

# Actionable Data Visualizations of Science & Technology

Katy Börner @katycns

Victor H. Yngve Distinguished Professor of  
Intelligent Systems Engineering & Information Science  
Director, Cyberinfrastructure for Network Science Center  
School of Informatics, Computing, and Engineering  
Indiana University Network Science Institute (IUNI)  
Indiana University, Bloomington, IN, USA  
+ 2018 Humboldt Fellow, TU Dresden, Germany



*Allen Institute for Brain Science, Seattle, WA | March 7, 2019*



# Overview

This talk will introduce a theoretical data visualization framework (DVL) meant to empower anyone to systematically render data into insights using temporal, geospatial, topical, and network analyses and visualizations.

Börner, Katy, Andreas Bueckle, and Michael Ginda. 2019. [Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments](#). *PNAS*, 116 (6) 1857-1864.

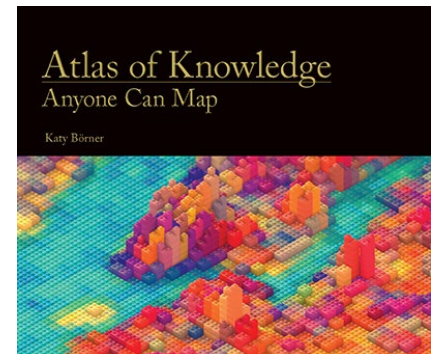
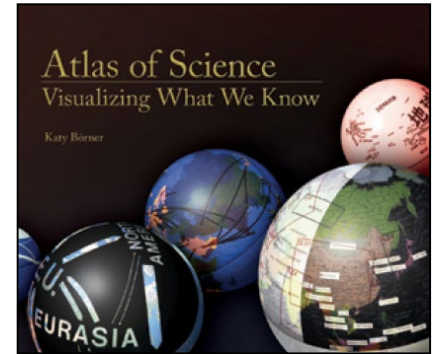
Exemplarily, the DVL is applied to

- map science and technology, see interactive data visualizations from the *Places & Spaces: Mapping Science* exhibit (<http://scimaps.org>) and recent *PNAS* special issue on *Modelling and Visualizing Science and Technology Developments* (<https://www.pnas.org/modeling>)

Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press.

Börner, Katy. 2010. [Atlas of Science: Visualizing What We Know](#). Cambridge, MA: The MIT Press.

- design reference systems and user interfaces within the Human BioMolecular Atlas Program (HuBMAP) (<https://commonfund.nih.gov/hubmap>) that support the exploration and communication of single-cell data—from the subcellular to the whole body level.





# Places & Spaces: Mapping Science Exhibit

1<sup>st</sup> Decade (2005-2014)

## Maps

Iteration I (2005)  
The Power of Maps



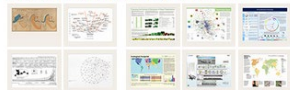
Iteration II (2006)  
The Power of Reference Systems



Iteration III (2007)  
The Power of Forecasts



Iteration IV (2008)  
Science Maps for Economic Decision Makers



Iteration V (2009)  
Science Maps for Science Policy Makers



Iteration VI (2010)  
Science Maps for Scholars



Iteration VII (2011)  
Science Maps as Visual Interfaces to Digital Libraries



Iteration VIII (2012)  
Science Maps for Kids



Iteration IX (2013)  
Science Maps Showing Trends and Dynamics



Iteration X (2014)  
The Future of Science Mapping



2<sup>nd</sup> Decade (2015-2024)

## Macroscopes

Iteration XI (2015)  
Macroscopes for Interacting with Science



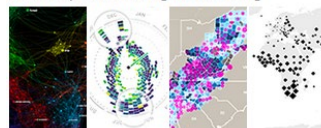
Iteration XIII (2017)  
Macroscopes for Playing with Scale



Iteration XII (2016)  
Macroscopes for Making Sense of Science



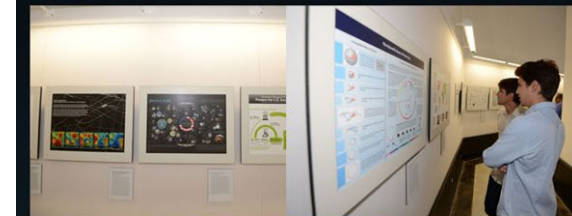
Iteration XIV (2018)  
Macroscopes for Ensuring our Well-being



3<sup>rd</sup> Decade (2015-2034)



101st Annual Meeting of the Association of American Geographers, Denver, CO. April 5th - 9th, 2005 (First showing of Places & Spaces)



University of Miami, Miami, FL. September 4 - December 11, 2014.



The David J. Sencer CDC Museum, Atlanta, GA. January 25 - June 17, 2016.



Duke University, Durham, NC. January 12 - April 10, 2015

<http://scimaps.org>

# The Structure of Science

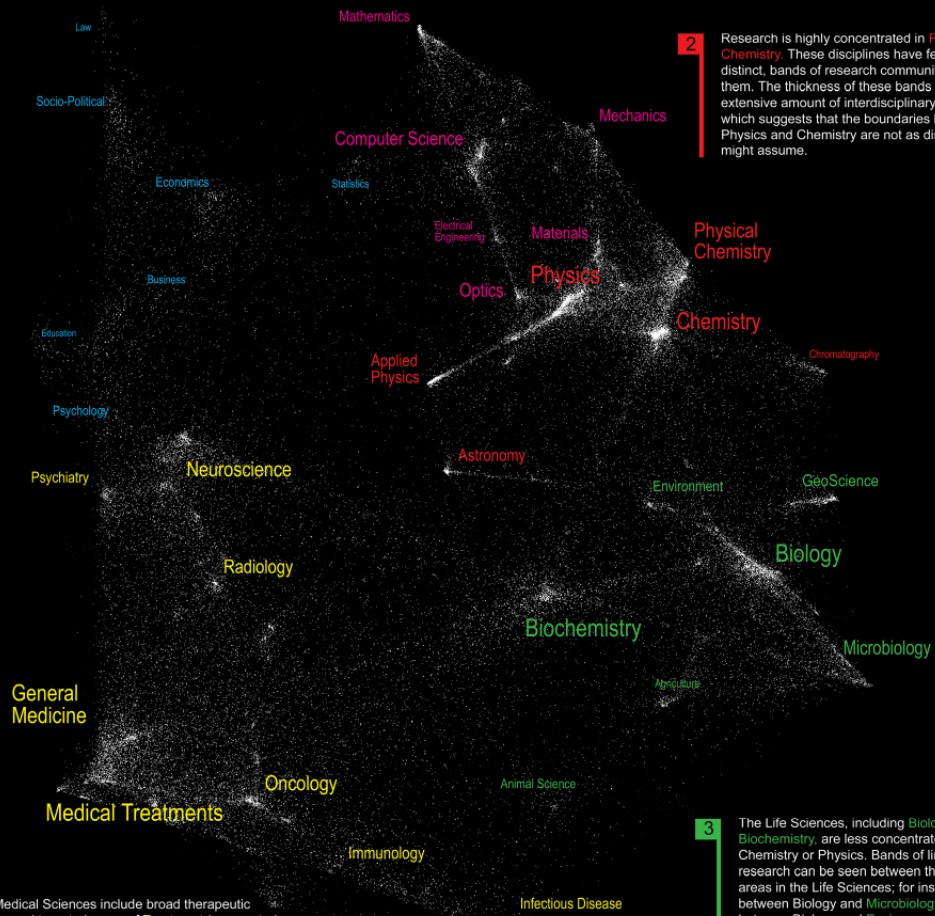
**5** The Social Sciences are the smallest and most diffuse of all the sciences. **Psychology** serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. **Statistics** serves as the link with Computer Science and Mathematics.

**1** **Mathematics** is our starting point, the purest of all sciences. It lies at the outer edge of the map. **Computer Science**, **Electrical Engineering**, and **Optics** are applied sciences that draw upon knowledge in Mathematics and Physics. These three disciplines provide a good example of a linear progression from one pure science (Mathematics) to another (Physics) through multiple disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communities that link them. Bands indicate interdisciplinary research.

**2** Research is highly concentrated in **Physics** and **Chemistry**. These disciplines have few, but very distinct, bands of research communities that link them. The thickness of these bands indicates an extensive amount of interdisciplinary research, which suggests that the boundaries between Physics and Chemistry are not as distinct as one might assume.

**3** The Life Sciences, including **Biology** and **Biochemistry**, are less concentrated than Chemistry or Physics. Bands of linking research can be seen between the larger areas in the Life Sciences; for instance between Biology and Microbiology, and between Biology and Environmental Science. Biochemistry is very interesting in that it is a large discipline that has visible links to disciplines in many areas of the map, including Biology, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.

**4** The Medical Sciences include broad therapeutic studies and targeted areas of **Treatment** (e.g. central nervous system, cardiology, gastroenterology, etc.) Unlike Physics and Chemistry, the medical disciplines are more spread out, suggesting a more multi-disciplinary approach to research. The transition into Life Sciences (via Animal Science and Biochemistry) is gradual.



We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

One of the first maps of science was developed at the Institute for Scientific Information over 30 years ago. It identified 41 areas of science from the citation patterns in 17,000 scientific papers. That early map was intriguing, but it didn't cover enough of science to accurately define its structure.

Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd) from the citation patterns in 800,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 96,000 research communities active in science in 2002. A research community is a group of papers (9 on average) that are written on the same research topic in a given year. Over time, communities can be born, continue, split, merge, or die.

The map of science can be used as a tool for science strategy. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore, exploit, abandon, or ignore.

We also envision the map as an educational tool. For children, the theoretical relationship between areas of science can be replaced with a concrete map showing how math, physics, chemistry, biology and social studies interact. For advanced students, areas of interest can be located and neighboring areas can be explored.

## Nanotechnology

Most research communities in nanotechnology are concentrated in **Physics**, **Chemistry**, and **Materials Science**. However, many disciplines in the Life and Medical Sciences also have nanotechnology applications.

## Proteomics

Research communities in proteomics are centered in **Biochemistry**. In addition, there is a heavy focus in the tools section of chemistry, such as **Chromatography**. The balance of the proteomics communities are widely dispersed among the Life and Medical Sciences.

## Pharmacogenomics

Pharmacogenomics is a relatively new field with most of its activity in **Medicine**. It also has many communities in **Biochemistry** and two communities in the Social Sciences.



# Science related Wikipedian ACTIVITY

This visualization explores the activity of science, math, and technology (SMT) related articles in the English-language Wikipedia (<http://en.wikipedia.org>). The central image shows 659,388 articles (circles). Overlaid is a 37 x 37 grid of relevant half-inch sized images.

Blue, green, and yellow circles represent the 3,599 math, 6,474 science, and 3,164 technology related articles respectively. The larger the size of a circle the higher the likelihood it is that type of article. The four corners show activity patterns of SMT articles.

**Article Edit Activity**  
Articles are size coded based on how frequently they have been edited from Feb. 6, 2001 to April 6, 2007. More consideration is given to current and major edits. Larger circles have been edited more frequently than smaller circles.

**2007 Major Edits**  
Articles are size coded based on how many major edits they received from January 1st, 2007 to April 6th, 2007. Larger circles have received more edits than smaller circles. The highest number of major edits was 2,627.

For the central image, each article is size coded based on the likelihood that it is math, science, or technology related.

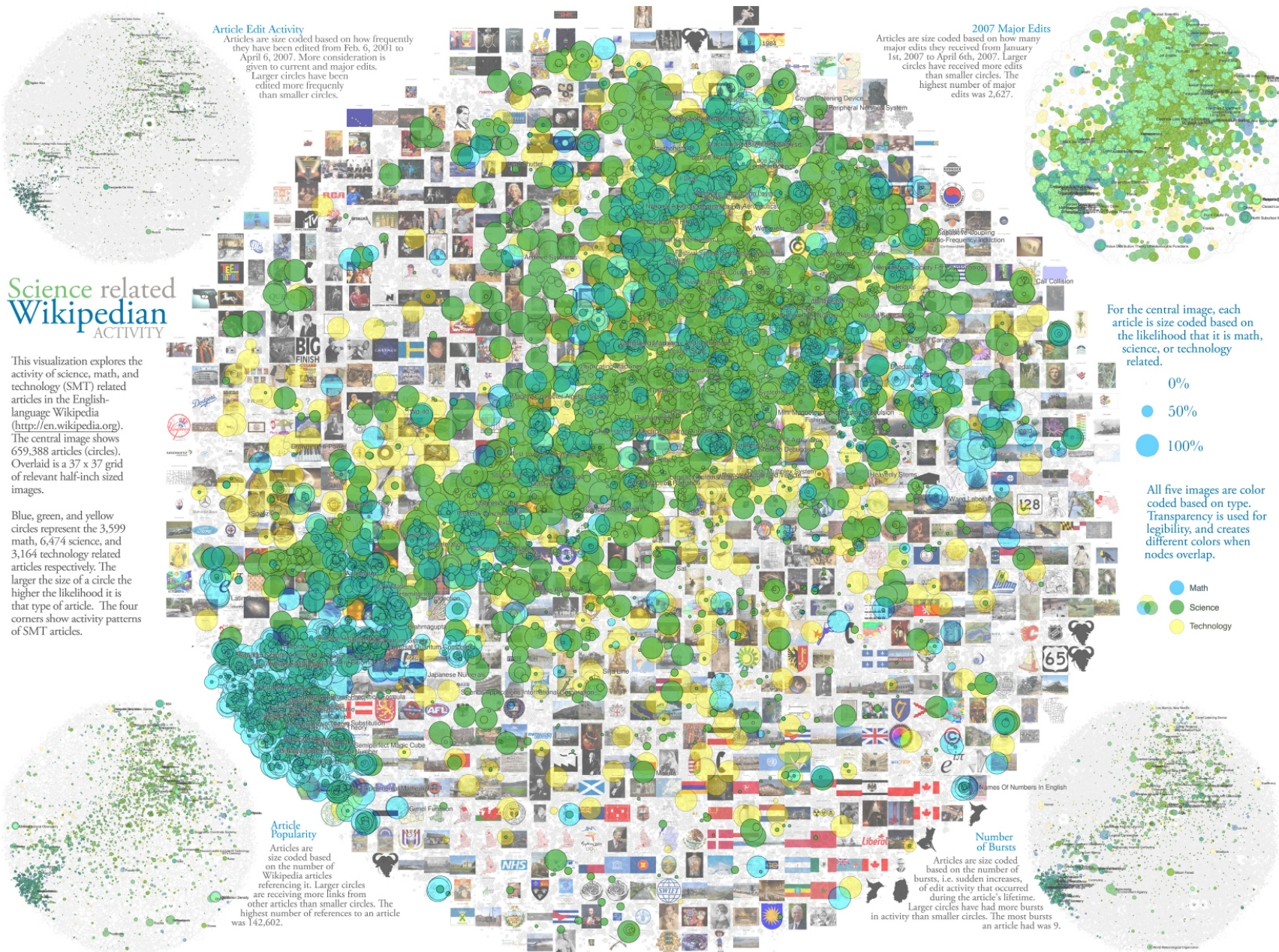
- 0%
- 50%
- 100%

All five images are color coded based on type. Transparency is used for legibility, and creates different colors when nodes overlap.

- Math
- Science
- Technology

**Article Popularity**  
Articles are size coded based on the number of Wikipedia articles referencing it. Larger circles are receiving more links from other articles than smaller circles. The highest number of references to an article was 142,602.

**Number of Bursts**  
Articles are size coded based on the number of bursts, i.e. sudden increases, of edit activity that occurred during the article's lifetime. Larger circles have had more bursts in activity than smaller circles. The most bursts an article had was 9.







# Diseasome

## The Human Disease Network

Explore online at <http://diseasome.eu>

### Statistics

# of Nodes: 516  
# of Edges: 1188  
Density: 0,0089  
Average Degree: 9,20  
Diameter: 15  
Average Shortest Path: 6,5

### Top 5 Diseases

1. Deafness
2. Leukemia
3. Colon Cancer
4. Retinitis Pigmentosa
5. Diabetes Mellitus

### Top 5 Genes

1. TP53
2. PAX6
3. FGFR2
4. RTN4
5. MSH2

### Description

The map presents a network of 516 diseases linked by 1188 known disorder-gene associations, indicating the common genetic origin of many diseases.

#### HOW TO USE THIS MAP

The map offers a rapid visual reference of the genetic links between disorders and a valuable global perspective for physicians, genetic researchers, and biomedical researchers alike. This new approach may lead to relatively efficient targeting to their affected genes, improve the understanding of the causes of disease, and the functions of particular genes.

#### NETWORK VISUALIZATION TECHNIQUES APPLIED

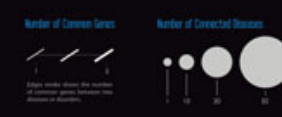
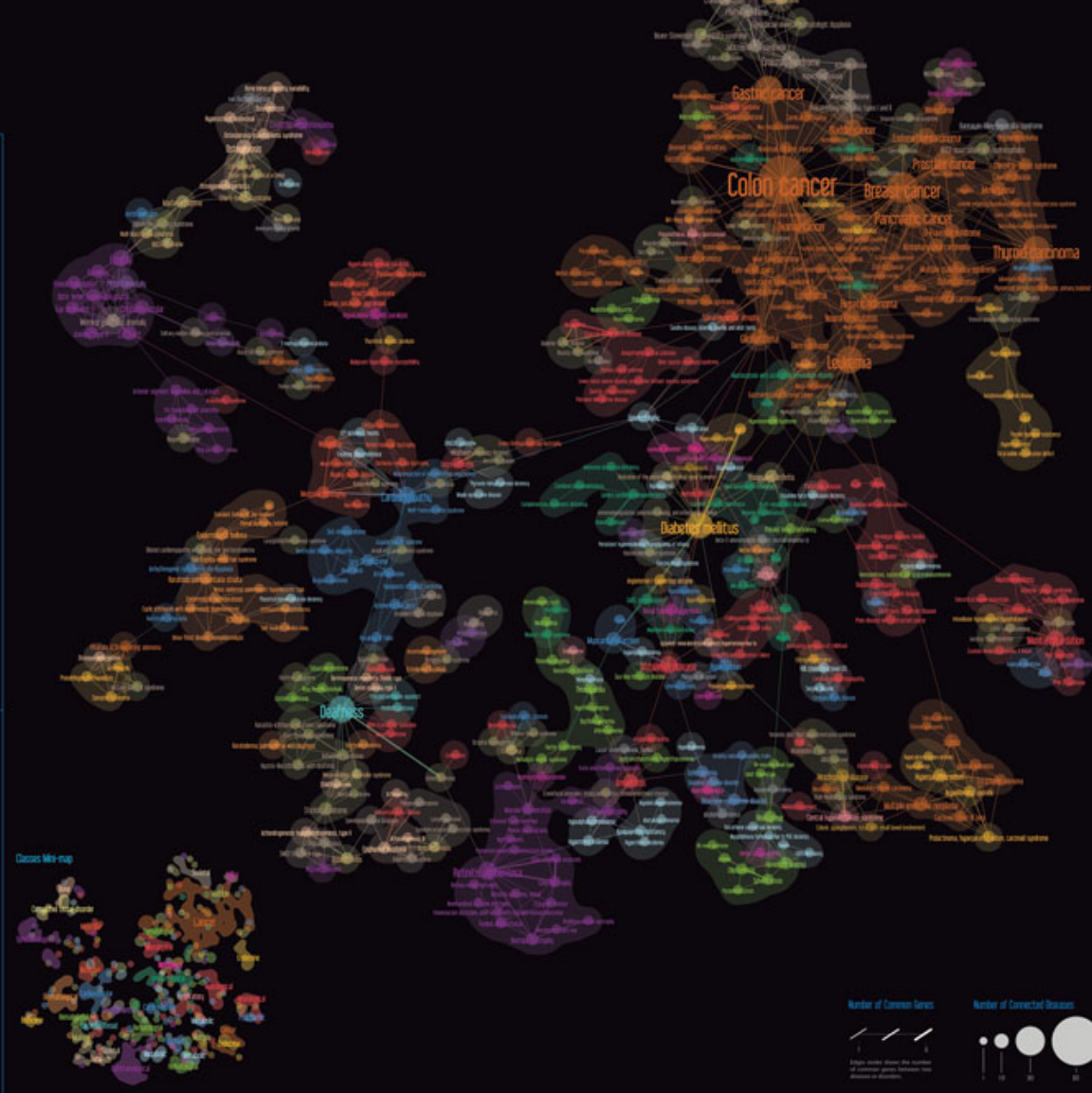
The map was done using the force-directed layout algorithm ForceAtlas2 in Gephi. Node color corresponds to the disorder class to which the disease belongs, and the size is proportional to its node degree, the overall number of links. Link's width is proportional to the number of genes that are involved in both diseases and colored with the average color between source and target nodes. Isolated diseases are not shown and only the great component has been kept. The Clusters filter map labels most remarkable disorder clusters and shows largest visual clusters.

The Disorder Class Interactions graph below shows the interaction level between disorder classes, representing the number of shared genes, up to 80.

#### NOTES

The network Disorder Network  
Bastin M, Heymann S, Valleron AJ, Chabot B, Toldu M, Barthelemy A-L (2017)  
PLoS Med 14(10): e1005465

### Disorder Class Interactions

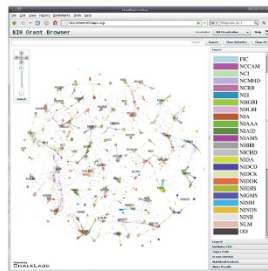




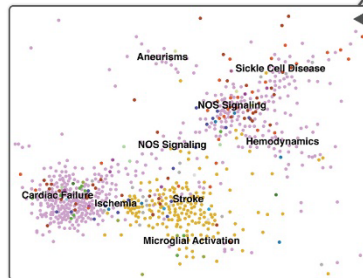
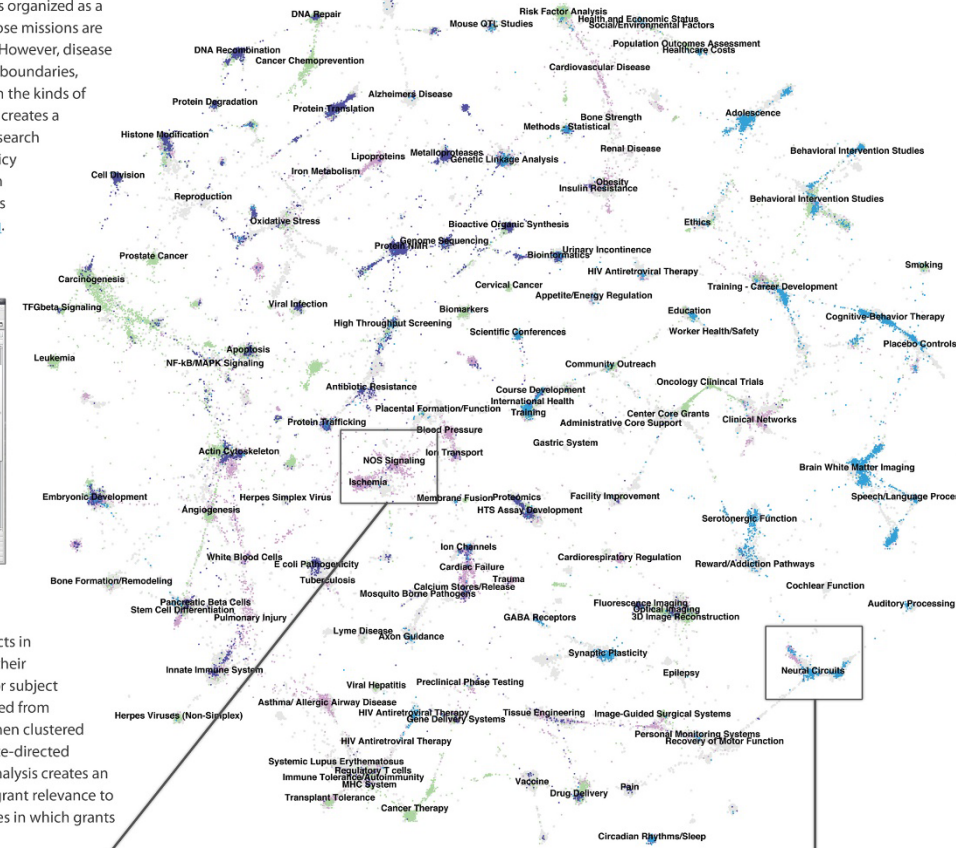
# A Topic Map of NIH Grants 2007

Bruce W. Herr II (Chalklabs & IU), Gully Burns (ISI), David Newman (UCI), Edmund Talley (NIH)

The National Institutes of Health (NIH) is organized as a multitude of Institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies flout scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each Institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is devised an interactive topic map for navigating this landscape, online at [www.nihmaps.org](http://www.nihmaps.org). Institute abbreviations can be found at [www.nih.gov/icd](http://www.nih.gov/icd).



Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analysis creates an interactive environment for assessing grant relevance to research categories and to NIH Institutes in which grants are localized.

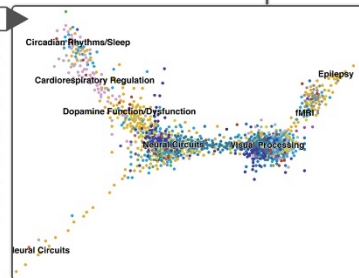


### Cardiac Diseases Research

An area of the map focused on cardiovascular function and dysfunction. Cardiac Failure (primarily funded by NHLBI) is typically clustered next to Stroke (NINDS), since these are the two major medical emergencies associated with ischemia, which results from a restricted blood supply. Also localized in this area are grants focused on Nitric Oxide (NOS) Signaling, a major biochemical pathway for vasodilation, and grants on Hemodynamics, Sickle Cell Disease, and Aneurysms.

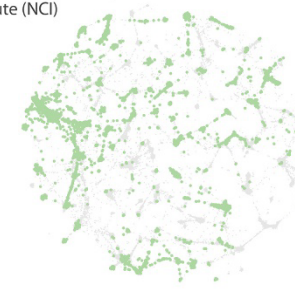
### Neural Circuits Research

An area of the map focused on neural circuits, which shows the diversity of topics and NIH Institutes that fund research in this area, such as: Cardiorespiratory Regulation, primarily funded by NHLBI; Visual Processing, primarily funded by NEI and Epilepsy, primarily funded by NINDS. For color coding, see legend in the upper-left inset.



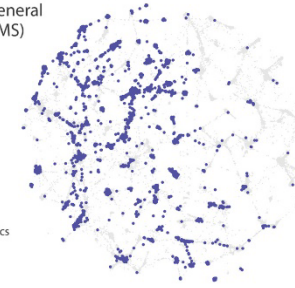
### National Cancer Institute (NCI)

- TOP 10 TOPICS
- 1 Oncology Clinical Trials
  - 2 Cancer Treatment
  - 3 Cancer Therapy
  - 4 Carcinogenesis
  - 5 Risk Factor Analysis
  - 6 Cancer Chemotherapy
  - 7 Metastasis
  - 8 Leukemia
  - 9 Prediction/Prognosis
  - 10 Cancer Chemoprevention



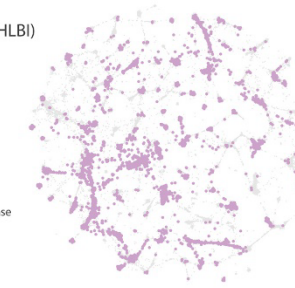
### National Institute of General Medical Sciences (NIGMS)

- TOP 10 TOPICS
- 1 Bioactive Organic Synthesis
  - 2 X-ray Crystallography
  - 3 Protein NMR
  - 4 Computational Models
  - 5 Yeast Biology
  - 6 Metalloproteases
  - 7 Enzymatic Mechanisms
  - 8 Protein Complexes
  - 9 Invertebrate/Zebrafish Genetics
  - 10 Cell Division



### National Heart, Lung, and Blood Institute (NHLBI)

- TOP 10 TOPICS
- 1 Cardiac Failure
  - 2 Pulmonary Injury
  - 3 Genetic Linkage Analysis
  - 4 Cardiovascular Disease
  - 5 Atherosclerosis
  - 6 Hemostasis
  - 7 Blood Pressure
  - 8 Asthma/ Allergic Airway Disease
  - 9 Gene Association
  - 10 Lipoproteins



### National Institute of Mental Health (NIMH)

- TOP 10 TOPICS
- 1 Mood Disorders
  - 2 Schizophrenia
  - 3 Behavioral Intervention Studies
  - 4 Mental Health
  - 5 Depression
  - 6 Cognitive-Behavior Therapy
  - 7 AIDS Prevention
  - 8 Genetic Linkage Analysis
  - 9 Adolescence
  - 10 Childhood



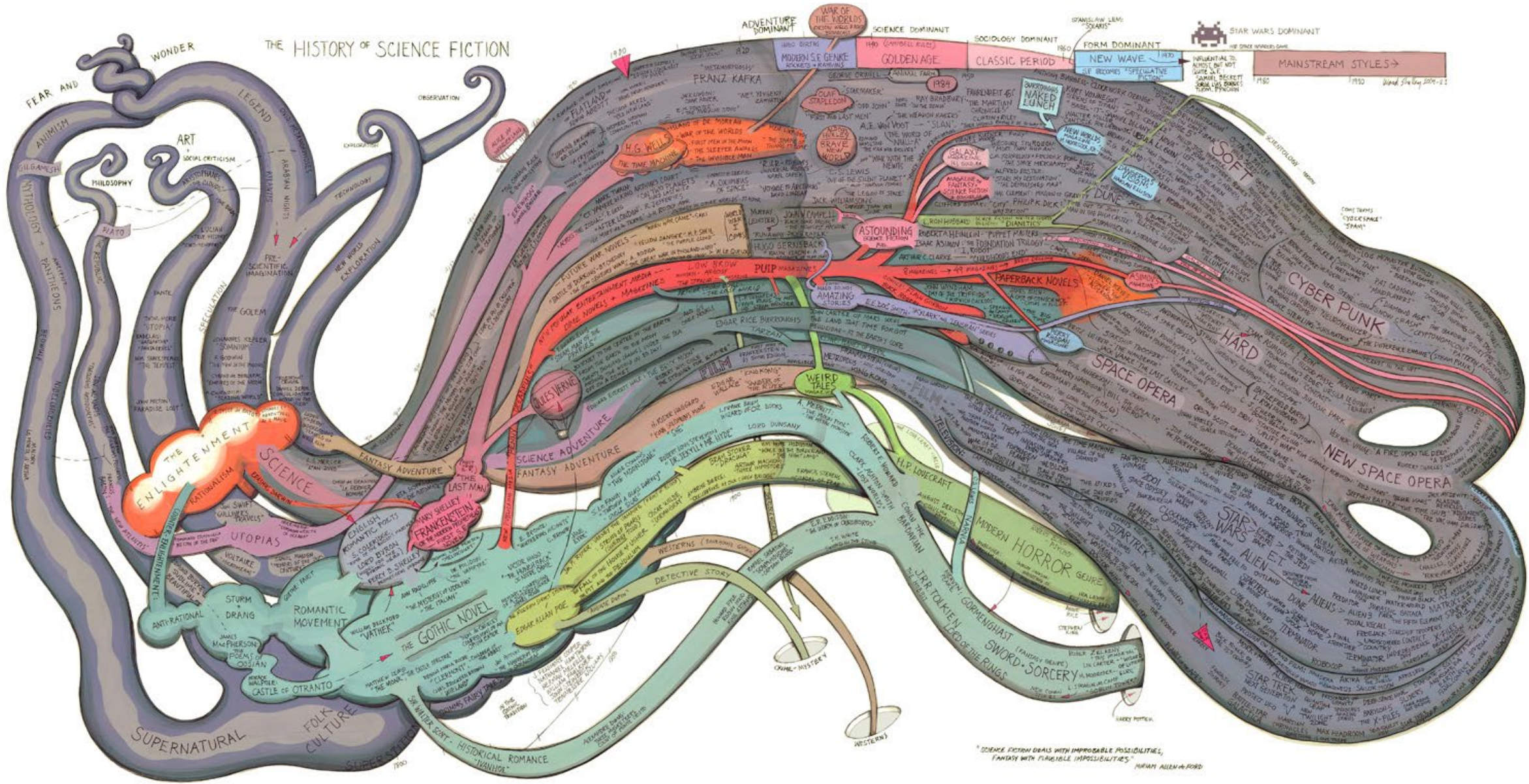


# Map of Scientific Collaborations from 2005-2009



Computed Using Data from Elsevier's Scopus





VII.10 History of Science Fiction - Ward Shelley - 2011



# Check out our **Zoom Maps** online!



**VII.10**  
History of Science Fiction, by Ward Shelley

BROOKLYN, NY, 2011  
Courtesy of Ward Shelley Studio

Ward Shelley is an artist identified with the Williamsburg scene in Brooklyn, New York, about art and culture. This map plots the science fiction literary genre from its nascence, emerging out of the data, here the narrative structure precedes and organizes the data. The monster whose tentacles are like trace roots to pre-historical sources and whose body is Romanticism, which birthed gothic fiction, source not only of Sci-Fi, but also of criticism, progressed through a number of distinct periods, which are charted, citing hundreds of authors and works.

**PLACES & SPACES**  
MAPPING IDEAS

Visit [scimaps.org](http://scimaps.org) and check out all our maps in stunning detail!

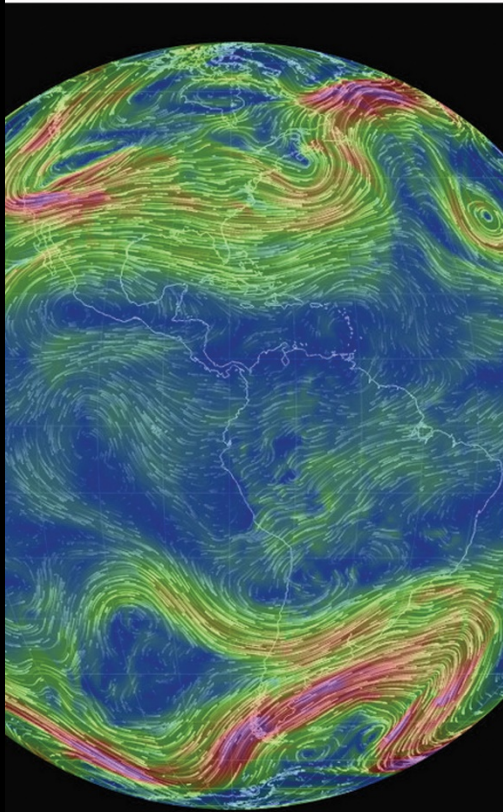




# MACROSCOPES FOR INTERACTING WITH SCIENCE



**PLACES &  
SPACES &**  
MAPPING SCIENCE  
scimaps.org



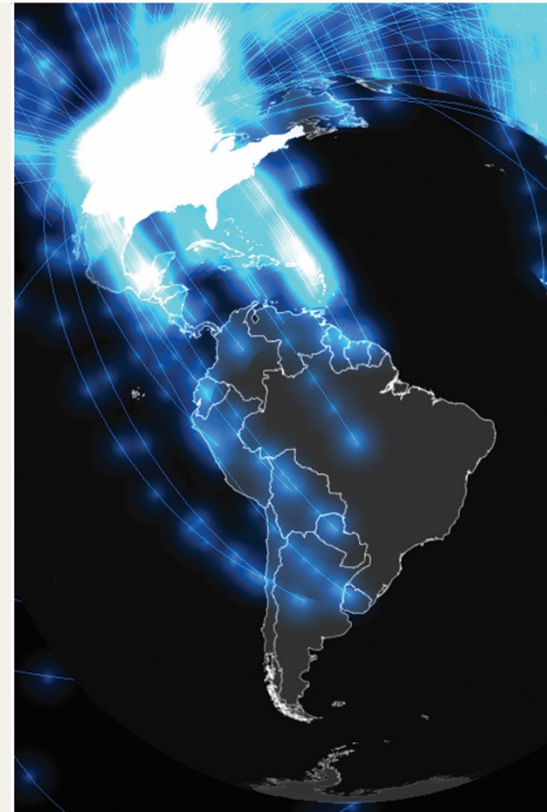
## Earth

*Weather on a worldwide scale*



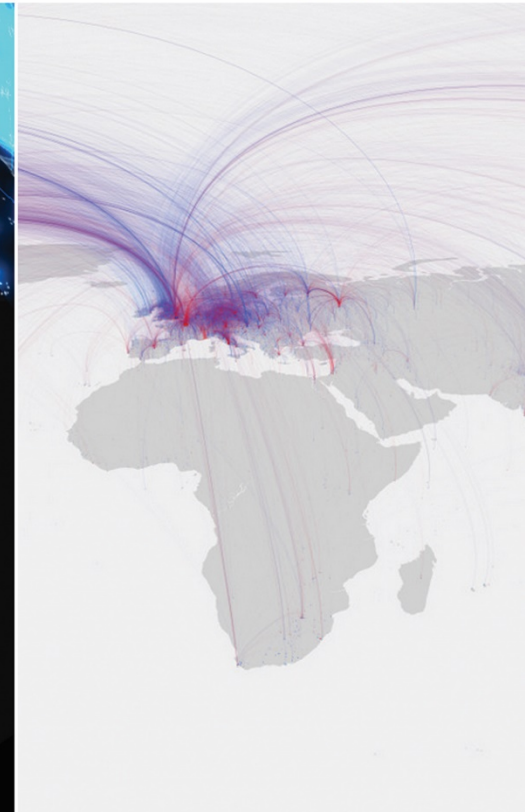
## AcademyScope

*Exploring the scientific landscape*



## Mapping Global Society

*Local news from a global perspective*



## Charting Culture

*2,600 years of human history in 5 minutes*

<http://idemo.cns.iu.edu/macroscope-kiosk>





# Modeling Science, Technology & Innovation Conference

WASHINGTON D.C. | MAY 17-18, 2016

[View Agenda](#)

Government, academic, and industry leaders discussed challenges and opportunities associated with using big data, visual analytics, and computational models in STI decision-making.

Conference slides, recordings, and report are available via <http://modsti.cns.iu.edu/report>







## Modeling and Visualizing Science and Technology Developments

National Academy of Sciences Sackler Colloquium, December 4-5, 2017, Irvine, CA

### Rankings and the Efficiency of Institutions

H. Eugene Stanley | Albert-László Barabási | Lada Adamic | Marta González | Kaye Husbands Fealing | Brian Uzzi | John V. Lombardi

### Higher Education and the Science & Technology Job Market

Katy Börner | Wendy L. Martinez | Michael Richey | William Rouse | Stasa Milojevic | Rob Rubin | David Krakauer

### Innovation Diffusion and Technology Adoption

William Rouse | Donna Cox | Jeff Alstott | Ben Shneiderman | Rahul C. Basole | Scott Stern | Cesar Hidalgo

### Modeling Needs, Infrastructures, Standards

Paul Trunfio | Sallie Keller | Andrew L. Russell | Guru Madhavan | Azer Bestavros | Jason Owen-Smith





## PROGRAMS

## Sackler Colloquia

- » About Sackler Colloquia
- » Upcoming Colloquia
- » Completed Colloquia
- » Sackler Lectures
- » Video Gallery
- » Connect with Sackler Colloquia
- » Give to Sackler Colloquia

## Cultural Programs

## Distinctive Voices

## Kavli Frontiers of Science

## Keck Futures Initiative

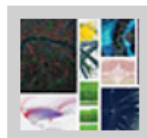
## LabX

## Sackler Forum

## Science &amp; Entertainment Exchange



## Modeling and Visualizing Science and Technology Developments



December 4-5, 2017; Irvine, CA

Organized by Katy Börner, H. Eugene Stanley, William Rouse and Paul Trunfio

### Overview

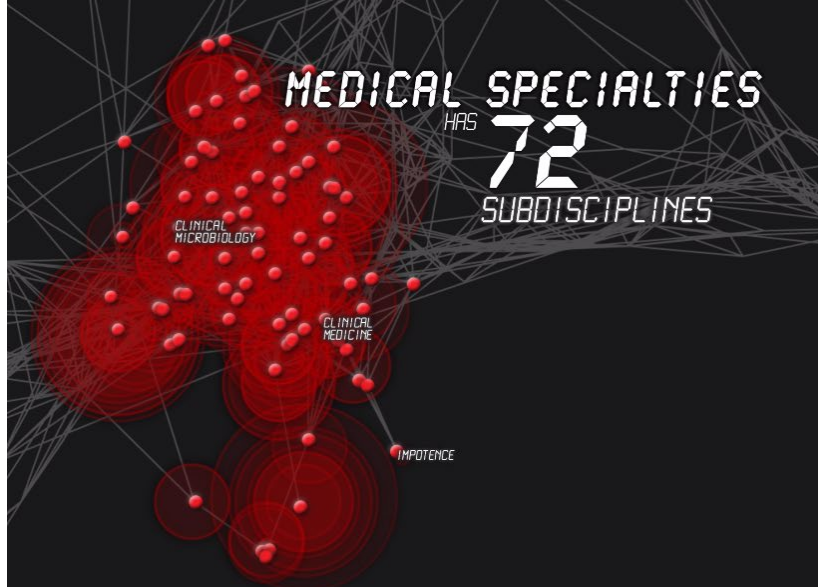
This colloquium was held in Irvine, CA on December 4-5, 2017.

This colloquium brought together researchers and practitioners from multiple disciplines to present, discuss, and advance computational models and visualizations of science and technology (S&T). Existing computational models are being applied by academia, government, and industry to explore questions such as: What jobs will exist in ten years and what career paths lead to success? Which types of institutions will likely be most innovative in the future? How will the higher education cost bubble burst affect these institutions? What funding strategies have the highest return on investment? How will changing demographics, alternative economic growth trajectories, and relationships among nations impact answers to these and other questions? Large-scale datasets (e.g., publications, patents, funding, clinical trials, stock market, social media data) can now be utilized to simulate the structure and evolution of S&T. Advances in computational power have created the possibility of implementing scalable, empirically validated computational models. However, because the databases are massive and multidimensional, both the data and the models tend to exceed human comprehension. How can advances in data visualizations be effectively employed to communicate the data, the models, and the model results to diverse stakeholder groups? Who will be the users of next generation models and visualizations and what decisions will they be addressing.

Videos of the talks are available on the [Sackler YouTube Channel](#).

<https://www.pnas.org/modeling>





# Science Forecast S1:E1





# Data Visualization Literacy Framework

Börner, Katy, Andreas Bueckle, and Michael Ginda. 2019. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *PNAS*, 116 (6) 1857-1864.



# Data Visualization Literacy (DVL)

Data visualization literacy (ability to read, make, and explain data visualizations) requires:

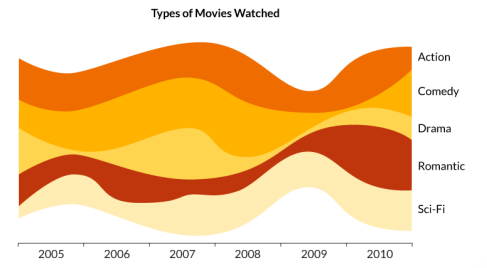
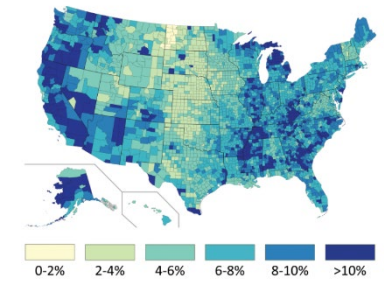
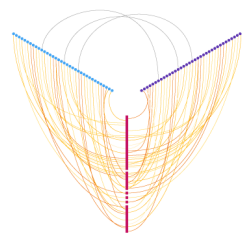
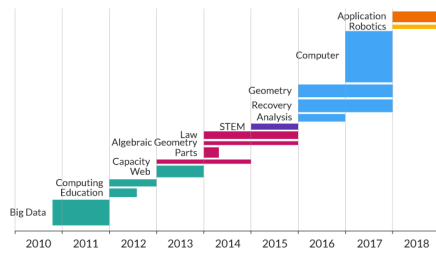
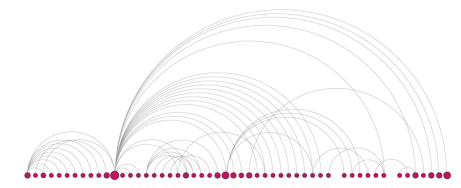
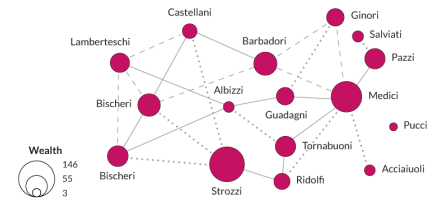
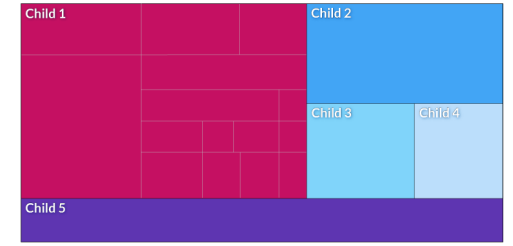
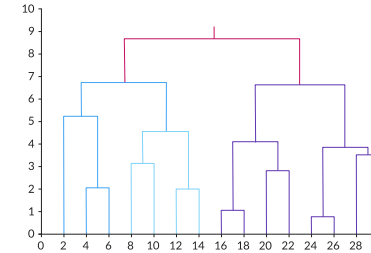
- literacy (ability to read and write text in titles, axis labels, legends, etc.),
- visual literacy (ability to find, interpret, evaluate, use, and create images and visual media), and
- mathematical literacy (ability to formulate, employ, and interpret math in a variety of contexts).

Being able to “read and write” data visualizations is becoming as important as being able to read and write text. Understanding, measuring, and improving data and visualization literacy is important to strategically approach local and global issues.

# Visualization Frameworks

MANY frameworks and taxonomies have been proposed to

- help organize and manage the evolving zoo of 500+ different data visualization types,
- provide guidance when designing data visualizations, and
- facilitate teaching.

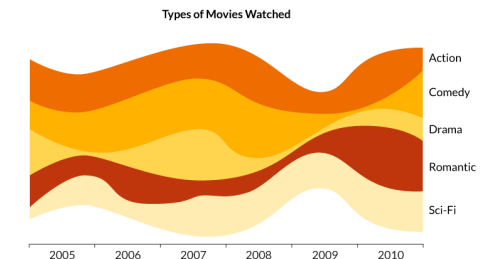
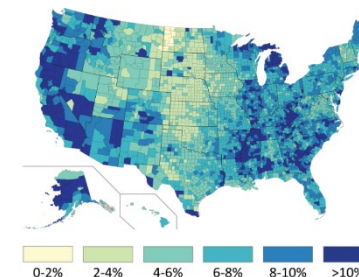
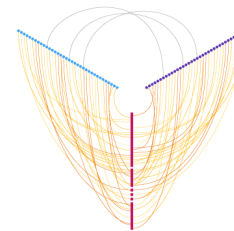
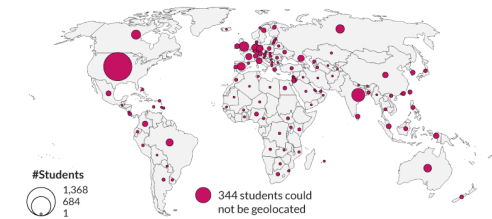
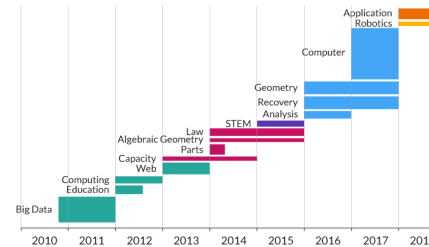
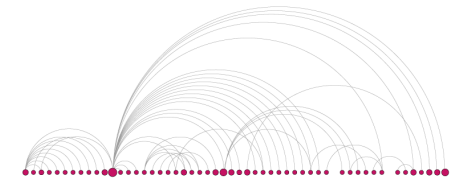
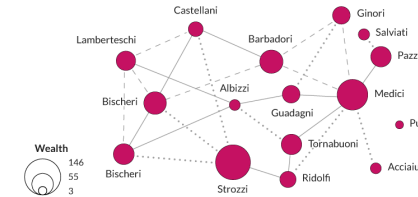
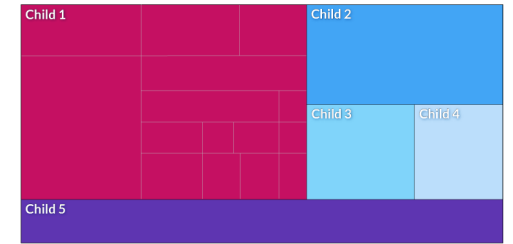
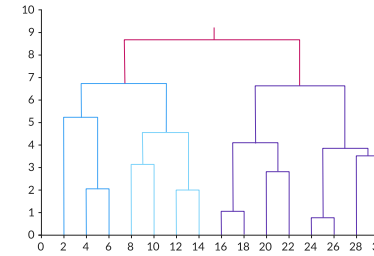




# Existing Visualization Frameworks

Organize data visualizations by

- User insight needs
- User task types
- Data to be visualized
- Data transformations
- Visualization technique
- Visual mapping transformations
- Interaction techniques
- Deployment options
- and other features ...



# DVL Framework: Desirable Properties

- Most existing frameworks focus on **READING**. We believe that much expertise is gained from also **CONSTRUCTING** data visualizations.
- Reading and constructing data visualizations needs to take human perception and cognition into account.
- Frameworks should build on and consolidate prior work in cartography, psychology, cognitive science, statistics, scientific visualization, data visualization, learning sciences, etc. in support of a de facto standard.
- Theoretically grounded + practically useful + easy to learn/use.
- Highly modular and extendable.



# DVL Framework: Development Process

- The initial DVL-FW was developed via an extensive literature review.
- The resulting DVL-FW typology, process model, exercises, and assessments were then tested in the *Information Visualization* course taught for more than 15 years at Indiana University. More than 8,500 students enrolled in the IVMOOC version (<http://ivmooc.cns.iu.edu>) over the last six years.
- The FW was further refined using feedback gained from constructing and interpreting data visualizations for 100+ real-world client projects.
- Data on student engagement, performance, and feedback guided the continuous improvement of the DVL-FW typology, process model, and exercises for defining, teaching, and assessing DVL.
- The DVL-FW used in this course supports the systematic construction and interpretation of data visualizations.

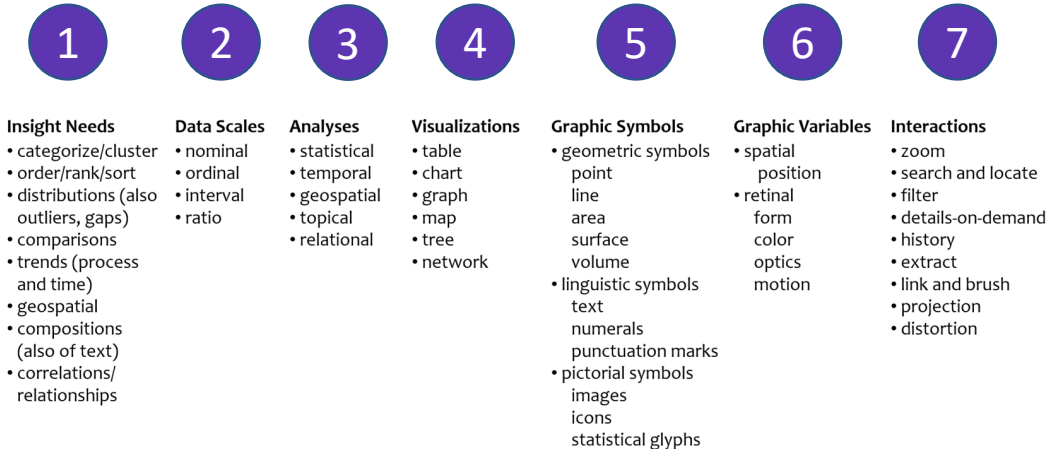


# Data Visualization Literacy Framework (DVL-FW)

Consists of two parts:

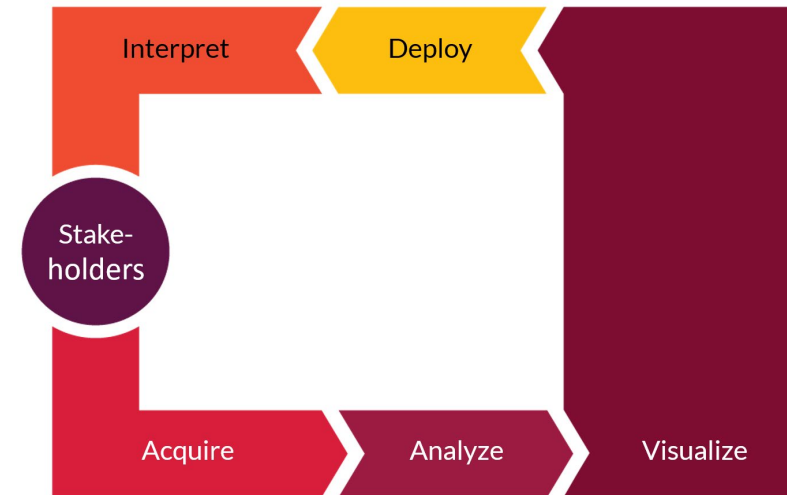
## DVL Typology

Defines 7 types with 4-17 members each.



## DVL Workflow Process

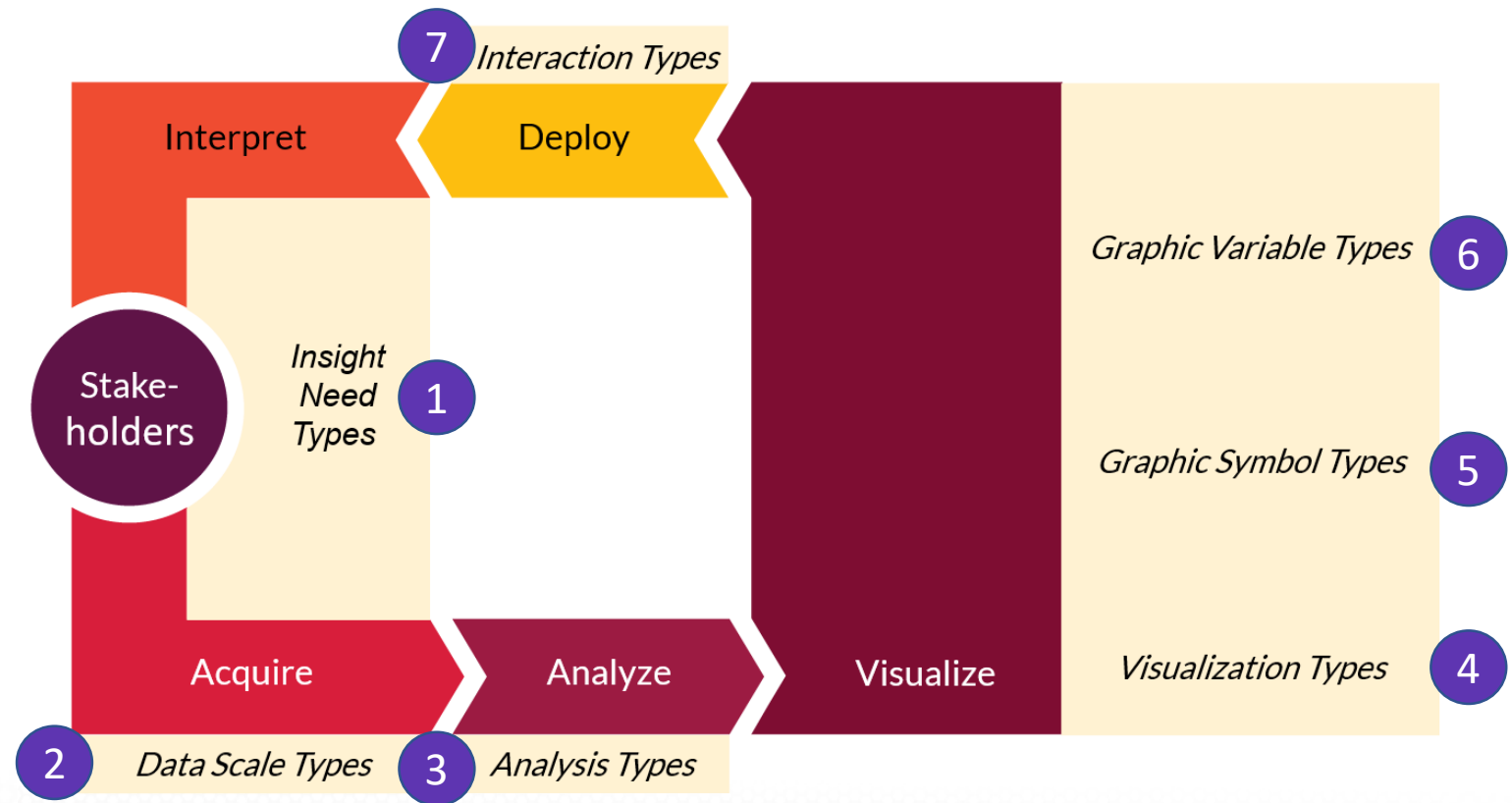
Defines 5 steps required to render data into insights.



# Data Visualization Literacy Framework (DVL-FW)

Consists of two parts *that are interlinked*:

**DVL Typology +  
DVL Workflow Process**





# Data Visualization Literacy Framework (DVL-FW)

Implemented in Make-A-Vis (MAV) to support learning via horizontal transfer, scaffolding, hands-on learning, etc.

The screenshot shows the Make-A-Vis interface with three main sections: Data, Make Visualization, and a visualization preview.

**Data Section:**

- ISI Publications: (CSV) Preprocessed-wos**

Title	Authors	Journal	Year	#Cites
[Progress bar]				

Total Records: 562
- Journals: (from ISI Publications)**

Name	#Papers	#Cites	First Year	Last Year
BMC EVOL BIOL	1	7	2006	2006
FEBS J	2	0	2005	2005
NAT PHYS	3	18	2005	2006

Total Records: 562

**Make Visualization Section:**

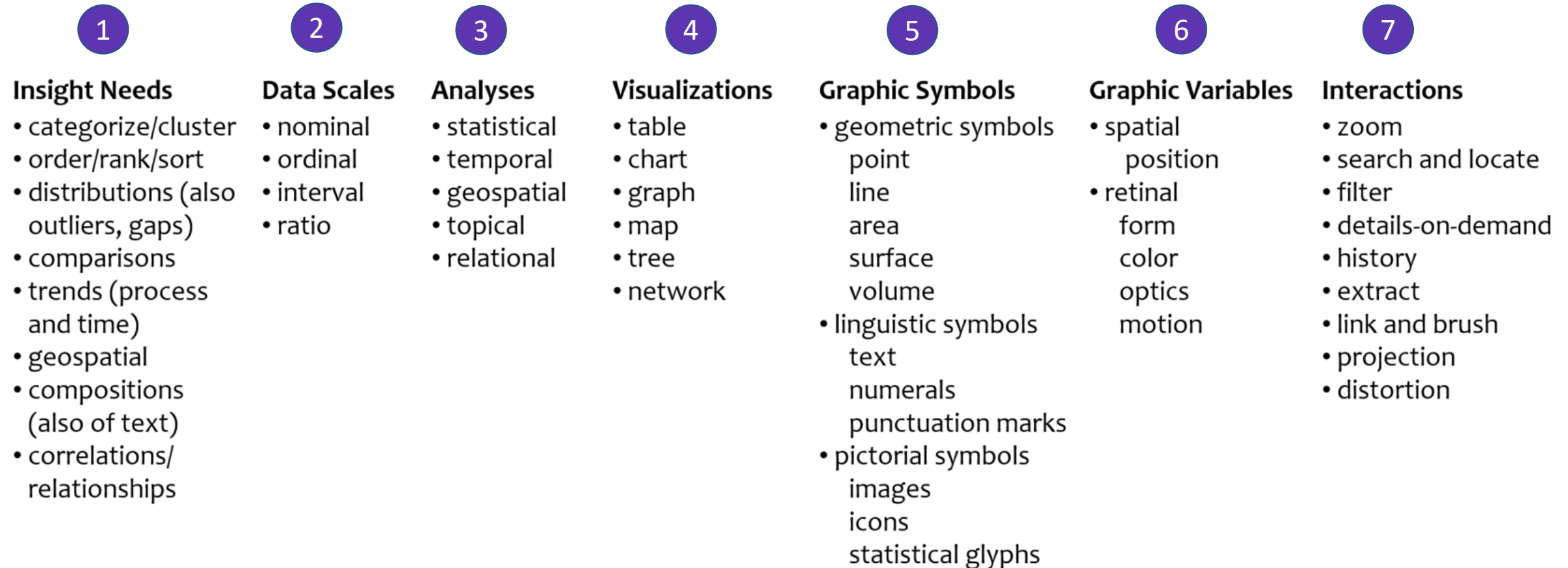
- Select Visualization Type:** Scatter Graph, Geomap, Scimap, Temporal Bar Graph (selected).
- Select Graphic Symbol Type(s):** (Dropdown menu)
- Select Graphic Variable Types:** (Dropdown menu)

**Temporal Bar Graph Preview:**

Temporal Bar Graph

Year	Machine	Big Data	Education	Building	Making	Computing	Web	Form	Smart	Capacity	Algebraic Geometry	Parts	Law	Stem	Analysis	Recovery	Geometry	Computer	Application	Robotics	
1998	1																				
1999		1																			
2000			1																		
2001				1																	
2002					1																
2003						1															
2004							1														
2005								1													
2006									1												
2007										1											
2008											1										
2009												1									
2010													1								
2011														1							
2012															1						
2013																1					
2014																	1				
2015																		1			
2016																			1		
2017																				1	

# Typology of the Data Visualization Literacy Framework



Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press. 25.



# Typology of the Data Visualization Literacy Framework

1

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 26-27.

Bertin, 1967	Wehrend & Lewis, 1996	Few, 2004	Yau, 2011	Rendgen & Wiedemann, 2012	Frankel, 2012	Tool: Many Eyes	Tool: Chart Chooser	Börner, 2014
selection	categorize			category				categorize/ cluster
order	rank	ranking					table	order/rank/ sort
	distribution	distribution					distribution	distributions (also outliers, gaps)
	compare	nominal comparison & deviation	differences		compare and contrast	compare data values	comparison	comparisons
		time series	patterns over time	time	process and time	track rises and falls over time	trend	trends (process and time)
		geospatial	spatial relations	location		generate maps		geospatial
quantity		part-to- whole	proportions		form and structure	see parts of whole, analyze text	composition	compositions (also of text)
association	correlate	correlation	relationships	hierarchy		relations between data points	relationship	correlations/ relationships



# Typology of the Data Visualization Literacy Framework

2

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press. 28-29.

# Data Scale Types

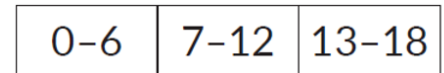
**Nominal:** A categorical scale, also called a nominal or category scale, is **qualitative**. Categories are assumed to be non-overlapping.



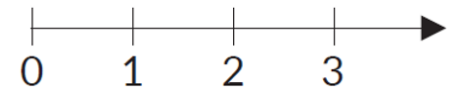
**Ordinal:** An ordinal scale, also called sequence or ordered, is **quantitative**. It rank-orders values representing categories based on some intrinsic ranking, but not at measurable intervals.



**Interval:** An interval scale, also called a value scale, is a **quantitative** numerical scale of measurement where the distance between any two adjacent values (or intervals) is equal, but the zero point is arbitrary.



**Ratio:** A ratio scale, also called a proportional scale, is a **quantitative** numerical scale. It represents values organized as an ordered sequence, with meaningful uniform spacing, and a true zero point.





# Data Scale Types - Examples

**Nominal:** Words or numbers constituting the “categorical” names and descriptions of people, places, things, or events.

**Ordinal:** Days of the week, degree of satisfaction and preference rating scores (e.g., using a Likert scale), or rankings such as low, medium, high.

**Interval:** Temperature in degrees or time in hours. Spatial variables such as latitude and longitude are interval.

**Ratio:** Physical measures such as height, weight, (reaction) time, or intensity of light; number of published papers, co-authors, citations.

# Data Scale Types - Examples

**Nominal:** Words or numbers constituting the “categorical” names and descriptions of people, places, things, or events.

Qualitative

**Ordinal:** Days of the week, degree of satisfaction and preference rating scores (e.g., using a Likert scale), or rankings such as low, medium, high.

Quantitative







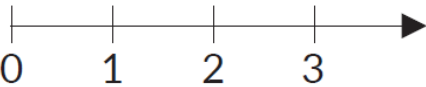
**Interval:** Temperature in degrees or time in hours. Spatial variables such as latitude and longitude are interval.

**Ratio:** Physical measures such as weight, height, (reaction) time, or intensity of light; number of published papers, co-authors, citations.



# Data Scale Types - Mathematical Operations

This table shows the logical mathematical operations permissible, the measure of central tendency, and examples for the different data scale types.

Data Scale Types	Logical Mathematical Operations				Measure of Central Tendency	Examples			
	= ≠	< >	+ -	x ÷					
Nominal	y				mode	  			
Ordinal	y	y			median	  			
Interval	y	y	y		arithmetic mean	<table border="1" data-bbox="1396 1043 1819 1115"> <tr> <td>0-6</td> <td>7-12</td> <td>13-18</td> </tr> </table>	0-6	7-12	13-18
0-6	7-12	13-18							
Ratio	y	y	y	y	geometric mean				

Qualitative

Quantitative

# Typology of the Data Visualization Literacy Framework

3

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press. 25.

# Analysis Types

- When: Temporal Data Analysis + Statistical
- Where: Geospatial Data Analysis
- What: Topical Data Analysis
- With Whom: Network Analysis



# Typology of the Data Visualization Literacy Framework

4

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

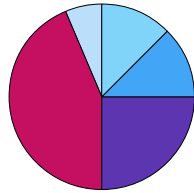
## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

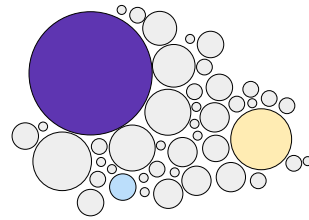
Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 30-31.

# Visualization Types

Chart

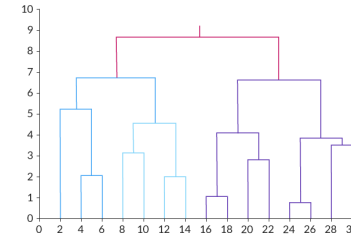


Pie Chart

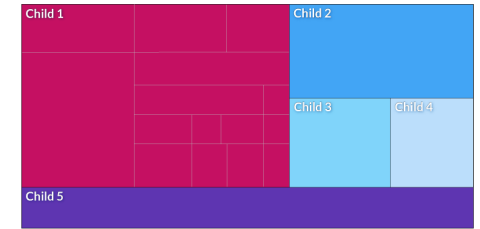


Bubble Chart

Tree

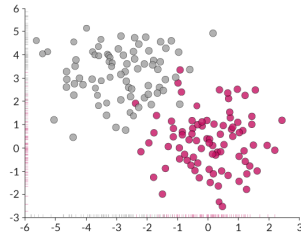


Dendrogram

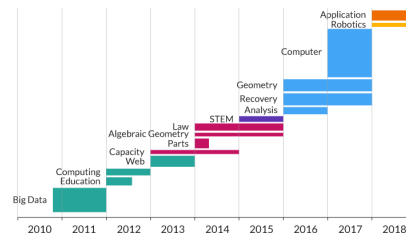


Tree Map

Graph

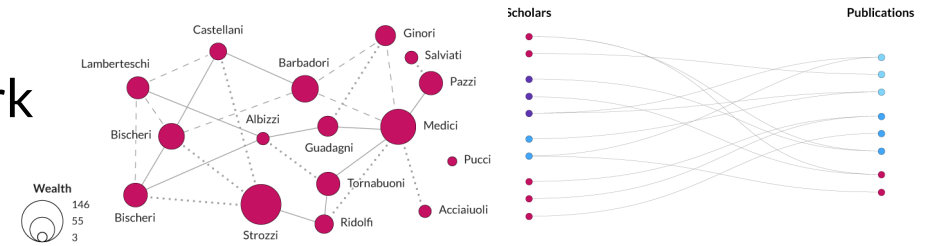


Scatter Graph



Temporal Bar Graph

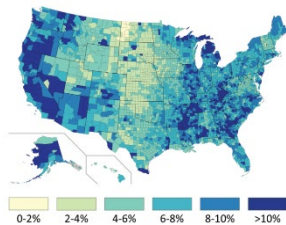
Network



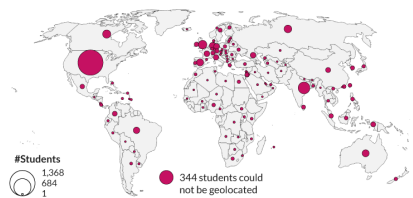
Dendrogram

Tree Map

Map



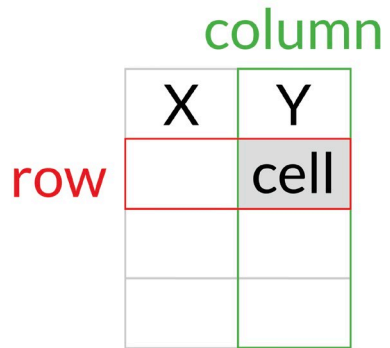
Choropleth Map



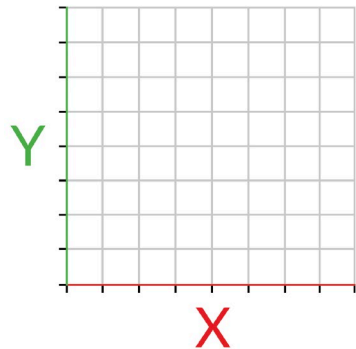
Proportional Symbol Map

# Visualize: Reference Systems

**Table**  
columns by rows



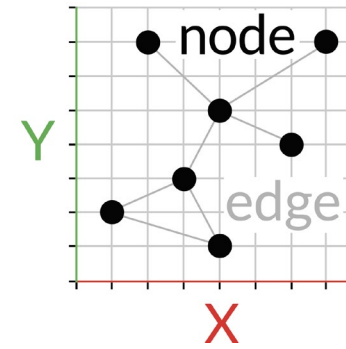
**Graph**  
x-y coordinates



**Map**  
latitude/  
longitude



**Network**  
local similarity



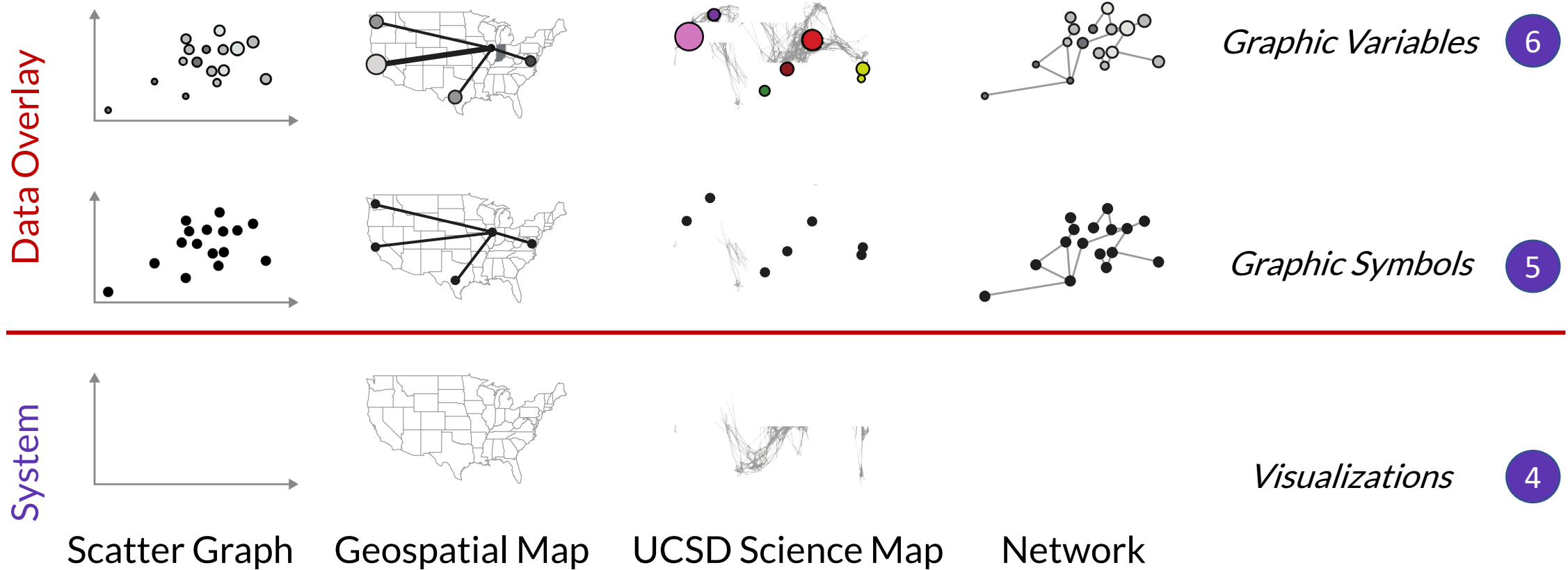
4

**Visualization Types**

- table
- chart
- graph
- map
- network layout



# Visualize: Reference Systems, Graphic Symbols and Variables



# Typology of the Data Visualization Literacy Framework

5

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 32-33.

# Typology of the Data Visualization Literacy Framework

6

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press. 34-35.



# Graphic Variable Types

**Position:** x, y; possibly z

Quantitative

**Form:**

- Size
- Shape
- Rotation (Orientation)

Quantitative

Qualitative

Quantitative

**Color:**

- Value (Lightness)



Quantitative

- Hue (Tint)



Qualitative

- Saturation (Intensity)



Quantitative

**Optics:** Blur, Transparency, Shading, Stereoscopic Depth

**Texture:** Spacing, Granularity, Pattern, Orientation, Gradient

**Motion:** Speed, Velocity, Rhythm

# Graphic Variable Types

**Position:** x, y; possibly z

**Form:**

- Size
- Shape
- Rotation (Orientation)

**Color:**

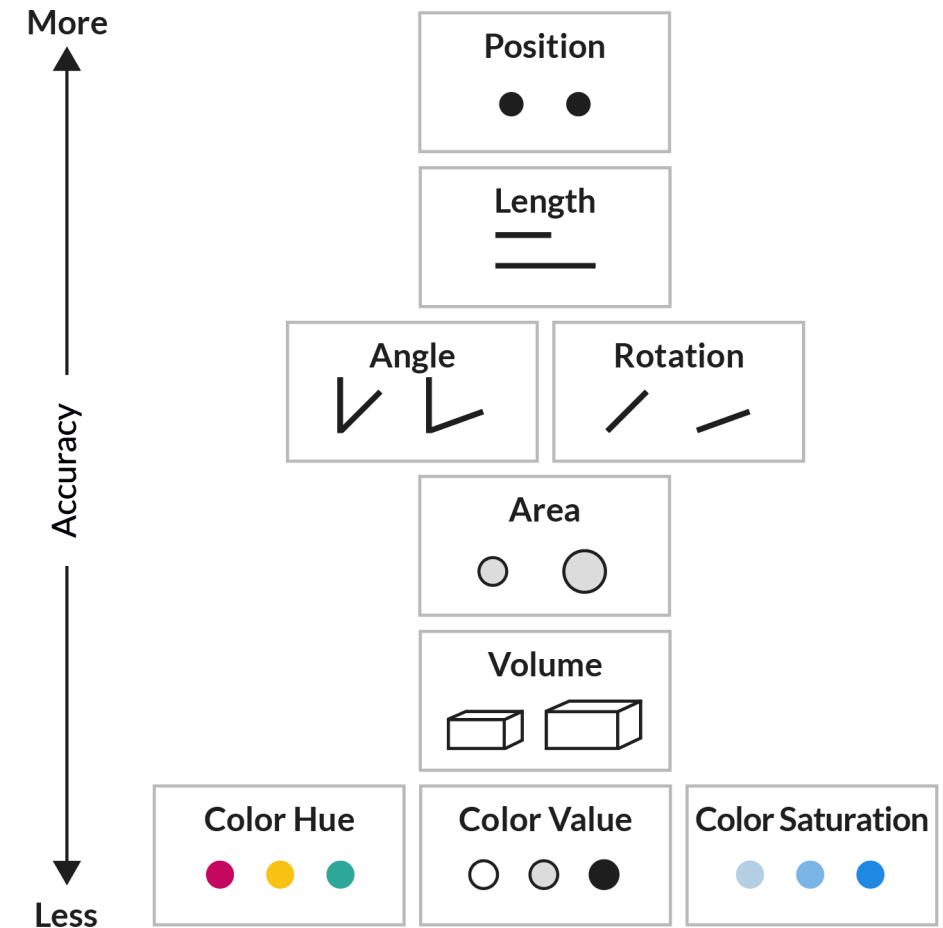
- Value (Lightness)
- Hue (Tint)
- Saturation (Intensity)



**Optics:** Blur, Transparency, Shading, Stereoscopic Depth

**Texture:** Spacing, Granularity, Pattern, Orientation, Gradient

**Motion:** Speed, Velocity, Rhythm



# Graphic Symbol Types

			Geometric Symbols		Linguistic Symbols	Pictorial Symbols
			Point	Line		
Spatial	Position	X Y				
		Retinal	Form	Size		
Shape					Text Text Text	
Color	Value				Text Text Text	
	Hue				Text Text Text	
	Saturation				Text Text Text	
Texture	Granularity					
	Pattern					
Motion Optics	Blur				Text Text Text	
	Speed					

Graphic Variable Types

See *Atlas of Knowledge* pages 36-39 for complete table.

**Qualitative**

Also called:  
Categorical Attributes  
Identity Channels

**Quantitative**

Also called:  
Ordered Attributes  
Magnitude Channels



# Graphic Variable Types Versus Graphic Symbol Types

			Geometric Symbols					Linguistic Symbols Text, Numerals, Punctuation Marks					Pictorial Symbols Images, Icons, Statistical Glyphs				
			Point	Line	Area	Surface	Volume										
Spatial	x	quantitative															
	y	quantitative															
	z	quantitative															
Form	Size	quantitative	NA (Not Applicable)														
	Shape	qualitative	NA														
	Rotation	quantitative	NA														
	Curvature	quantitative	NA														
	Angle	quantitative	NA														
	Closure	quantitative	NA														
	Value	quantitative															
	Hue	qualitative															
Color	Saturation	quantitative															
	Value	quantitative															
Texture	Spacing	quantitative															
	Granularity	quantitative															
	Pattern	qualitative															
	Orientation	quantitative	NA														
	Gradient	quantitative															
	Blur	quantitative															
	Transparency	quantitative															
	Shading	quantitative															
	Stereoscopic Depth	quantitative	Point in foreground .. background	Line in foreground .. background	Area in foreground .. background	Surface in foreground .. background	Volume in foreground .. background	Text in foreground .. background					Icons in foreground .. background				
	Speed	quantitative															
Motion	Velocity	quantitative															
	Rhythm	quantitative	Blinking point slow .. fast	Blinking line slow .. fast	Blinking area slow .. fast	Blinking surface slow .. fast	Blinking volume slow .. fast	Blinking text slow .. fast					Blinking icons slow .. fast				

See *Atlas of Knowledge* pages 36-39 for complete table.

# Typology of the Data Visualization Literacy Framework

7

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

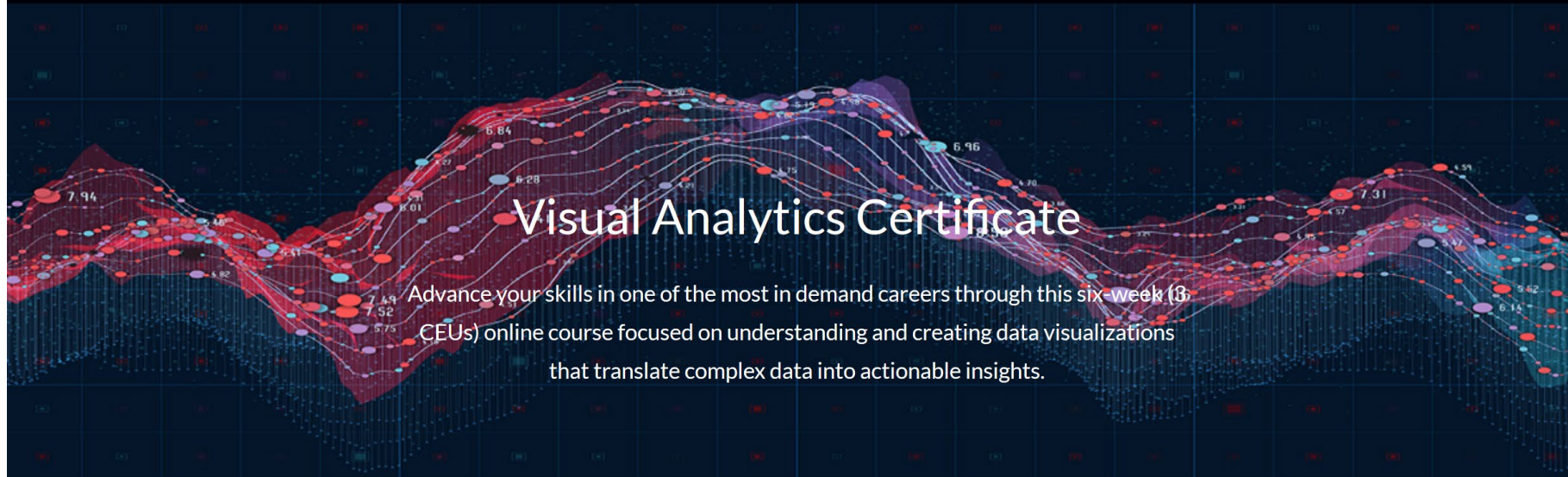
## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press. 26, 68-69.



# Visual Analytics Certificate

Advance your skills in one of the most in demand careers through this six-week (3 CEUs) online course focused on understanding and creating data visualizations that translate complex data into actionable insights.

DOWNLOAD FLYER

REGISTER FOR MAR 25-MAY 3, 2019



### Learn from Experts

Connect with industry professionals and leading researchers.



### Evolve Yourself

Gain forever knowledge and skill-up in powerful data visualization tools.



### Make a Difference

Embrace data-driven decision-making in your personal and professional life.

<https://visanalytics.cns.iu.edu>



# High-Resolution, Functional Mapping of Voxel, Vector, and Meta Datasets

within  
**Human BioMolecular Atlas Program (HuBMAP)**

<https://commonfund.nih.gov/HuBMAP>



# High-Resolution, Functional Mapping of Voxel, Vector, and Meta Datasets within the [Human BioMolecular Atlas Program \(HuBMAP\)](#)

MC-Indiana Team:

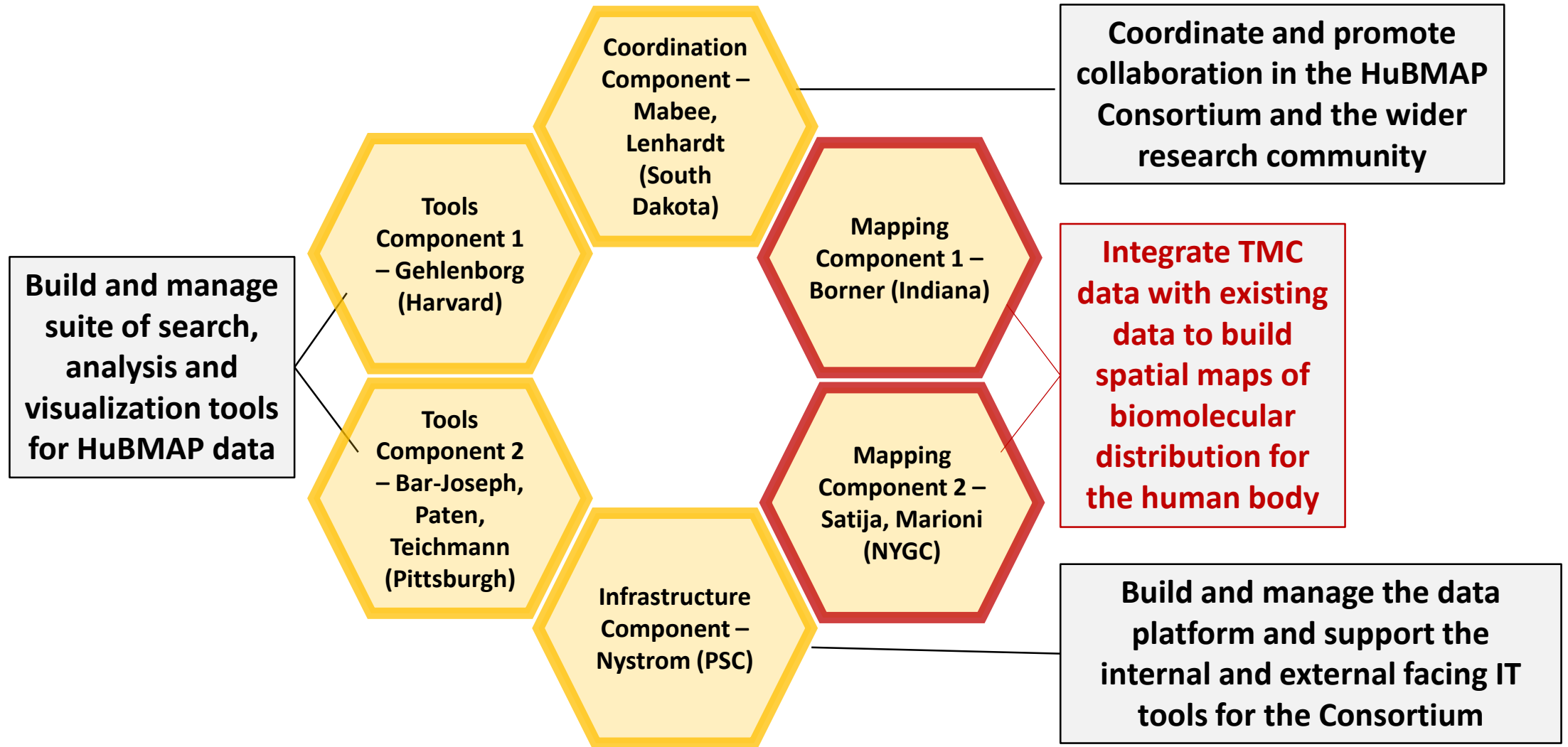
**Katy Börner, Bruce Herr II, Paul Macklin, Randy Heiland & Ellen Quardokus**

Intelligent Systems Engineering, SICE, Indiana University, Bloomington, IN

**Griffin Weber**, Harvard Medical School, Boston, MA

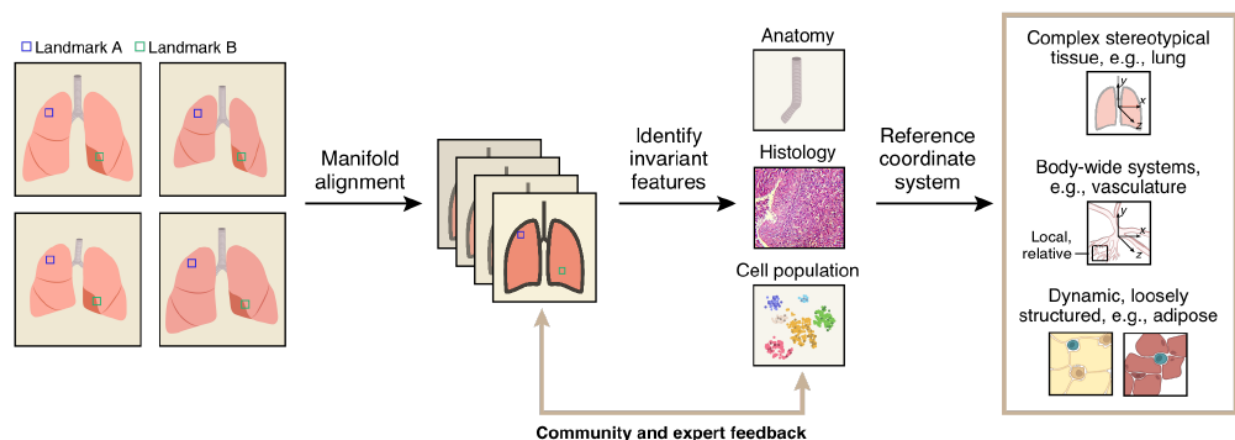
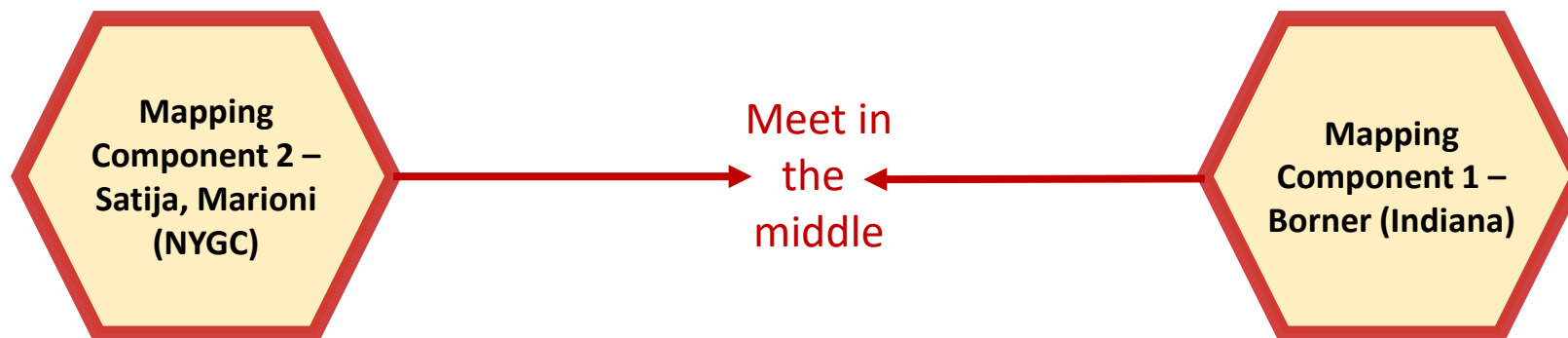
**Samuel Friedman**, Opto-Knowledge Systems, Inc.

# HuBMAP: HIVE

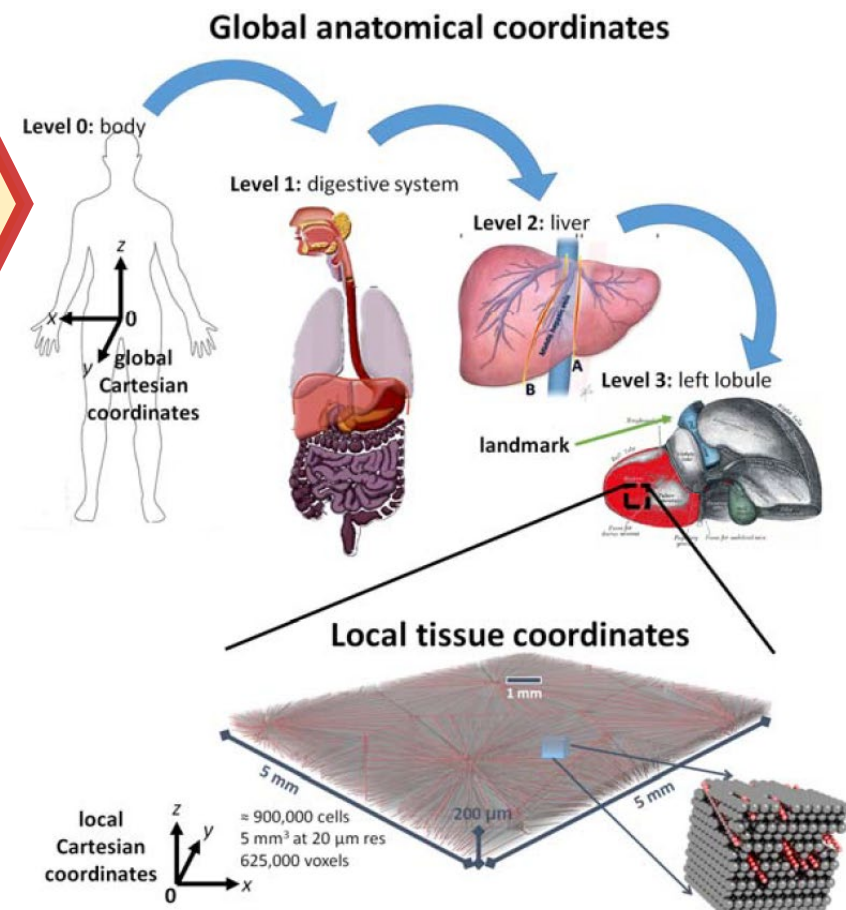




# HuBMAP: HIVE Mapping Components (MC)



**Figure 1: Overview of MC strategy to iteratively generate a reference coordinate system.** We will work with TMC to annotate initial datasets with key features and ontologies, which will serve as ‘landmarks’ to align images across individuals. We will modify and adapt our strategy to diverse tissues, retaining a probabilistic framework that represents uncertainty due to measurement, and inter-individual variation.



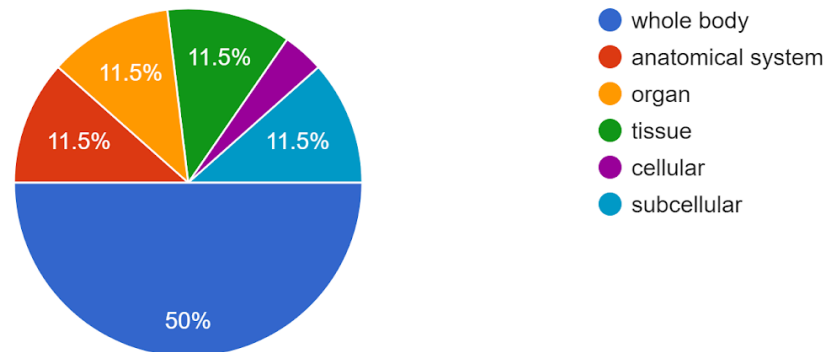
**Fig. 5. CCF concept**, navigating through the global anatomical coordinate system to insert a synthetic tissue sample (from PhysiCell<sup>4</sup>) into the left liver lobule with a local coordinate system.

# User Needs Analysis: Gathering User Stories

- Data was collected during the HIVE kickoff meeting held October 11-12, 2018.
- 26 user stories were provided from NIH and HIVE award project participants
- Data collection took place via a Google Form.
- Please note that many stakeholders are not represented here. Additional data collection continues.

At what level would you like to enter the map:

26 responses



*As a computational biologist, I would like to use HubMAP to integrate data across platforms, modalities, and organs so that I can*

*genes,  
ell  
common*

*As a cell biologist, I would like to use HubMAP to analyze data so that I can develop research in individual labs*

# Mapping Components (MC): Spatial maps of biomolecular data

Given anatomical and molecular data, develop and validate

## 1.) Terminologies/Ontologies (Semantics)

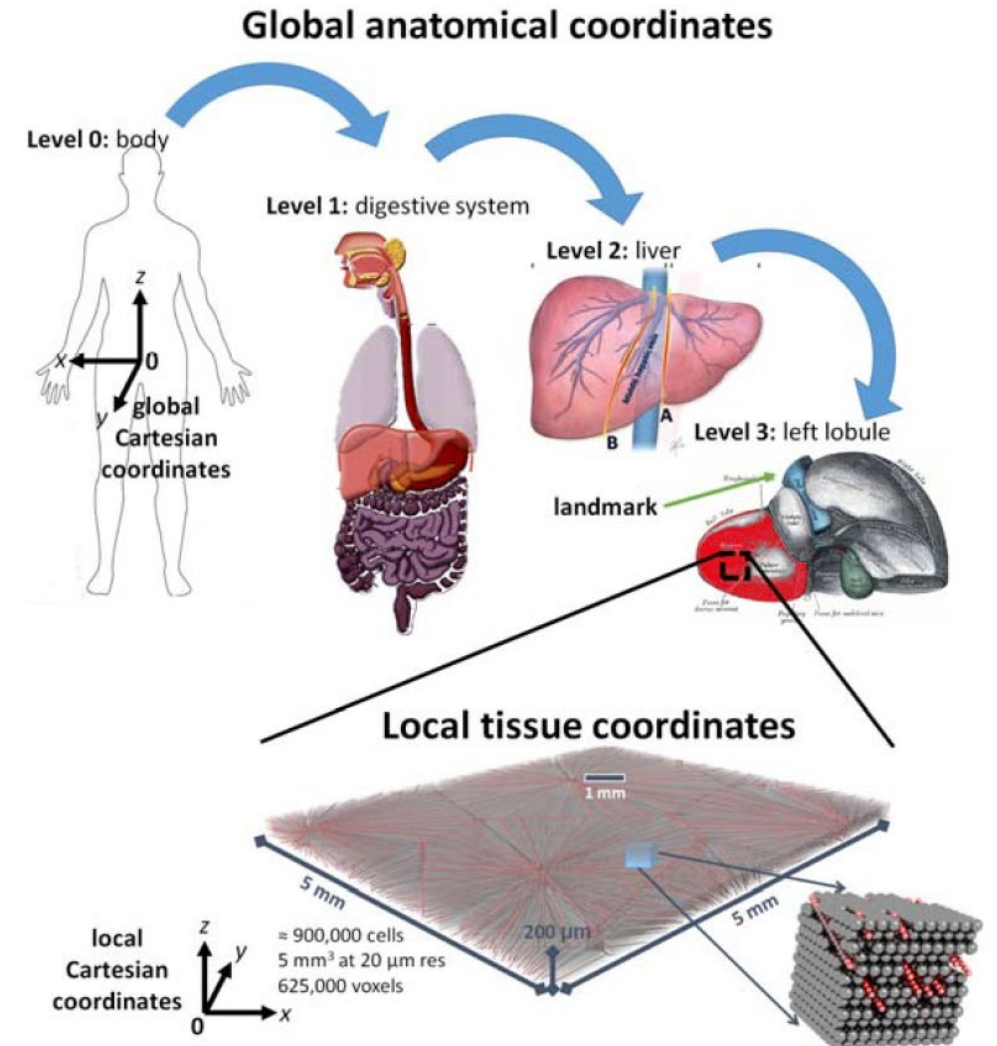
- Reference concepts, e.g., organs, organ parts, cell types, cell states
- Fiduciary concepts: Well defined landmarks that can be provided by TMCs and used by MC to spatially orient data with respect to 3D structures

## 2.) 3D Spatial Models interlinked with terminology/ontology

- Across levels (gross anatomy/organ, tissue, cell level) using hierarchical containment to localize the sample within the body
- Make landmarks visible in 3D models

## 3.) Interface for semantic and spatial search, filter, review, download of data.

- Use ontology for query expansion (semantic search), semantic browsing, and as controlled vocabulary (e.g., turning on/off male/female or different cell states).
- Use 3D models for spatial browsing, confirmation of proper tissue registration, exploring cell context, etc.



**Fig. 5. CCF concept**, navigating through the global anatomical coordinate system to insert a synthetic tissue sample (from PhysiCell<sup>4</sup>) into the left liver lobe with a local coordinate system.

# 1.) Terminologies/Ontologies (Semantics)

## Goals/Needs

- Create an ontologically based coordinate system, i.e., semantic terminology/ontology works with a 3D spatial model
- Respect existing ontologies and data standards that are machine readable, e.g., UBERON or FMA. NIH also has Common Data Elements (CDE) and other standards for clinical data, but elements need structure and to be machine readable.
- Metadata: Aim to unify across TMCs with common terms and usage
- Landmarks data: Identify what annotations are feasible for different organs/TMCs
- Molecular “image” data: Work with any any interoperable data standards (X, Y data; slice # or Z; possibly time; many well defined “color” channels), e.g., OMETiff (see work by Data Science WG)

## Challenges & Possible Solutions

- Using massive ontologies—Create multiple “slim” ontologies or subsets of ontologies (an approach used in GO), annotate terms with “slim” annotations, and merge the slim ontologies for our ontology.
- Many ontologies not suitable for UI navigation—Use full ontology for query expansion/filter but show only a “slim” version containment/navigation-tree in user interface
- Ever expanding list of terms to include—Create software to accommodate automatically growing term list. Ontology should work with tools like <https://www.mbfbioscience.com/tissue-mapper> to permit ontology and 3D spatial model integration; will need panel of anatomists/pathologists to agree on names for new 3D structures discovered by humans or via ML



# 1.) Terminologies/Ontologies (Semantics) cont.

## Relevant Ontologies

### Anatomic/Phenotypic

- Uberon
- Foundational Model of Anatomy (FMA) (has anatomical terms NOT in Uberon)
- Human Phenotype Ontology (HPO)
- Phenotype and Trait Ontology (PATO)
- Organ specific: Kidney Tissue Atlas Ontology (KTAO) and LungMAP

### Tissue/Data Collection

- Biological Spatial Ontology (BSPO)
- Ontology of Biomedical Investigations (OBI)
- EDAM (Bioinformatics concepts)

### (Sub-)Cellular

- Cell Ontology (CL)
- Gene Ontology (GO)
- Chemical Entities of Biological Interest (ChEBI)
- RNA Ontology (RNAO)
- Protein Ontology (PR)
- Cell Behavior Ontology (CBO)

### Metadata







- Basic Formal Ontology (BFO)
- Information Artifact Ontology (IAO)
- Ontology of units of Measure (OM)
- Provenance, Authoring and Versioning ontology (PAV)
- VIVO (Identifying researchers)

### MeSH and NCI Thesaurus

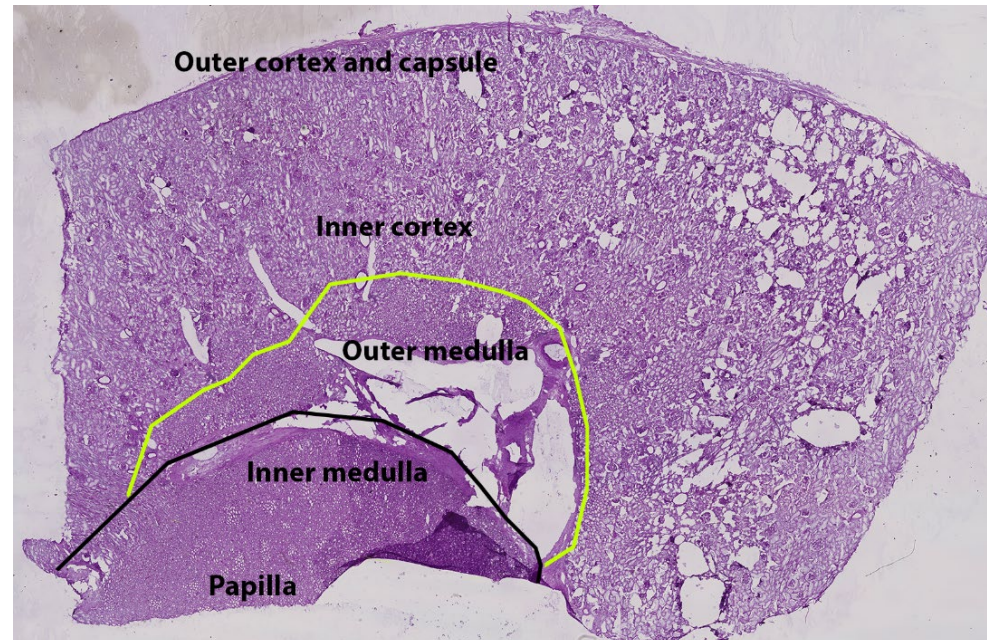
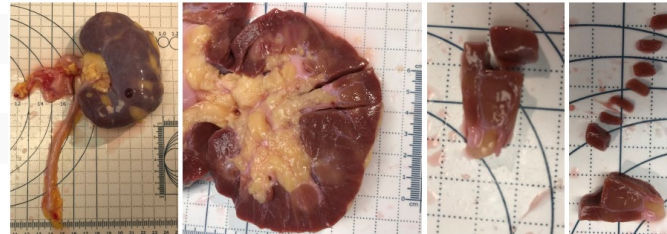
# 2.) 3D Spatial Models interlinked with terminology/ontology

**Kidney: Jeff Spraggins et al., VU**

See data on Globus, BIOMIC\_patient-64354

-  BIOMIC\_patient-64354\_clinical\_and\_spatial\_metadata.xlsx
-  BIOMIC\_patient-64354\_data\_guide.pptx
-  BIOMIC\_patient-64354\_overview.png
-  BIOMIC\_patient-64354\_Sample-20-Histology.tif
-  neg\_ion\_mode\_section
-  pos\_ion\_mode\_section

BUKMAP, Zhang Group



**Heart: Shin Lin, UW**

Year 1: Tissue data for 1-2cm cubed volumes from 9 sites for 1 heart from 1 individual.

Terminology; Coordinates and photos to spatialize



<u>Sites</u>	<u>Distinctive features</u>
1. LV, apex	
2. LV, free wall 3 cm from apex	
3. septum, 3cm from apex including LAD	major arterial vessel, Purkinje fiber CM
4. RV, free wall 3 cm from apex	
5. RA appendage	
6. RA, SA node to AV node	pacemaker CM
7. LA, appendage	
8. LA, PV inflow	
9. Posterior, adjacent to coronary sinus	major venous vessel

**Sternocostal surface labels:** Sinuatrial (SA) nodal branch, Atrial branch of right coronary a., Right coronary a., Anterior cardiac vv., Small cardiac v., Right (acute) marginal branch of right coronary a., Interventricular septal branches, Aorta, Left, Cir, Great cardiac v., Anterior interventricular branch (left anterior descending) of left coronary a.

**Diaphragmatic surface labels:** Oblique v. of left atrium (of Marshall), Great cardiac v., Circumflex branch of left coronary a., Left marginal branch, Coronary sinus, Posterolateral a., Middle cardiac v., Interventricular septal branches, Sinuatrial (SA) nodal branch, Small cardiac v., Right coronary a., Inferior (posterior) interventricular (posterior descending) branch of right coronary a., Right marginal branch.

# 3.) Interface for semantic and spatial search, filter, review, download

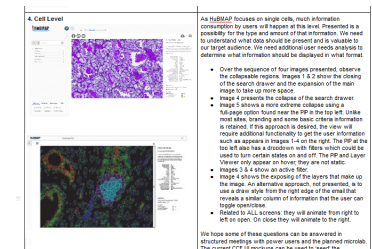
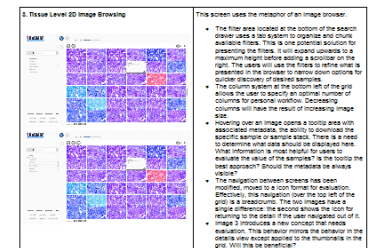
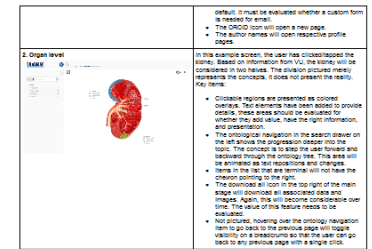
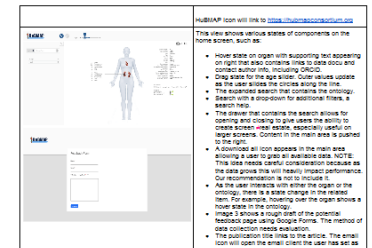
CCF UI Spec v0.5.0 v2

The initial 'user story' features a researcher interested to search for, filter out, review, and download biomolecular data in the context of the whole human body or in spatial relationship to specific organ(s), tissues, or cell types. The researcher is also able to learn more about how the data was acquired, to connect with data authors, and to submit questions and comments on the CCF UI.

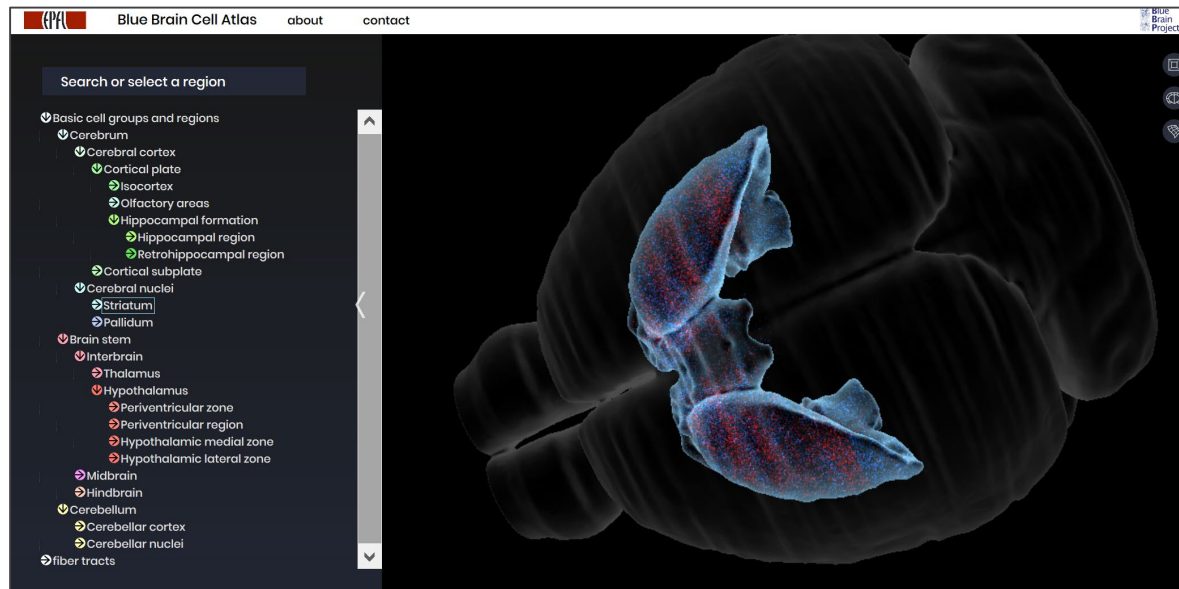
Consequently, the UI will support

- Search (ontology supported semantic search) and filter by ontology, anatomy, and metadata
- Visual browsing of tissue samples and metadata at the whole body, organ, tissue, and cell level
- Connect with data authors to inquire about technology details.
- Data download at the whole body, organ, tissue, and cell level.
- Submit questions and comments on the CCF UI.

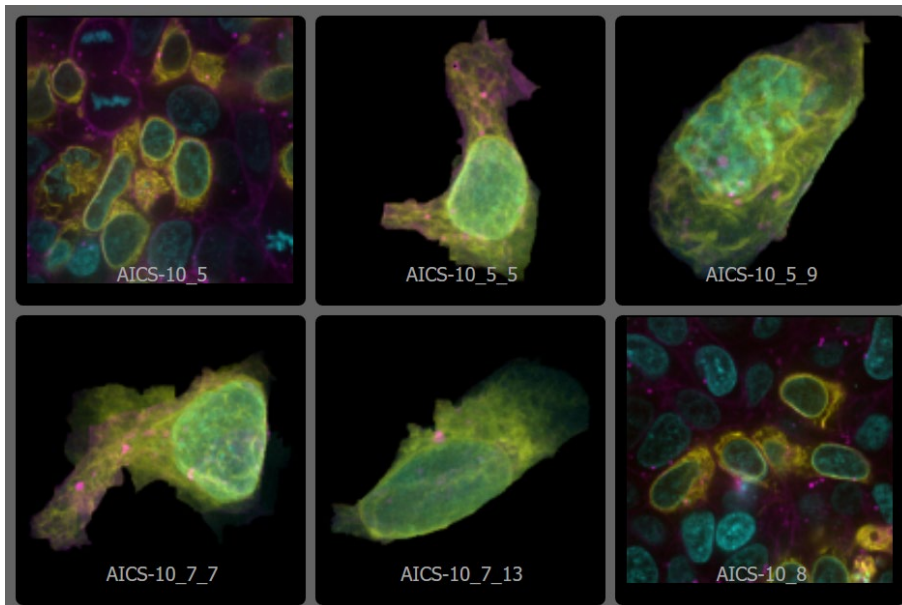
In the initial 9m, proof of concept versions of the whole body, organ, tissue and cell level views will be implemented.



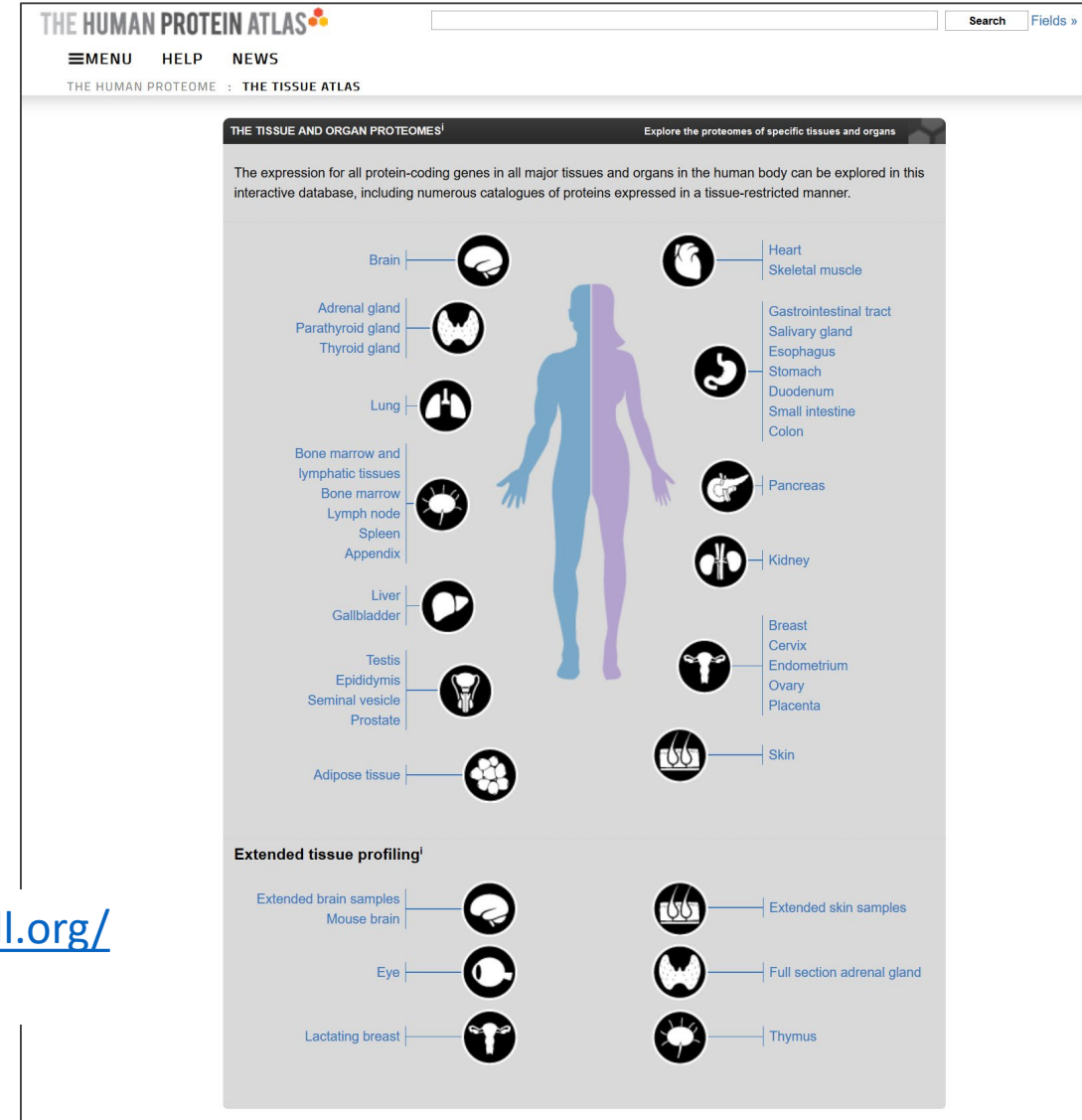




<https://bbp.epfl.ch/nexus/cell-atlas>



<https://www.allencell.org/3d-cell-viewer.html>



<https://www.proteinatlas.org/humanproteome/tissue>



# References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255. <http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl\_1). [http://www.pnas.org/content/vol101/suppl\\_1](http://www.pnas.org/content/vol101/suppl_1)

Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. <http://scimaps.org/atlas>

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press. <http://scimaps.org/atlas2>

