Linking Research and Education in Digital Libraries

TPDL 2011 Workshop

28-29 September, Berlin

"Living In the KnowlEdge Society: the double duty of a librarian"

by Edward A. Fox

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

Outline

- Acknowledgments
- Digital libraries, 5S
- LIKES, CTRnet
- Curricular efforts
- Book efforts

Acknowledgements

- Mentors (Licklider, Kessler, Salton)
- Virginia Tech, CS, Digital Library Research Lab
- NSF and other sponsors
- Students, colleagues, co-investigators
- Monika Akbar, Yinlin Chen, Marcos André Gonçalves, Doug Gorton, Tarek Kanan, Nadia Kozievitch, Spencer Lee, Jonathan Leidig, Yi Ma, Uma Murthy, Sung Hee Park, Rao Shen, Venkat Srinivasan, Ricardo Torres, Xiaoyan Yu, ...
- Barbara Wildemuth, Jeffrey Pomerantz, Sanghee Oh, Seungwon Yang

Acknowledgements – 2

- VT Digital Library Research Laboratory
  - Digital Libraries
  - Information Retrieval
  - Human-Computer Interaction
  - Multimedia / Hypermedia / ePublishing
  - Computing-related education

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

Selected DL Projects - 2

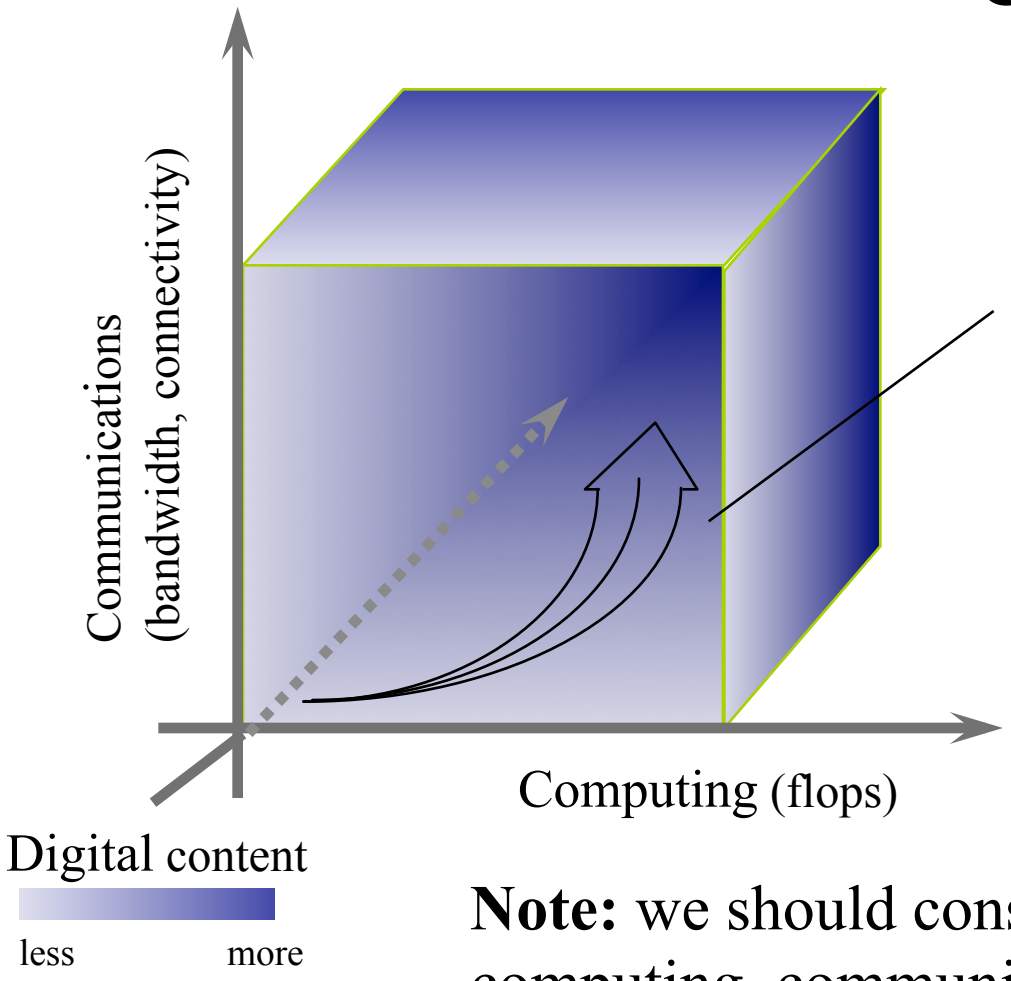
- CINET: Network Science Middleware
  - NSF SDCI 1032677
  - Simulation, Cyberinfrastructure
  - Metadata-based Generation
- Establishing a Qatari Arabic-English DL Institute: NPRP 4 - 029 - 1 – 007 pending
- Fingerprint Analysis/Distortion/Training DLs
  - National Inst of Justice, BAE Systems
- ETD Analysis, Extraction, Classification

2 duties of digital librarians

- Raise ability of patrons
  - Literacy -> facility -> computational thinking
  - University graduates -> general public
- Make content accessible now & forever
  - Discover, explore, search, browse (libraries)
  - Preserve (archives and museums)

- What to do as a digital librarian?
- How to prepare to be a digital librarian?

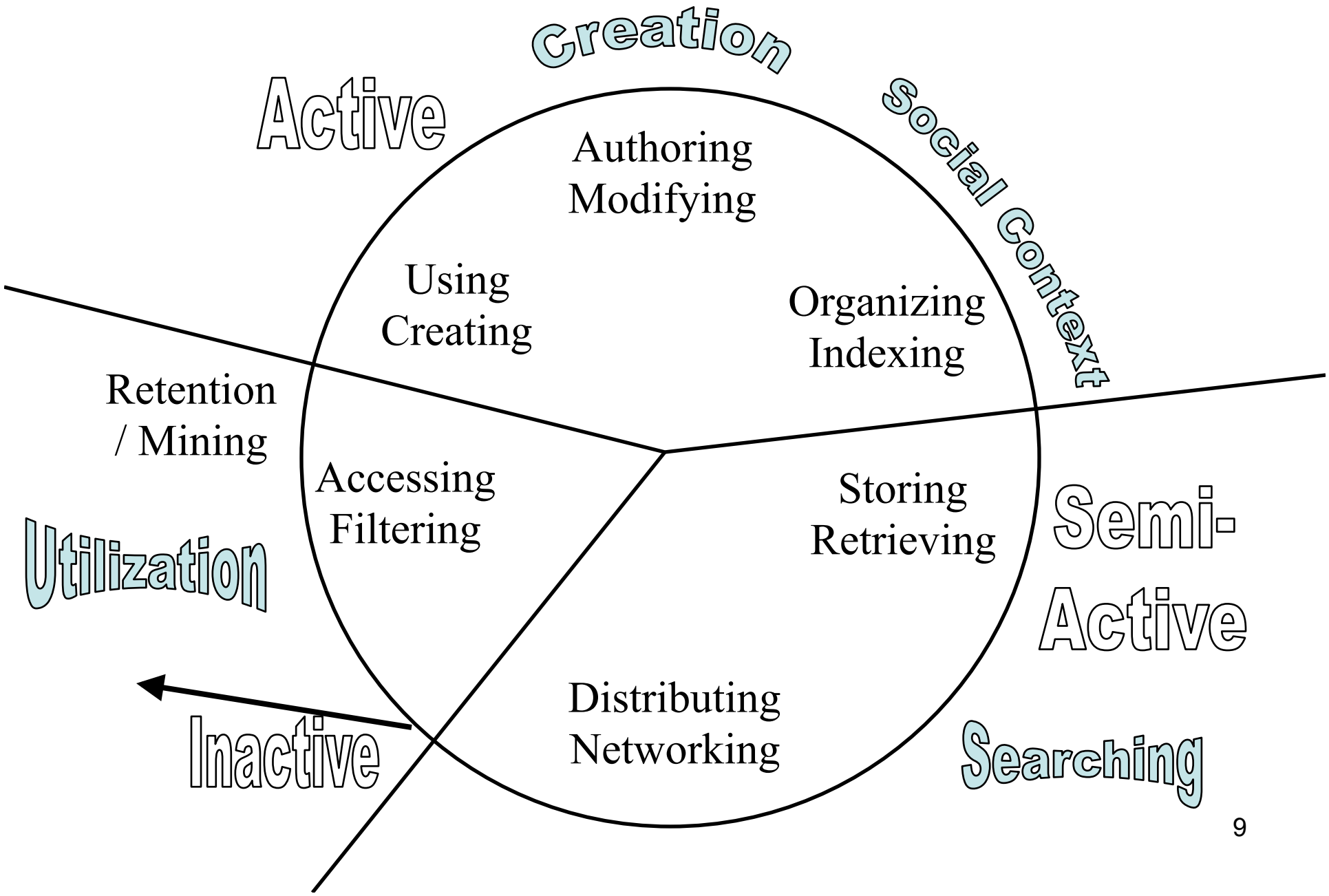
Locating Digital Libraries in Computing and Communications Technology Space



Digital Libraries technology trajectory: ***intellectual access to globally distributed information***

Note: we should consider 4 dimensions: computing, communications, content, and community (people)

Information Life Cycle

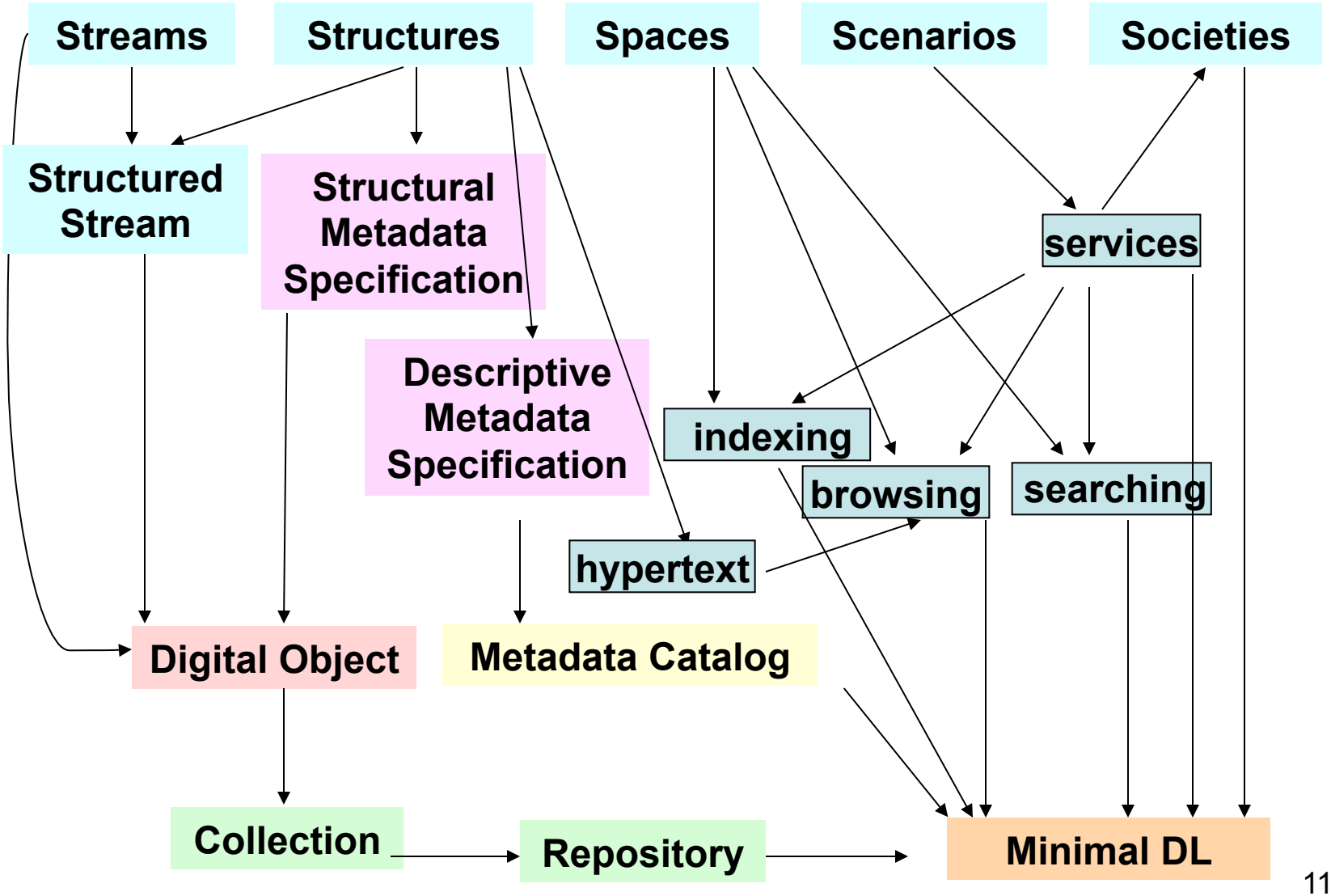


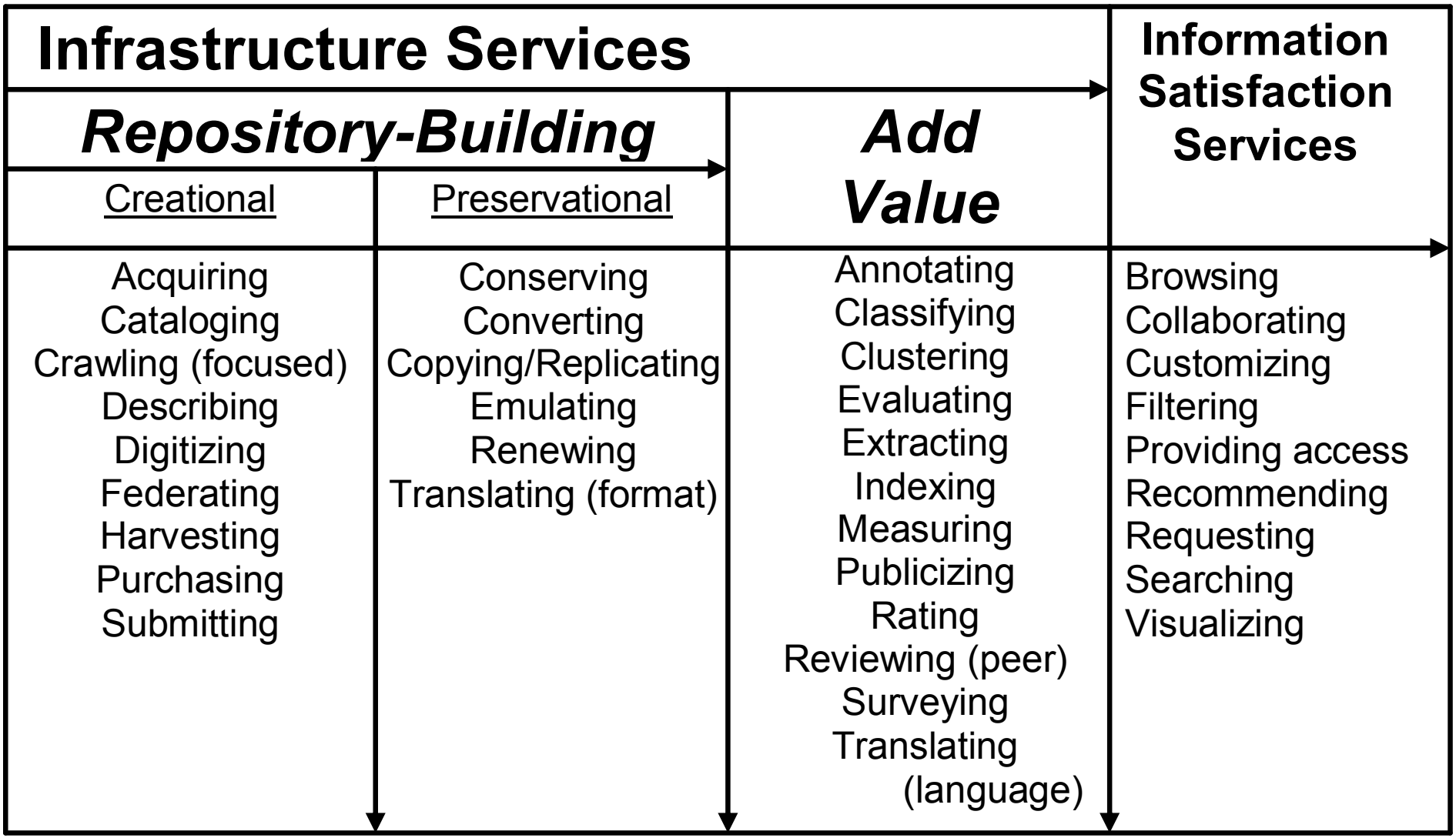
Informal 5S & DL Definitions

DLs are complex systems that

- help satisfy info needs of users (**societies**)
- provide info services (**scenarios**)
- organize info in usable ways (**structures**)
- present info in usable ways (**spaces**)
- communicate info with users (**streams**)

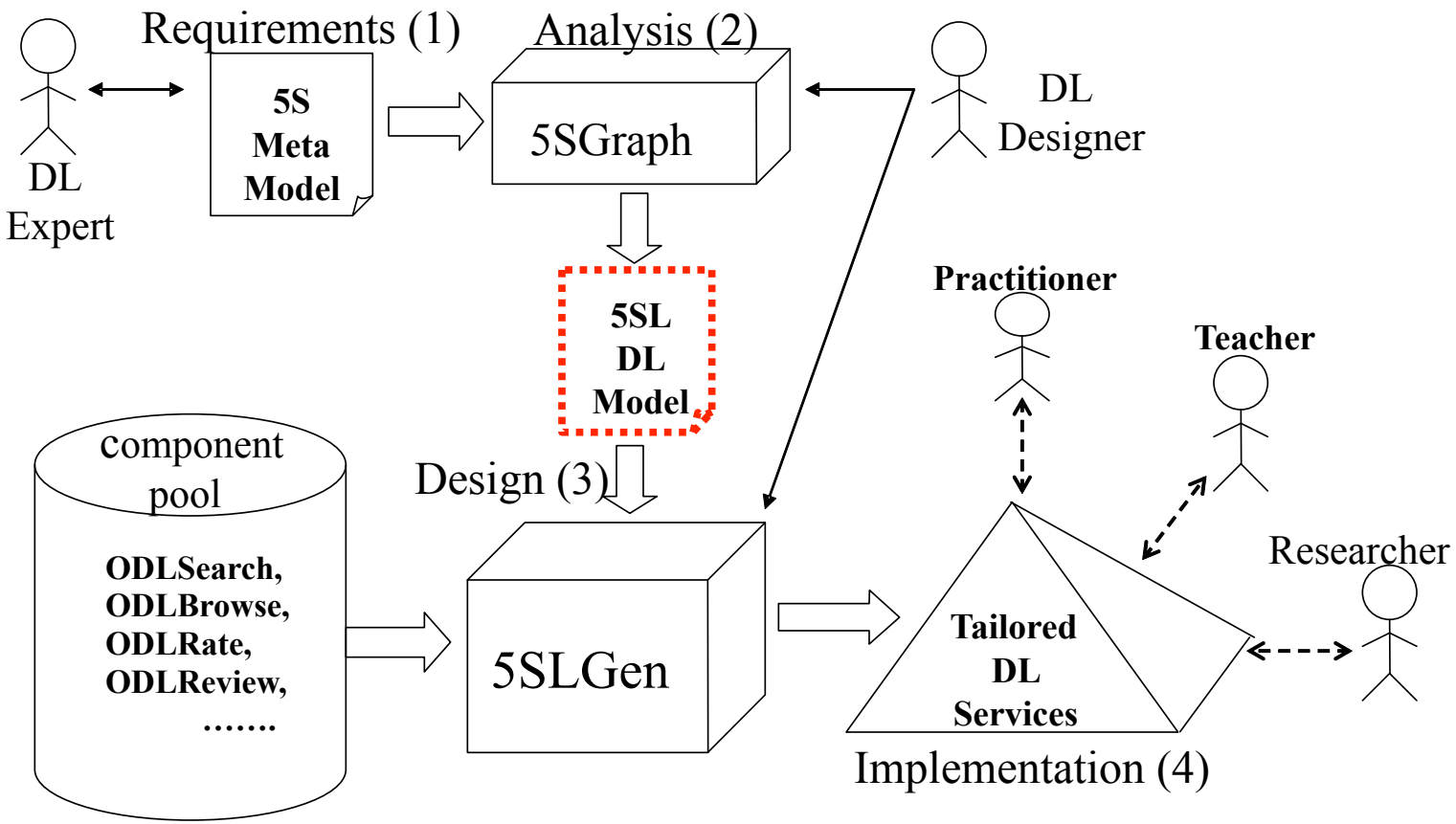
A Minimal DL in the 5S Framework



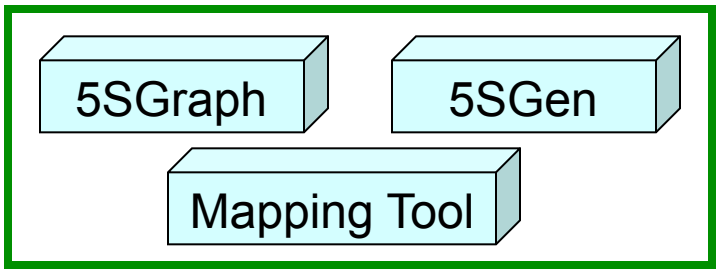

Infrastructure Services
Repository-Building
Creational
Preservational
Add Value
Information Satisfaction Services
Acquiring
Cataloging
Crawling (focused)
Describing
Digitizing
Federating
Harvesting
Purchasing
Submitting
Conserving
Converting
Copying/Replicating
Emulating
Renewing
Translating (format)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

Quality Dimensions

DL Concept	Dimensions of Quality
Digital object	Accessibility Pertinence Preservability Relevance Similarity Significance Timeliness
Metadata specification	Accuracy Completeness Conformance
Collection	Completeness Impact Factor
Catalog	Completeness Consistency
Repository	Completeness Consistency
Services	Composability Efficiency Effectiveness Extensibility Reusability Reliability



5SSuite for DL R&D



Digital Libraries --- Objectives

- World Lit.: 24hr / 7day / from desktop
- Integrated “super” information systems: 5S: Table of related areas and their coverage
- Ubiquitous, Higher Quality, Lower Cost
- Education, Knowledge Sharing, Discovery
- Disintermediation -> Collaboration
- Universities Reclaim Property
- Interactive Courseware, Student Works
- Scalable, Sustainable, Usable, Useful

DL Overview
Why of Global Interest?

- **National projects** can preserve antiquities and heritage: cultural, historical, linguistic, scholarly
- Knowledge and information are essential to economic and technological **growth, education**
- DL - a **domain for international collaboration**
  - wherein all can **contribute** and **benefit**
  - which leverages investment in **networking**
  - which provides useful **content** on Internet & WWW
  - which will **tie nations and peoples together** more strongly and through **deeper understanding**

Outline

- Acknowledgments
- Digital libraries, 5S
- LIKES, CTRnet
- Curricular efforts
- Book efforts

Living In the KnowlEdge Society (LIKES)

North Carolina A & T
Santa Clara University
Villanova University
Virginia Tech

NSF CPATH: CCF-0722259,
0722276,
0722289,
and 0752865

LIKES Workshops

1. SCU, 2007: Defining Problems & Applications of Knowledge Society
2. NC A&T, 2008: Biology, Geography, Music, Physics, Statistics, University Studies
3. VT, 2008: defining key terms related to knowledge society, identifying key computing concepts, mapping disciplinary needs with computing concepts
4. Villanova, 2009: interdisciplinary connections, modules/ tools/pedagogy/assessment
5. VT, 2009: Curricular Guidelines Connecting Computing with Other Disciplines
6. Durham, 2010: biology, chemistry, physics, computational science, business/social sciences

LIKES Vision

Build a community leading the way to change how computing concepts are taught in both computing-related disciplines and the disciplines of the broader workforce and society.

Overall LIKES Objectives

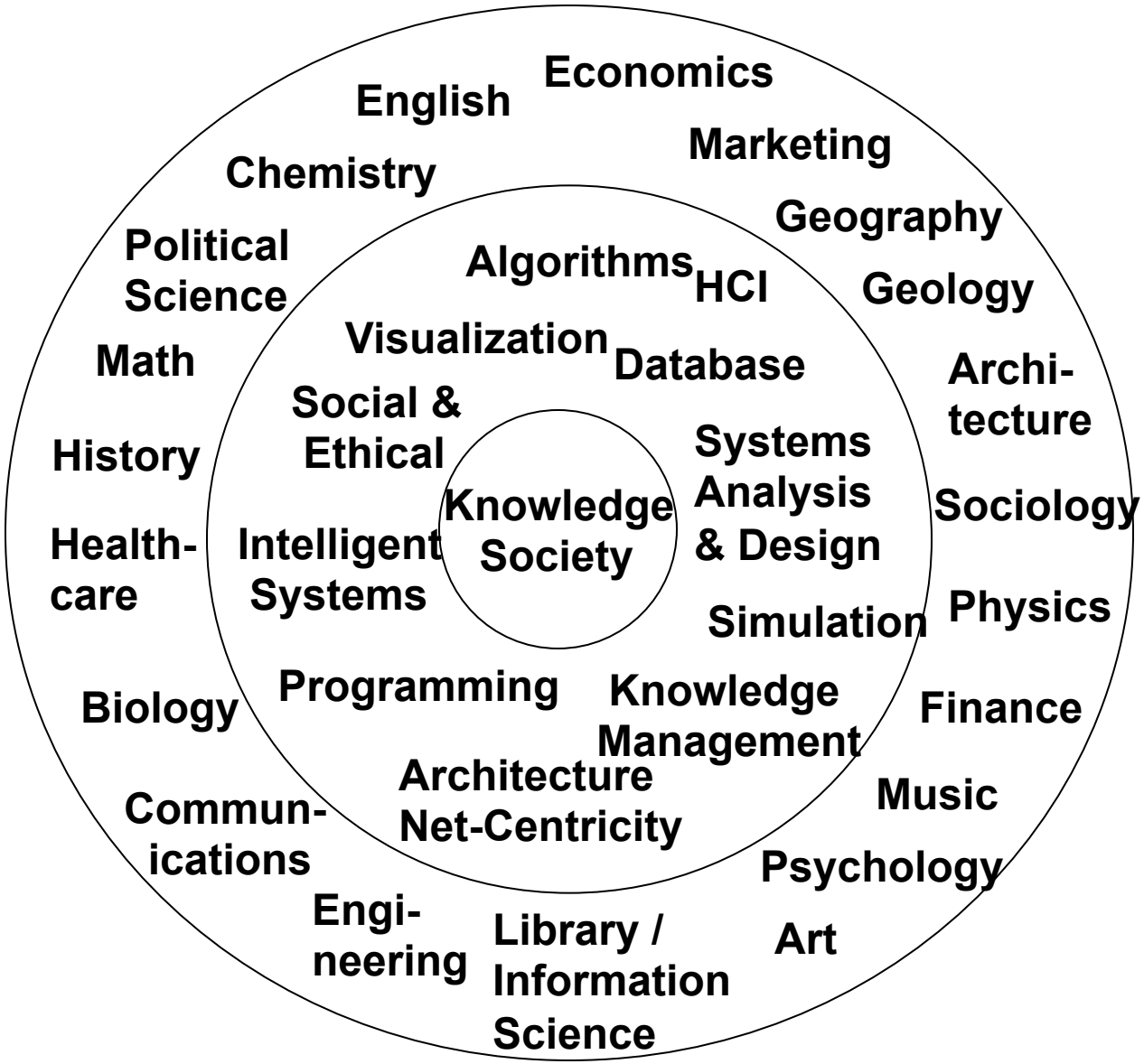
- Students should have the ability to apply
  - Computing concepts
  - Methods
  - Computational thinking skills

to the needs of the emerging
knowledge society,
in modern times and
in the future

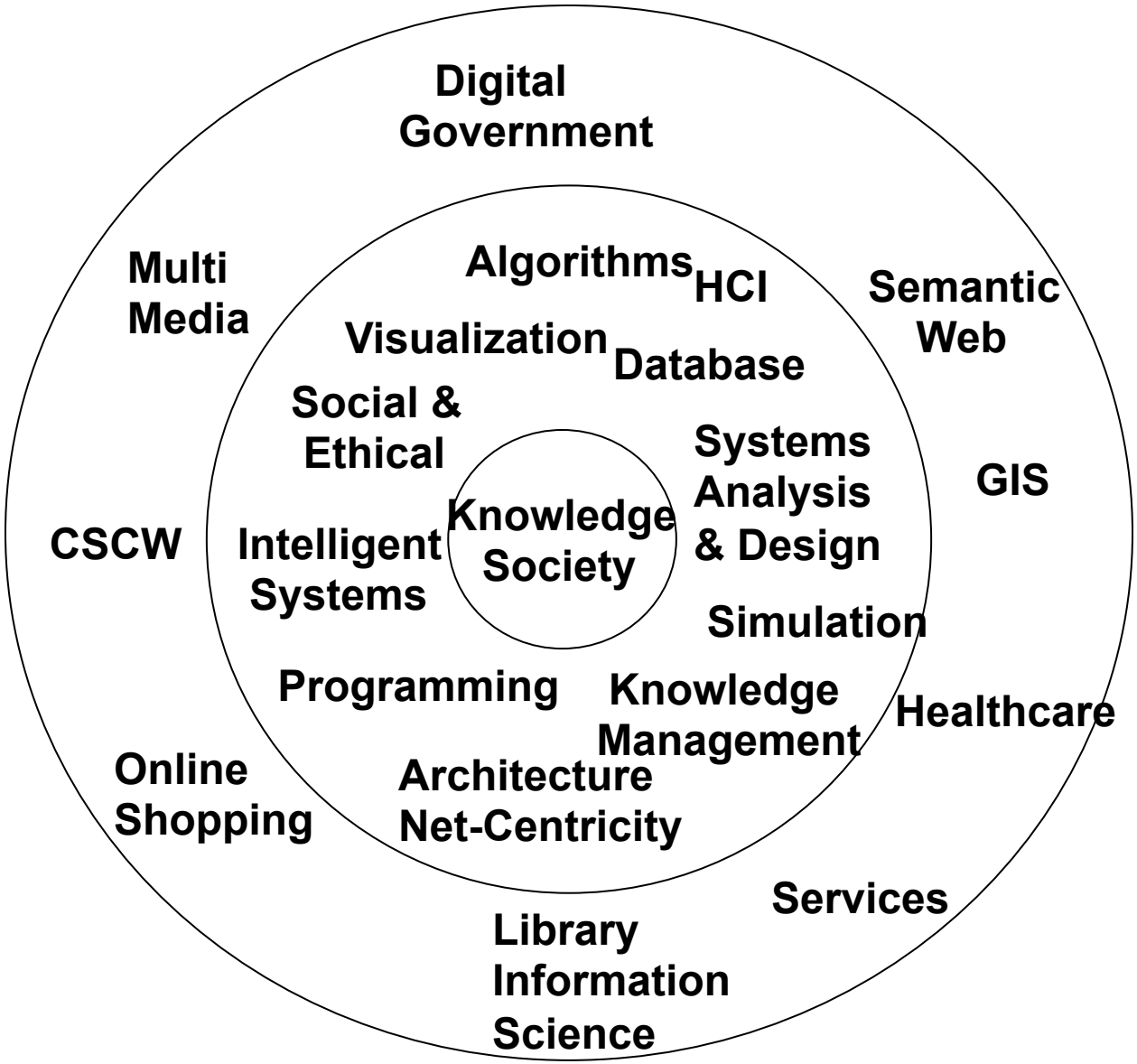
Computing Concepts of Broad Interest (to other disciplines)

1. data, information, and knowledge
2. algorithms, analysis, problem solving, programming, work flows, and software engineering
3. interaction, interfaces, graphics, games, visualization, and virtual environments
4. modeling and simulation

Disciplines

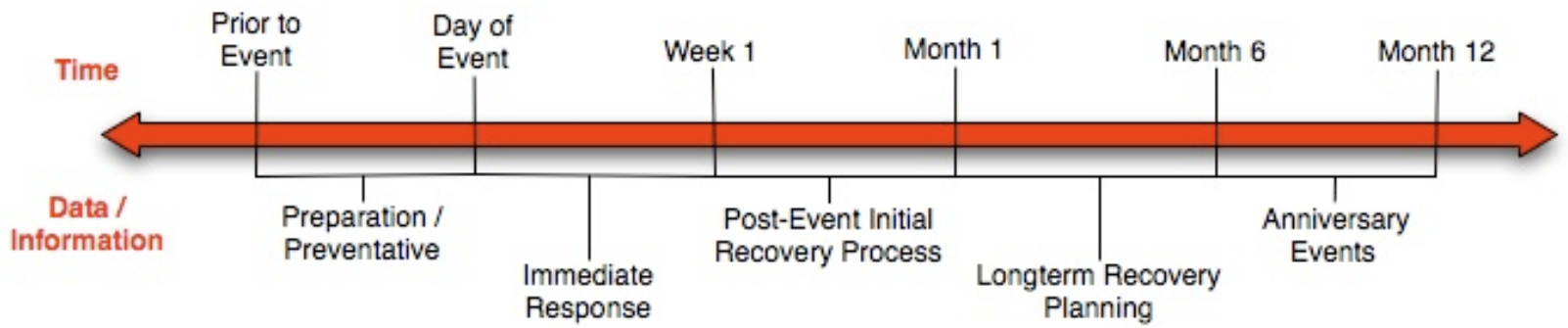


Applications

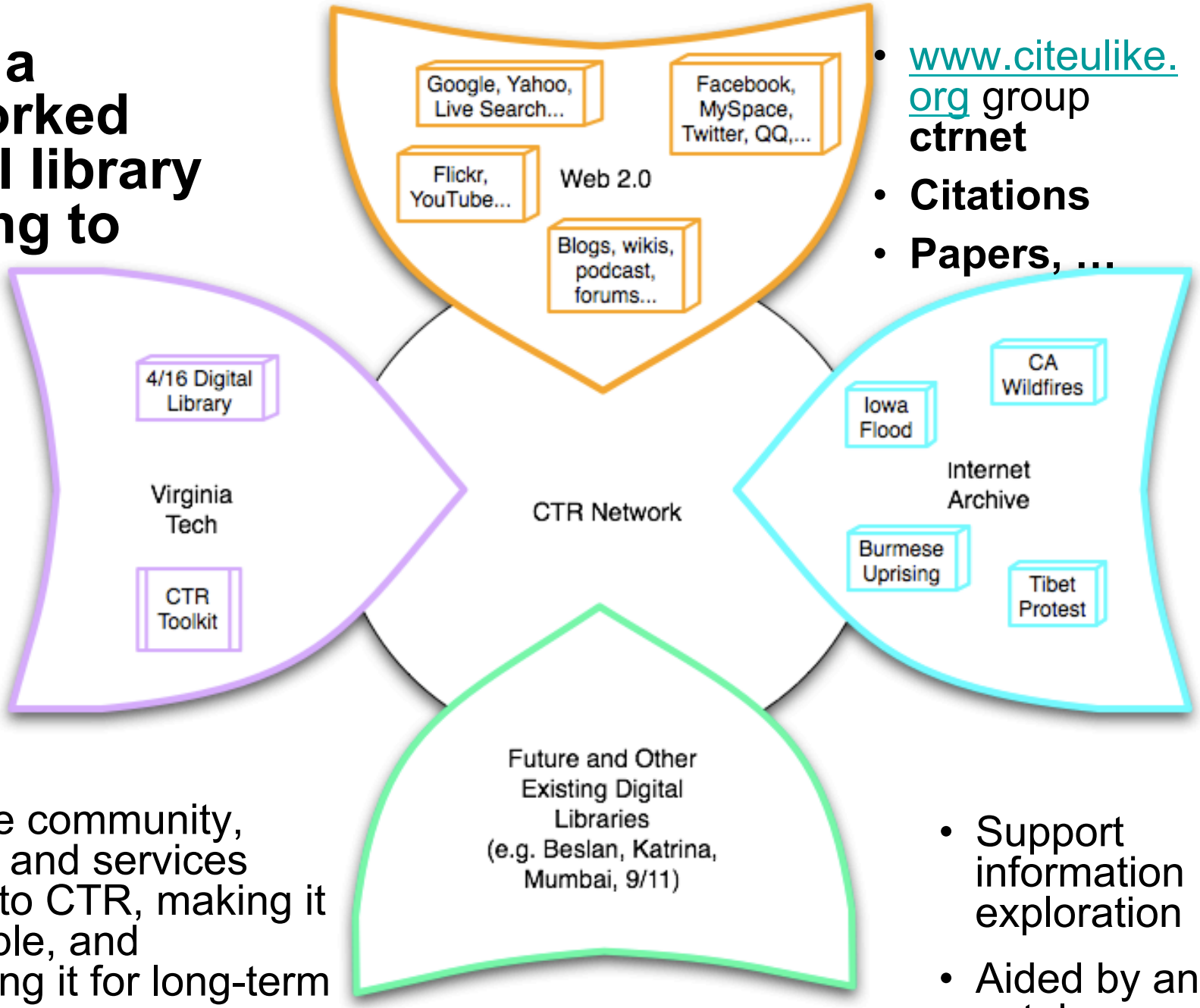


Crisis, Tragedy, and Recovery

- 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**



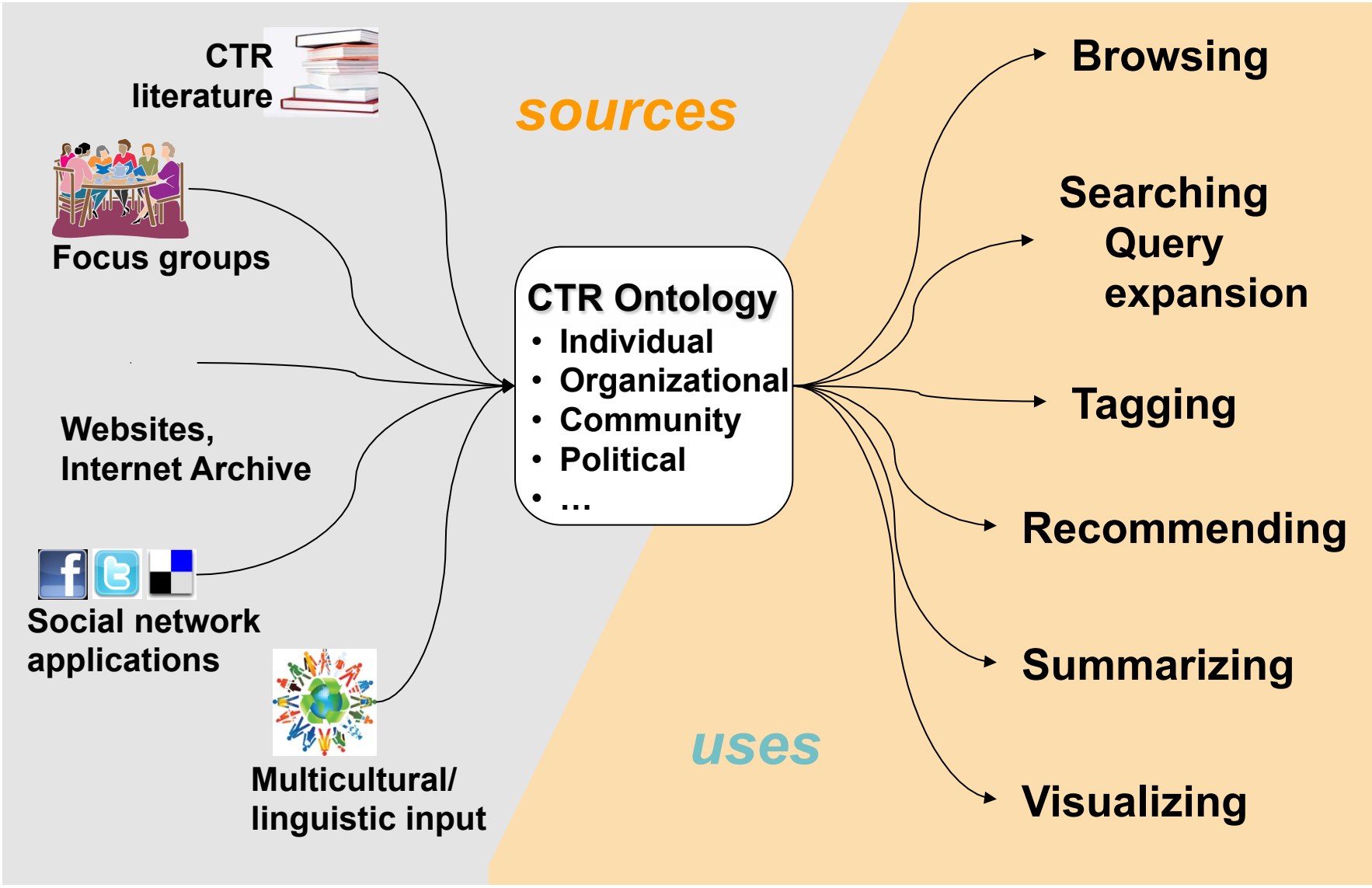
- www.citeulike.org group **ctrnet**
- **Citations**
- **Papers, …**

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

www.ctrnet.net

- Support information exploration
- Aided by an ontology--

Goals for Ontology for CTR



Generalizing CTRnet

- Precursor: www.dl-vt-416.org
- Sequel: Event Archiving
  - Related to work of national libraries to archive internet of interest (as in plans for Qatar Digital Library Institute)
  - Related to International Internet Preservation Consortium (IIPC, www.netpreserve.org)
  - Related to Archive-it (www.archive-it.org) and its Spontaneous Events collections (see http://www.archive-it.org/public/topic.html?topic=spontaneousEvents)

Outline

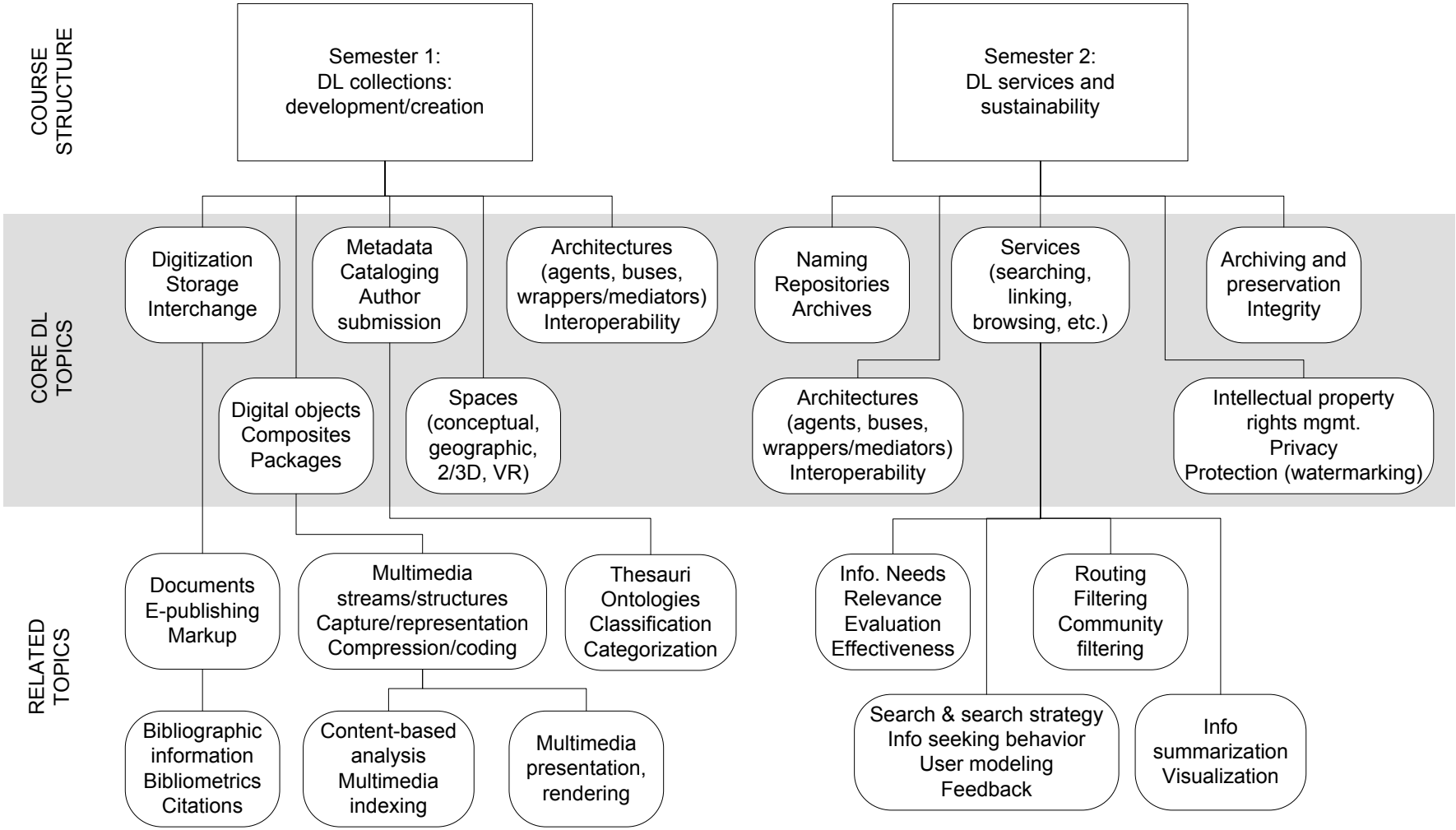
- Acknowledgments
- Digital libraries, 5S
- LIKES, CTRnet
- Curricular efforts
- Book efforts

CS2008 Information Management Areas beside DL#	
Information models*	Transaction processing
Database systems*	Distributed DBs
Data modeling*	Physical DB design
Indexing	Data mining
Relational DBs	Information storage and retrieval#
Query languages	Hypermedia
Relational DB design	Multimedia systems

* Core components

#DL moved to ISR for 2013

DL Curriculum Framework



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-a (10-c): Conceptual frameworks, models, theories, defns**
- **Module 1-b: History of digital libraries and library automation**
- **Module 2-c (8-d): File Formats, Transformation, and Migration**
- **Module 3-b: Digitization**
- **Module 3-e (7-e): Web publishing**
- **Module 4-b: Metadata**
- **Module 5-a: Architecture overviews**

DL Curric. Project - 3

- **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**
- **Module 7-a: Indexing and searching**
- **Module 7-a(1): Image retrieval**

DL Curric. Project - 4

- **Module 7-b: Reference Services**
- **Module 7-c: Recommender systems**
- **Module 7-d: Routing**
- **Module 7-f: Crawling**
- **Module 7-g: Personalization**
- **Module 8-a: Preservation**
- **Module 8-b: Web archiving**
- **Module 9-c: Digital library evaluation, user studies**

DL Curric. Project – 5 Modules and Cloud Instances

IR	Apache Solr	Lemur
IR	WordNet	R
IR	NLTK	SEDNA XML DB
IR	CLUTO	Weka
IR	TREC Eval	Hadoop Map-Reduce
MM	Media Computation	Audacity
MM	PureData	Fingerprint

Module Development – What?

- Digital Libraries
- Information Retrieval tools (cloud)
- Multimedia tools (cloud)

- Biometrics Training
  - Especially fingerprint analysis

Module Development – Who?

- Experts
  - DL
  - Biometrics
- Teams in a 6000-level DL Course: 4
- Teams in a 5000-level IR Course: 5 (+5)
- Teams in a 4000 MM Course: 4

Outline

- Acknowledgments
- Digital libraries, 5S
- LIKES, CTRnet
- Curricular efforts
- Book efforts

Old Planned Book: Parts

- Ch. 1. Introduction (Motivation, Synopsis)
- Part 1 – The “Ss”
- Part 2 – Higher DL Constructs
- Part 3 – Advanced Topics
- Appendix

Old Planned Book: Part 2

- Part 2 – Higher DL Constructs
  - Ch. 7: Collections
  - Ch. 8: Catalogs
  - Ch. 9: Repositories and Archives
  - Ch. 10: Services
  - Ch. 11: Systems
  - Ch. 12: Case Studies

Old Planned Book: Part 3 …

- Part 3 – Advanced Topics
  - Ch. 13: Quality
  - Ch. 14: Integration
  - Ch. 15: How to build a digital library
  - Ch. 16: Research Challenges, Future Perspectives
- Appendix
  - A: Mathematical preliminaries
  - B: Formal Definitions: Ss
  - C: Formal Definitions: DL terms, Minimal DL
  - D: Formal Definitions: Archeological DL
  - E: Glossary of terms, mappings

Book(s) for 2012

- Morgan-Claypool lecture manuscript for Synthesis digital library series – planned

- Book for CS6604, Digital Libraries, Fall 2011
  - See highlights in next slides
  - Being considered by publishers for 2012 release
  - Accompanied by slides and other supplemental materials
  - Available for test use for classes in January

Book Draft Chapter Authors

- Monika Akbar, Pranav Angara, Yinlin Chen, Lois M. Delcambre, Noha Elsherbiny, Eric Fouh, Marcos André Gonçalves, Nádia P. Kozievitch, Spencer Lee, Jonathan Leidig, Lin Tzy Li, Mohamed Magdy Gharib Farag, Uma Murthy, Sung Hee Park, Rao Shen, Venkat Srinivasan, Ricardo da Silva Torres, and Seungwon Yang

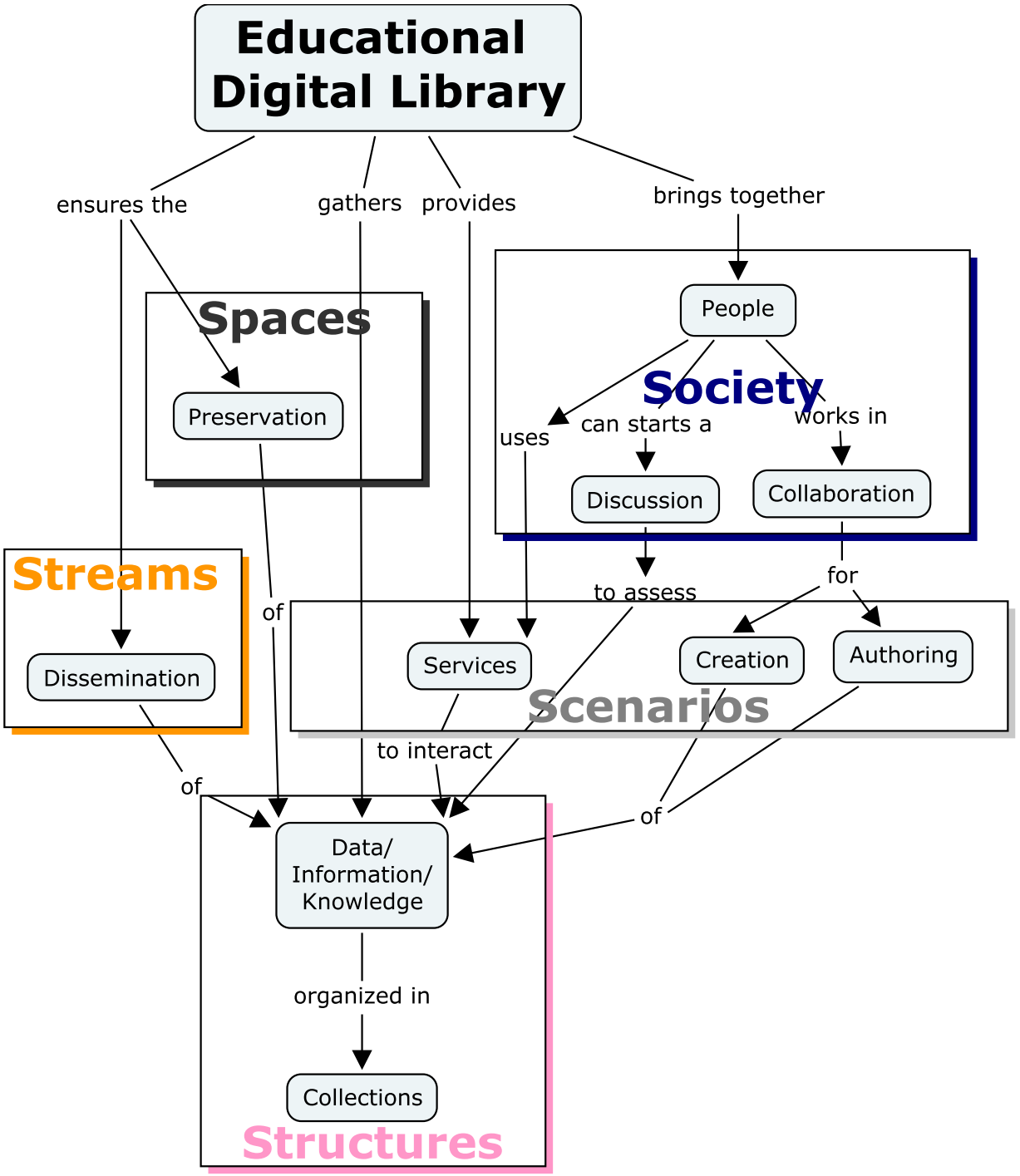
Book Draft - 1

- **Frontmatter**
- **1 Basic Concepts**
  - 1 Introduction
  - 2 Exploration
  - 3 Evaluation
- **2 Advanced Concepts**
  - 4 Complex objects
  - 5 Integration
  - 6 Subdocuments
  - 7 Ontologies
  - 8 Classification

Book Draft - 2

- **3 Applications**
  - 9 Content-based Image Retrieval (CBIR)
  - 10 Online Communities and Social Networks
  - 11 Education
  - 12 Bioinformatics, Scientific, and Simulation DLs
  - 13 Geospatial Information
  - 14 Security
  - 15 Text Extraction

- **Backmatter:** References, Mathematical Preliminaries, Glossary


Educational Digital Library
ensures the
gathers
provides
brings together
Spaces
Preservation
People
Society
uses
can starts a
works in
Discussion
Collaboration
Streams
Dissemination
of
to assess
for
Services
Creation
Authoring
Scenarios
to interact
of
of
Data/ Information/ Knowledge
organized in
Collections
Structures

Summary

- Acknowledgments
- Digital libraries, 5S
- LIKES, CTRnet
- Curricular efforts
- Book efforts

Questions?
Discussion?

Thank You!

(fox@vt.edu)