CNS Projects: Visual Analytics

Katy Börner
Victor H. Yngve Distinguished Professor of ISE and Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing and
Indiana University Network Science Institute
Indiana University, USA

Guest Lecture in E500

Sept 10, 2018

Mapping Longitudinal Scientific Progress, Collaboration and Impact of the Alzheimer's Disease Neuroimaging Initiative (ADNI)

Xiaohui Yao[1,3,4], Jingwen Yan[1,3,4], Michael Ginda[2,3], Katy Börner[2,3], Andrew J Saykin[1,3], Li Shen[1,3,4], for the Alzheimer's Disease Neuroimaging Initiative*

[1] Center for Neuroimaging, Indiana University School of Medicine

[2] Cyberinfrastructure for Network Science Center, Indiana University Bloomington

[3] Indiana University Network Science Institute

[4] School of Informatics and Computing, Indiana University

*Data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in data analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

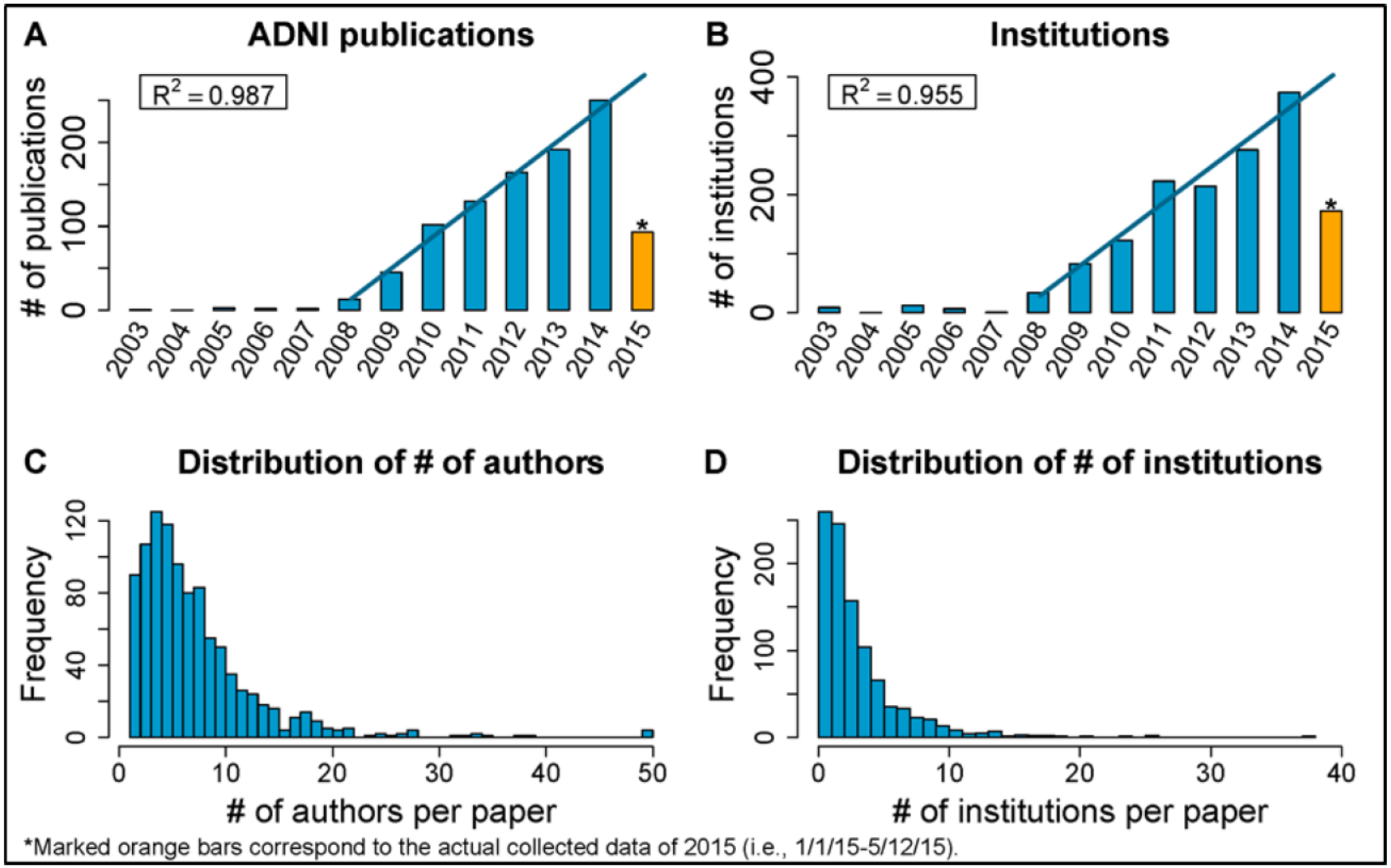


Figure 1. Statistics for ADNI publications between 01/01/2003 and 05/12/2015. (A) Growth of ADNI publications on the year-by-year basis; line indicates a linear regression prediction for the 2015 number using data from 2008 to 2014. (B) Growth of institutions involved in ADNI publications; line indicates a linear regression prediction for the 2015 number using data from 2008 to 2014. (C) Distribution of number of authors per paper. (D) Distribution of number of institutions per paper.

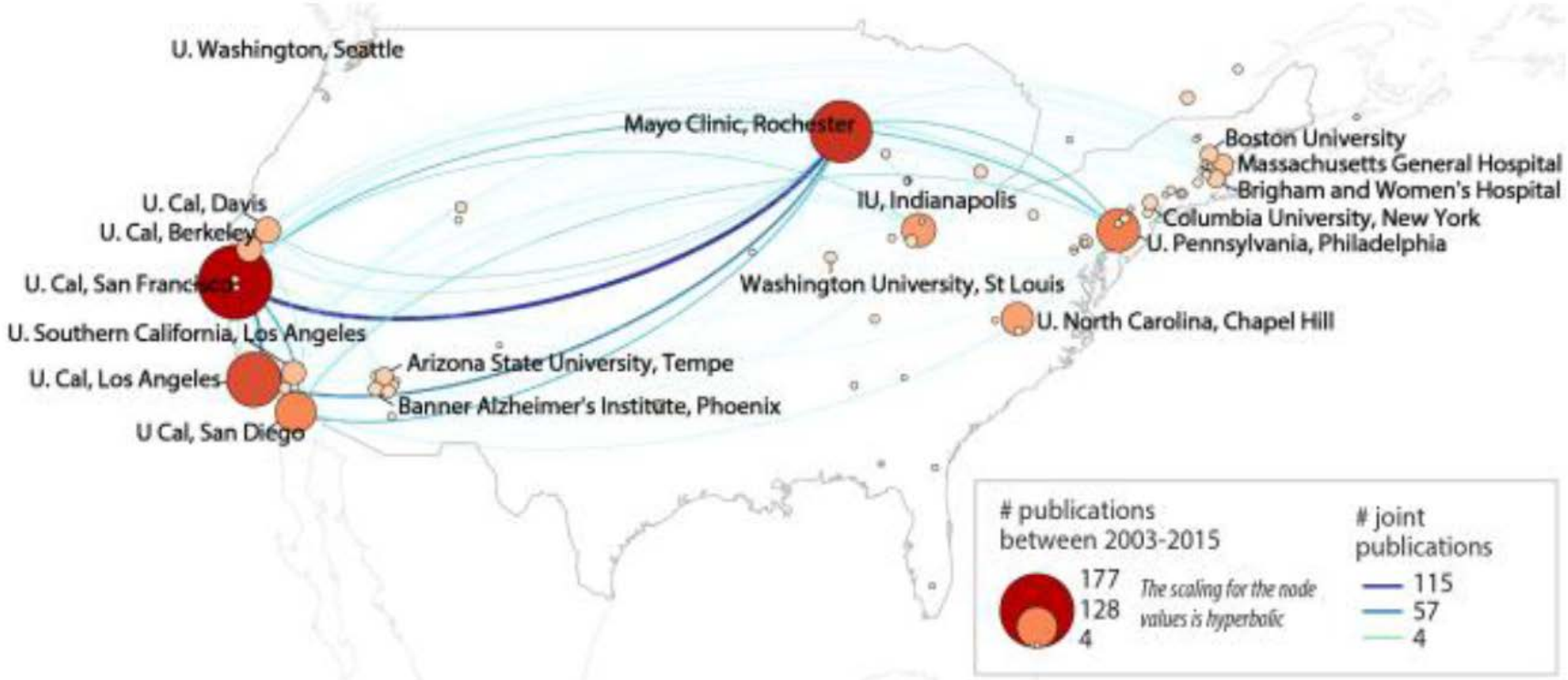
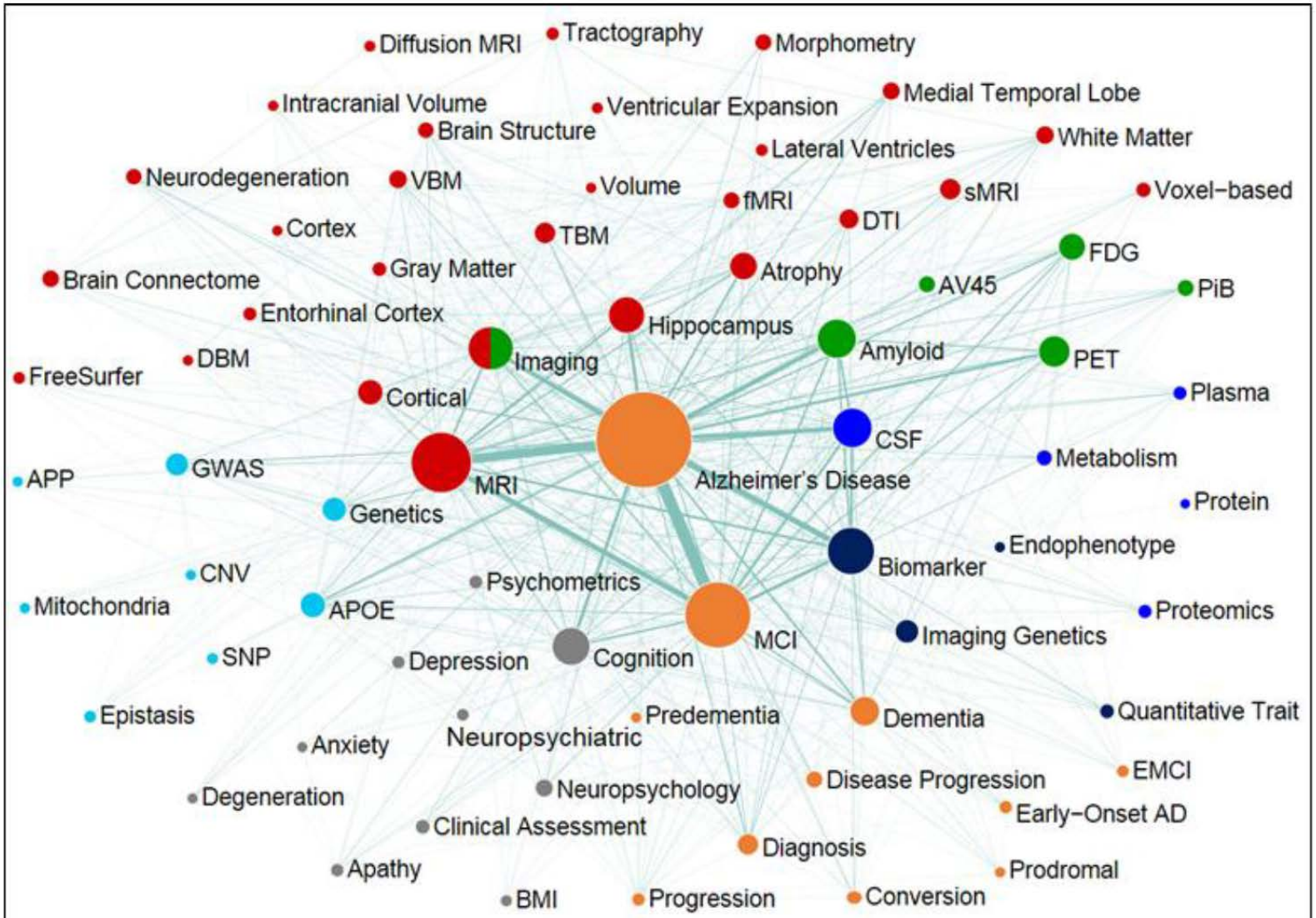


Figure 3: Co-affiliation network overlaid on a geospatial map shows collaborating organizations affiliated with ADNI in North American based on co-authored publications. Only organizations with at least 4 publications are shown; organizations with at least 30 publications or that are a Core ADNI research institution have been labeled in the network. Organization relationships (edges) with four or more co-authorships are shown.



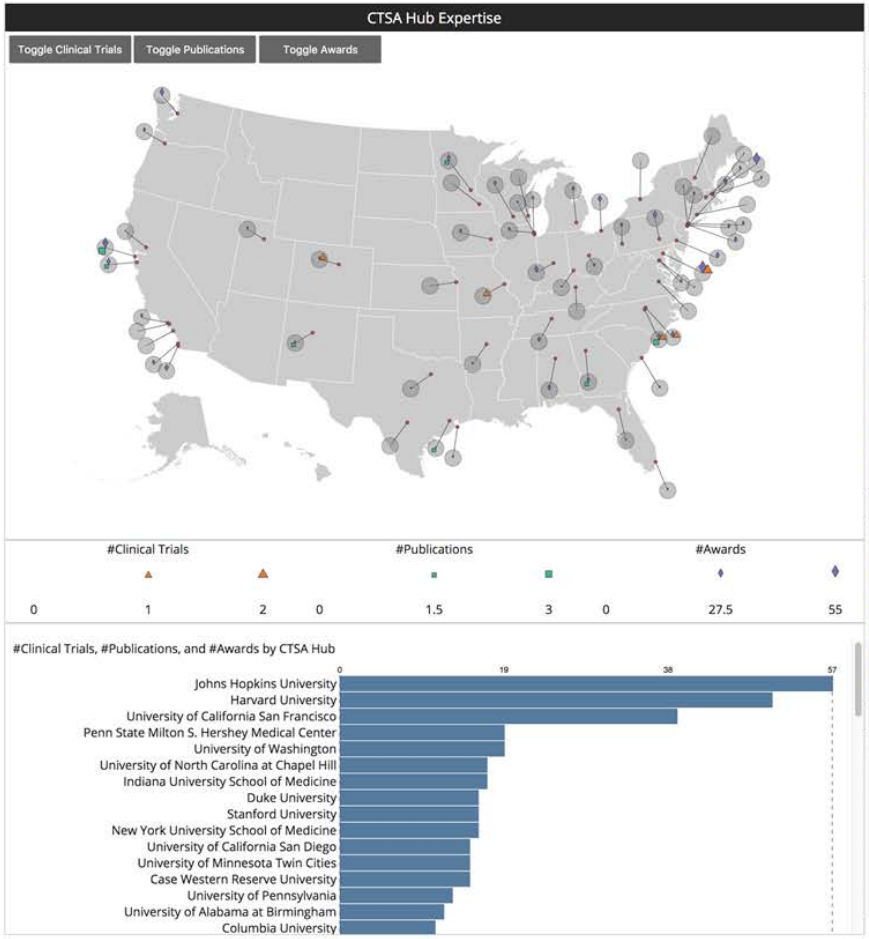
Supplemental Figure 5: Keyword co-occurrence network focused on major ADNI themes. Nodes represent keywords relevant to major ADNI themes, including MRI, PET, other biological biomarkers, clinical and neuropsychological assessment, genetics, and disease and progression. Edges denote the joint appearance of keywords in a publication. Nodes are colored based on the themes they belonged to, and those across three or more themes are colored in dark blue. Both nodes and edges were scaled proportionally based on Bezier curve. Only nodes with degree > 2 are shown.

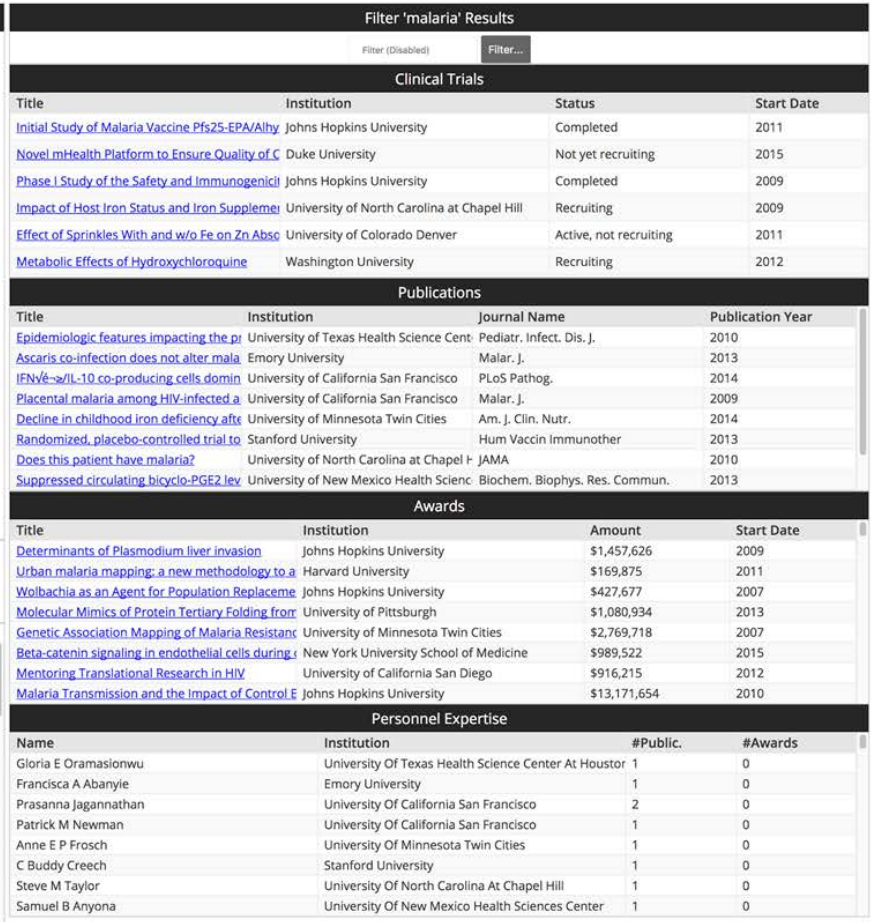
“Expertise Visualization” for NIH’s Clinical and Translational Science Awards (CTSA) Program Hubs. The online service lists key experts, publications, funding awards and clinical trials that match user-specified search queries. *Collaborative work with Intelligent Automation, Inc. work; NIH SBIR Phase II project entitled “SMS-VAT: A Scalable Multi-Scale Visual Analytical Tool.”* Sept. 14 - Aug. 16.

Visualization: IAI Expertise Visualization

Project: IAI

demo.cns.iu.edu/client/iai/expertise.html?set=malaria



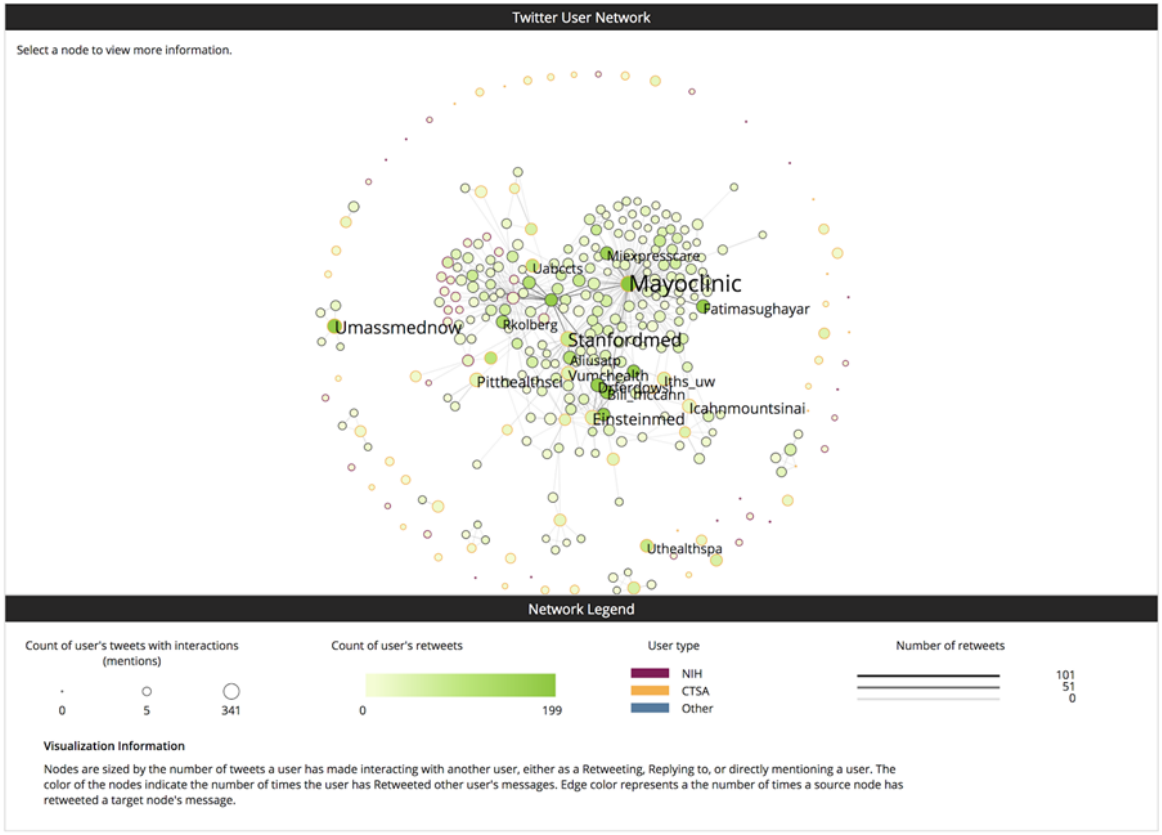


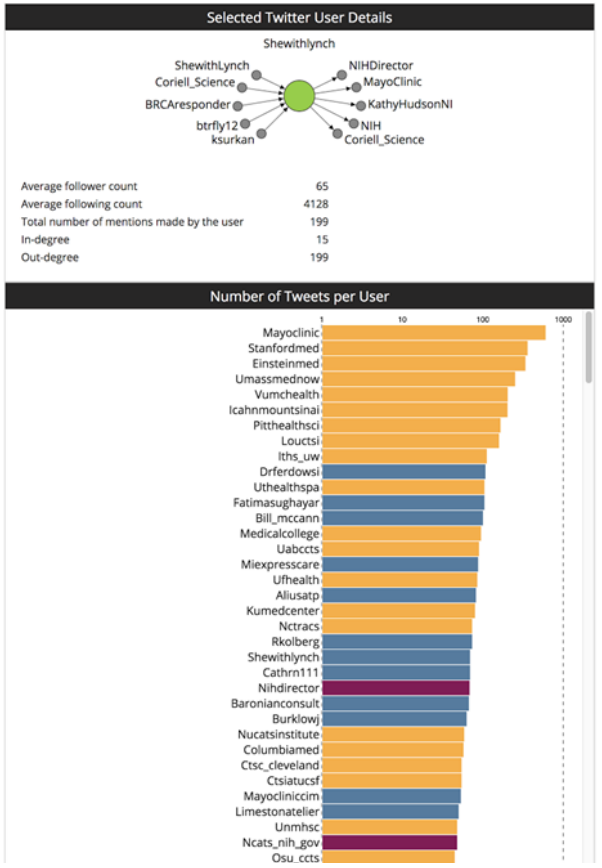
This visualization is based on publication datasets retrieved from the Scholarly Database at IU, and is used to identify relevant experts, publications, clinical trials, and awards that match a search term.

Visualization: Twitter Network

Project: IAI

demo.cns.iu.edu/client/iai/twitter.html





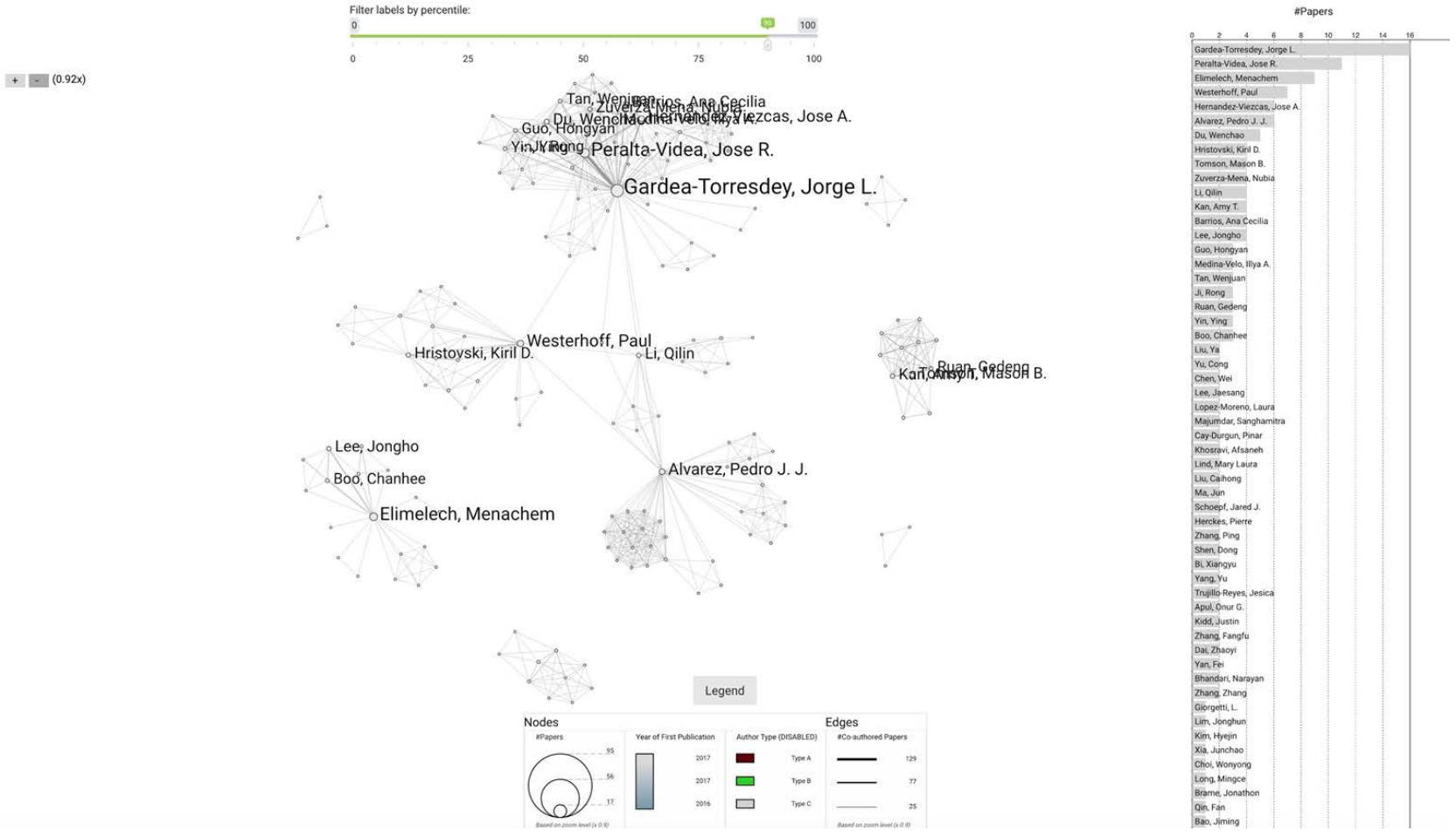
This visualization shows CTSA hub and NIH activities on Twitter. The accounts and tweets associated with CTSA and NIH were collected between August 2015 - Sept. 2015, then processed and analyzed to create a social network based on the interaction behaviors of users. The layout of the user network is force-directed, meaning that nodes that are close to each other have a stronger connection to one another indicating which Twitter users interact with each other most often.

“Engineering Observatory” that facilitates near real-time monitoring of Engineering Research Centers (ERCs) in support of informed decision making. Relevant data streams comprising course data, publications, patents, scientific datasets, code will need to be federated. Data mining and visualization web services will be provided for different stakeholders (NSF staff, researchers, students) to increase their understanding of temporal, geospatial, topical, and network patterns and trends in engineering. User evaluations will be performed to validate and optimize the new functionality. *This collaborative work with the nanoHub team at Purdue University is funded by NSF.* Dec 15 – Nov 17.

Visualization: Co-Authorship Network

Project: ERC

nanohub.org/citations/curate

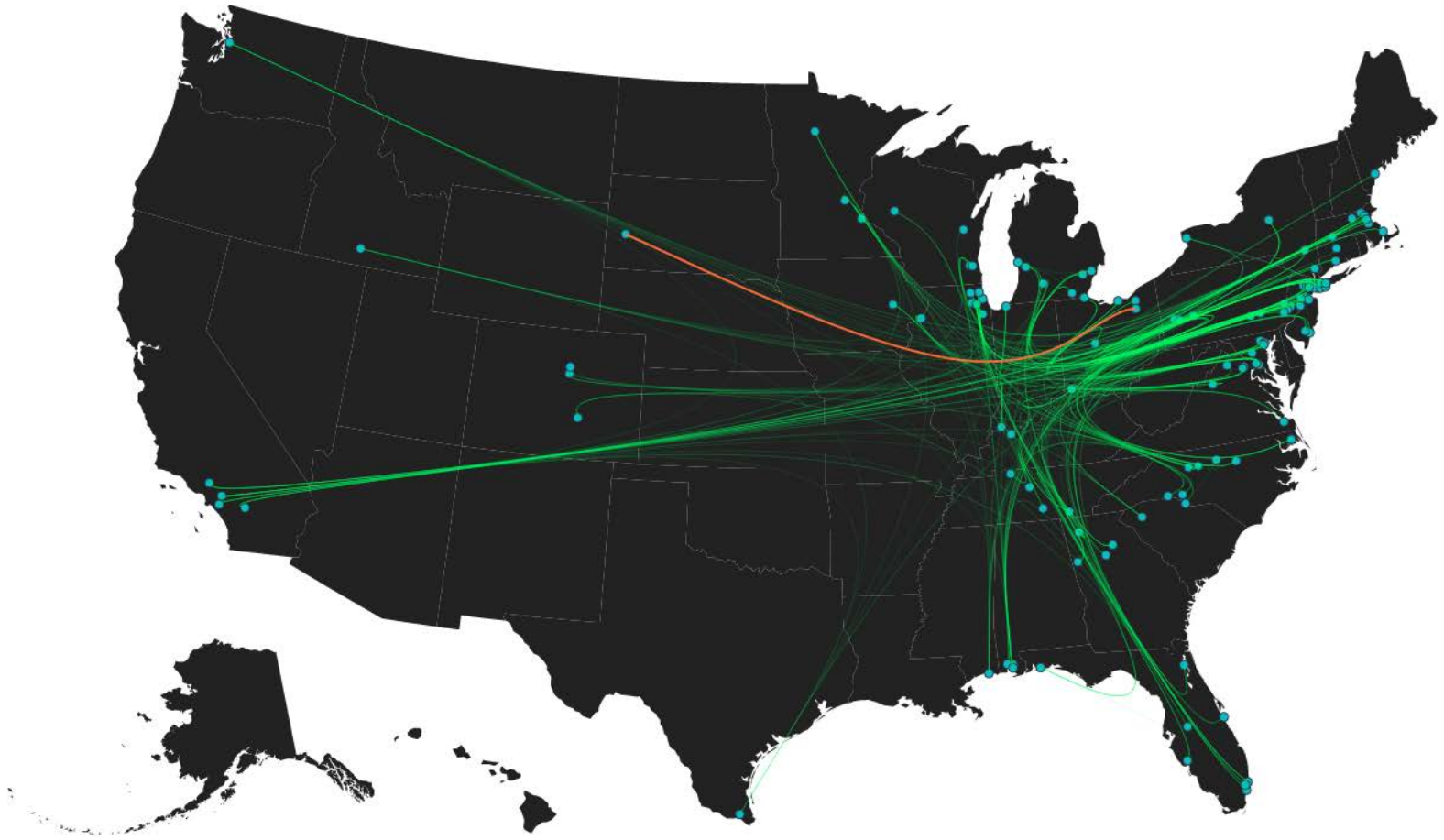


This figure shows the collaboration network of one ERC, based on co-authorship extracted from bibliography files. Each node is an author, and 2 authors are connected if they have authored a publication together.

Visualization: Geographic co-authorship visualization

Project: ERC

nanohub.org/citations/curate

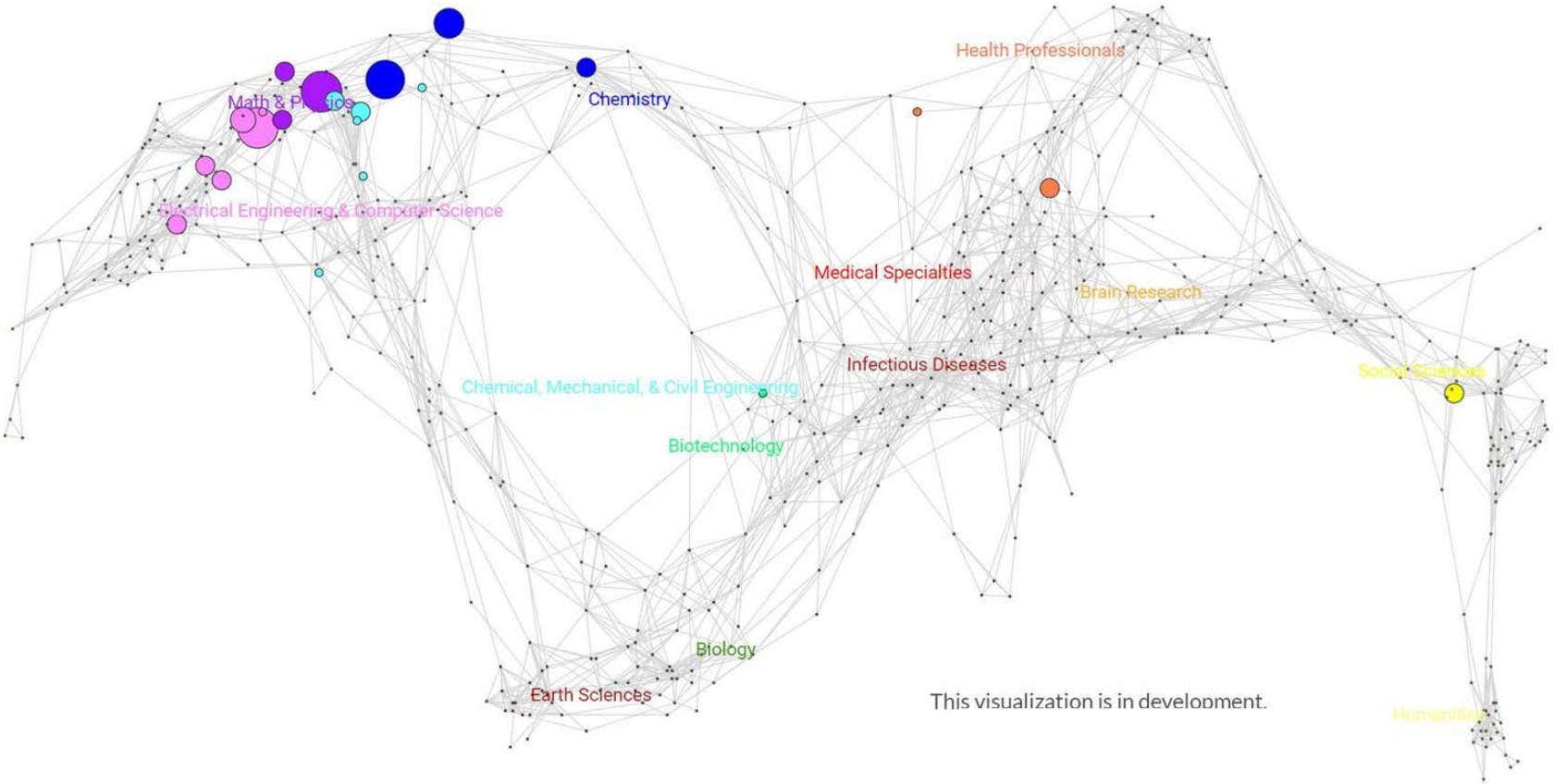


This map shows the co-authorship network overlaid on a geospatial map of the US. each node represents an author and two authors are connected if they have authored a paper together.

Visualization: UCSD Map of Science

Project: ERC

nanohub.org/citations/curate

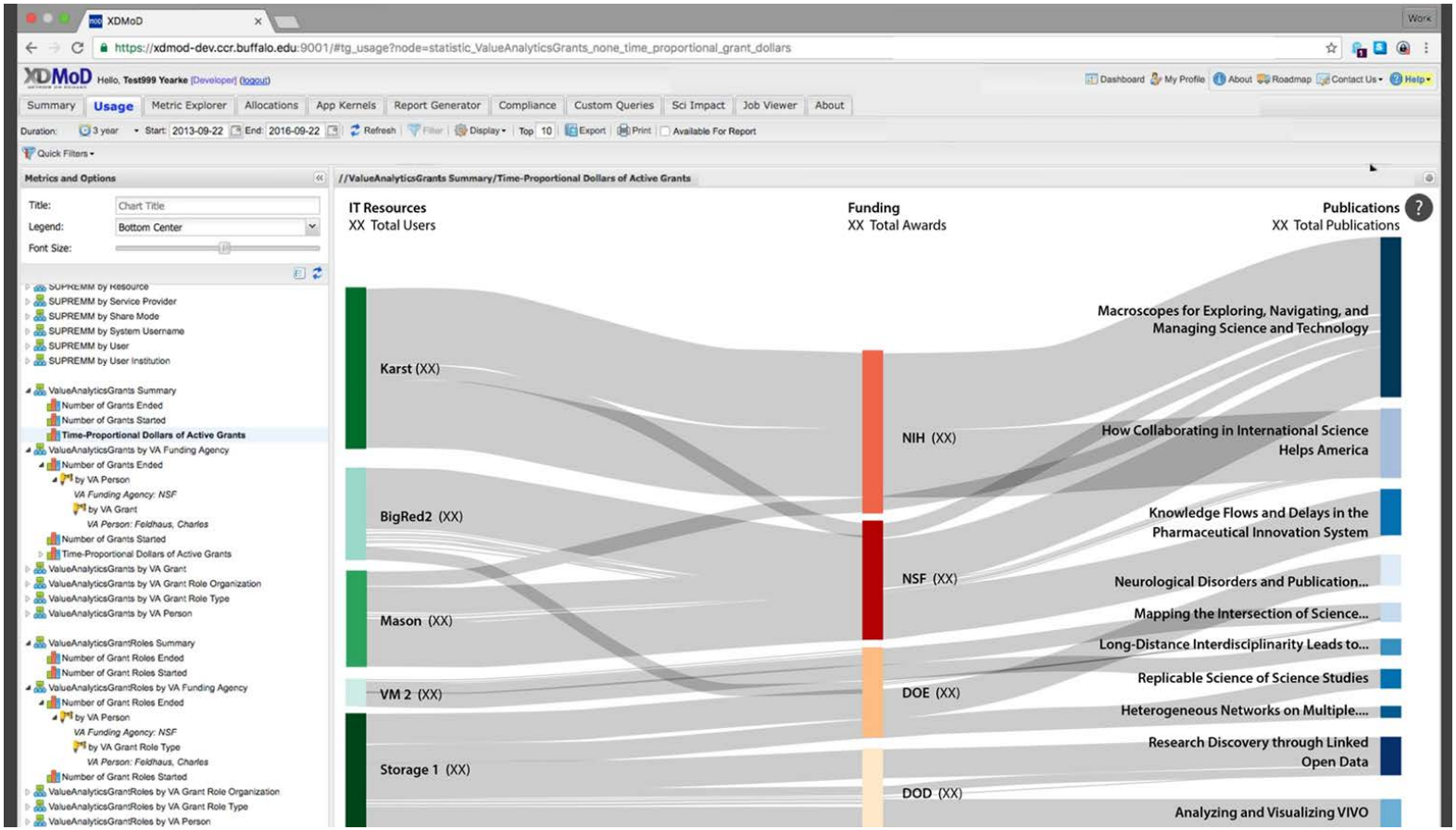


This organizes and visually represents 554 sub disciplines of science and their relationships to one another. Sub disciplines are grouped into 13 overarching disciplines that are color coded (red for medicine, green for biology, etc.) and labelled. Using a journal name based or keyword based mapping process, data overlays can be computed. For example, expertise profiles for an individual or an institution are generated by reading a bibtex or EndNote file with relevant publications, identifying unique journal names, and overlaying geometric symbols such as circles atop the sub discipline(s) that are associated with each journal. This Map of science can be used to explore, understand, and communicate the expertise profiles of an institute or nation.

“XDMoD Value Analytics” aims to improve our understanding of the interplay between compute resource availability, resource consumption, and scientific outputs. The overall goal of this line of research is to provide data-driven, objective input to regular evaluation exercises but also to support near-real time proactive management and resource allocation decision making related to optimizing the usage of advanced computing infrastructure. *This collaborative work with UITS@IUB is funded by an NSF EAGER Award.* April 16 - March 18.

Visualization: Sankey Diagram

Project: XDMoD



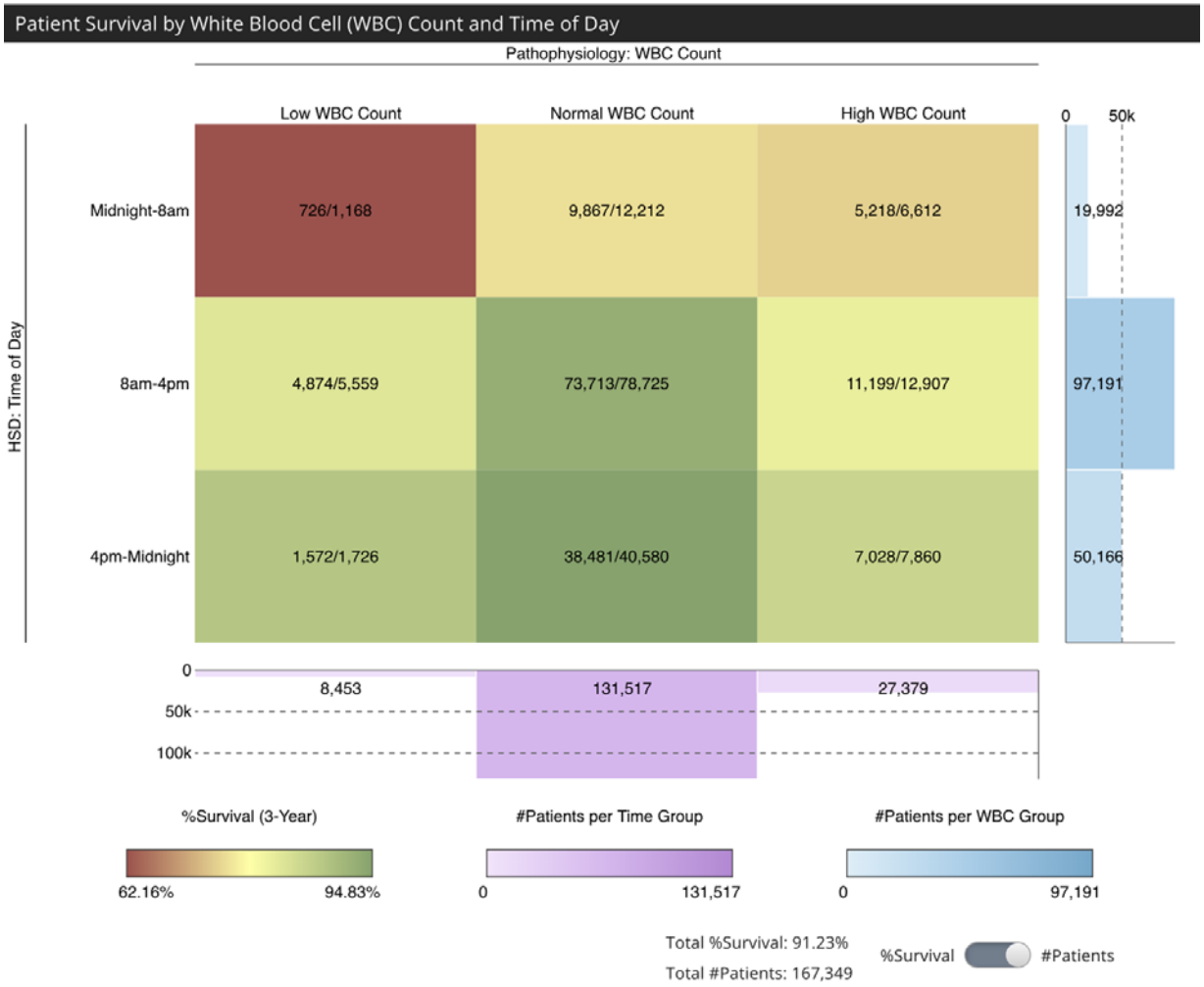
This Sankey diagram displays a multivariate analysis of the relationship between IT resources, funding agencies, and publications. The width of each line represents grant dollars awarded to researchers. The configuration model allows for easy metric switching.

“Visualizing Healthcare System Dynamics in Big Biomedical Science.” NIH U01CA198934 (Griffin Weber, Harvard University, Katy Börner) June 15 - May 18.

Visualization: Heatmap

Project: HSD

demo.cns.iu.edu/client/hsd/static/heatmap_group.html



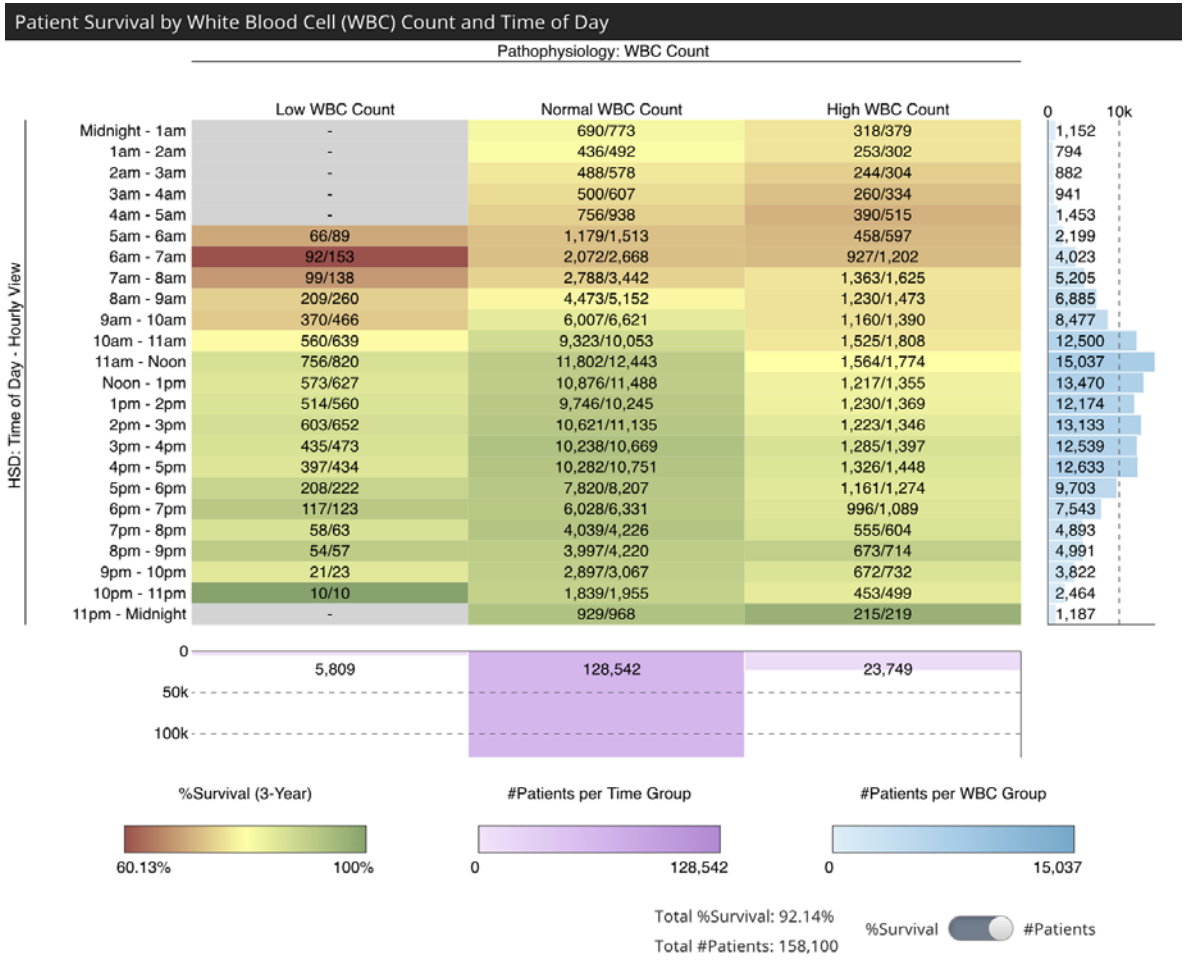
This visualization shows how white blood cell (WBC) laboratory tests correlate with three-year survival rates. The HSD dimension of the data (rows) is the time of the day of the test; and three-year survival rate (numbers and colors in the boxes) is an outcome variable. **Aggregation level for the HSD time of day are shown—three 8 hour blocks**. The lowest survival rates are for patients with a low WBC value in the morning (specifically at 6am).

In this project, we created data visualizations to explain HSD to users and to help them incorporate it into in their research.

Visualization: Heatmap

Project: HSD

demo.cns.iu.edu/client/hsd/static/heatmap_hour.html



This visualization shows how white blood cell (WBC) laboratory tests correlate with three-year survival rates. The HSD dimension of the data (rows) is the time of the day of the test; and three-year survival rate (numbers and colors in the boxes) is an outcome variable. **Aggregation level for the HSD time of day are shown— 24 hourly blocks on the right.**. The lowest survival rates are for patients with a low WBC value in the morning (specifically at 6am).

In this project, we created data visualizations to explain HSD to users and to help them incorporate it into in their research.

M: The Human Body Atlas: High-Resolution, Functional Mapping of Voxel, Vector, and Meta Datasets

Katy Börner, Bruce Herr II, Paul Macklin & Randy Heiland
Intelligent Systems Engineering, SICE, Indiana University, Bloomington, IN

Griffin Weber, Harvard Medical School, Boston, MA
Samuel Friedman, Opto-Knowledge Systems, Inc.

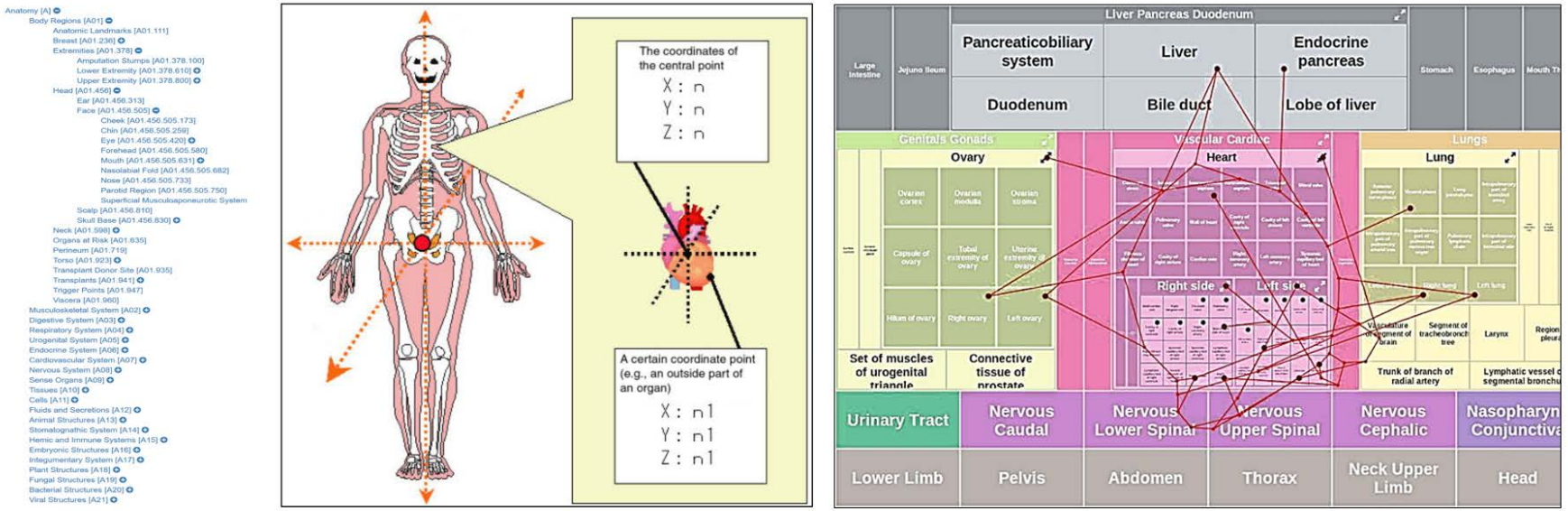
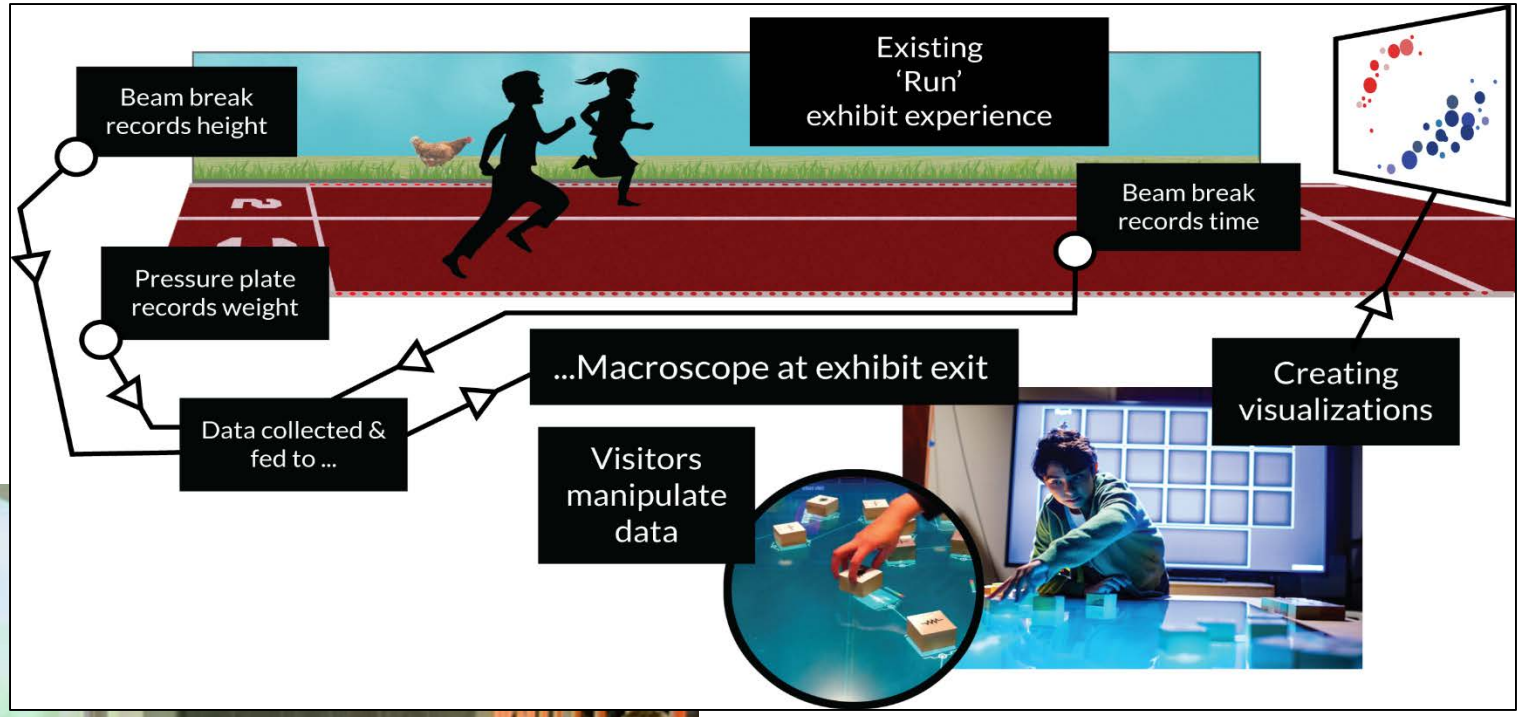
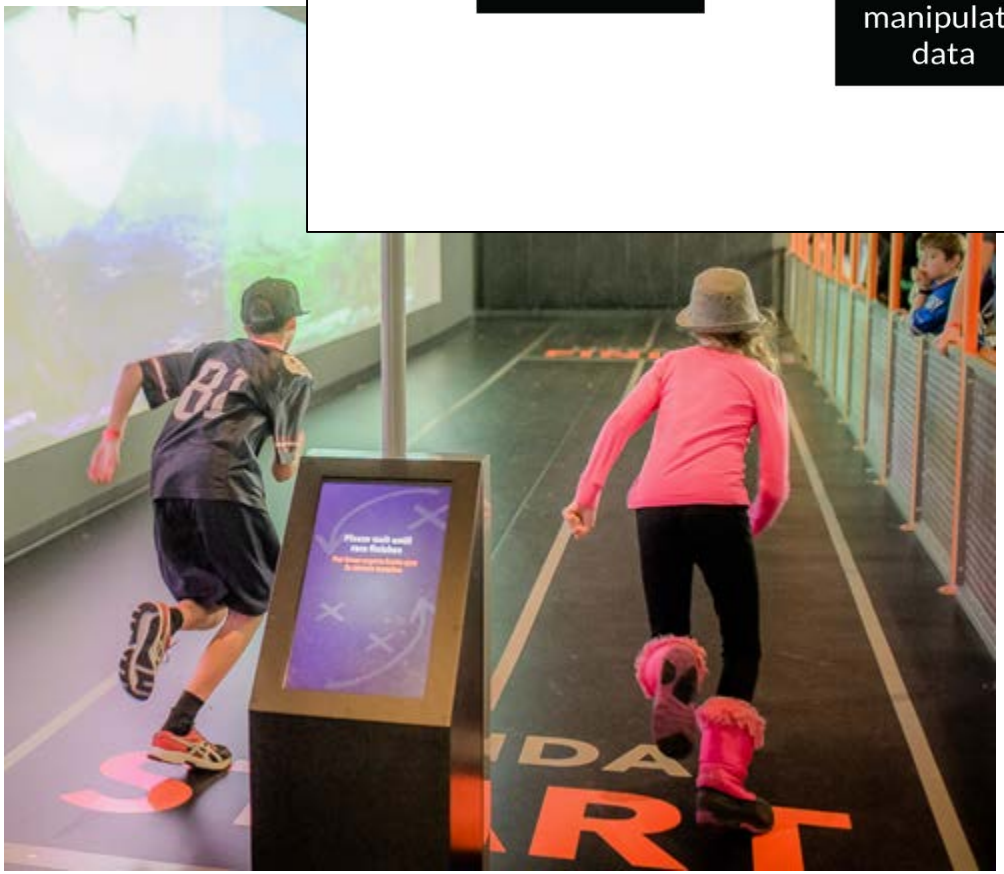


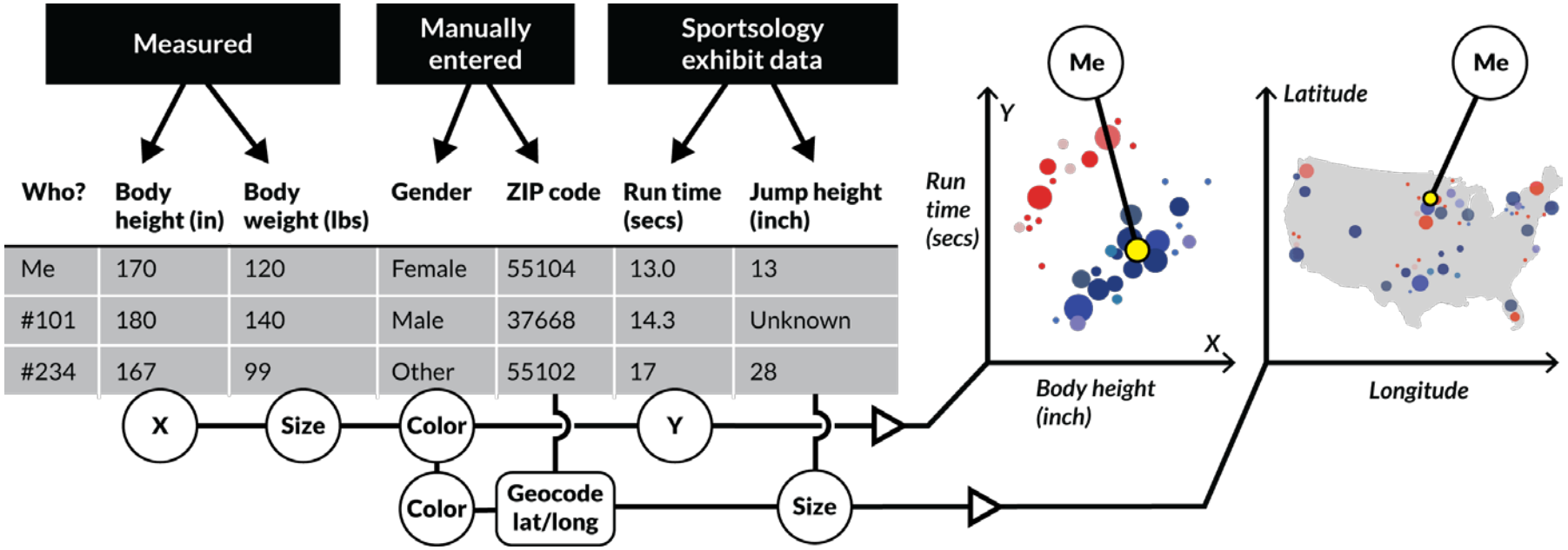
Fig. 7: CCF Ontology Browser (left), spatial map of human body, adopted from http://www.natureinterface.com/e/ni04/P056-059/ (middle), functional ApiNATOMY map[1,3]. (right)

“Visualizing Data Visualization Literacy: Research and Tools that Advance Public Understanding of Scientific Data.” NSF AISL 1713567 Award (Katy Börner, Kylie Peppler, Joseph Heimlich, Bryan Kennedy, and Stephen M Uzzo) Sept. 17 – Aug. 20.





Sketch of the *Run* exhibit including data collection (top) and macroscope add-on that lets interested visitors explore more complex data visualizations using table-top displays.

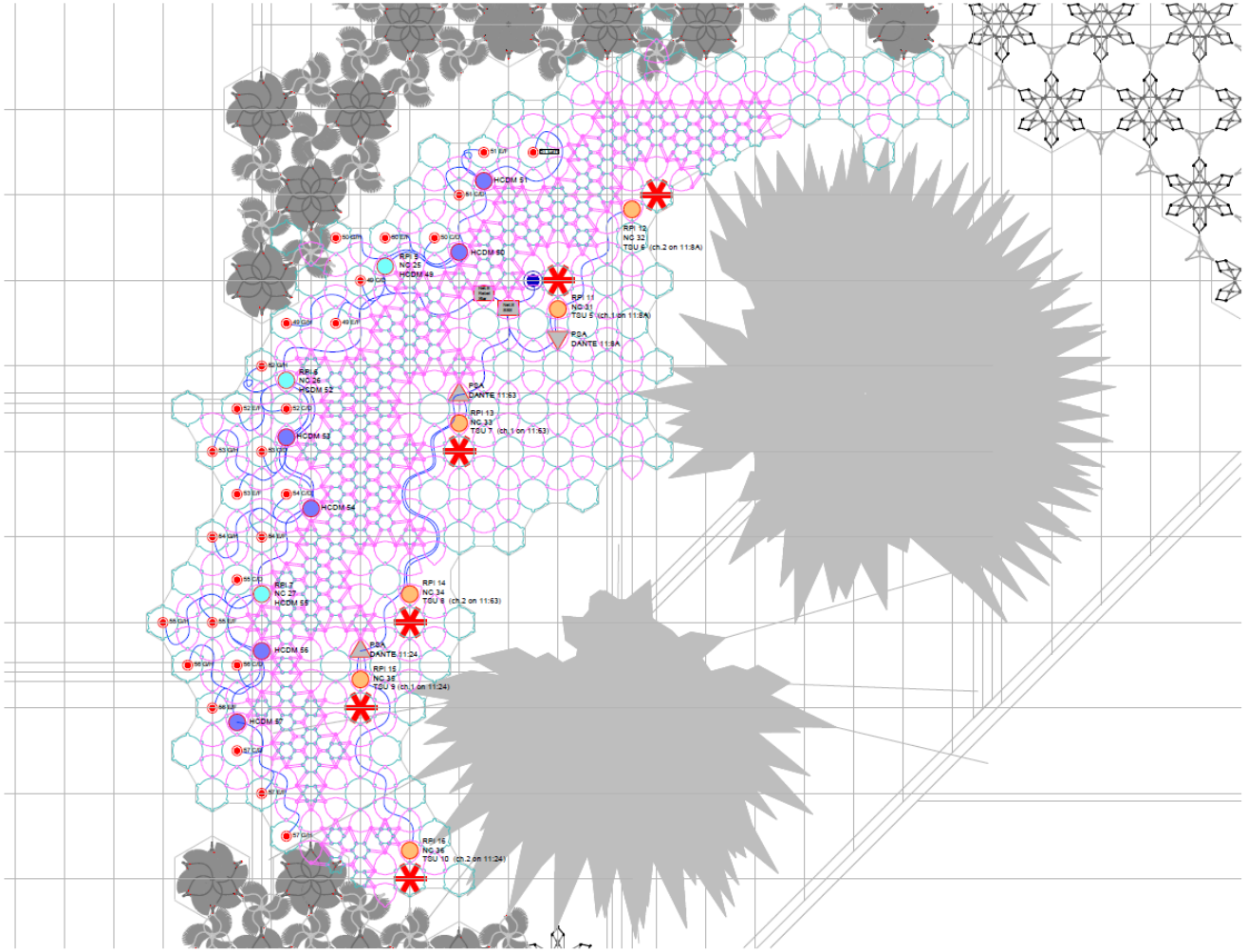


xMacroscope general setup and activity—Raw data on left is converted to visualization on right by dragging and dropping (or connecting) column headers to axes, paint buckets, size, and shape.

Using Big Data for Visualizing Living Architectures. International Research Partnership Grant, University of Waterloo, Canada (Philip Beesley, Dana Kulic, Katy Börner) Feb 17 - Jan 18.


AMATRIA
Sentient Architecture

https://cns.iu.edu/amatria.html



PBAI / LASG

Philip Beesley
Architect Inc.
213 Sterling Road Suite 200
Toronto, Canada
M6R2B2
web: philipbeesleyarchitect.com
web: livingarchitecturesystems.com
tel: 416 766 8284

By	Date	Status	Rev By	Rev Date
JP	18/04/02	As-Built	TB	18/06/13

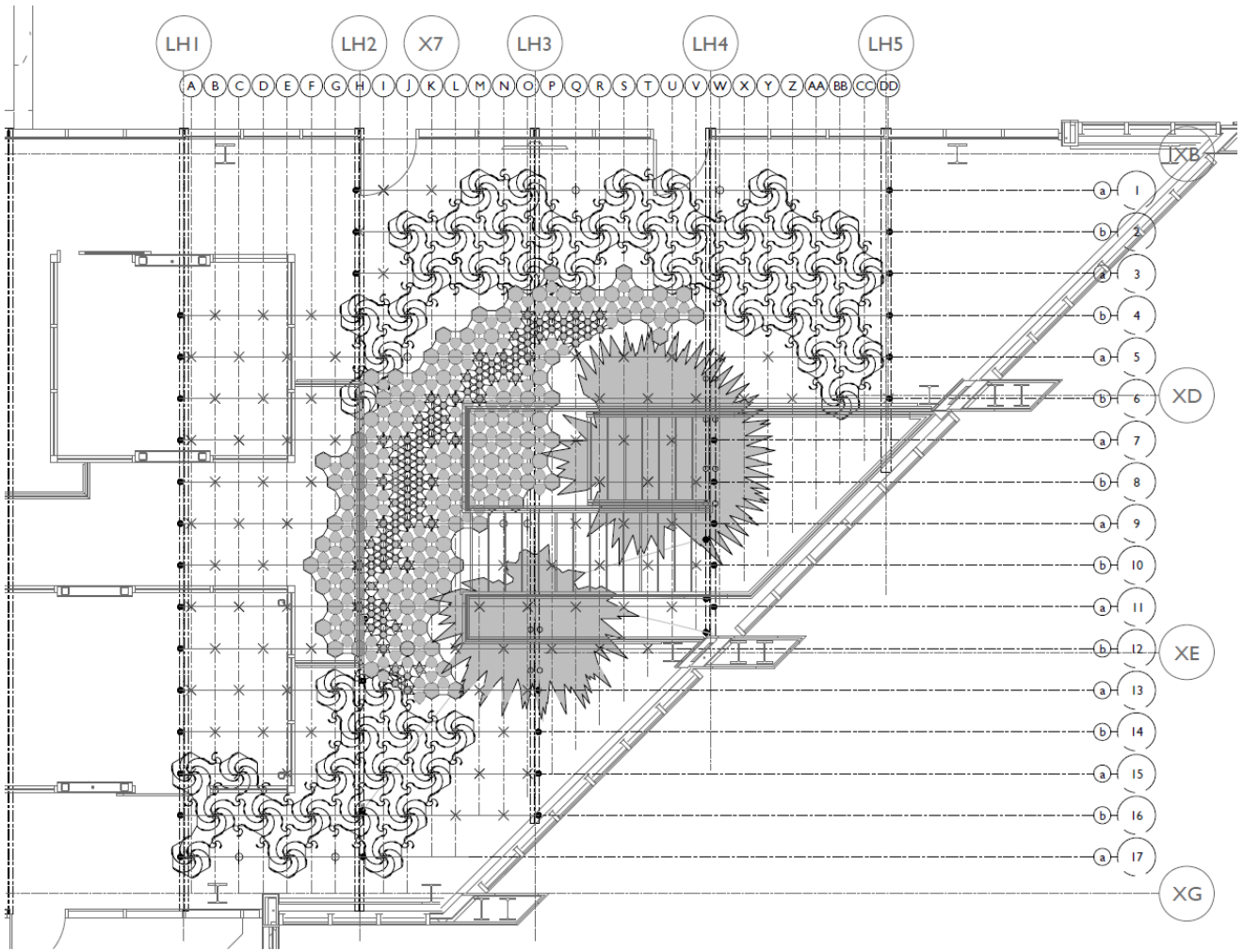
Notes

Phase
Maintenance Manual

Project
17540 Luddy Hall - Amatria

Drawing Title
Device Locator - Sparfield Canopy

Sheet
M102



Living Architecture
Systems Group/
Philip Beesley
Architect Inc.

213 Sterling Road Suite 200
Toronto, Canada
M6R2B2

web: lasg.ca
tel: 416 766 8284

By	Date	Status	Rev By	Rev Date
TB	18/02/09	DRAFT	MF	18/02/09
TB	18/02/13	DRAFT	VF	18/02/13
TB	18/02/16	DRAFT	PB	18/02/16

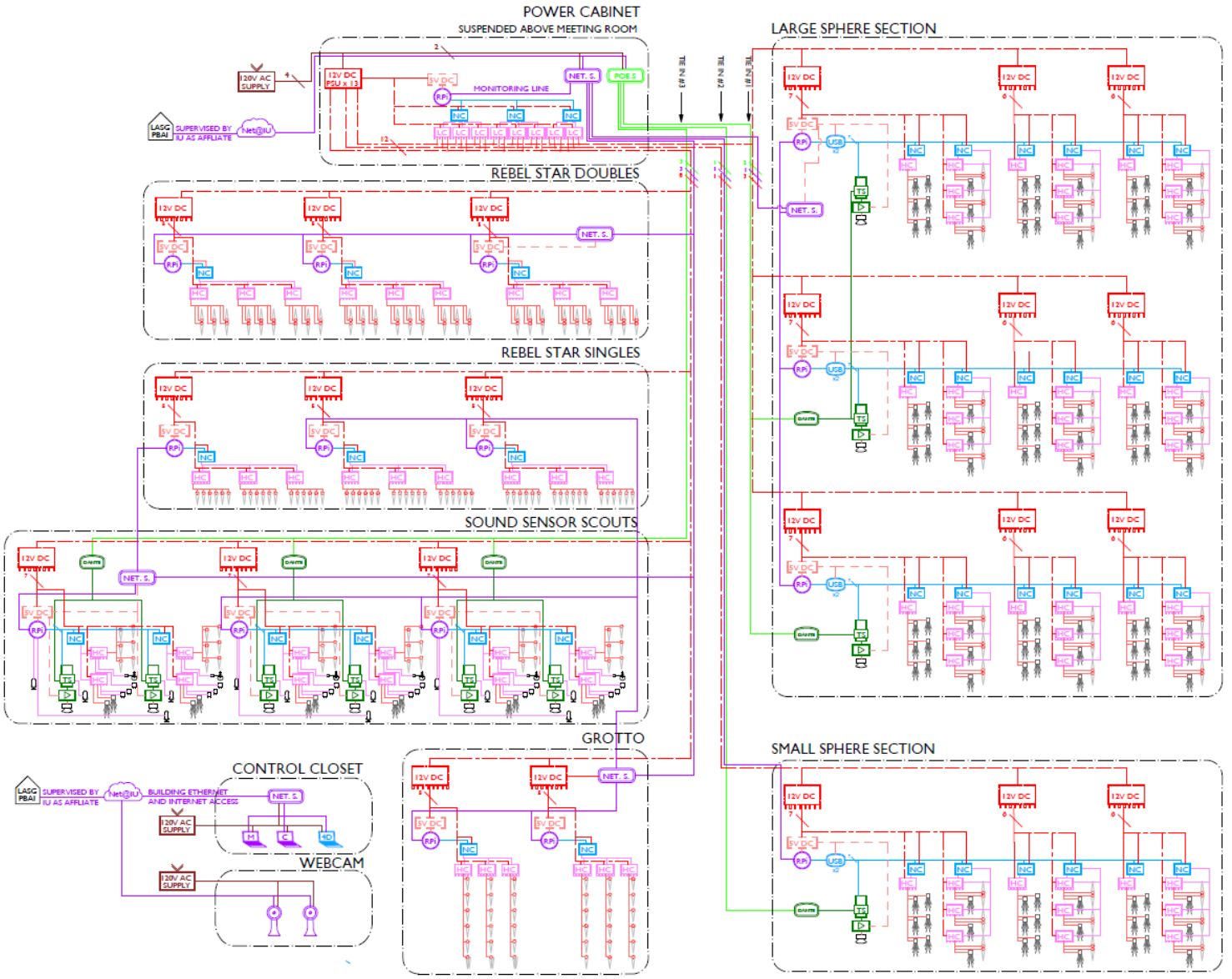
Notes

Phase
Schematic Design

Project
17540 Luddy Hall

Drawing Title
Sculpture Master Plan

Sheet
A102



Living Architecture
Systems Group/
Philip Beesley
Architect Inc.

213 Sterling Road Suite 200
Toronto, Canada
M6R2B2
web: lasg.ca
tel: 416 766 8284

By	Date	Status	Rev By	Rev Date
MH	01/03/08	Draft		

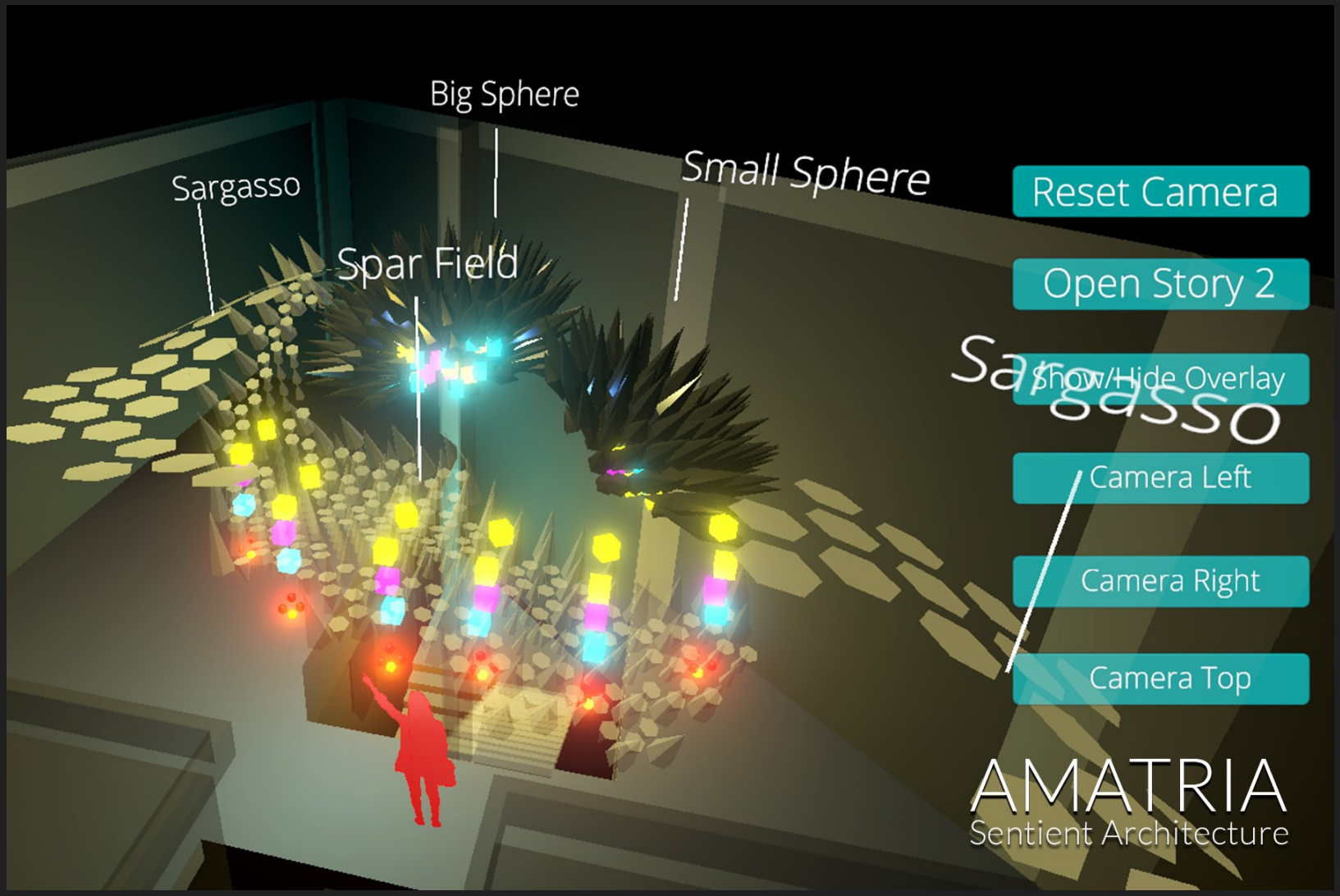
DRAWING LEGEND

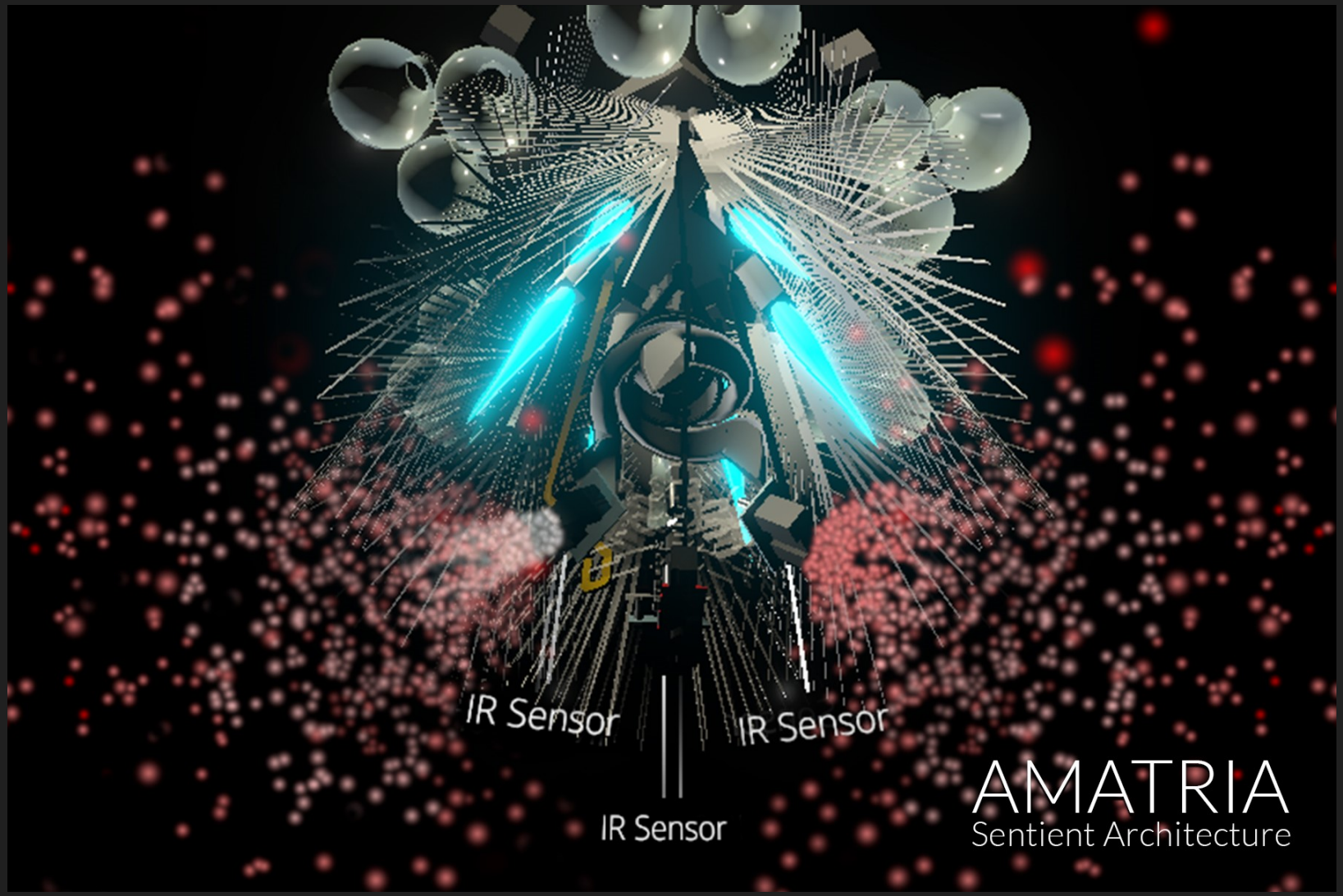
Phase
Design
Development

Project
17540 Luddy Hall

Drawing Title
Electronic Schematic

Sheet
IS10••


Big Sphere
Sargasso
Small Sphere
Spar Field
Reset Camera
Open Story 2
Show/Hide Overlay
Sargasso
Camera Left
Camera Right
Camera Top
AMATRIA
Sentient Architecture

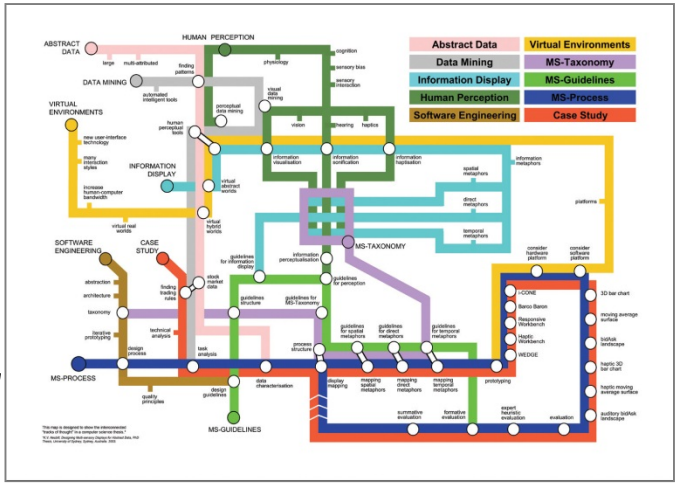

IR Sensor
IR Sensor
IR Sensor
AMATRIA
Sentient Architecture

Register for free: http://ivmooc.cns.iu.edu.

Different Question Types

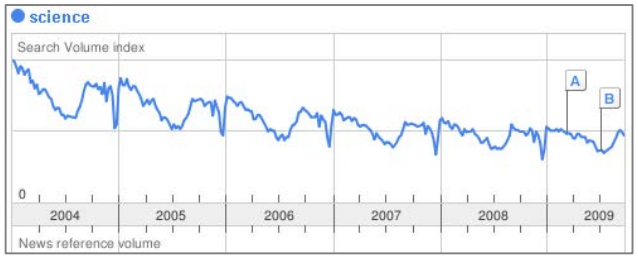
Terabytes of data

Descriptive & Predictive Models



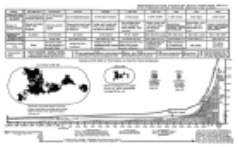
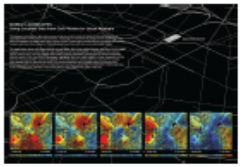
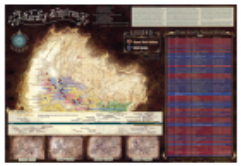
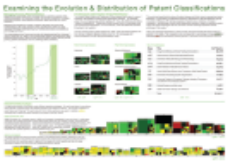
Find your way

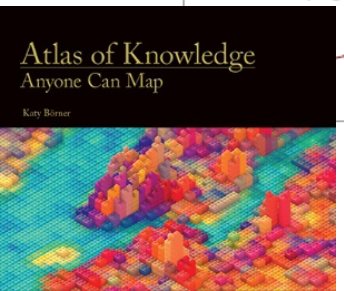
Find collaborators, friends



Identify trends

Tasks

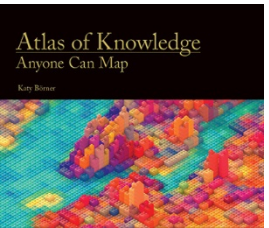
TYPES	LEVELS		
	MICRO: Individual Level about 1–1,000 records page 6	**MESO: Local Level** about 1,001–100,000 records page 8	**MACRO: Global Level** more than 100,000 records page 10
Statistical Analysis page 44	 Knowledge Cartography page 135	 Productivity of Russian life sciences research teams page 105	 Number of scientists versus population and R&D costs versus GNP. page 103
WHEN: Temporal Analysis page 48	 Visualizing decision-making processes page 95	 Key events in the development of the video tape recorder page 85	 Increased travel and communication speeds page 83
WHERE: Geospatial Analysis page 52	 Cell phone usage in Milan, Italy page 109	 Victorian poetry in Europe page 137	 Ecological footprint of countries page 99
WHAT: Topical Analysis page 56	 Evolving patent holdings of Apple Computer, Inc. and Jerome Lemelson page 89	 Evolving journal networks in nanotechnology page 139	 Product space showing co-export patterns of countries page 93
WITH WHOM: Network Analysis page 60	 World Finance Corporation network page 87	 Electronic and new media art networks page 133	 World-wide scholarly collaboration networks page 157



See *Atlas of Science: Anyone Can Map*, page 5

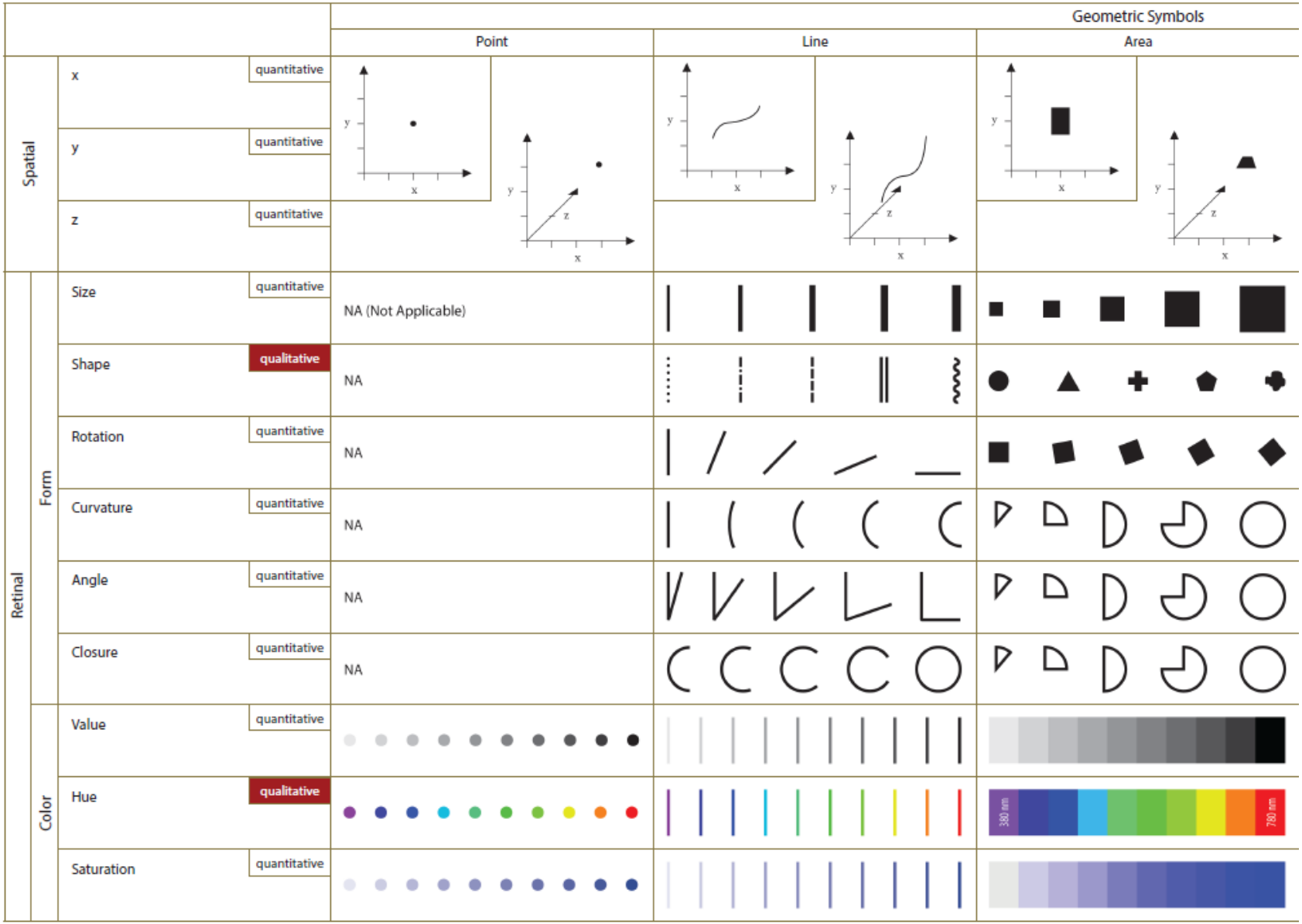
Visualization Framework

Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships	• nominal • ordinal • interval • ratio	• table • chart • graph • map • network layout	• geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs	• spatial position • retinal form color optics motion	• overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion

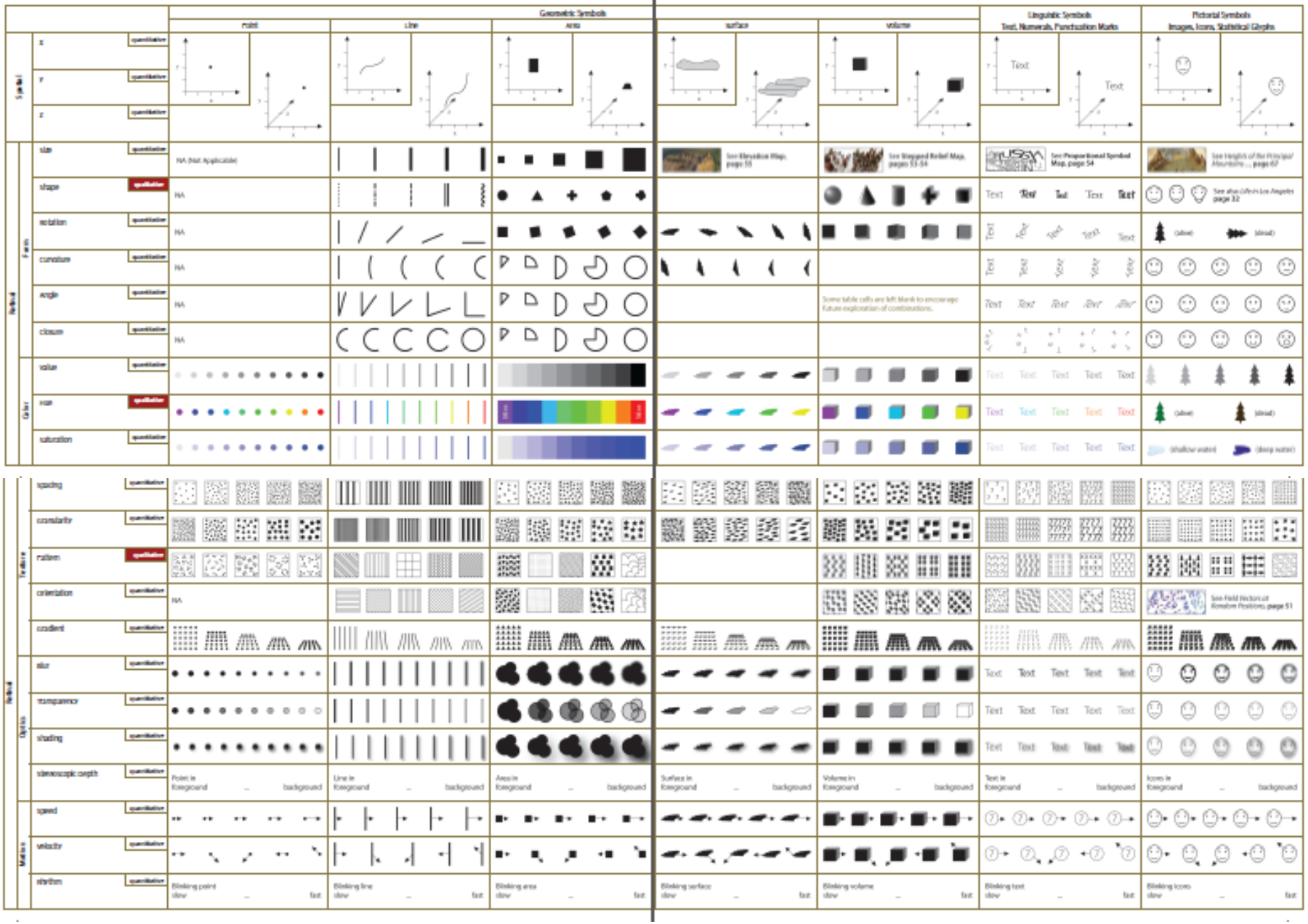


See page 24

Graphic Variable Types Versus Graphic Symbol Types

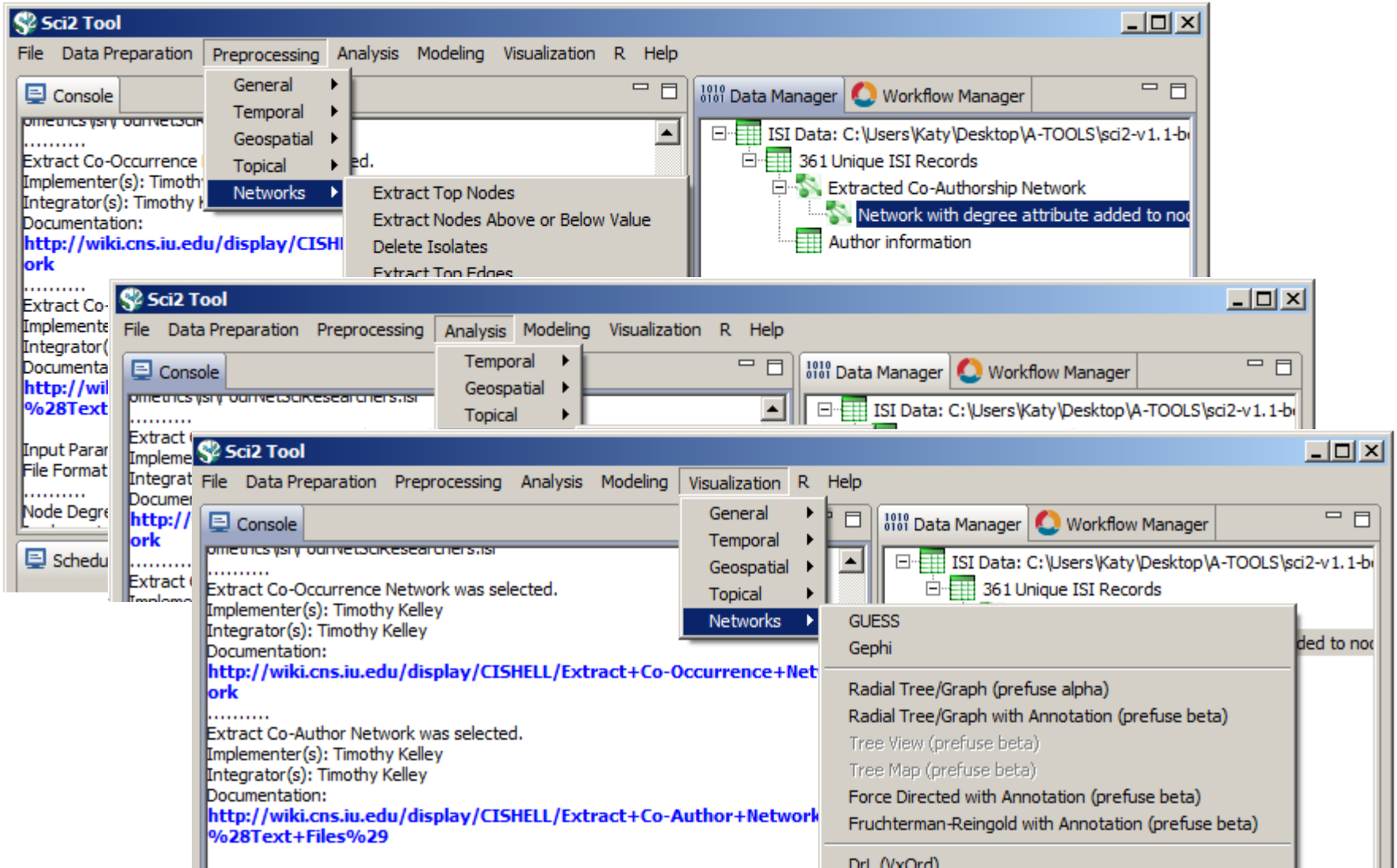


Graphic Variable Types Versus Graphic Symbol Types



Sci2 Tool Interface Components Implement Vis Framework

Download tool for free at http://sci2.cns.iu.edu



Load **One** File and Run **Many** Analyses and Visualizations

Times Cited	Publication Year	City of Publisher	Country	Journal Title (Full)	Title	Subject Category	Authors
12	2011	NEW YORK	USA	COMMUNICATIONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONAL SCIENCE	Advancing the Science of Team Science	Research & Experimental Medicine	Falk-Krzesinski, HJ|Borner, K|Contractor, N|Fiore, SM|Hall, KL|Keyton, J|Spring, B|Stokols, D|Trochim, W|Uzzi, B
13	2010	WASHINGTON	USA	SCIENCE TRANSLATIONAL MEDICINE	A Multi-Level Systems Perspective for the Science of Team Science	Cell Biology |Research & Experimental Medicine	Borner, K|Contractor, N|Falk-Krzesinski, HJ|Fiore, SM|Hall, KL|Keyton, J|Spring, B|Stokols, D|Trochim, W|Uzzi, B

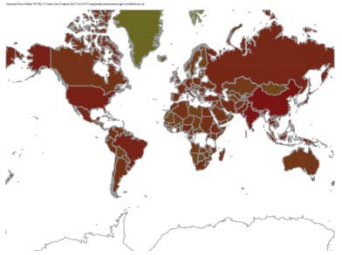
Statistical Analysis–p. 44

Location	*Count*	*# Citations*
Netherlands	13	292
United States	9	318
Germany	11	36
United Kingdom	1	2

Temporal Burst Analysis–p. 48

Geospatial Analysis–p. 52

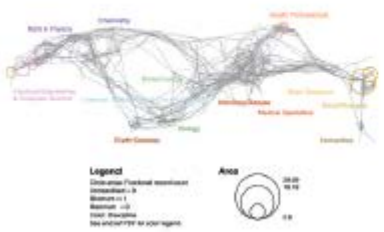
Geospatial Analysis–p. 52



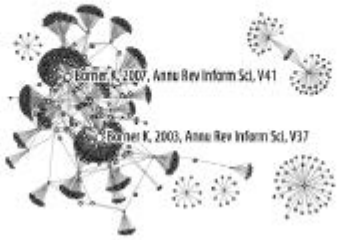
Load **One** File and Run **Many** Analyses and Visualizations

Times Cited	Publication Year	City of Publisher	Country	Journal Title (Full)	Title	Subject Category	Authors
12	2011	NEW YORK	USA	COMMUNICATIONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONAL SCIENCE	Advancing the Science of Team Science	Research & Experimental Medicine	Falk-Krzesinski, HJ|Borner, K|Contractor, N|Fiore, SM|Hall, KL|Keyton, J|Spring, B|Stokols, D|Trochim, W|Uzzi, B
13	2010	WASHINGTON	USA	SCIENCE TRANSLATIONAL MEDICINE	A Multi-Level Systems Perspective for the Science of Team Science	Cell Biology |Research & Experimental Medicine	Borner, K|Contractor, N|Falk-Krzesinski, HJ|Fiore, SM|Hall, KL|Keyton, J|Spring, B|Stokols, D|Trochim, W|Uzzi, B

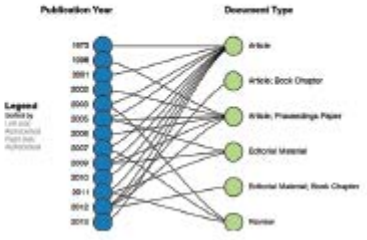
Topical Analysis–p. 56



Paper Citation Network–p. 60

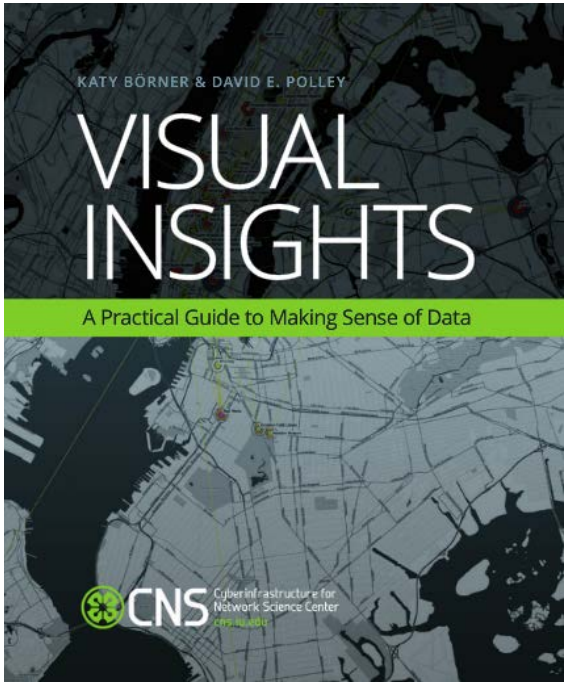


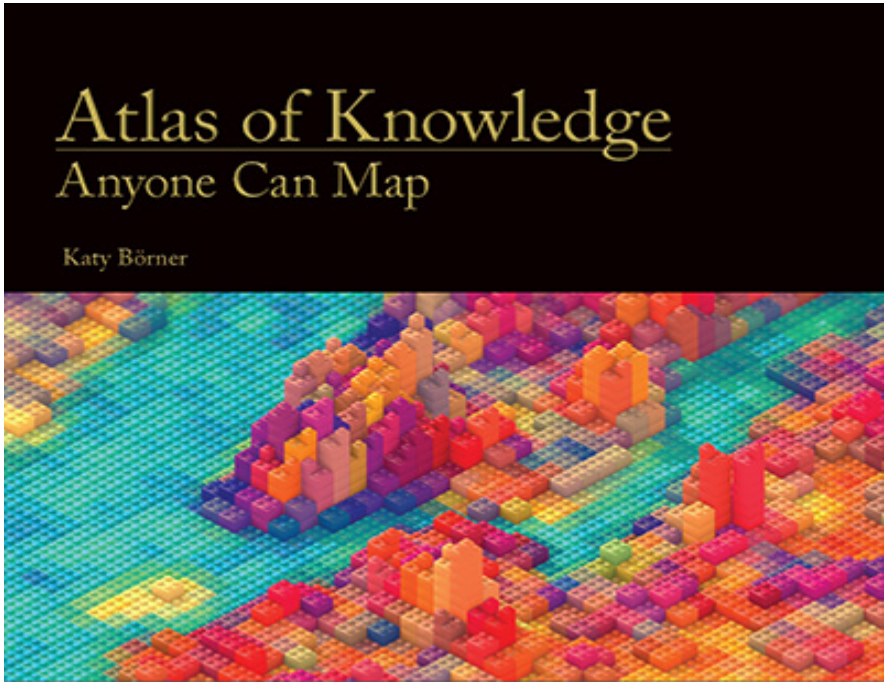
Bi-Modal Network–p. 60



Co-author and many other bi-modal networks.

Books Used in the IVMOOC





Teaches timely knowledge:

Advanced algorithms, tools, and hands-on workflows.

Teaches timeless knowledge:

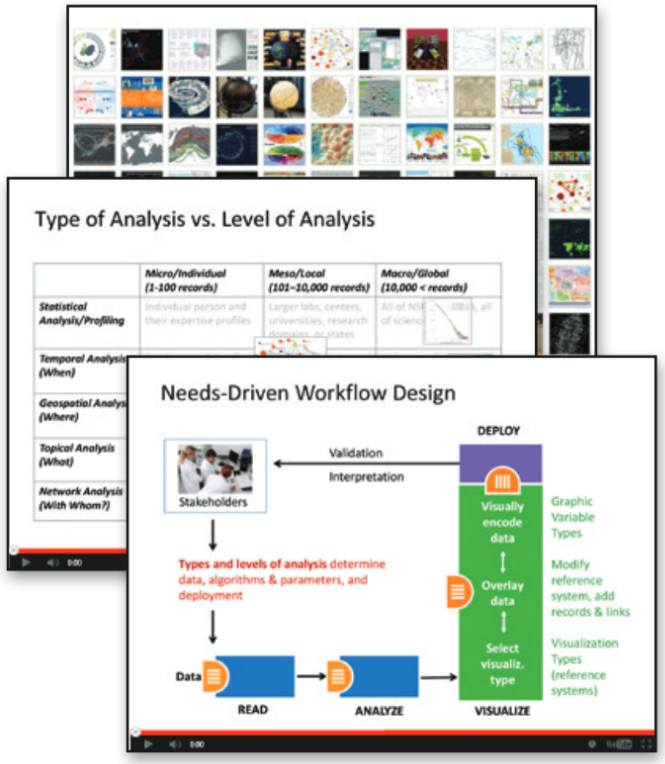
Visualization framework—exemplified using generic visualization examples and pioneering visualizations.

S637/E583/IVMOOC Information Visualization

Spring 2019

Learn to Harness the Power of Data

CNS launched the inaugural Information Visualization MOOC (IVMOOC) in January 2013, attracting participants from more than 100 countries. The course provides an overview about the state of the art in information visualization, emphasizing a user-needs-driven process. Anyone interested in generating visualizations would benefit from the course, and there are also opportunities to work with real-world clients on a variety of data visualization projects.



ENGR-E484/E584 | Fall 2019

Scientific Visualization

Instructor: William R. Sherman, shermanw@indiana.edu

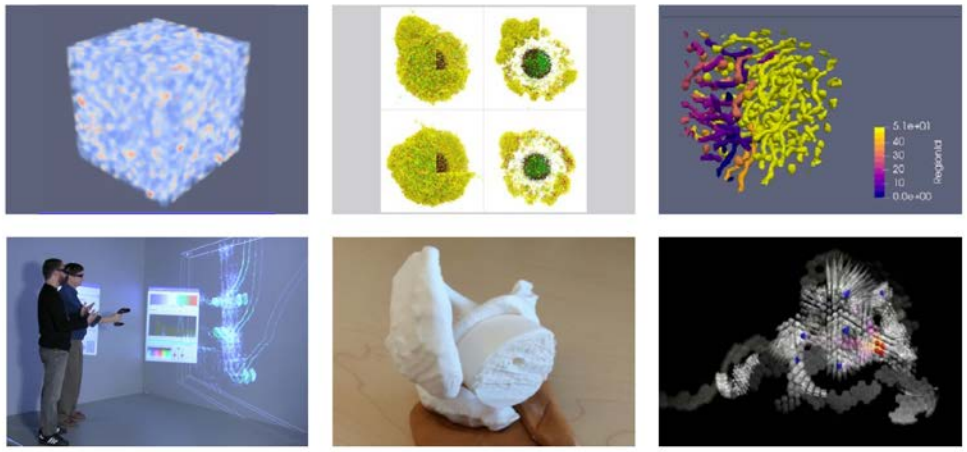
Monday/Wednesday 4:00–5:15 p.m.

Visualization Lab, Luddy Hall 4012

This 3-credit course teaches basic principles of human cognition and perception; techniques and algorithms for designing and critiquing scientific visualizations in different domains (neuro, nano, bio-medicine, IoT, smart cities); hands-on experience using modern tools for designing scientific visualizations that provide novel and/or actionable insights; 3D printing and augmented reality deployment; and teamwork/project management expertise.

Topics covered:

- Scientific visualization: Past, present, and future trends
- Human cognition and perception
- Techniques and algorithms for neurological sciences, nanotechnologies, bio-medicine, IoT, etc.
- Virtual and augmented reality visualizations
- 3D printing deployment
- Choosing and working with clients
- User and task analysis
- Client-oriented projects



INTELLIGENT SYSTEMS ENGINEERING

All papers, maps, tools, talks, press are linked from http://cns.iu.edu
These slides will soon be at http://cns.iu.edu/presentations.html

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience