

Hypatia Seminar
11 November 2019, Virginia Tech

“Information - Friendly”

by Edward A. Fox

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

Dept. of Computer Science, www.cs.vt.edu

Acknowledgements

- Mentors (MIT: Licklider, Kessler; Cornell: Salton)
- Virginia Tech, CS (ECE), Digital Library Research Laboratory (DLRL)
- Department Head 2008-2015, Barbara Ryder, VP for ACM, many awards
- NSF and many other sponsors
- Students, colleagues, co-investigators (selected):
- Ghaleb Abdulla, Monika Akbar, Hamed Alhoori, Pranav Angara, Bipasha Banerjee, Abigail Bartolome, Warren Bickel, Boots Cassel, Saurabh Chakraborty, Prashant Chandrasekar, Yinlin Chen, Kiran Chitturi, Lois Delcambre, Noha ElSherbiny, Alexandre Falcao, Weiguo Fan, Eric Fouh, Chris Franck, Rick Furuta, Ayush Ganotra, Lee Giles, Marcos André Gonçalves, Doug Gorton, Islam Harb, S.M. Shamimul Hasan, Bill Ingram, Palakh Jude, Tarek Kanan, Ola Karajeh, Andrea Kavanaugh, Farnaz Khaghani, Nadia Kozevitch, Abhinav Kumar, Harinni Kumar, Spencer Lee, Sunshin Lee, Jonathan Leidig, Lin Tzy Li, Liuqing Li, Yi Ma, Yufeng Ma, Mohamed Magdy, Madhav Marathe, Uma Murthy, Pranav Nakate, Lucy Nowell, Sung Hee Park, Denilson Pereira, Sagnik Ray Choudhury, Naren Ramakrishnan, Chandan Reddy, Rao Shen, Clifford Shaffer, Steve Sheetz, Don Shoemaker, Ziqian Song, Venkat Srinivasan, Amirsina Torfi, Ricardo Torres, Xinyue Wang, Zhiwu Xie, Xiaoyan Yu, Xuan Zhang, ...
- DL Curriculum: Sanghee Oh, Jeffrey Pomerantz, Barbara Wildemuth, Seungwon Yang

Working in CS Areas with Less Gender Imbalance

1. Information retrieval / digital libraries
(closely related to library and information science, move toward iSchools)
2. Human-computer interaction (including social networks and informatics)
3. Collaboration with health, medical, biological and bio-informatics areas (+Reiki)
4. Collaboration with humanities and social/behavioral sciences (e.g., NLP / computational linguistics)

Current Funded Grants

1. IMLS LG: Opening Books and the National Corpus of Graduate Research, Bill Ingram, Jian Wu (ODU), E. A. Fox
2. IMLS RE: Continuing Education to Advance Web Archiving, Zhiwu Xie, E. A. Fox; GWU, LANL, IA, ODU, U. Waterloo
3. NIH: The Social Interactome of Recovery: Social Media as Therapy Development, Warren K. Bickel (VTCRI), E. A. Fox ...
4. NSF CRISP : Coordinated, Behaviorally-Aware Recovery for Transportation and Power Disruptions (CBAR-tpd), Clemson: Pamela Murray-Tuite; VT: E. A. Fox, Kris Wernstedt; U. Mich: Seth Guikema
5. NSF I-Corps: Automated Summarization Technology, E. A. Fox, Naren Ramakrishnan
6. NSF IIS: Global Event and Trend Archive Research (GETAR), VT: E. A. Fox, Chandan Reddy, Andrea L. Kavanaugh, Donald J. Shoemaker; Internet Archive: Jefferson Bailey
7. NSF OAC: CINES: A Scalable Cyberinfrastructure for Sustained Innovation in Network Engineering and Science, Madhav Marathe, C. Amelink, N. Ramakrishnan, E. A. Fox, ...

Design of

INFORMATION

Access Extraction Representation Retrieval Systems Technology Theory Viz

Libraries

Archives

Multimedia

Hypermedia

Text

WWW

Hypertext

Images

Search Engine

Crawling

Webpage

Links

Videos

Machine Learning

Mining

Analytics

Deep Learning

Relational

NLP

AI

Statistics

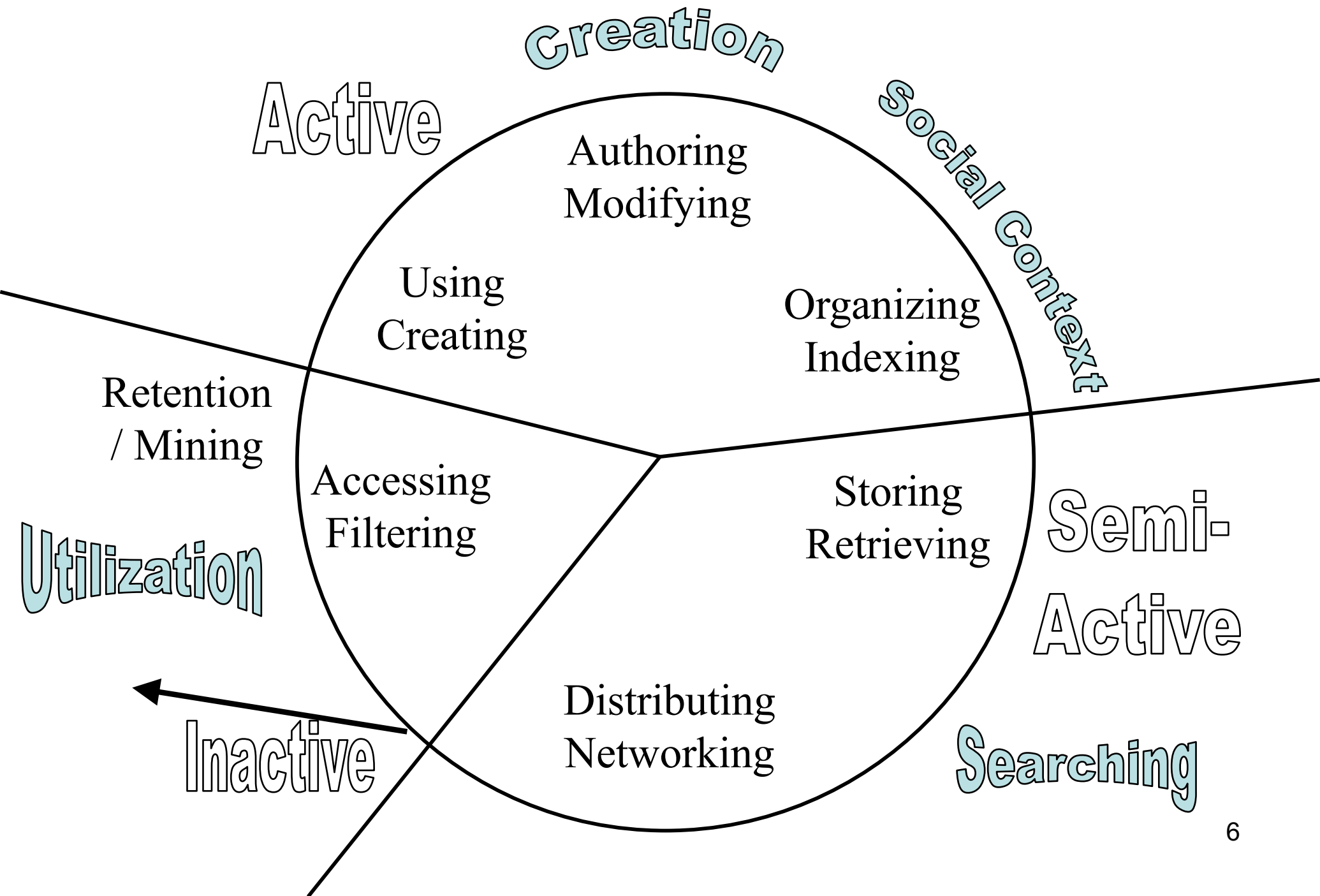
Database

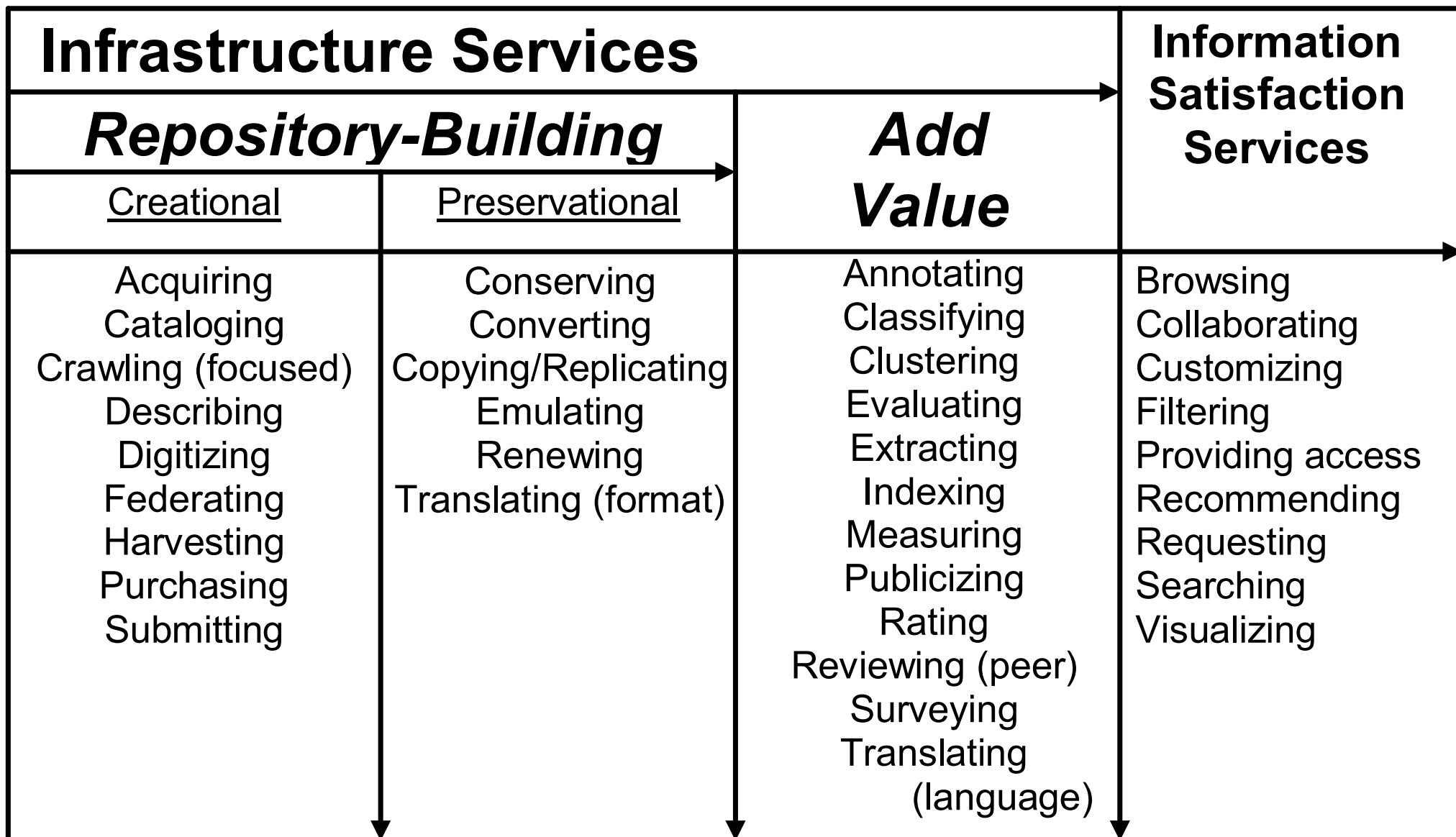
Tables

Data

Knowledge 5

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**)

People

Creation and
Preservation

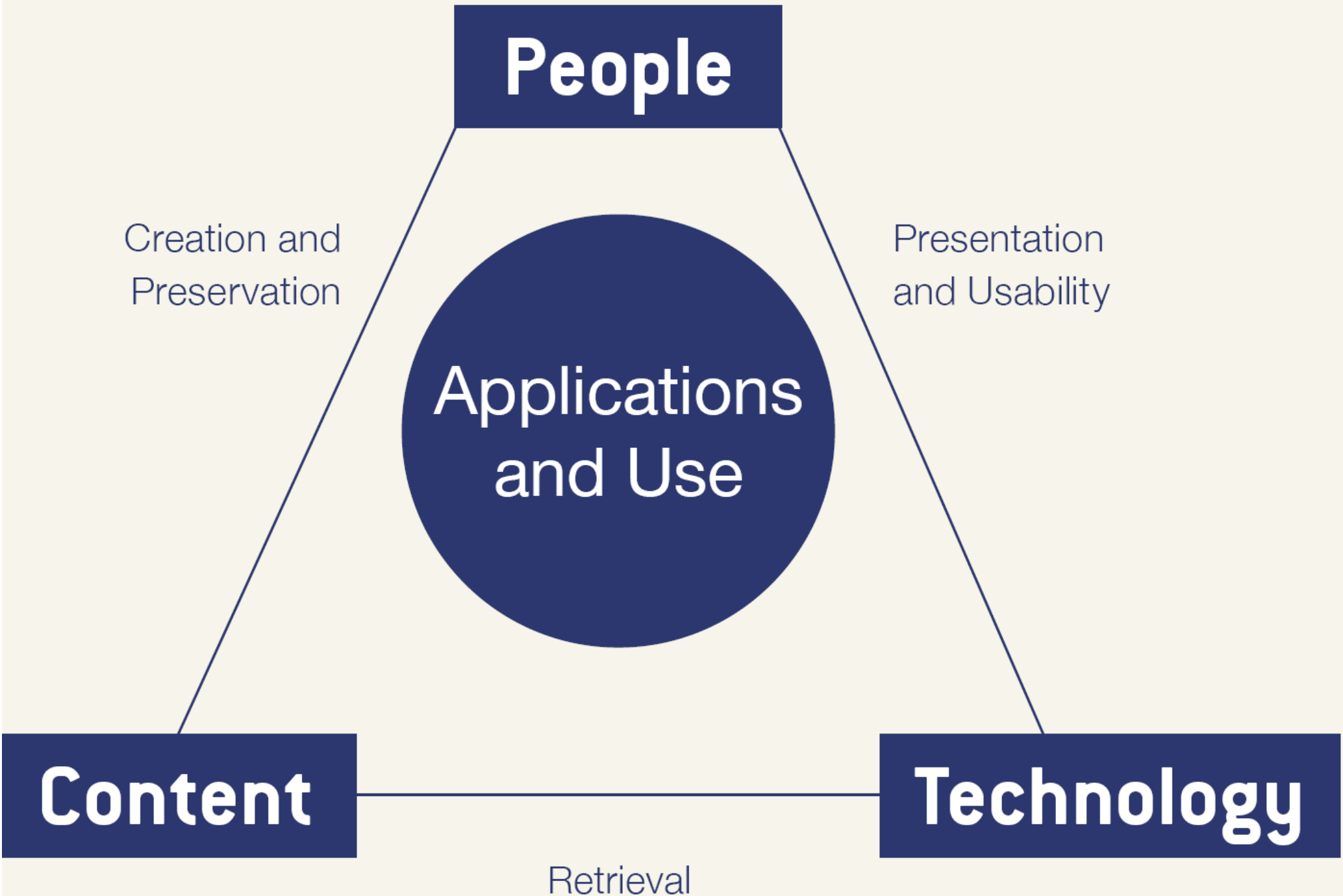
Presentation
and Usability

**Applications
and Use**

Content

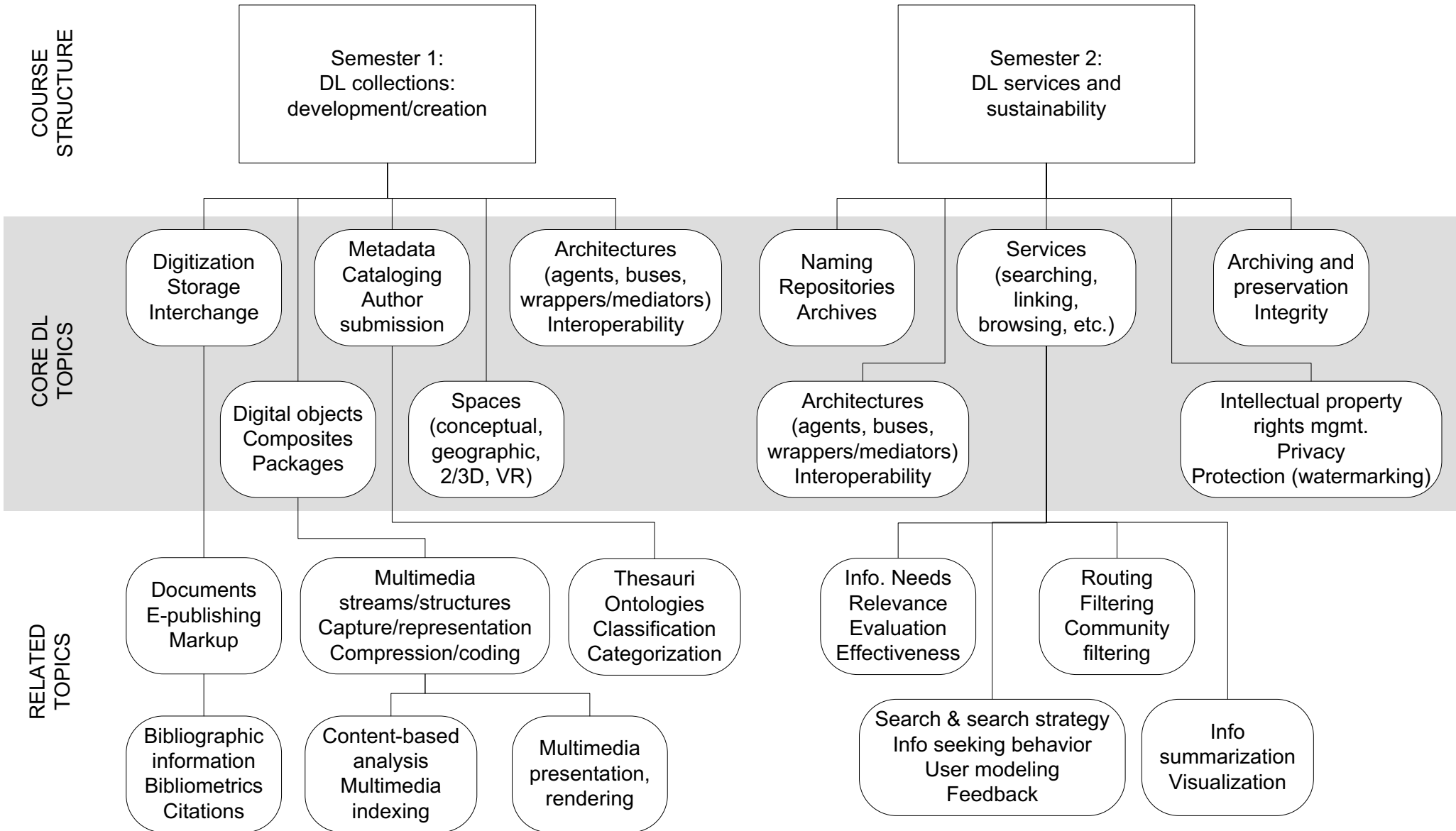
Technology

Retrieval



Introduction

DL Curriculum Framework



Communication Analysis in the Social Interactome

Abigail Bartolome, Advised by Dr. Edward A Fox

NIH Grant: 1R01DA039456-01

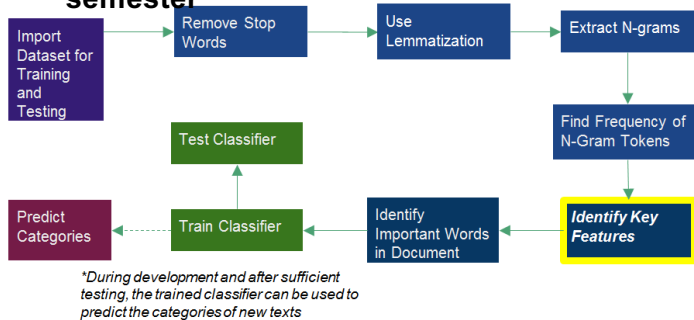
The Social Interactome of Recovery: Social Media as Therapy Development

Acknowledgements to Dr. Chris Franck, Prashant Chandrasekar, Lexie Mellis

Virginia Tech CS 4994, April 2016

Text Classification

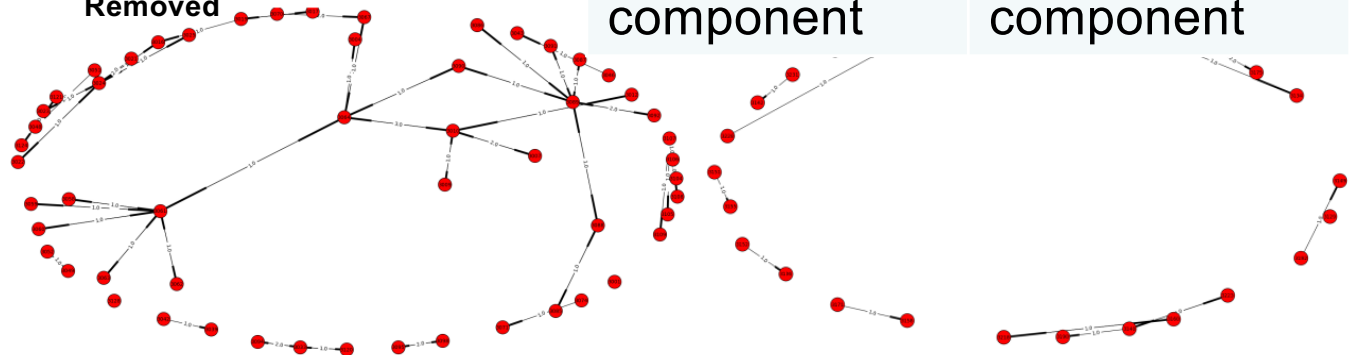
- Multinomial, naïve-Bayes classification considers the count for each feature name in making classifications
- Training the classifier: built a corpus of 150 documents– 75 of which were sentences that were clearly indicative of belonging to a success story and 75 of which were sentences that were not indicative of a success story
- Acknowledgements to Victoria Worrall for her efforts on this classifier last semester



Network Structures

- Queried the Friendica database to see who the participants wrote text to and who the participants received text from
- Generated graph of the private messaging communication in the lattice social network

Lattice Network with Administrator Removed



Lattice Network	Small-world Network
128 participants	128 participants
22 users in the most connected component	4 users in the most connected component

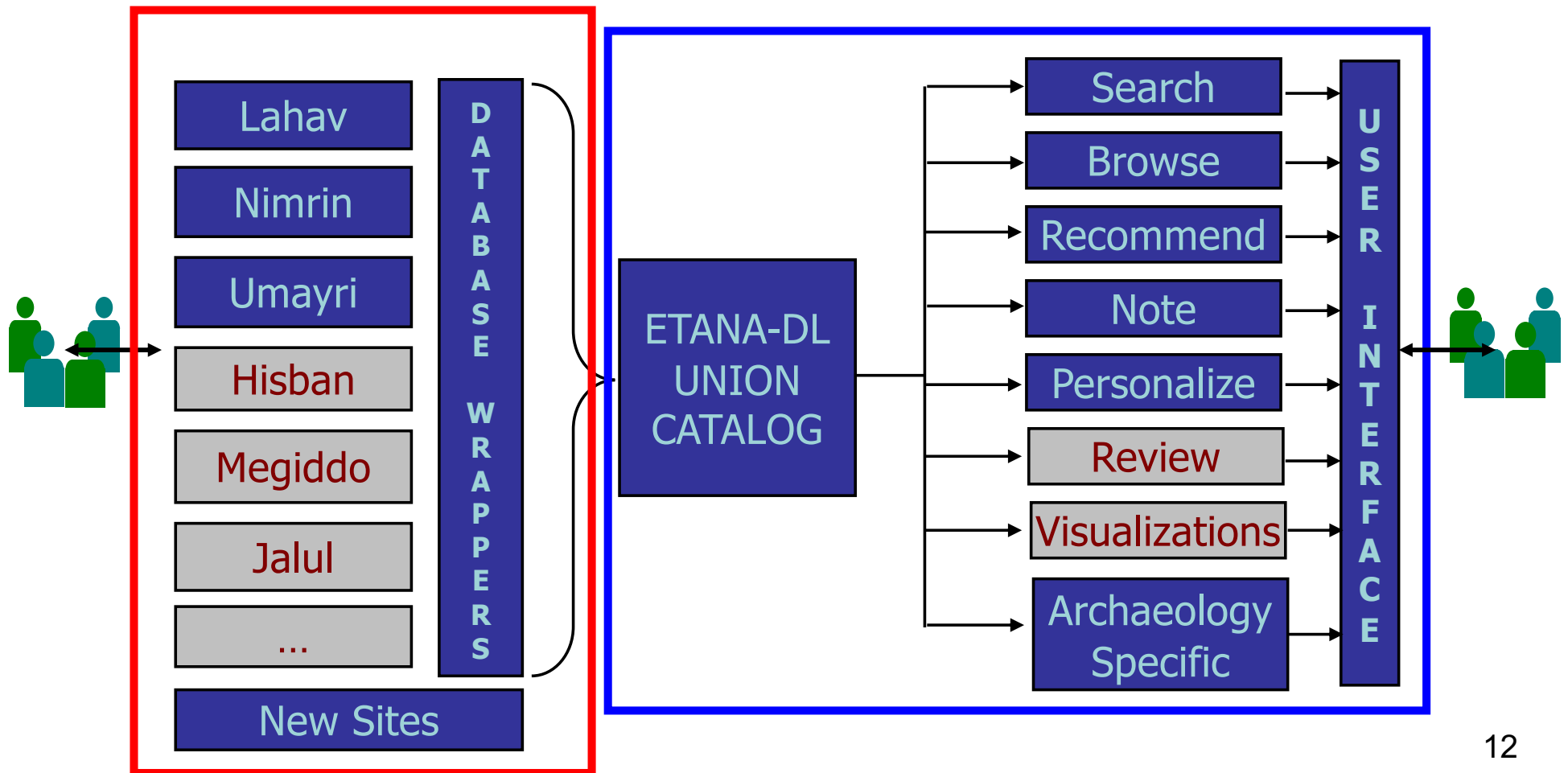
Samples of Story Classification

"Since being in recovery I have not been around any drugs or alcohol but if I had to, such as a wedding or something I wouldn't have a problem saying that I don't drink or I'm in recovery." => success

'Drove very drunk.' => not_success

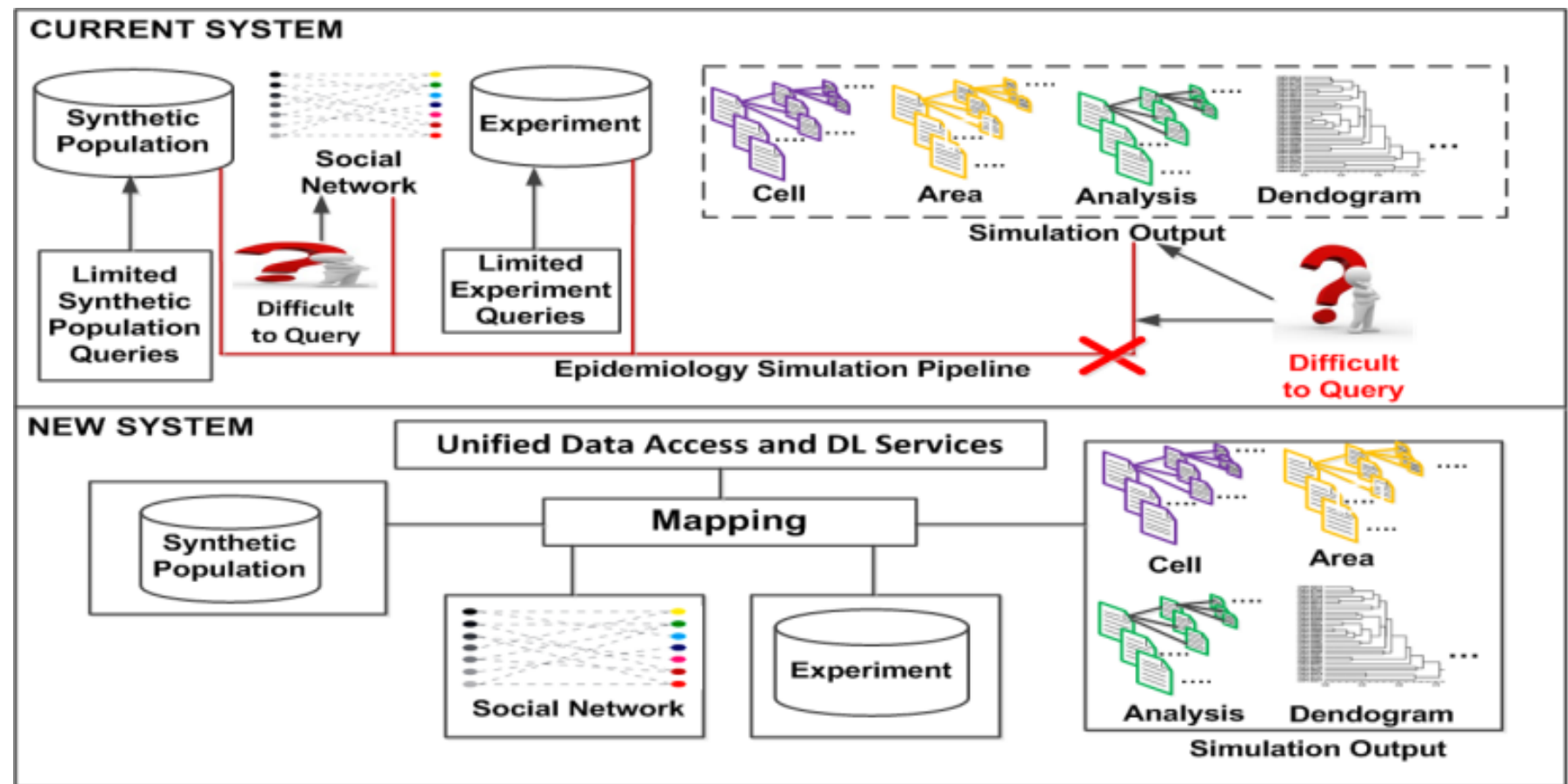
ETANA-DL Architecture

DigBase and DigKit



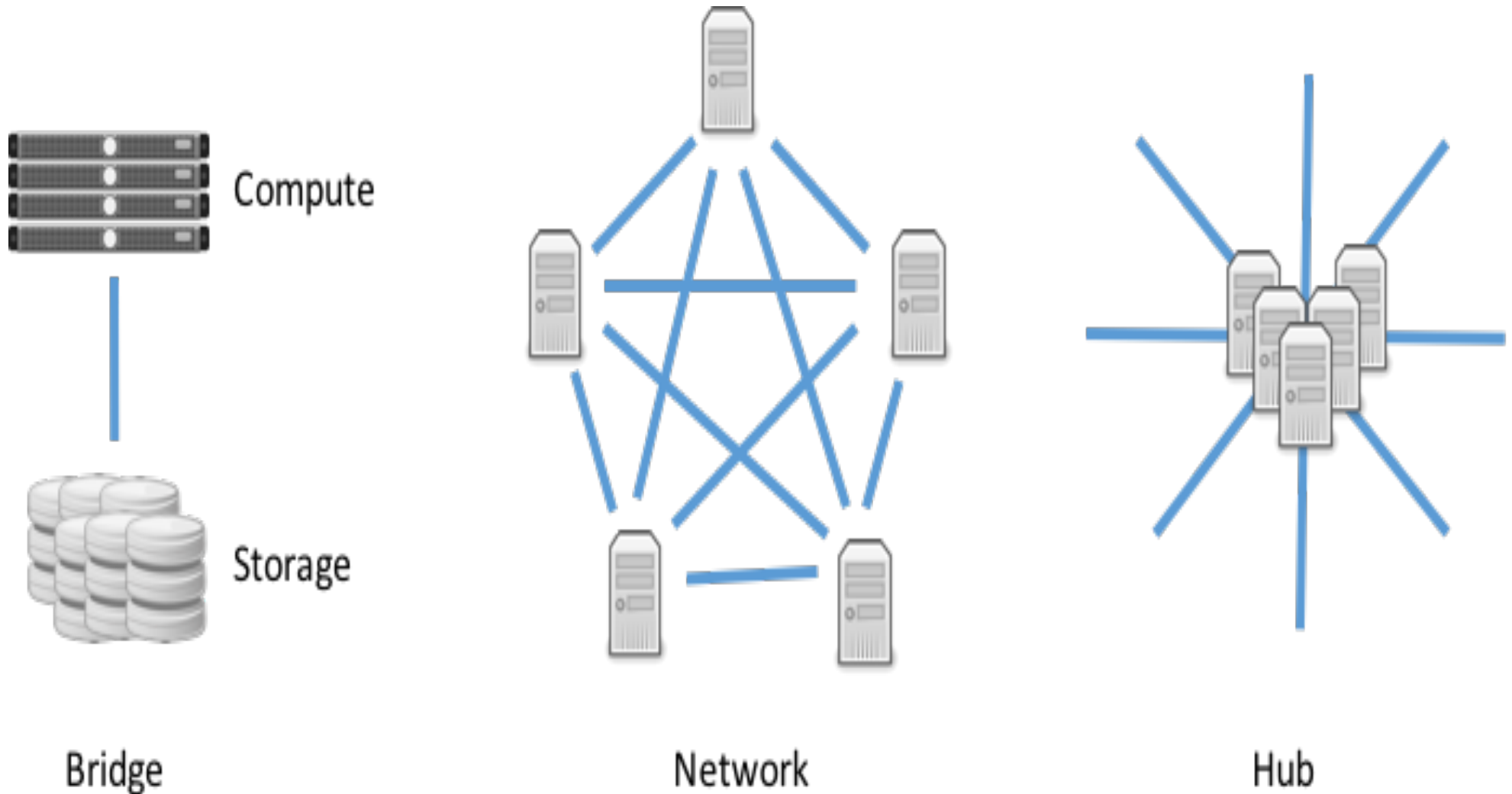
Data Mapping Framework in a Digital Library with Computational Epidemiology Datasets

S.M.Shamimul Hasan, Sandeep Gupta, Edward A. Fox, Keith Bisset, Madhav Marathe --- Virginia Tech (CS, BI)

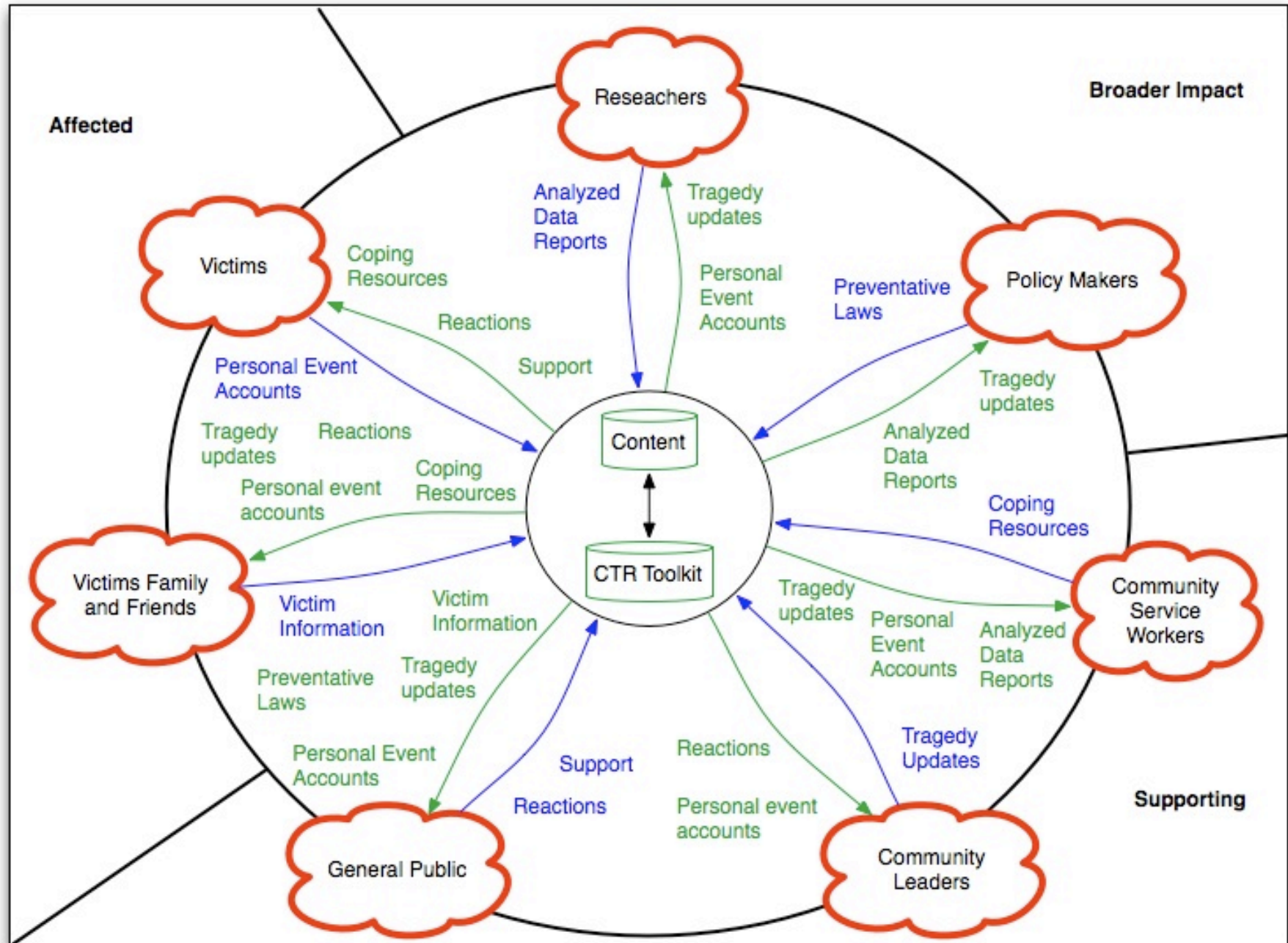


IMLS: Developing Library Cyberinfrastructure Strategy for Big Data Sharing and Reuse

3 patterns for Library Big Data Services



Integrated Digital Event Archive and Library (IDEAL) stakeholders



Who is involved in a WMB ?

- Fix water pipe
 - Water utility
 - city/town utility
- Traffic
 - Police
- Affected
 - Citizen
- Others ...



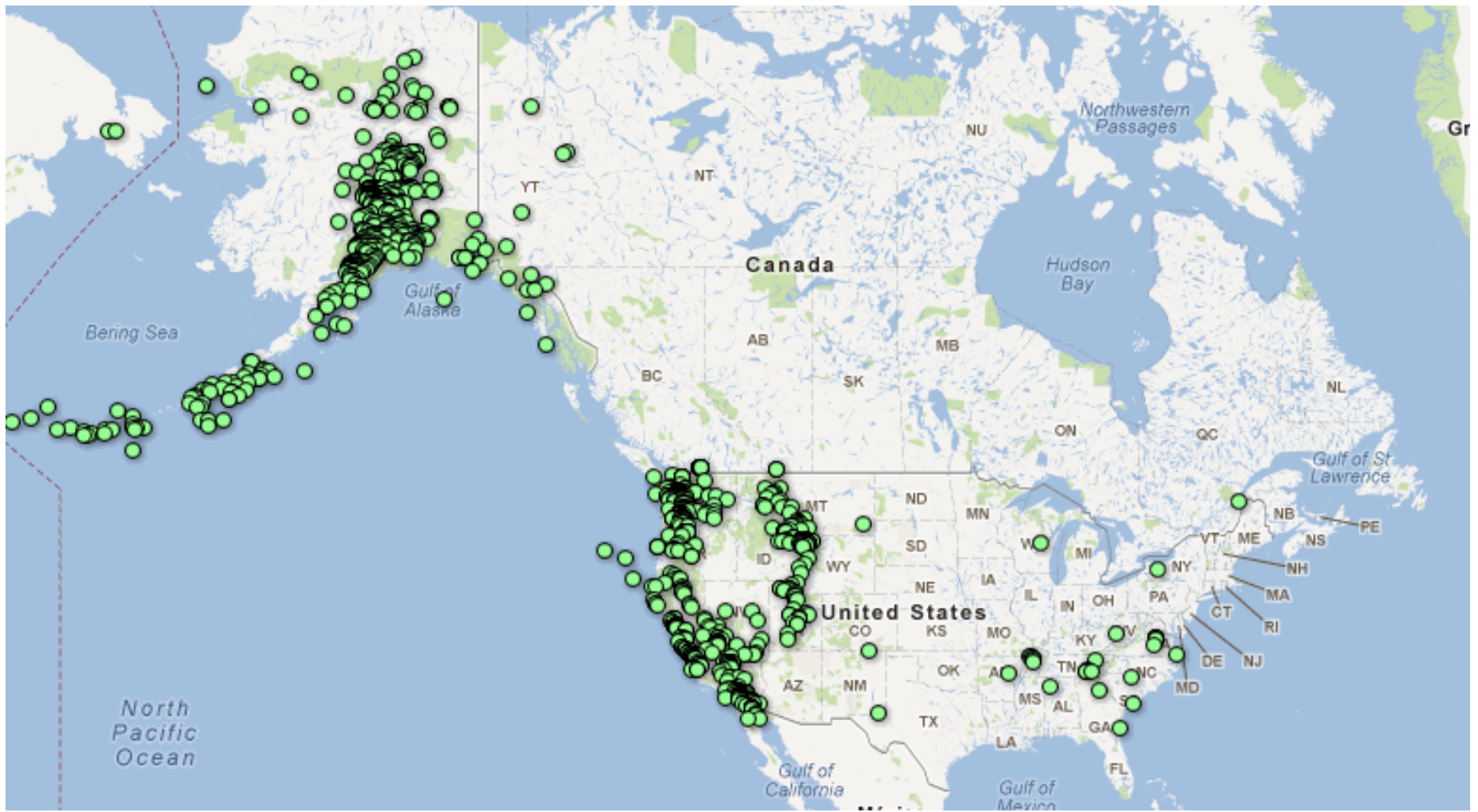
Lakewood, NJ, June.
2014



West Philadelphia, PA, June.
2015

What Causes Water Main Breaks? Earthquakes (USGS)

Mar. 1 – Apr. 5, 2012

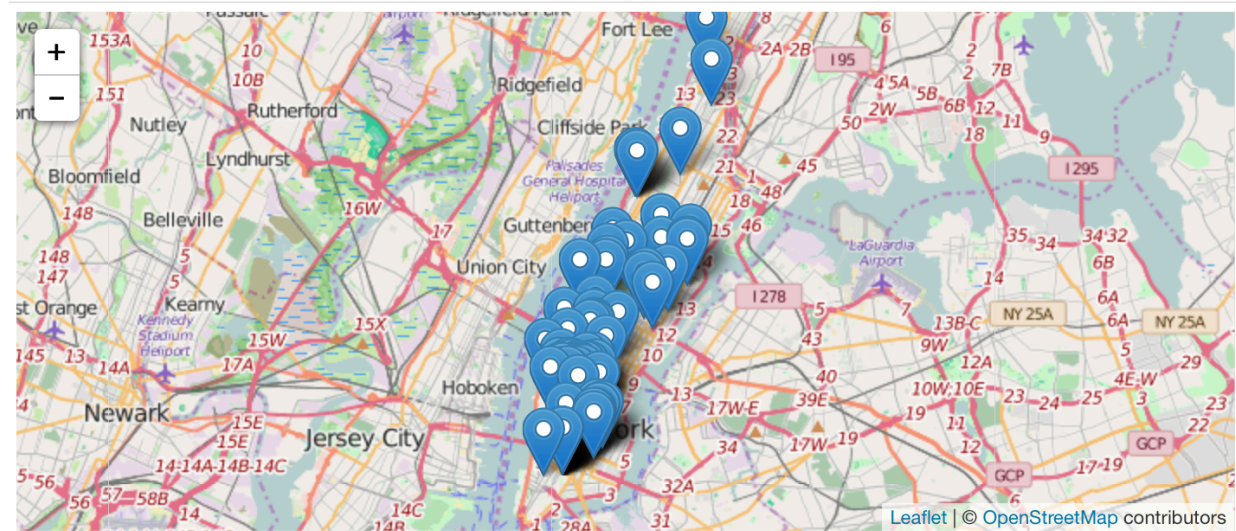


- ie_hashtags_ss**
- NYC (4)
 - MTA (3)
 - nyc (3)
 - 1010wins (2)
 - mta (2)
 - 14thst (1)
 - NYCT (1)
 - NewYorkNews (1)
 - badcommute (1)
 - closed (1)
 - Show more...

- ie_mentions_ss**
- ABC7NY (2)
 - ConEdison (2)
 - NYPDnews (2)
 - 4rings1121 (1)
 - AmyFreeze7 (1)
 - DNAinfo (1)
 - FDNY (1)
 - Gothamist: (1)
 - Local4News: (1)
 - MTA (1)
 - Show more...

- iso_language_code**

Marker Map



- Columbus (6)
 - Denver (49)
 - Show more...
- geo_type**
- Point (80)
 - (0)

Grid Results

- Filter fields
- All (34) / Current (2)
- Field Name
 - _version_
 - archivesource
 - cleantext
 - col_name
 - colnum
 - created_at

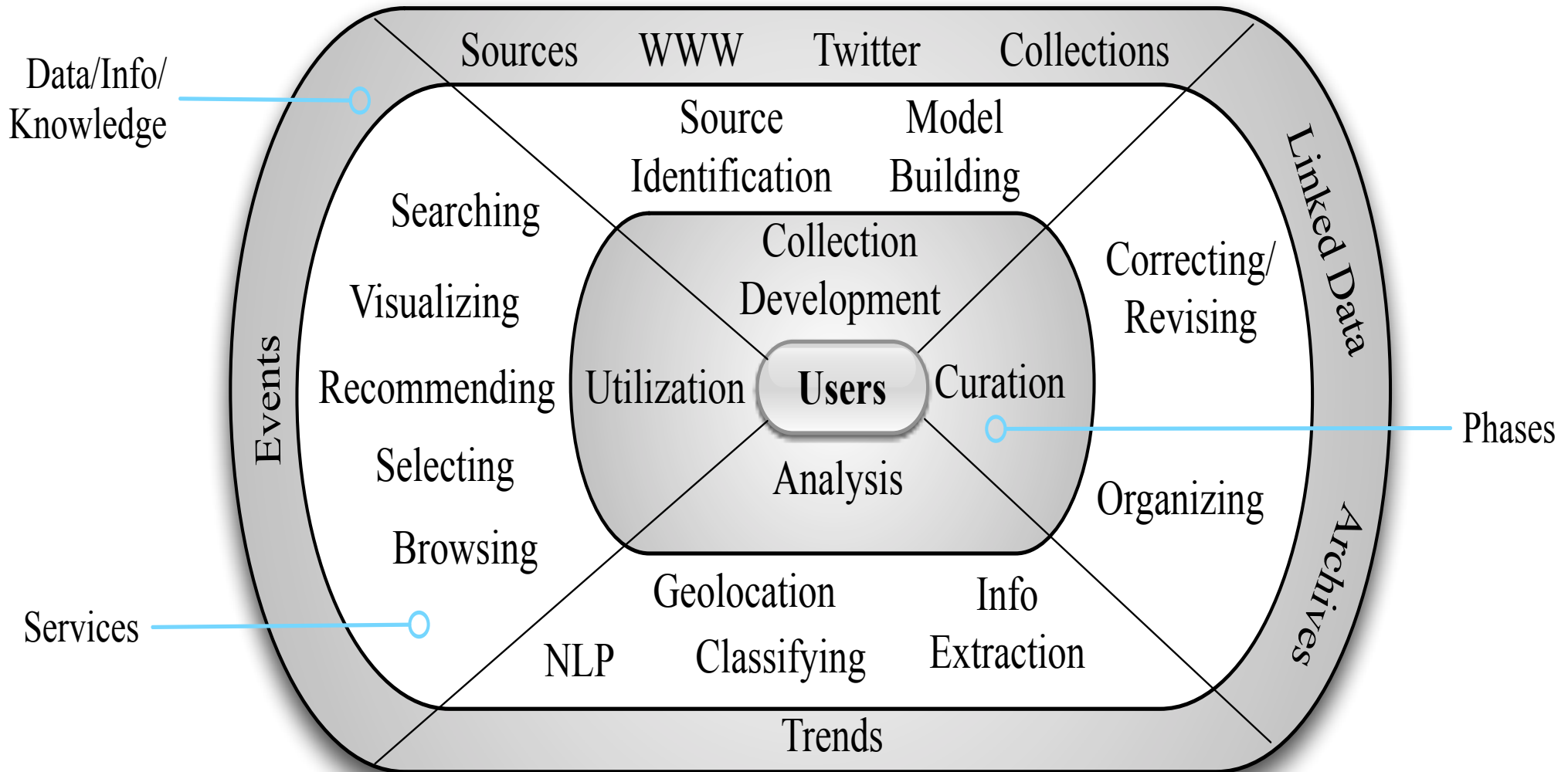
Showing 1 to 100 of 80 results

text

- ▶ West Village water main break causing traffic problems: The main broke near Greenwich and C
- ▶ Reroutes due to water main bk at West 4th. Use numbered lines if possible. #MTA http://t.co/e/
- ▶ Rep from Comptroller @scottmstringer's office at scene of water main break on 13th St and 5th
- ▶ Watch out world. I've got a water main break, and the pix stitch app. âœ“âœ“âœ“
- ▶ Looks like a water main break at 14th and 6th, lots of MTA emergency presence at W 4th and
- ▶ @watermainbreak (@ Water Main Break) http://t.co/tOTOeRAFyC
- ▶ Between my apartment, the NYU bookstore pipe burst, and this Union Square water main brea

GETAR Architecture

Global Event and Trend Archive Research



Area	Description	Investigators	Courses
Core Research			
Analysis, access	Databases, HCI, information visualization, machine learning, ontologies, statistics	Fox, Franck, Huang, North, Sheetz	BIT 4524, 4544, 4614; CMDA/CS/ STAT3654; CS5764
Library, information, data	Archives, big data, curation, data management, decision support, exploring, knowledge engineering, searching	Fox, French, Nicholls, Speer, Thomas, Zobel	CS4624, 5604, 6604; FOR3604; GRAD5134
NLP	Arabic, document analysis, errors, information extraction, summarization, topic identification	Eubank, Fox, Rozovskaya	CS4624, 4984, 5984, 6804
Applied Research Across Disciplines			
Geospatial	Car crashes, data structures, GIS, maps, queries, traffic, tweets, weather and crops	Baird, Lu, Sforza	GEOG1115, 1116
Simulation	NDSSL: epidemiology, diffusion in networks, planning response	Eubank, Lewis, Swarup	GBCB5874, 7994
Climate change	Adaptation, biodiversity, conservation, ecology, ecosystems, effects on plants& animals, environment, sea-level rise	Bukvic, Jelesko, Kalkstein, Quinn	GEOG2994, 4974, 4994; PPWS4994
Economics	Development, families, game theory, Middle East, smart cities, social networks	Ball, Korkmaz, Salehi-Isfahani	ECON3004
Energy	Green engineering, nuclear policies	Avey, McGinnis	ENGR3124
History	Globalization, Soviet history	Ewing	HIST1214, 1215, 2124, 3394, 3554
Innovation	CIE: entrepreneurship, impact of resources, industry collaboration, social and technology based ventures	Junkunc	MGT3064, 3074, 4064, 4094
Resilience	Concentrations, dependency, disasters, evacuations, planning, policy, relocation, supply chains, urban and regional, vulnerability	Bohland, Bukvic, Lawrence, Murray-Tuite, Zobel	CEE5620, 5660; GRAD5134
Sociology	Crises, global issues, social inequality, social movements, social participation, violence, social networks, communication behavior and effects (incl. in Maasai society)	Baird, Kavanaugh, Shoemaker, Wimberley	SOC2034, 2044, 3304, 3504, 3854, 4354, 4424, 4444, 4764, 5424
Political Science	National security, world politics, nuclear policies	Avey, Nicholls	PSCI1004, 1024, 1034, 2034, 2054, 2064, 3114, 3514-6, 3524, 3544, 3564, 3515-6, 3624, 3634, 3684-5, 3794, 4734, 5254, 5264, 5284, 5384, 5424, 5444, 5464, 5474, 5514, 5524, 5584, 5624, 5634, 5644

GETAR:

Areas, Investigators, Courses

Where Can You Fit in CS?

CS Looking Outward:

- Interaction: Games, Graphics, HCI, VR/AR
- Programming: Algorithms, Languages, Problem Solving, Workflows
- Simulation: Agents, Modeling: Epidemiology
- KID: Knowledge, Information, Data: AI, Machine Learning

CS – Looking Inside:

- HPC <-> PC <-> GPU
- Networking
- Programming
 - Algorithms,
 - Languages,
 - Problem Solving
 - Workflows
- Systems
- Theory

Women in CS Faculty

1. Jin-Hee Cho
2. Hoda Eldardiry
3. Margaret Ellis
4. Andrea Kavanaugh
5. Aisling Kelliher
6. Wenjing Lou
7. Na Meng
8. Tanushree Mitra
9. Deborah Tatar
10. Danfeng (Daphne) Yao
11. Liqing Zhang

Questions?

At least learn: Computational Thinking

Volunteers and students from high school to Ph.D.

<http://fox.cs.vt.edu>

Feel free to contact: fox@vt.edu