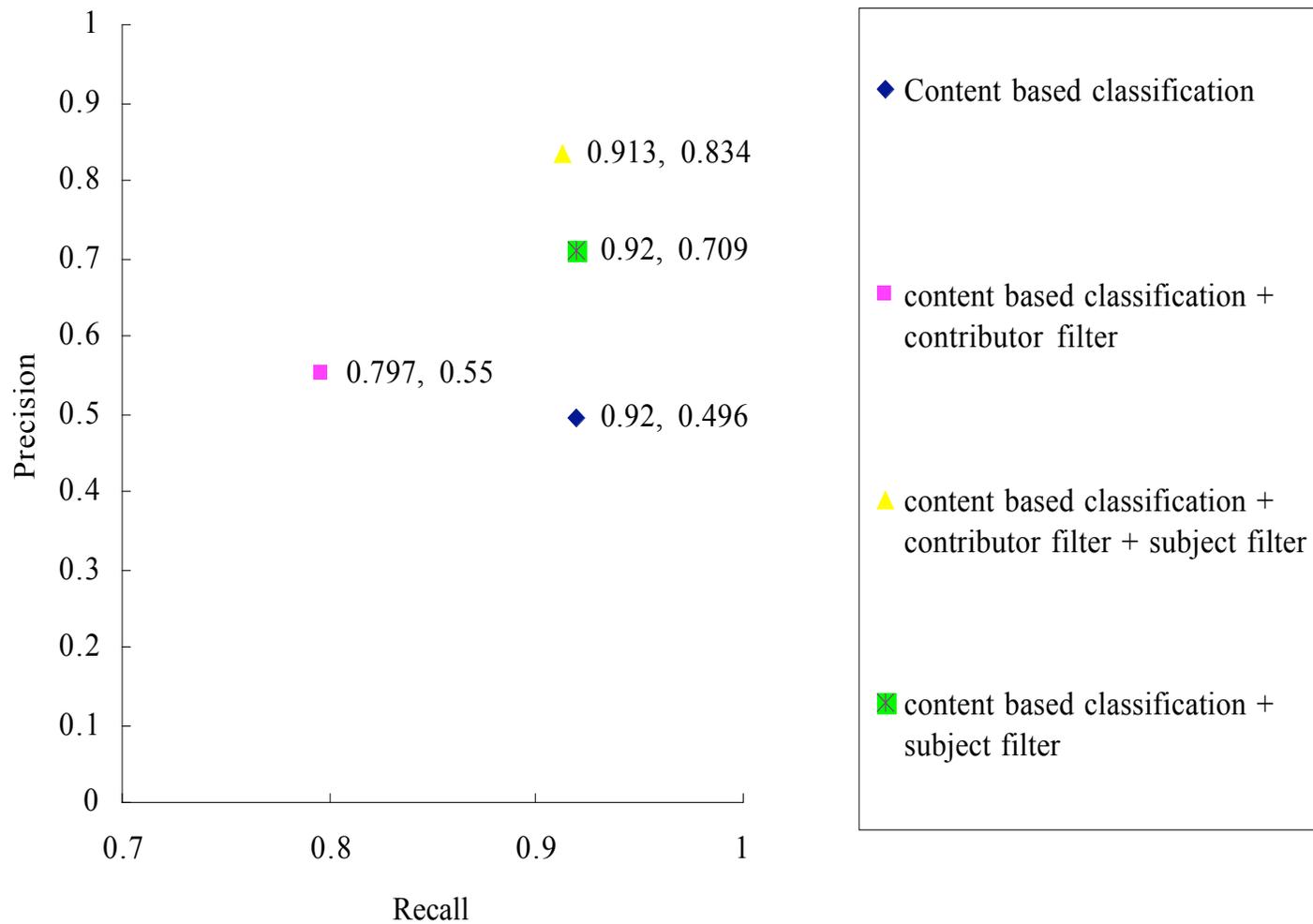- **Selected Links**

Figure 5. Experiments results in Precision Recall format

Slide from Baoping Zhang

Outline

- **OAI Static Repository Model (reminder)**
- **Focus on Education**
- **CITIDEL (including NCSTRL) and NSDL**
- **NDLTD (as complex case study)**
- **Automatic Classification from NDLTD to CITIDEL**
- **Selected Links**

Selected Links - http://fox.cs.vt.edu

- CITIDEL
  - www.citidel.org
- NCSTRL
  - www.ncstrl.org
- NDLTD
  - www.ndltd.org and etdguide.org
- NSDL
  - www.nsdl.org
- Virginia Tech Digital Library Research Laboratory (DLRL)
  - http://www.dlib.vt.edu (5S, 5SL, AmericanSouth.Org, CSTC, ENVISION, MARIAN, NDLTD, NSDL, OAI, ODL)