

# CS1944

Virginia Tech, Blacksburg, VA 24061 USA  
28 October 2014

## KID (Knowledge, Information, Data) Track

by Edward A. Fox

- fox@vt.edu    <http://fox.cs.vt.edu>
- Dept. of Computer Science, Virginia Tech
- Blacksburg, VA 24061 USA
- <http://fox.cs.vt.edu/talks/2014/20141028FoxCS1944KIDtrack.pdf>

# Acknowledgements

- Faculty (slides included)
  - Doug Bowman
  - Kirk Cameron
  - Naren Ramakrishnan
- Current IDEAL Ugrad Students
  - Jieun Chon (CS)
  - Alex Cummins (CS)
  - Megan Eyler (Political Science)
  - So Hyun Jo (CS – see visualization)
  - Ah Young Kim (CS)
  - Nikhil Plassman (CS)
  - J. Braeden Sebastian (CS)
- Grants: NIH 1R01DA039456-01 (Addiction Recovery), NSF 1319578 (IDEAL), NSF 1141209 (Computational Linguistics), ...

# Outline

- KID: courses, capstones, approach
- Jobs
- DAC slides (Ramakrishnan)
- HCI slides (Bowman)
- SeeMore slides (Cameron)
- Fox slides
- Discussion

# KID

- Knowledge, Information, Data
- <http://www.cs.vt.edu/undergraduate/tracks/kid>
- People: Fox, Harrison, Huang (in Jan.), Lu (N. VA, graduate), Murali, Prakash, Ramakrishnan (Arlington)



# KID Courses

- Background in some areas:
  - CS3414 Numerical Methods; also Prob/Stat
- 4000 level (capstones in italics)
  - 4244 Internet Software Development
  - 4604 Data Base Management Systems
  - *4624 Multimedia, Hypertext, Information Access (Audio, Images, Video, Web, Search Engines)*
  - *4634 Design of Information*
  - 4804 Artificial Intelligence
  - *4984 Computational Linguistics*

# Computational Linguistics (and Big Data)

- <http://fox.cs.vt.edu/CS4984CL.htm>
- Fall 2014 -> probably each fall
- Large text collections (corpora)
- AI: Natural language processing (Python)
- (English) text processing, parsing, mining, extracting, classifying, clustering, generating
- Hadoop cluster, MapReduce
- Problem/Project Based Learning in Teams
  - Generate the best summary for an event corpus<sup>6</sup>

# KID Approach

- Reality (Nature, Life, People)
- Observations (Numbers, Strings, Media)
- Patterns (Grammars, Rules, Distributions)
- Models (HMM, Prob. Inference, Power Laws)
- Representations (Series, Tables, Graphs)
- Processing (Transform, Analyze, Compress, Synthesize, Generate/Report, Predict)
- Presentation/Interaction (Visualize, Explore)

# Work Opportunities -1

- Areas
  - Biometrics, Data Analytics, Data/Text Mining, Information Systems, Machine Learning, Multimedia (Audio, Graphics, Image, Music, Speech, Video), Natural Language Processing, Recommendation, Search Engines, Social Networks, Web
- Jobs
  - Data Scientist, Database Administrator, Information Scientist, Knowledge Engineer

# Work Opportunities -2

- Areas
  - Business Consulting, Finance, GIS, Intelligence, Learning, Libraries, Media
- Organizations
  - Apple, Bloomberg, Facebook,
  - Google, IARPA, LinkedIn,
  - Microsoft, NSA, Oracle,
  - Twitter, USGS

# ***Slides adapted from Dr. Ramakrishnan***

## **Discovery Analytics Faculty:**

### ***Computer Science***

**Edward A. Fox  
Lenwood Heath  
Bert Huang  
Chang-Tien Lu  
T.M. Murali  
Chris North  
B. Aditya Prakash  
Naren Ramakrishnan  
Ravi Tandon  
Layne T. Watson**

### ***Statistics***

**Leanna House  
Scotland Leman**

### ***ECE***

**A. Lynn Abbott  
Dhruv Batra  
Devi Parikh**

### ***MATH***

**Layne T. Watson**

### ***Areas***

**Visual Analytics  
Spatial Databases  
Temporal Data  
Mining  
Social Networks  
Image Processing  
Prob Reasoning  
Bayesian Statistics  
Information Theory  
Computer Vision  
Info Visualization**

### ***Recent faculty awards***

**Dhruv Batra: Microsoft Azure for Research, ARO Young Investigator Award,  
NSF CAREER Award, Google Faculty Research Award  
Devi Parikh: ARO Young Investigator Award, Google Faculty Research Award  
Aditya Prakash: NSA Science of Security Lablet Award**

# Discovery Analytics Center

- Research center focused on “big data” analytics and knowledge discovery in key areas of national interest
  - intelligence analysis, sustainability, forecasting, and health informatics
- Grant and contract funding
  - Federal: NSF, NIH, NEH, DARPA, DTRA, IARPA, DHS, US Army, ONR
  - Industry: General Motors, HP Labs, NEC Labs, Google
- Composition
  - 15 academic + research faculty
  - 75 PhD students



**Blacksburg (Kelly Hall)**



**Blacksburg (Torg Hall)**



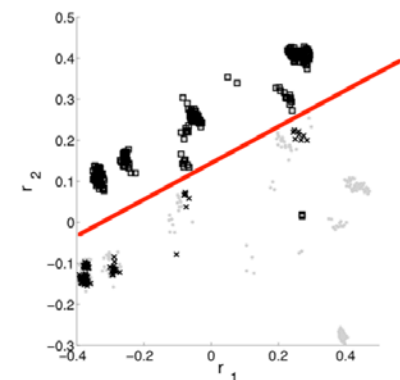
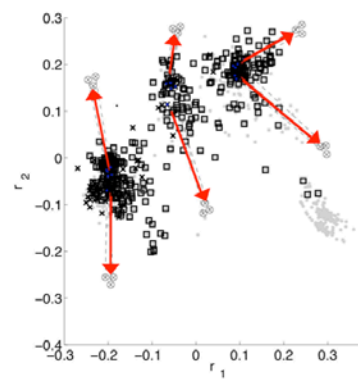
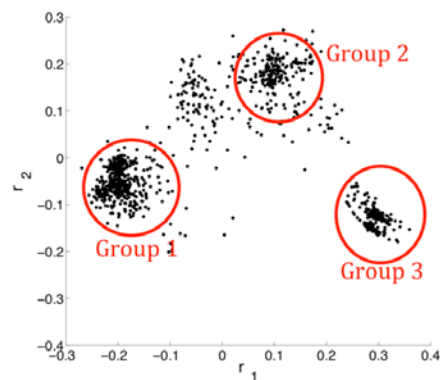
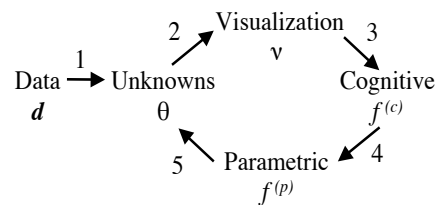
**NCR (Arlington)**



**NCR (Falls Church)**

# Multi-scale V2PI

- Supported by NSF BIGDATA (\$1M)
  - Chris North (CS; PI), Yong Cao (CS), Scotland Leman (STAT), Leanna House (STAT)
- Novel combination of data analytics and human-computer interaction
  - Data analytics: machine learning of model inputs, via model inversion
  - HCI: humans interact directly with model outputs in sensemaking loop





# Initial prototype for text analytics

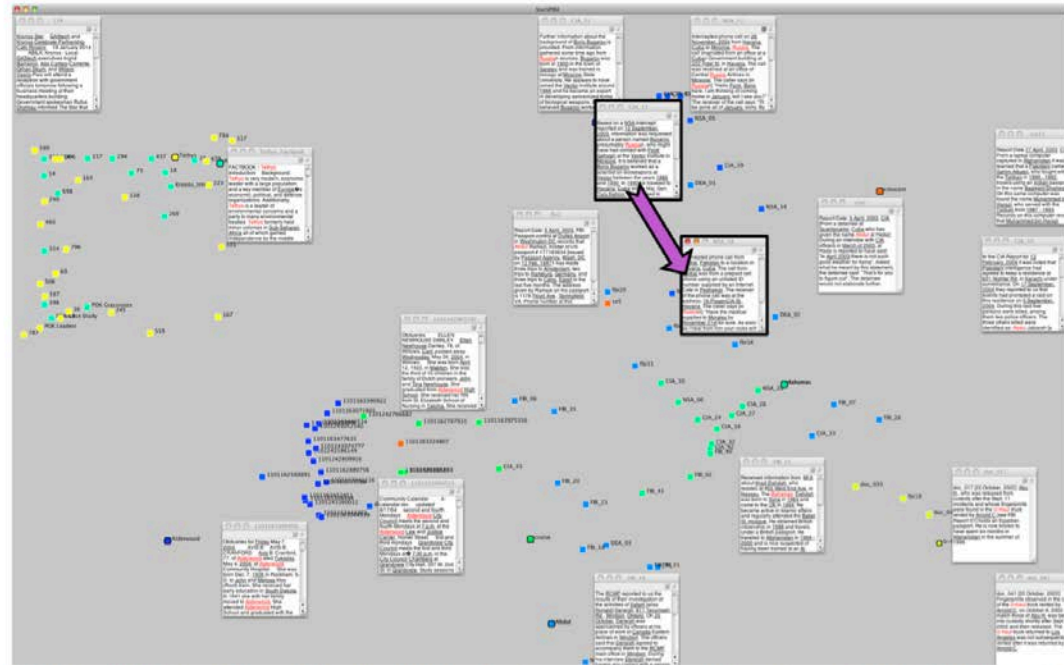


Figure 2. Initial visualization of documents, and user's interaction (re-organizing two documents)

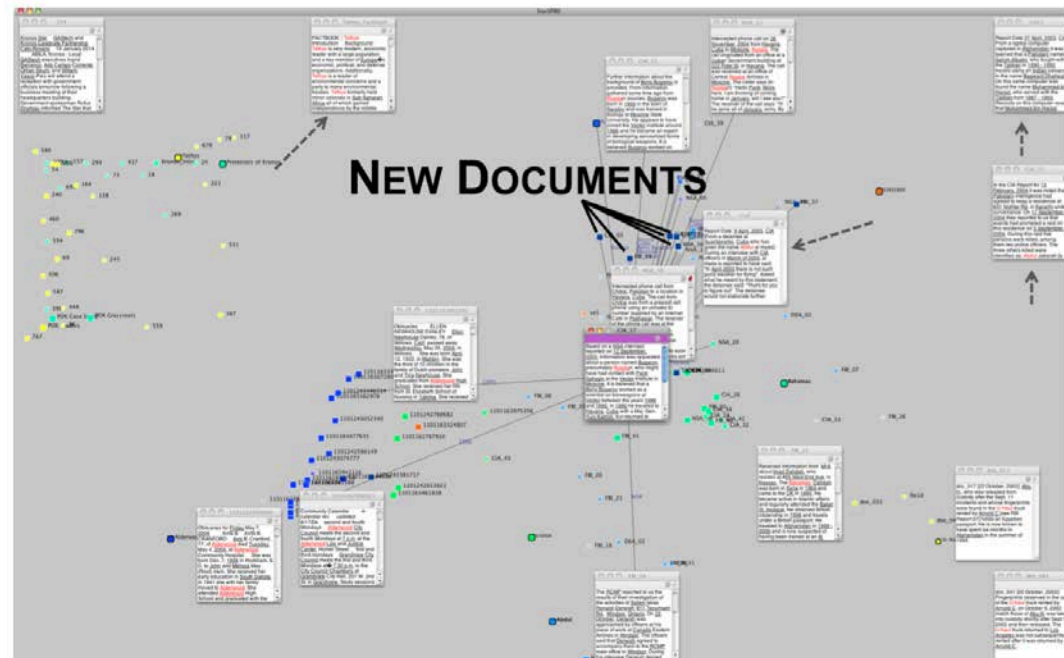


Figure 3. Updated visualization, highlighting newly added documents and newly moved documents

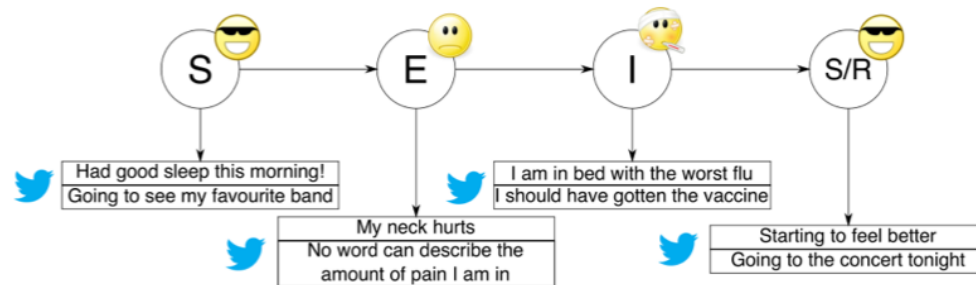
# Propagation in networks

## (Aditya Prakash)

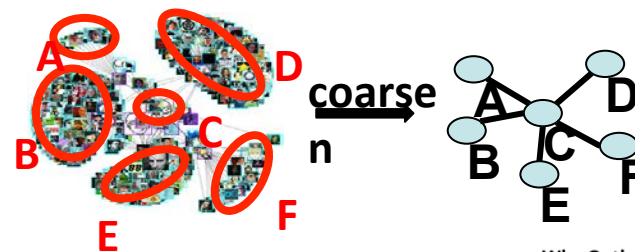
- Effective Immunization



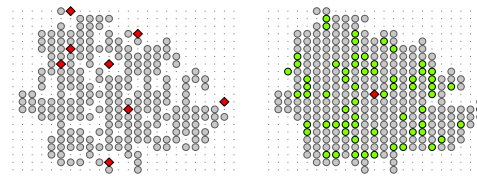
- Efficient Flu Surveillance



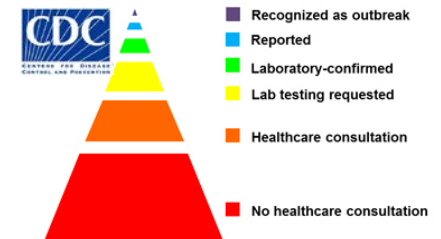
- Meaningful large graph summarization



- Handling uncertain data



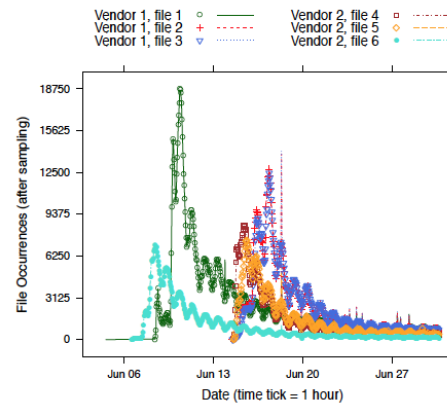
Why Outbreak Surveillance Is Not Complete



# Graph Mining for Cybersecurity

## (Aditya Prakash)

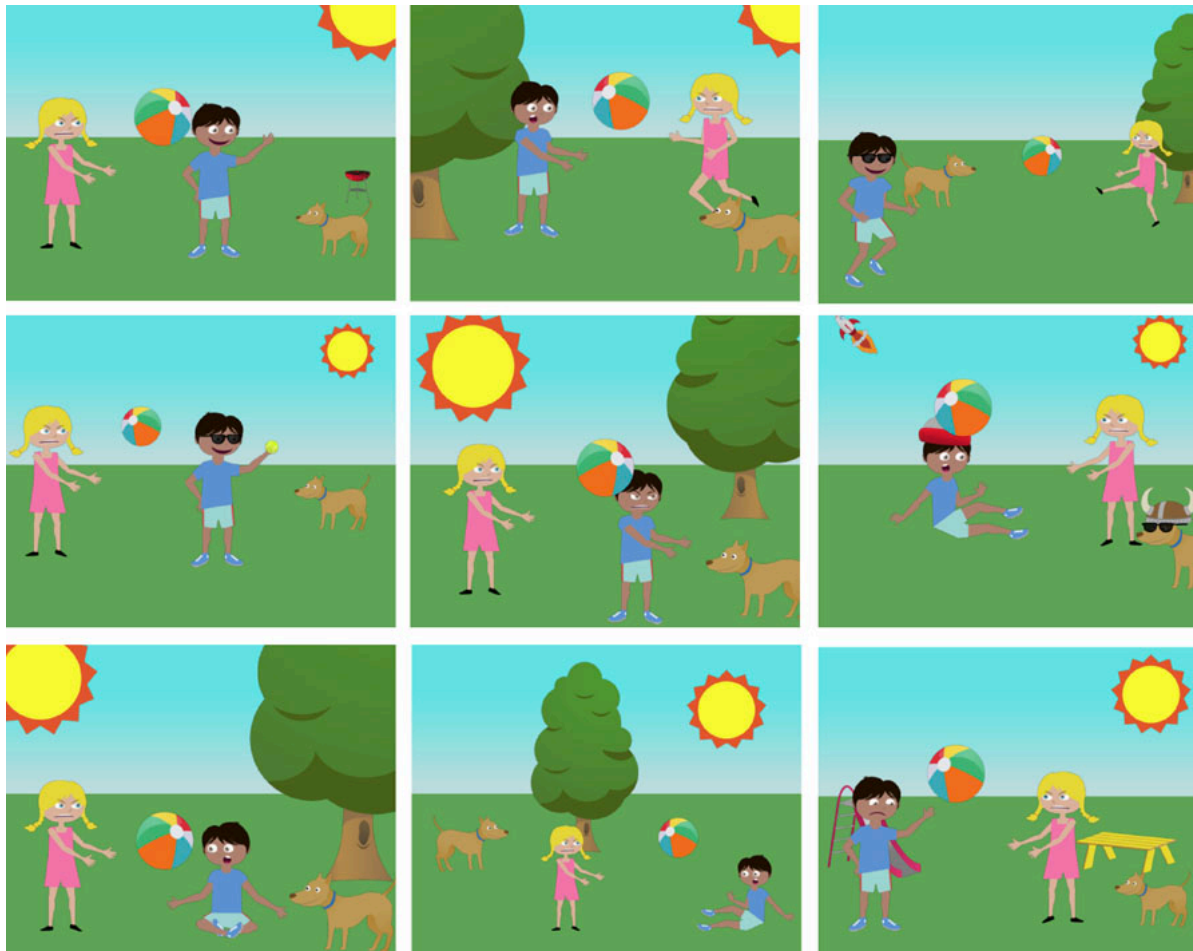
- Modeling malware patterns
- Discovering risky cyberhuman behaviors



# Learning via visual abstractions

(Devi Parikh)

- Teaching computers “context”



Hats



Food



Animals



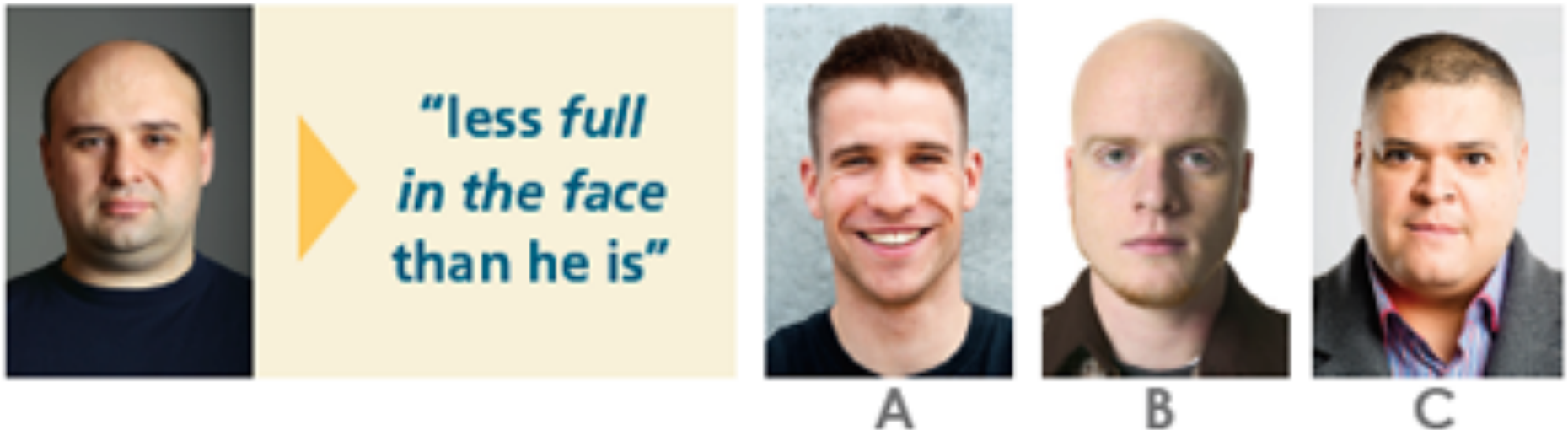
Large objects



# Computer vision

(Devi Parikh)

- “Relative attributes”
  - Using human-machine communication to make image search more natural

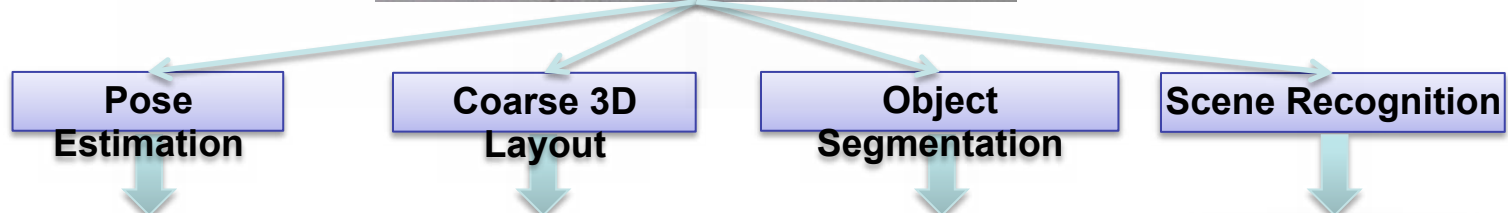


# Holistic scene understanding

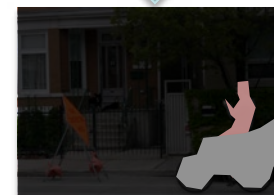
(Dhruv Batra)

- Develop multiple AI modules that talk to each other
  - A module operating in isolation often produces nonsensical results





Hypothesis #1



Hypothesis #2



Hypothesis #M



Consistent

Consistent

Consistent

Contradiction

# The EMBERS Project

- **\$16.8M competitively awarded by the IARPA Open Source Indicators (OSI) program**
  - **Fundamental research (6.1) to develop an early warning system for population-level events using publicly available data**
    - **Goal: Increase time for analysts to respond/analyze/decide**
- **Event classes: civil unrest, disease outbreaks, elections, political crises**
- **Collaboration between computer scientists, statisticians, epidemiologists, social scientists, and political scientists**
- **OSI program geographical focus: Latin America, now MENA**
- **Three initially funded teams downselected → VT**
- **EMBERS has been sending warnings from Nov 2012**



I A R P A



# EMBERS as a “big data” system

**Runs autonomously on the Amazon cloud**



Over **10,000** warnings delivered  
Average 45-50 warnings/day

## Rich diversity of data sources

News	Blogs	Twitter	Facebook	Foursquare	Wikipedia
Humidity	Temperature	OpenTable	Food prices	Stocks	Currencies
ICEWS	GDELT	Parking lot imagery	Routing traffic	Search volume	Economic indicators

# Key events that we forecasted

- Riots after impeachment of Paraguay president (2012)
- The “Brazilian Spring” (June 2013)
- Hantavirus outbreaks in Argentina and Chile (2013)
- Venezuelan student uprising (Feb 2014)
- Recent elections in Panama, Colombia (2014)



Adapted from Slides on

# Human-Computer Interaction & Creative Computing

Dr. Doug A. Bowman

[bowman@vt.edu](mailto:bowman@vt.edu)



# Without HCI, you get this ...

Fac: Order# 99004234 New 99031927 Report Selection OCB SSF View Dupe Load View Routing Print Call Canceled

Mode From SC To SC Air ADT ADT Find CAV# 100670861

Charges: 761.50  
Discount: 0%  
SubTotal: 761.50  
Accessorial: 40.00  
DV: 0 0.00

Charges: 761.50  
Discount: 0%  
SubTotal: 761.50  
Accessorial: 40.00  
DV: 0 0.00

FSC: CAX 2.50% 38.00  
Total: 839.58  
Balance: 839.58

Addend  
Closed  
Post

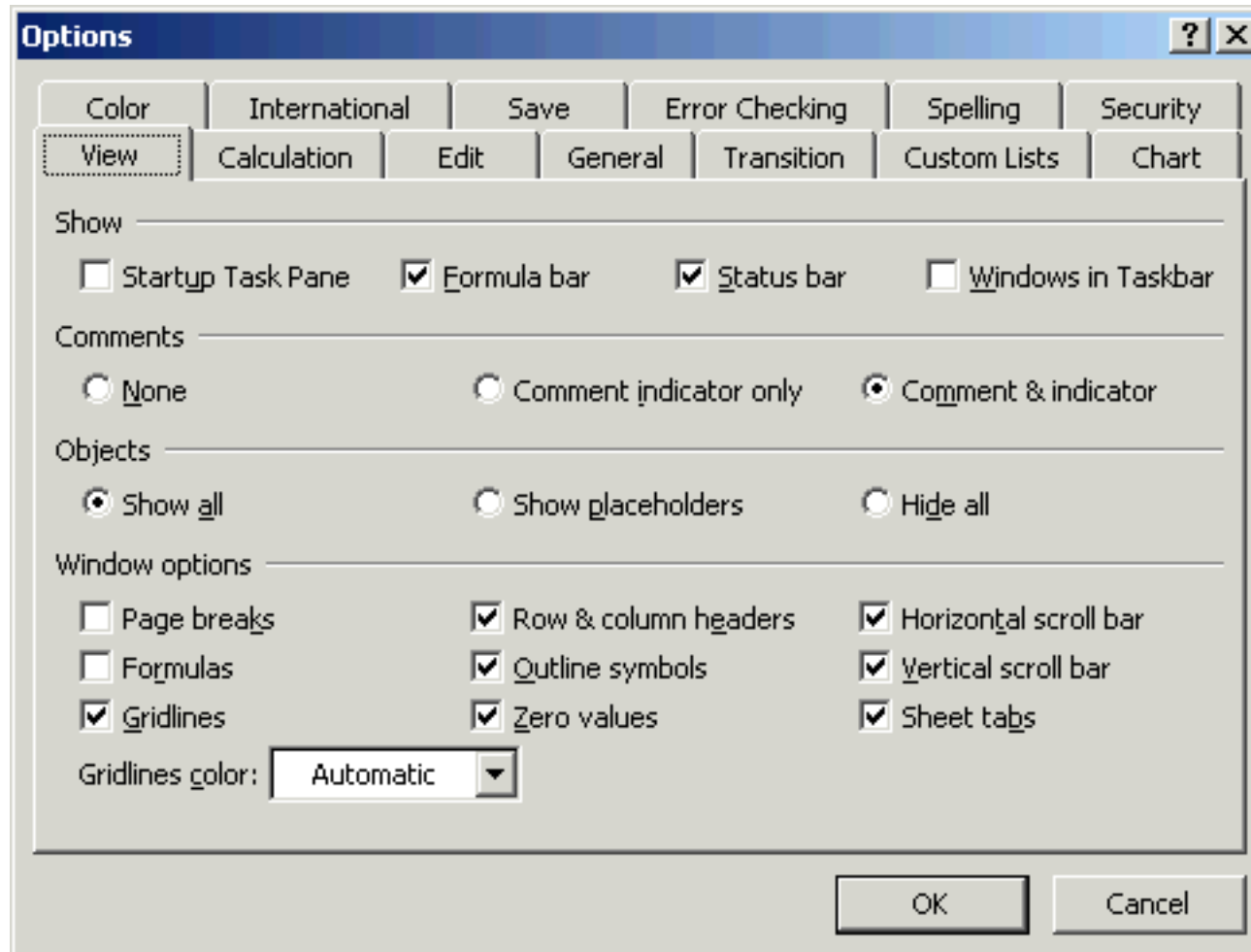
Non-Freight  
Manifest Hold  
Print Hold

SAVED

Units	Type	H Description	Stored	AcWT	Dimensions	ChgWt	Rate	Charge
1	CRATE	CRATE	91	94	97 25x25x30	97	50.00	40.50
1	2MAN	2 MAN P&D					40.00	40.00
2	CRATE	CRATE	500		1,426 50x48x48	1,426	50.00	713.00
0							0.00	0.00

Accs: 140.00 DV: 0 10.00 591 94 152 1,523 761.50

... or this ...





... or this ...



**23 Butikker**

[Gratis el-retur](#)



**www.ARNGREN.net**

Teknologi & Gadgets

103518

< Frithjof i sin Fly-  
Bil

Søk Arngren.net

Slik Betaler du

Bygg ditt eget  
2-seters Helikopter.  
kr. 398.000.-

[Index](#)

**Index**

- Akvarium
- Alarm
- Alkotester
- ATV (el.)
- Betale
- Bil (elektrisk) gas
- Bilbane
- Conrad-elektronikk
- Digital-Kikkert
- Disko-Lys
- Dummy-Kamera
- DVD-Spiller
- Elektronikk & DAB
- EL-Scooter & Bil
- Figurer
- Golf-biler (m/skilt)
- Hobby & RC
- Hoverpod
- HP-Måler (Bil)
- Isbitmaskin
- Kamera (trådløst)
- Kino (bærbart)
- Kompass (bærbart)
- Laser-Jammer (Bil)
- Luft-Jekk
- Lykt (oppladbar)
- Mobil-telefon-1, 2
- Motorsykkel-Mini
- Omformer (110V)
- Omsetter (44 spole)
- PC-mini (9") 2
- Rakett-Fly
- RC Produkter
- Robot-Hund
- Robot-Stevsuger
- Roboter
- Solcelle-Ryggesek
- Star Wars
- Sykkel-PC
- Togbane (til PC)
- Traktor (elektrisk)
- TV-Ur & Armb. Ur
- Walkie Talk

**Trådløs Video-Dørtелефон & Monitor**  
kr. 1798.-

**El-Bil for norske forhold**  
(Vi vant elbil kontrakten med Posten) kr. 99.998.-

**44 Språk-Talende Oversetter**  
m/Norsk fra kr. 298.-

**V22 Osprey 2.4 Ghz**  
kr. 998.- (4ch)

**RC Shark 450 II**  
kr. 2999.-

**Dirt-bike**  
3.499.-

**Original Enzo-Ferrari**  
kr. 3.998.-

**El-ATV**  
6.998.-

**Elektrisk-ATV**  
7998.-

**Elektriske-Biler til Barn, Ungdom & Voksne**  
3999.-

**1298.-**

**2499.-**

**PC-9 Propell-fly 2.4 Ghz**  
kr. 1398.- fra Art-Tech

**Nyhet**

**86 cm lang**

**Elektronikk**

**Jagerfly 4ch 100cm langt**

**m/ Jet-motor**  
kr. 999.-

**Helikopter 4ch 40cm**  
499.-

**Spion-Kamera**  
fra kr. 399.-

**Elektronikk**

**Trådløst Mikro Farge-Kamera sett**  
kr. 399.-

**Styreenhet & Fordelelere til Bil**  
fra kr. 3500.-

**RC-Tank i Metall (55 cm) med Luftkanon**  
fra kr. 998.-

**RC-Produkter**

**Avatar Nyhet**

**Forbruker Elektronikk**

**Elektriske-Kjøretøy**

**Disko-Lys**

**Rakett-Fly**

**Roboter**

**CONRAD**

**Video**

**Fjernstyrte Produkter**

**Kjøpsloven klikk her.**  
[Angreskjema klikk her.](#)

**Elbil; Arngren-Comarth**  
2WD/4WD. 2 eller 4 seter  
kr. 79.998.-

**Avatar-Gunship**  
299.-

**GYRO**

**4-CH**

**Video**

**Ta alltid ut 230Vac Adapteren**  
når du ikke er tilstede, eller sover

**Gratis Mobil-telefon-Spill**

**Video**

**3, 4 & 6 ch**

**Airwolf Helikopter**  
fra kr. 399.-

**RC Herrier VTOL (Lette)**  
kr. 699.-

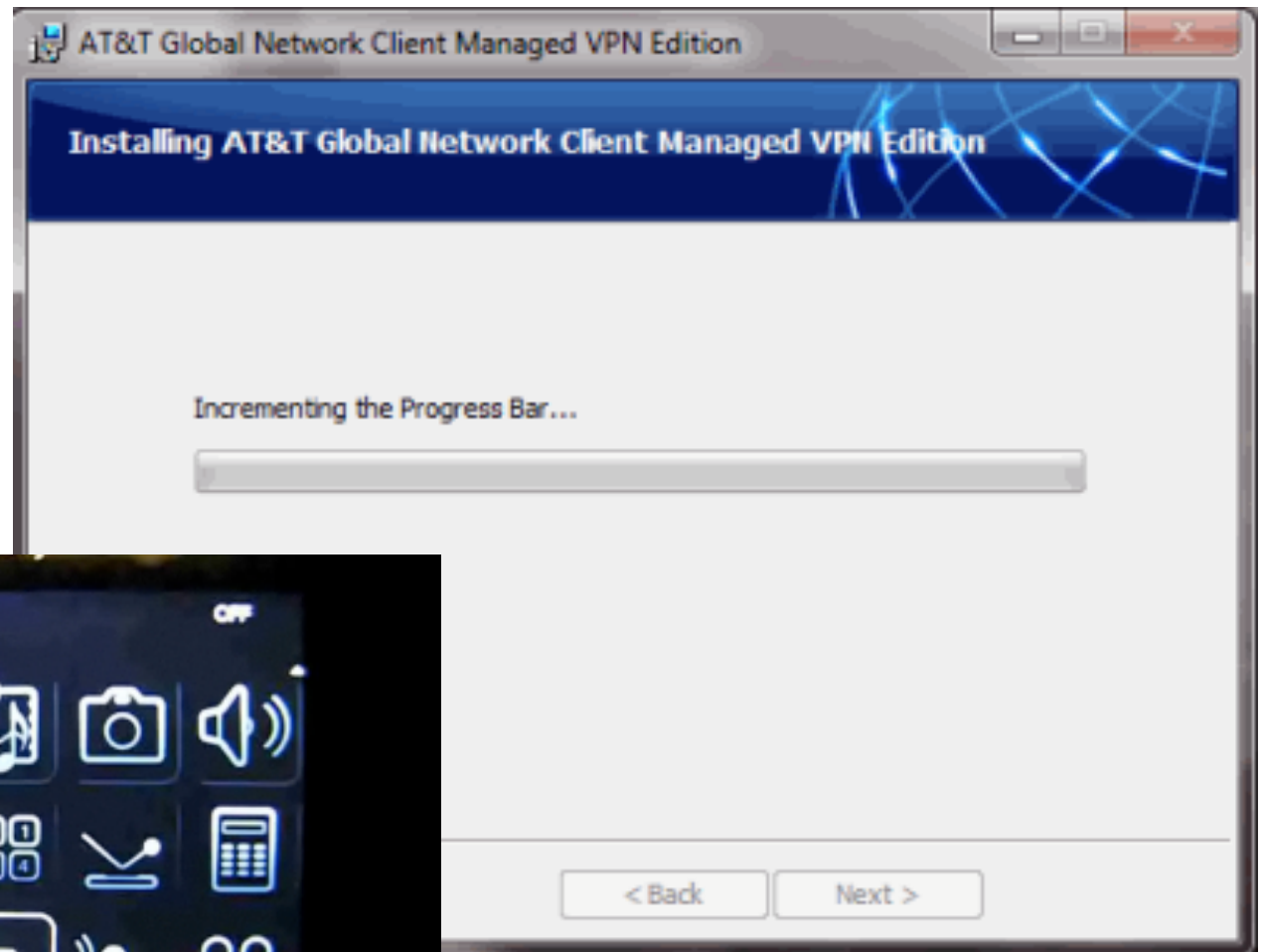
**Kikkert-lengste & kraftigste**

**Hus-Alarm m/nr. sender**  
kr. 599.-

**RC Fly-Båt**  
70 cm lang  
fra kr. 1299.-

Go to "http://www.arngren.net/elektronikk.html"

... or these!





# HCI is everywhere





# HCI in the news

- Using Virtual Worlds to 'Soft Control' People's Movements in the Real One
- Bypassing the Password
- Data Glove Could Help Diagnosis and Treatment of Arthritis
- Microsoft Builds a Browser for Your Past
- Hopkins Researchers Aim to Uncover Which Mobile Health Applications Work
- Guiding Robot Planes With Hand Gestures
- Futurist: We'll Someday Accept Computers as Human

# HCI careers

- User experience / user interface designer
- Interaction designer
- Usability engineer / analyst
- Information architect
- Human factors engineer
- User researcher

# HCI People in CS @ VT



# Faculty

some faculty who research and teach creative CS

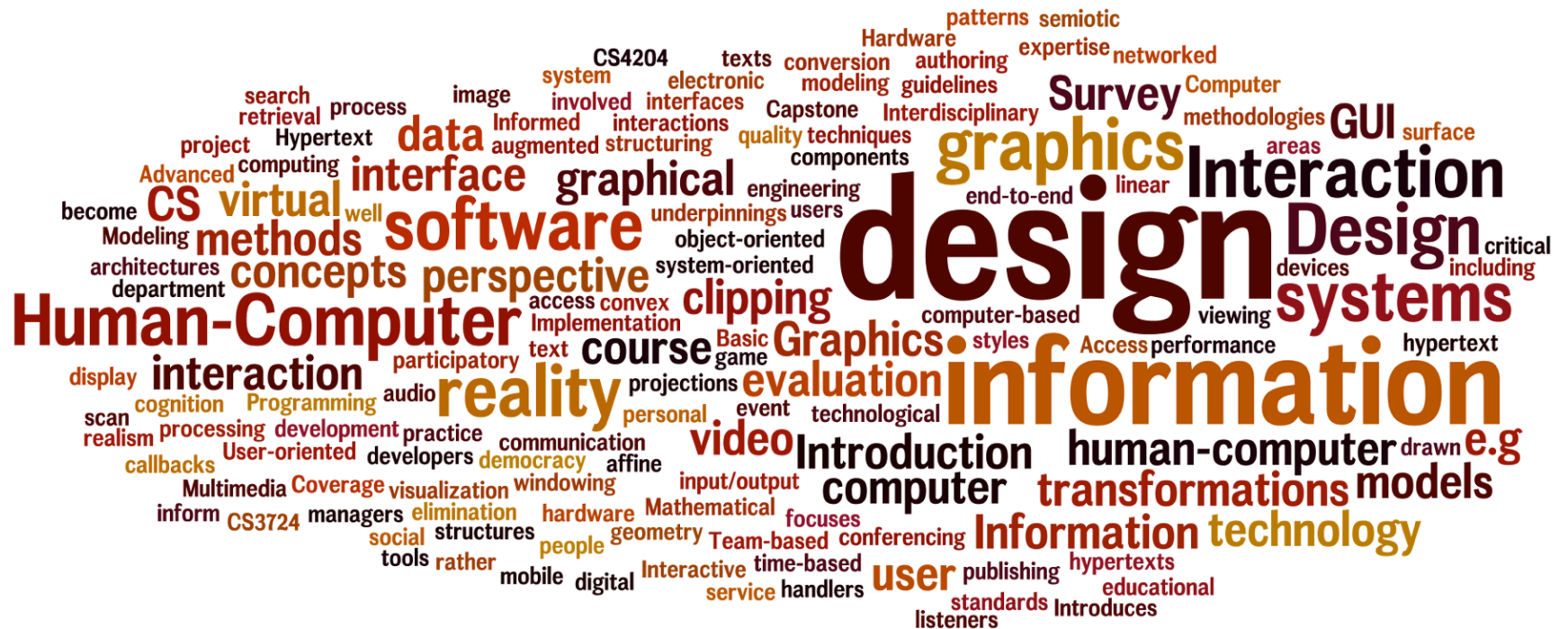
- Yong Cao - computer animation and games
- Ed Fox -- multimedia
- Steve Harrison -- information arts
- Dane Webster (SOVA) -- character animation
- Ico Bukvic (Music) -- digital instruments  
("laptop orchestra")

# HCI labs



# HCI courses

- **Introduction to Human-Computer Interaction**
- **Introduction to GUI Programming and Graphics**
- **Computer Graphics**
- **Design of Information**
- **Human-Computer Interaction Capstone**



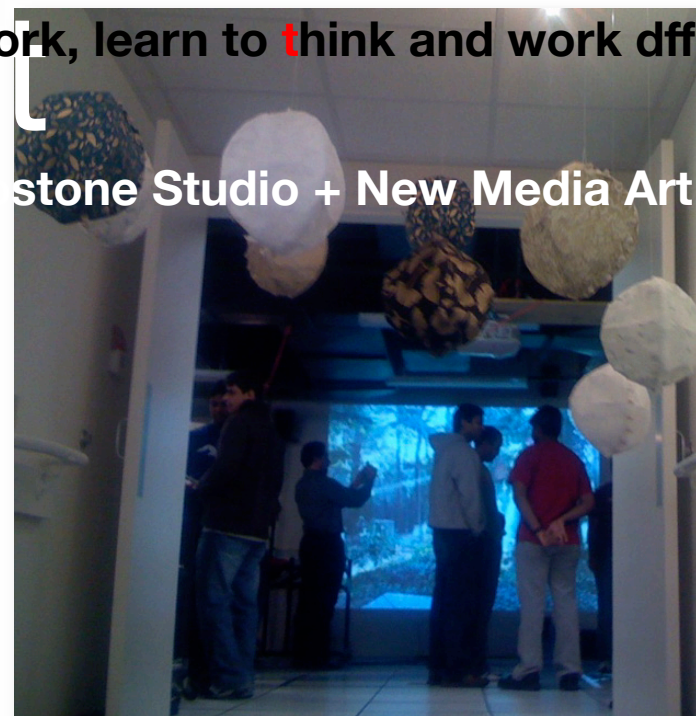
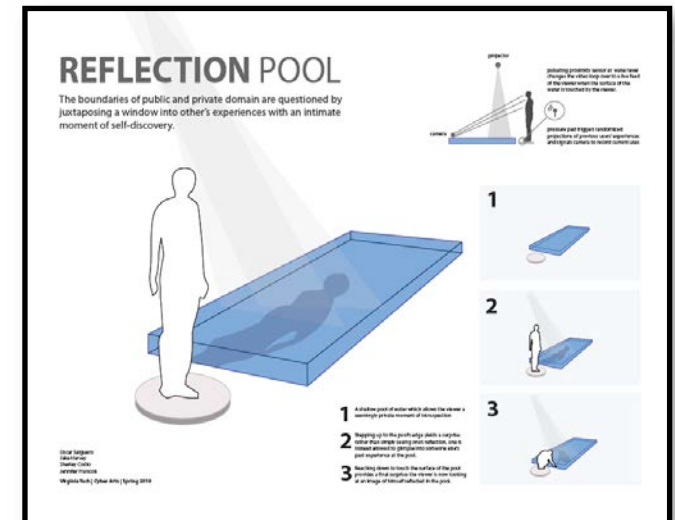
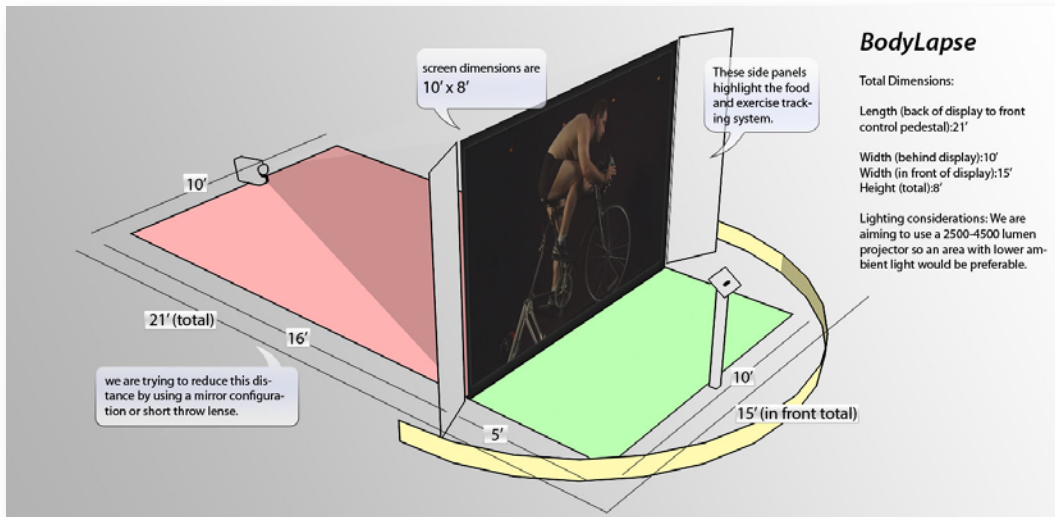
# CS 3724: Unlike other CS classes ...

- No equations (well, maybe one)
- No proofs
- No algorithms
- Some programming, but focus on *design*
- Multi-disciplinary
  - psychology
  - graphic design
  - industrial engineering

# Current topics in HCI

- Large displays
- Small displays
- Human-robot interaction
- Alternative I/O
- Ubiquitous computing
- VR & AR
- Speech recognition
- Video conferencing
- Sustainable interaction design
- Gesture interfaces
- Body-based interaction
- Brain-computer interface





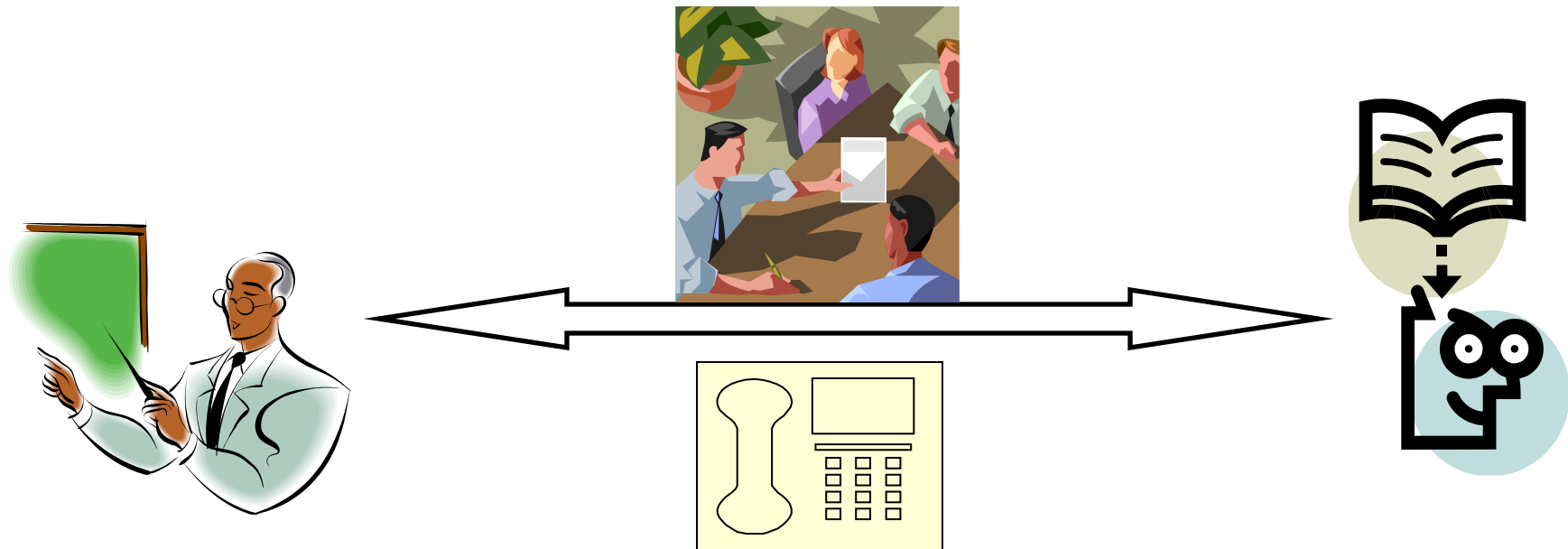
# SeeMore (Kirk Cameron)

- <http://scape.cs.vt.edu/seemore/> slides from Scalable Performance Lab (Scape)
- [https://www.youtube.com/watch?v=JyjB\\_v5rukQ](https://www.youtube.com/watch?v=JyjB_v5rukQ) at SIGGRAPH 2014

# Slides by Fox

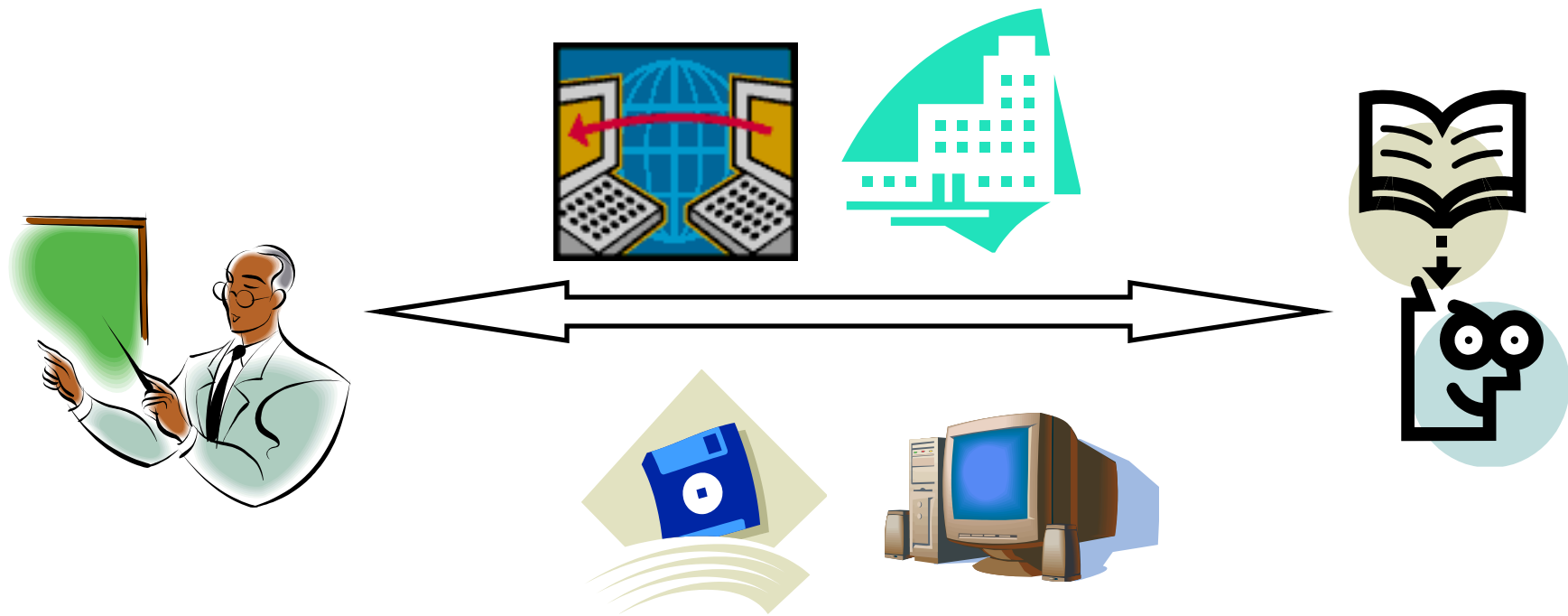
- Introduction
- Current grants
  - Addiction Recovery
  - Integrated Digital Event Archive & Library
  - Qatar: Electronic Library Institute
- Digital Libraries
- Selected projects: Ensemble, ETD classification, ETANA-DL (archaeology)

# Synchronous Scholarly Communication



Same time, Same or different place

# Asynchronous, Digital Library Mediated Scholarly Communication



Different time and/or place

# Digital Libraries Shorten the Chain from

Author

Editor

Reviewer

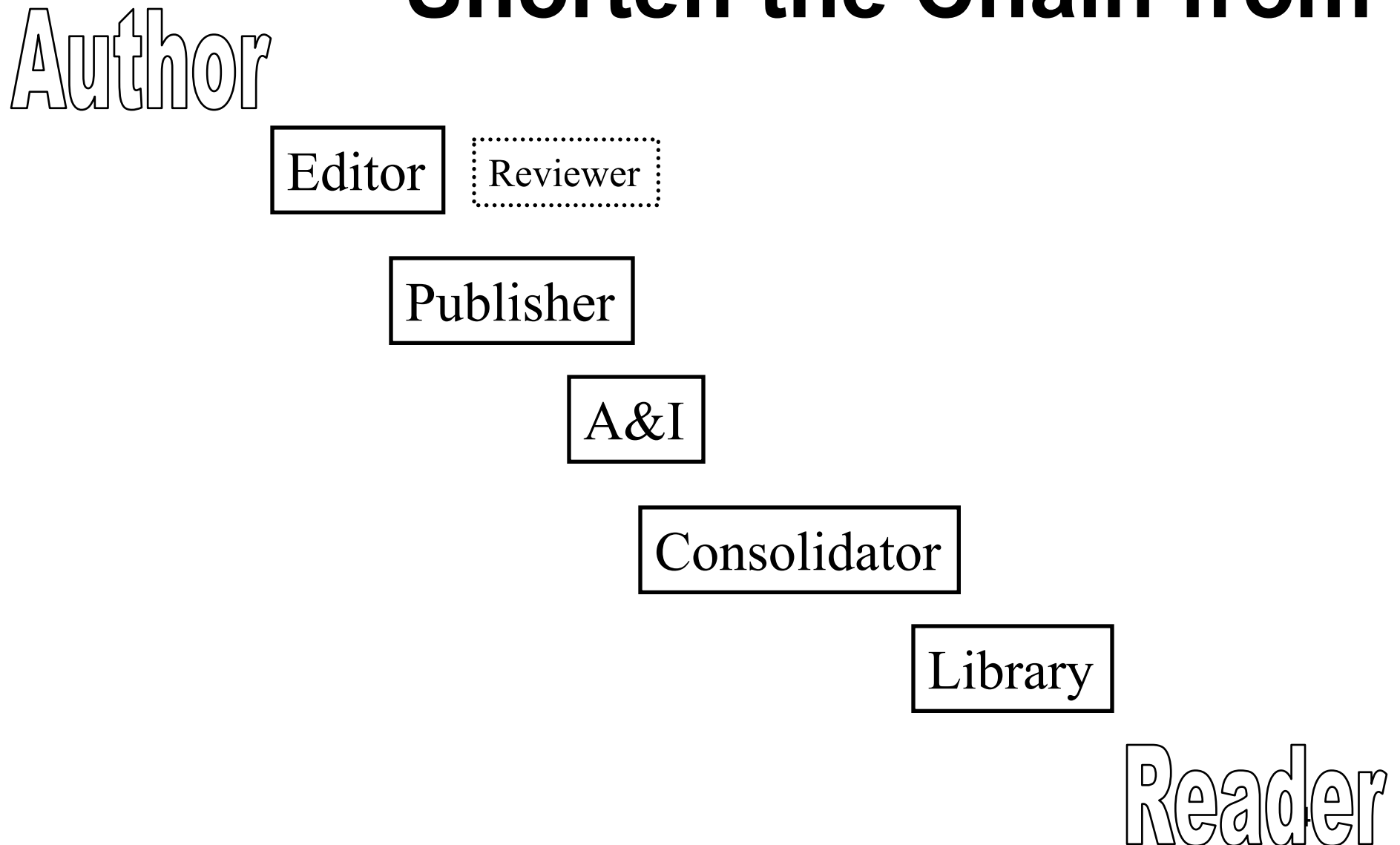
Publisher

A&I

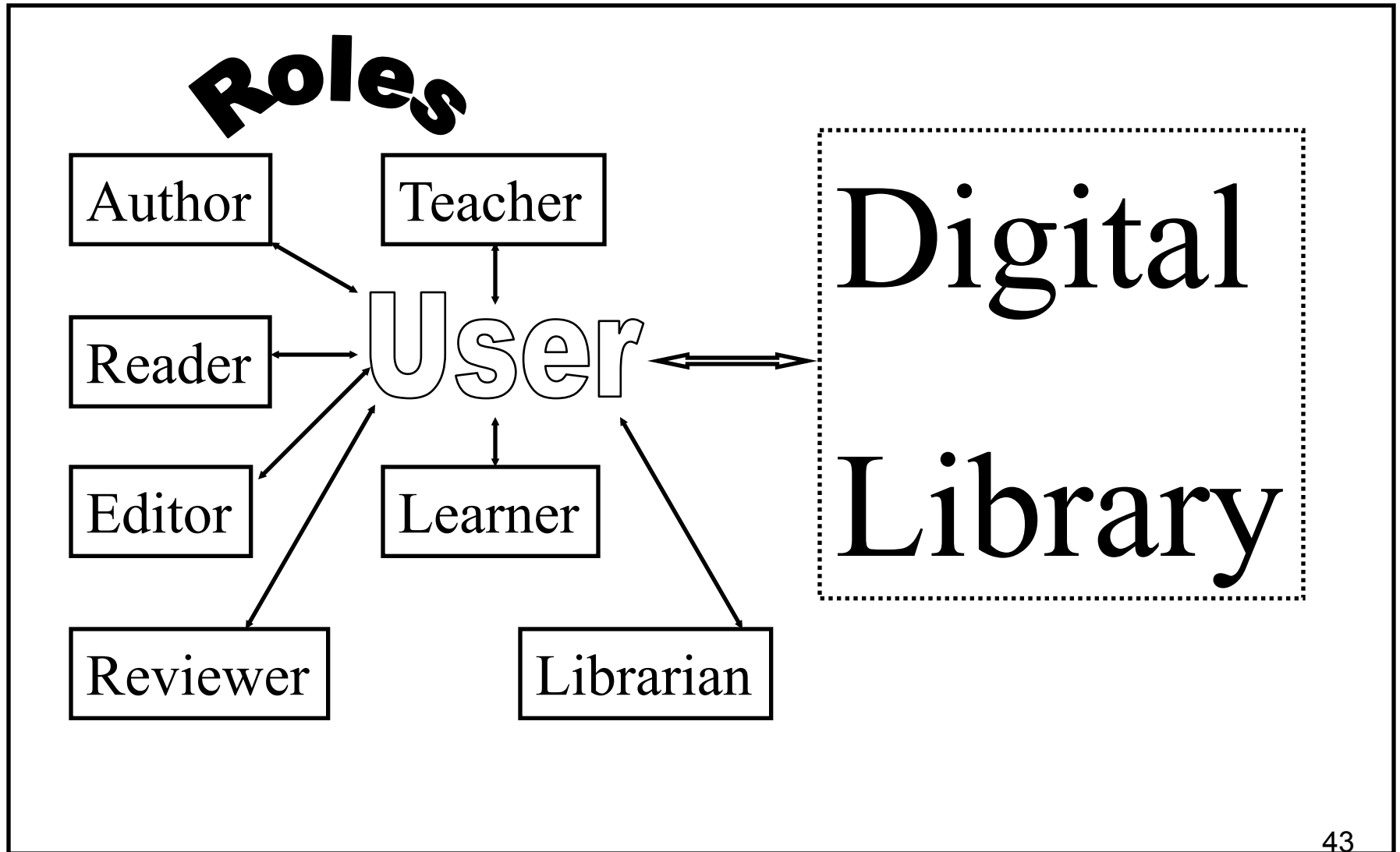
Consolidator

Library

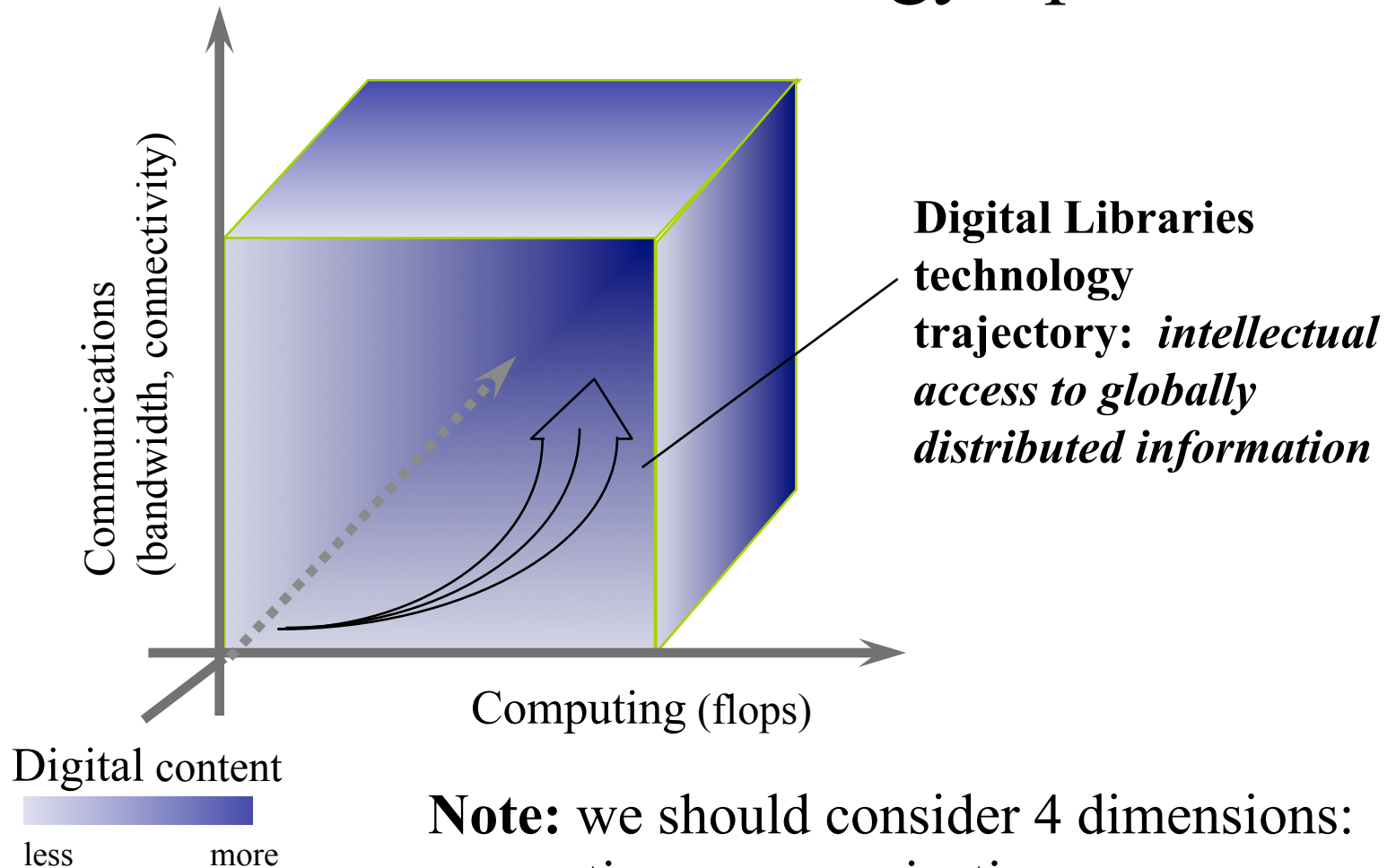
Reader



# DLs Flatten the World



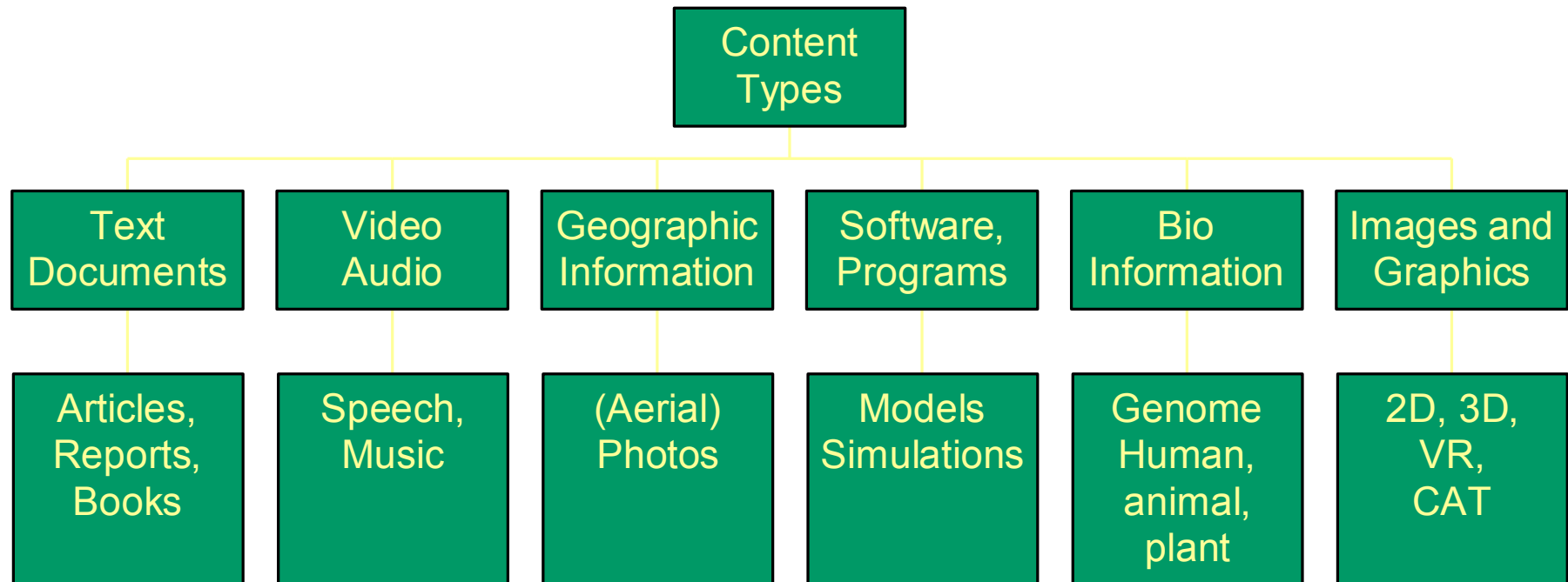
# Locating Digital Libraries in Computing and Communications Technology Space



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



# Digital Library Content



# Degree of Structure

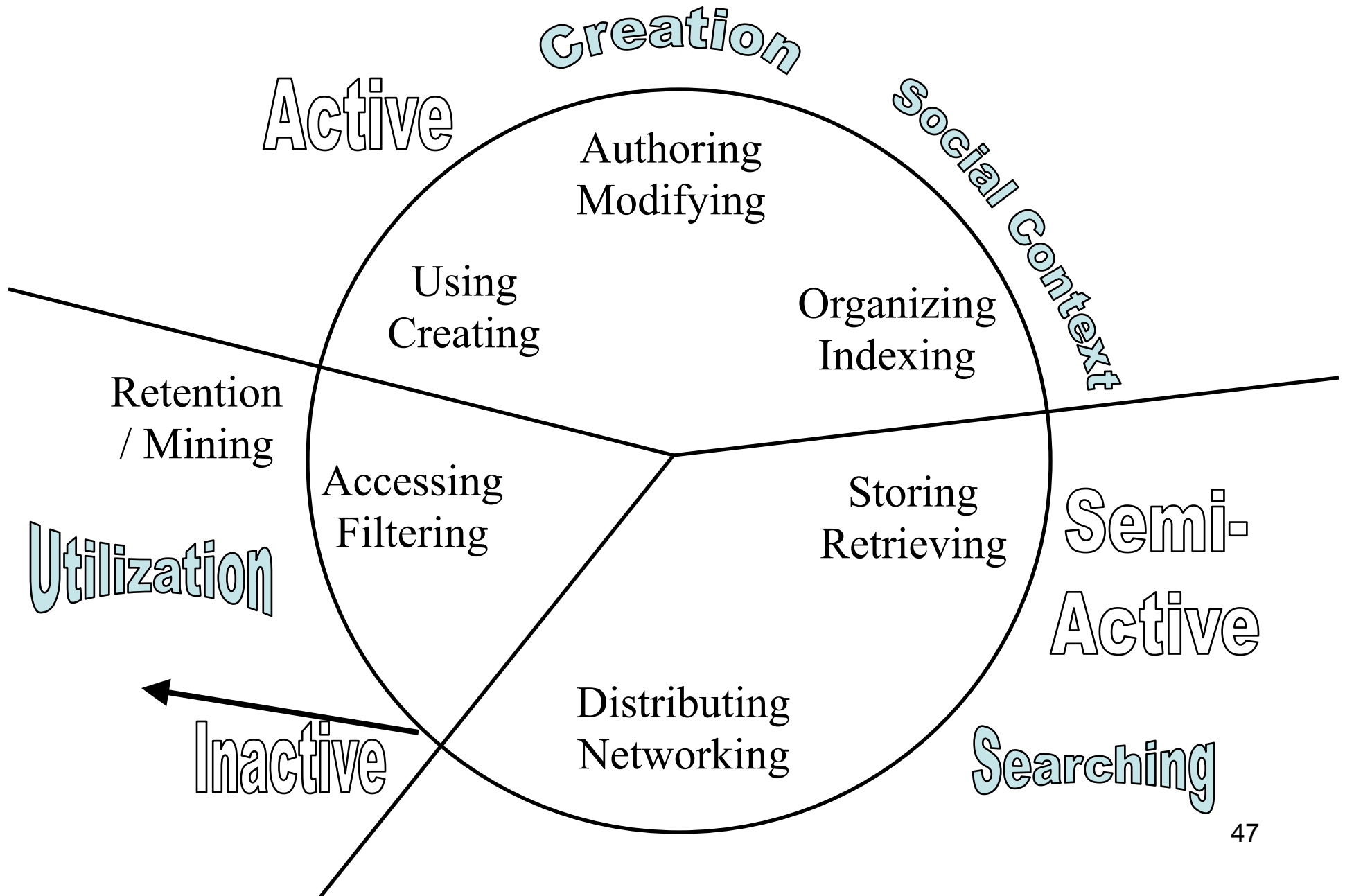
Web

DLs

DBs



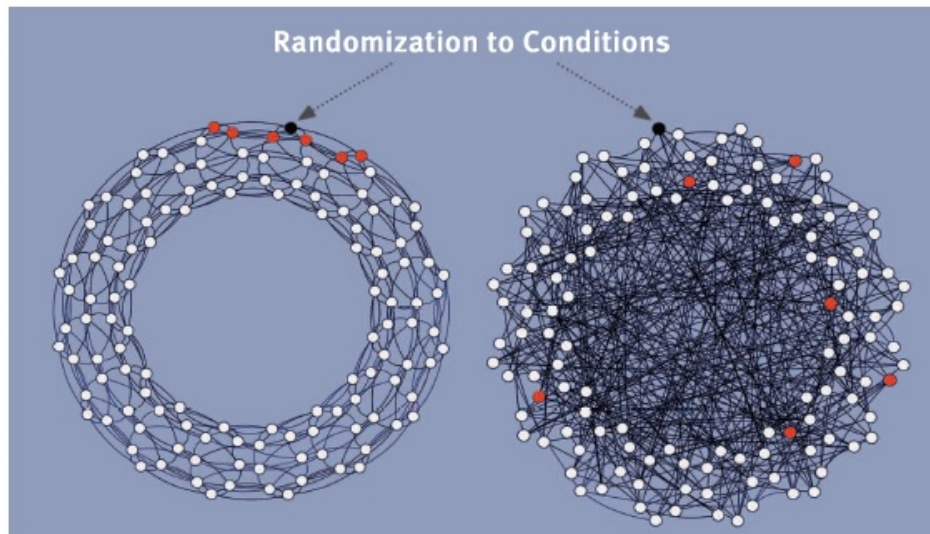
# Information Life Cycle



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

# The Social Interactome of Recovery: Social Media as Therapy Development

- NIH, \$1.7M, to VTCRI/VT, Warren Bickel PI
- National Institute of Drug Abuse
- Friendica (local controlled version of Facebook)
- 2 large experiments X 3 replicates
  - Effects of social network topology
  - Effects of controls on social network communication
- Natural language processing, network analysis, HCI



Randomization of participants to clustered-lattice and random network conditions (Adapted from Centola 2010). Each network has  $N=128$  nodes and each node has  $Z=6$  edges. In both cases the black node has 6 red nodes as neighbors.



# Integrated Digital Event Archiving & Library (IDEAL)

<http://www.eventsarchive.org>

(includes proposal and 1 year report to NSF)

# Haiti Photographs, Content Based Image Retrieval Evaluation


Facebook WebMail: Welcome Eva tool - User-orient...

← → ↻ 🏠 ☆ [http://localhost/eva/codes/view\\_images\\_feedback.php](http://localhost/eva/codes/view_images_feedback.php) ▶ 📄 🔧

## User-oriented evaluation








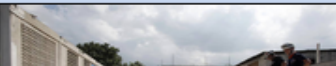

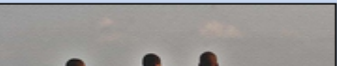
Evaluating query 1 of 1

Query image:



Name:  
HumanGrandGoave1.jpg

### RESULTS

 HumanLeogane1.jpg	 grand-goave2.jpg	 HumanGrandGoave2.jpg	 sacrecoeur-church.jpg	 petit-goave3.jpg
				

Menu 🖨️ 📁 📄 📶 🔊 🔌 📶 📶 Tue Apr 27, 11:14 AM

Eva tool - User-orient... jennifer - File Browser HaitiDisasterPhotoFin...



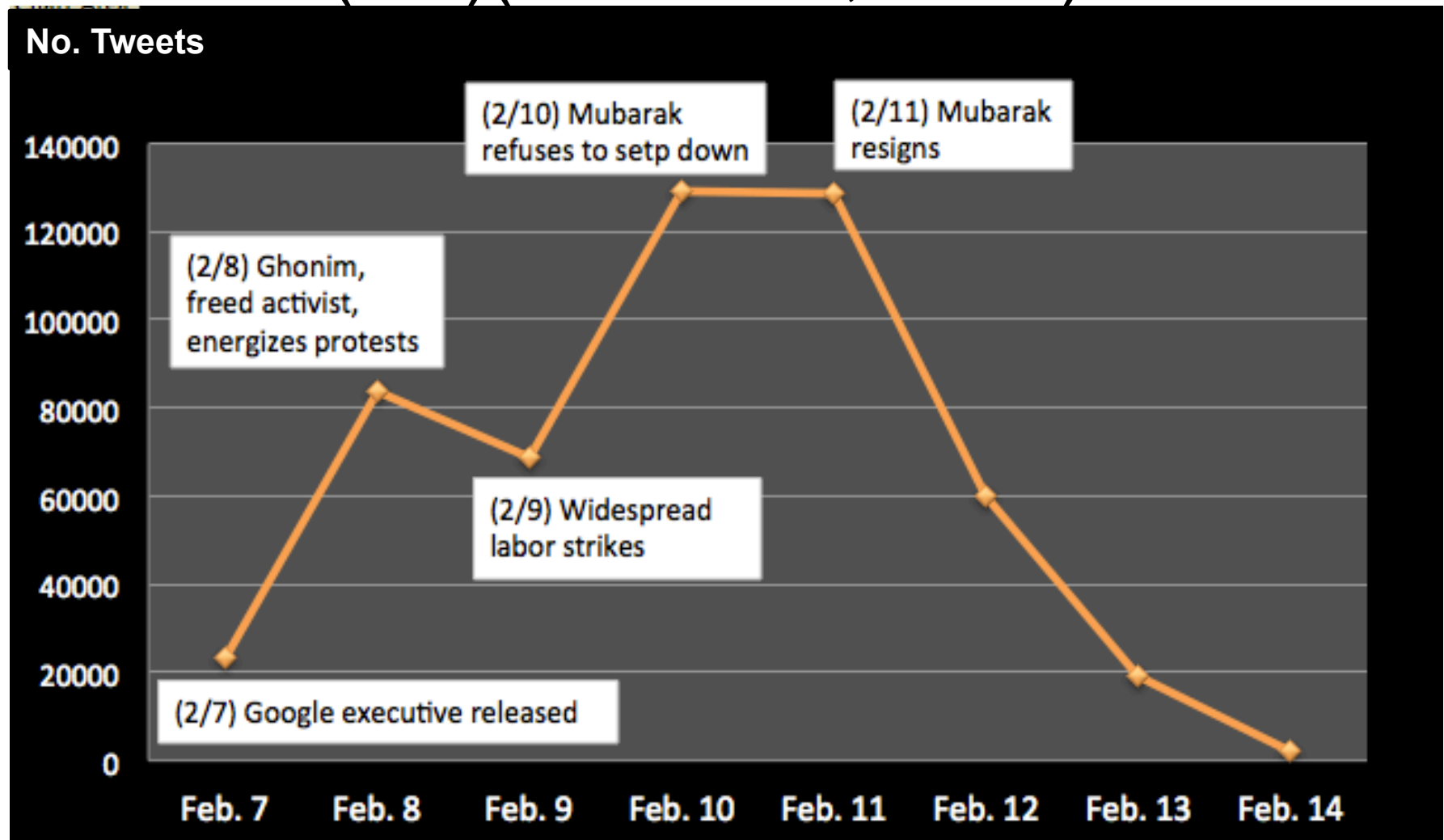
# Precursor: Crisis, Tragedy, and Recovery network (CTRnet)

Collect, analyze, and visualize disaster information with a

DL	Collect	Analyze	Visualize
Content	Web sites, images	Image similarity	Organize images by similarity
	Tweets	Content, user profiles	Patterns, frequencies
	Facebook content	Usage of social media (SM)	SM use
	Focus group interviews/surveys	Usage of SM	SM use/needs
Technology	Crawler	CBIR algorithm	CBIR visualization interface
	Online tools, scripts, APIs	NLP toolkit, SQL	Graphics
	Facebook app	Spreadsheets	
	Brainstorming tool	Brainstorming tool	

# Social Media Use in Political Crisis

## (1/2)(2/7 - 2/14, 2011)



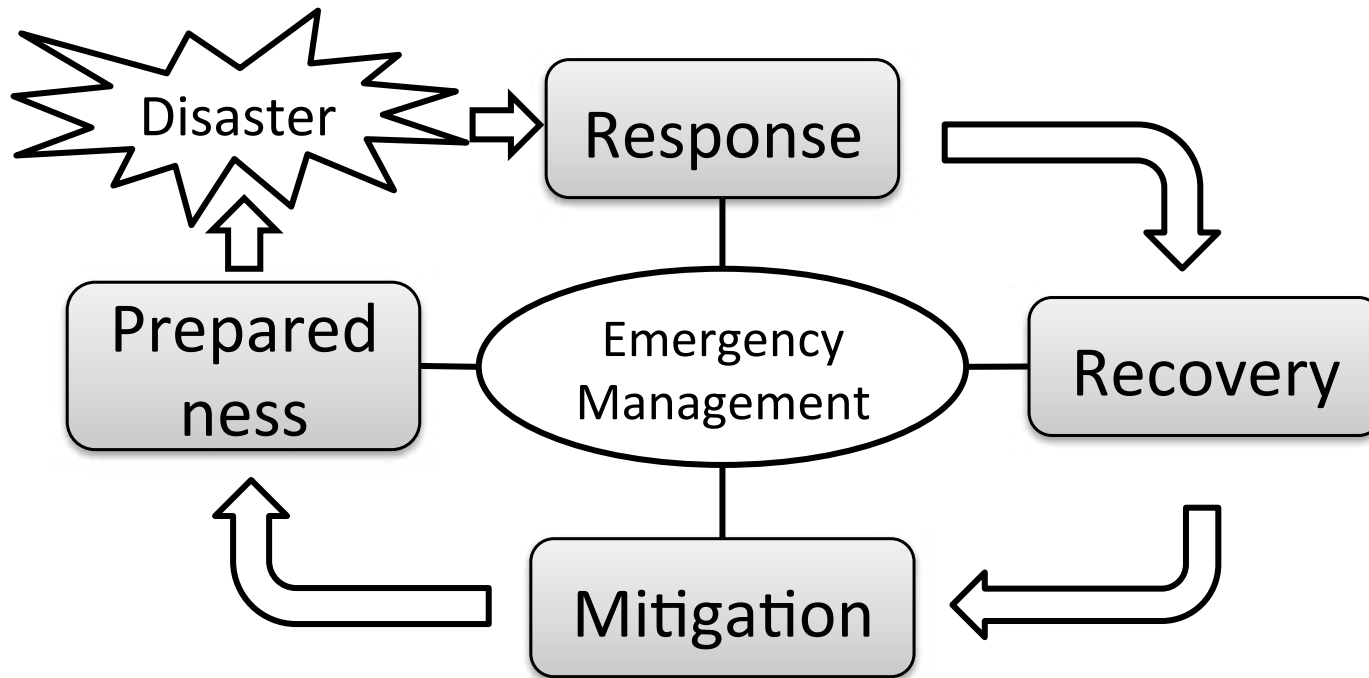
► Total 514,782 tweets

# Social Media Use in Political Crisis

## (2/2)

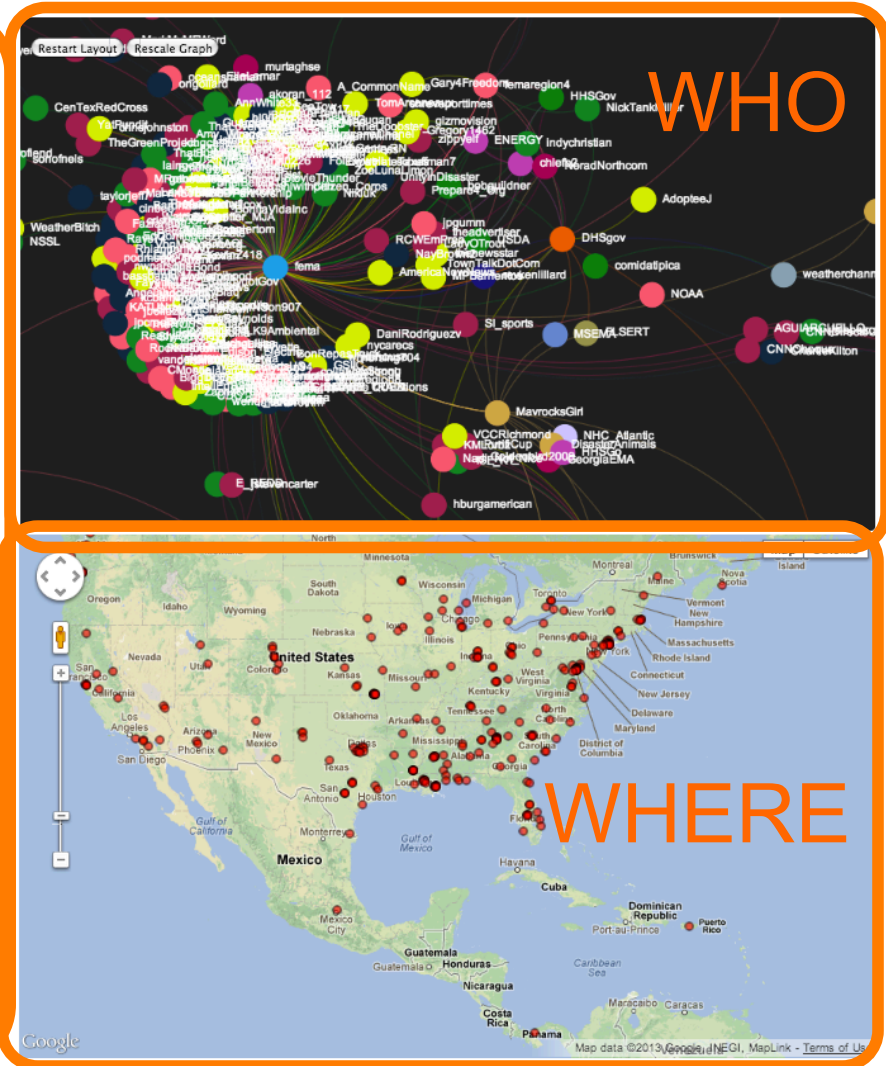
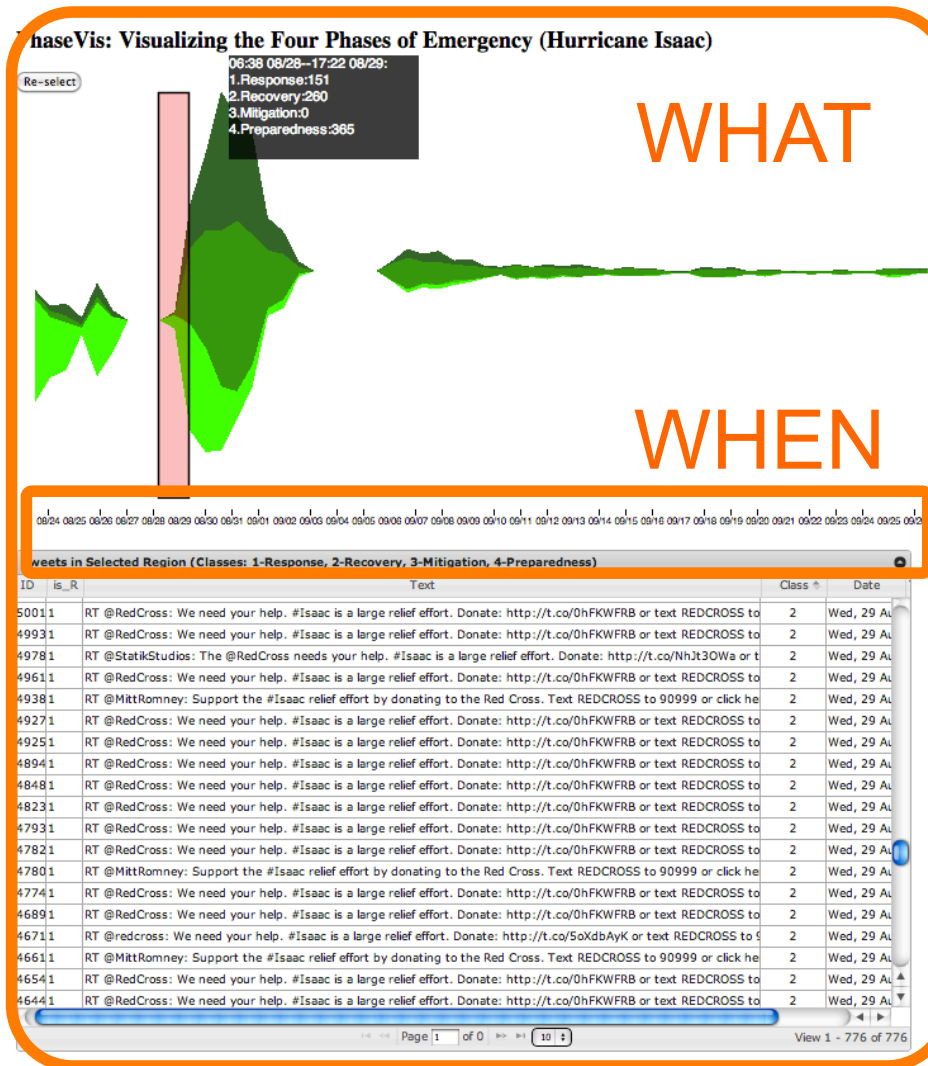
- Opinion Leadership in Egypt Uprising 2011
  - 514,782 tweets (one week around Mubarak's resignation)
  - Total 79,000 unique users
    - Presumably posting from Egypt → 4,710
    - Individuals excluding organizations → 3,675
  - Opinion leaders
    - 500-27,000 followers in top 10% (365) individuals
    - Bios: blogger/activist, writer/reporter, lawyer/executive director, social media consultant,... → 'elite' type actors
- This has led to other studies, surveys, publications

# Visualizing Emergency Phases in Tweets (ISCRAM 2013) (1/2)



Four phases of emergency management model

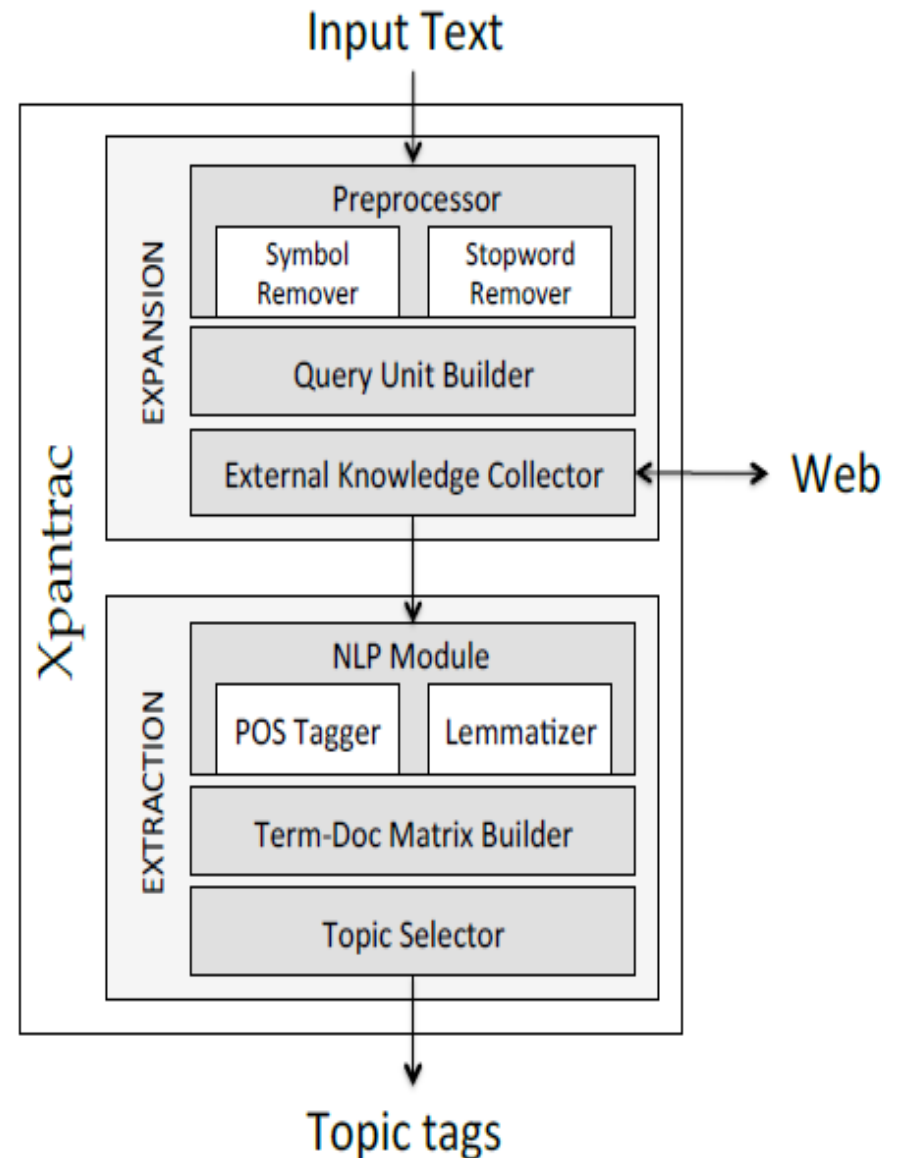
# Visualizing Emergency Phases in Tweets (2/2)



# Topic Tagging of Webpages: Xpantrac

Seungwon Yang dissertation

- Input: text file
- Build query
  - ◆ Every 5 words, 1 word overlap
- Send query to search API
  - Web search (Seungwon)
  - Wikipedia, our collection(s):  
CS4624 Spring 2014: Sloane Neidig, Samantha Johnson, David Cabrera, Erika Hoffman
- Find topics in retrieved documents
  - ◆ Frequency of words
- Select most frequent as “topics”
- Output: topics



# Water Main Break Visualization

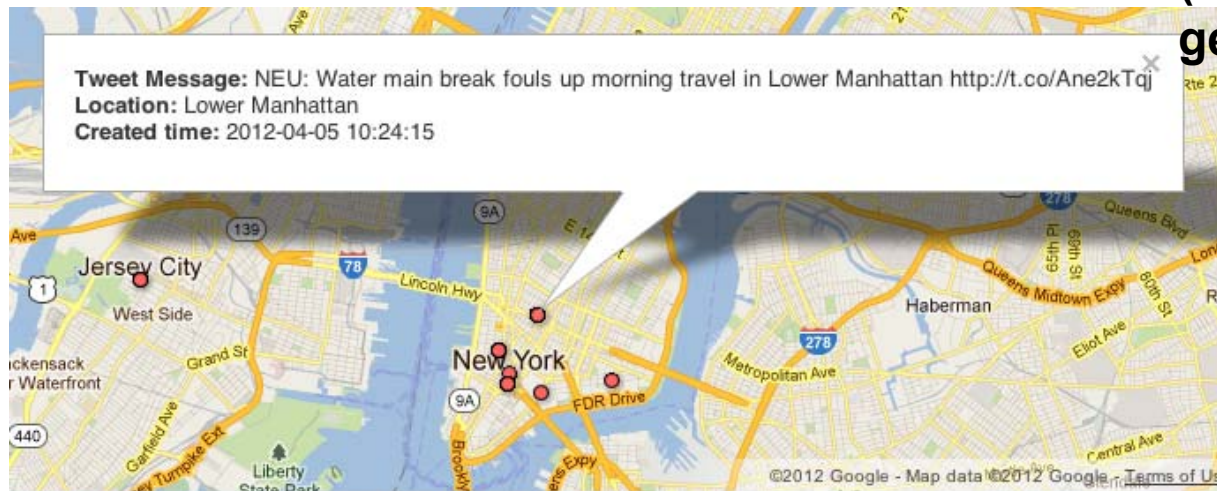
Sunshin Lee: leading to current tweet geo-location research

Dataset Keyword	Total tweets	# of tweets which have GPS information (percentage)
water main break	13,382	156 (1.17 %)
water pipe leak	967	1 (0.10 %)

← **Tweets collected with keywords**

Location information type	# of tweets
GPS data (longitude, latitude)	36 (1.08 %)
Location information extracted from text	1,473 (44.19 %)

← **Selected tweets with location information (lat/long, geonames)**



← **Event locations displayed with details**



# Web Archives

- 13 TB of IA Collections, e.g., 2013: Boko Haram attack, Boston Marathon blast, Global Emergency Overview, Texas fertilizer plant explosion

Category	No. of Archives
Accidents (plane crash, building collapse, ferry sinking)	11
Bombings	4
Earthquakes (Japan)	12
Fires	2
Floods	4
Hurricanes (Sandy), Tsunami, Cyclones, Typhoons	8
Shootings	17

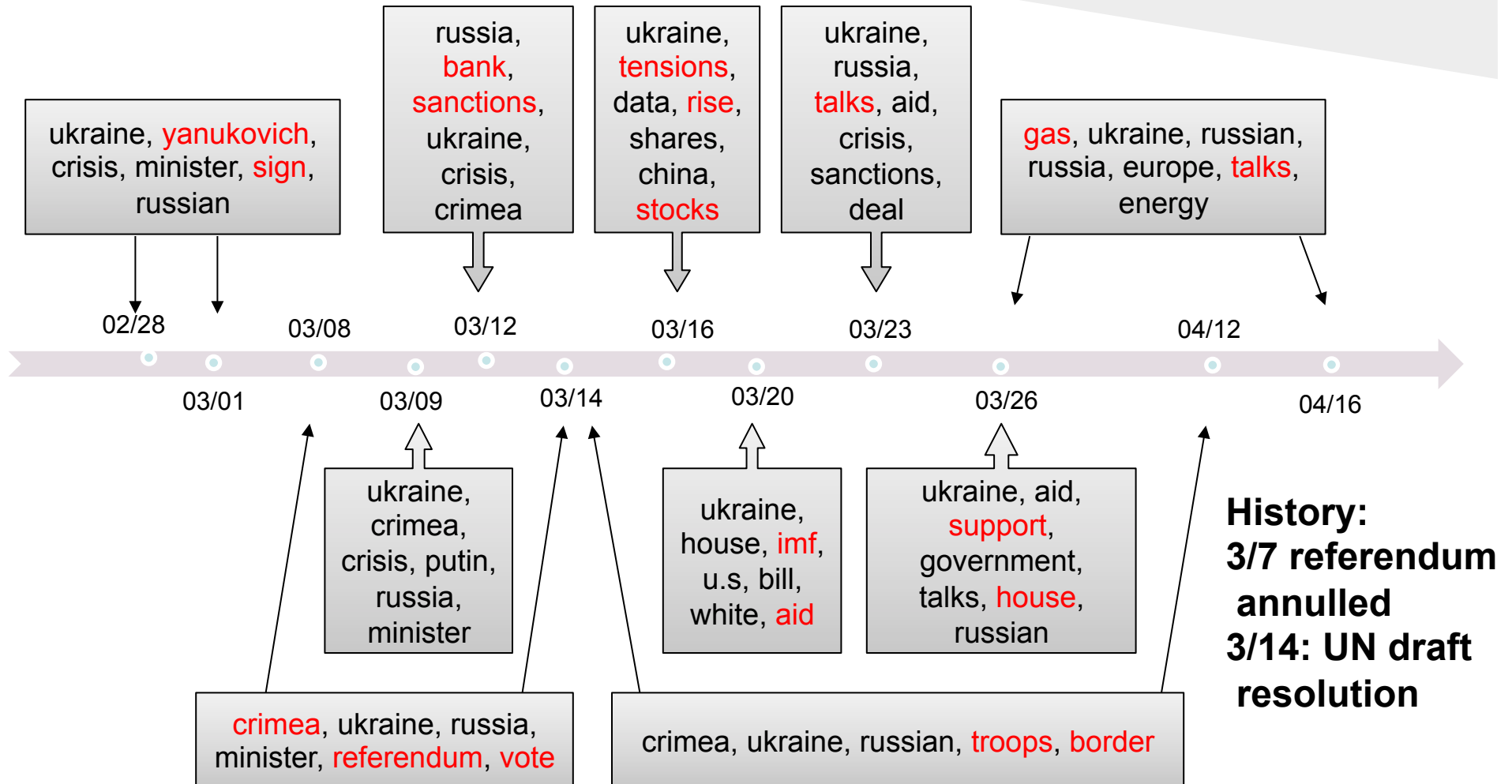
# Tweet Collections

- 442 Event-specific and general collections
- Total of 915 million tweets, from streaming API, using hashtags and keywords

Category	No. of collections			
Accident (transportation)	33			
Bombing	8			
Community	10	servers	# of collections	# of tweets
Earthquake	18	cinnamon	54	46,373,297
Fire	6	lancelot	44	344,519,093
Flood	11	spare05	272	249,318,681
General (including health)	67	spare07	44	108,801,519
Hurricane, Tsunami	39	virginia	28	166,113,764
Political (Middle East, Iran)	40	Total	442	915,126,354
Shooting	29			

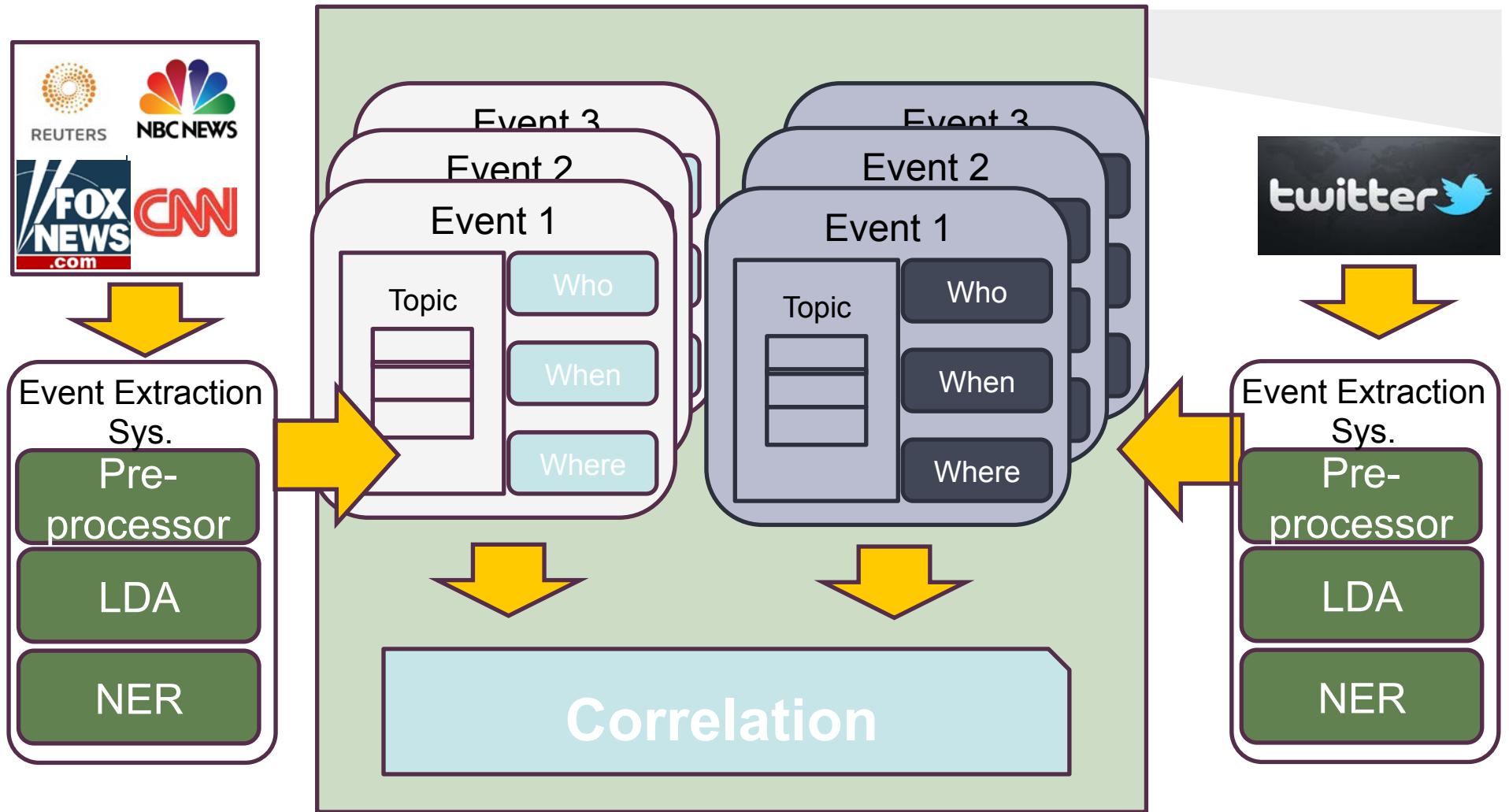
# Extracted News Events on a Time Line

CS6604 Spring 2014: Tianyu Geng, Wei Huang, Ji Wang, Xuan Zhang



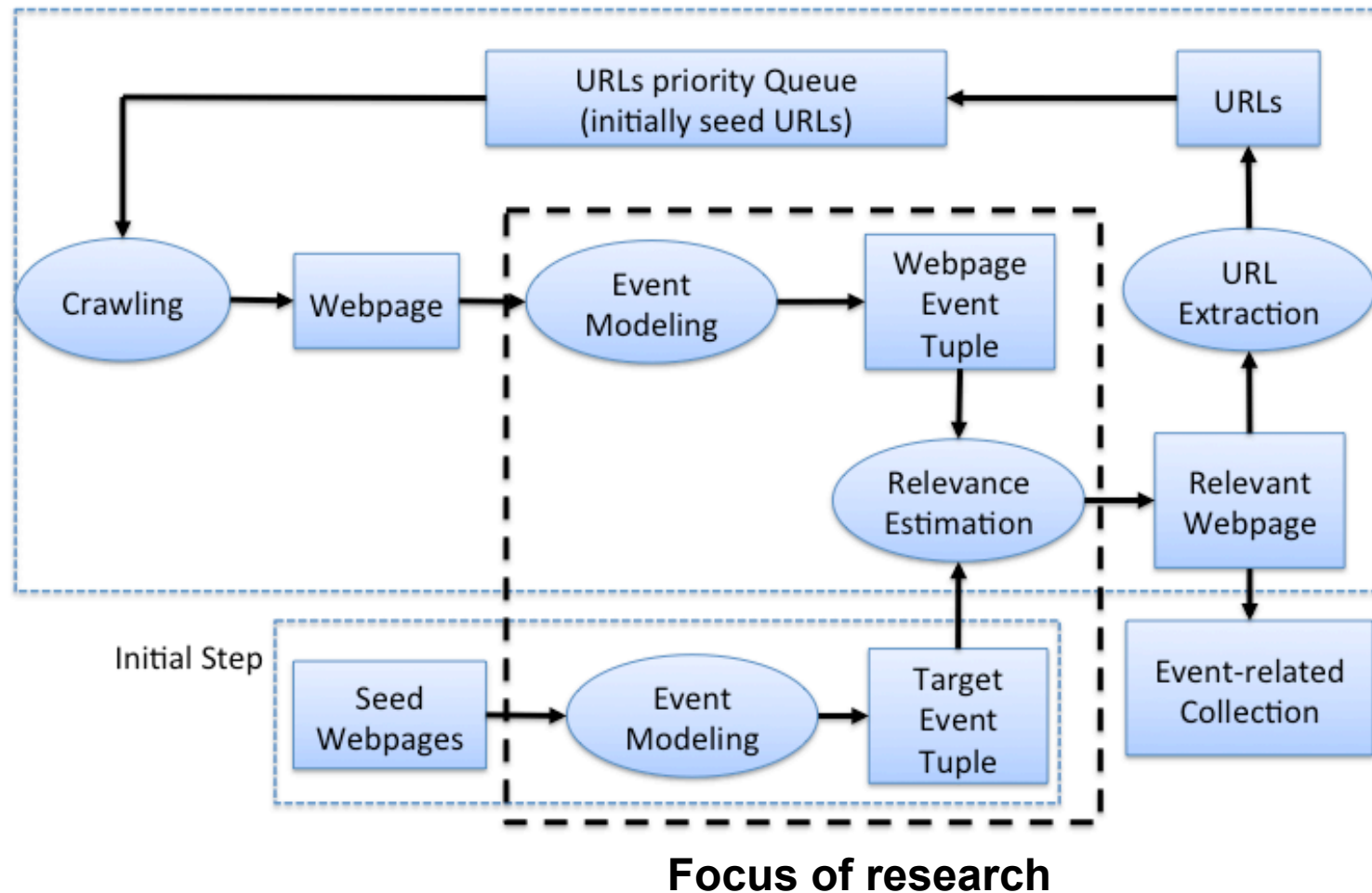
# News-Tweet Architecture

CS6604 Spring 2014: Tianyu Geng, Wei Huang, Ji Wang, Xuan Zhang



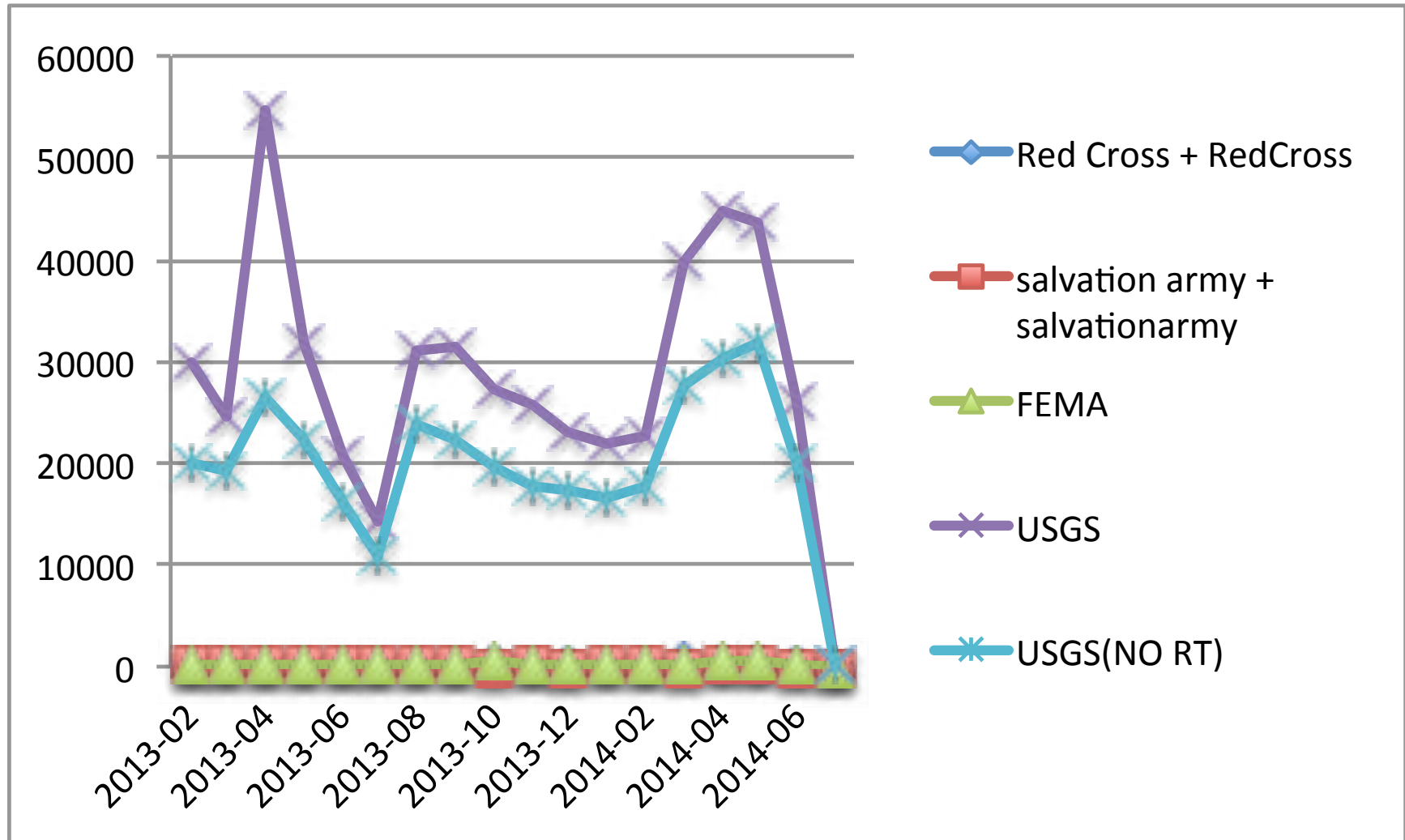
# Event Focused Crawler

Mohamed Magdy



# Earthquake stats by keyword

from So Hyun Jo Ugrad Research 10/27/2014



# Qatar

## Project Objectives/Aims

- A. Research and prototype digital library systems and infrastructure for Qatar, focusing initially on Qatari information related to government and scholarly activities.

*Leverage the crawling engine from Penn State's SeerSuite software infrastructure, and extend it beyond its current focus on English to support Arabic-English collections, and to cover a broad range of scholarly disciplines, and all types of government information.*

- B. Research and build the digital library community in Qatar, supporting digital library use, services, collection development, tailored systems, and advancing toward a Knowledge Society.

*Study scholarly activities, and engage in community building in Qatar, so DLs can be tailored to specific domains and to the unique needs of Qatar. Through workshops, a consulting center at the proposed Institute, and collaborative efforts with libraries and museums in Qatar, we will identify particular needs and uses, and tailor collections, systems, and services, to lead toward the Qatari Knowledge Society.*



# 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**)

# **DL Book 1:**

## **Theoretical Foundations**

**(Fox, Goncalves, Shen)**

- ~180 pages
- 1) Introduction
- 2) Exploration (searching, browsing, visualizing)
- A) Mathematical Preliminaries
- B) Minimal DL
- C) Archaeological DLs
- D) 5S Results: Lemmas, Proofs, 5SSuite
- E) Glossary
- Bibliography (263 references)

# **DL Book 2: Key Issues Regarding Digital Libraries**

(Shen, Goncalves, Fox)

- ~110 pages
- 1) Evaluation
  - Intro, Related Work, Formalization
  - Digital Objects, Metadata Specs & Format, Collections, Metadata Catalogs, Repositories, Services, Case Study: 5SQual
- 2) Integration
  - Intro, Related Work
  - Case Study: ETANA-DL

# **DL Book 3: Technologies**

## **(Fox, Torres)**

- ~180 pages
- 1) Complex Objects
- 2) Subdocuments
- 3) Ontologies
- 4) Classification
- 5) Text Extraction
- 6) Security

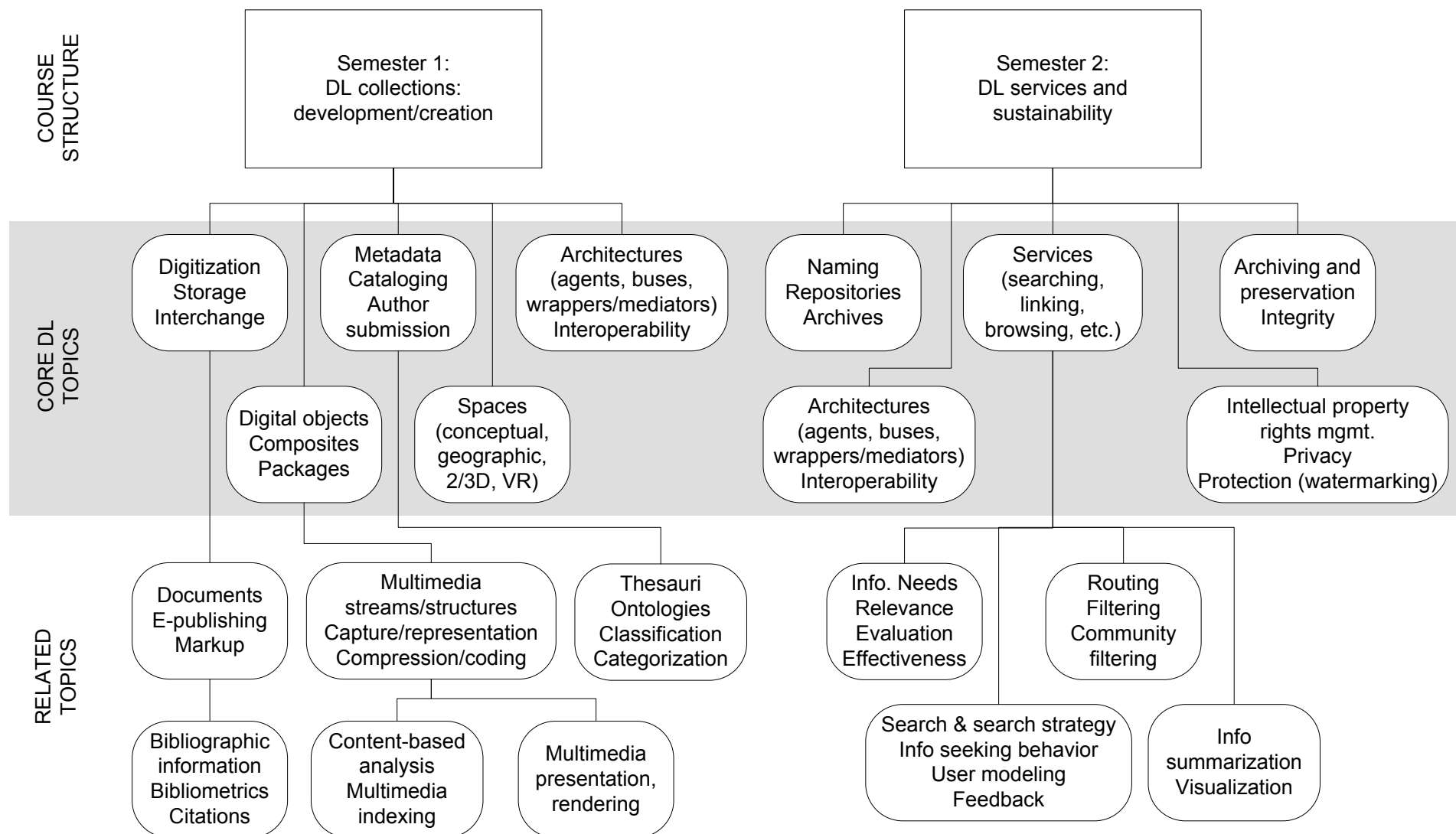
# **DL Book 4: Applications**

**(Fox, Leidig)**

- ~160 pages
- 1) Content-Based Image Retrieval (CBIR)
- 2) Education
- 3) Social Networks
- 4) Bioinformatics, Scientific, & Simulation DLs
- 5) Geospatial Information

# Introduction

## DL Curriculum Framework



# Wikiversity Modules

- Table 1: Core DL Curriculum (2 slides)
  - Table 2: Information Retrieval Packages
  - Table 3: LucidWorks Big Data Software
  - Table 4: Multimedia Software
- 
- [http://en.wikiversity.org/wiki/  
Curriculum\\_on\\_Digital\\_Libraries](http://en.wikiversity.org/wiki/Curriculum_on_Digital_Libraries)



iPad

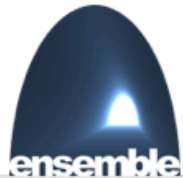
1:17 AM

100%

Ensemble: A Digital Library for Computing Education I

www.computingportal.org/

Google



Connecting Computing Educators



ENSEMBLE HOME

LOGIN

SEARCH

ABOUT

CONTACT



Browse  
our Collections



Explore  
our Communities



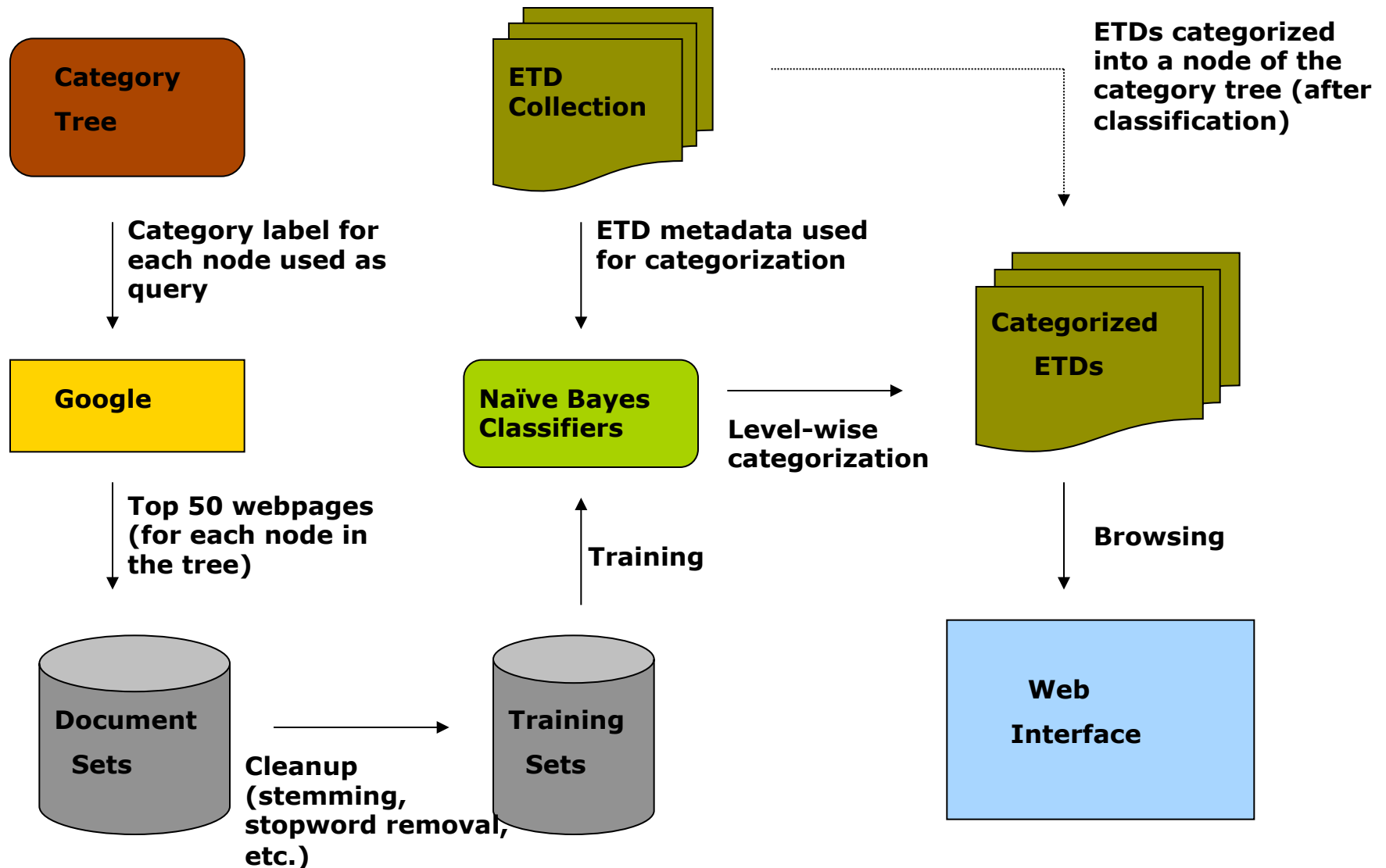
Try  
our Tools

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

**www.computingportal.org**

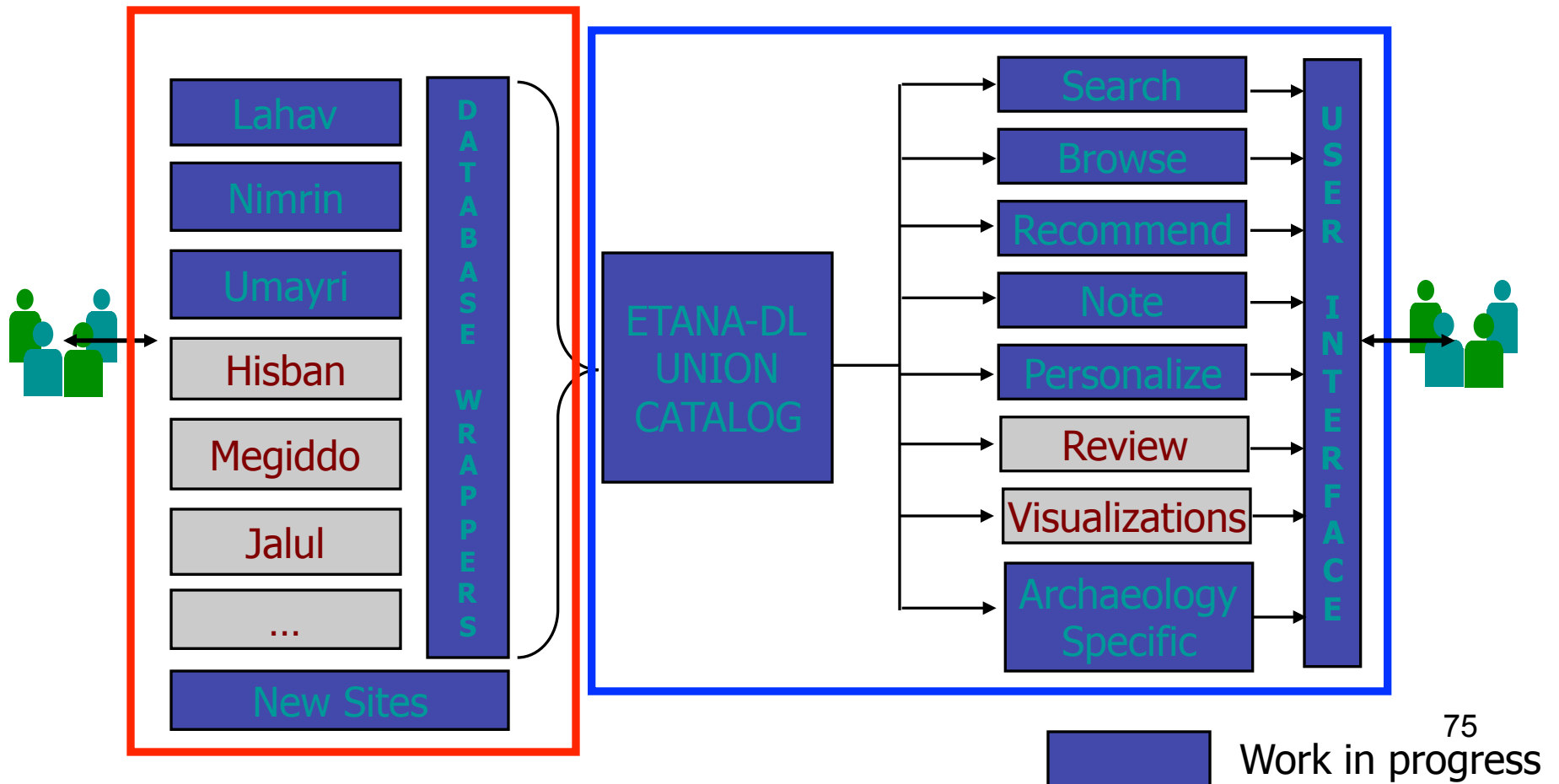
# ETD Classification:

# Algorithm Pipeline



# ETANA-DL Architecture

## DigBase and DigKit



# Area Screen

Print SearchByNumber ToEdit Exit StandBy

AREA:2000/K

ID	Map
1	K-4

Local Stratigraphy:

#	Stratum	Period
1	K1	IAII
2	K2	IAII
3	K3	IAII
4	K4	IAI-IAIIA
5	K5	IAI

Stratigraphic report:  
Area K: in . . .

K-4

TYPE: ☒ AR ☐ VS ☐ LB ☐ FI Show Finds

structure part  
transport & communica  
vessel  
waste

tool/weapon  
tool/weapon: metallurgy

Search by:

SUBTYPE Total: 119

Show locus

View ceramics map

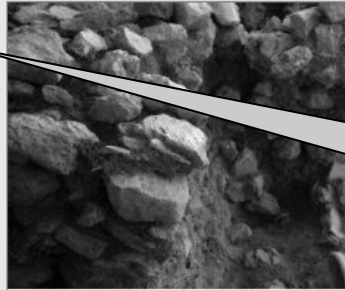
ACCESS: VIEW F2: DO IT PgUp/PgDn: Pages Ctrl + Space: LookUp F5/F6: Cards

# Locus Screen: Pictures

**MICADO** Search ToEdit

LOCUS: **00/K/025** PERIOD:

VIEW ALL



# **1** from 3

File: MG3763S Picture ID: 3763

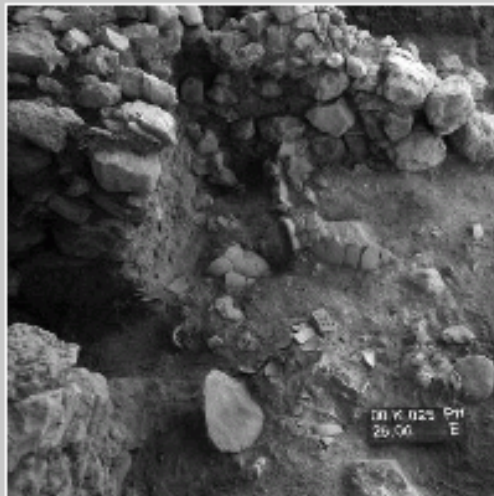
View all

LOCUS: 2000/K/025

PERIOD:

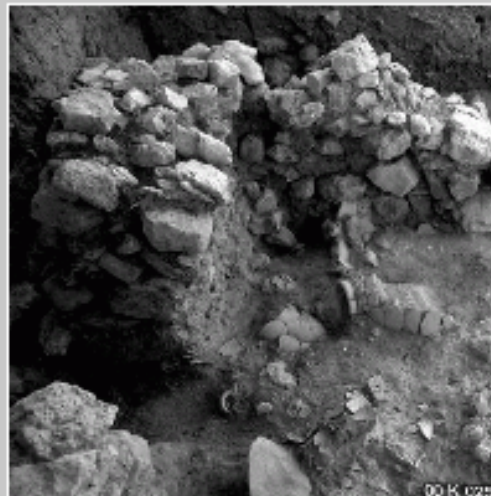
SQUARES: P/11

:



1

MG3763S



2

MG3766S



3

MG3769S

Post Message

Items recommended  
on the basis of  
similar interests

Other users also showed interest in the following items

[Umayri](#) Bone ID 1268 [FIELD F](#) [SQUARE 6L98](#) [LOCUS 077](#) [PAIL 0](#) COUNT 2

LOCUS PERIOD PER AGE -

BONE - ANIMAL Cattle

COMMENTS -

[\[View complete record\]](#) [\[Add item to Items of Interest\]](#) [\[Mark Item\]](#)

[Lahav](#) Figurine ID 2029

*Horse from Horse and Rider*

[Field IV](#) [Area K7](#) [Locus K7004.1](#) [Basket IV.K7.13](#)

**Dimensions** 6.94 cm. h; 3.28 cm. w; 2.84 cm. d

**Dig** Israel Antiquities Authority Accession No. 1998-811.

[\[View complete record\]](#) [\[Add item to Items of Interest\]](#) [\[Mark Item\]](#)



[Lahav](#) Figurine ID 2032

*Horse from Horse and Rider*

[Field IV](#) [Area K9](#) [Locus K7004.1](#) [Basket IV.K7.13](#)

**Dimensions** 6.03 cm. hlong; 3.22 cm. w; 3.32 cm. d

**Dig** Israel Antiquities Authority Accession No. 1998-829.

[\[View complete record\]](#) [\[Add item to Items of Interest\]](#) [\[Mark Item\]](#)

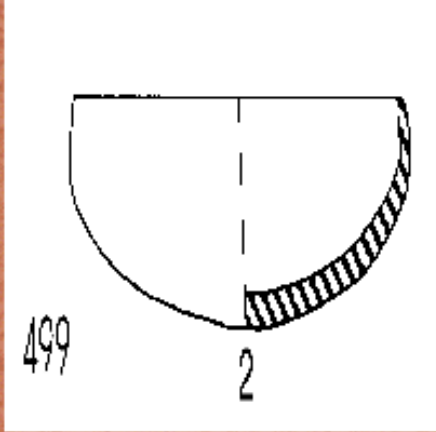


## Recommendations



# Visual Browsing

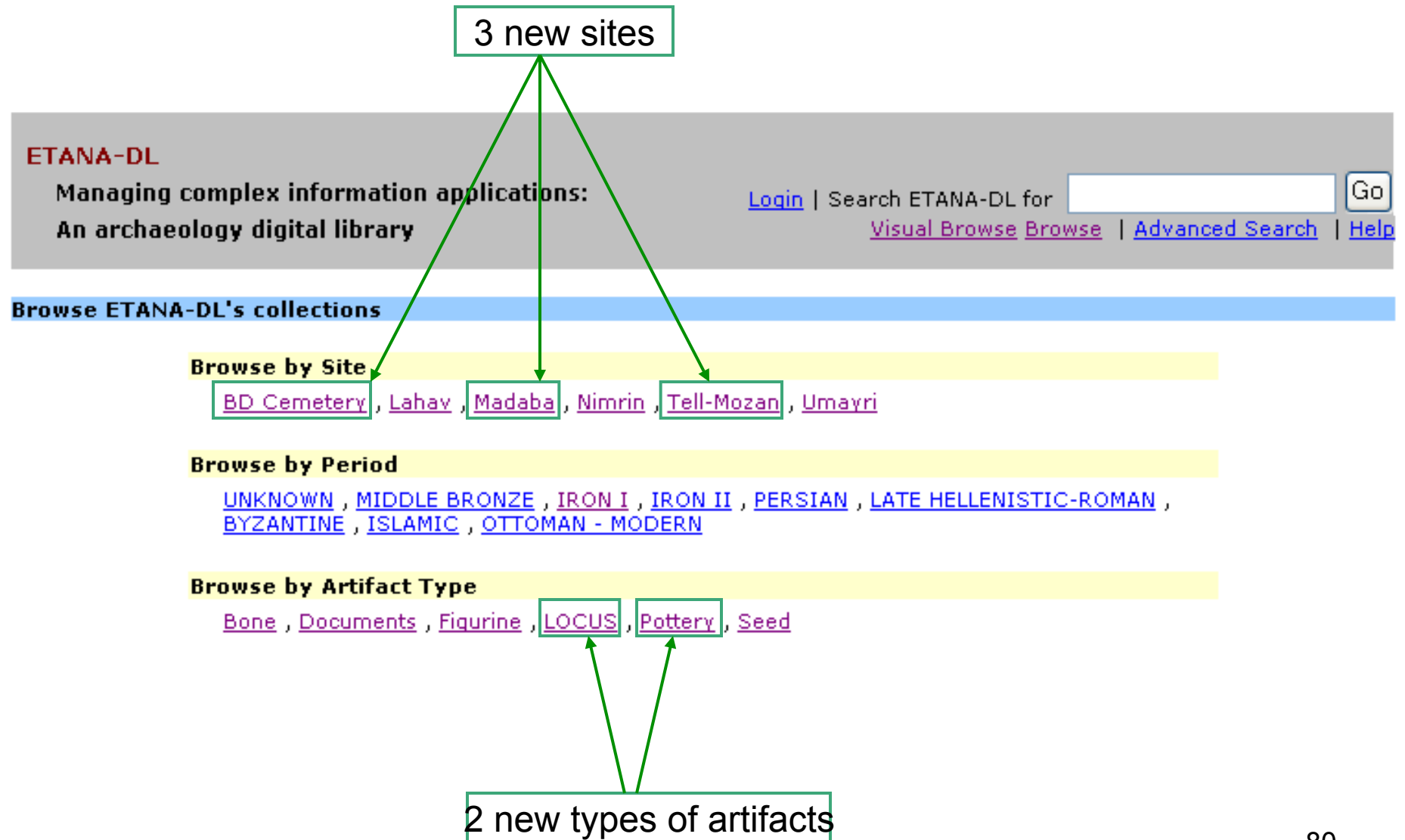
## Bab edh-Dhra' Cemetery

BD Cemetery Pottery Record			
Pottery Image Quick View			
Registration Number	Registered	Lapp	none
	Unregistered	EDSP	499
Pottery Type	Chronological Sequence	EB II	
	Basic Category	Small bowls and Saucers	
	Rim Treatment	unavailable	
	Handle Type	unavailable	
Location	Tomb Area	A	
	Tomb Number	056	
	Direction	No Direction	
	Vessel Number	025	
	Neck Join	unavailable	

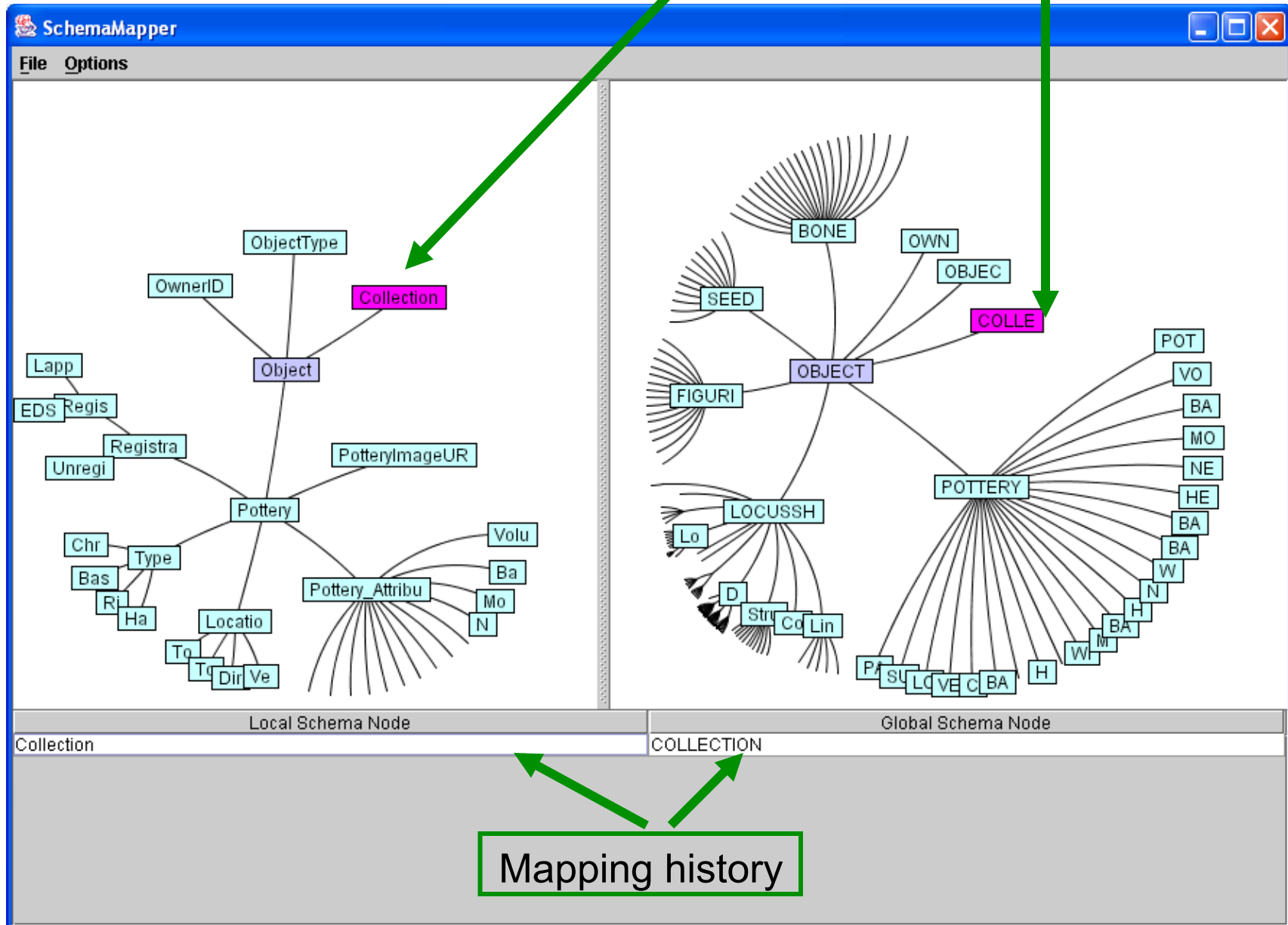
Pottery # 25

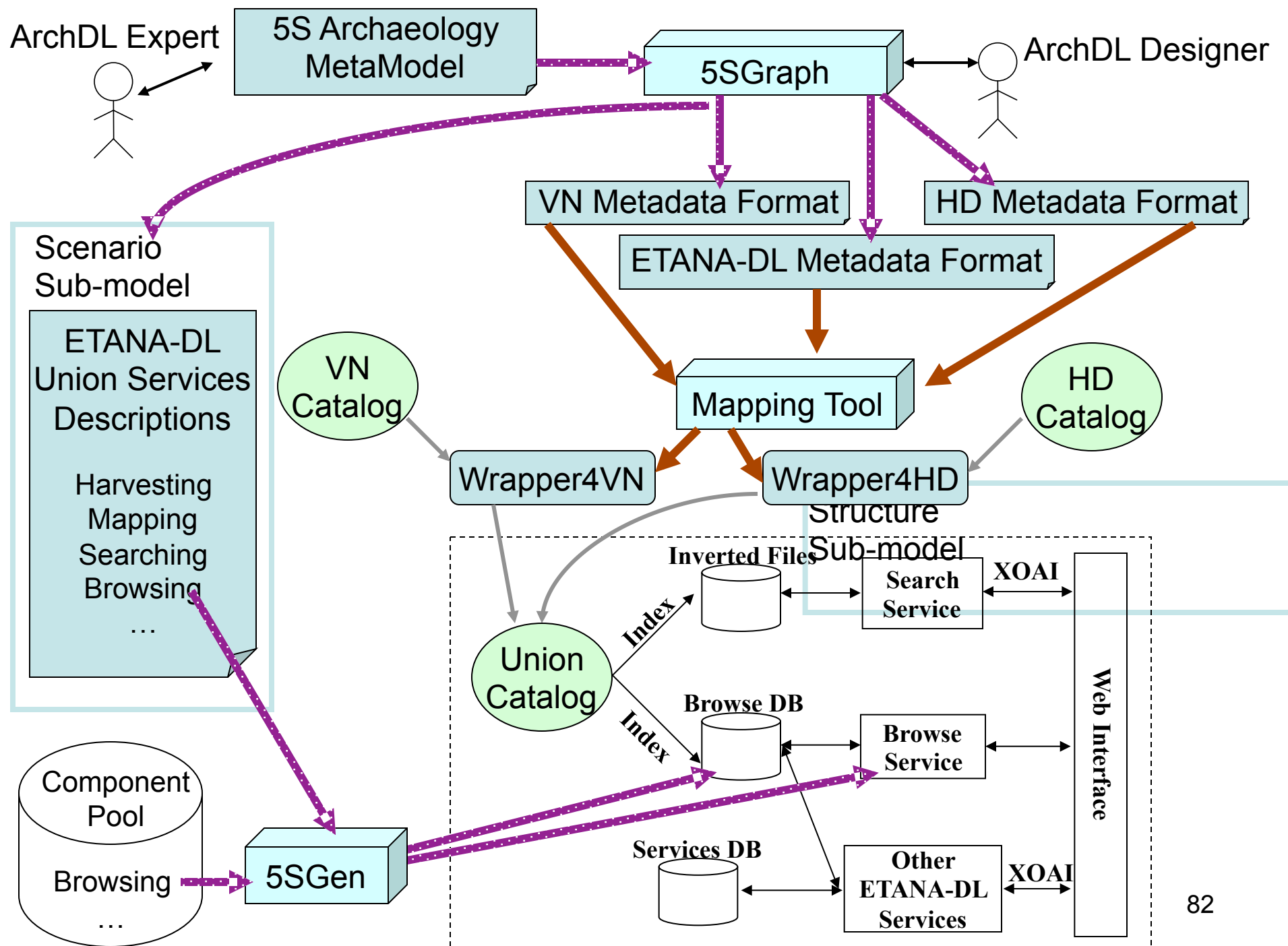


# ETANA-DL Multi-dimensional Browsing

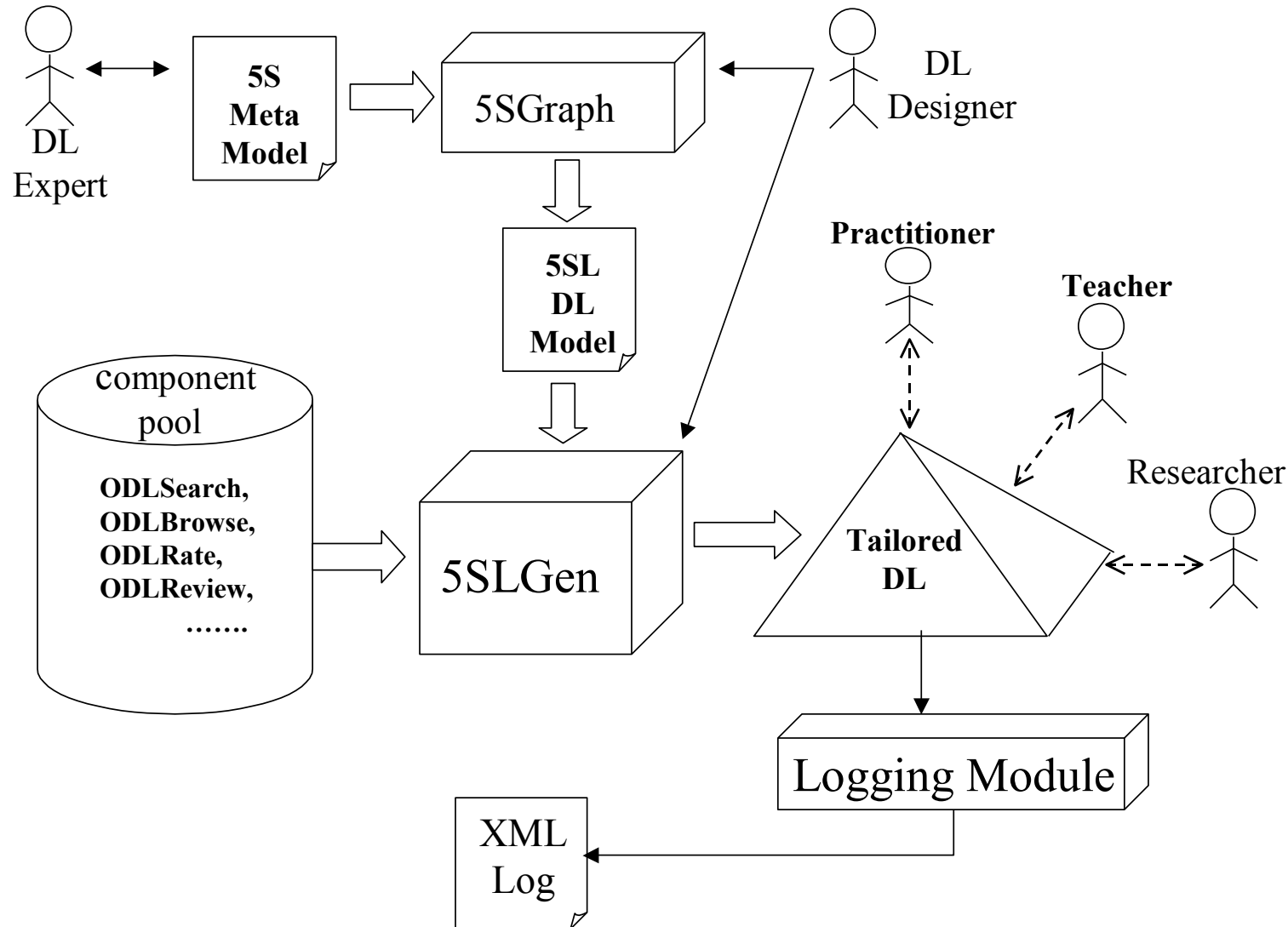


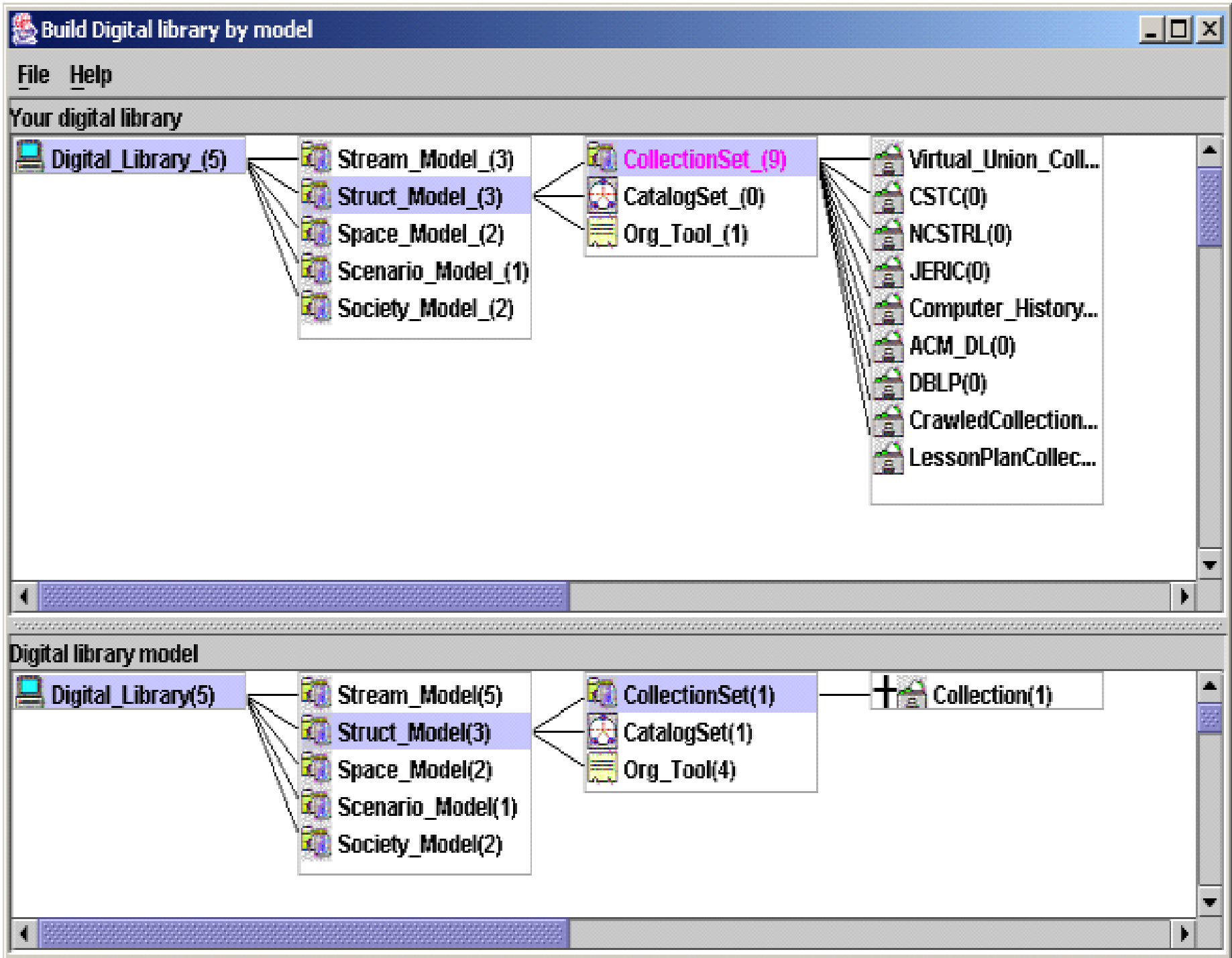
Mapping confirmation





# 5SSuite: Tools/Applications

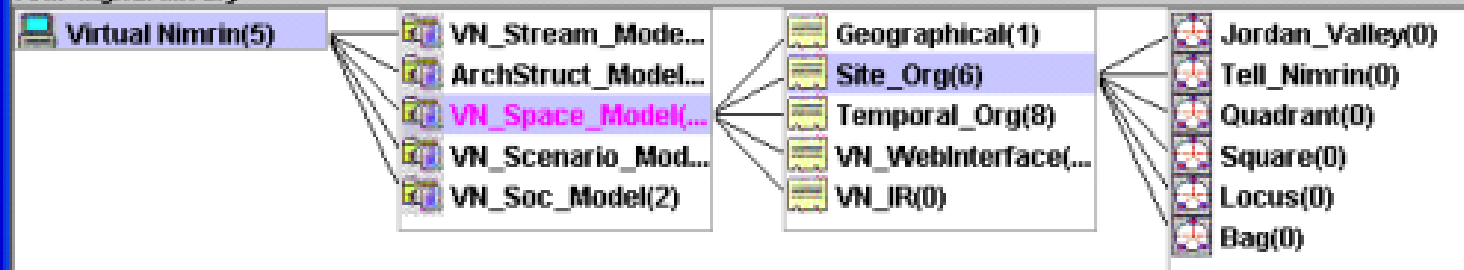




# Build Archaeological Digital Library by Model

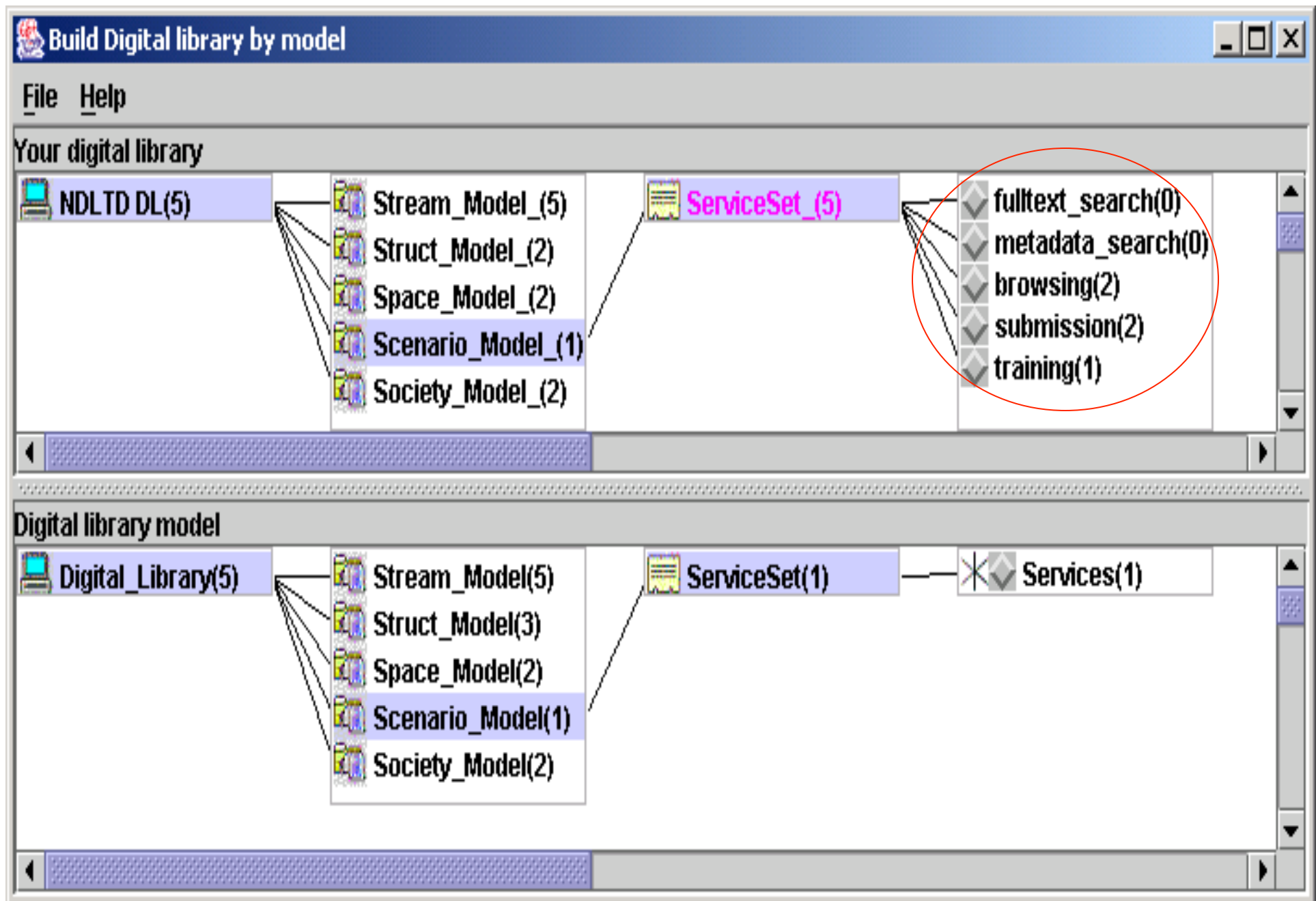
File Help

## Your digital library

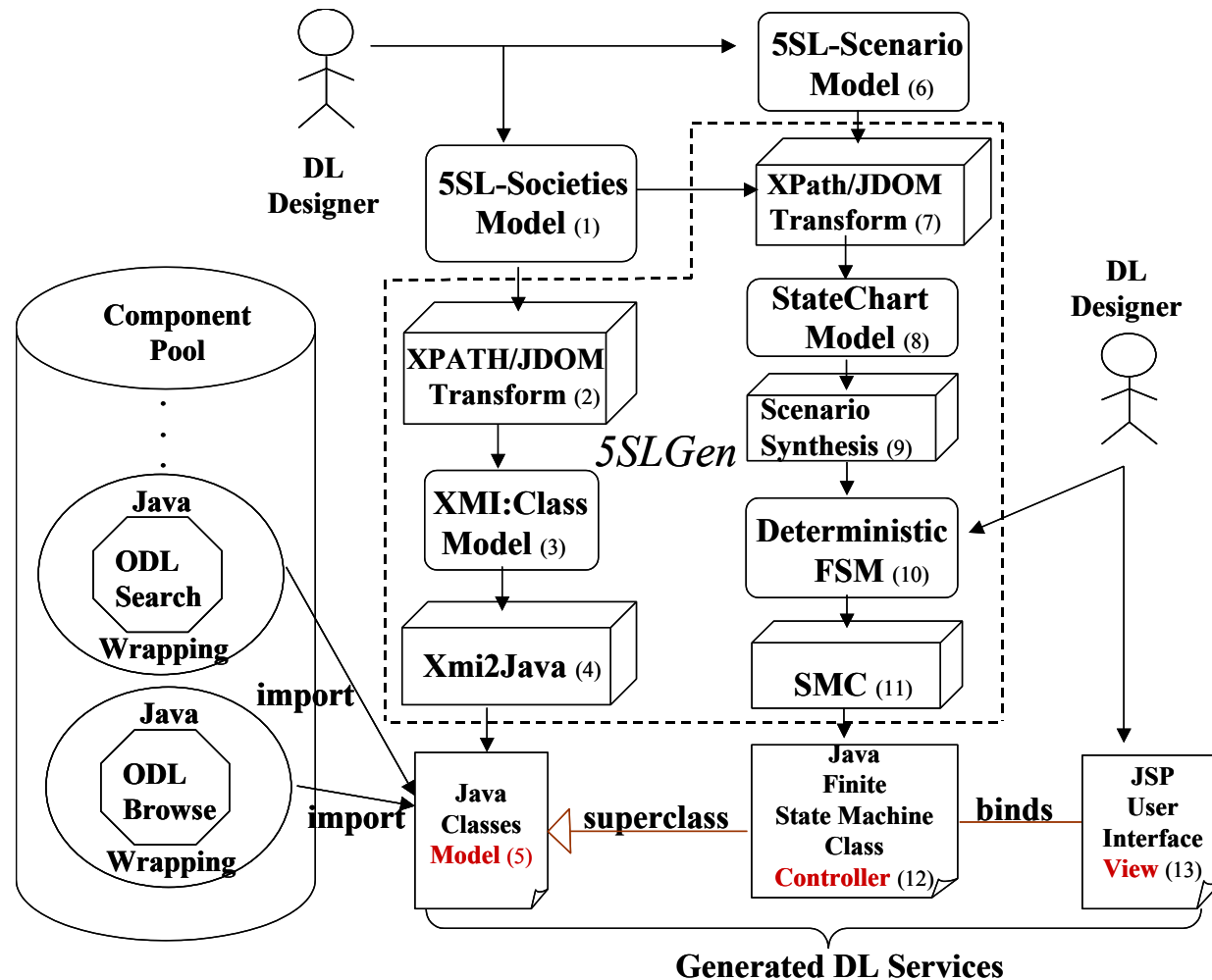


## Digital library model





# 5SLGen – Version 2: ODL, Services, Scenarios





Questions?  
Discussion?

Thank You!

Ed Fox

2160G Torgersen Hall

[fox@vt.edu](mailto:fox@vt.edu)

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