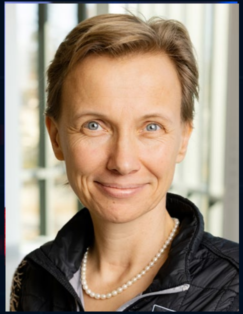
Insightful and Actionable Visual Analytics

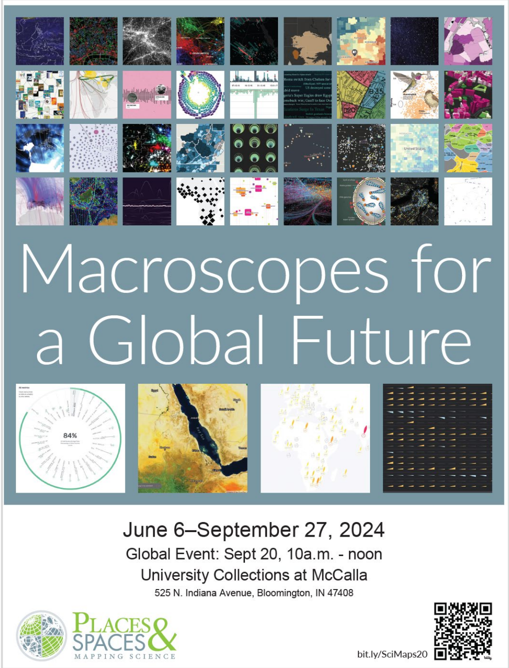
Katy Börner @katycns

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

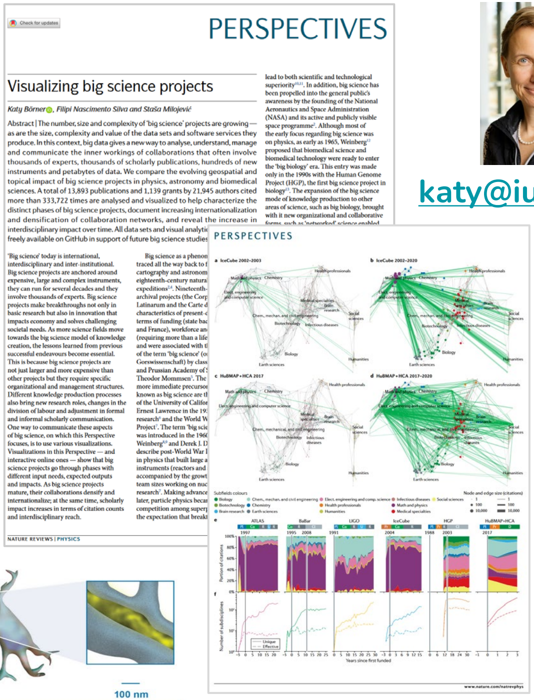


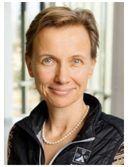
National Graduate Institute for Policy Studies (GRIPS), Minato, Tokyo

November 11, 2024



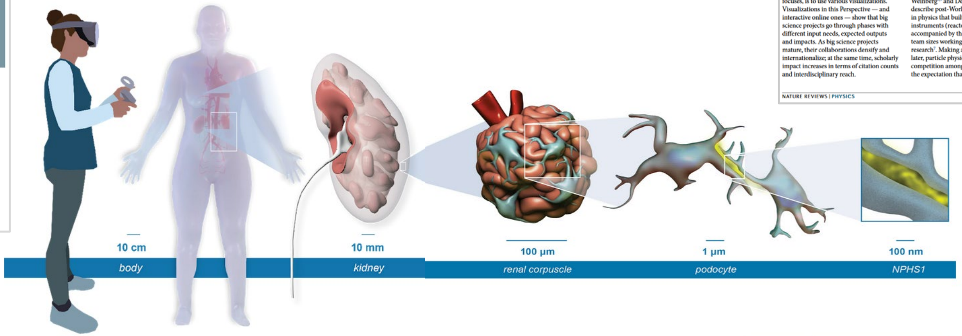
Mapping Science, Education, Technology, Jobs, Policy Making interactions





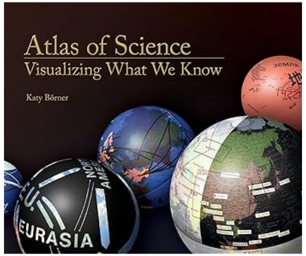
katy@iu.edu

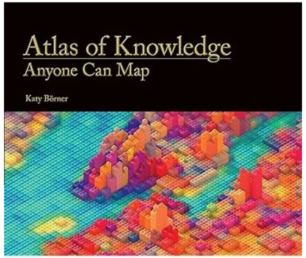
Mapping the Human Body at single-cell resolution, across scales, in 3D

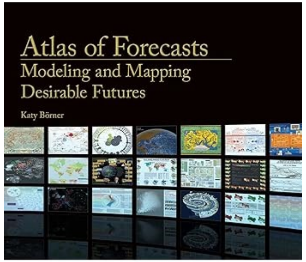


https://scimaps.org

https://humanatlas.io



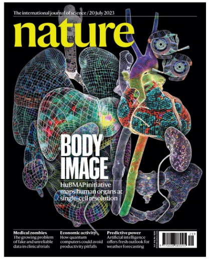


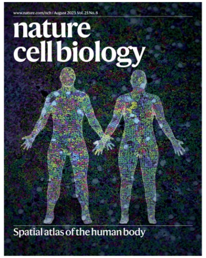


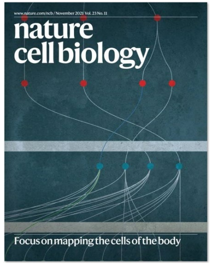
NIH National Institutes of Health

NSF

CIFAR

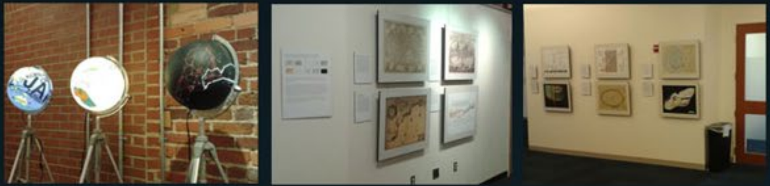






101st Annual Meeting of the Association of American Geographers, Denver, CO.
April 5th - 9th, 2005 (First showing of Places & Spaces)

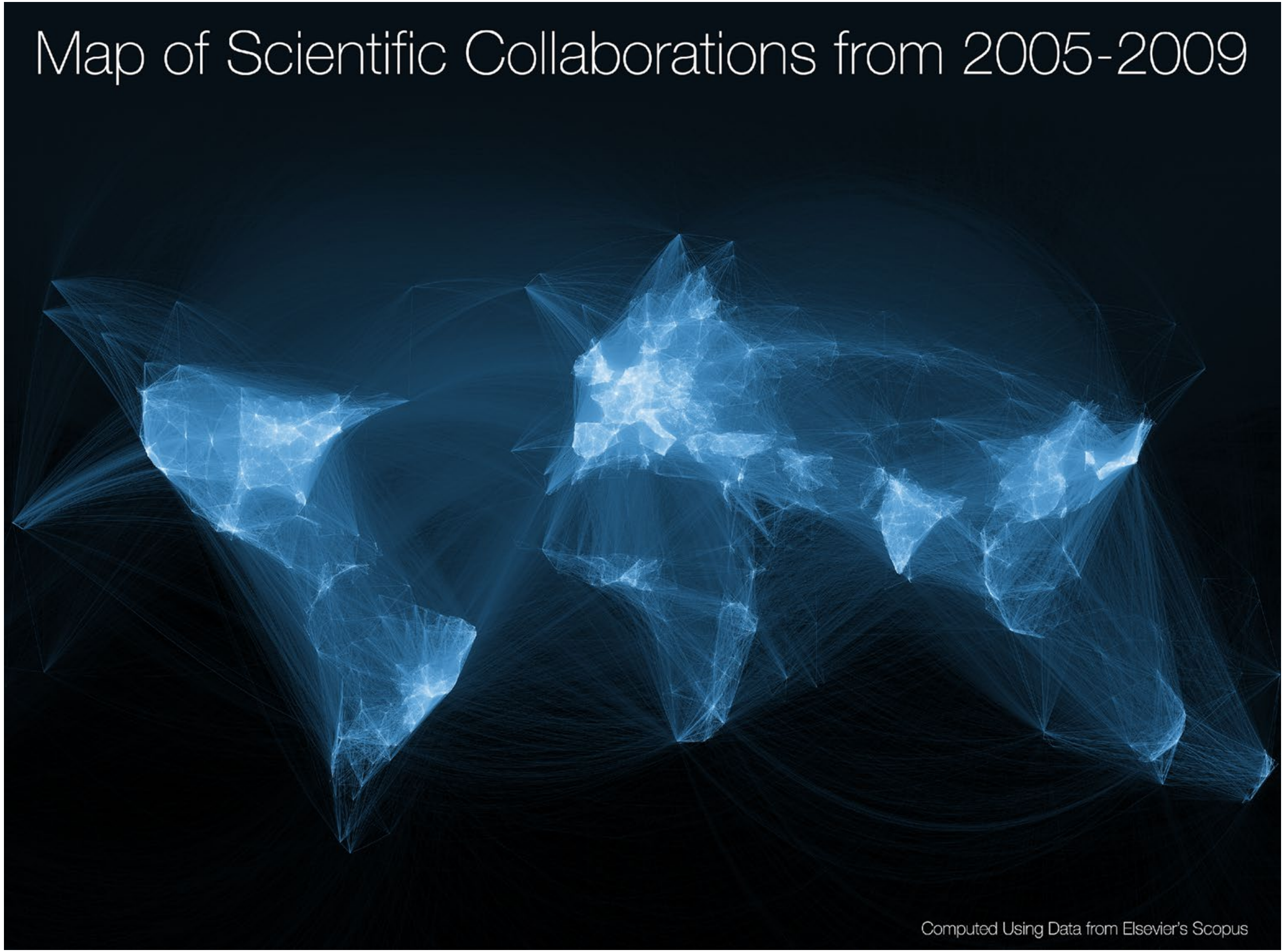
University of Miami, Miami, FL.
September 4 - December 11, 2014.



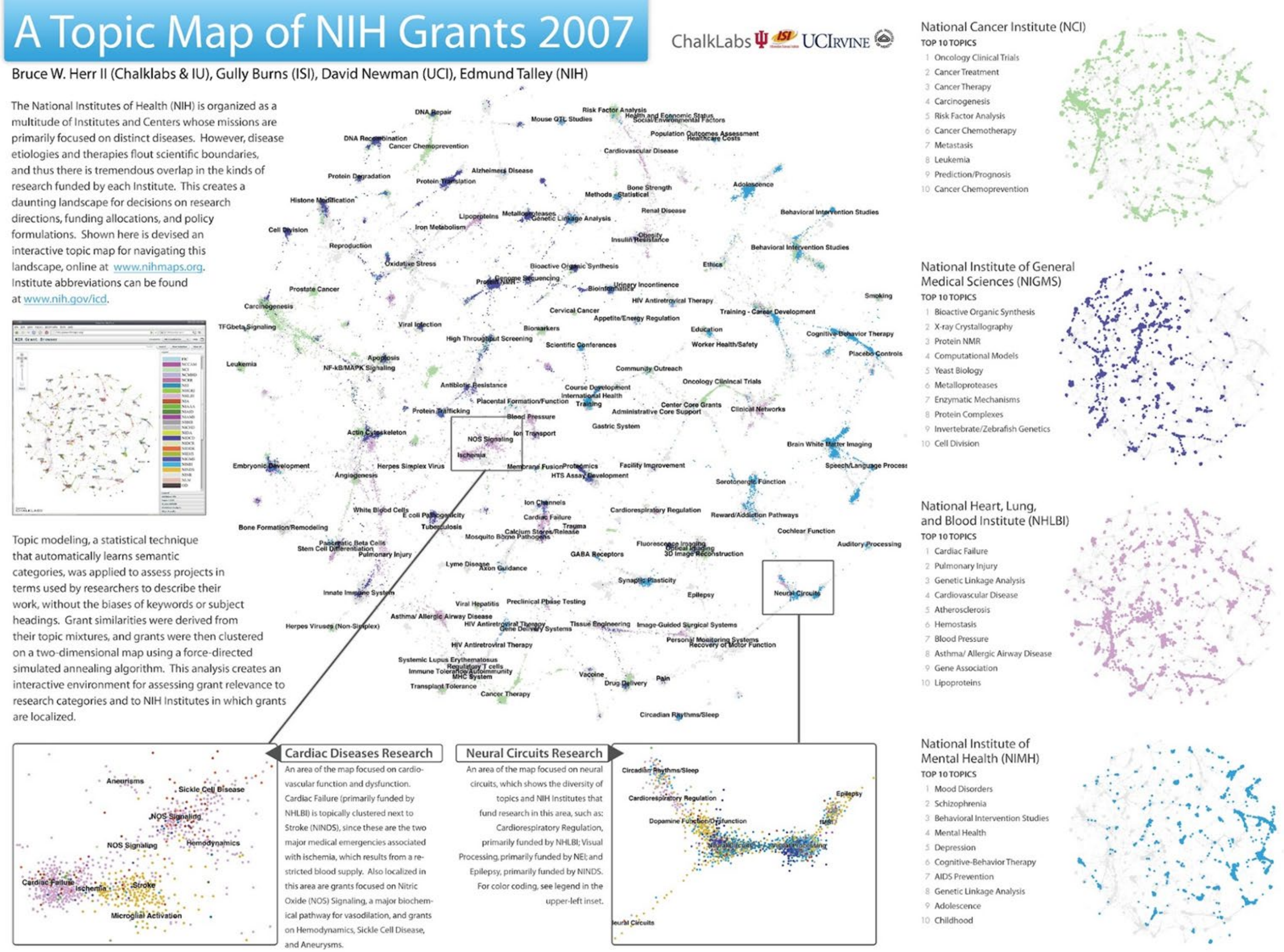
Duke University, Durham, NC.
January 12 - April 10, 2015

http://scimaps.org

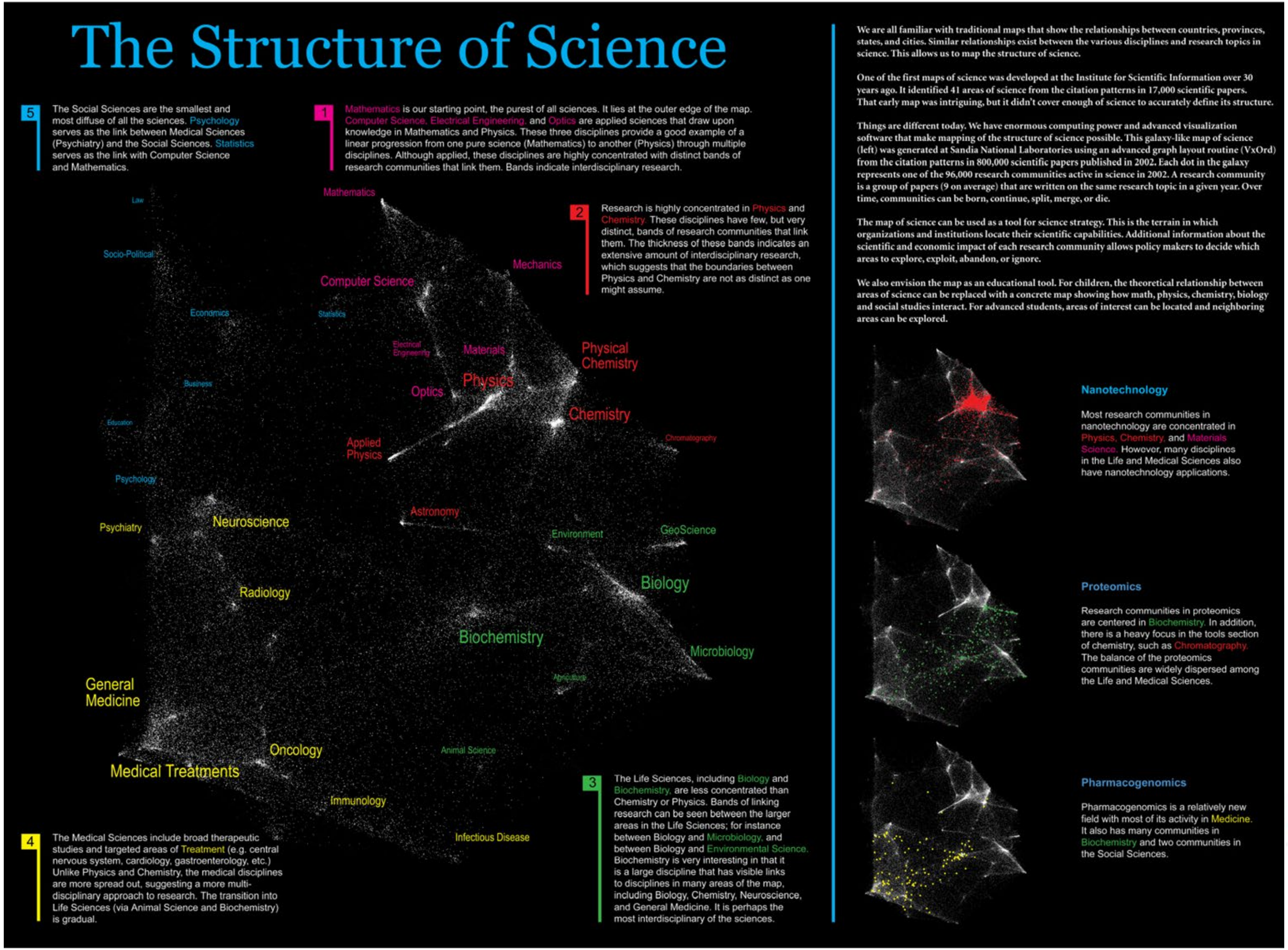
The David J. Sencer CDC Museum, Atlanta, GA.
January 25 - June 17, 2016.



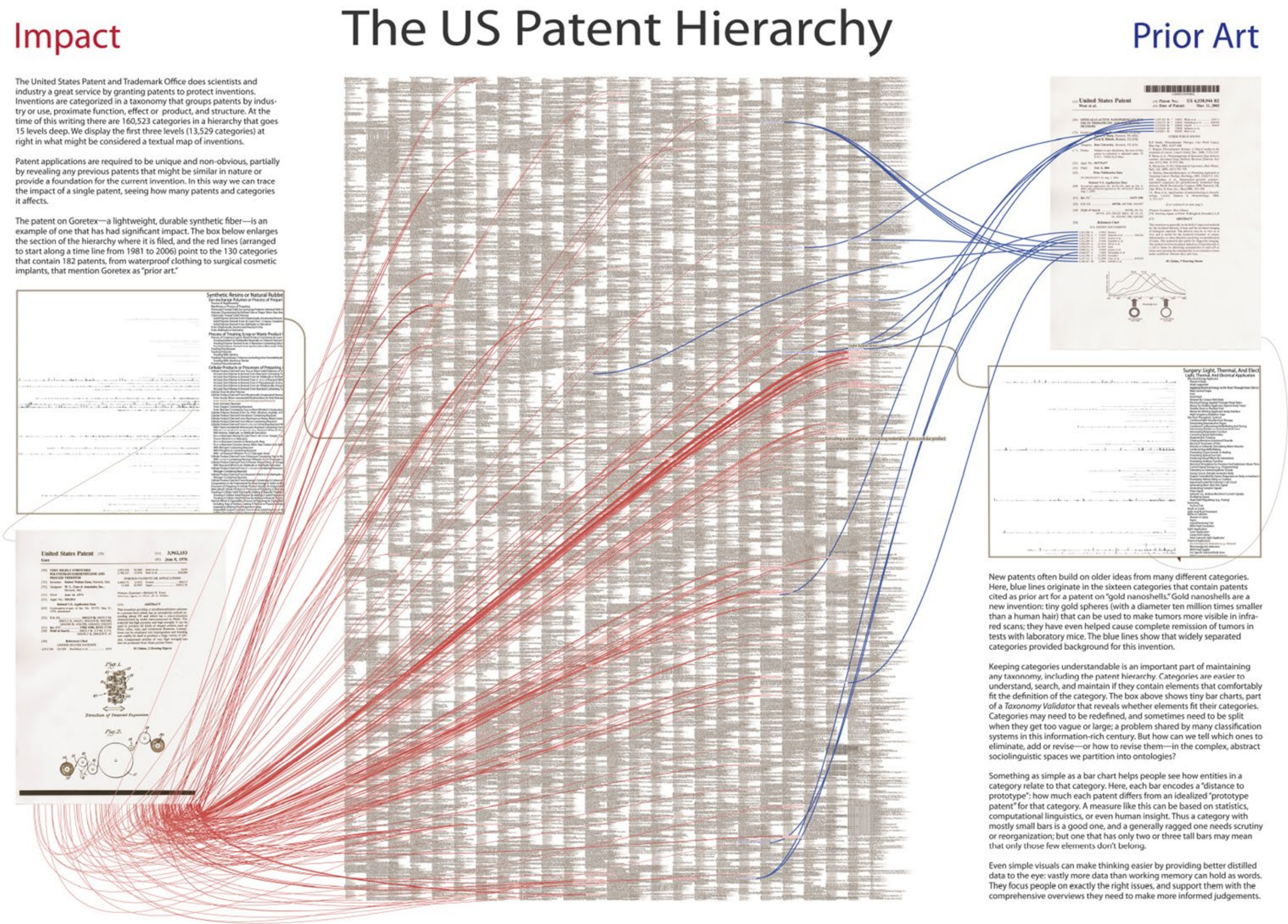
VII.6 Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012



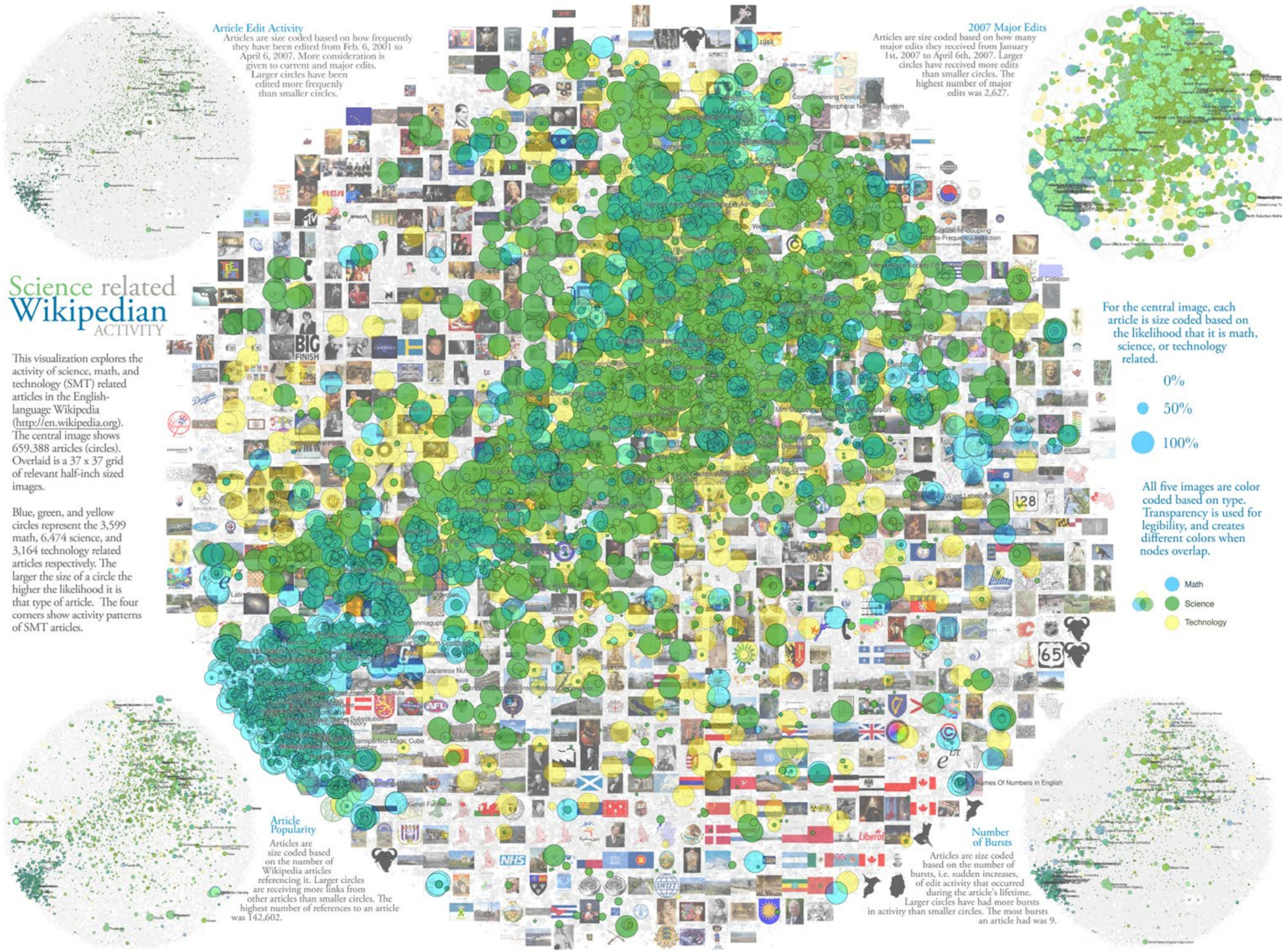
V.7 *A Topic Map of NIH Grants 2007* - Bruce W. Herr II, Gully A.P.C. Burns, David Newman, and Edmund Talley – 2009



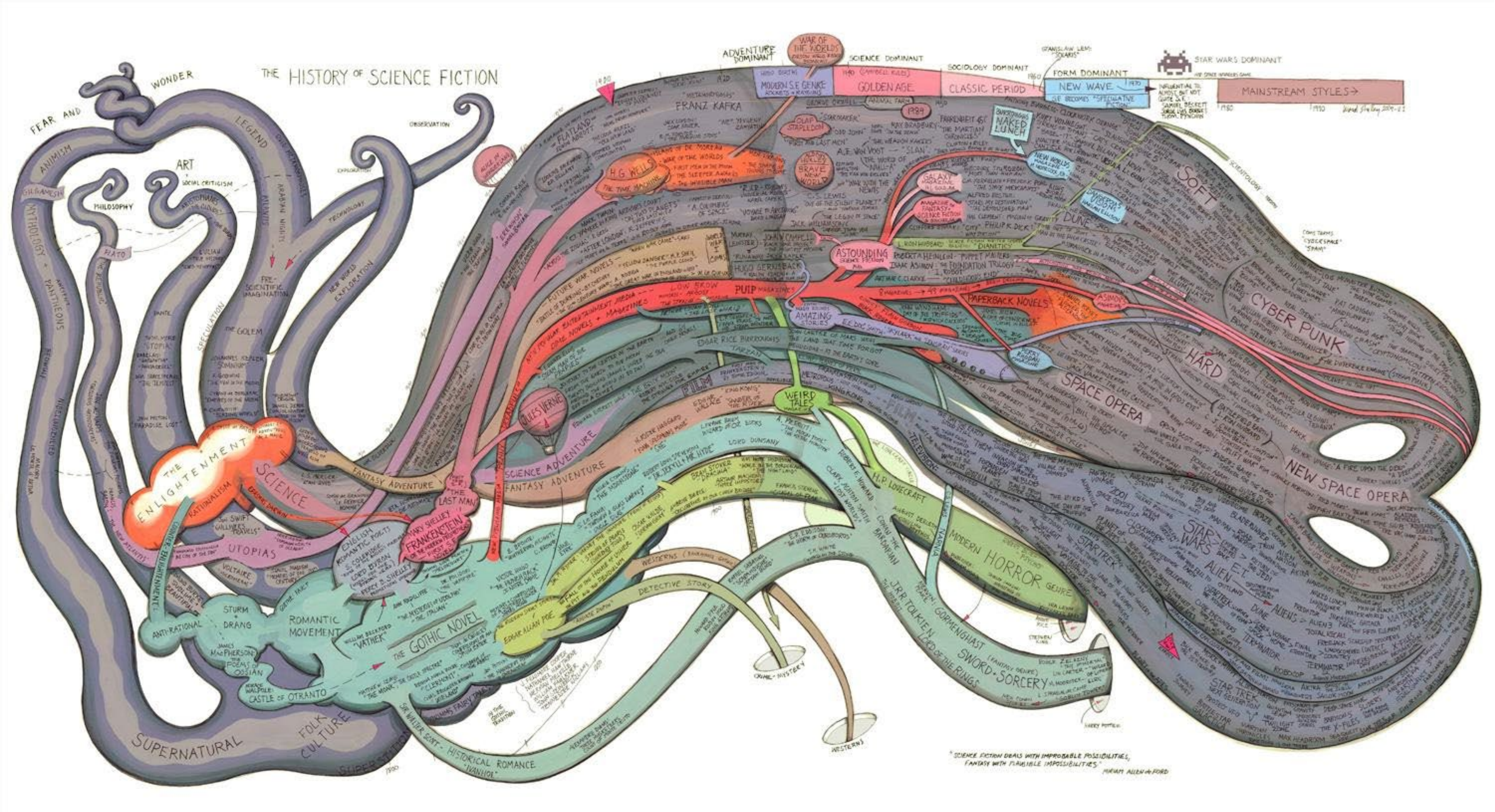
I.10 *The Structure of Science* - Kevin W. Boyack and Richard Klavans - 2005



II.8 *Taxonomy Visualization of Patent Data* - Katy Börner, Elisha F. Hardy, Bruce W. Herr II, Todd Holloway, and W. Bradford Paley - 2006



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



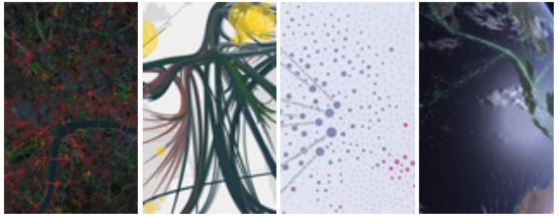
VII.10 *History of Science Fiction* - Ward Shelley - 2011

Check out our Zoom Maps online!

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

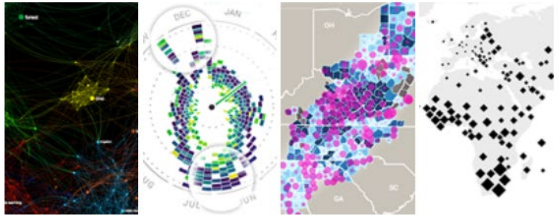
Iteration XI (2015)
Macroscopes for Interacting with Science

Iteration XII (2016)
Macroscopes for Making Sense of Science

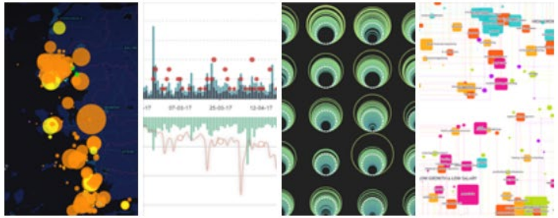


Iteration XIII (2017)
Macroscopes for Playing with Scale

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



Iteration XV (2019)
Macroscopes for Tracking the Flow of Resources



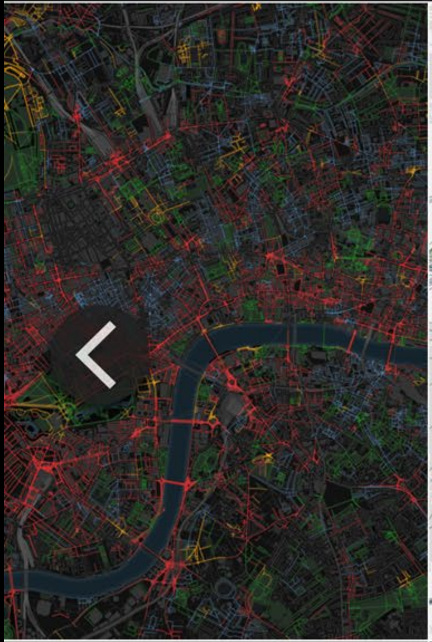
Iteration XVI (2020)
Macroscopes for Harnessing the Power of Data

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

About *Places & Spaces*

PLACES & SPACES
MAPPING SCIENCE

About macroscopes

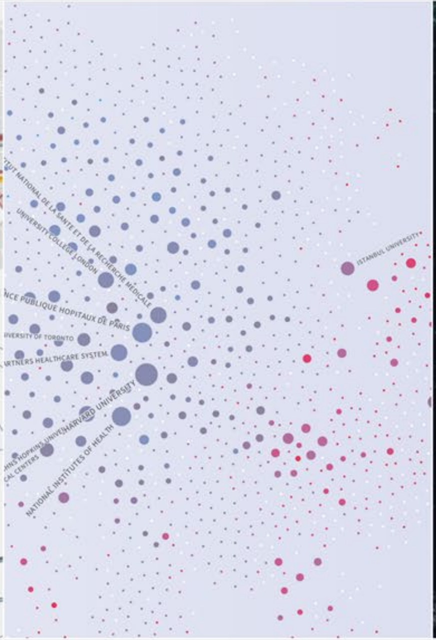


Smelly Maps

Charting urban smellscapes

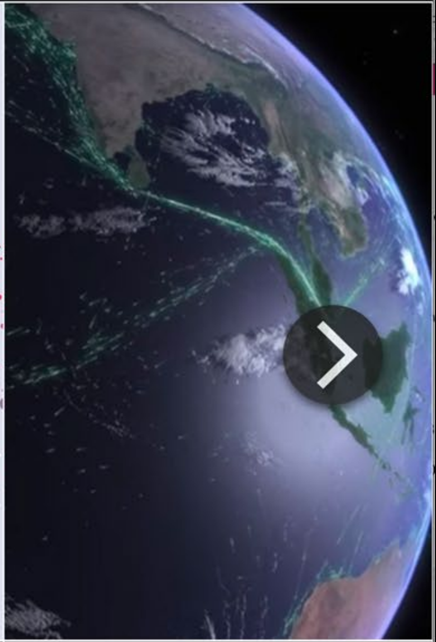
HathiTrust

Storehouse of knowledge



Excellence Networks

Publish or perish together



FleetMon Explorer

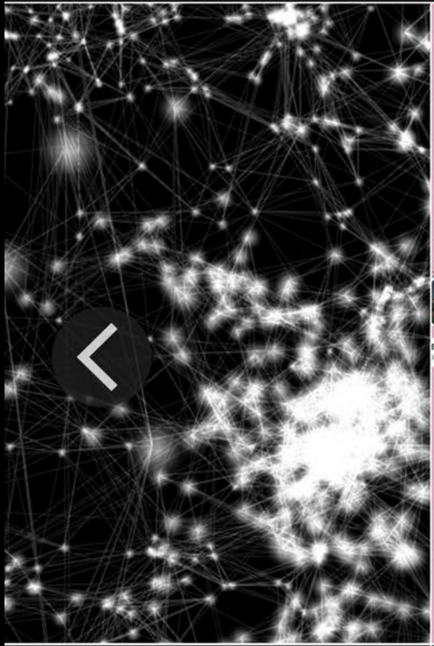
Tracking the seven seas

Smelly Maps – Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello – 2015

About *Places & Spaces*

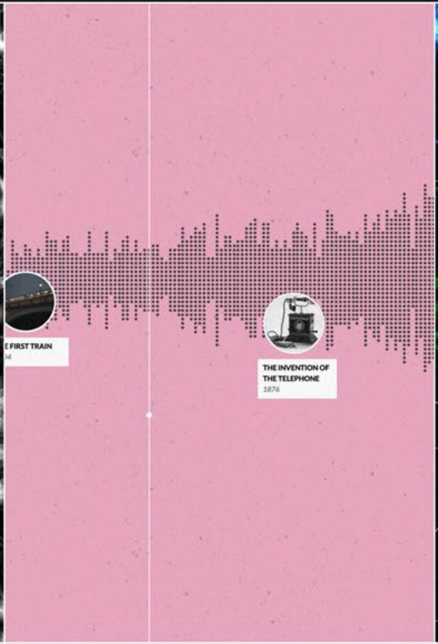
PLACES & SPACES
MAPPING SCIENCE

About macroscopes



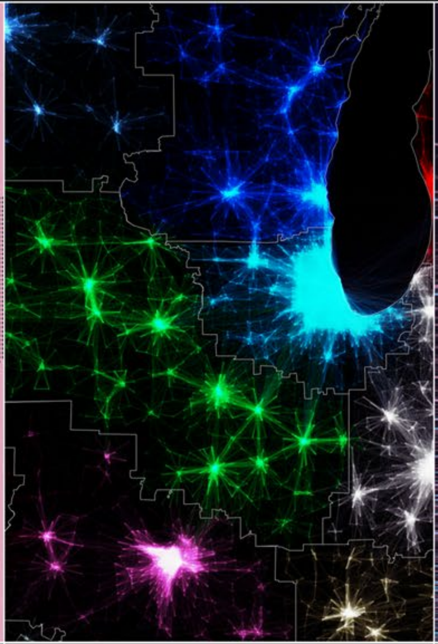
The Cosmic Web

And the network behind it



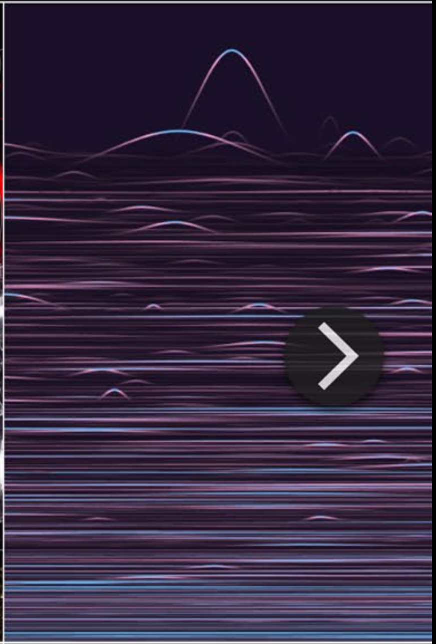
Histography

An interactive timeline



Megaregions of the US

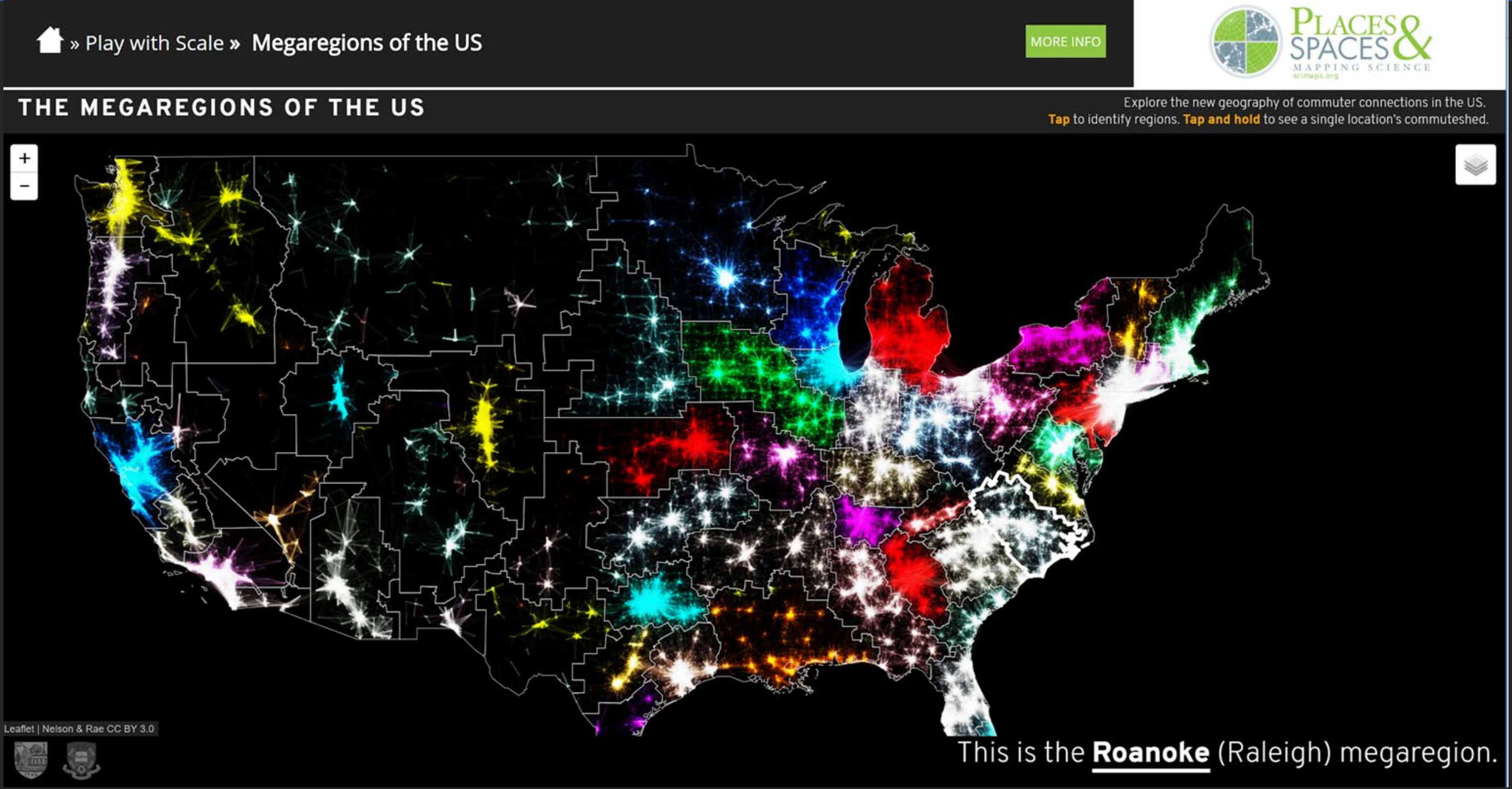
Mapping commuter patterns



Science Paths

The random impact rule

7 1

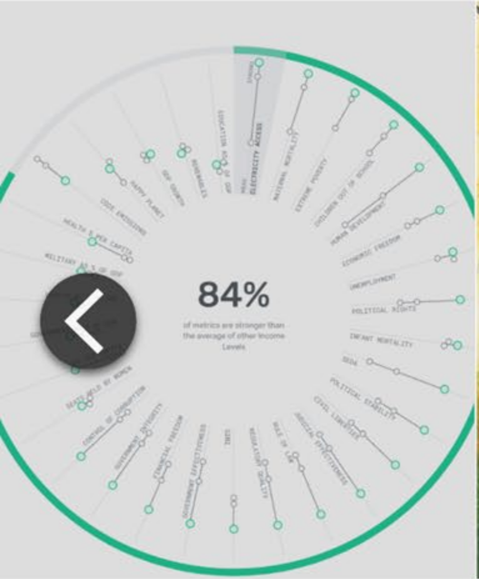


Megaregions of the US –Garrett Dash Nelson and Alasdair Rae – 2016

About *Places & Spaces*

PLACES & SPACES
MAPPING SCIENCE

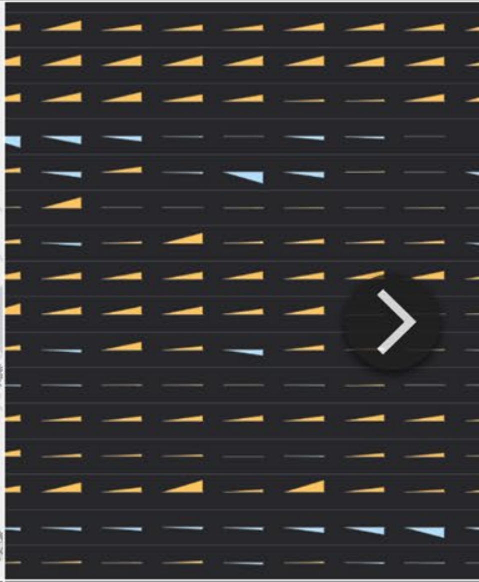
About macroscopes



How Do We Compare?
Using metrics for global good

River Runner
Don't stop that drop

The Whole Picture
The cost of connectedness



The Shape of Change
A global progress report

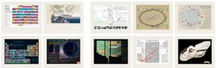
Places & Spaces: Mapping Science Exhibit

1st Decade (2005-2014)

Maps

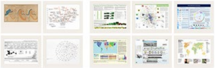
Iteration I (2005)
The Power of Maps

Iteration II (2006)
The Power of Reference Systems



Iteration III (2007)
The Power of Forecasts

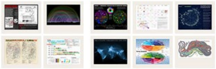
Iteration IV (2008)
Science Maps for Economic Decision Makers



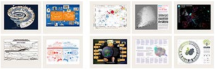
Iteration V (2009)
Science Maps for Science Policy Makers

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

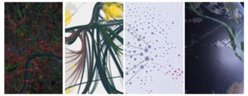
2nd Decade (2015-2024)

Macroscopes

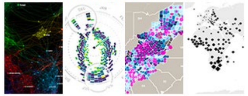
Iteration XI (2015)
Macroscopes for Interacting with Science

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



100
MAPS
in large format, full color, and high resolution.

248
MAPMAKERS
from fields as disparate as art, urban planning, engineering, and the history of science.

43
MACROSCOPE MAKERS
including one whose job title is "Truth and Beauty Operator."

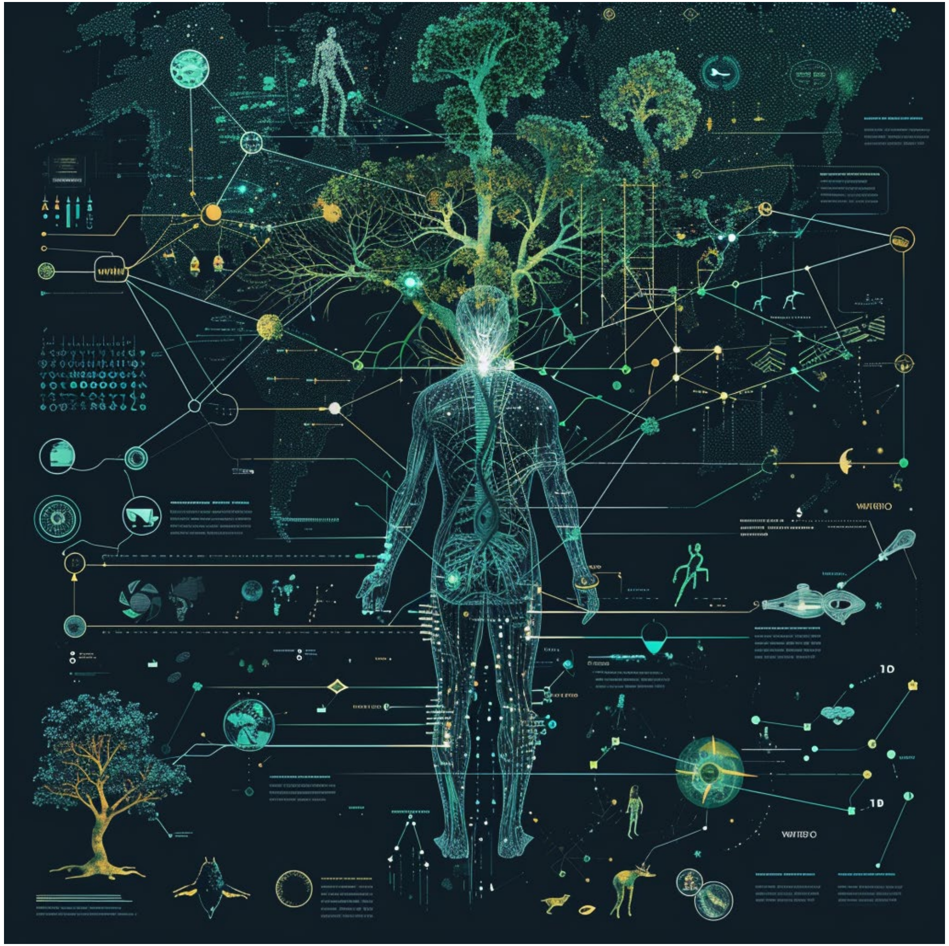
20
MACROSCOPES
for touching all kinds of data.

382
DISPLAY VENUES
from the Cannes Film Festival to the World Economic Forum.

354
PRESS ITEMS
including articles in *Nature*, *Science*, *USA Today*, and *Wired*.

http://scimaps.org

3rd Decade of *Places & Spaces*: Envisioning Intelligences (2025-2034)



wermind using https://www.midjourney.com

https://scimaps.org

Envisioning Intelligences

Including

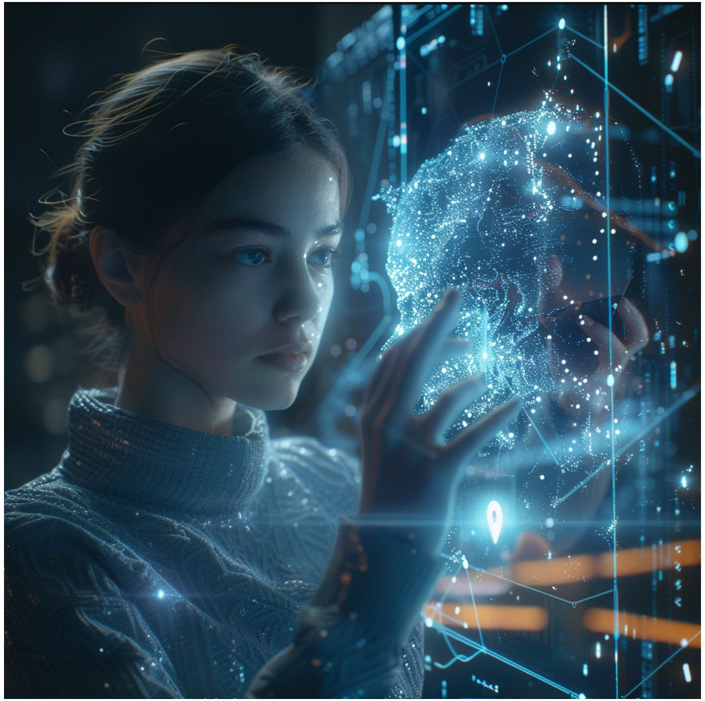
- linguistic, kinesthetic, communication, musical, emotional, and other intelligences by biological and technological life forms

with a focus on

- collaboration & coordination across life forms and intelligence types

to inspire discussion about

- existing and future sensors & actuators, memory & reasoning, exploration & communication, plus shared goals & desirable futures.



.master_of_code_global using
https://www.midjourney.com

Acknowledgements

Exhibit Curators

The exhibit team: Lisel Record, Katy Börner, and Todd Theriault.

Plus, we thank the more than 250 authors of the 100 maps and 16 interactive macroscopes.

http://scimaps.org

Exhibit Advisory Board

Gary Berg-Cross
Cognitive psychologist (PhD, SUNY-Stony Brook). Potomac, MD, USA

André Skupin
Associate Professor of Geography at San Diego State University, California

Donna J. Cox, MFA, Ph.D.
Director of the Advanced Visualization Laboratory at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, IL, USA

Moritz Stefaner
Freelance designer on the crossroads of data visualization, information aesthetics, and user interface design in Germany

Bonnie DeVarco
Media X Distinguished Visiting Scholar at Stanford University, Palo Alto, CA, USA

Olga Subirós
Curator of Big Bang Data and Founder of Olga Subirós Studio in Barcelona, Spain

Peter A. Hook
Head of Digital and Scholarly Services and LawArXiv Administrator, Cornell Law Library. Ithaca, NY, USA

Stephen Uzzo
Vice President of Science and Technology for the New York Hall of Science

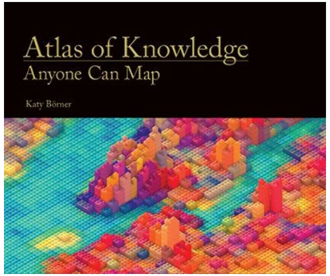
Francis Harvey
Professor of Visual Communication in Geography at the Liebnitz Institute for Regional Geography, Leipzig University, Germany

Benjamin Wiederkehr
Founding Partner and Managing Director of Interactive Things in Zürich, Switzerland

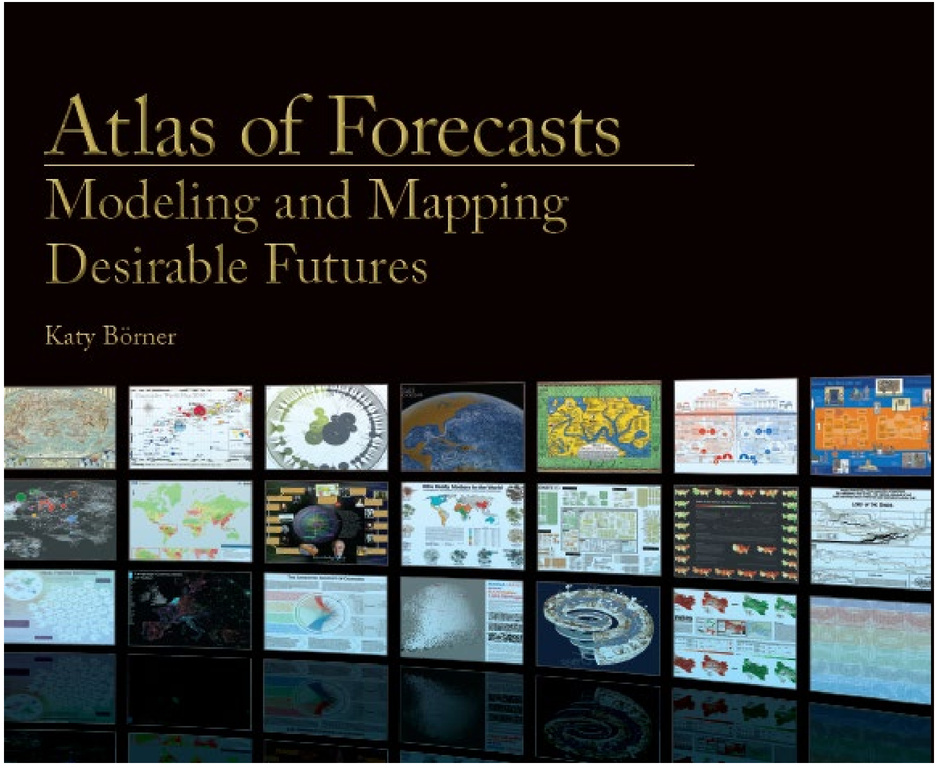
Lev Manovich
Professor, The Graduate Center, City University of New York; Director, Software Studies Initiative (big data, digital humanities, visualization)

Atlas Trilogy

2010



2015



2021

https://mitpress.mit.edu/books/atlas-forecasts

About | Location | Scientific Committee Members | Agenda | Photo Gallery | Registration & Logistics | Submission Information | Doctoral Forum | FAQ

About ISSI 2023

https://cns-iu.github.io/workshops/2023-07-02_issi

The ISSI 2023 Conference welcomes scientists, research managers, authorities and information professionals to engage in an open and proactive debate of the current status and advancements of informetric and scientometric theories and their deployment. Research, development and practice in library and information science, data mining, information retrieval, history of science and philosophy of science are the main focus but contributions on the analysis and visualization of education, job market, and policy data are most welcome.

Indiana University in Bloomington, IN, USA is the main organizer of the conference in close collaboration with scientific committee members and under the auspices of ISSI – the International Society for Informetrics and Scientometrics (http://www.issi-society.org). The conference language is English.

ISSI

International Society for Scientometrics and Informetrics

iConference 2025

Living in an AI-gorithmic world

Virtual Academic Program: March 11 - 14, 2025

Onsite Academic Program in Indiana Bloomington, USA: March 18 - 22, 2025

https://www.ischools.org/iconference

Hosted by Indiana University

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Data Visualization Literacy

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

Börner, Katy (2015) *Atlas of Knowledge: Anyone Can Map*. The MIT Press.

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.

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 17 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

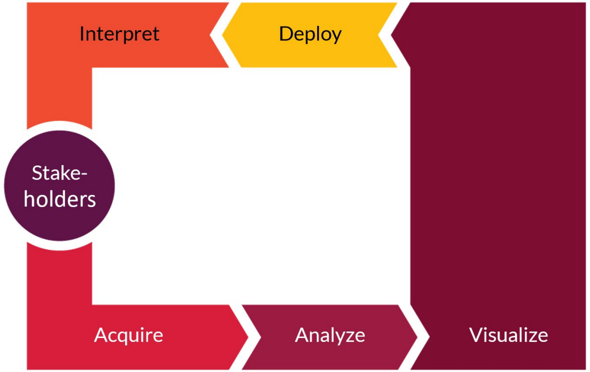
4 **Visualizations**
- table
- chart
- graph
- map
- tree
- network

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

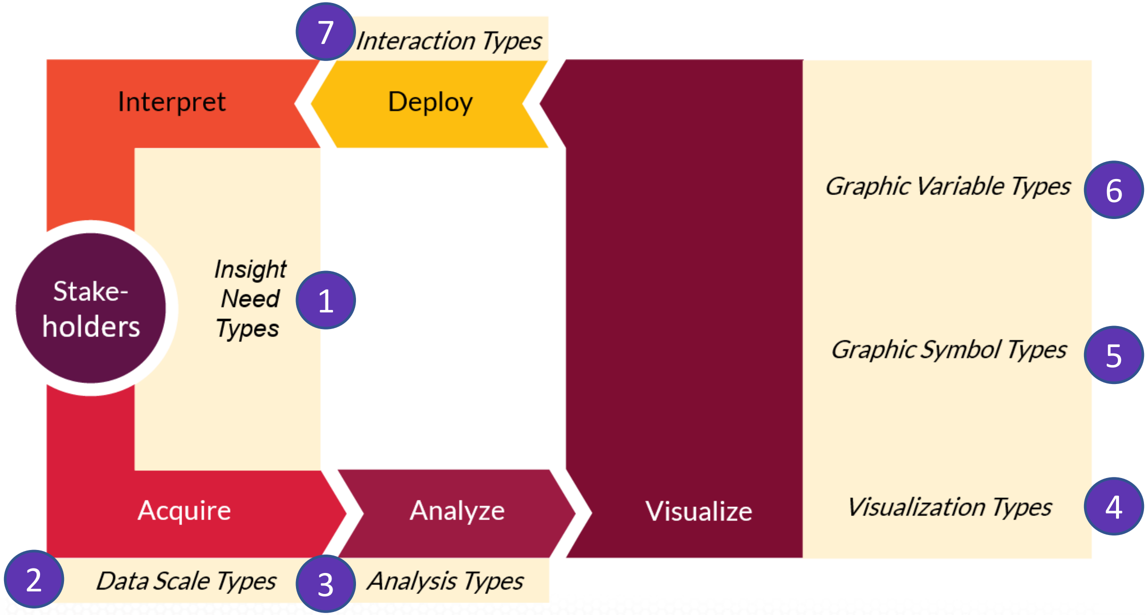
DVL Workflow Process
Defines 5 steps required to render data into insights.

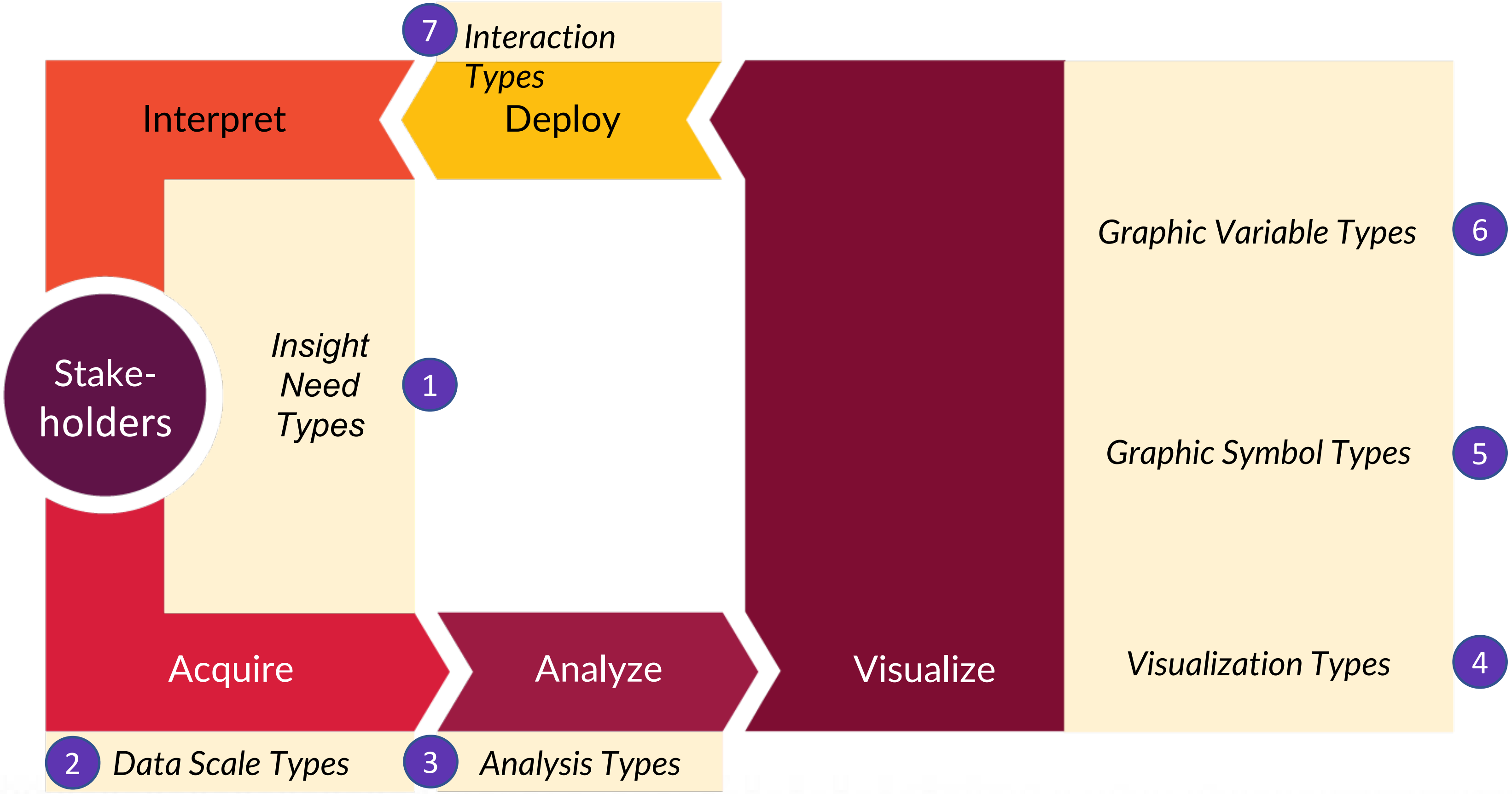


Data Visualization Literacy Framework (DVL-FW)

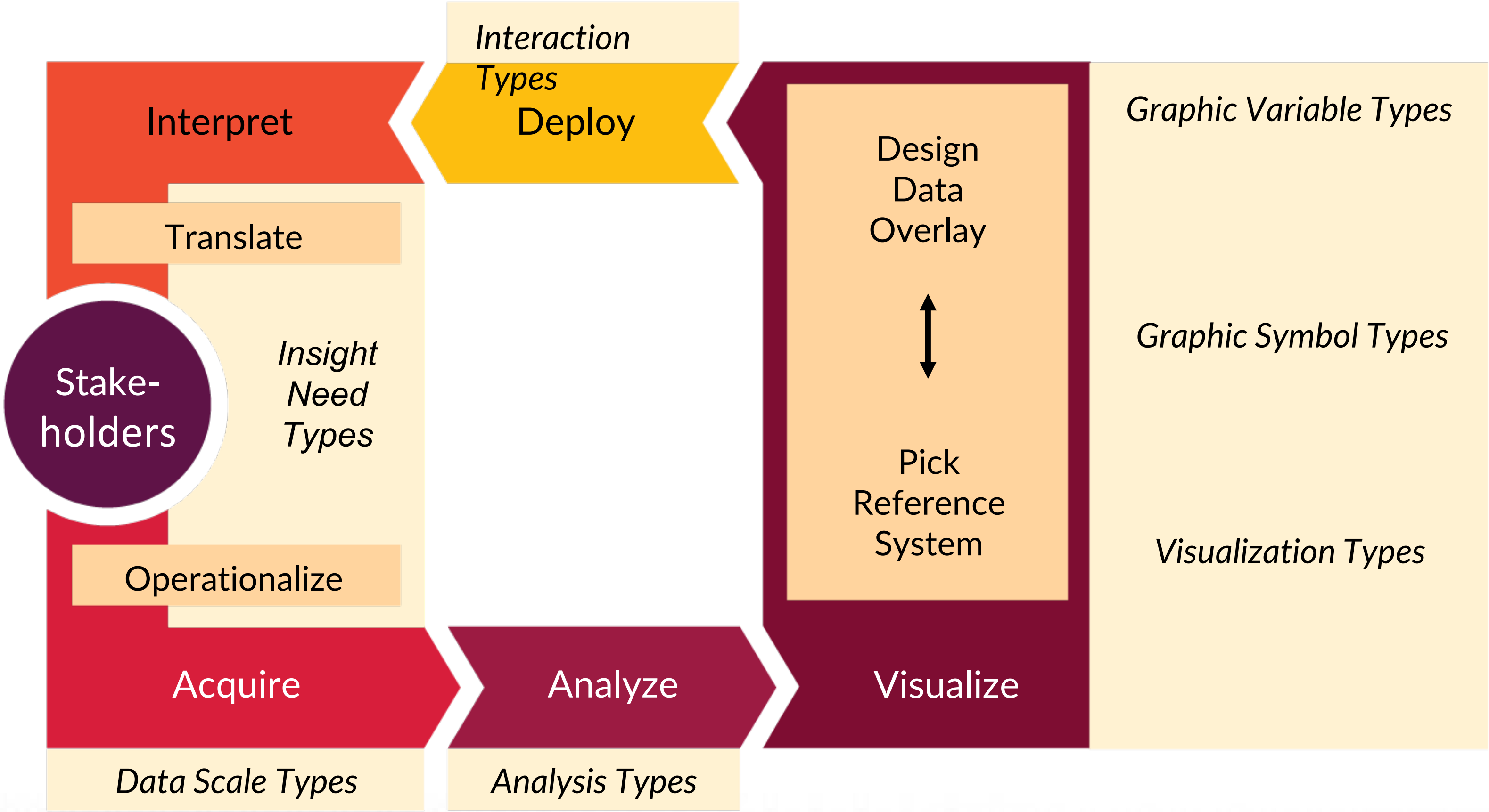
Consists of two parts that are interlinked:

DVL Typology +
DVL Workflow
Process

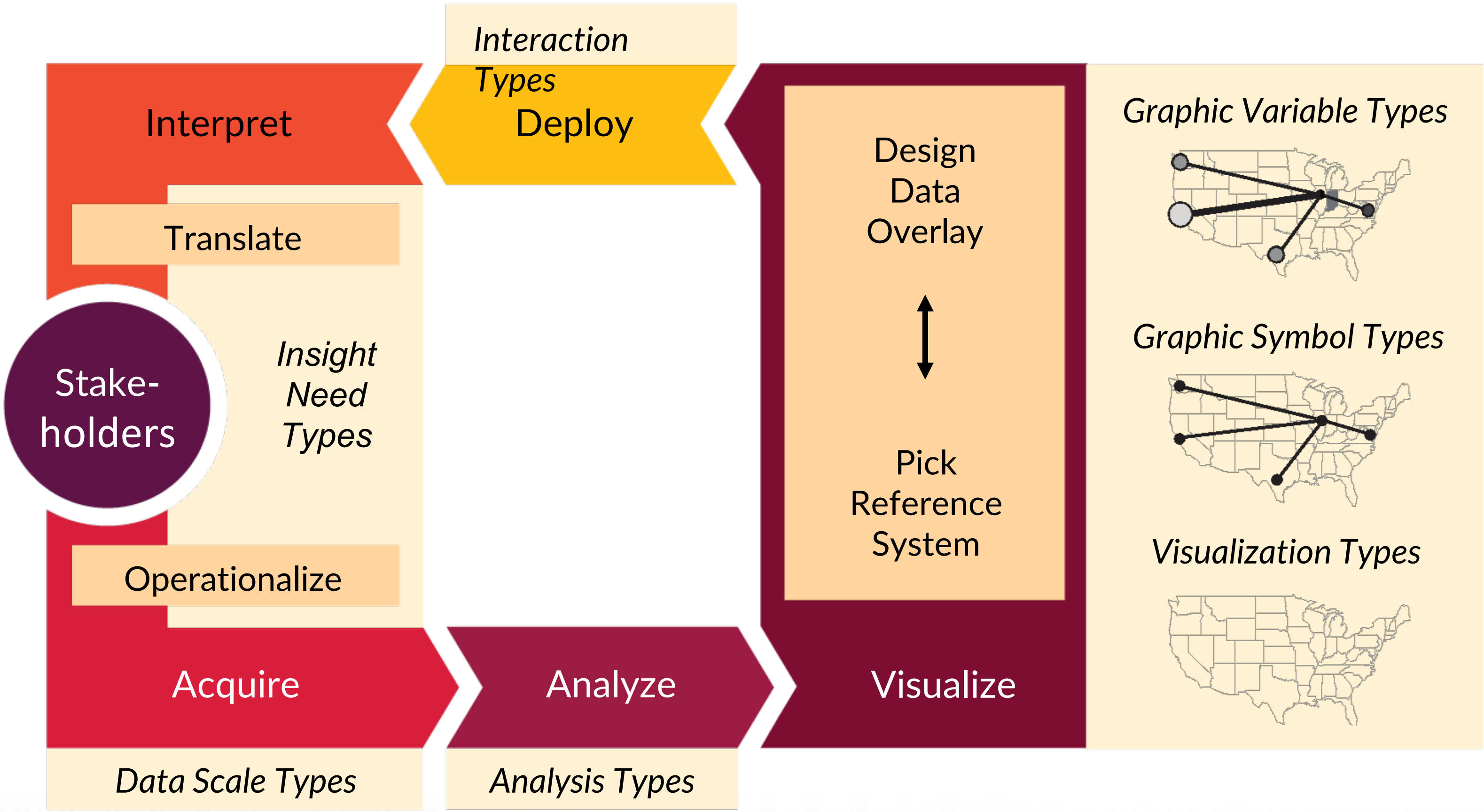




CNS Cyberinfrastructure for Network Science Center



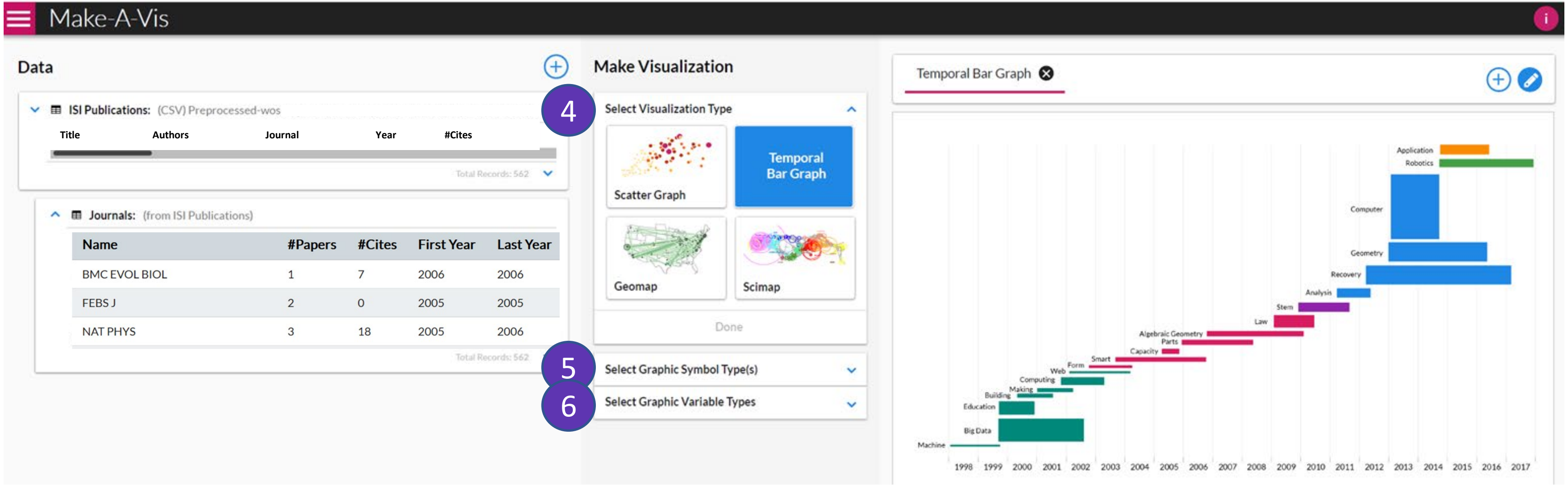
CNS Cyberinfrastructure for Network Science Center



Interaction Types
Interpret
Deploy
Translate
Stake-holders
Insight Need Types
Operationalize
Design Data Overlay
Pick Reference System
Graphic Variable Types
Graphic Symbol Types
Visualization Types
Acquire
Analyze
Visualize
Data Scale Types
Analysis Types

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	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. 26-27.

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

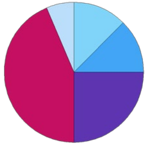
4

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
	• ratio	• topical	• map	area	form	• details-on-demand
• comparisons		• relational	• tree	surface	color	• history
• trends (process and time)			• network	volume	optics	• extract
				• linguistic symbols	motion	• link and brush
• geospatial				text		• projection
• compositions (also of text)				numerals		• distortion
				punctuation marks		
• correlations/ relationships				• pictorial symbols		
				images		
				icons		
				statistical glyphs		

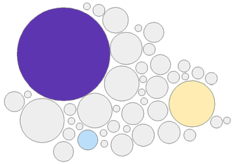
Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 30-31.

Visualization Types

Chart

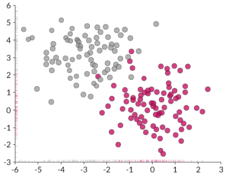


Pie Chart

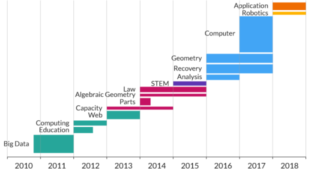


Bubble Chart

Graph

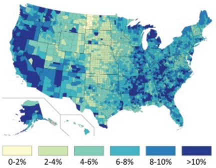


Scatter Graph

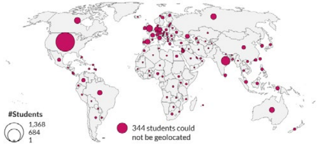


Temporal Bar Graph

Map

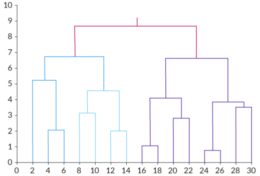


Choropleth Map

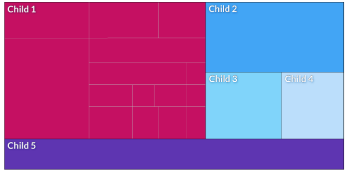


Proportional Symbol Map

Tree

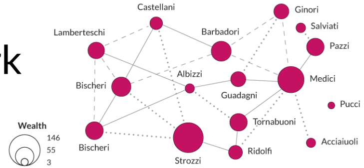


Dendrogram

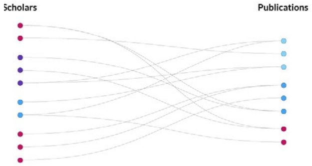


Tree Map

Network



Force-Directed Network Layout



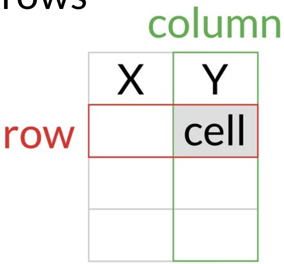
Bimodal Network Layout

Visualize: Reference Systems

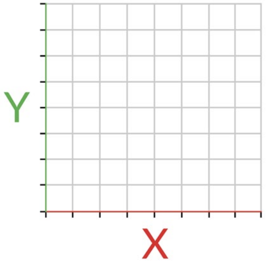
4

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

Table
columns by rows

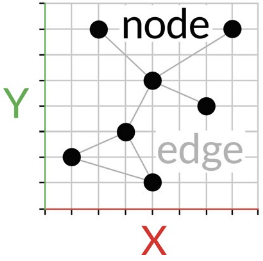


Graph
x-y coordinates

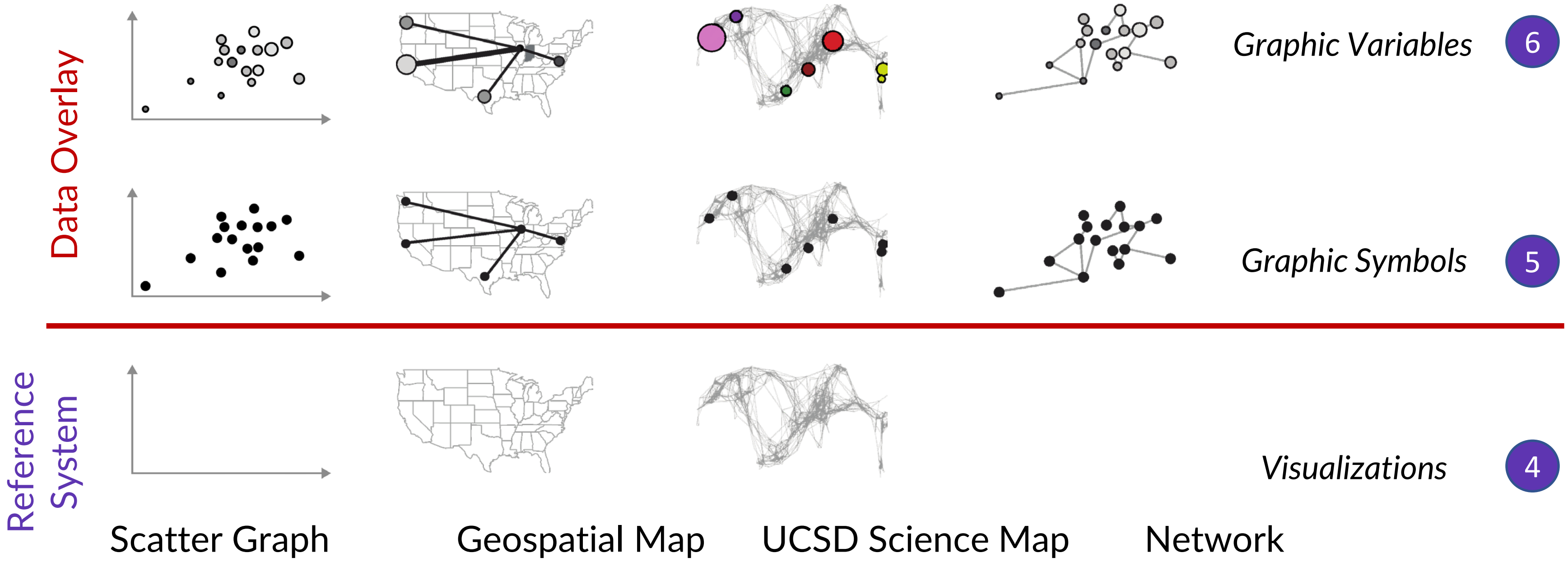


Map
latitude/ longitude

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
- 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

Qua

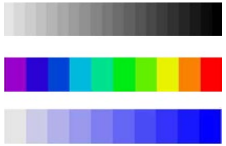
Form:

- Size
- Shape
- Rotation (Orientation)

Quantitative

Color:

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



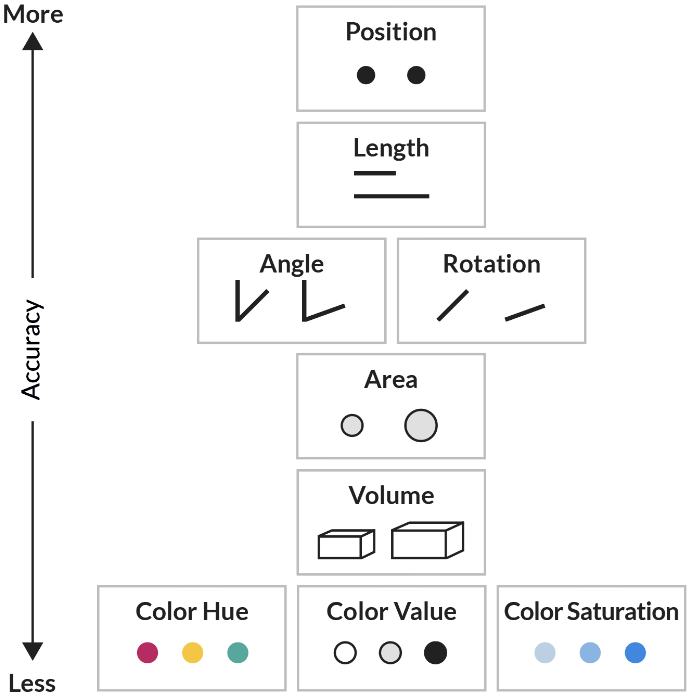
Qua

Qua

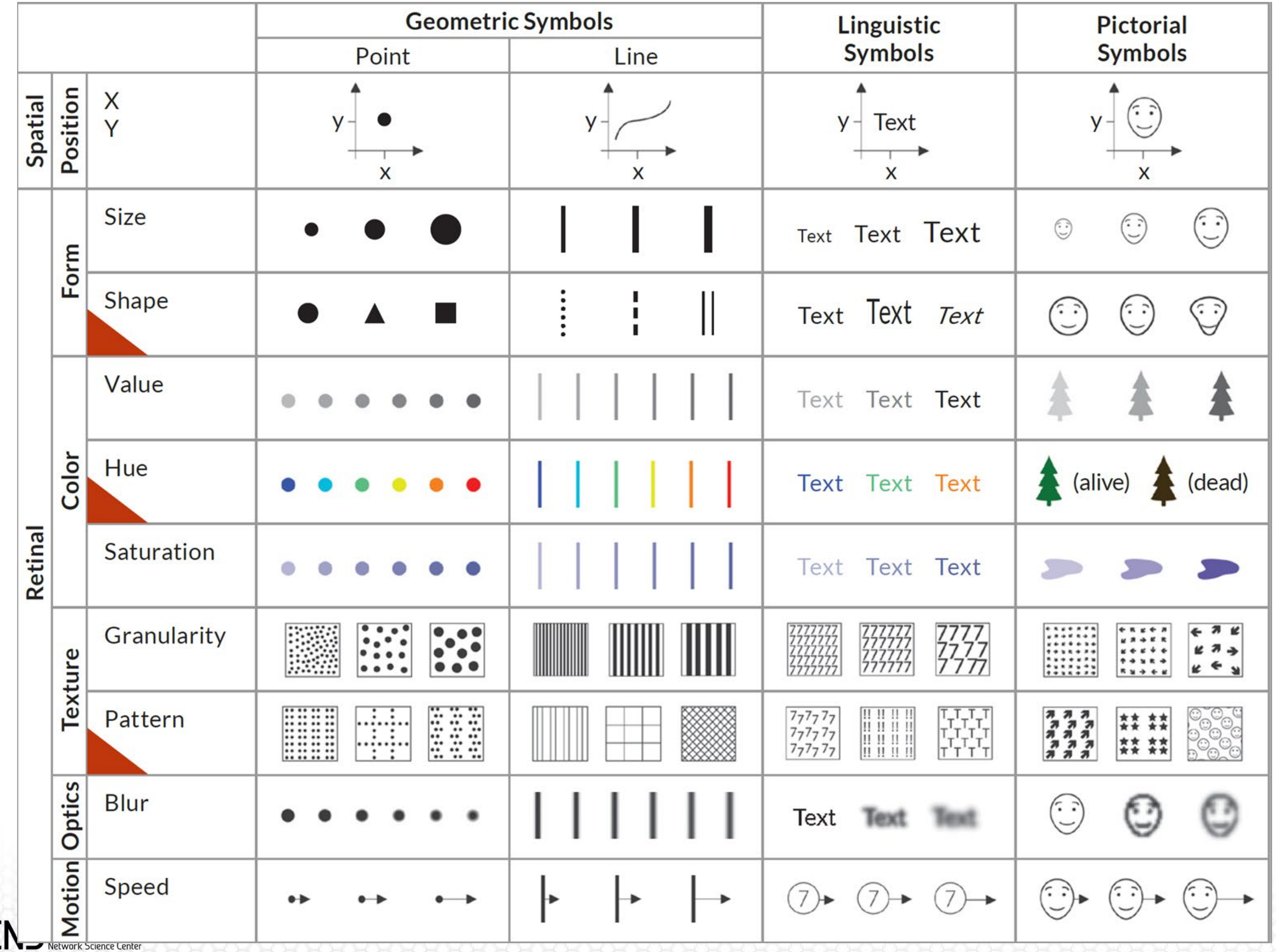
Optics: Blur, Transparency, Shading, Stereoscopic Depth

Texture: Spacing, Granularity, Pattern, Orientation, Gradient

Motion: Speed, Velocity, Rhythm



Graphic Variable Types



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

Qualitative

Also called:

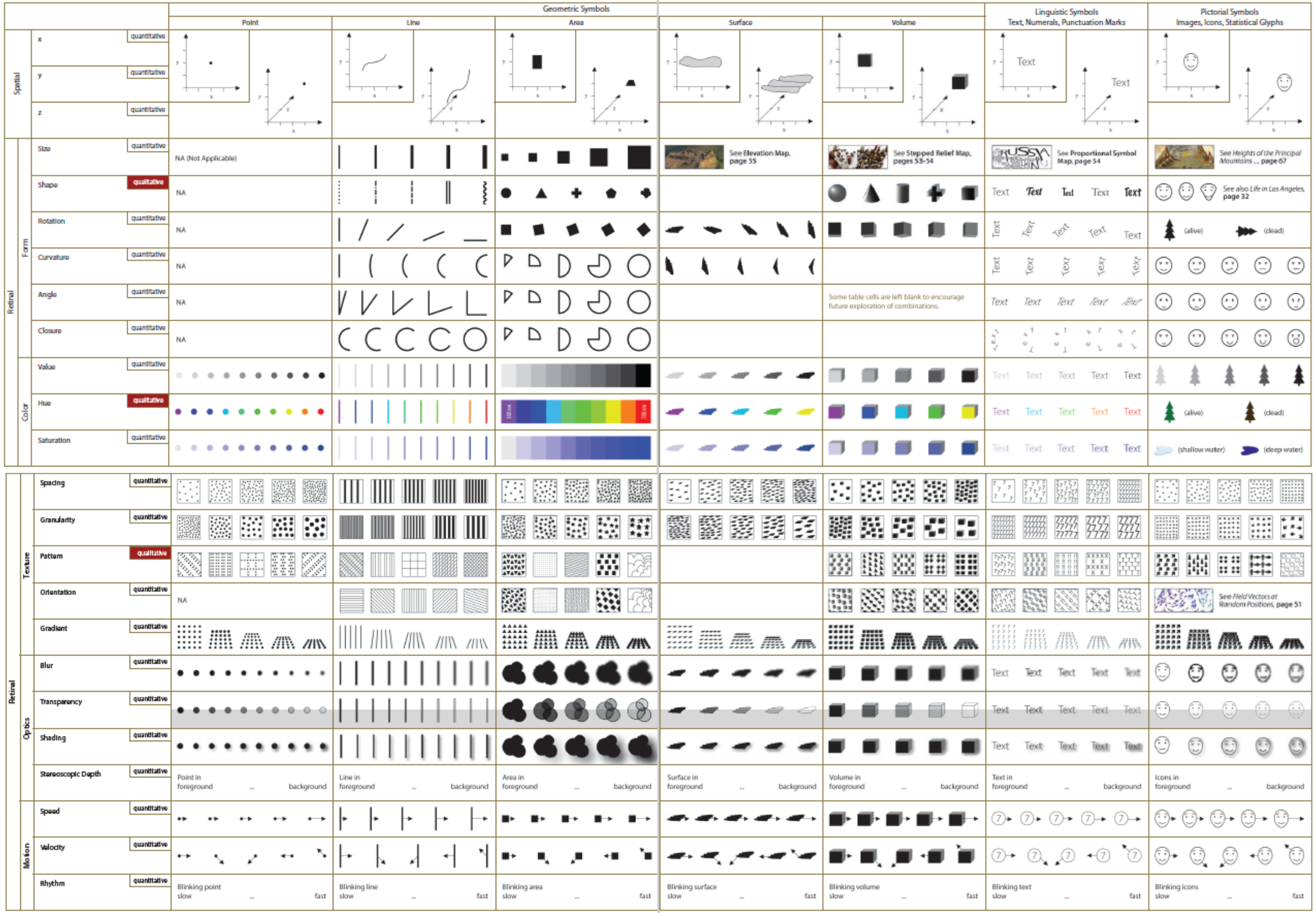
Categorical Attributes
Identity Channels

Quantitative

Also called:

Ordered Attributes
Magnitude Channels

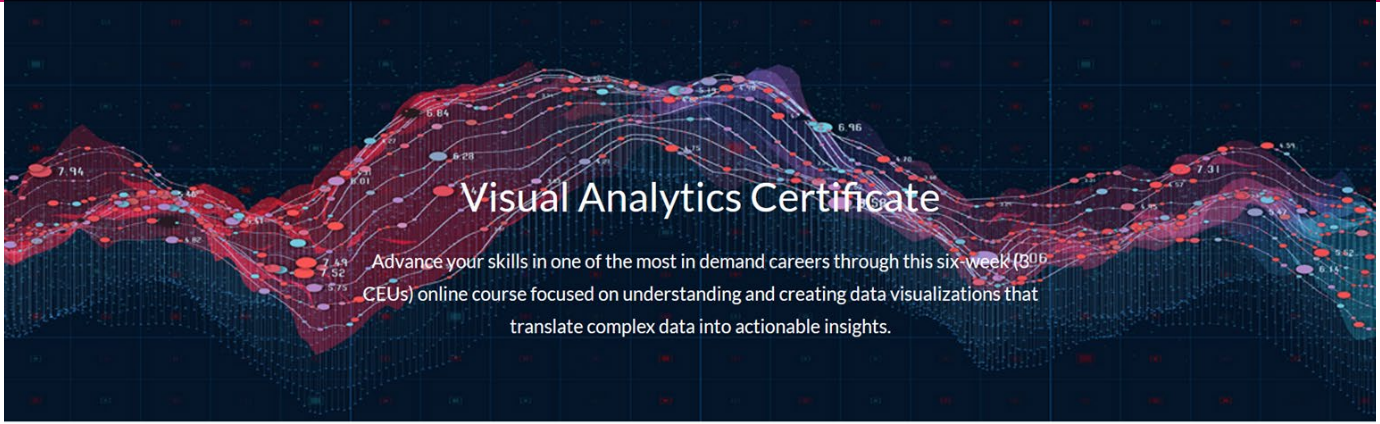
Graphic Variable Types Versus Graphic Symbol Types



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

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
LUDDY SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING



FLYER | REGISTER FOR JAN 9–FEB 19, 2023 | FAQS

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

US Employers which have sent employees include
The Boeing Company, Eli Lilly, DOE, CDC, NSWC Crane.

Visual Analytics Certificate: Submit a Case Study

Collaborate with us on solutions that make a true difference. Submit a project description and we collaborate with your team on efficient solutions plus training materials that speed up adoption by key decision makers.

SUBMIT

Human Reference Atlas Literature

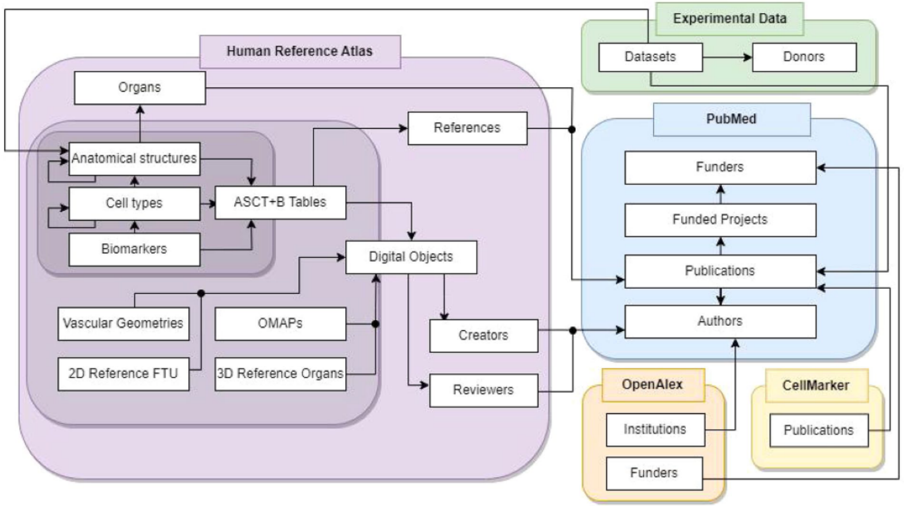


Fig. 1 Overview of the HRAlit database. HRA data types are linked to experimental data and to PubMed data using HuBMAP IDs of the HRA digital objects; experimental dataset IDs, donor IDs, and publication DOIs; PubMed publication PMIDs, author IDs, funding IDs, institution IDs, and funder IDs.

Kong, Yongxin, and Katy Börner. 2024. "Publication, funding, and experimental data in support of Human Reference Atlas construction and usage". *Scientific data* (574): https://doi.org/10.1038/s41597-024-03416-8.

www.nature.com/scientificdata

scientific **data**

Check for updates

OPEN

DATA DESCRIPTOR

Publication, funding, and experimental data in support of Human Reference Atlas construction and usage

Yongxin Kong[1,2] & Katy Börner[1]

Experts from 18 consortia are collaborating on the Human Reference Atlas (HRA) which aims to map the 37 trillion cells in the healthy human body. Information relevant for HRA construction and usage is held by experts, published in scholarly papers, and captured in experimental data. However, these data sources use different metadata schemas and cannot be cross-searched efficiently. This paper documents the compilation of a dataset, named HRAlit, that links the 136 HRA v1.4 digital objects (31 organs with 4,279 anatomical structures, 1,210 cell types, 2,089 biomarkers) to 583,117 experts; 7,103,180 publications; 896,680 funded projects, and 1,816 experimental datasets. The resulting HRAlit has 22 tables with 20,939,937 records including 6 junction tables with 13,170,651 relationships. The HRAlit can be mined to identify leading experts, major papers, funding trends, or alignment with existing ontologies in support of systematic HRA construction and usage.

Background & Summary

Constructing an atlas of the healthy human body is a massive undertaking due to the multiscale, biological complexity of human physiology. Since March 2020, international experts funded by the National Institutes of Health and/or supported by the Human Cell Atlas have been collaborating on the construction of a Human Reference Atlas (HRA)[1]. The 5th release of the HRA (v1.4) was published in June 2023 and comprises 31 organs with 4,279 unique anatomical structures, 1,210 unique cell types, 2,089 unique biomarkers linked to 32 Anatomical Structures, Cell Types, plus Biomarkers (ASCT + B) tables, 21 two-dimensional functional tissue units (FTU), and 65 three-dimensional, anatomically correct reference organs[2]. A total of 101 experts created and 99 experts reviewed (158 unique experts with ORCID IDs) the HRA digital objects across all releases and compiled 420 papers with DOIs that provide scholarly evidence for the anatomical structures, cell types, and biomarkers in the 31 ASCT + B tables.

As the HRA grows in the number of organs and data types it captures, it becomes important to use data-driven decision making to ensure systematic and efficient collaboration of scholars from different areas of research and development; federation of experimental data from different laboratories and data portals across scales (whole body to subcellular); and strategic foresight when setting data acquisition, tool development, and funding priorities.

In parallel to atlas construction, many high-quality experimental datasets are becoming available via data portals developed and served by Human BioMolecular Atlas Program (HuBMAP)[3], Cellular Senescence Network (SenNet)[4], Kidney Precision Medicine Project (KPMP)[5,6], GenitoUrinary Developmental Molecular Anatomy Project (GUDMAP)[7], the Genotype-Tissue Expression (GTEx)[8], or CZ CELLxGENE[9]. However, the portals use different metadata schemas and few provide DOIs for papers and only some offer API access—searching for data across portals is difficult or impossible.

Moreover, HRA relevant data is published in scholarly papers. Each month, more than 80,000 papers are published in PubMed making it difficult to keep track of expertise, methods, data, or code. *Scientific Data* papers typically focus on ontologies[10–12] or experimental data[13–15] while science of science studies commonly focus on

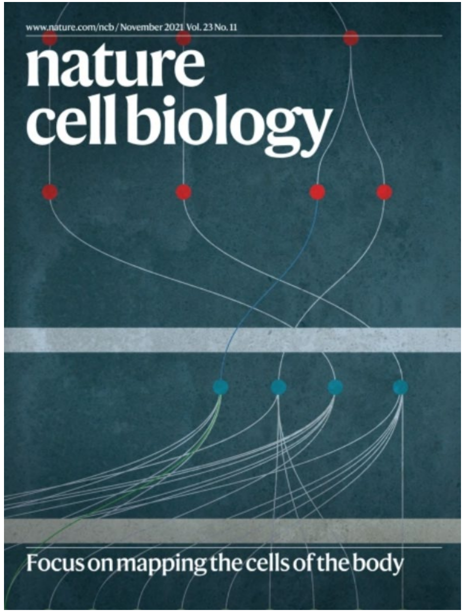
[1]Department of Intelligent Systems Engineering, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, 47408, USA. [2]School of Information Management, Sun Yat-sen University, Guangzhou, 510006, China. e-mail: yokong@iu.edu; katy@iu.edu

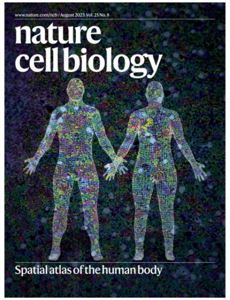
Human Reference Atlas (HRA)

The Human Reference Atlas (HRA)

1. defines the 3D space and shape of anatomical structures and cell types that are of biomedical relevance plus the biomarkers used to characterize them. Anatomical structures, cell types and biomarkers are validated and represented in/added to ontologies (Uberon/FMA, CL, HGNC).
2. defines how new datasets can be mapped to the HRA, e.g., spatially using the Visible Human CCF or Vasculature CCF (or both, see next slide), via ASCT+B ontology terms/IDs, or via gene expression data as in Azimuth.
3. it is
   - authoritative (there exists expert agreement and it was validated by data),
   - computable (supports API queries, UIs),
   - published as LOD (connected to gene, disease, and other ontologies and data),
   - open (anyone can use the HRA data and code), and
   - continuously evolving (e.g., as new technologies become available).

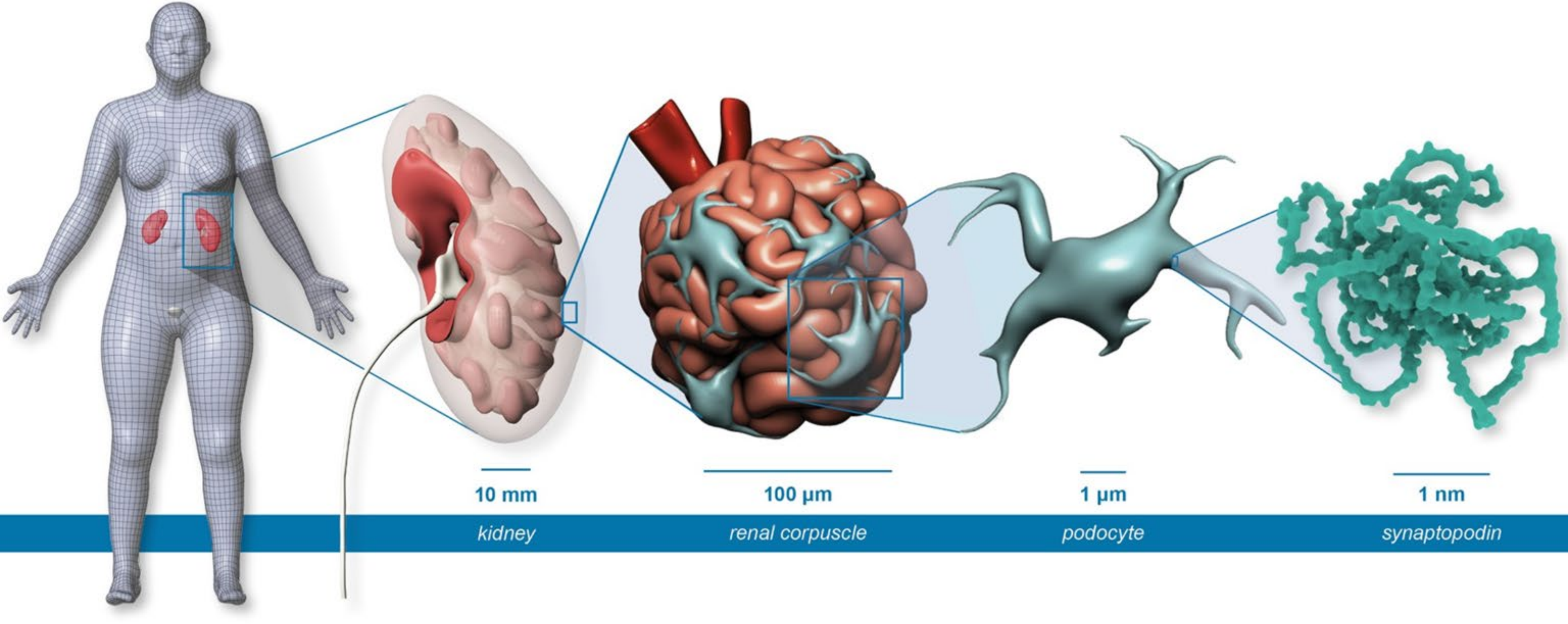
https://www.nature.com/articles/s41556-021-00788-6





Human Reference Atlas

A multiscale, high-resolution, three-dimensional, ontologically aligned atlas of anatomical structures and cells in the healthy human body



HRA-focused HuBMAP HIVE Marker Paper

Paper is now accepted as a ‘Resource’ paper in *Nature Methods*.

The preprint is at https://www.biorxiv.org/content/10.1101/2024.03.27.587041v3

Thanks go to all 170+ Core and HRA Team authors who made this possible.

It is our hope that this joint paper helps align efforts and optimize data formats, APIs.

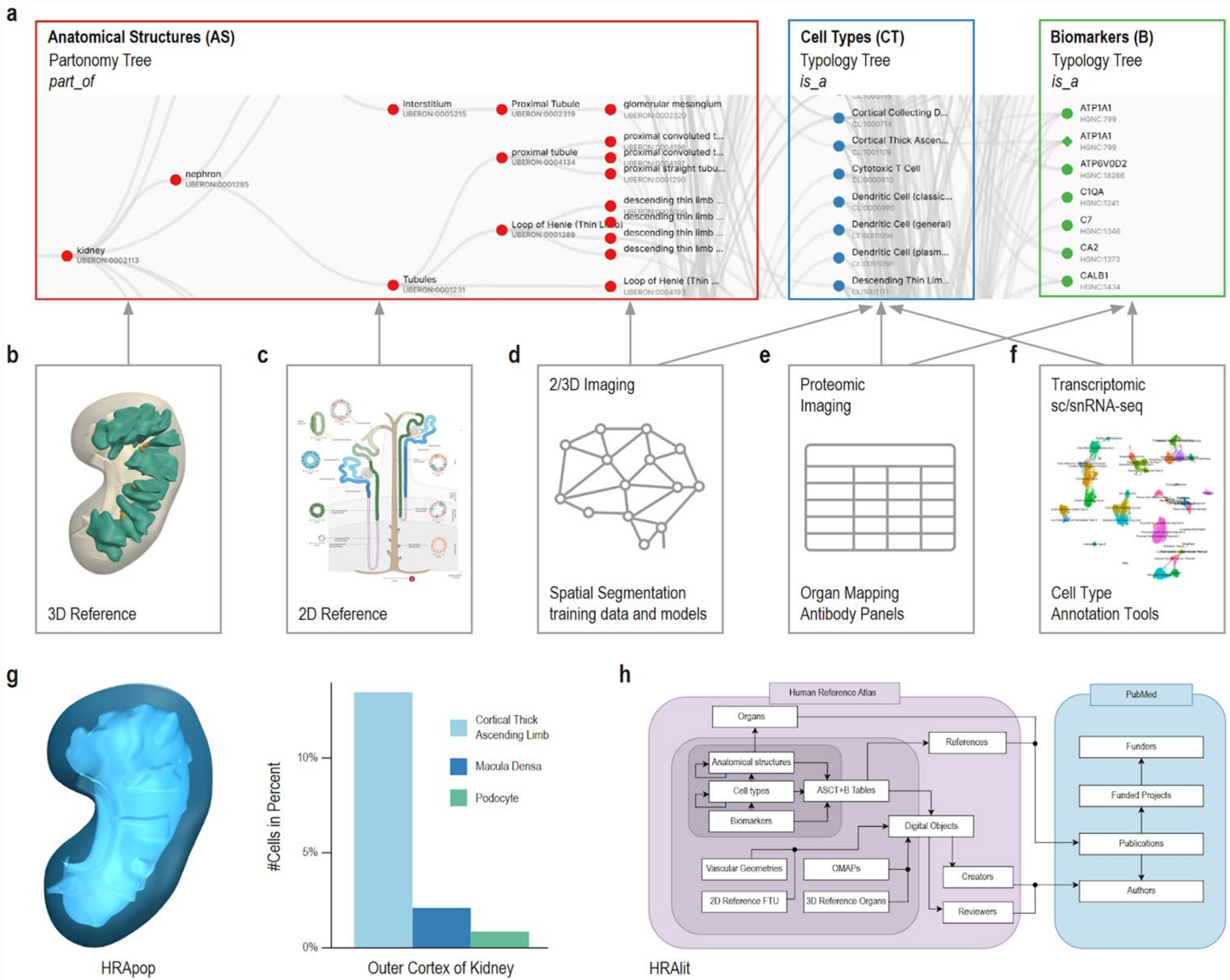


Figure 1: Human Reference Atlas (HRA) components and linkages. a. The anatomical structures, cell types and biomarkers (ASCT+B) tables document the nested *part_of* structure of organs (e.g., cells that make up functional tissue units, successively larger anatomical structures, an entire organ such as the kidney, which is *part_of* the body). The cells that make up (are *located_in*) each of the anatomical structures are organized in a multi-level cell type typology with 'cell' at the root and more and more specialized child nodes. The biomarkers used to *characterize* cell types might have one of five types: genes, proteins, metabolites, proteoforms, and lipids organized in a biomarker typology. Gray arrows indicate crosswalks that connect other HRA DOs to ASCT+B tables. **b.** The HRA 3D reference objects represent the shape, size, location, and rotation of 1,218 3D anatomical structures of 356 types for 65 organs with crosswalks to ASCT+B tables. Shown are 'renal papilla' and 'renal pyramid' in the kidney. **c.** 2D reference illustrations document the shape, size, and spatial layout of 3,726 2D cells of 131 types for 22 FTUs in 10 organs with crosswalks to ASCT+B tables. Shown is the kidney nephron. **d.** Labeled training data exist for FTUs in five organs with crosswalks (gray arrows) to anatomical structures and cell types in the ASCT+B tables. **e.** 13 Organ Mapping Antibody Panels (OMAPs) are linked to 197 Antibody Validation Reports (AVRs) and there exist crosswalks to cell types and biomarkers in ASCT+B tables. **f.** 10 Azimuth references for healthy adult organs plus crosswalks to cell types and biomarkers in ASCT+B tables. **g.** Cell type populations from single cell experimental data exist for 74 3D anatomical structures across 23 organs with 13 unique UBERON IDs in the HRA. Shown is the 'outer cortex of kidney' on left and a bar graph that plots the percentage of cells for three cell types in this anatomical structure on right. **h.** The HRAlit database links HRA DOs to existing ontologies (e.g., Uberon, CL), expert ORCID, publication evidence, funding, and experimental data used for HRApop computation.

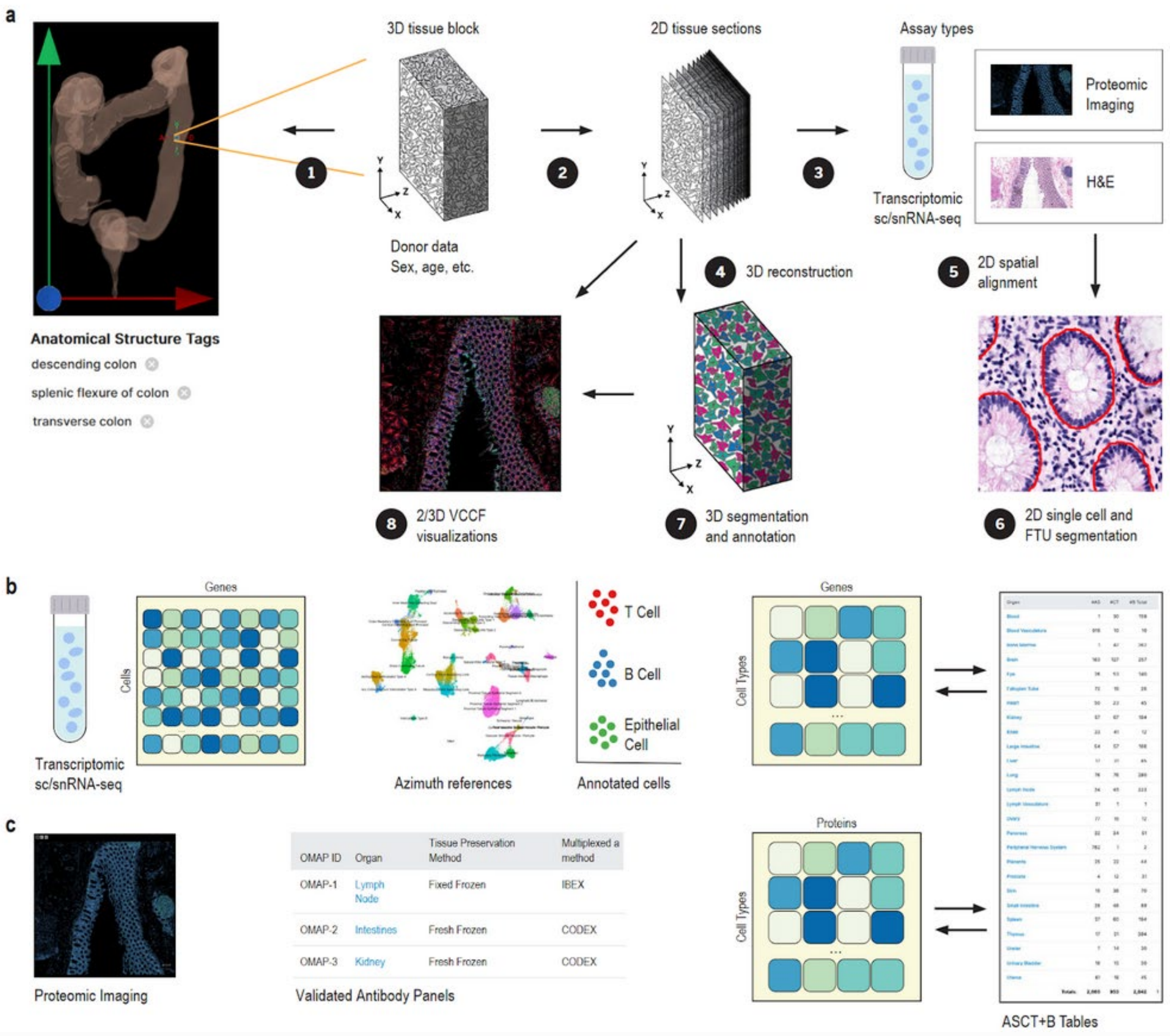


Figure 2: Mapping experimental data to the HRA. a. A 3D tissue block is spatially registered and semantically annotated using the Registration User Interface or the Millitome, see (1). A smaller part of the tissue block might be used for sc/snRNA-seq analysis (not shown) or cut into tissue sections (2). Tissue sections are analyzed using the very same or different assay types (3). Shown are single cell transcriptomics (e.g., sc/snRNA-seq), OMAP-aligned spatial proteomics (e.g., CODEX, Cell DIVE), and high resolution hematoxylin and eosin (H&E) stained histology images. Spatial alignment of different assay types for the very same or different tissue sections is non-trivial (5). H&E data is used to segment functional tissue units (FTUs) using trained machine learning models (6). 3D reconstruction of tissue volumes is accomplished by aligning data from multiple serial tissue sections computationally (4) followed by 3D segmentation and annotation (7). 2D or 3D data is analyzed to identify the distance of different cell types to the vasculature (VCCF Visualizations) as a multiscale common coordinate framework from which no other cell is very distant (8). **b.** Single cell/nucleus data (sc/snRNA-seq) is stored as a cell by gene matrix; cell types are annotated using Azimuth or other cell type annotation tools; results are aggregated to cell type by gene biomarker expression value matrices that are aligned with the ASCT+B tables; and are used in diverse HRA user interfaces (e.g., Exploration User Interface and FTU Explorer). **c.** OMAP-aligned spatial data generated using validated antibody panels linked to AVRs are analyzed to compute cell type by protein biomarker expression value matrices that are aligned with the ASCT+B tables using semi-automated workflows.

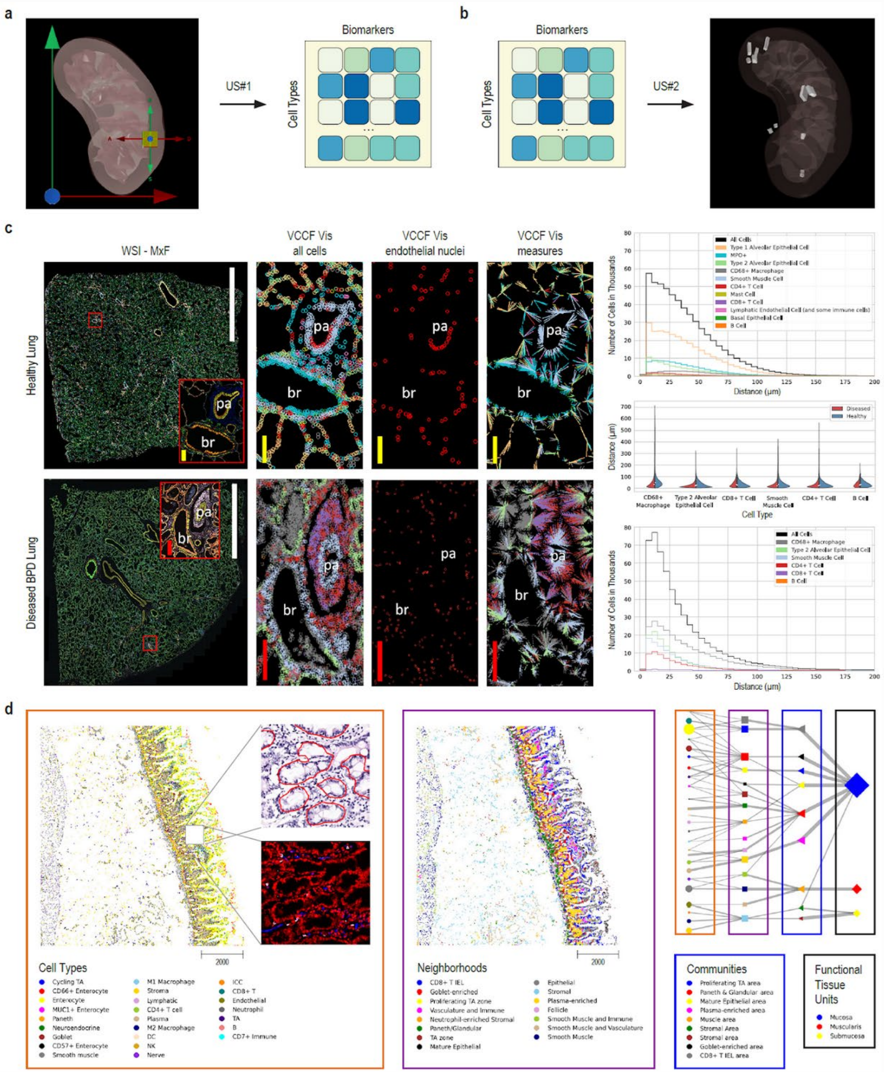


Figure 3: Human Reference Atlas Usage. a. User story #1 (US#1) lets a user define a 3D volume inside of the HRA reference body using the RUI and it predicts cell type populations and mean expression values for cell types in that volume. **b.** User story #2 (US#2) reads cell type population data and predicts the 3D origin of tissue, shown as a collection of extraction sites that have a similar cell type population. **c.** HRA can be used to compare the distribution of parenchymal cells including endothelial, epithelial, and muscle that compose the blood vessels, airways and gas exchanging functional lung structures, and resident immune cells including macrophages, to local vasculature (VCCF Visualizations) in healthy (top) and diseased (bottom) lung using multiplexed immunofluorescence microscopy images with bronchiole (br) and an accompanying small pulmonary artery (pa). Scale bar legend: white: 5 mm, red: 200 µm, yellow: 100 µm. The graphs on the right show distance distributions for cell types present in the healthy lung (top) and diseased BPD lung (bottom); the violin plot (middle) shows a comparison between distance distributions for cell types common in both datasets. **d.** Multi-level cell neighborhoods can be computed to analyze and communicate the structure and function of FTUs; tissue image with cell type annotations and zoom into H&E with FTU segmentations (red outlines) and zoom into the multiplexed image (CODEX) is shown in left, neighborhoods are given in the middle; hierarchy of FTUs, neighborhoods, communities, and cell types are shown on the right.

- HRA-API:
  - Production: https://apps.humanatlas.io/api/
  - OpenAPI Specification: https://apps.humanatlas.io/api/hra-api-spec.yaml
  - JavaScript Library: hra-api
- HRA-API Client Libraries:
  - JavaScript: @hra-api/js-client
  - TypeScript: @hra-api/ts-client
  - Angular 17+: @hra-api/ng-client
  - Python 3.6+: hra-api-client
- HRA Knowledge Graph (HRA-KG):
  - Production: https://lod.humanatlas.io
  - SPARQL Endpoint: https://lod.humanatlas.io/sparql
- grlc APIs (SPARQL queries plus OpenAPI endpoints more info):
  - grlc queries using the HRA-KG SPARQL endpoint: https://grlc.io/api-git/hubmapconsortium/ccf-grlc/subdir/hra/
  - grlc queries using the Ubergraph SPARQL endpoint: https://grlc.io/api-git/hubmapconsortium/ccf-grlc/subdir/ubergraph/
  - More queries available at: https://github.com/hubmapconsortium/ccf-grlc/

See also https://humanatlas.io/api and https://apps.humanatlas.io/api

Is it possible to create a multiscale map of the human body?

The human body has 37 trillion cells. In this incredibly complex and dynamic system, changes at the molecular level alter cell behaviour, tissue architecture and the function of organs; in the other direction, changes in the environment can feed into the body and cause molecular changes. The CIFAR MacMillan Multiscale Human Team seeks to understand the system at all these scales to create an unprecedented map of the human body and help drive medical advances.

The fundamental challenge the researchers will address in the CIFAR MacMillan Multiscale Human program is how to integrate data from various spatial and temporal scales across the body while accounting for diversity in the 8 billion strong global population. To do this, the program will bring together researchers with diverse disciplines and a shared goal — to understand, map, and communicate the multiscale human.

Founded

2023

Supporters

The MacMillan Family Foundation

CIFAR Contact

Amy Cook

IMPACT CLUSTERS

The CIFAR MacMillan Multiscale Human program is part of the following CIFAR Impact Clusters: Decoding Complex Brains and Data; Exploring Emerging Technologies and Shaping the Future of Human Health.

https://events.cifar.ca/website/78533/venue/
https://cifar.ca/research-programs/cifar-macmillan-multiscale-human/

Q&A