CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Actionable Data Visualizations of Science & Technology

Katy Börner @katycns

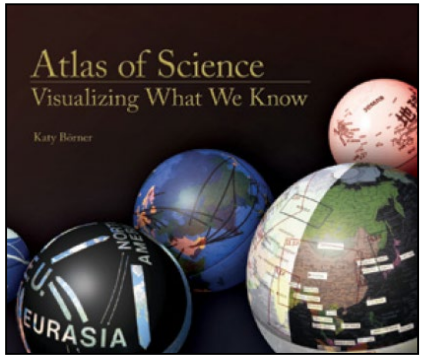
Victor H. Yngve Distinguished Professor of
Intelligent Systems Engineering & Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics, Computing, and Engineering
Indiana University Network Science Institute (IUNI)
Indiana University, Bloomington, IN, USA
+ 2018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

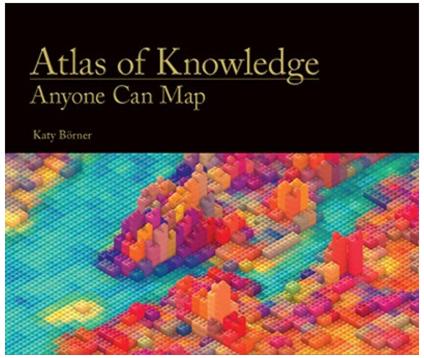


Exemplarily, the DVL is applied to

- 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

1st Decade (2005-2014)

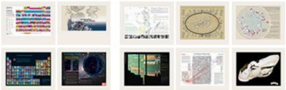
2nd Decade (2015-2024)

3rd Decade (2015-2034)

Maps

Iteration I (2005)
The Power of Maps

Iteration II (2006)
The Power of Reference Systems



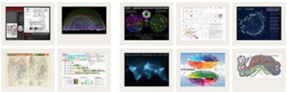
Iteration III (2007)
The Power of Forecasts

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

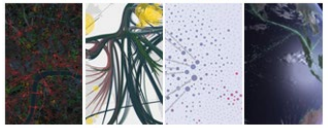
Iteration X (2014)
The Future of Science Mapping

Macroscopes

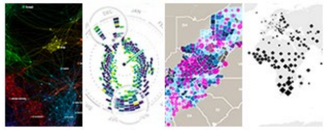
Iteration XI (2015)
Macroscopes for Interacting with Science

Iteration XIII (2017)
Macroscopes for Playing with Scale

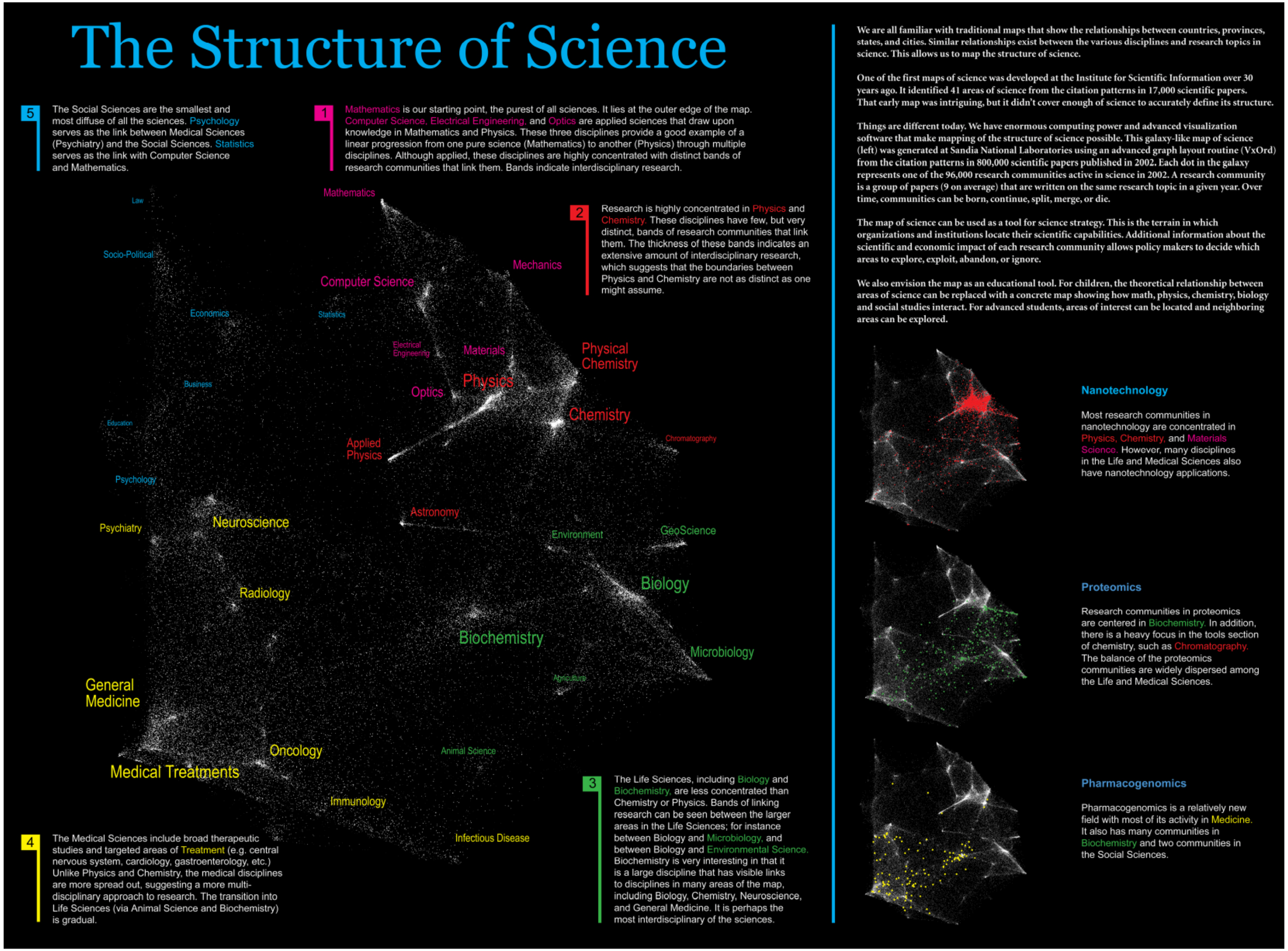
Iteration XII (2016)
Macroscopes for Making Sense of Science



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

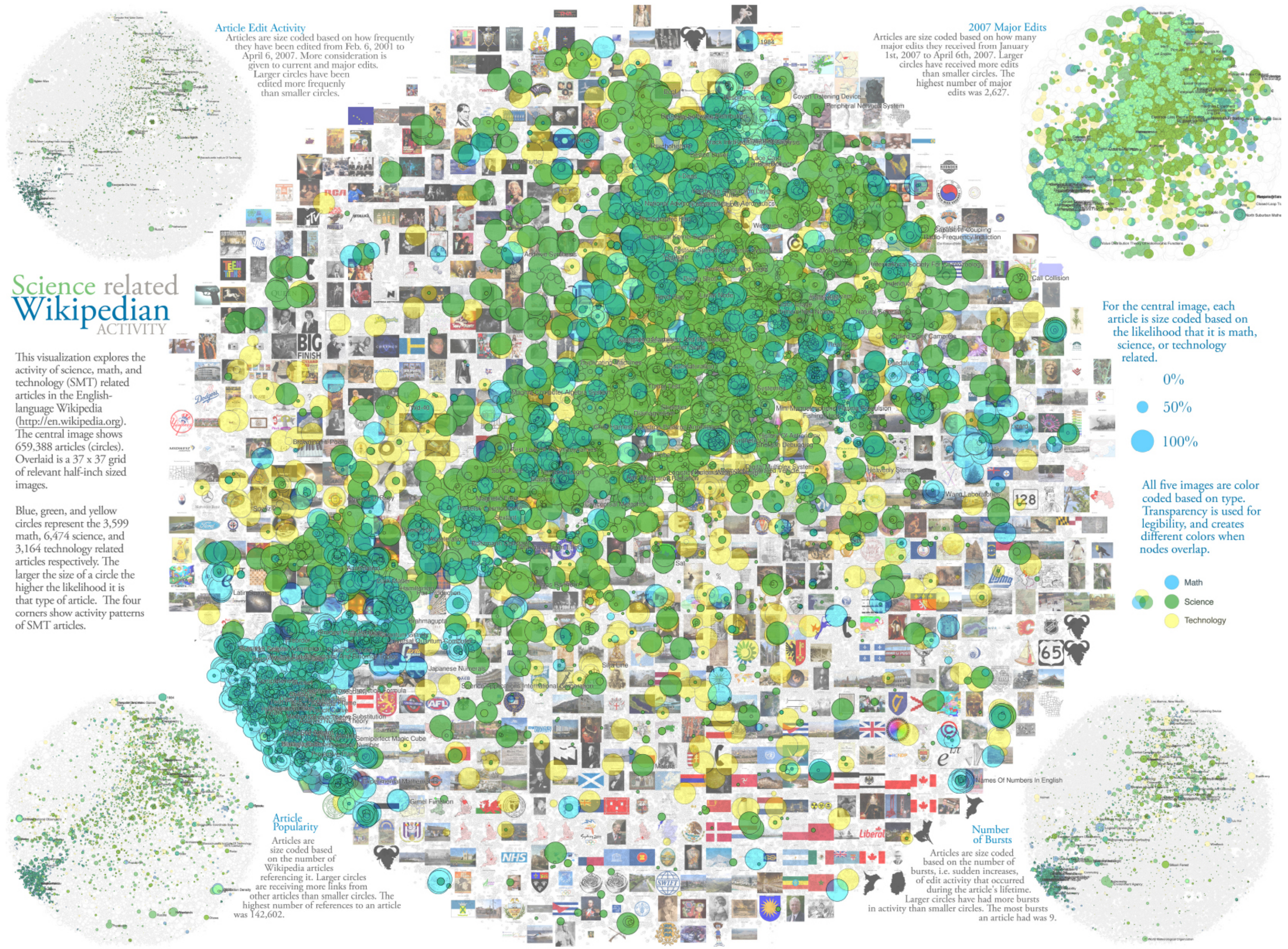


http://scimaps.org



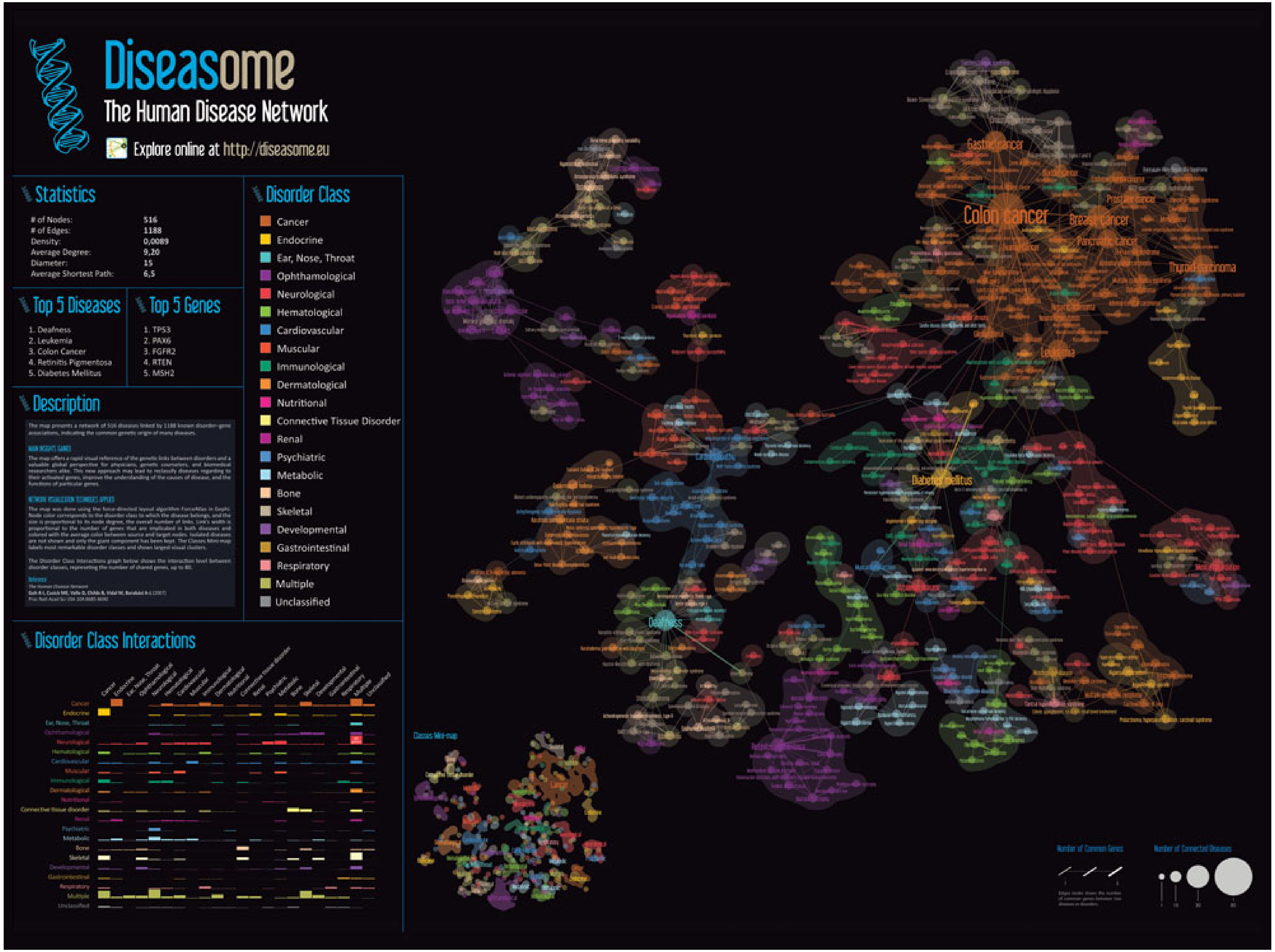
I.10 *The Structure of Science* - Kevin W. Boyack and Richard Klavans - 2005

http://scimaps.org



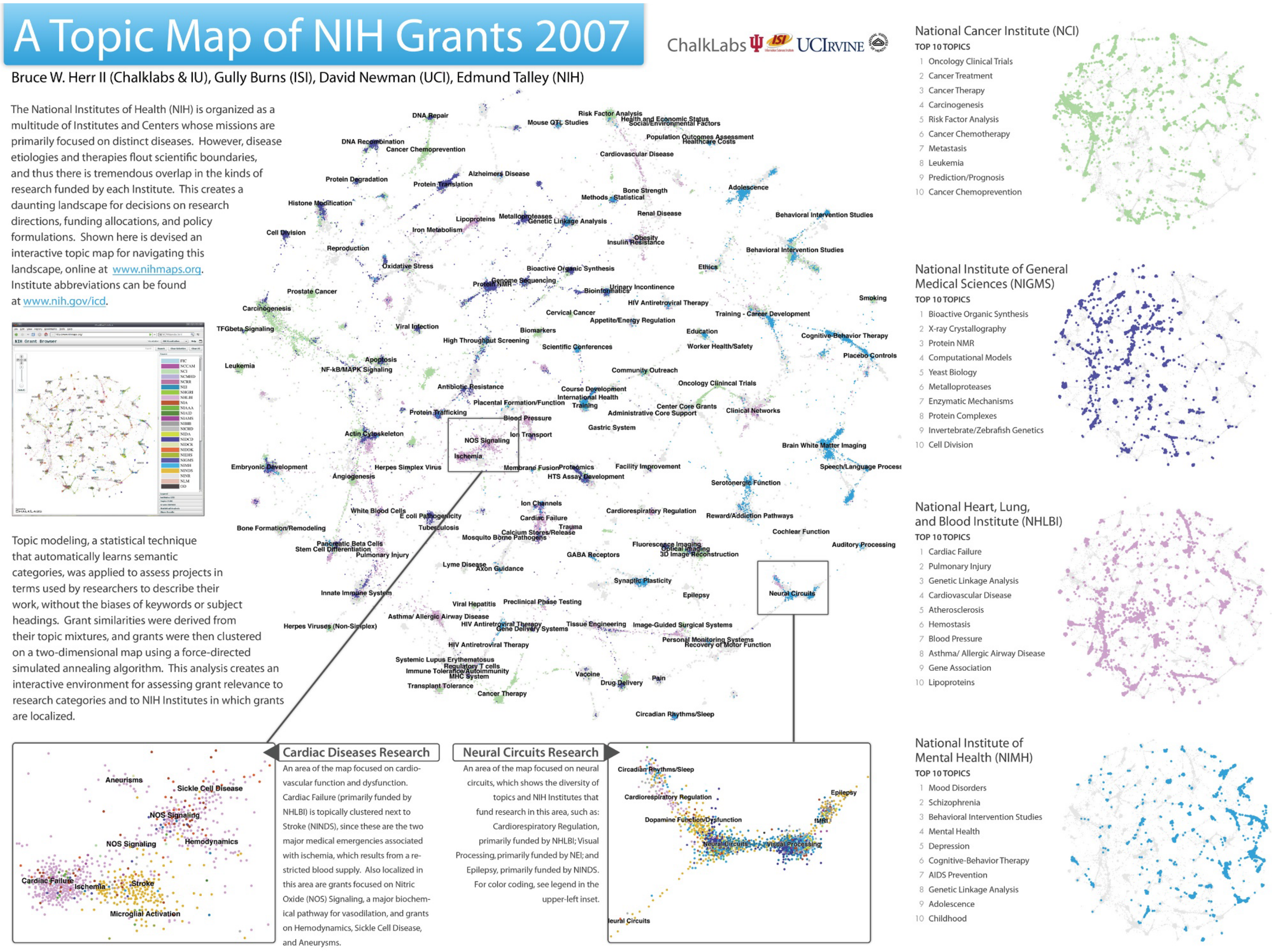
III.8 *Science-Related Wikipedian Activity* - Bruce W. Herr II, Todd M. Holloway, Elisha F. Hardy, Katy Börner, and Kevin Boyack - 2007

http://scimaps.org



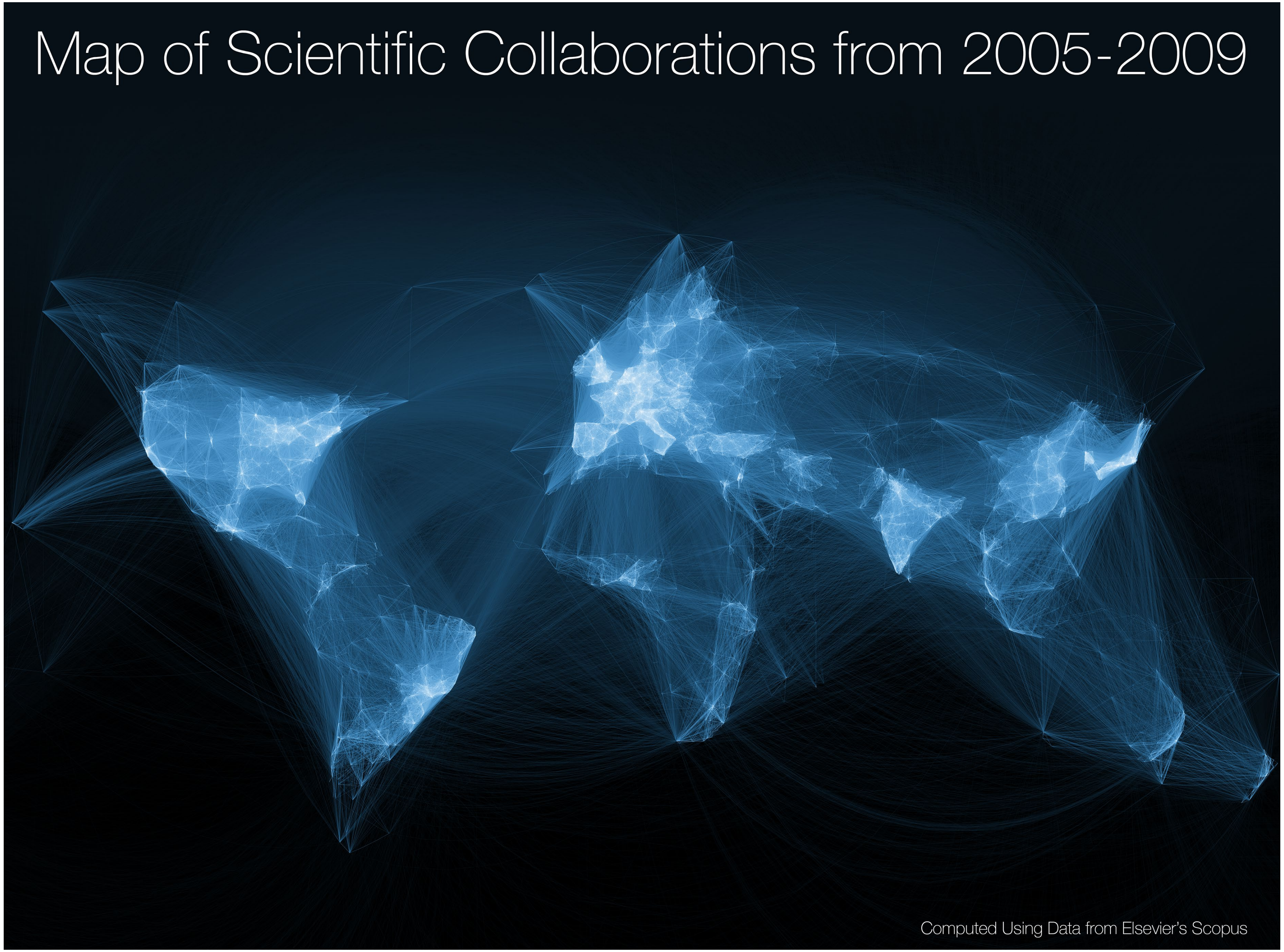
VI.3 *Diseasome: The Human Disease Network* - Mathieu Bastian and Sébastien Heymann - 2009

http://scimaps.org



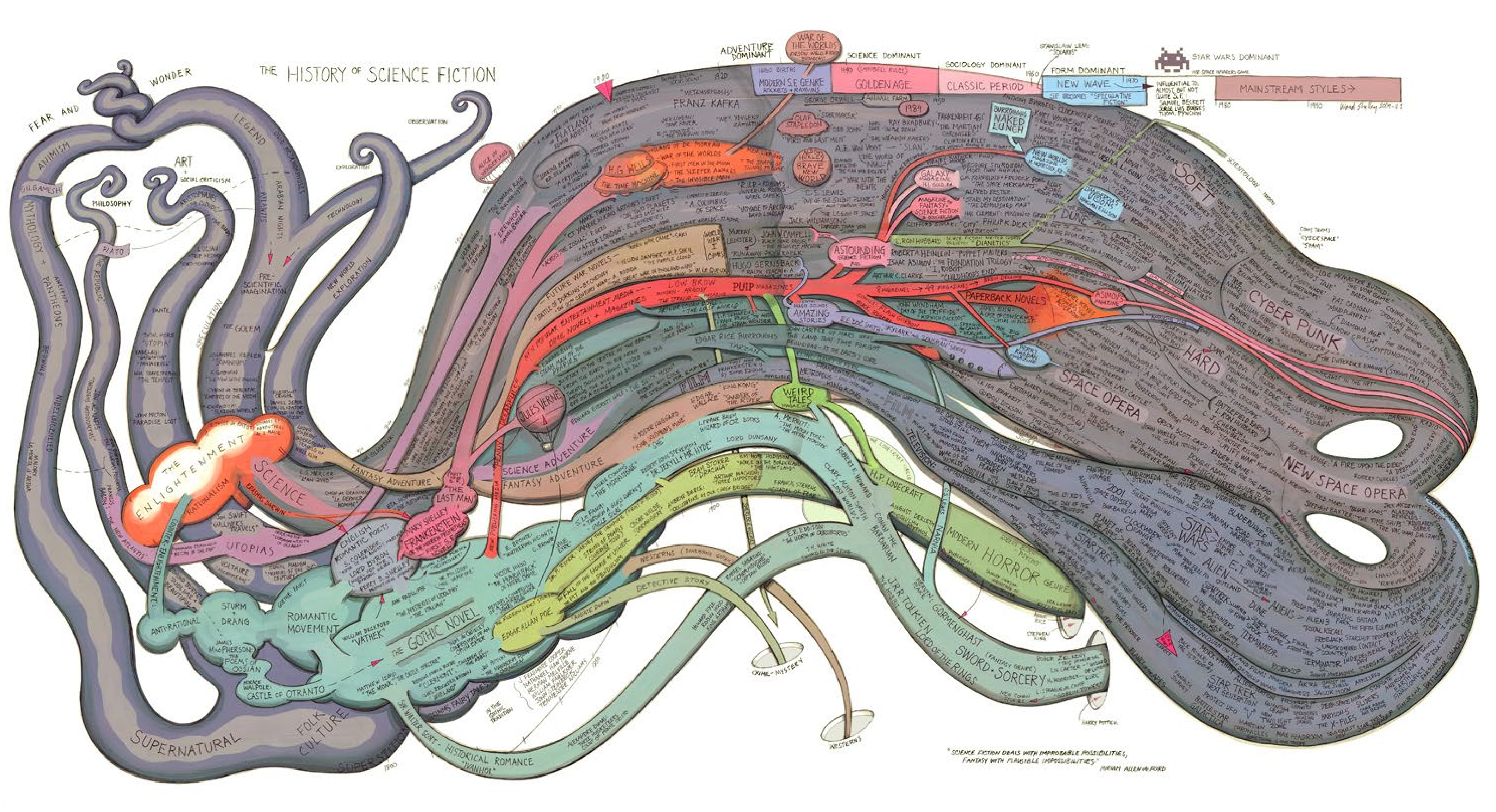
V.7 *A Topic Map of NIH Grants 2007* - Bruce W. Herr II, Gully A.P.C. Burns, David Newman, and Edmund Talley - 2009

http://scimaps.org



VII.6 Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012

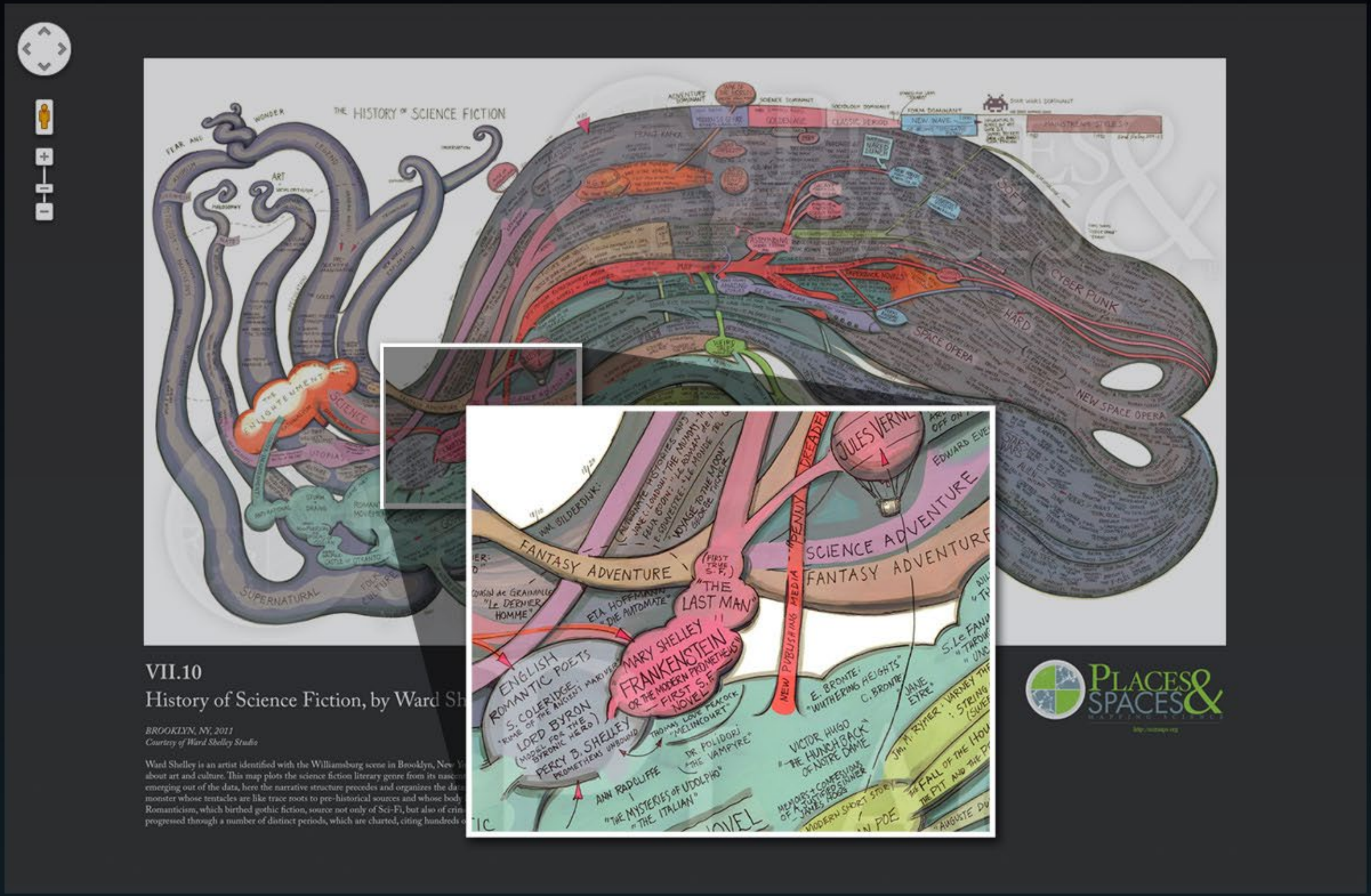
http://scimaps.org



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

http://scimaps.org

Check out our Zoom Maps online!



Visit scimaps.org and check out all our maps in stunning detail!

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

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

#SacklerModVisST

Arthur M. Sackler COLLOQUIA OF THE NATIONAL ACADEMY OF SCIENCES

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

nasonline.org/Sackler-Visualizing-Science

Arthur M. Sackler COLLOQUIA OF THE NATIONAL ACADEMY OF SCIENCES Clarivate Microsoft

Programs | Arthur M. Sackler Colloquia | Completed Colloquia | Modeling and Visualizing Science and Technology Developments | Share

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

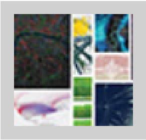
LabX

Sackler Forum

Science & Entertainment Exchange

Arthur M. Sackler COLLOQUIA

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

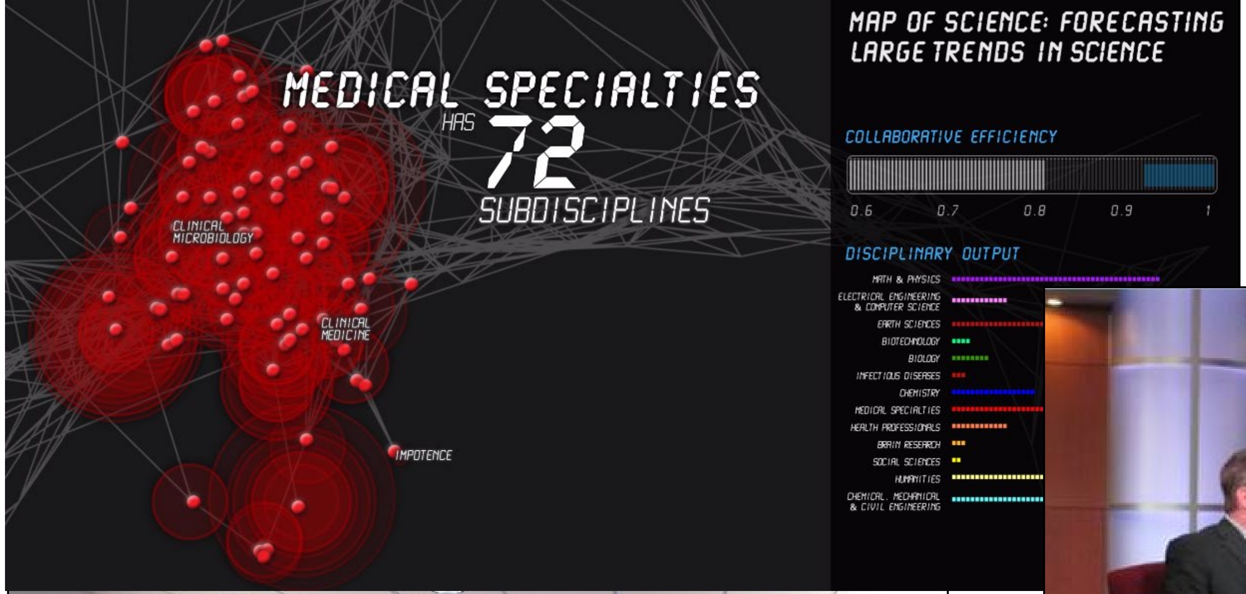
This colloquium was held in Irvine, CA on December 4-5, 2017.

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



CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

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:

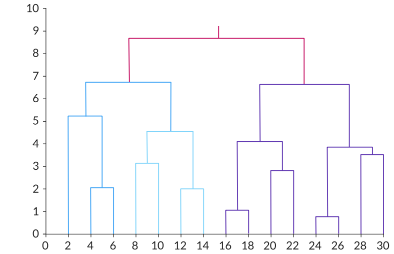
- literacy (ability to read and write text in titles, axis labels, legends, etc.),
- visual literacy (ability to find, interpret, evaluate, use, and create images and visual media), and
- mathematical literacy (ability to formulate, employ, and interpret math in a variety of contexts).

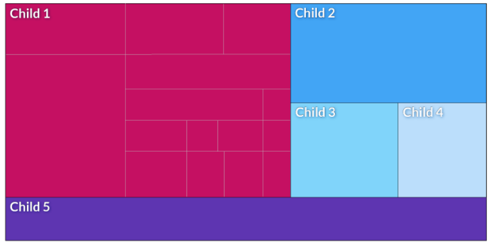
Being able to "read and write" data visualizations is becoming as important as being able to read and write text. Understanding, measuring, and improving data and visualization literacy is important to strategically approach local and global issues.

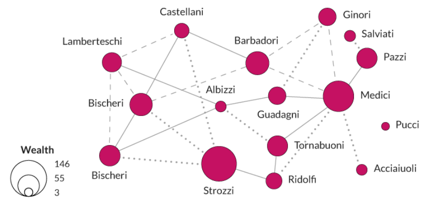
Visualization Frameworks

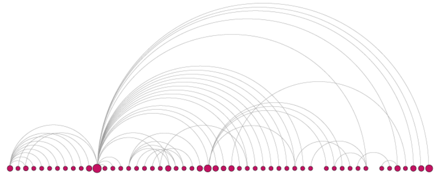
MANY frameworks and taxonomies have been proposed to

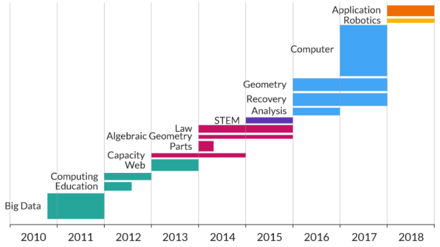
- help organize and manage the evolving zoo of 500+ different data visualization types,
- provide guidance when designing data visualizations, and
- facilitate teaching.

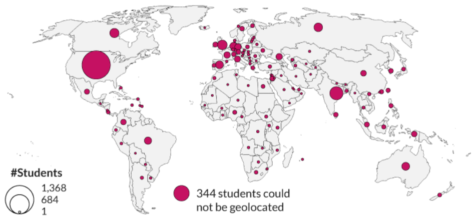


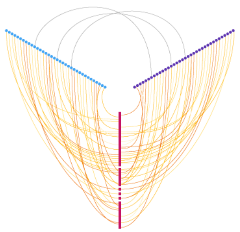


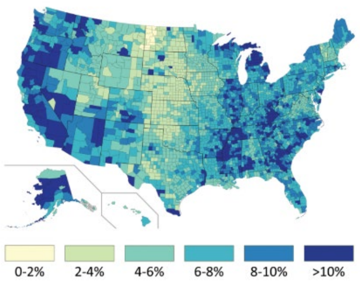


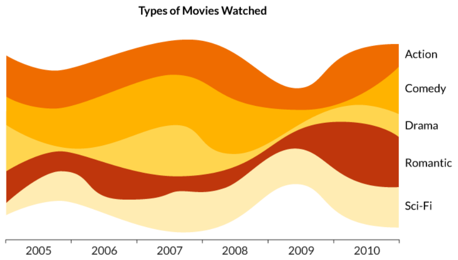








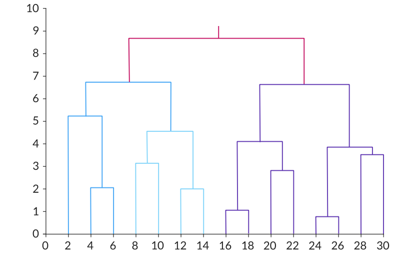


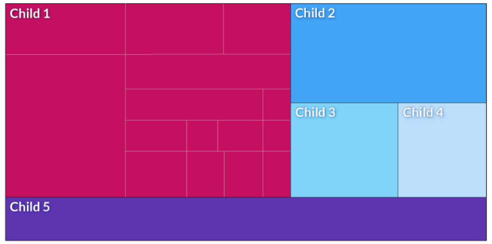


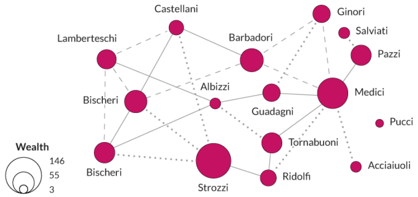
Existing Visualization Frameworks

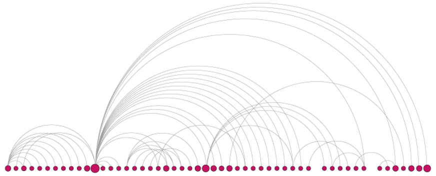
Organize data visualizations by

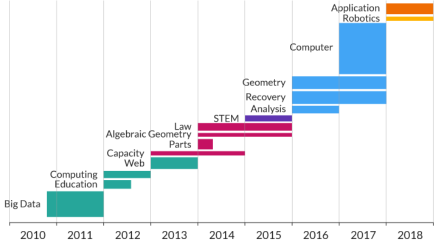
- User insight needs
- User task types
- Data to be visualized
- Data transformations
- Visualization technique
- Visual mapping transformations
- Interaction techniques
- Deployment options
- and other features ...

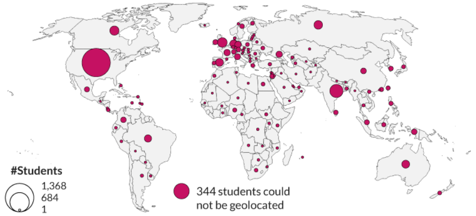


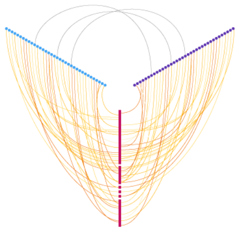


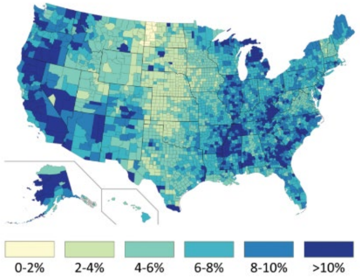


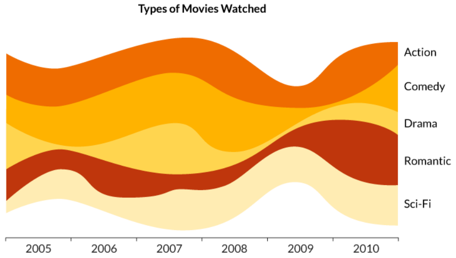












DVL Framework: Desirable Properties

- 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.

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

2 Data Scales
- nominal
- ordinal
- interval
- ratio

3 Analyses
- statistical
- temporal
- geospatial
- topical
- relational

4 Visualizations
- table
- chart
- graph
- map
- tree
- network

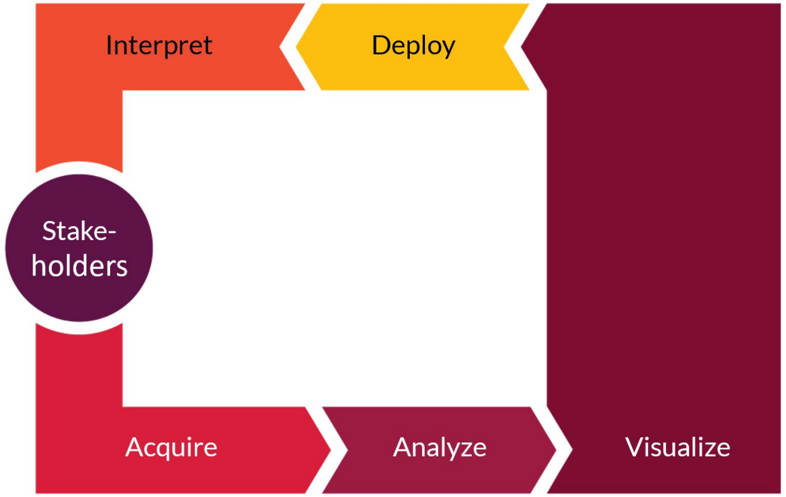
5 Graphic Symbols
- geometric symbols
  point
  line
  area
  surface
  volume
- linguistic symbols
  text
  numerals
  punctuation marks
- pictorial symbols
  images
  icons
  statistical glyphs

6 Graphic Variables
- spatial
  position
- retinal
  form
  color
  optics
  motion

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

DVL Workflow Process

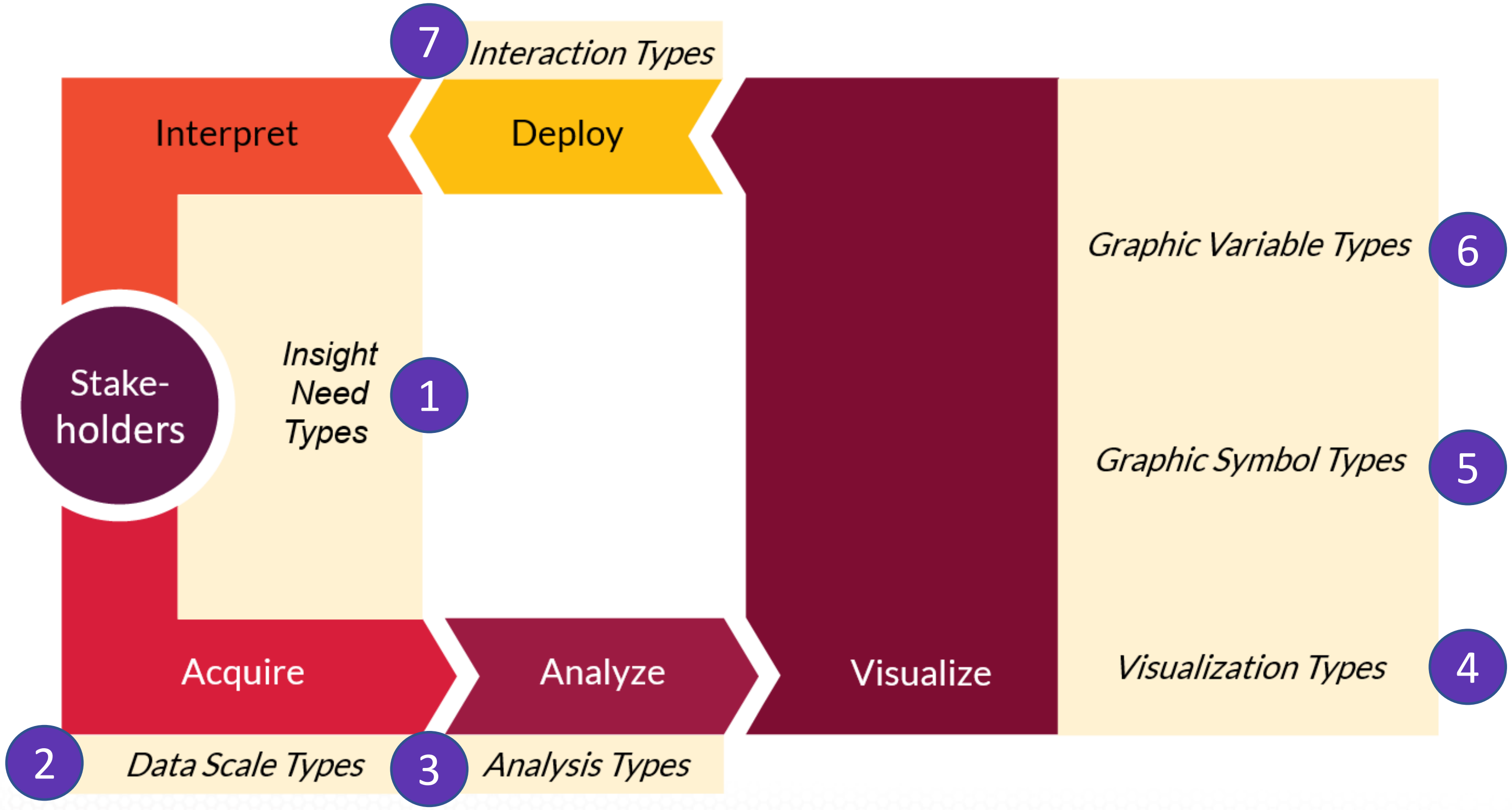
Defines 5 steps required to render data into insights.



Data Visualization Literacy Framework (DVL-FW)

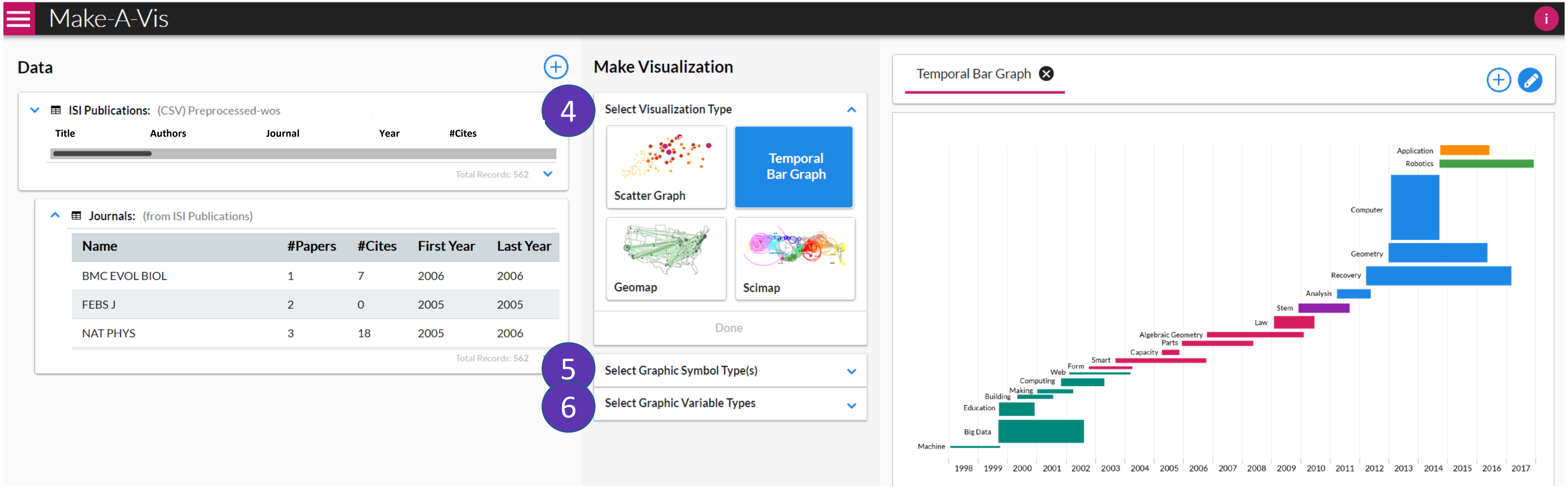
Consists of two parts that are interlinked:

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.



Typology of the Data Visualization Literacy Framework

1	2	3	4	5	6	7
Insight Needs	**Data Scales**	**Analyses**	**Visualizations**	**Graphic Symbols**	**Graphic Variables**	**Interactions**
• 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	• statistical • temporal • geospatial • topical • relational	• table • chart • graph • map • tree • network	• 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	• 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.

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	Data Scales	Analyses	Visualizations	Graphic Symbols	Graphic Variables	Interactions
• categorize/cluster	• nominal	• statistical	• table	• geometric symbols	• spatial	• zoom
• order/rank/sort	• ordinal	• temporal	• chart	point	position	• search and locate
• distributions (also outliers, gaps)	• interval	• geospatial	• graph	line	• retinal	• filter
• comparisons	• ratio	• topical	• map	area	form	• details-on-demand
• trends (process and time)		• relational	• tree	surface	color	• history
• geospatial			• network	volume	optics	• extract
• compositions (also of text)				• linguistic symbols	motion	• link and brush
• correlations/ relationships				text		• projection
				numerals		• distortion
				punctuation marks		
				• pictorial symbols		
				images		
				icons		
				statistical glyphs		

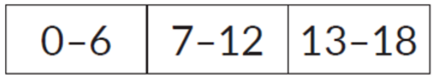
Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 28-29.

Data Scale Types

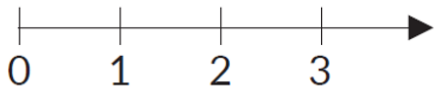
Nominal: A categorical scale, also called a nominal or category scale, is qualitative. Categories are assumed to be non-overlapping.

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	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	Data Scales	Analyses	Visualizations	Graphic Symbols	Graphic Variables	Interactions
• 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	• statistical • temporal • geospatial • topical • relational	• table • chart • graph • map • tree • network	• 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	• 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

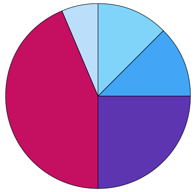
4

Insight Needs	Data Scales	Analyses	Visualizations	Graphic Symbols	Graphic Variables	Interactions
• categorize/cluster	• nominal	• statistical	• table	• geometric symbols	• spatial	• zoom
• order/rank/sort	• ordinal	• temporal	• chart	point	position	• search and locate
• distributions (also outliers, gaps)	• interval	• geospatial	• graph	line	• retinal	• filter
• comparisons	• ratio	• topical	• map	area	form	• details-on-demand
• trends (process and time)		• relational	• tree	surface	color	• history
• geospatial			• network	volume	optics	• extract
• compositions (also of text)				• linguistic symbols	motion	• link and brush
• correlations/ relationships				text		• projection
				numerals		• distortion
				punctuation marks		
				• pictorial symbols		
				images		
				icons		
				statistical glyphs		

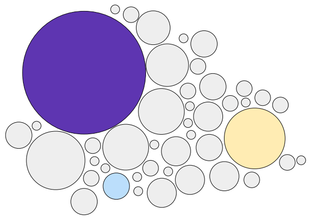
Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 30-31.

Visualization Types

Chart

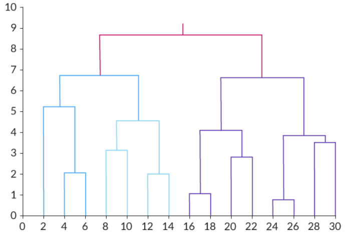


Pie Chart

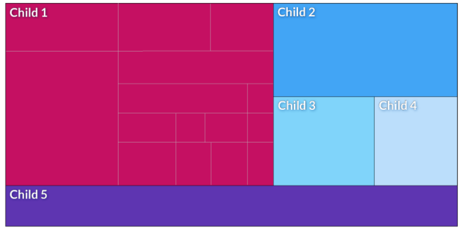


Bubble Chart

Tree

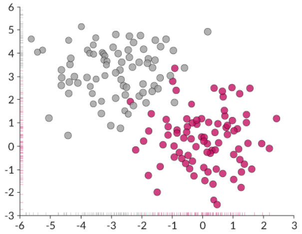


Dendrogram

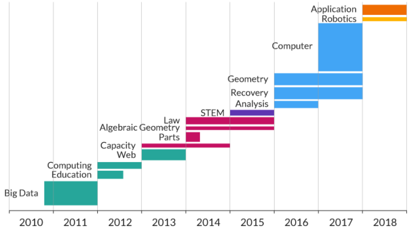


Tree Map

Graph

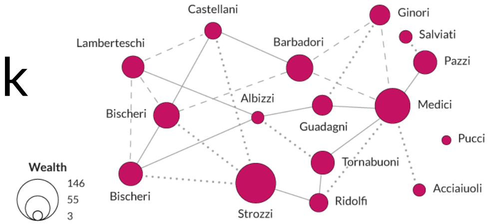


Scatter Graph

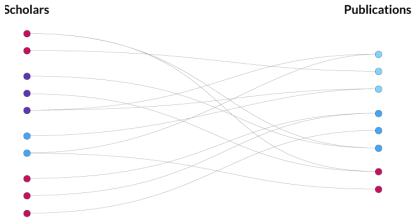


Temporal Bar Graph

Network

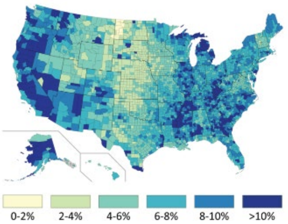


Dendrogram

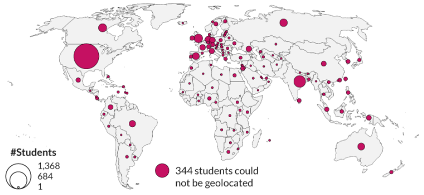


Tree Map

Map



Choropleth Map



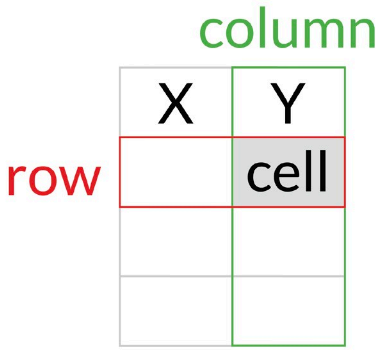
Proportional Symbol Map

Visualize: Reference Systems

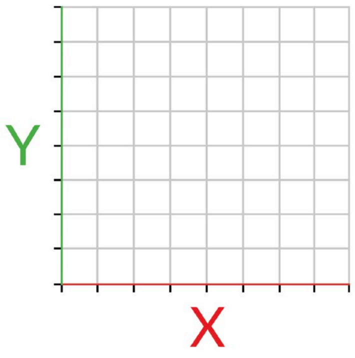
4

Visualization Types
- table
- chart
- graph
- map
- network layout

Table
columns by rows

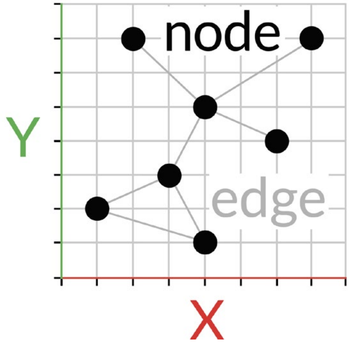


Graph
x-y coordinates

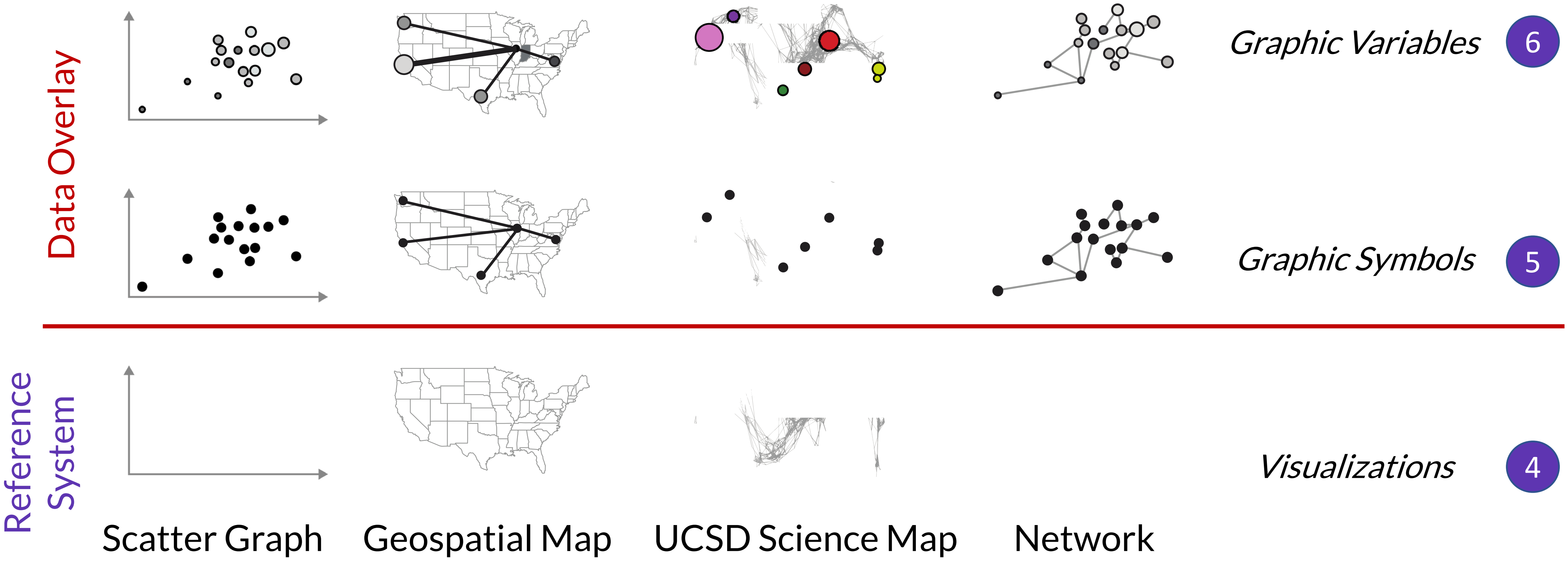


Map
latitude/
longitude

Network
local similarity



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	Data Scales	Analyses	Visualizations	Graphic Symbols	Graphic Variables	Interactions
• 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	• statistical • temporal • geospatial • topical • relational	• table • chart • graph • map • tree • network	• 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	• 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.

CNS Cyberinfrastructure for Network Science Center

Graphic Variable Types

Position: x, y; possibly z — Quantitative

Form:

- Size — Quantitative
- Shape — Qualitative
- Rotation (Orientation) — 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

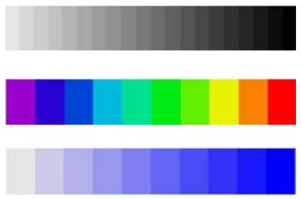
Position: x, y; possibly z

Form:

- Size
- Shape
- Rotation (Orientation)

Color:

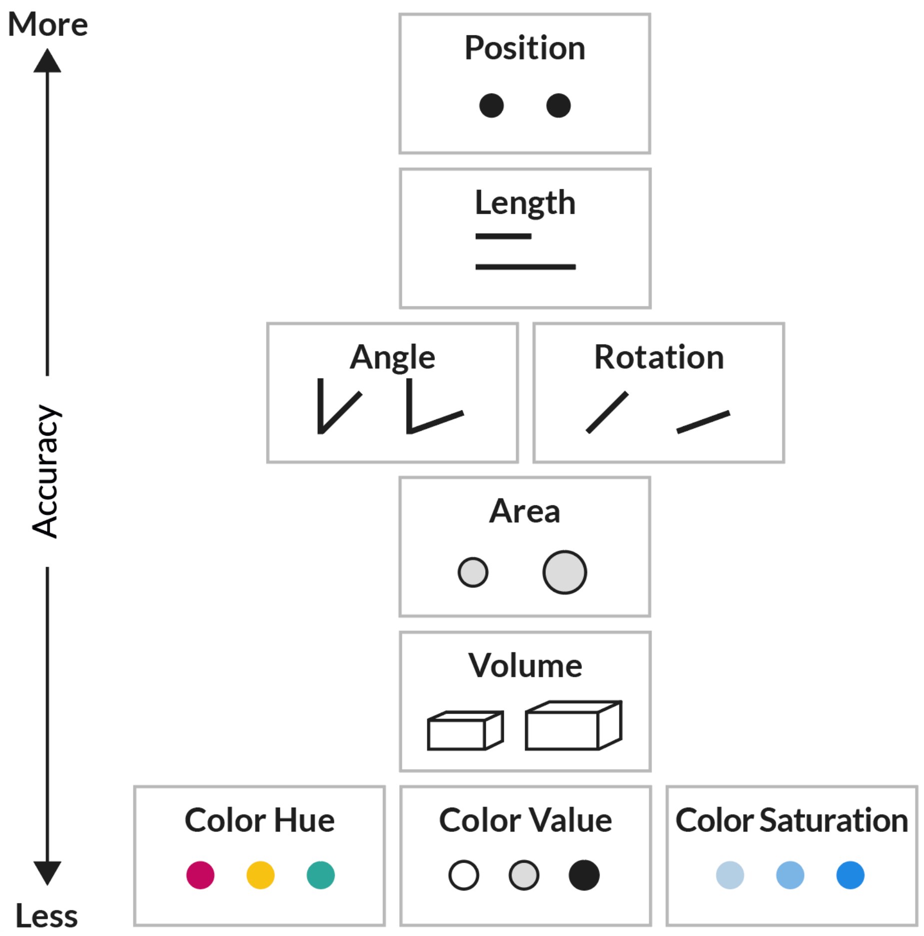
- Value (Lightness)
- Hue (Tint)
- Saturation (Intensity)



Optics: Blur, Transparency, Shading, Stereoscopic Depth

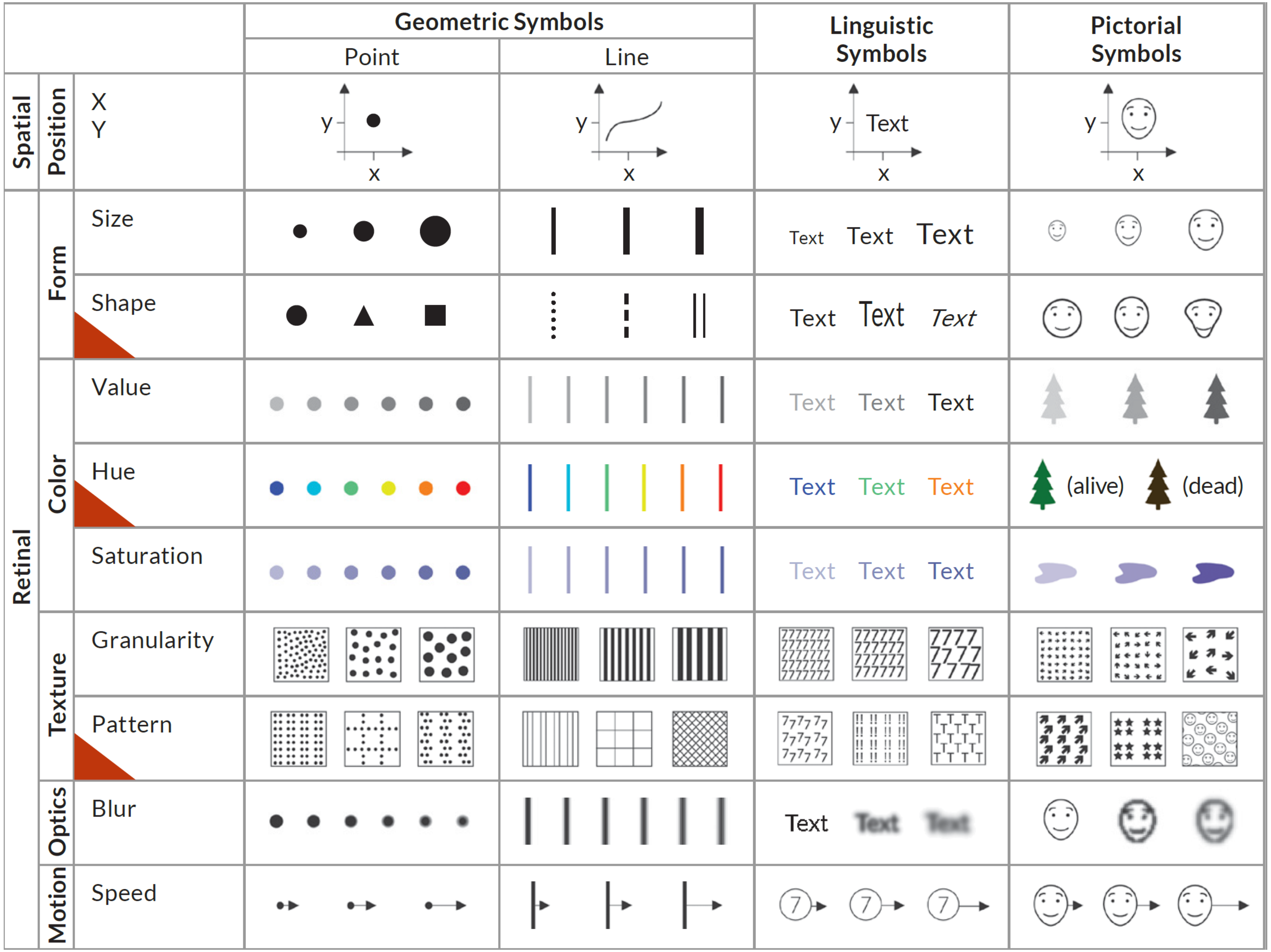
Texture: Spacing, Granularity, Pattern, Orientation, Gradient

Motion: Speed, Velocity, Rhythm



Graphic Symbol Types

Graphic Variable Types



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

Qualitative

Also called:

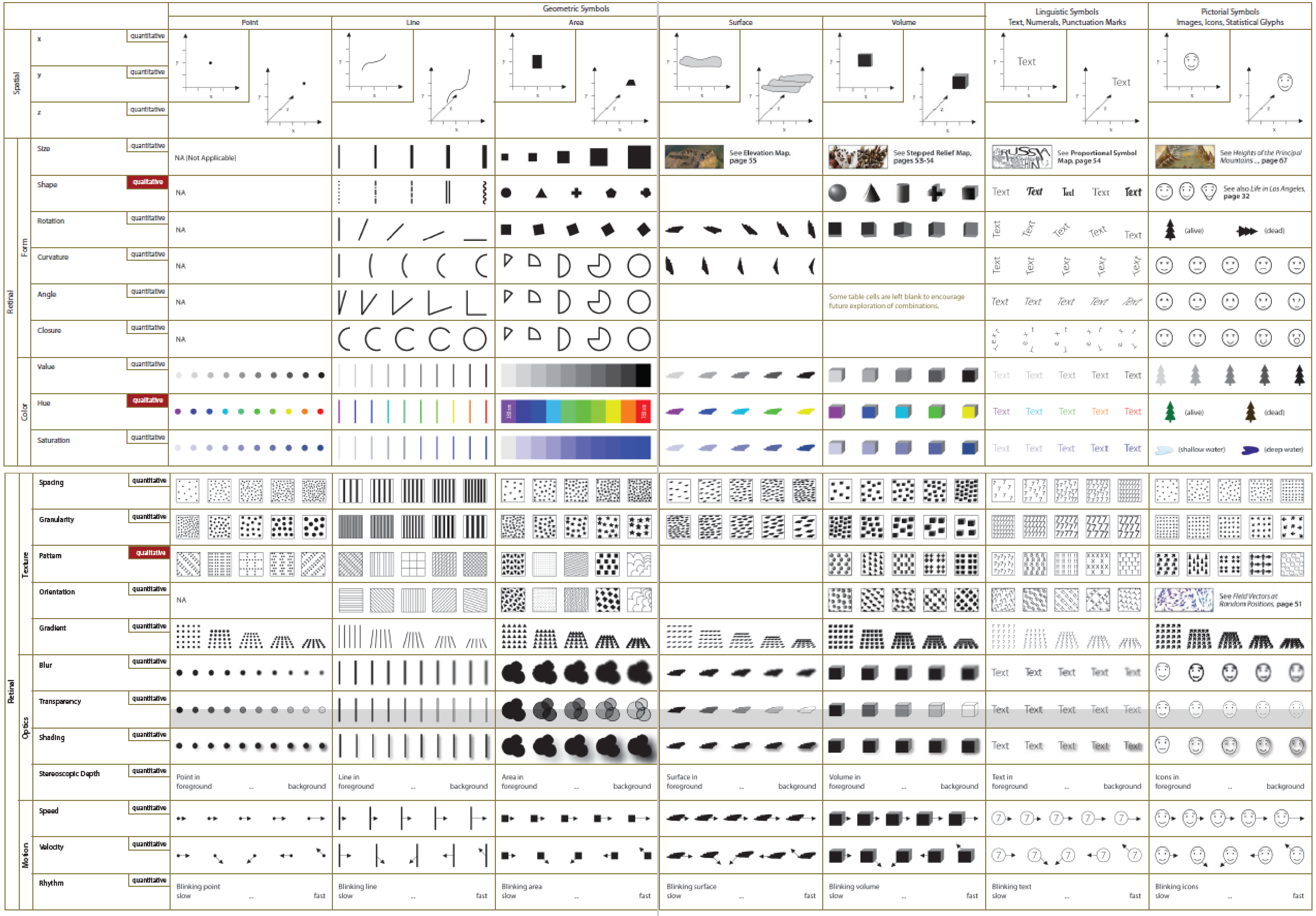
Categorical Attributes
Identity Channels

Quantitative

Also called:

Ordered Attributes
Magnitude Channels

Graphic Variable Types Versus Graphic Symbol Types



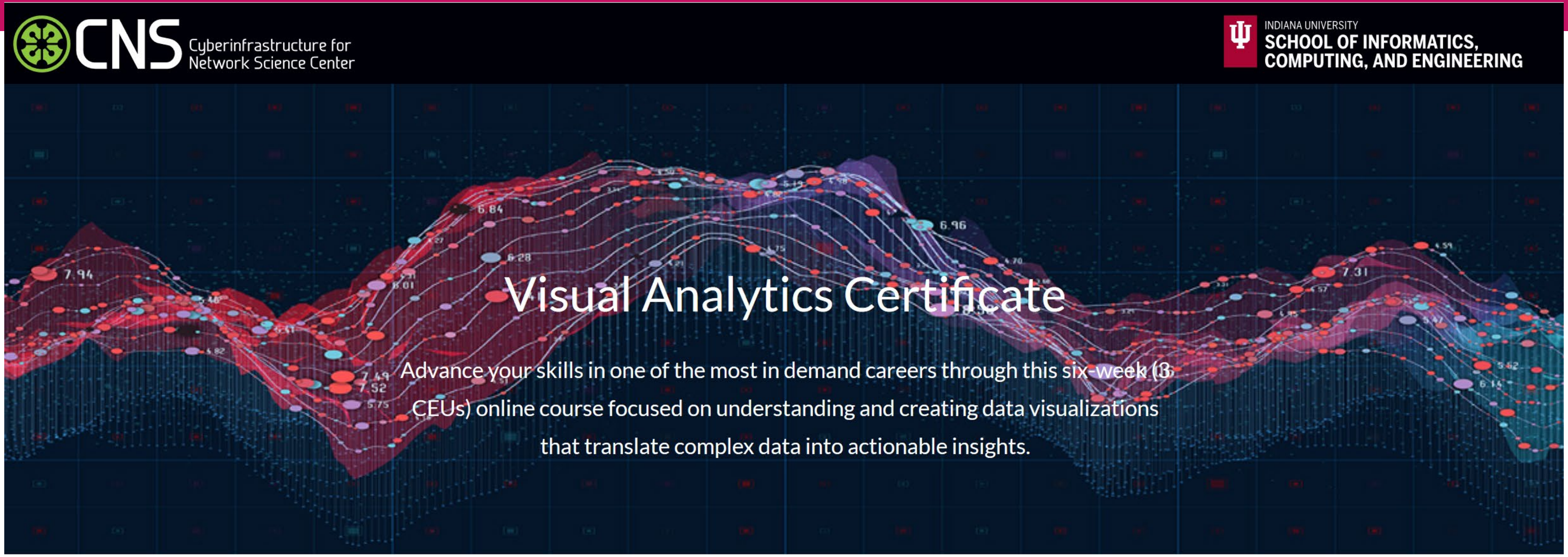
See *Atlas of Knowledge* pages 36-39 for complete table.

Typology of the Data Visualization Literacy Framework

7

Insight Needs	Data Scales	Analyses	Visualizations	Graphic Symbols	Graphic Variables	Interactions
• 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	• statistical • temporal • geospatial • topical • relational	• table • chart • graph • map • tree • network	• 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	• 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.



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

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

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:

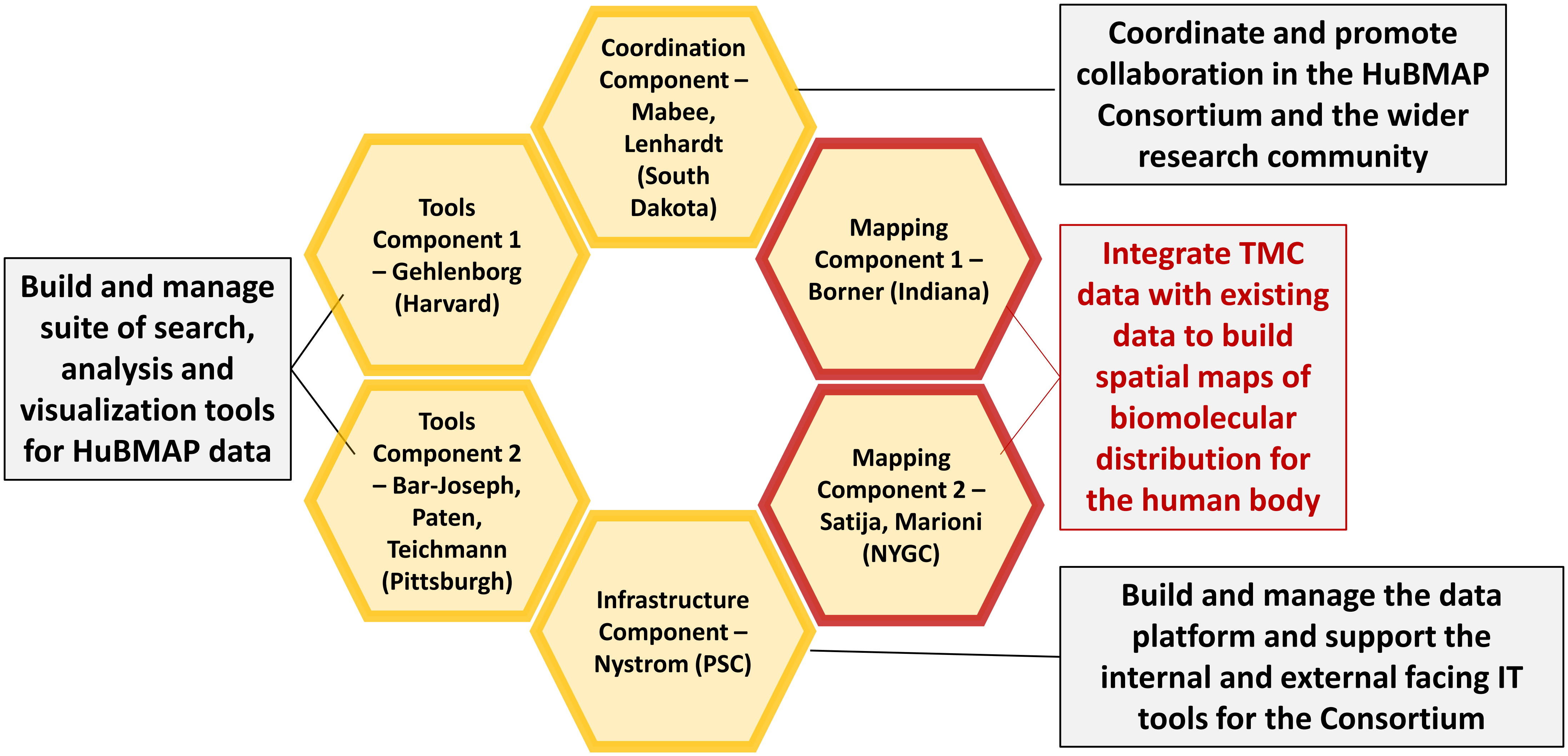
Katy Börner, Bruce Herr II, Paul Macklin, Randy Heiland & Ellen Quardokus

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

Griffin Weber, Harvard Medical School, Boston, MA

Samuel Friedman, Opto-Knowledge Systems, Inc.

HuBMAP: HIVE


Coordination Component – Mabee, Lenhardt (South Dakota)
Coordinate and promote collaboration in the HuBMAP Consortium and the wider research community
Tools Component 1 – Gehlenborg (Harvard)
Tools Component 2 – Bar-Joseph, Paten, Teichmann (Pittsburgh)
Build and manage suite of search, analysis and visualization tools for HuBMAP data
Mapping Component 1 – Borner (Indiana)
Mapping Component 2 – Satija, Marioni (NYGC)
Integrate TMC data with existing data to build spatial maps of biomolecular distribution for the human body
Infrastructure Component – Nystrom (PSC)
Build and manage the data platform and support the internal and external facing IT tools for the Consortium

HuBMAP: HIVE Mapping Components (MC)

Mapping Component 2 – Satija, Marioni (NYGC)

Meet in the middle

Mapping Component 1 – Borner (Indiana)

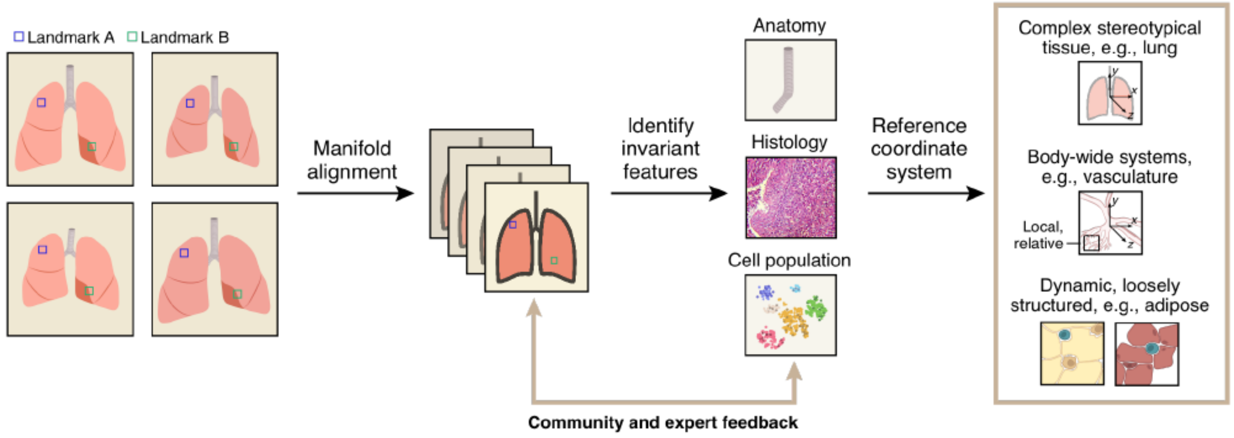


Figure 1: Overview of MC strategy to iteratively generate a reference coordinate system. We will work with TMC to annotate initial datasets with key features and ontologies, which will serve as 'landmarks' to align images across individuals. We will modify and adapt our strategy to diverse tissues, retaining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

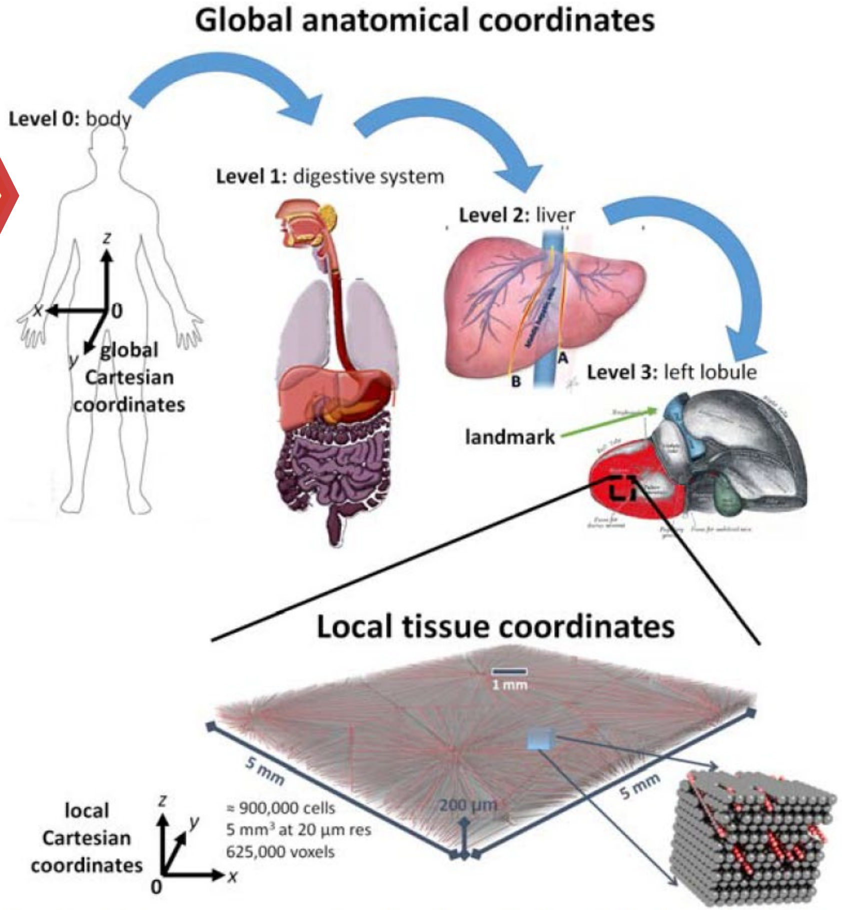


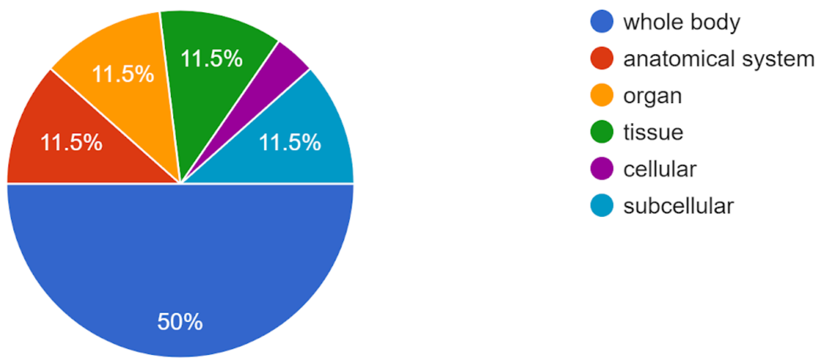
Fig. 5. CCF concept, navigating through the global anatomical coordinate system to insert a synthetic tissue sample (from PhysiCell[4]) into the left liver lobe with a local coordinate system.

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.

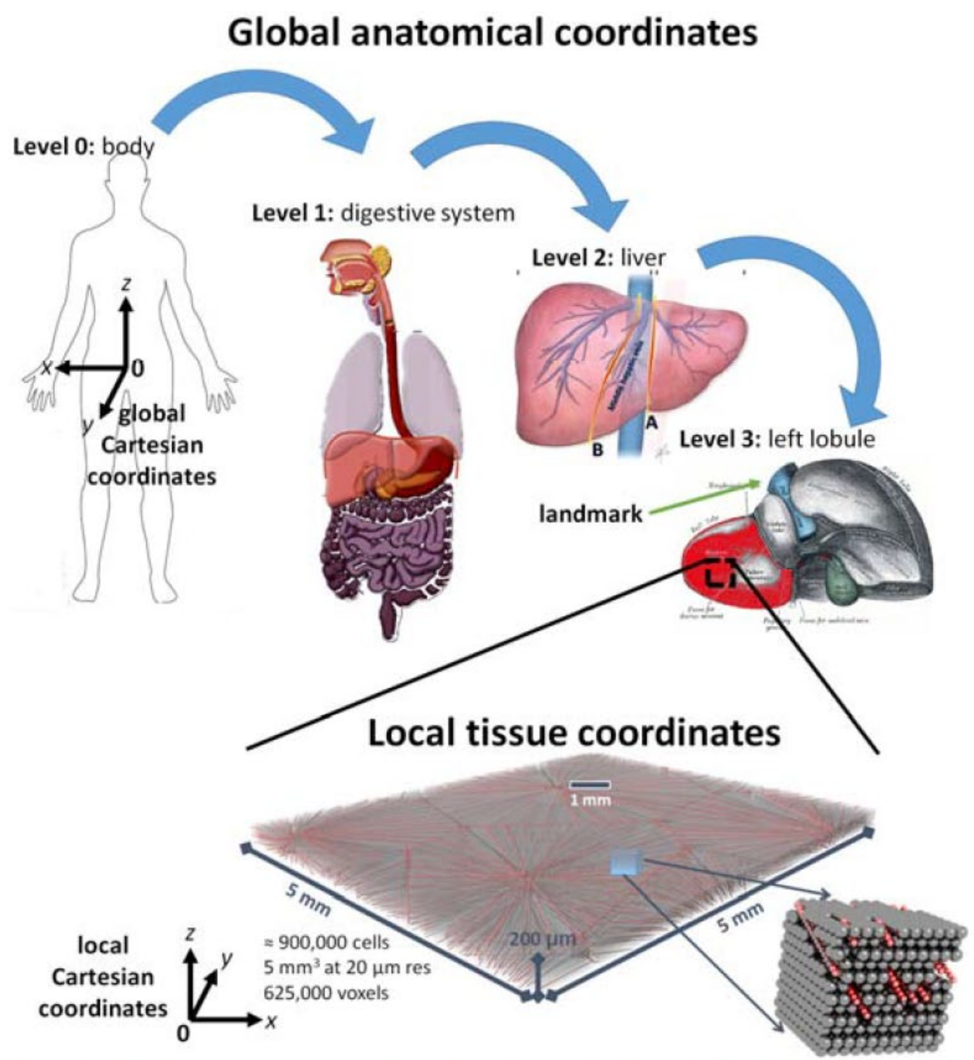


Fig. 5. CCF concept, navigating through the global anatomical coordinate system to insert a synthetic tissue sample (from PhysiCell[4]) into the left liver lobe with a local coordinate system.

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

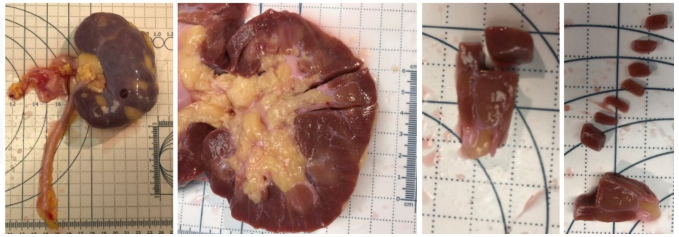
2.) 3D Spatial Models interlinked with terminology/ontology

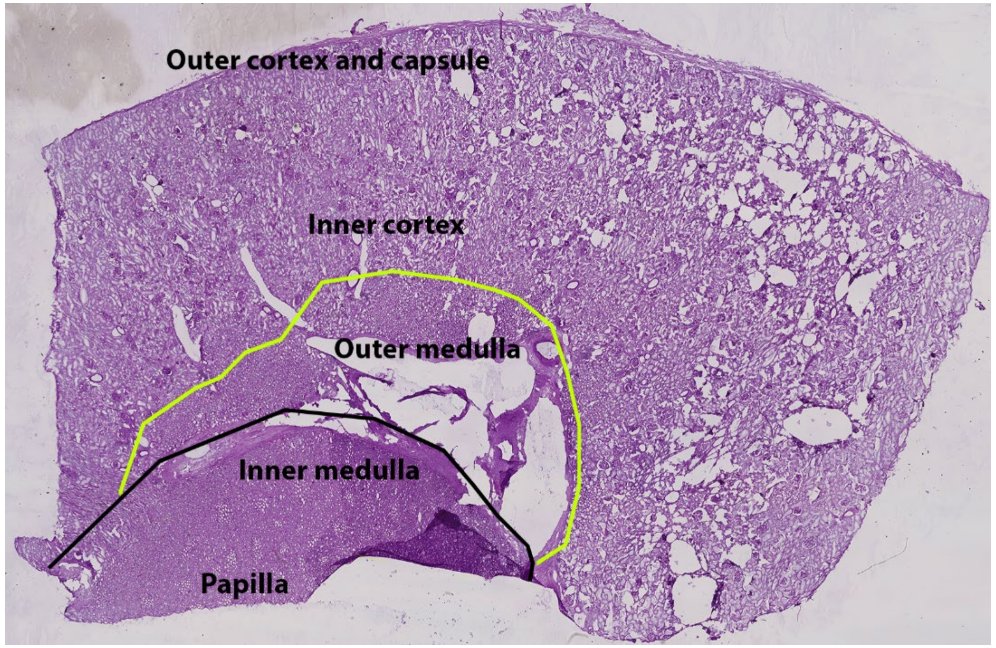
Kidney: Jeff Spraggins et al., VU

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

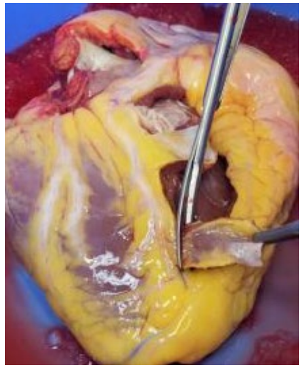




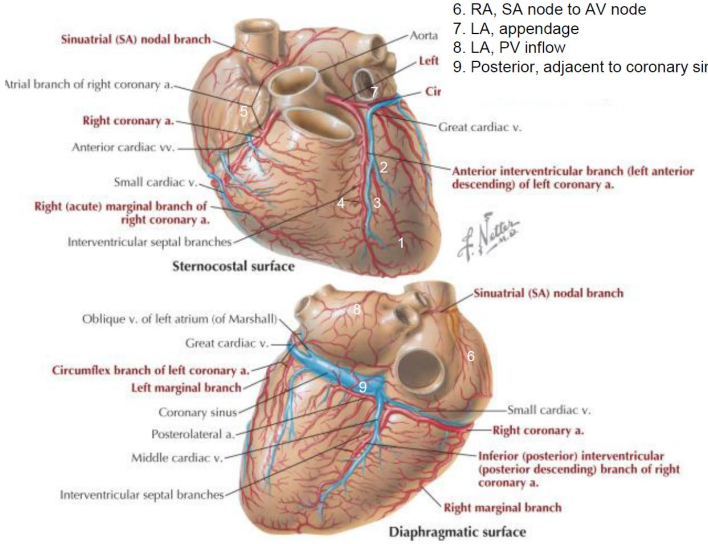
Heart: Shin Lin, UW

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

Terminology; Coordinates and photos to spatialize



Sites	Distinctive features
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



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

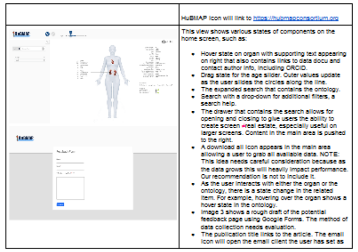
CCF UI Spec v0.5.0 v2

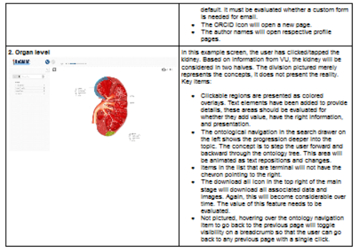
The initial 'user story' features a researcher interested to search for, filter out, review, and download biomolecular data in the context of the whole human body or in spatial relationship to specific organ(s), tissues, or cell types. The researcher is also able to learn more about how the data was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

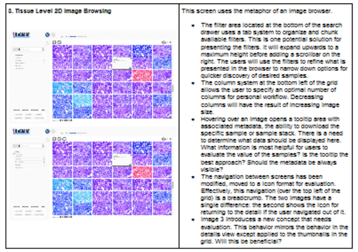
Consequently, the UI will support

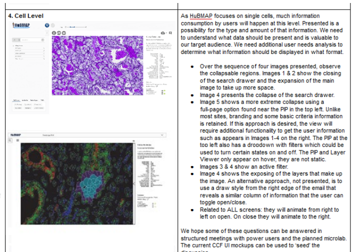
- Search (ontology supported semantic search) and filter by ontology, anatomy, and metadata
- Visual browsing of tissue samples and metadata at the whole body, organ, tissue, and cell level
- Connect with data authors to inquire about technology details.
- Data download at the whole body, organ, tissue,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

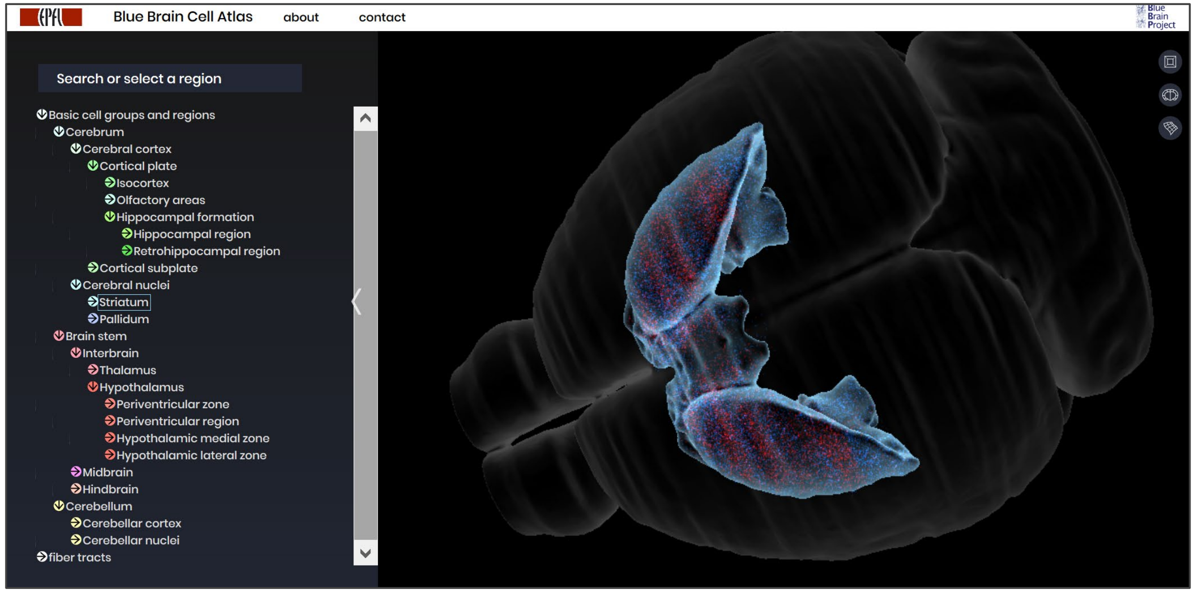
In the initial 9m, proof of concept versions of the whole body, organ, tissue and cell level views will be implemented.



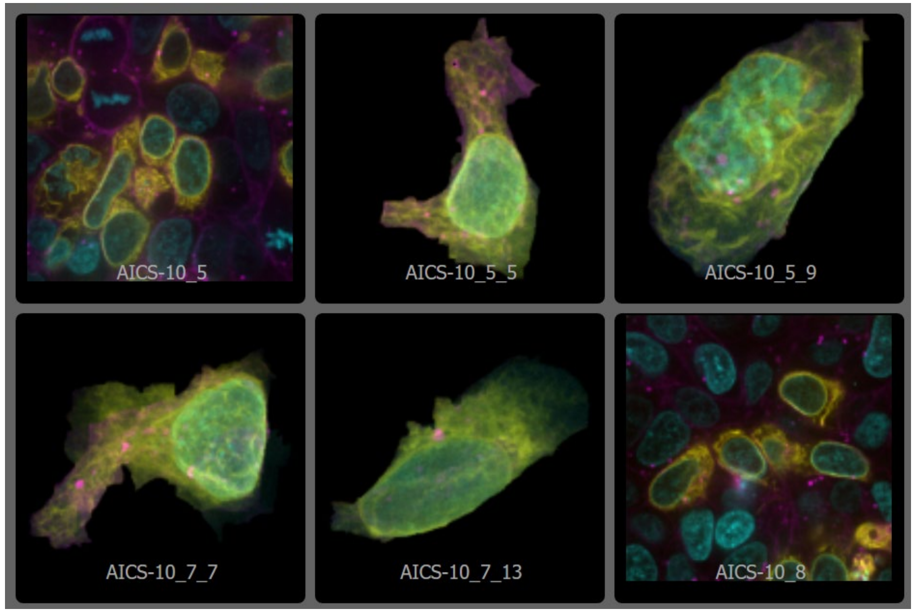




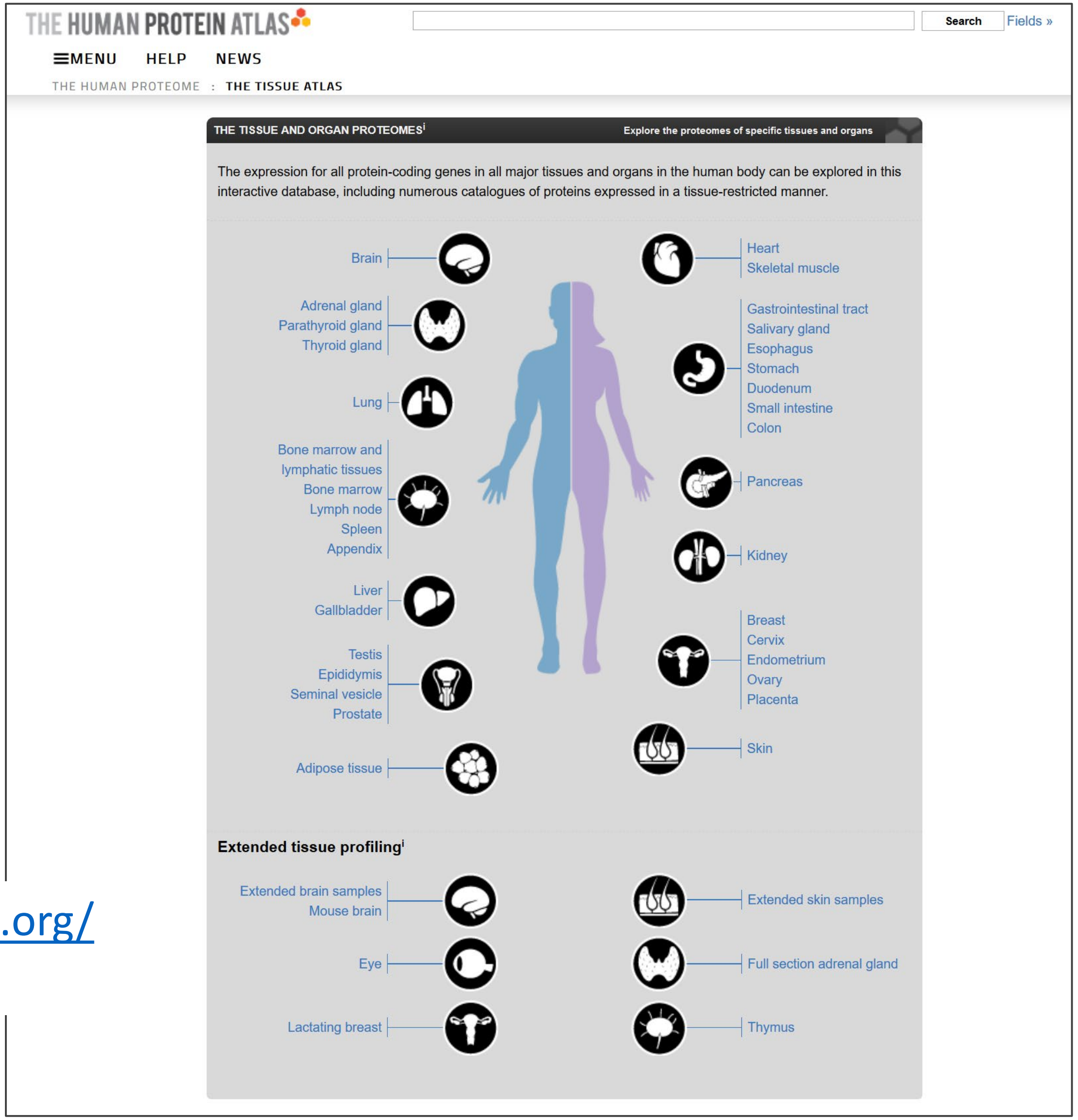




https://bbp.epfl.ch/nexus/cell-atlas



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

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