

Libraries In the Digital Age (LIDA)

May 24-28, 2010 Zadar, Croatia

“Digital Libraries, Scholarship, and the Humanities”

by Edward A. Fox

- fox@vt.edu <http://fox.cs.vt.edu/talks/2010/>
- Dept. of Computer Science, Virginia Tech
- Blacksburg, VA 24061 USA

Acknowledgements

- Mentors (Licklider, Kessler, Salton)
- Virginia Tech, CS, Digital Library Research Laboratory
- NSF and other sponsors
- Students, colleagues, co-investigators:
- Monika Akbar, Yinlin Chen, Spencer Lee, Venkat Srinivasan, Seungwon Yang, ...
- Boots Cassel, Gary Marchionini, Jeffrey Pomerantz, Barbara Wildemuth, Andrea Kavanaugh, Naren Ramakrishnan, Steve Sheetz, Don Shoemaker, ...

Scholarship

DL curriculum
Ensemble
Living In the KnowlEdge
Society (LIKES): Book,
Part on Humanities

Humans/ Humanities

Needs
Areas
Approach

**Supporting
Scholarship
About Humans
Through
Content &
Services**

Content

Digitized
ETDs
Born digital
(digital
natives)

DLs:
CTRnet
Ensemble
ETANA
NDLTD

Services

Flattening
(cooperation,
collaboration)
Exploring
(search, browse)
Interoperability 3
Integration

Outline

1. Humanism, Humanities
2. Selected DL Projects
3. Basic DL Concepts, Scholarship
4. NDLTD
5. Services Centered

Part 1 – Humanism, Humanities

- Humanism
 - Need for
- Humans
 - Principles
- Humanities
 - Areas
 - About

Humanism

- Human, humanist, humanistic, humanism
- Geographical Perspective
 - Asian, Greek, Roman, European, ...
- Historical Perspective
 - Ancient, Classical Period, Renaissance, ...
- Related
 - Knowledge, Literature, Philosophy, Religion
 - Education, Inclusion, Life Stance
 - ...

Need for Humanism

- Global population, tipping point
- Destruction of biodiversity, ecology, environment -> global weirding, extinctions
- War, terrorism, poverty -> education, ...
- Mental health -> creativity, compassion
- Tom Friedman: Hot, Flat and Crowded
- Jared Diamond: Collapse: How Societies Choose to Fail or Succeed

Humans

- Health care crisis: population, aging, cost, specialization; stress, mental health, ...
- ADHD, PTSD, “Negative emotions”, ...
- Wholistic: To nurture or treat the entire person - physical, emotional, mental, and spiritual. (East vs. West, “fusion”)
- Reductionism -> Integration
- Mental discipline, meditation, reflection
- Individual, Group, Organization, Society, World: Collaboration, cooperation, peace, ...⁸.

Usui Reiki Principles

Version a:

For today only

- **anger not,**
- **worry not.**

Be humble, and

**With Gratitude work on
yourself.**

Be Compassionate.

Version b:

Just for today

- **do not worry,**
- **do not anger.**

**Treat others with respect
and forgiveness.**

Work with integrity.

Be grateful.

Humanities – 1 (Areas?)

- Anthropology, Archaeology?
- Communication?
- Performing arts?
 - Music
 - Theatre
 - Dance
- Visual arts?
 - Painting
 - Sculpture
 - Media types, Multimedia, hypermedia

Humanities – 2 (Areas)

- Classics
- Culture
- History
- Languages, Linguistics, English
- Law
- Literature
- Philosophy
- Religion

Humanities – 3 (About)

- Studying:
 - Human condition
- Approach:
 - Analytical
 - Critical
 - Speculative

Part 2 – Selected DL Projects

- Digital Library Curricular Resources
 - NSF IIS-0535057 & 0535060
- CTRnet (Crisis, Tragedy & Recovery Net)
 - NSF IIS-0916733
- Ensemble (Computer Science Education)
 - NSF DUE-0840719
- Digital Preserve
 - NSF IIS-0910183 & 0910465
 - <http://slurl.com/secondlife/Digital%20Preserve/140/126/29>

DL Curric. Project - 1

- NSF awards to VT and UNC-CH
- CS and LIS
- Project server: <http://curric.dlib.vt.edu/>
- Wikiversity:
http://en.wikiversity.org/wiki/Curriculum_on_Digital_Libraries

DL Curric. Project - 2

- **Module 1-b: History of digital libraries and library automation**
- **Module 2-c: File Formats, Transformation, and Migration**
- **Module 3-b: Digitization**
- **Module 4-b: Metadata**
- **Module 5-a: Architecture overviews**

DL Curric. Project - 2

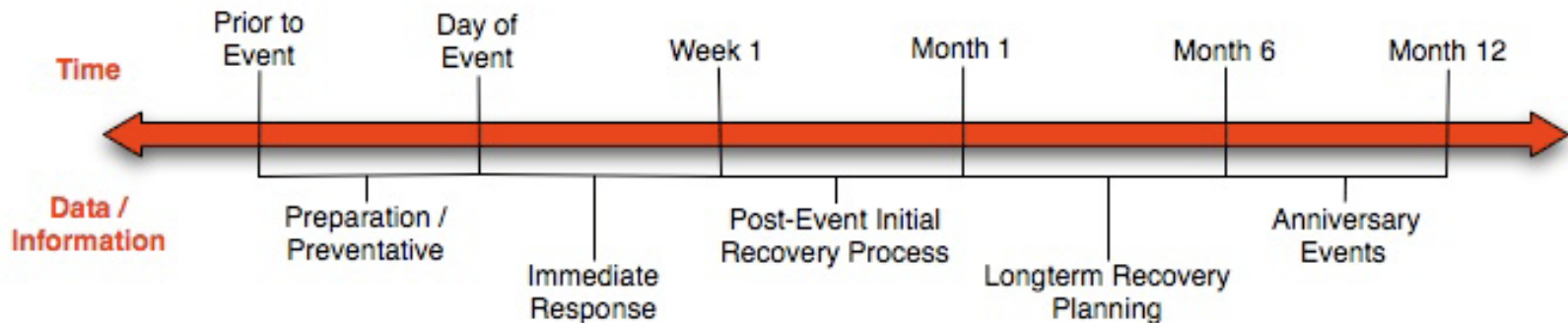
- **Module 5-b: Application software**
- **Module 5-d: Protocols**
- **Module 6-a: Information needs/relevance**
- **Module 6-b: Online information seeking behaviors and search strategies**
- **Module 6-d: Interaction design and usability assessment**

DL Curric. Project - 3

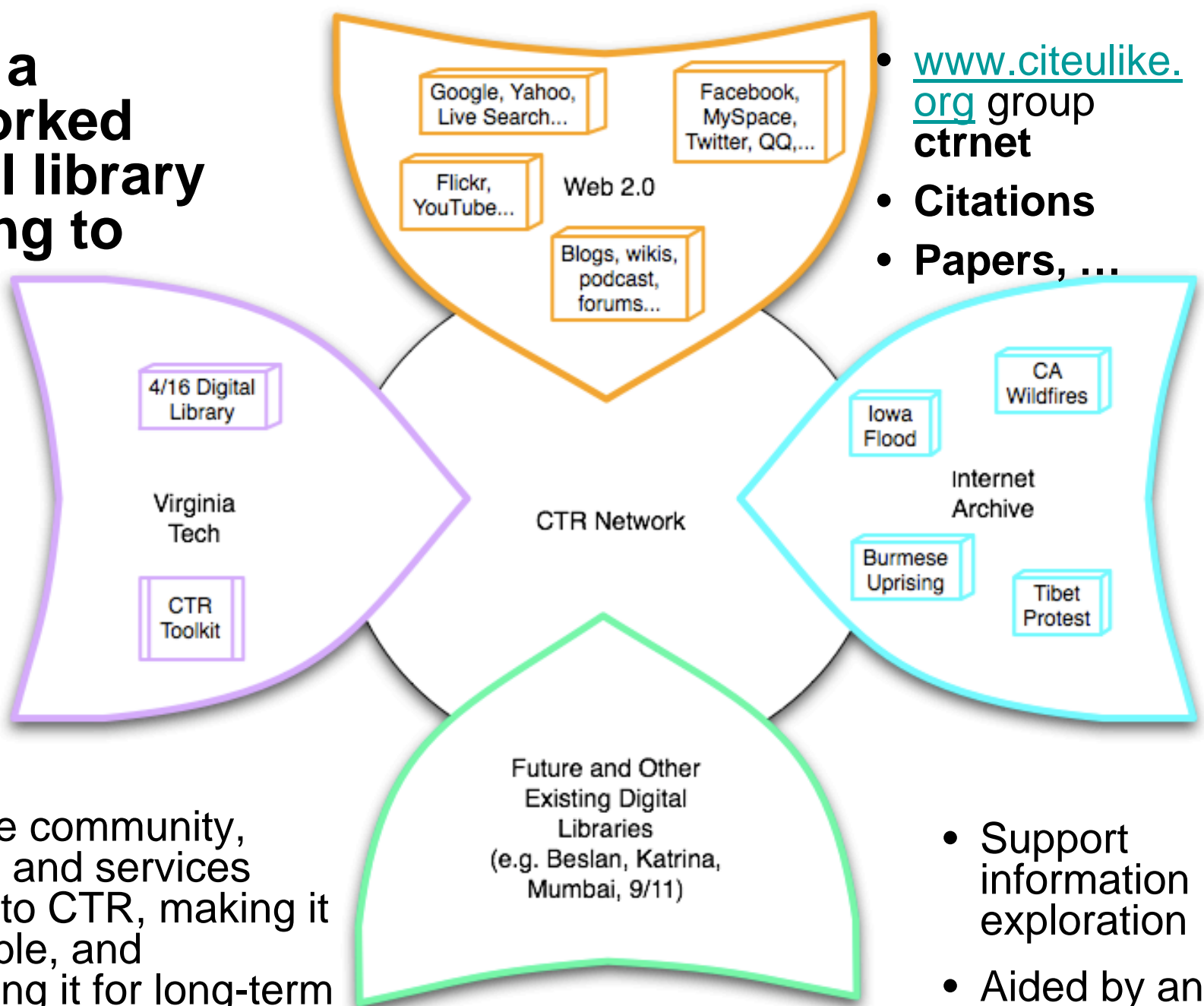
- **Module 7-b: Reference Services**
- **Module 7-g: Personalization**
- **Module 8-b: Web Archiving**
- **Module 9-c: Digital library evaluation, user studies**

Crisis, Tragedy, and Recovery

- Human tragedies that result from man-made and natural events affect humans and communities significantly.
- During and after a tragic event, there are a series of needs that have to be addressed.
 - Compounded by communication failures and a confusing plethora of data and information



- **Build a networked digital library relating to CTR**



- www.citeulike.org group ctrnet

- Citations

- Papers, ...

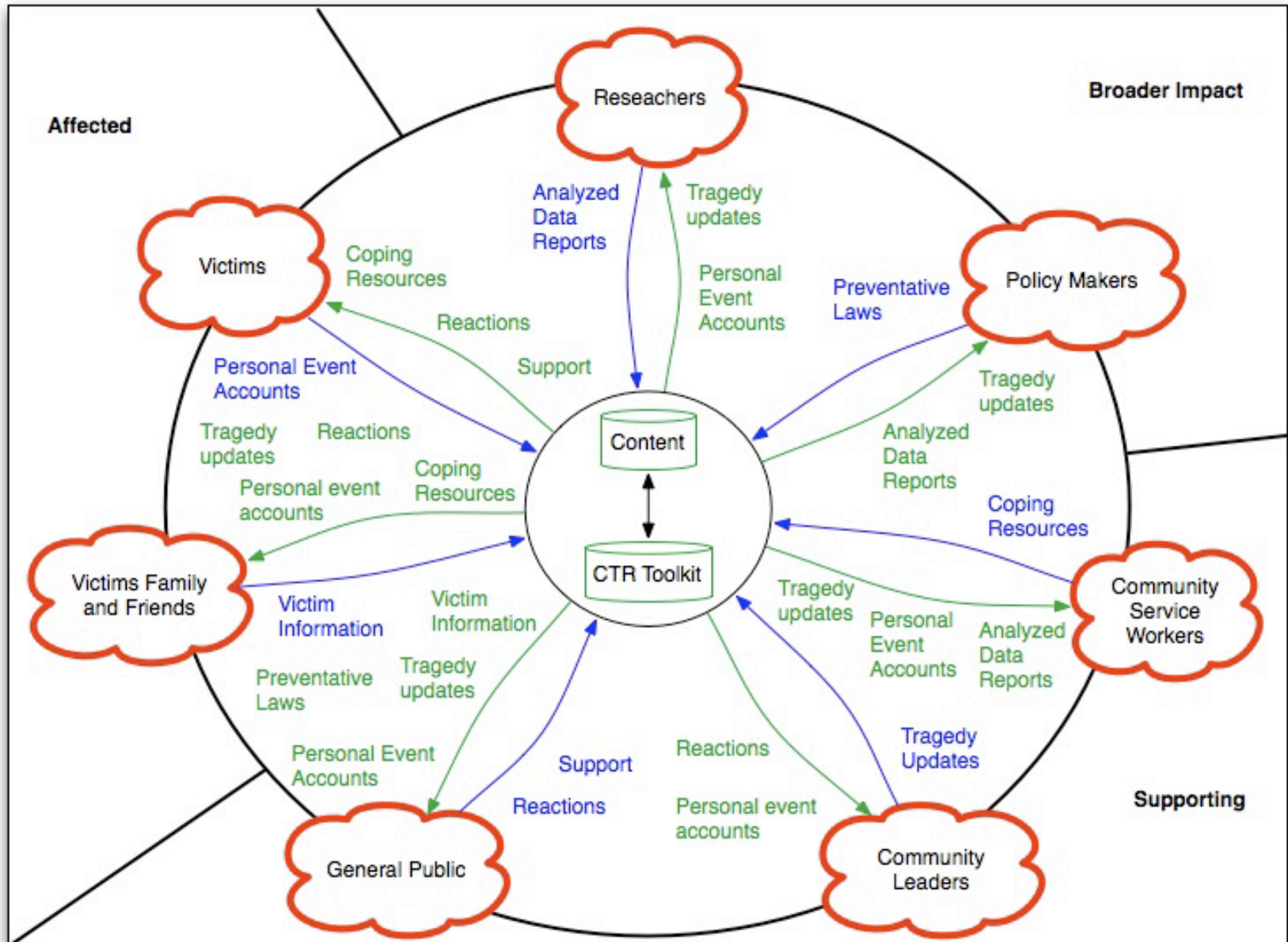
- Integrate community, content, and services relating to CTR, making it accessible, and preserving it for long-term reuse

- Support information exploration

- Aided by an ontology

www.ctrnet.net

CTR stakeholders



Haiti Photographs, Content Based Image Retrieval Evaluation

Facebook

WebMail: Welcome

Eva tool - User-orient...

+

←

→

↺

🏠

☆

http://localhost/eva/codes/view_images_feedback.php

▶


📄

🔧

User-oriented evaluation








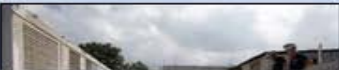


Evaluating query 1 of 1

Query image:



Name:
HumanGrandGoave1.jpg

RESULTS

 <div>HumanLeogane1.jpg</div>	 <div>grand-goave2.jpg</div>	 <div>HumanGrandGoave2.jpg</div>	 <div>sacrecoeur-church.jpg</div>	 <div>petit-goave3.jpg</div>
				

Menu

Eva tool - User-orient...

jennifer - File Browser

HaitiDisasterPhotoFin...

🔍

📁

📄

🔊

🔌

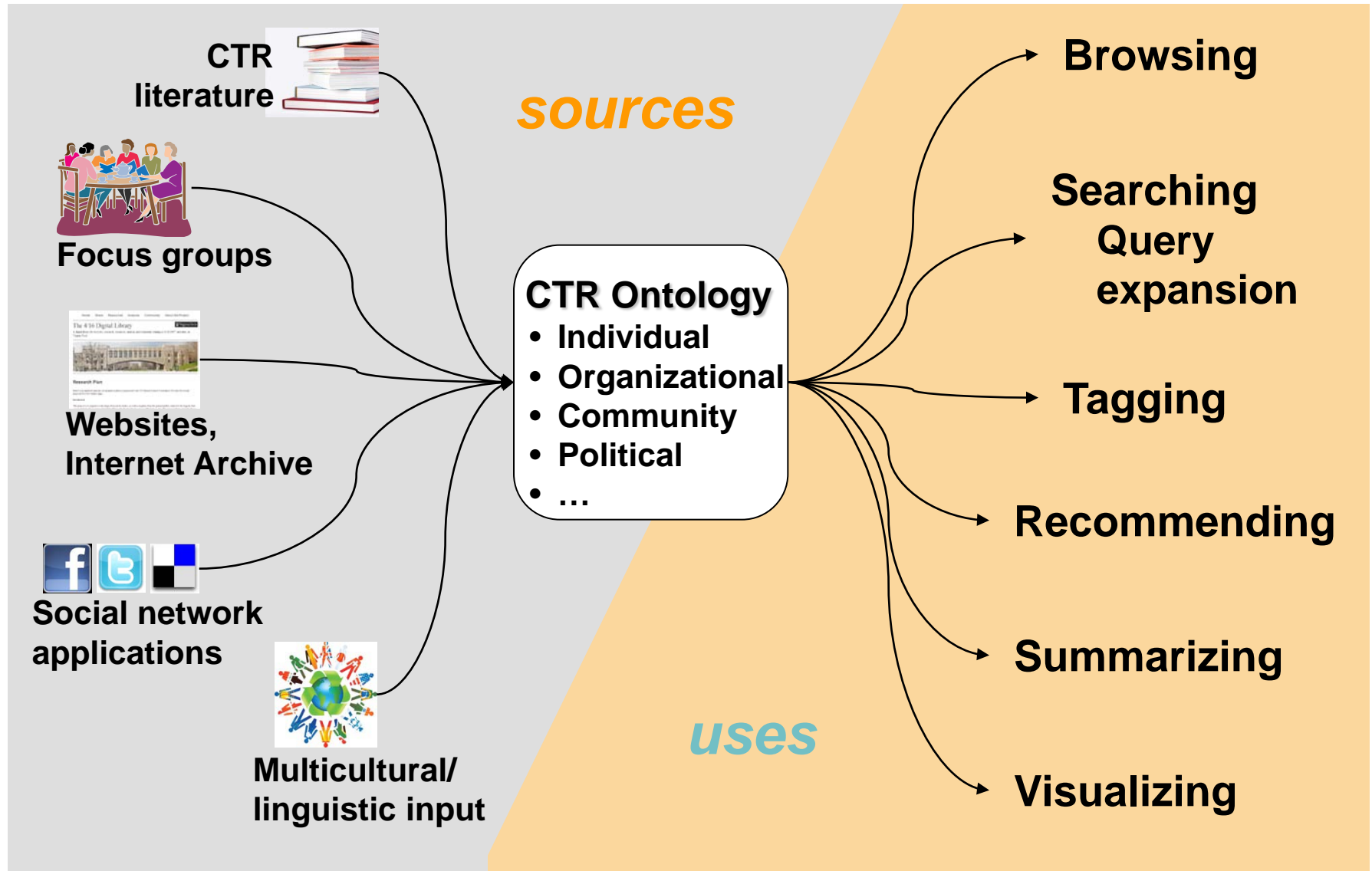
📶

Tue Apr 27, 11:14 AM

CTR Ontology

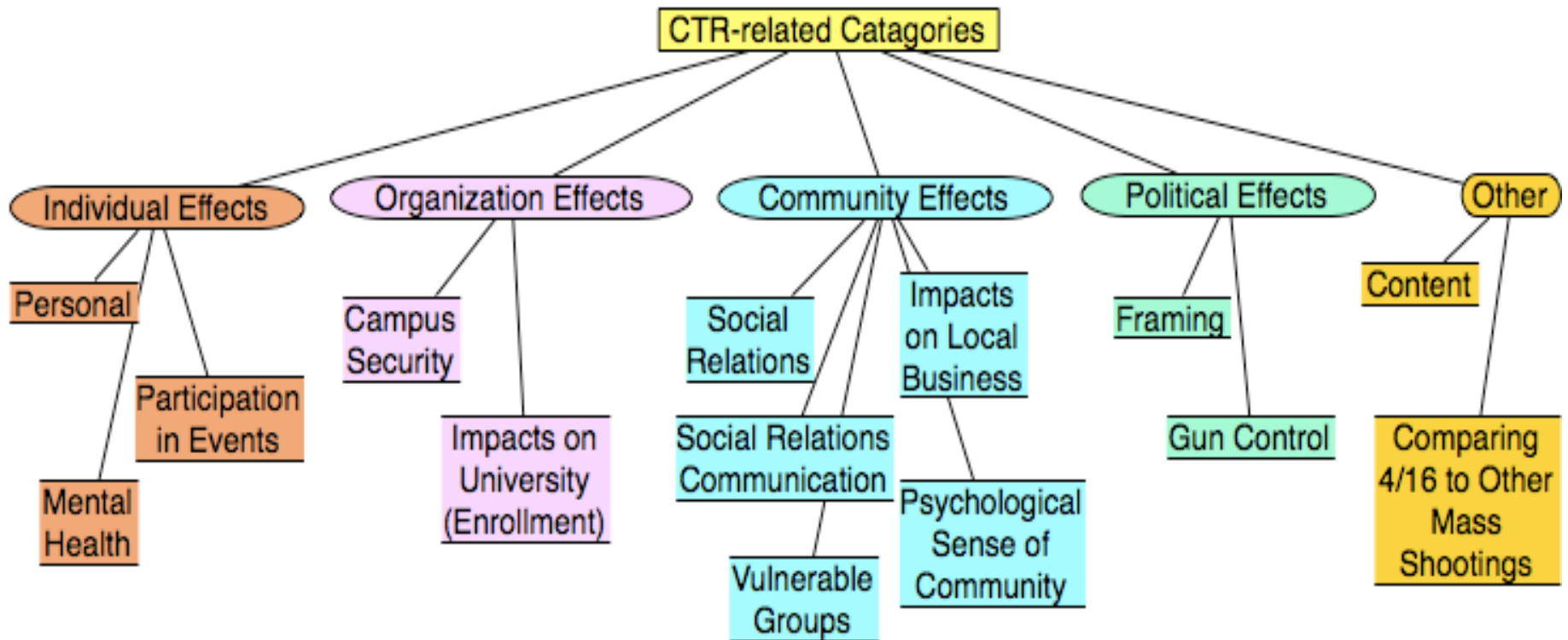
- An ontology is
 - Collection of words and phrases
 - Graph of relationships, sometimes mostly hierarchical as in a taxonomy
 - Covering the key concepts and terms in the discipline
 - Supporting differing views on the field
- Purposes
 - Describes documents
 - As tags, descriptors, keywords
 - Supports navigation, learning, and research

Goals for Ontology for CTR



Categories from focus group study

Results from focus group interviews following the April 16, 2007 tragedy at Virginia Tech

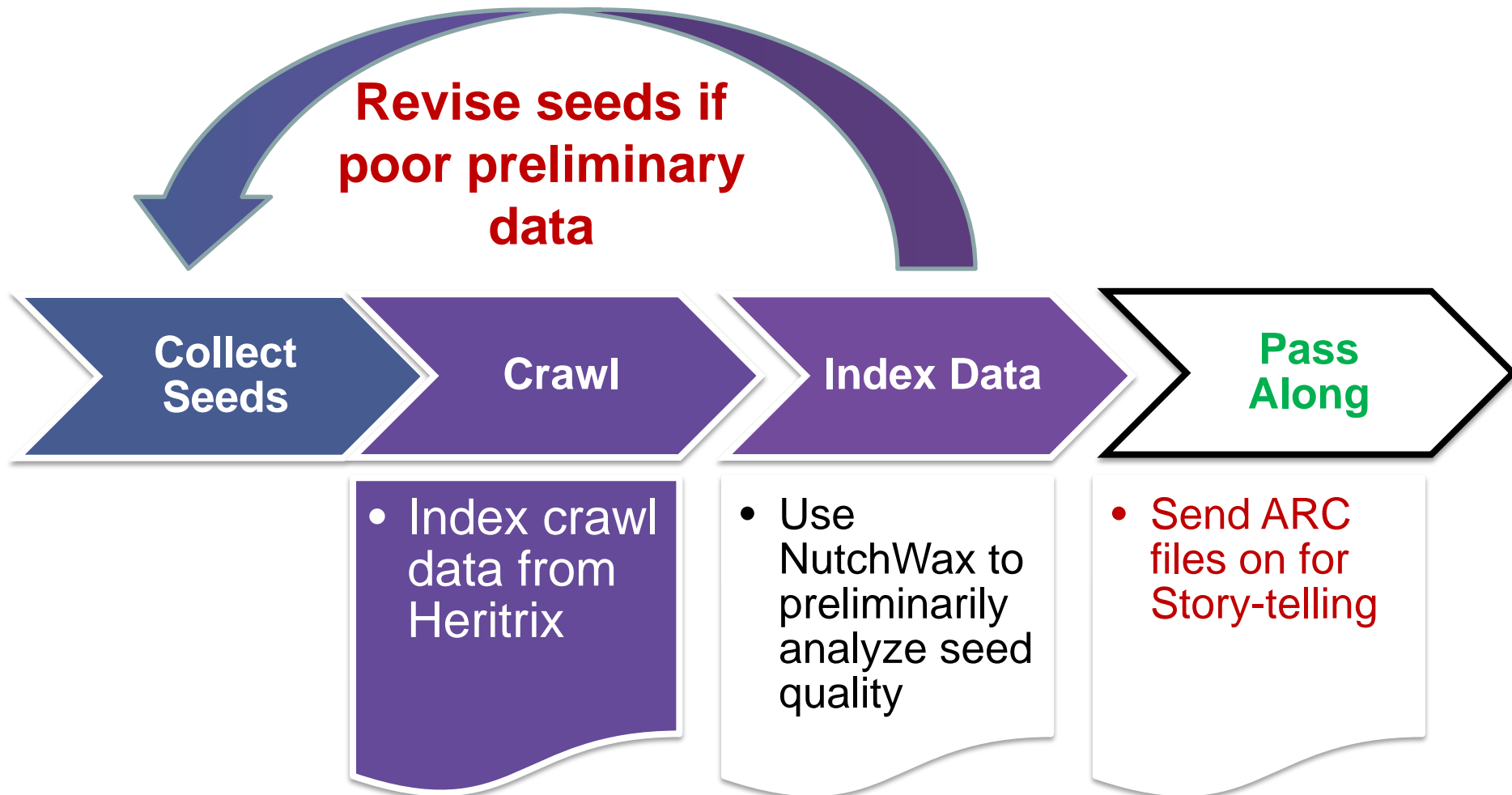


CTR keyword pairs

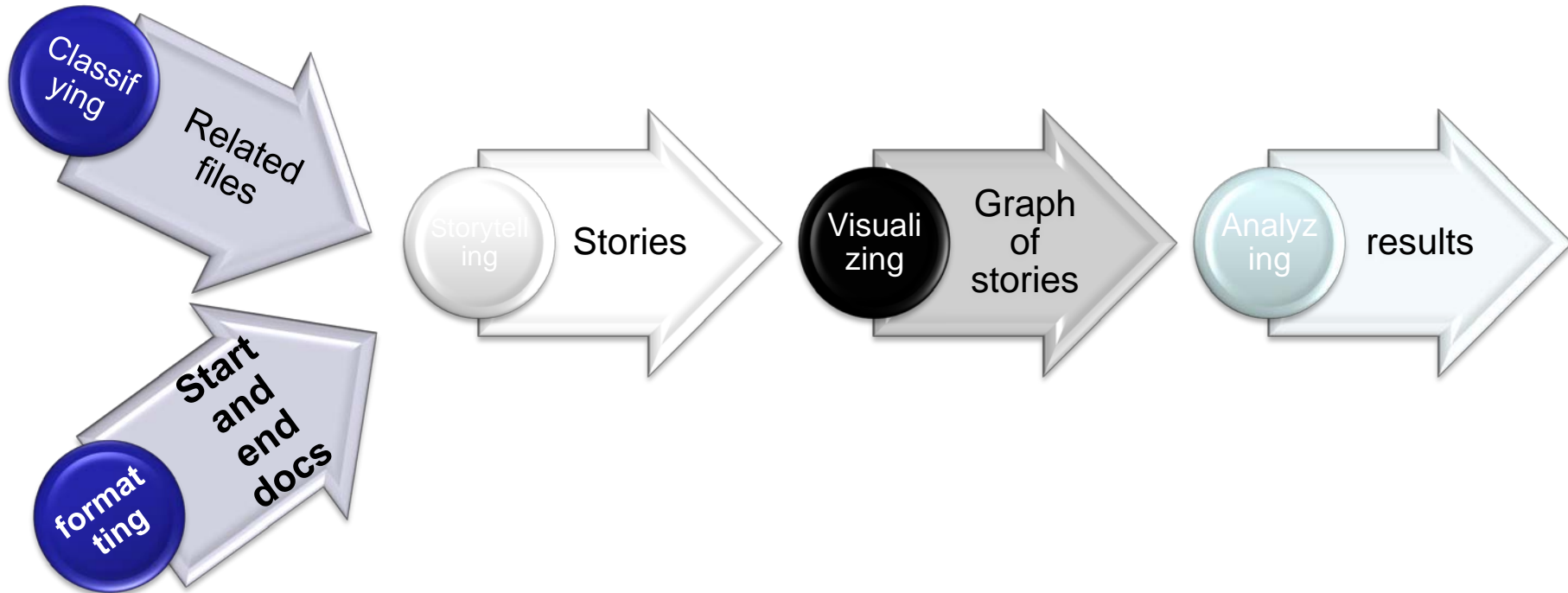
Extracted top keyword pairs from ISCRAM proceedings using the N-gram statistics package

emergency response	decision support	information systems	teams participants
decision making	data models	disaster monitoring	teams maps
command teams	disaster plan	crisis management	sms text-message
flood alerts	information seeking	situational awareness	disaster registry
physical communication	human disaster	teams access	decision preference

Preliminary Web Data Analysis

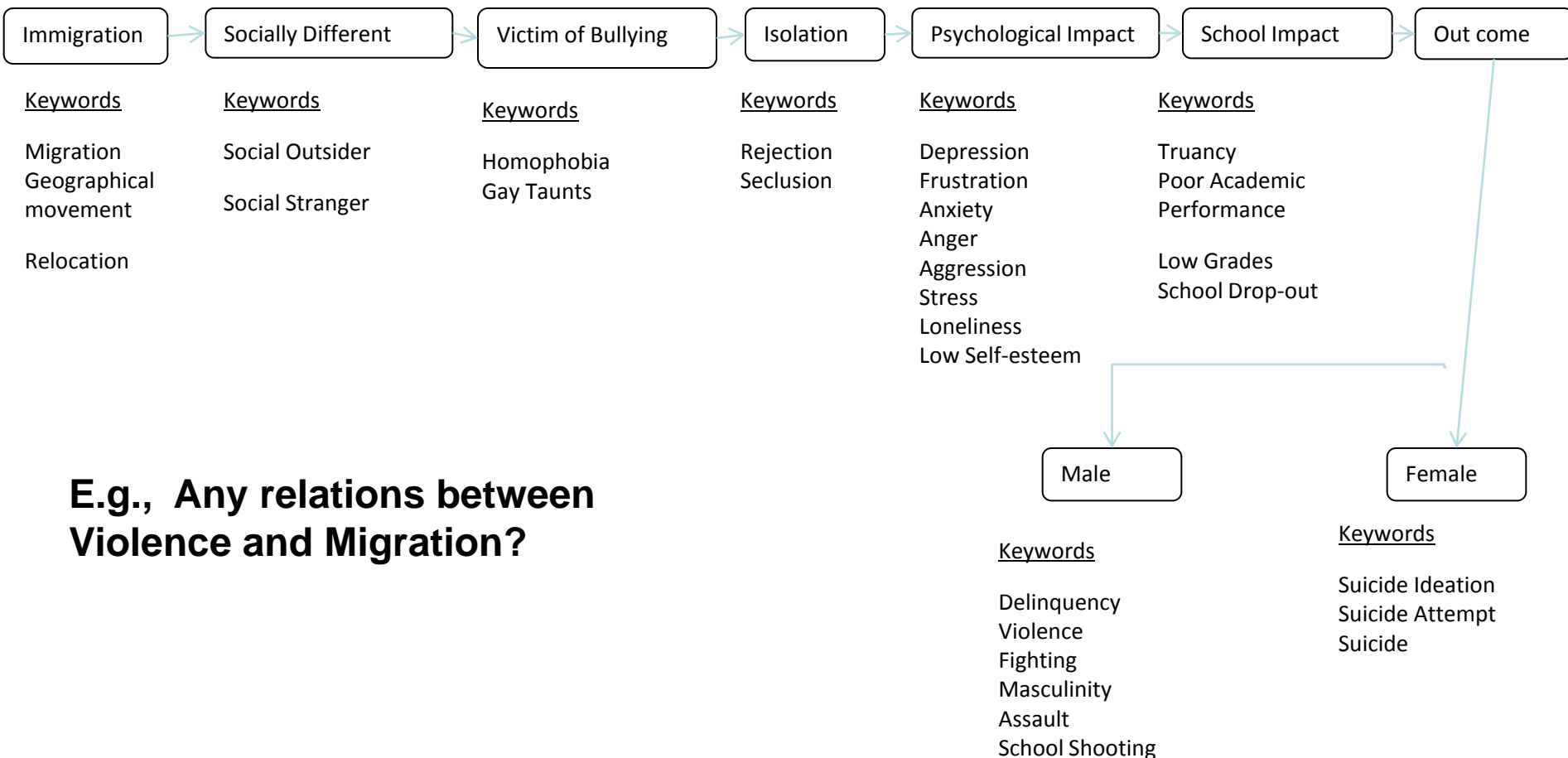


Data Cleaning, Extracting, Filtering, Storytelling



- Goal: Discover relationships between tragic events.
 - Examples: Shooting, study pressure, and mental disorders

Hypothesis of school shooting



E.g., Any relations between Violence and Migration?



Connecting Computing Educators

[ENSEMBLE HOME](#)[LOGIN](#)[SEARCH](#)[ABOUT](#)[CONTACT](#)

Browse
our Collections



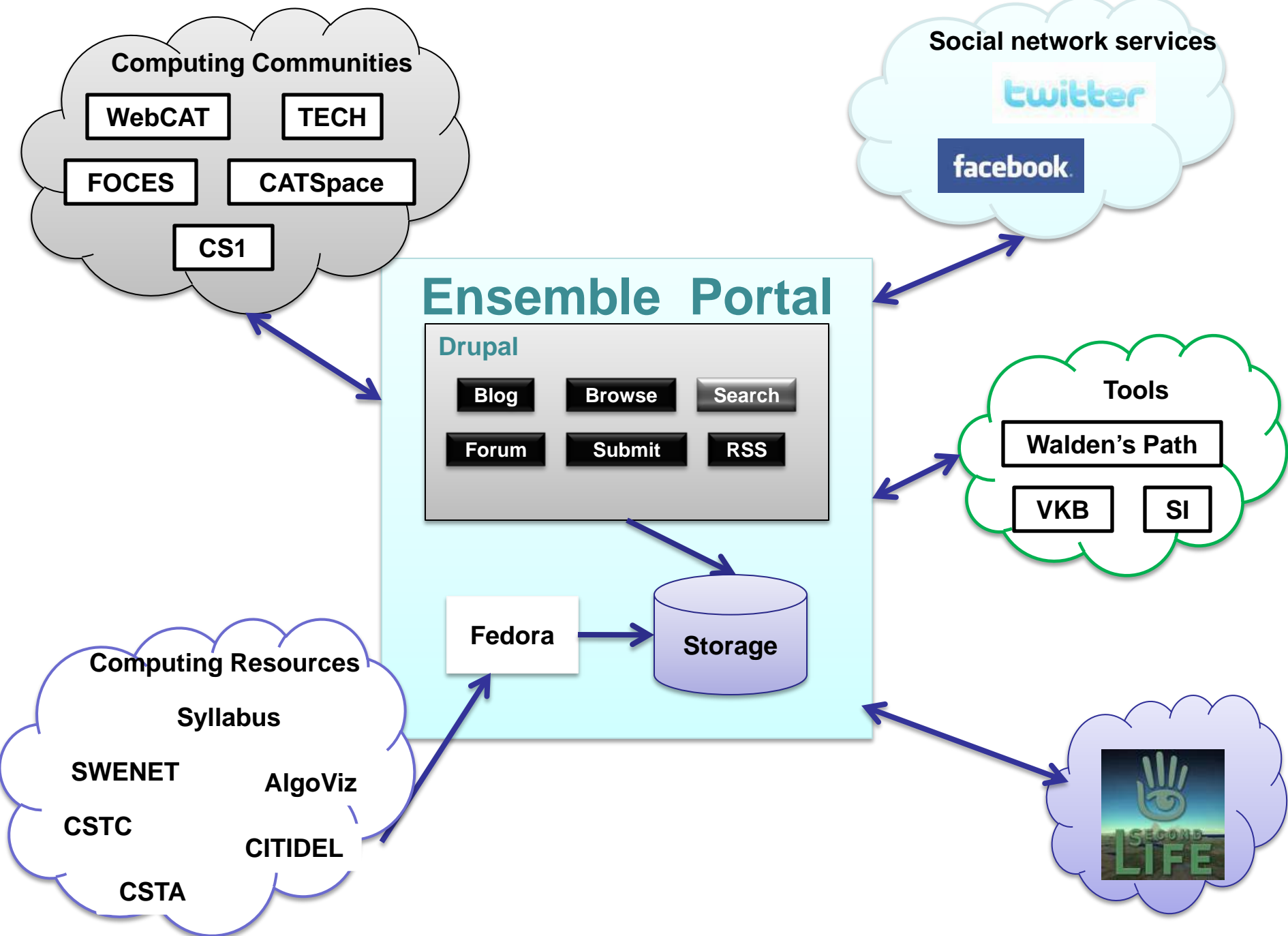
Explore
our Communities



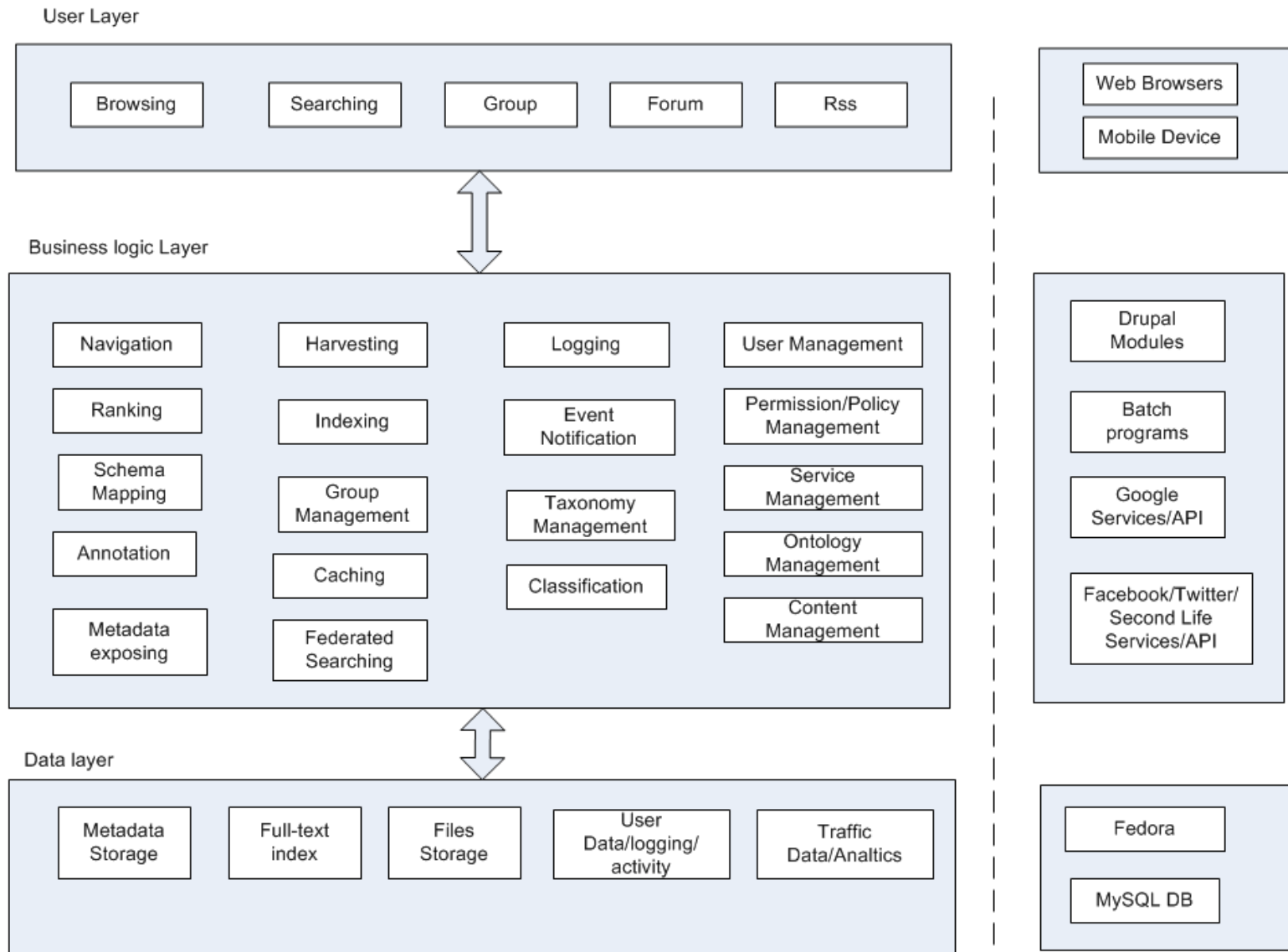
Try
our Tools

Powered by **Drupal**, an open source content management system. Arch images by criminalintent, auntiaura, and geishaboy500. CC BY-SA 2.0

www.computingportal.org



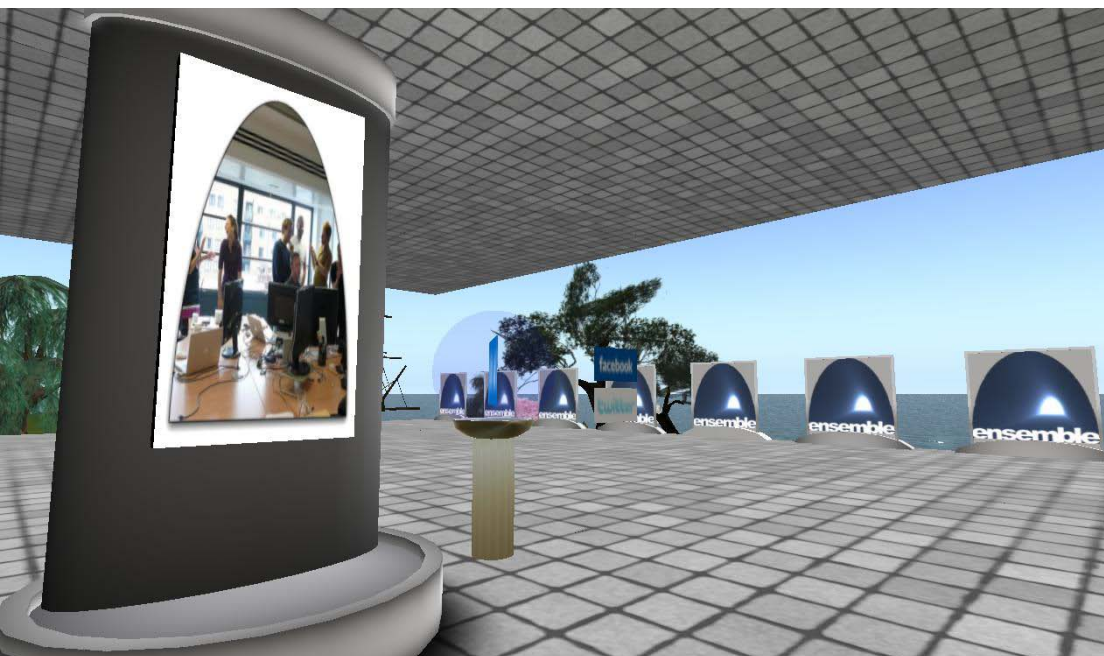
Ensemble Portal Architecture





Ensemble in Second Life

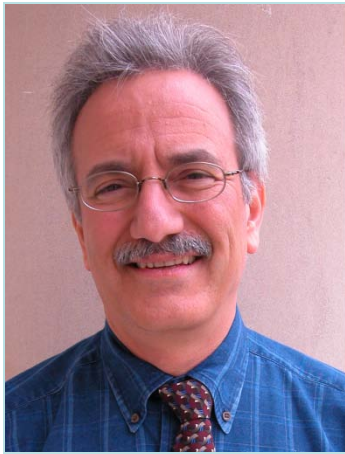
<http://slurl.com/secondlife/Educators%20Coop%204/66/236/28>



The Ensemble Pavilion offers:

- teleports to other computing sites in Second Life like the *Digital Preserve*
- hyperlinks to related computing websites
- RSS readers with feeds from computing and computing education blogs
- membership in the Ensemble Computing group in Second Life, Facebook, and Twitter

Selected Digital Preserve Personnel



Gary Octagon
Gary Marchionini



EdFox Rieko
Edward Fox



zamfir Paule
Spencer Lee



mantruc Martian
Javier Velasco-Martin



Uma Aldrin
Uma Murthy



Krad Proto
Seungwon Yang

DP areas

Poster Building

- 18 posters on display
- Poster view tips
- Video screen



Cafe

- Beverages
- Screens
- Discussion areas

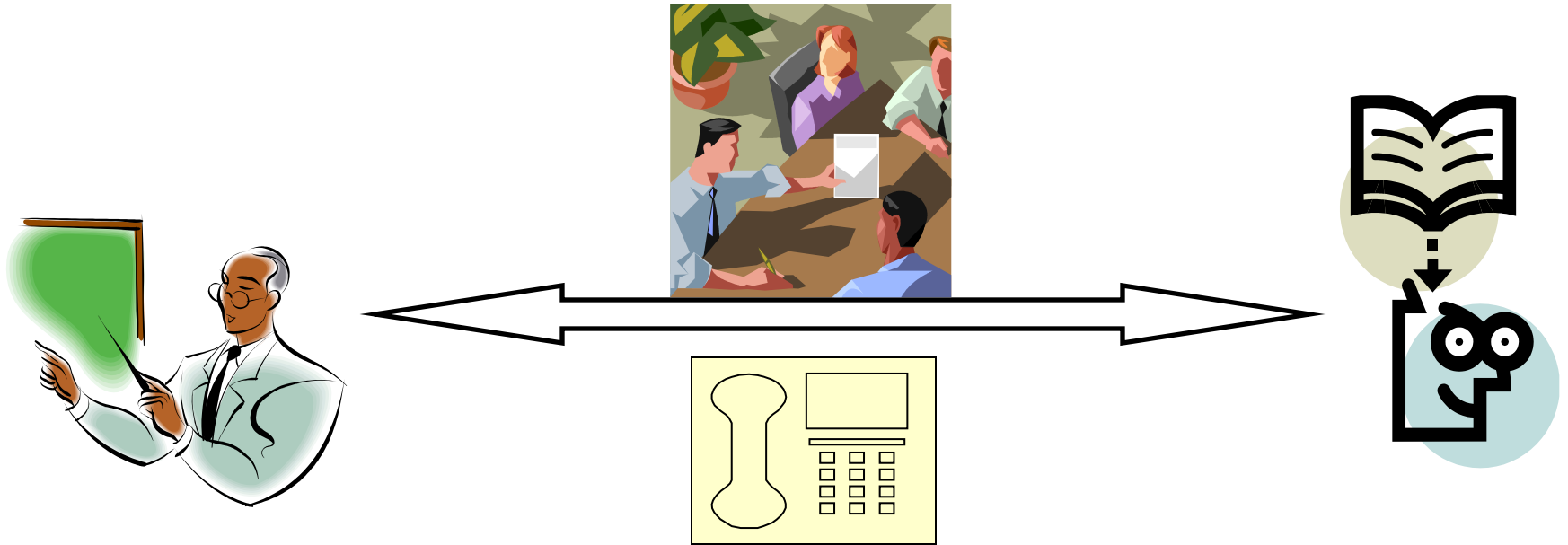
Part 3 – Basic DL Concepts

- Digital Library Scope
- OAI
 - Harvesting
 - Repositories
- Space-related Perspectives of Computing
 - Distributed
 - Cloud ...
- 5S -> Content + Society

DL Scope

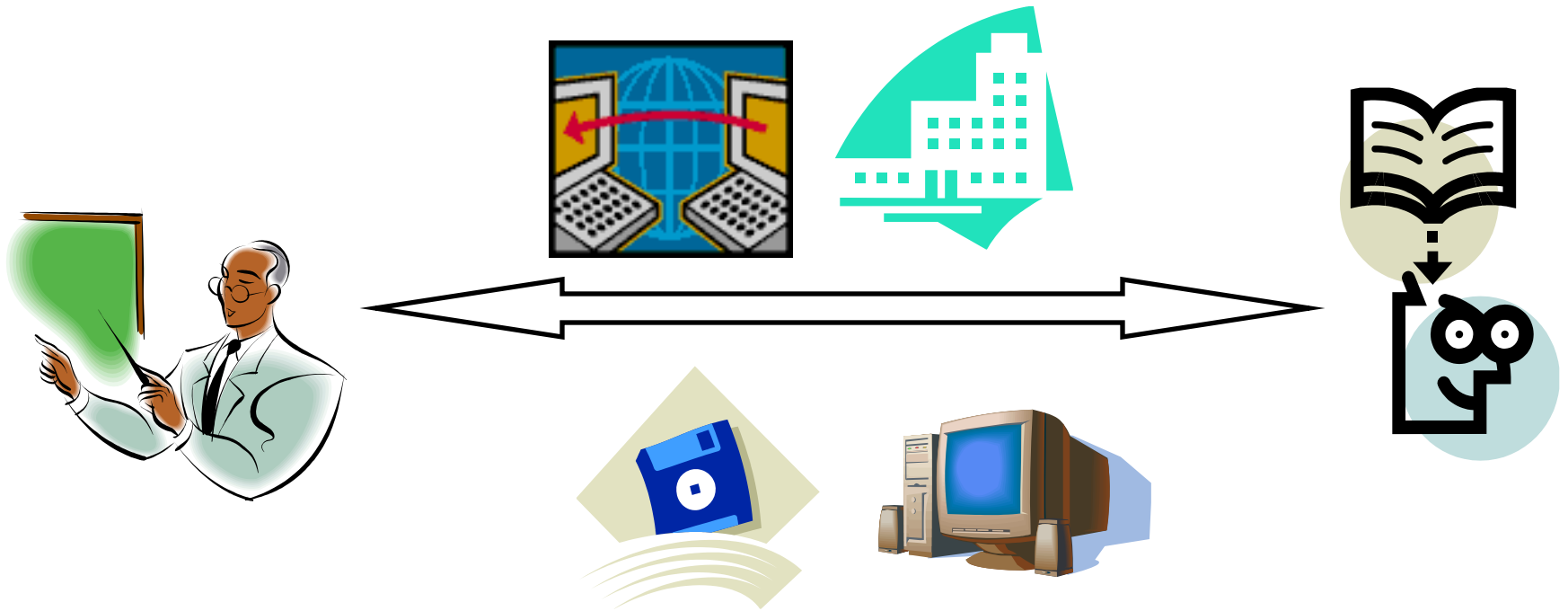
- Institutional repositories
- Open archives
- Electronic/virtual libraries
- Content management systems
- Courseware management systems
- Personal information management systems
- Cloud/ubiquitous/... computing

Synchronous Scholarly Communication

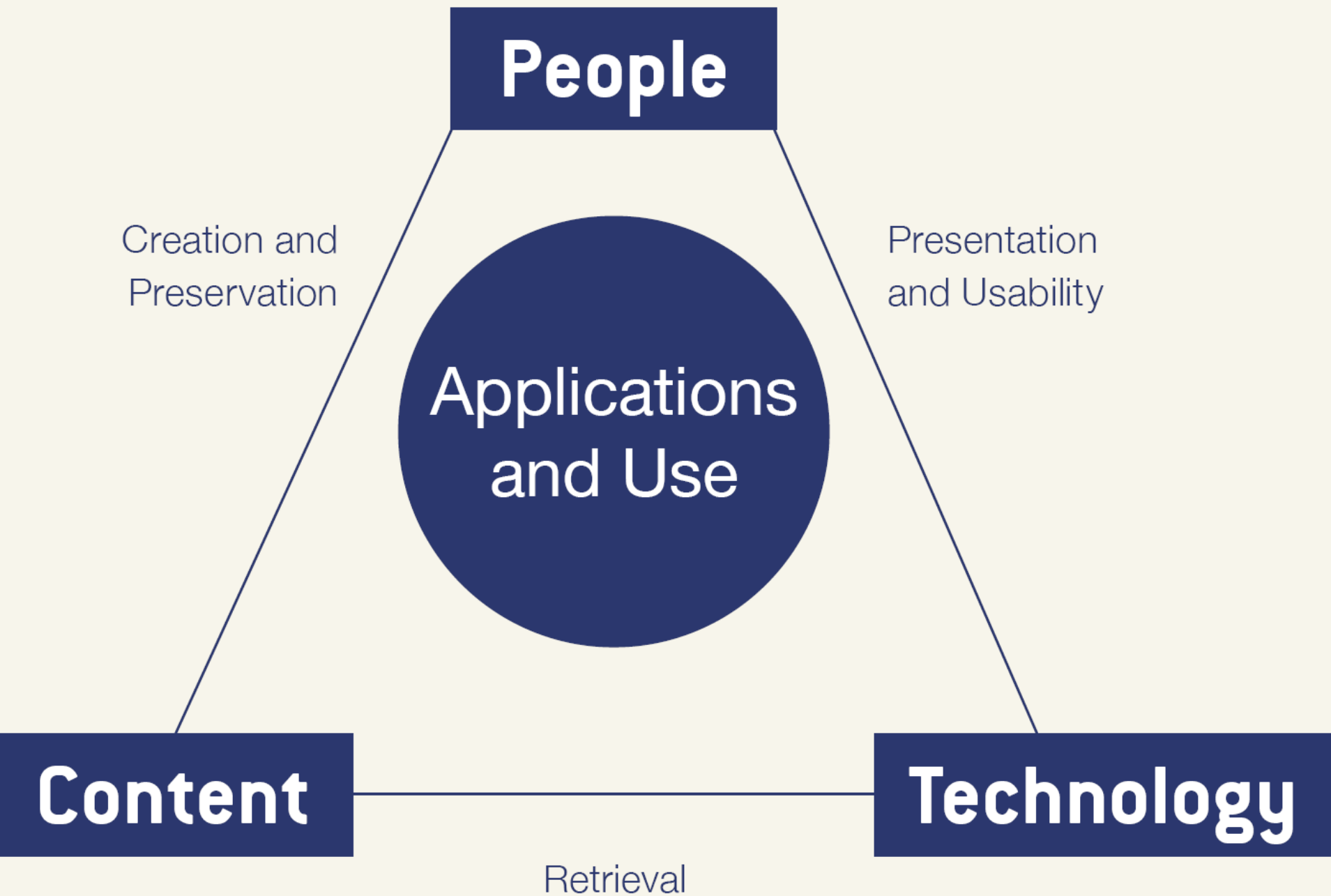


Same time, Same or different place

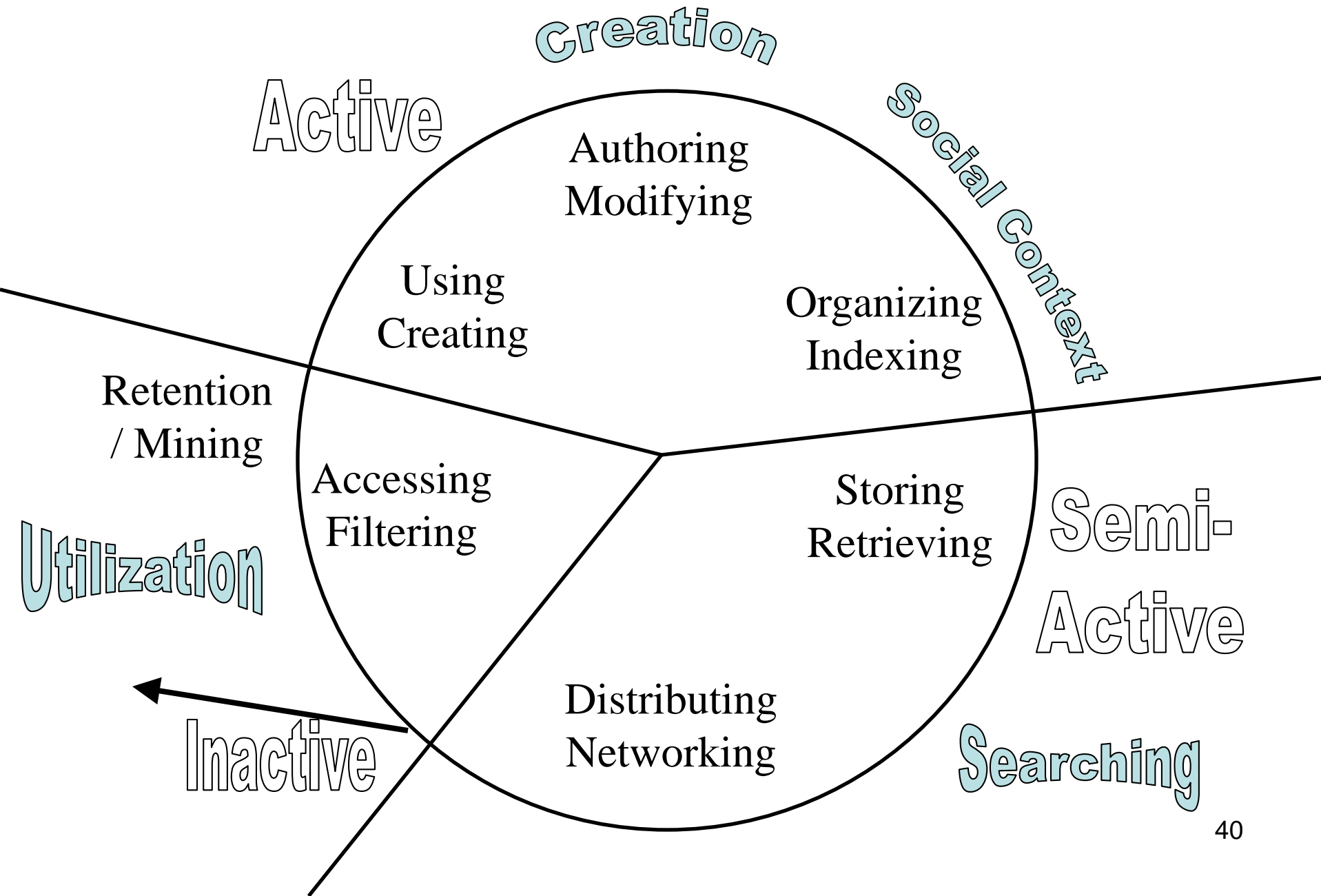
Asynchronous, Digital Library Mediated Scholarly Communication



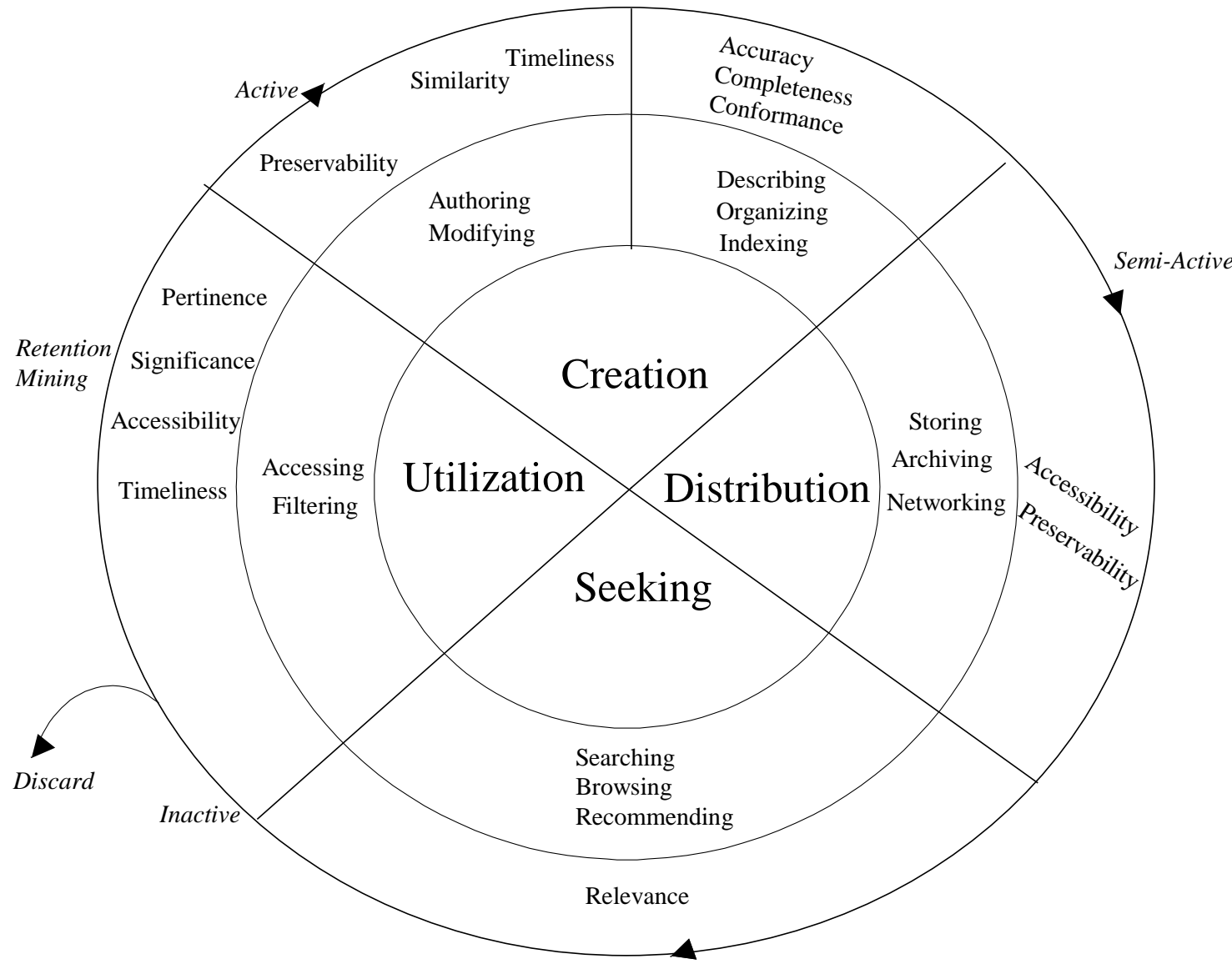
Different time and/or place



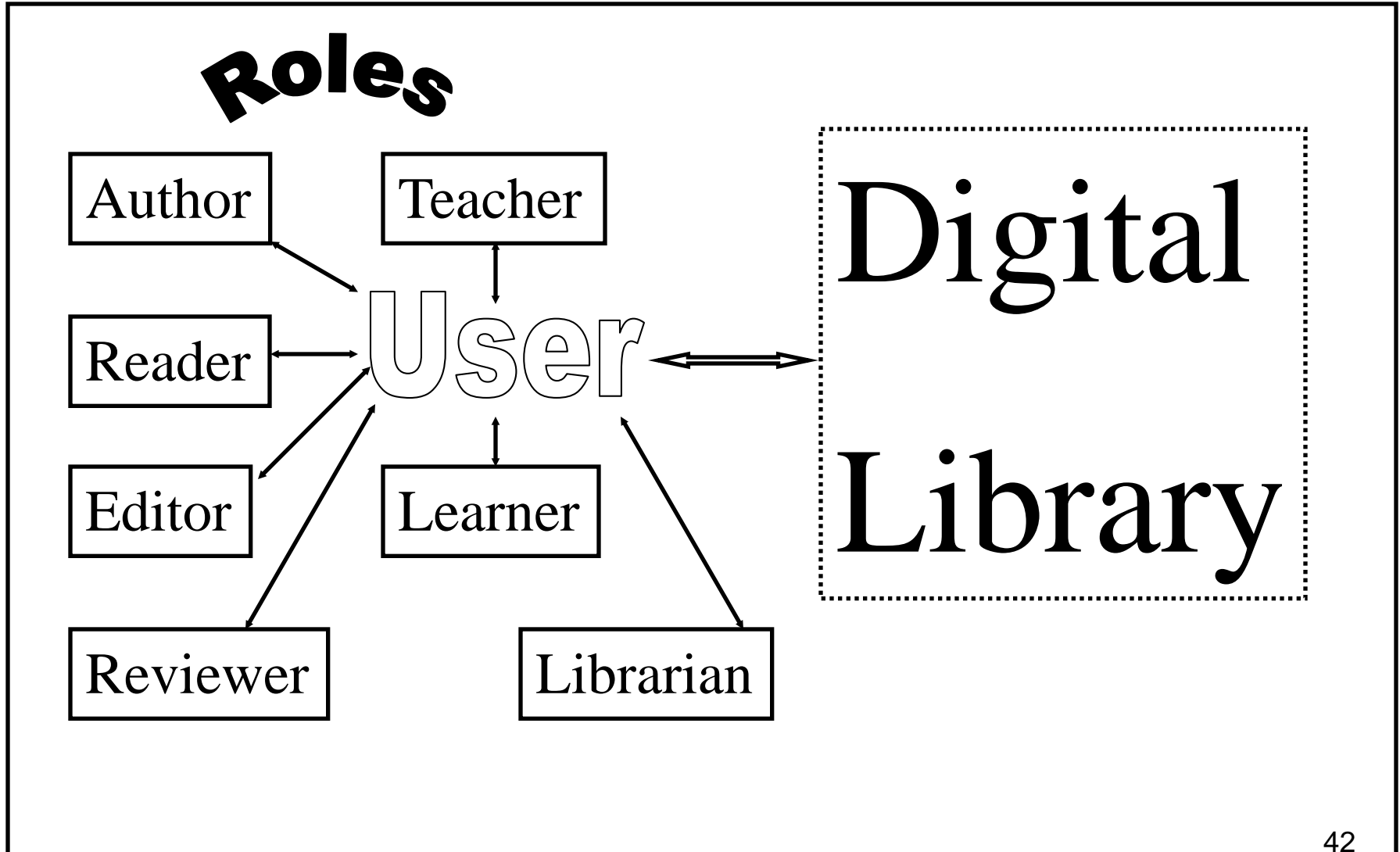
Information Life Cycle



Quality and the Information Life Cycle



DLs Shorten the Chain to



Degree of Structure

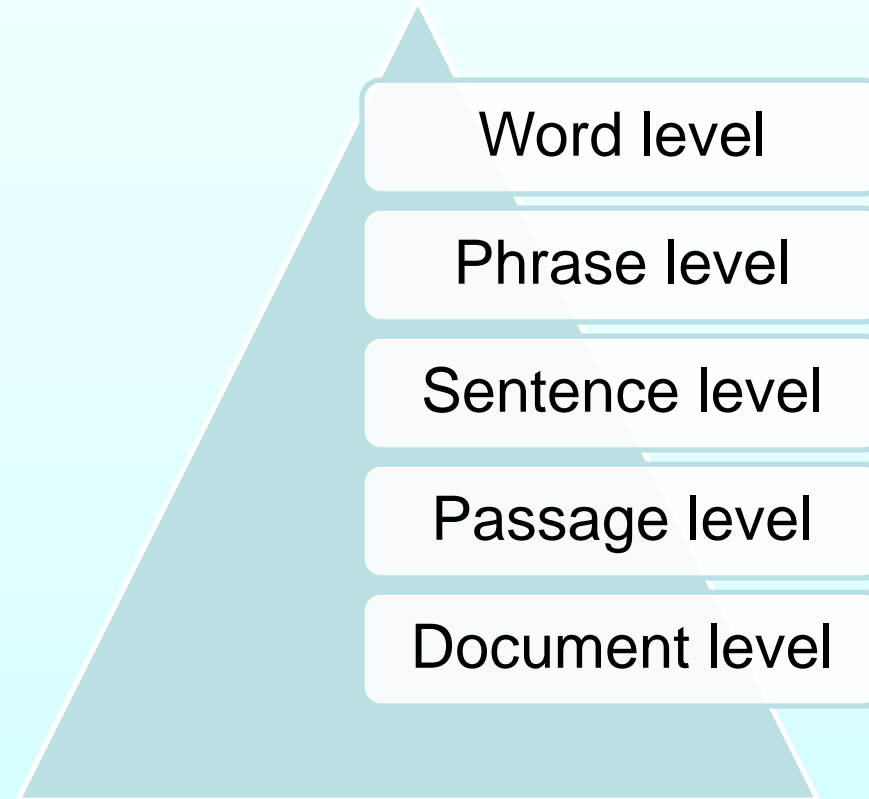
Web

DLs

DBs

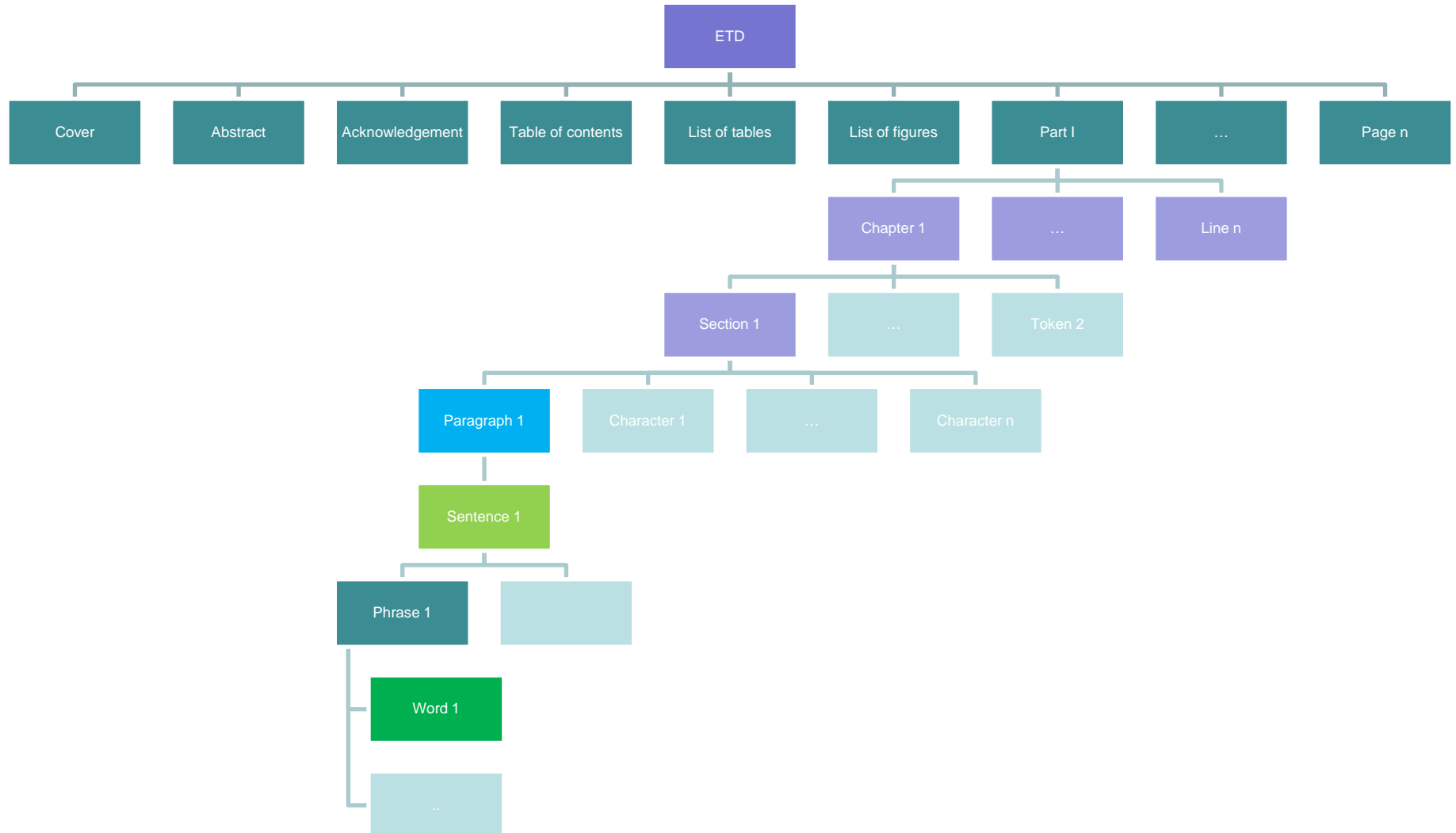


Example of Granularity of Information Structure

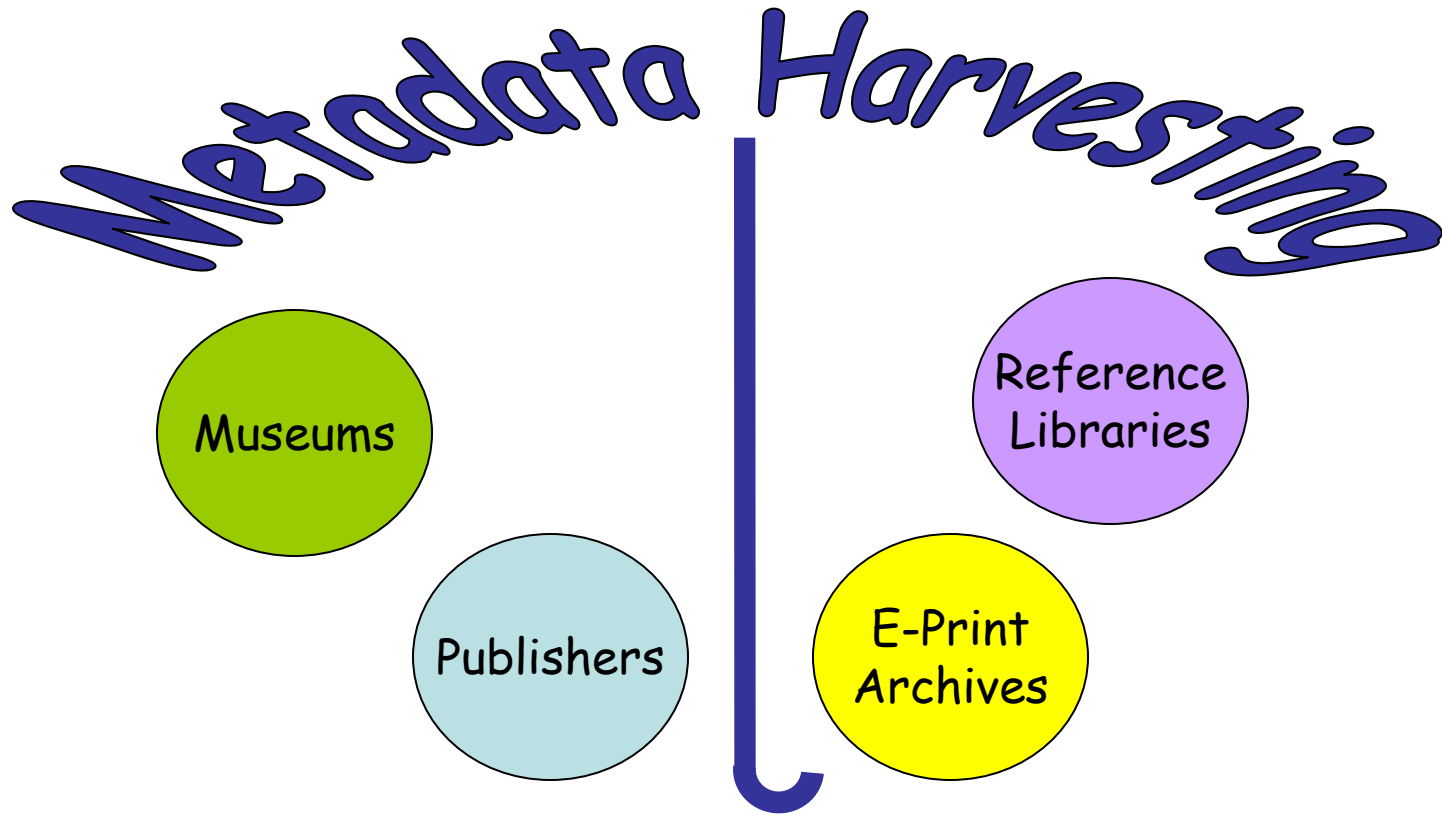


**Example of Structural Level
of Text Information**

ETD Logical Hierarchy



OAI = Technical Umbrella for Practical Interoperability...



...that can be exploited by different communities

OAI – Repository Perspective

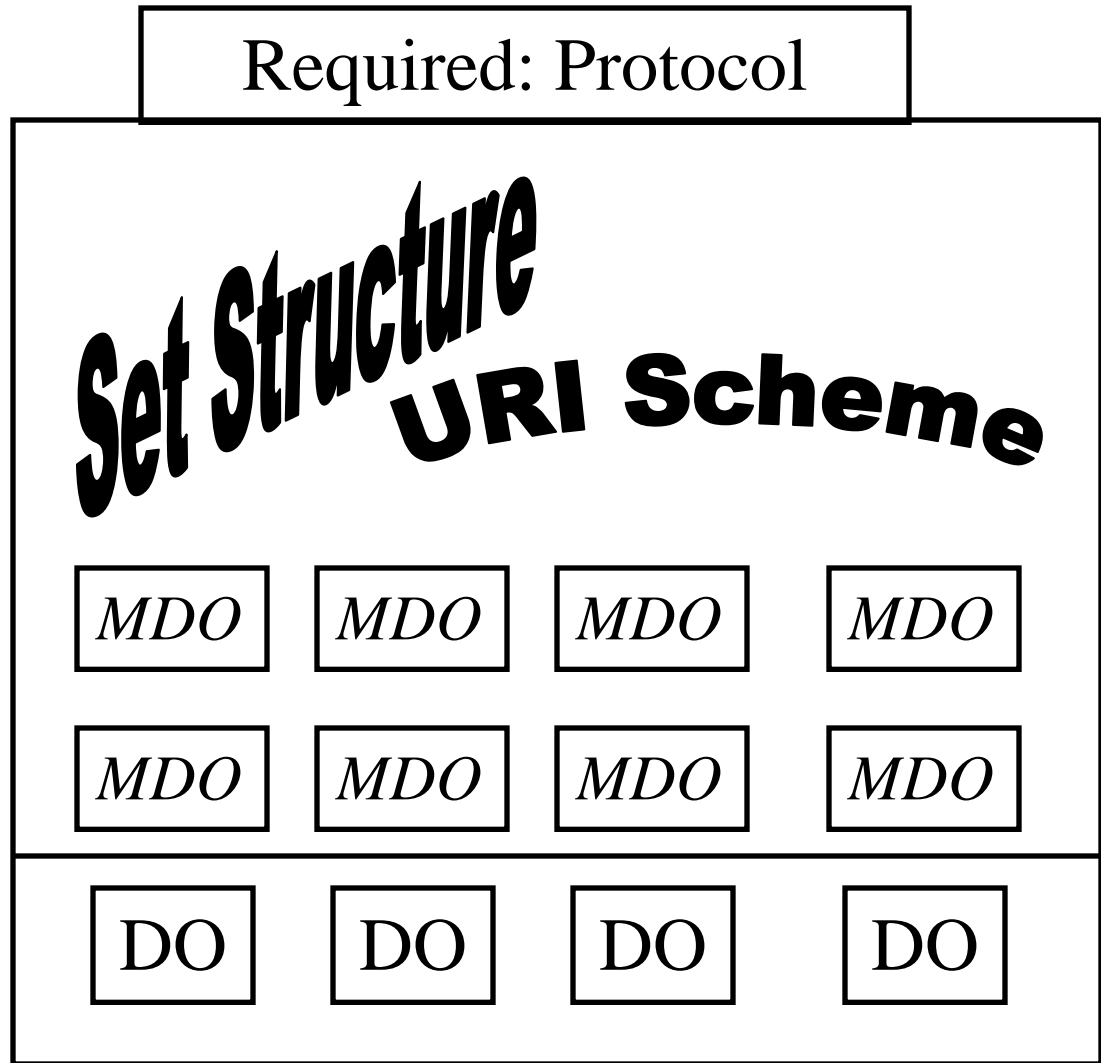
Glossary:

DC=Dublin Core

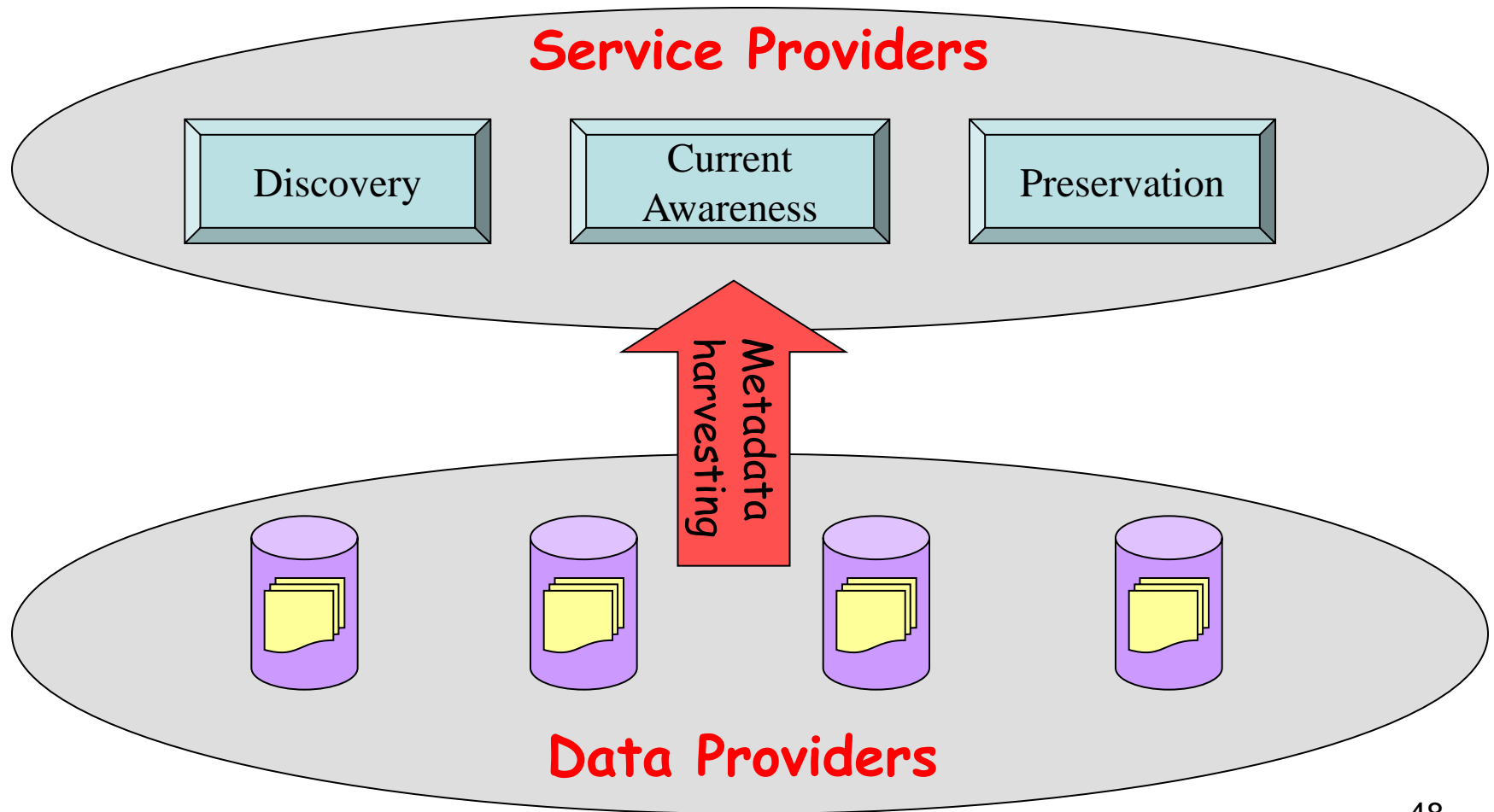
MDO=Metadata Object

DO=Digital Object

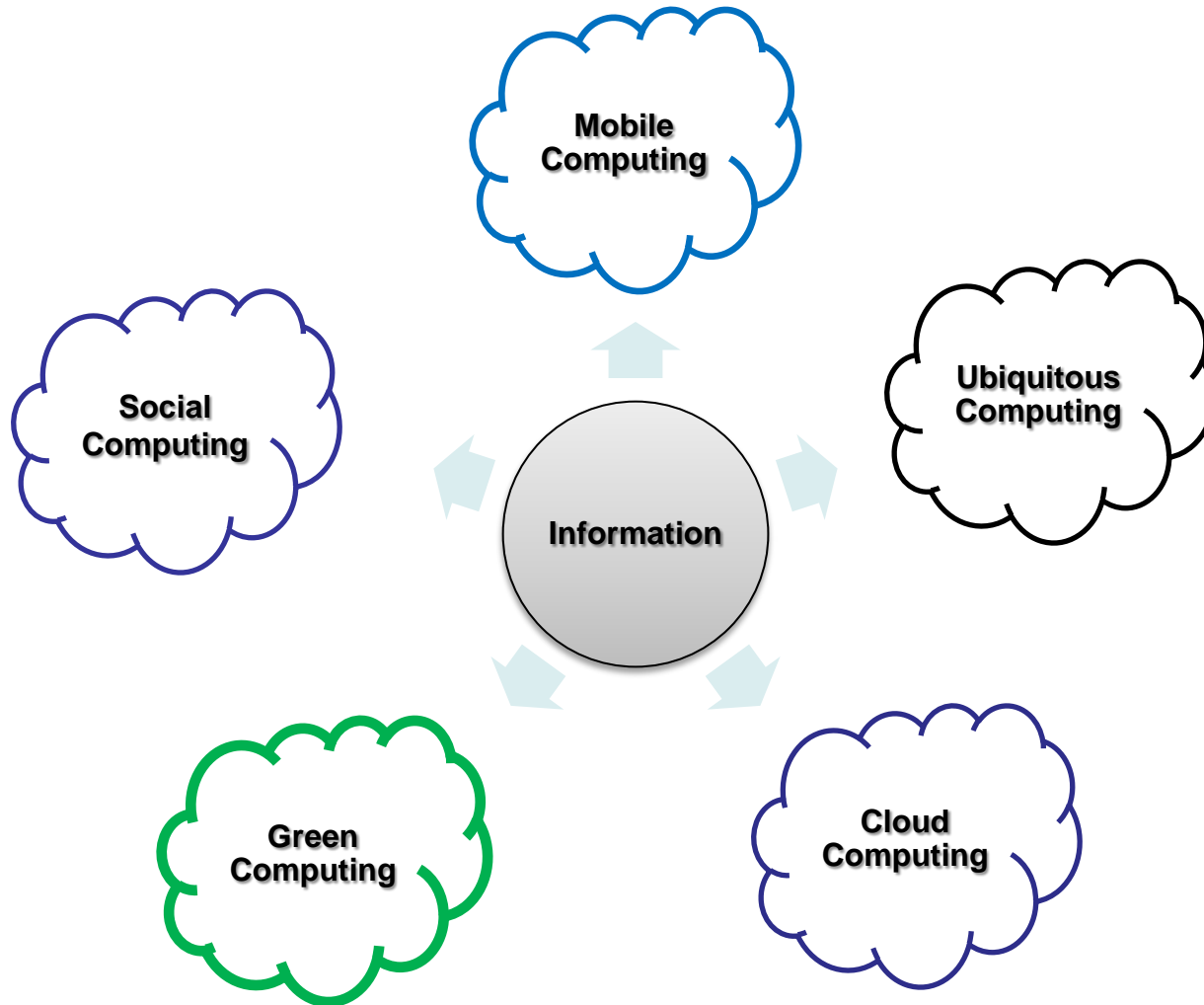
Required: DC



The World According to OAI



Space-related Computing



5S Layers

Societies

Scenarios

Spaces

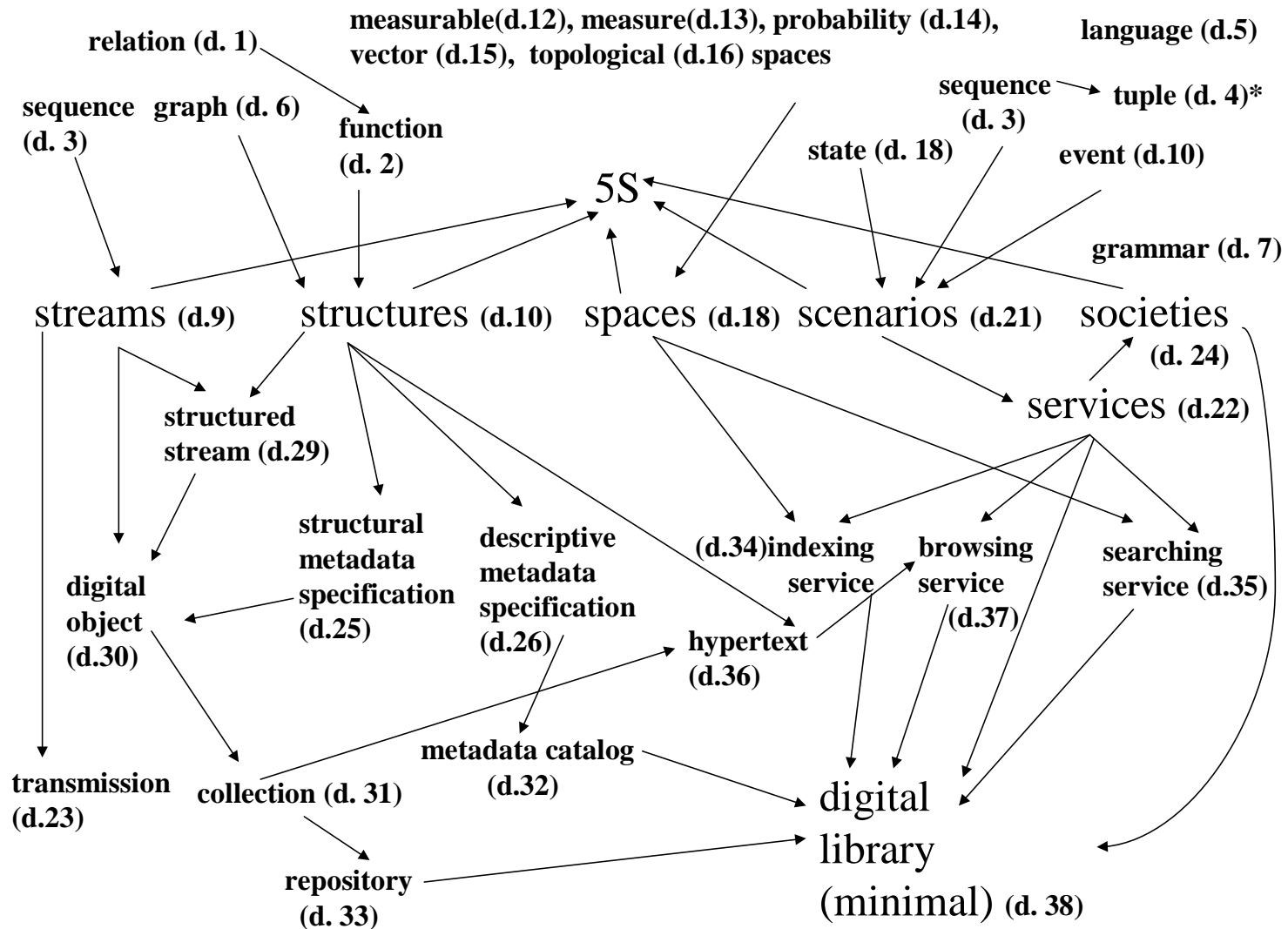
Structures

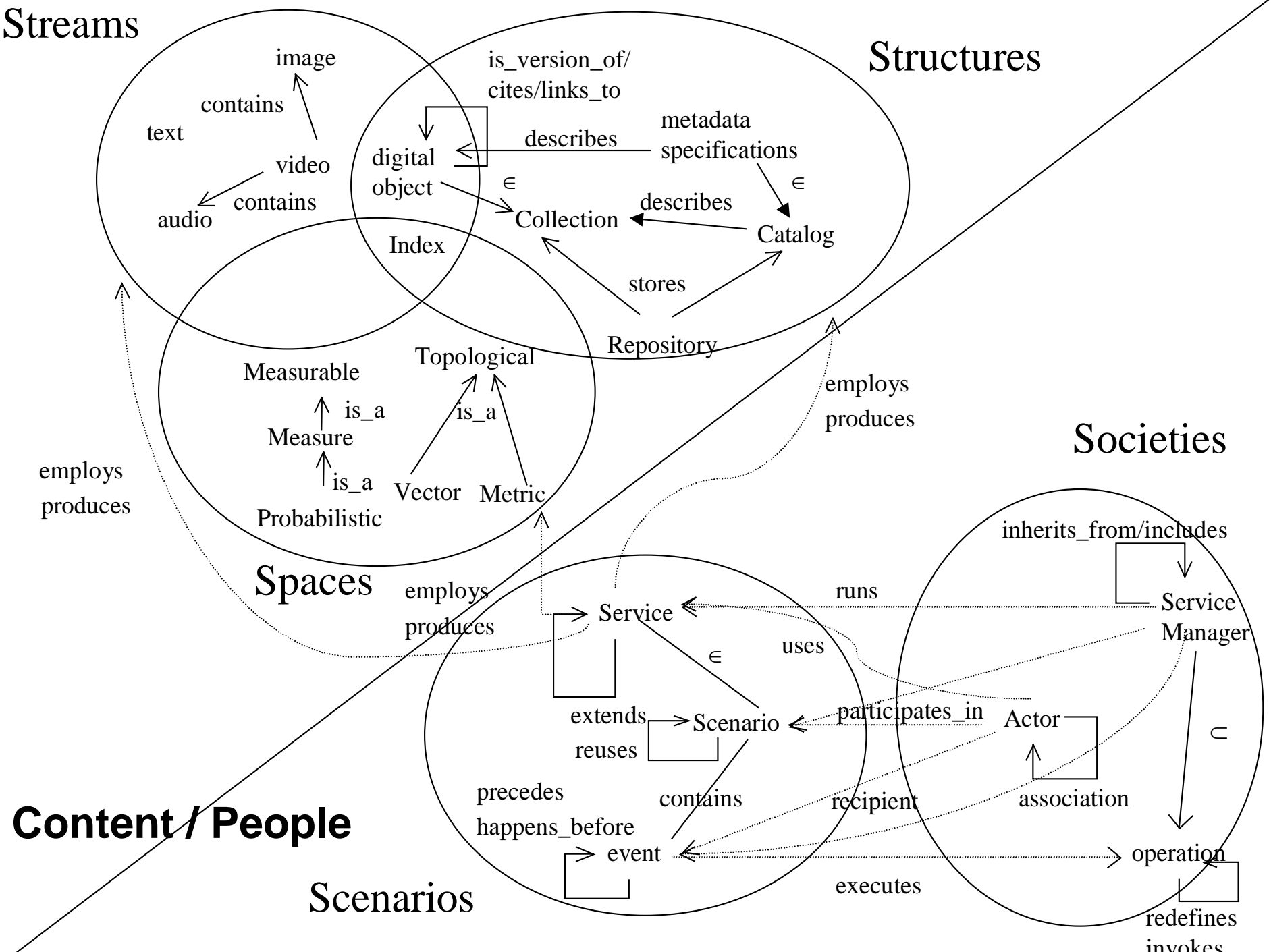
Streams

5S Contextualized

- Societies/communities/users served
- Scenarios/services supported
- Management of physical/conceptual/
feature spaces
- Use of structures/organizational devices
- Streams of content and communication

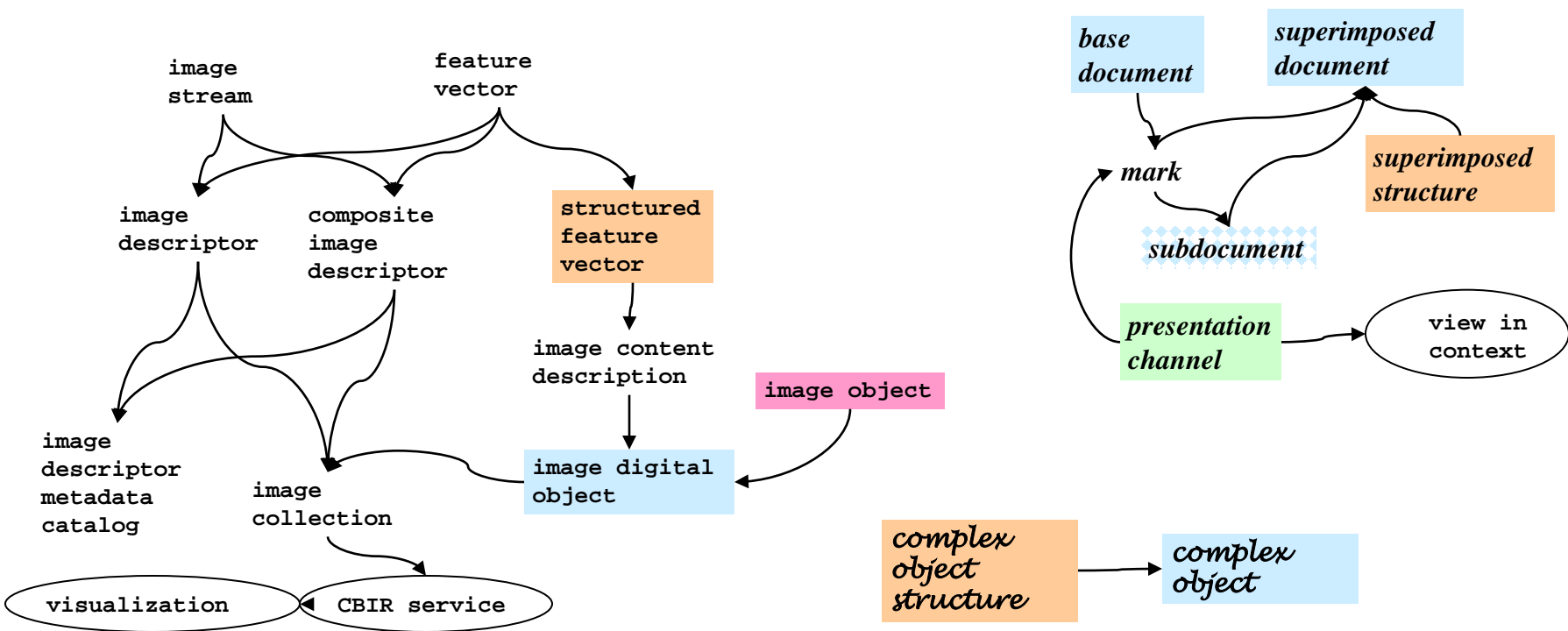
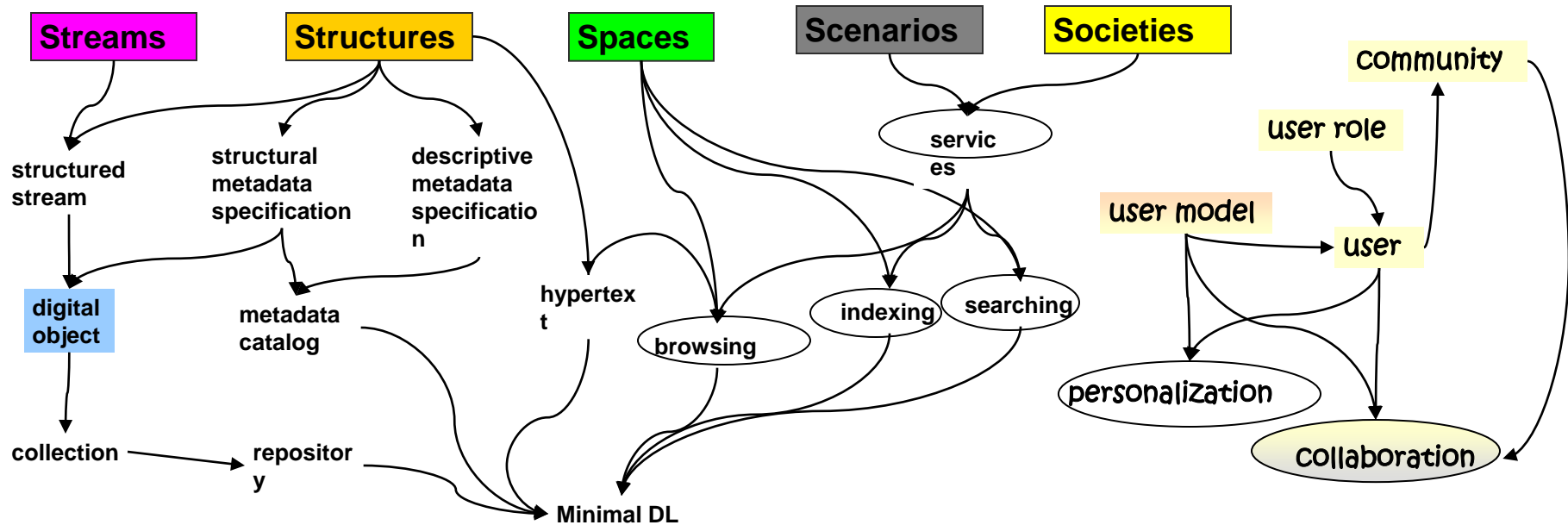
5S and DL formal definitions and compositions (April 2004 TOIS)



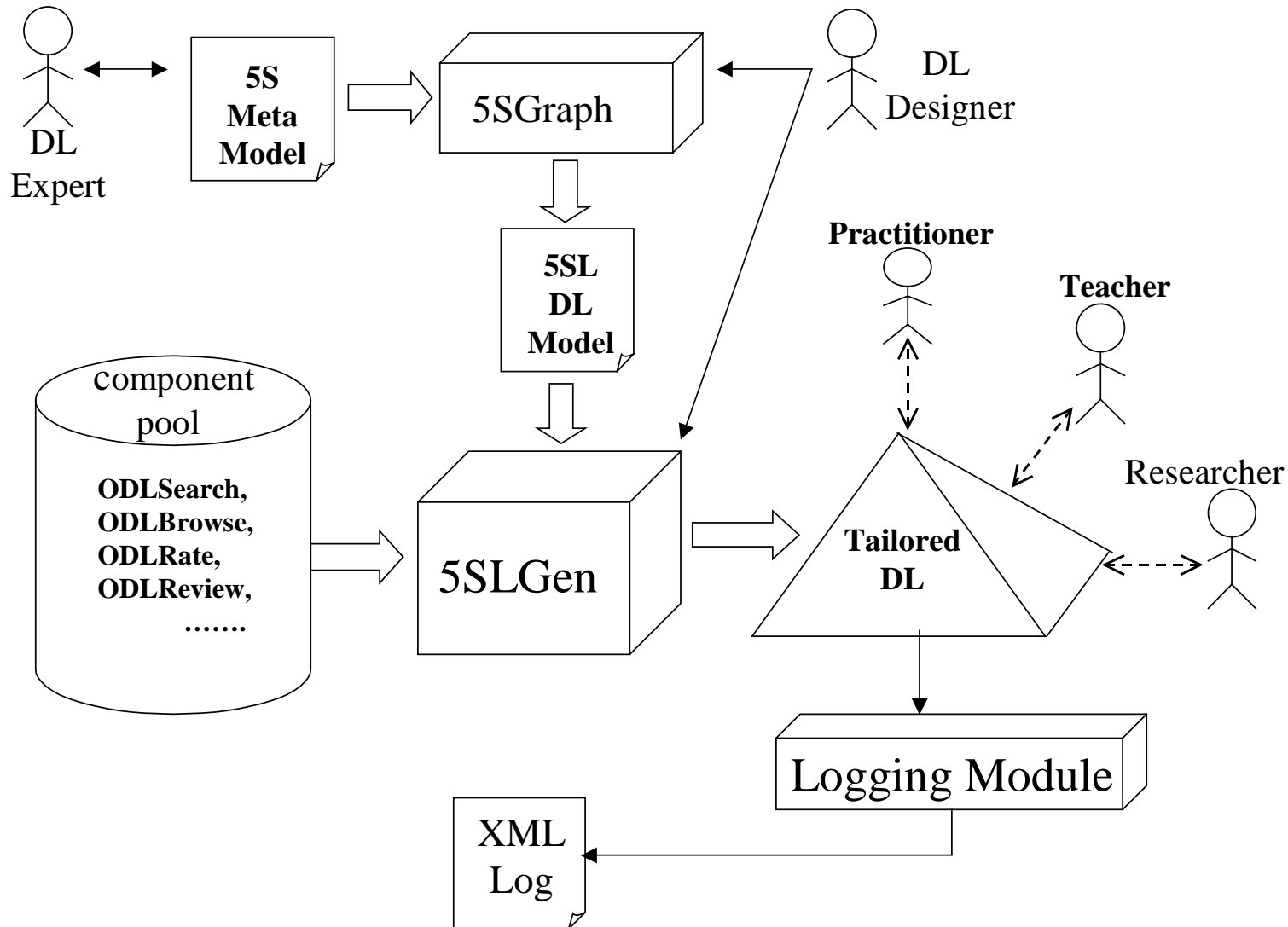


Extending 5S

- Higher DL Constructs
 - Collections
 - Catalogs
 - Repositories and Archives
 - Systems
 - Case Studies
- Specialized views and services



Tools/Applications



Content Centered

- Genre
 - Gray literature
 - Report, courseware
 - Posters, demos, tutorials, panels, debates
- Format (Humanities needed TEI)
- Presentation (Enhanced expressivity)
- Preservation (Rationalizing digitization)

Society Centered

- Society, community, group, user
- Web 2.0, Social networking
- Computer-supported cooperative work
- User modeling
 - Authors, committee/peers, readers
- Economics / culture
 - Free: but who actually pays, how, implications
 - Low cost: prepaid, but what of preservation
 - Repository hierarchy: group, institution, nation

Part 4 – NDLTD

- Roles
- Member Support
- Workflow
- Content
- Support for
- Union Catalog

The Networked Digital Library of Theses and Dissertations

www.NDLTD.org

Advocate:

Training Authors

Expanding Access

Preserving Knowledge

Improving Graduate Education

Enhancing Scholarly Communication

Empowering Students & Universities

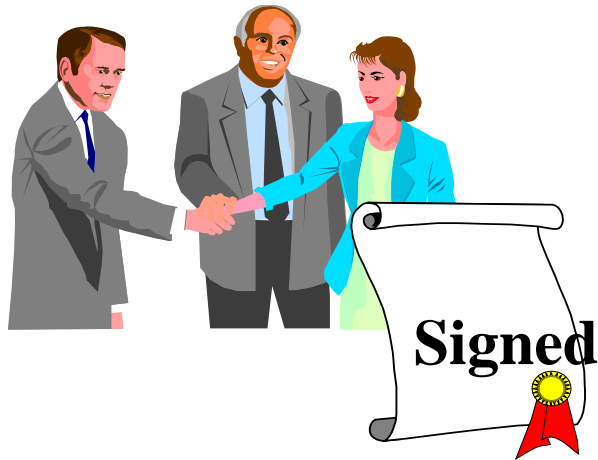
Leader of the Worldwide ETD

(Electronic Thesis and Dissertation) Initiative

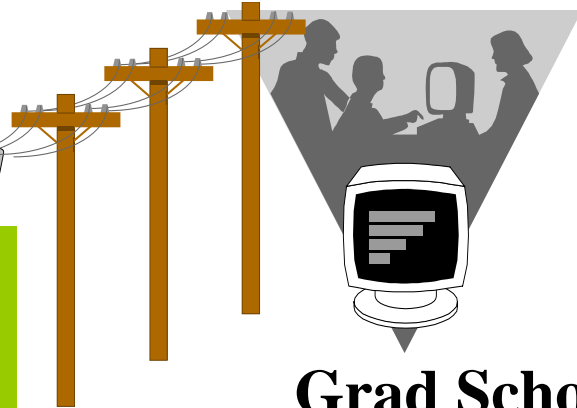
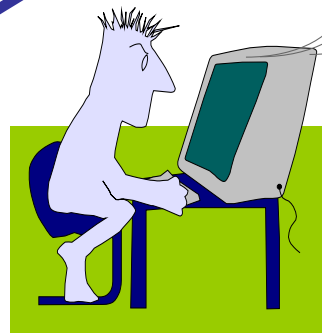
Member Support

- Advocacy for ETD activities
- Annual conference
- ETD-L – listserv for discussion
- Union catalog: improving local data
- Best X ETDs in Y lists
- Services for access
- Information for ETD projects
 - Standards, documentation, best practices
- Connect with: institutional repositories, ...

Student Gets Committee Signatures and Submits ETD

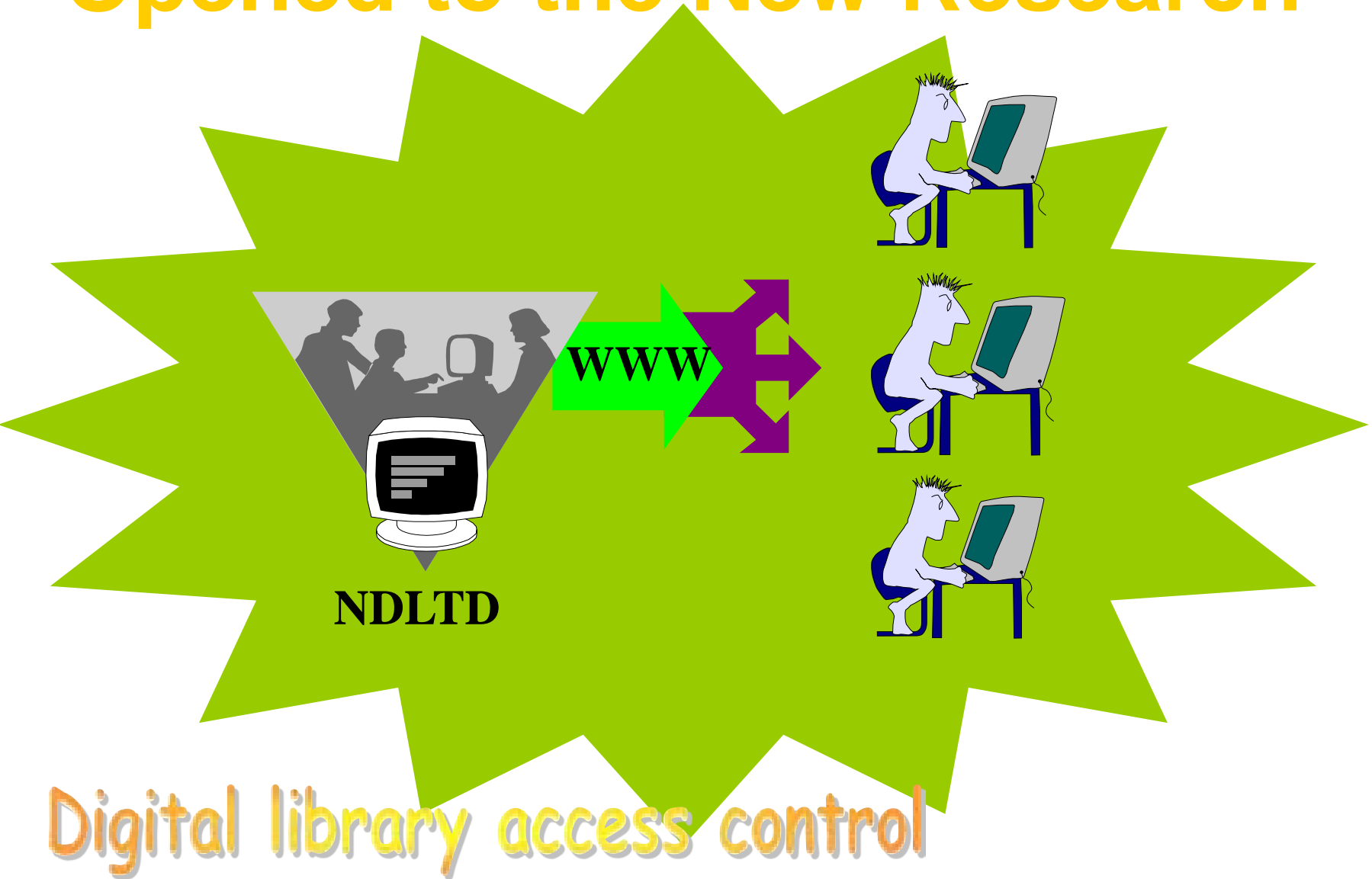


**Approval
form**



Grad School

Library Catalogs ETD, Access is Opened to the New Research



Support For Students

- Learn about
 - Electronic publishing
 - Concepts, computational thinking
 - Tools and services
 - Intellectual property rights
 - Digital libraries
- More visibility
 - Discover (potential) collaborators
 - More citations

Support For Universities

- Improve graduate education
- Save time, space, and money
- Increase visibility of research
 - Through open access
- Expand capabilities with digital libraries
 - Including institutional repositories
- Collaborate in region, nation, and internationally through the ETD movement

Union catalog statistics 5/2010

- WorldCat has 13.4M records w. 502 MARC field indicating “thesis”
- 915K records are ETDs (MARC 502, 856)
- 630K ETDs added to WorldCat in 21 months
- Uncertainty
 - ETDs in Union Catalog not in WorldCat
 - Union Catalog duplicates
- OCLC data provider; services provided by Scirus, VTLS, ...

Part 5 – Services Centered

- Taxonomy
- Interoperability, integration, packaging
 - HTML5
- Collaboration, annotation, recommending
- Indexing, CBIR
- Categorizing, browsing
- Roles of librarians

Infrastructure Services			Information Satisfaction Services
<i>Repository-Building</i>		<i>Add Value</i>	
<u>Creational</u>	<u>Preservational</u>		
Acquiring Cataloging Crawling (focused) Describing Digitizing Federating Harvesting Purchasing Submitting	Conserving Converting Copying/Replicating Emulating Renewing Translating (format)	Annotating Classifying Clustering Evaluating Extracting Indexing Measuring Publicizing Rating Reviewing (peer) Surveying Translating (language)	Browsing Collaborating Customizing Filtering Providing access Recommending Requesting Searching Visualizing

DL.Org Functionality WG

Dagobert Soergel – Sci. Lead:

Functions where Interoperability is important

Behind the scene	For users
Feature extraction	Federated search
Classification / clustering	Incorporating content from other places on the fly
Sharing authority files	Display and visualization
Log file analysis	Timelines
Sharing user profiles	Maps
Harvesting , aggregating	Playing videos
Shared storage and backup	Same look-and-feel browse

Sub-functions of *search*

Quick Search	Advanced Search
Enter a query and click search Enter keywords or phrases for selected field Limit results to Search subscribed titels Clear	Enter a query and click search Enter keywords or phrases for selected fields Select keyword from a list Select Boolean operator (explicit) Define phrase match (explicit) Clear Search within results Limit results to (preselection) Sort by (preselection) Select display options Display X results per page Display search history

Sub-functions of *annotate*

Select object to be annotated

(need to indicate selection method)

Mark region in the object

(many different methods depending on the object)

Select type of annotation

(highlight, mark with special meaning, text, image, sound)

If text, image, sound

Specify relationship to object to be annotated

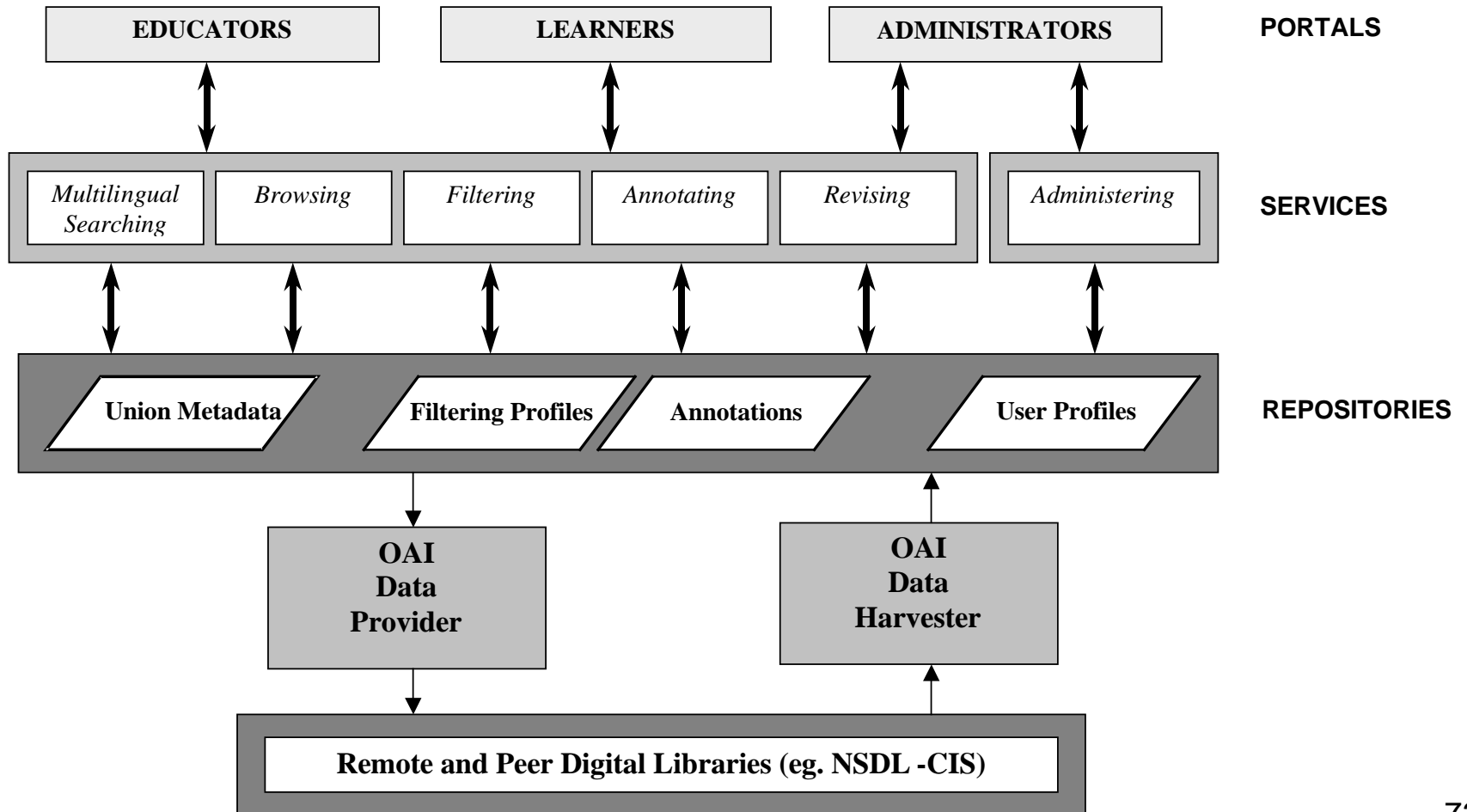
Select or create the annotating object

(possibly specifying a region)

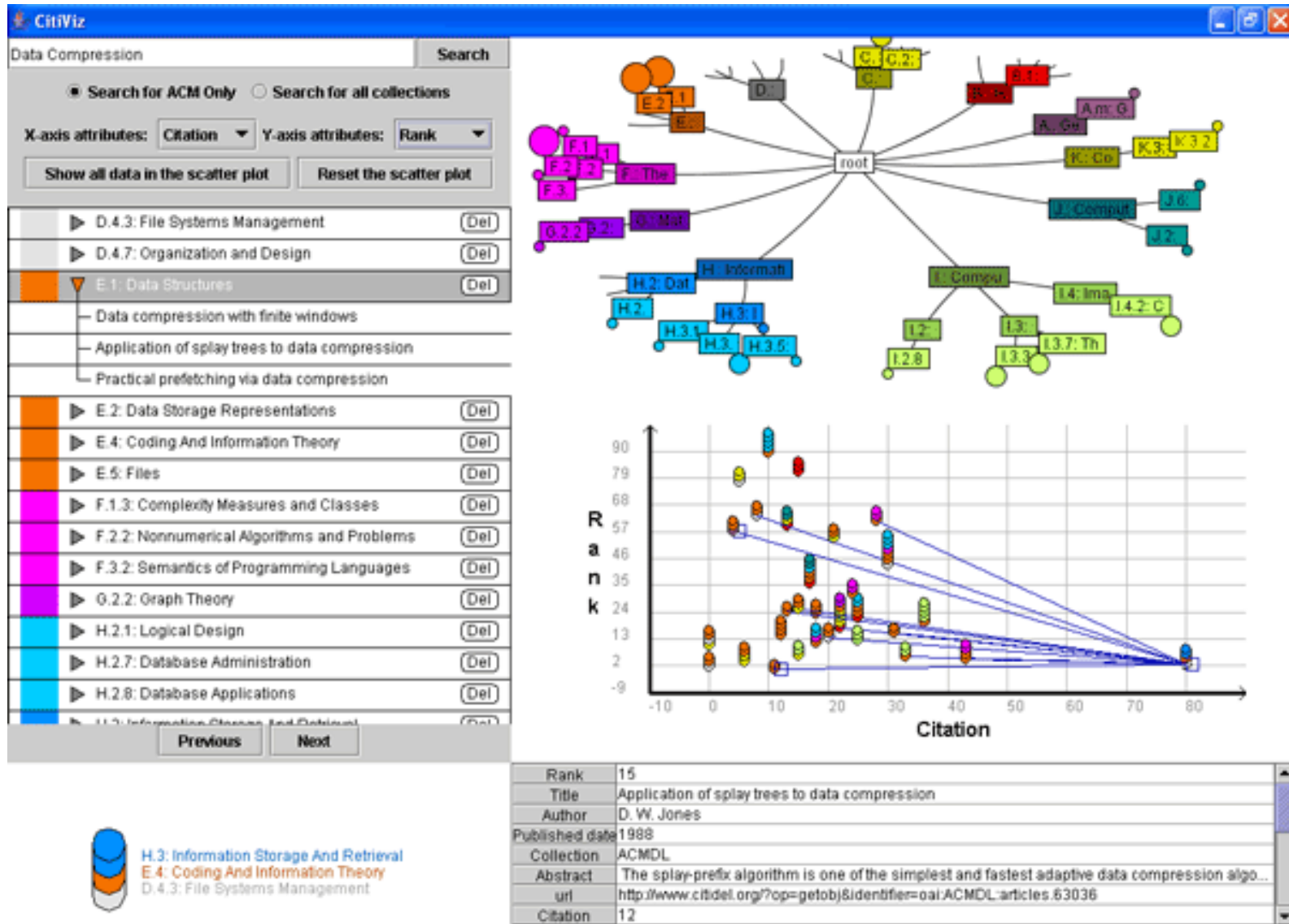
Annotating within one system

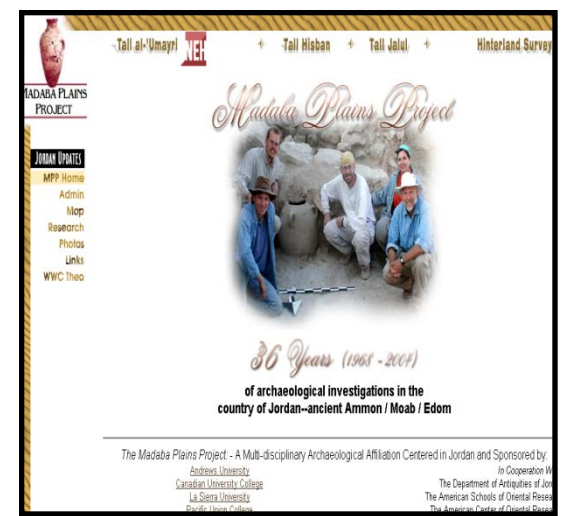
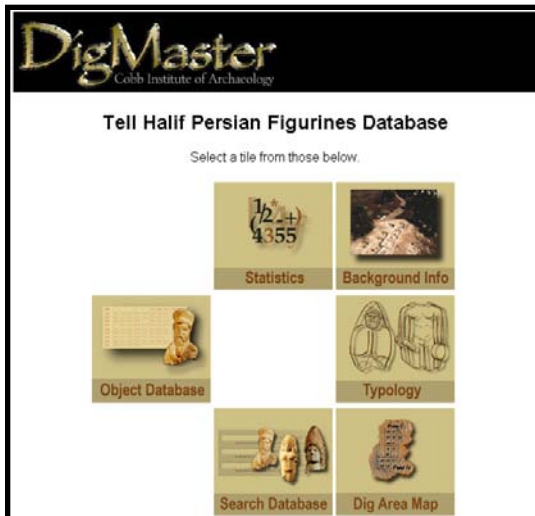
Annotating across systems

Digital library architecture for local and interoperable CITIDEL services

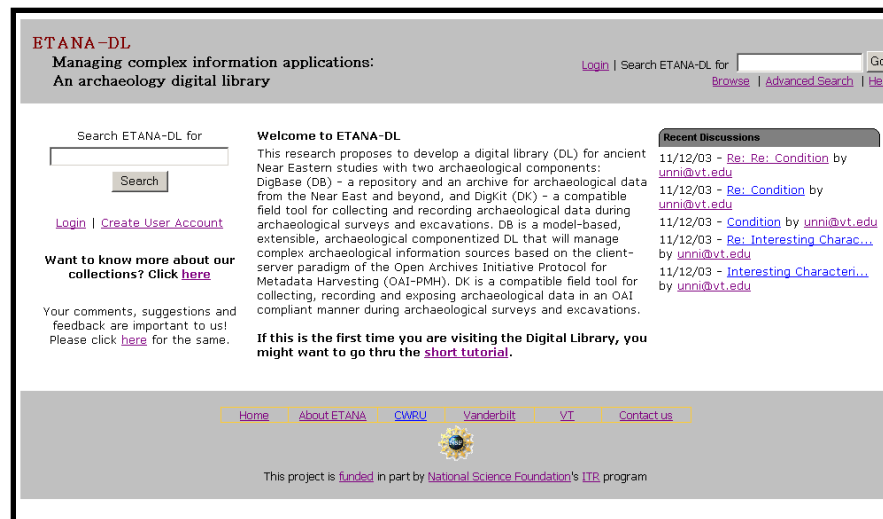


Example of Union Service: CitiViz

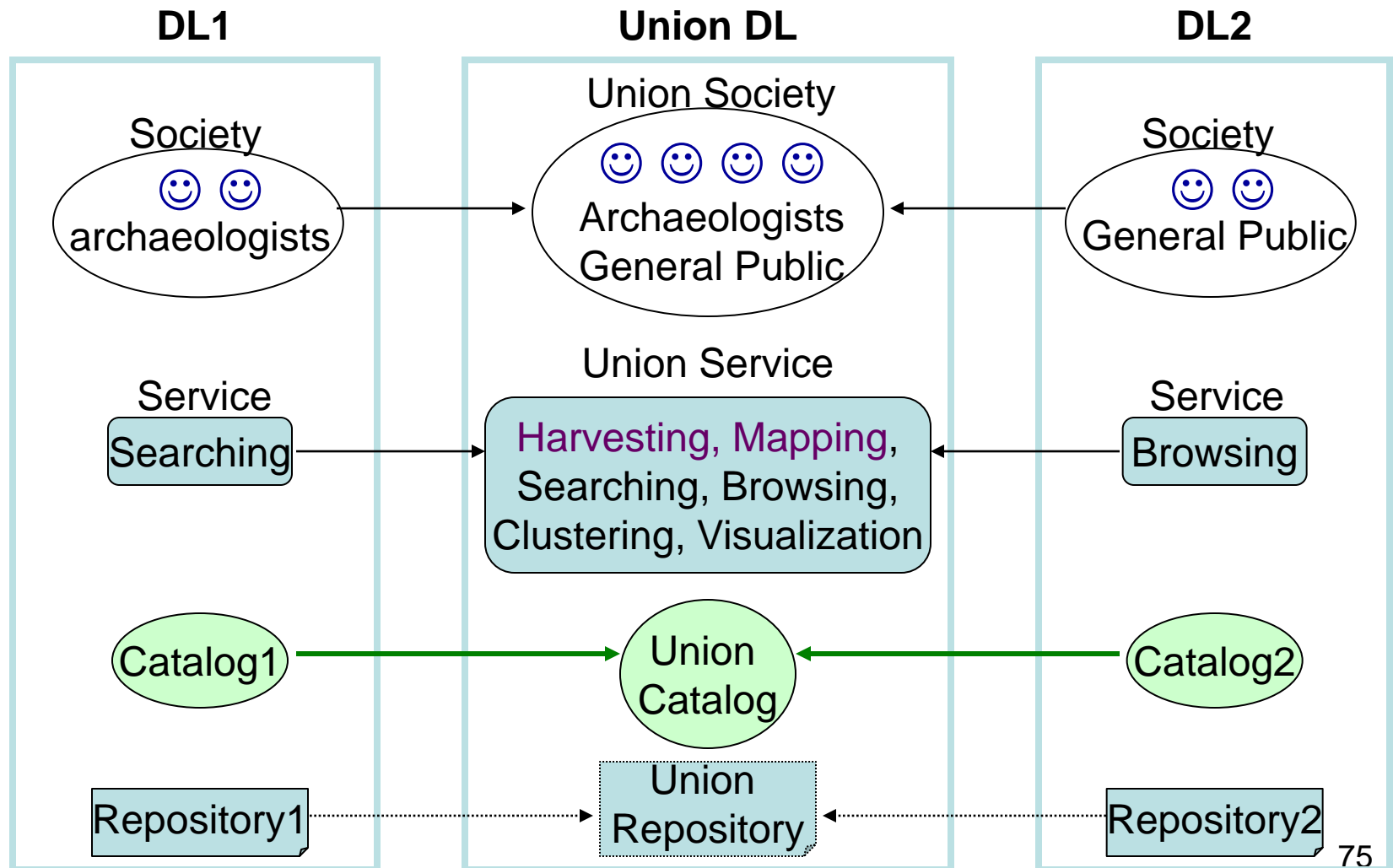




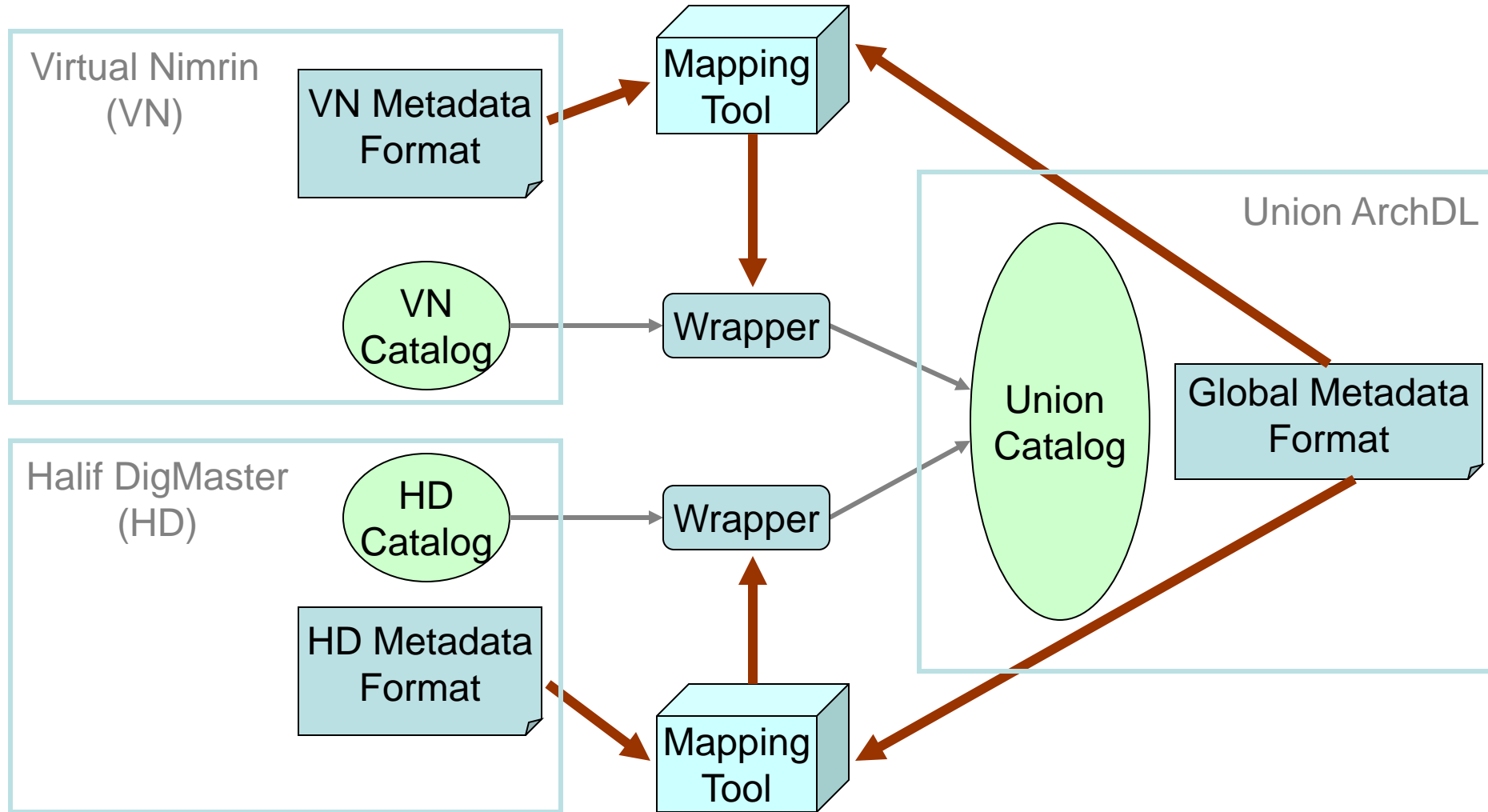
ETANA.org



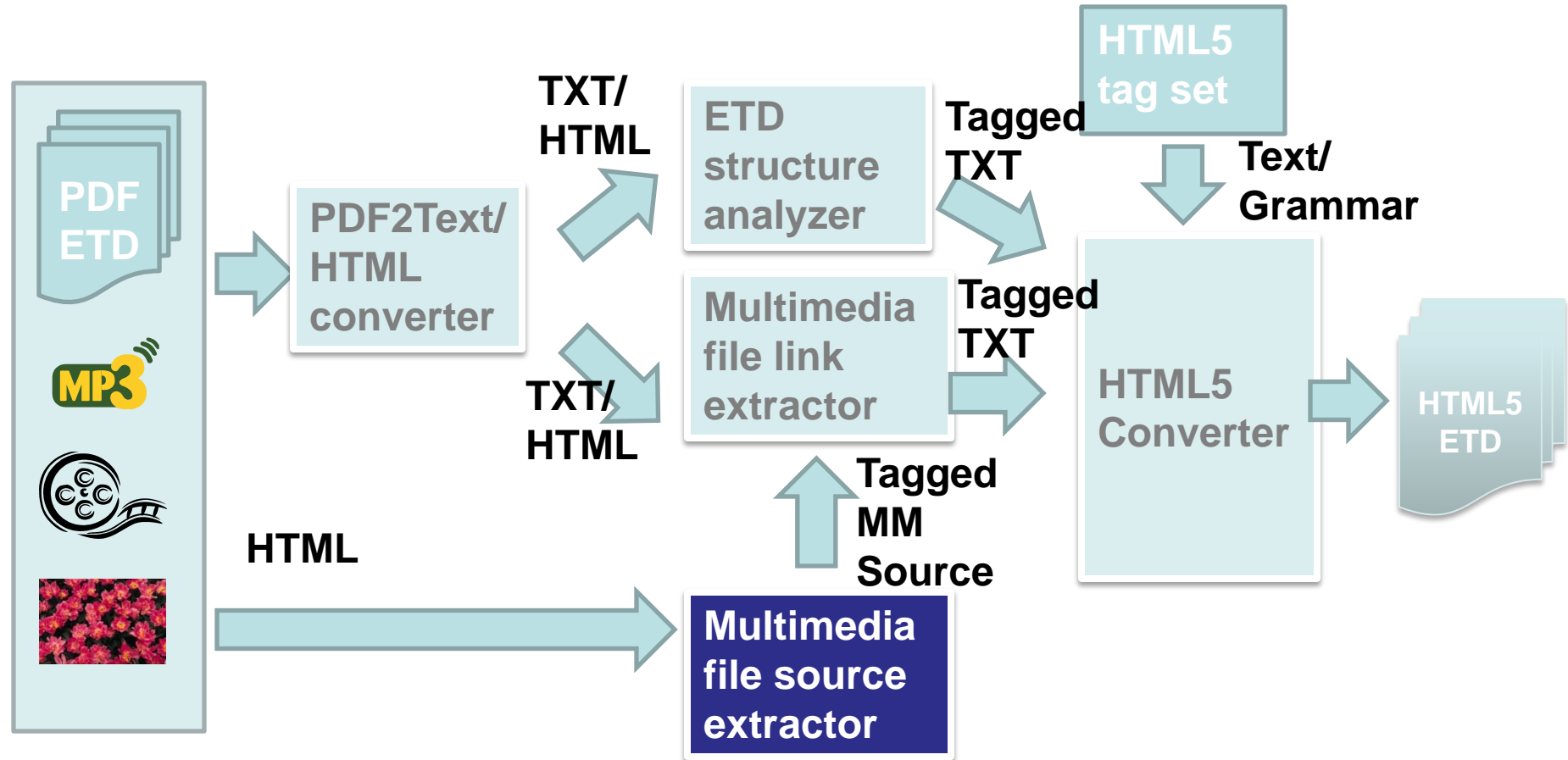
Architecture of a Union DL (ETANA.org)



Union Catalog Integration

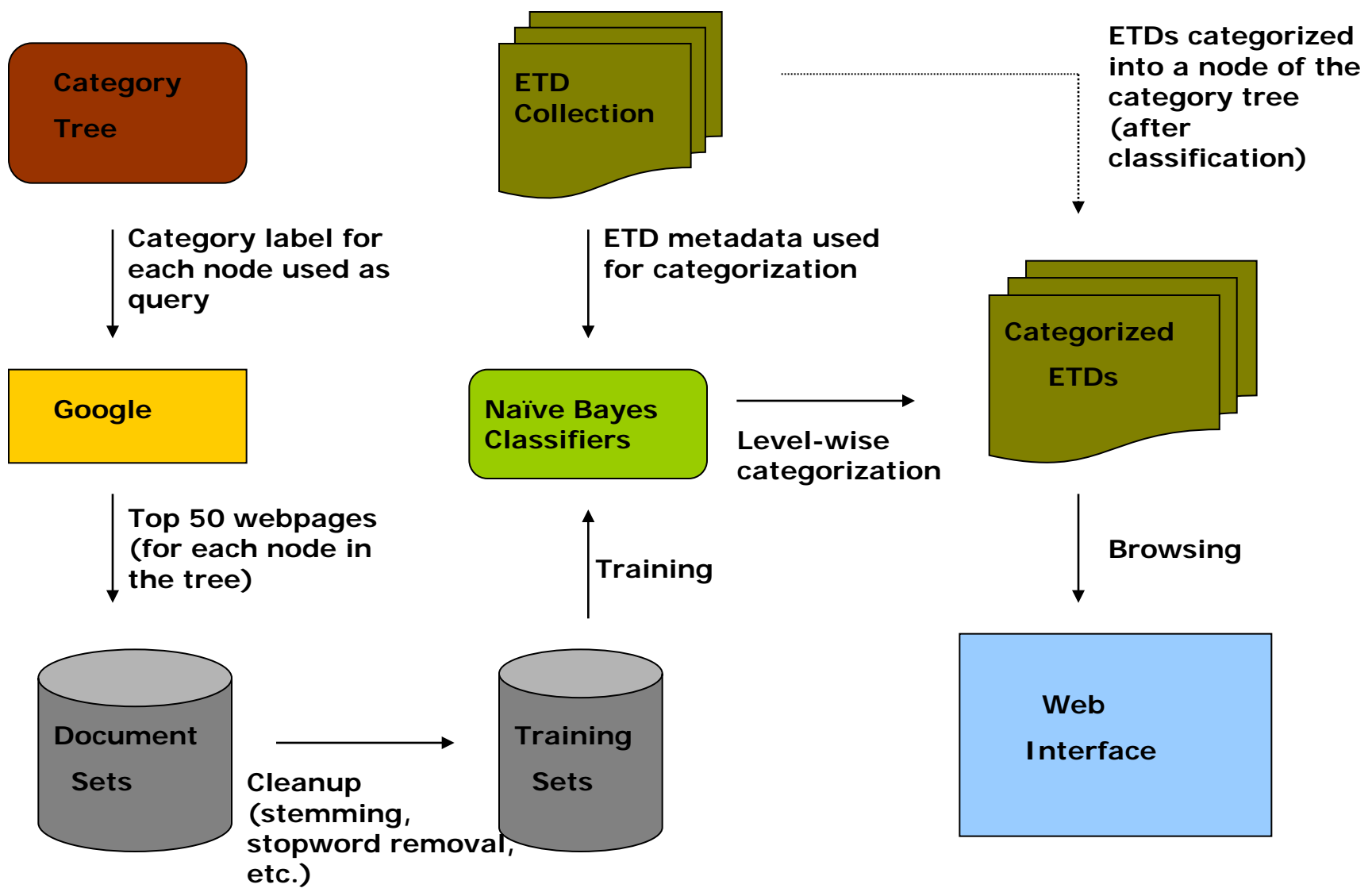


HTML5 Structuring Flowchart



ETD Classification:

Algorithm Pipeline



Digital Librarians

- Community oriented
- Collection management
- Customized services
- Principles:
 - Openness
 - Expansion
 - Interoperation, integration, communitization

Scholarship

DL curriculum
Ensemble
Living In the KnowlEdge
Society (LIKES): Book,
Part on Humanities

Humans/ Humanities

Needs
Areas
Approach

**Supporting
Scholarship
About Humans
Through
Content &
Services**

Content

Digitized
ETDs
Born digital
(digital
natives)

DLs:
CTRnet
Ensemble
ETANA
NDLTD

Services

Flattening
(cooperation,
collaboration)
Exploring
(search, browse)
Interoperability 80
Integration

Summary

- Humanism, Humanities
 - Selected DL Projects
 - Basic DL Concepts, Scholarship
 - ND LTD
 - Services Centered
-
- Questions and Comments?
 - <http://fox.cs.vt.edu/talks/2010/>
 - Remember: (ACM) JOCCH !!