

Insightful and Actionable Visual Analytics

Katy Börner @katycns

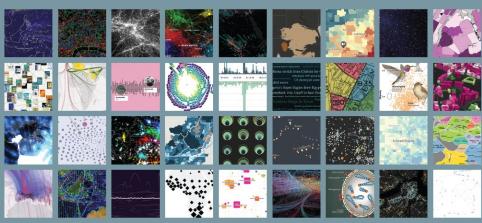
Victor H. Yngve Distinguished Professor of
Intelligent Systems Engineering & Information Science
Director, Cyberinfrastructure for Network Science Center
Luddy School of Informatics, Computing, and Engineering
Indiana University, Bloomington, IN, USA



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

November 11, 2024





Macrosopes for a Global Future



June 6–September 27, 2024

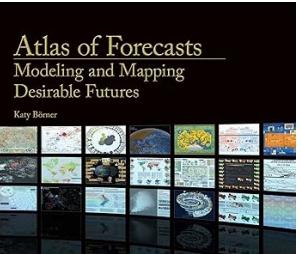
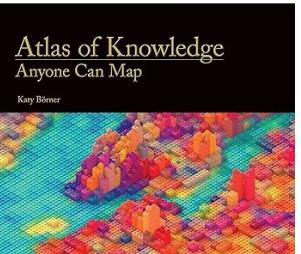
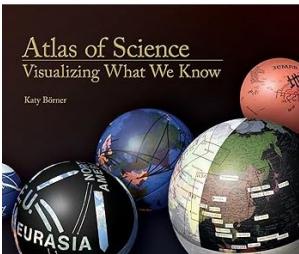
Global Event: Sept 20, 10a.m. – noon
University Collections at McCalla

525 N. Indiana Avenue, Bloomington, IN 47408

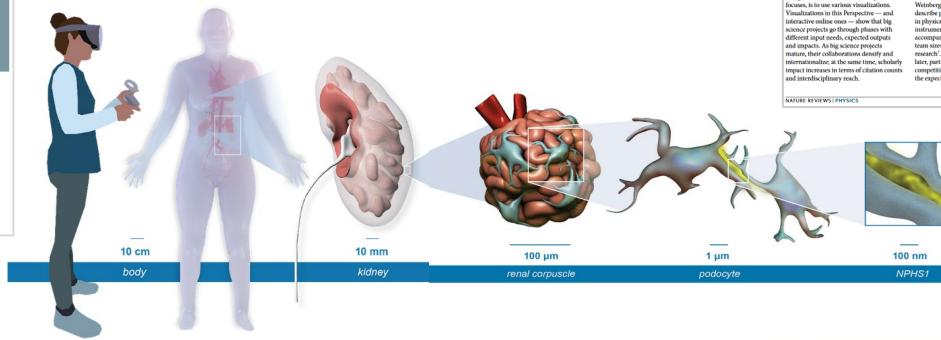


bit.ly/SciMaps20

<https://scimaps.org>



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



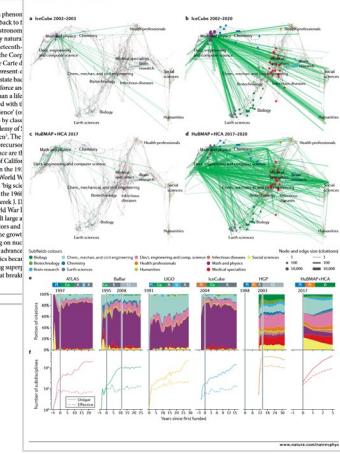
PERSPECTIVES

Visualizing big science projects

Katy Börner, Filipe Nascimento Silva and Stéfano Melotti

Abstract | These numbers, sizes and types of “big science” projects are growing, as are the number, size and value of the hardware and software tools they produce. In this context, big data gives a new way to analyse, understand, manage and communicate the inner workings of collaborations that often involve thousands of experts, thousands of scholarly publications, hundreds of new instruments and analysis tools. This article highlights the social and political impact of big science projects in physics, astronomy and biomedical sciences. A total of 1,893 publications and 1,139 grants by 1,945 authors cited more than 10,000 times between 2002 and 2020. The article also highlights the distinct phases of big science projects, document increasing internationalization and diversification of collaboration networks, and reveal the increase in interdisciplinary research. All data are freely available online.

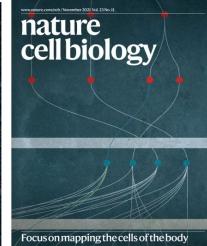
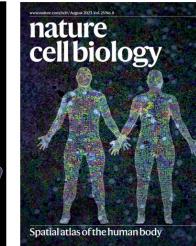
PERSPECTIVES



katy@iu.edu



<https://humanatlas.io>





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



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



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



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

<http://scimaps.org>

Map of Scientific Collaborations from 2005-2009

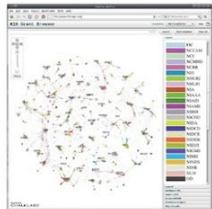


Computed Using Data from Elsevier's Scopus

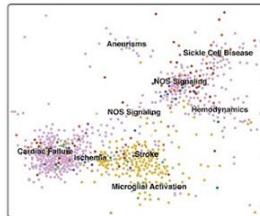
A Topic Map of NIH Grants 2007

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

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



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

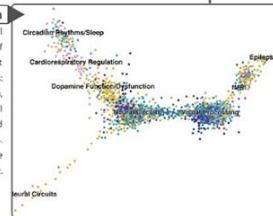


Cardiac Diseases Research

An area of the map focused on cardiovascular function and dysfunction. Cardiac Failure (primarily funded by NHLBI) is topically clustered next to Stroke (NINDS), since these are the two major medical emergencies associated with ischemia, which results from a restricted blood supply. Also localized in this area are grants focused on Nitric Oxide (NO) Signaling, a major biochemical pathway for vasodilation, and grants on Hemodynamics, Sickle Cell Disease, and Aneurysms.

Neural Circuits Research

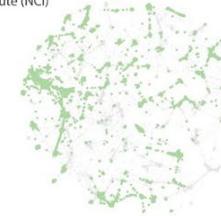
An area of the map focused on neural circuits, which shows the diversity of topics and NIH Institutes that fund research in this area, such as: Cardiorespiratory Regulation, primarily funded by NHLBI; Visual Processing, primarily funded by NEI; and Epilepsy, primarily funded by NINDS. For color coding,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



National Cancer Institute (NCI)

TOP 10 TOPICS

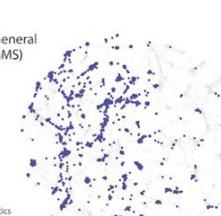
1. Oncology Clinical Trials
2. Cancer Treatment
3. Cancer Therapy
4. Carcinogenesis
5. Risk Factor Analysis
6. Cancer Chemotherapy
7. Metastasis
8. Leukemia
9. Prediction/Prognosis
10. Cancer Chemoprevention



National Institute of General Medical Sciences (NIGMS)

TOP 10 TOPICS

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



National Heart, Lung, and Blood Institute (NHLBI)

TOP 10 TOPICS

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



National Institute of Mental Health (NIMH)

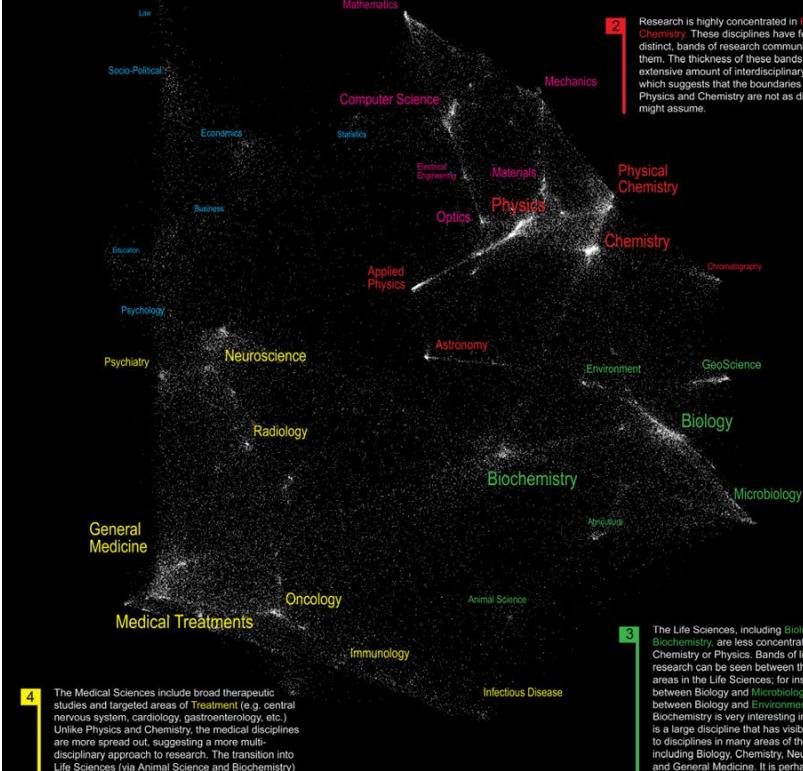
TOP 10 TOPICS

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



The Structure of Science

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



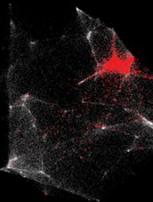
We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

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

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

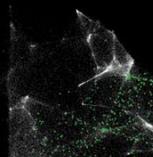
The map of science can be used as a tool for science strategy. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore,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

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



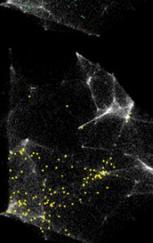
Nanotechnology

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



Proteomics

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



Pharmacogenomics

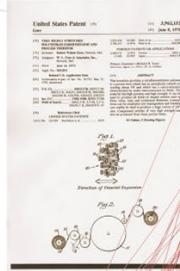
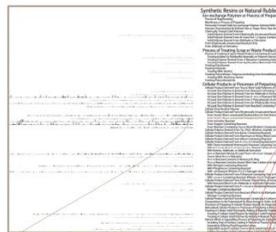
Pharmacogenomics is a relatively new field with most of its activity in **Medicine**. It also has many communities in **Biochemistry** and two communities in the **Social Sciences**.

Impact

The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry or use, proximate function, effect or product, and structure. At the time of this writing there are 160,523 categories in a hierarchy that goes 15 levels deep. We display the first three levels (13,529 categories) at right in what might be considered a textual map of inventions.

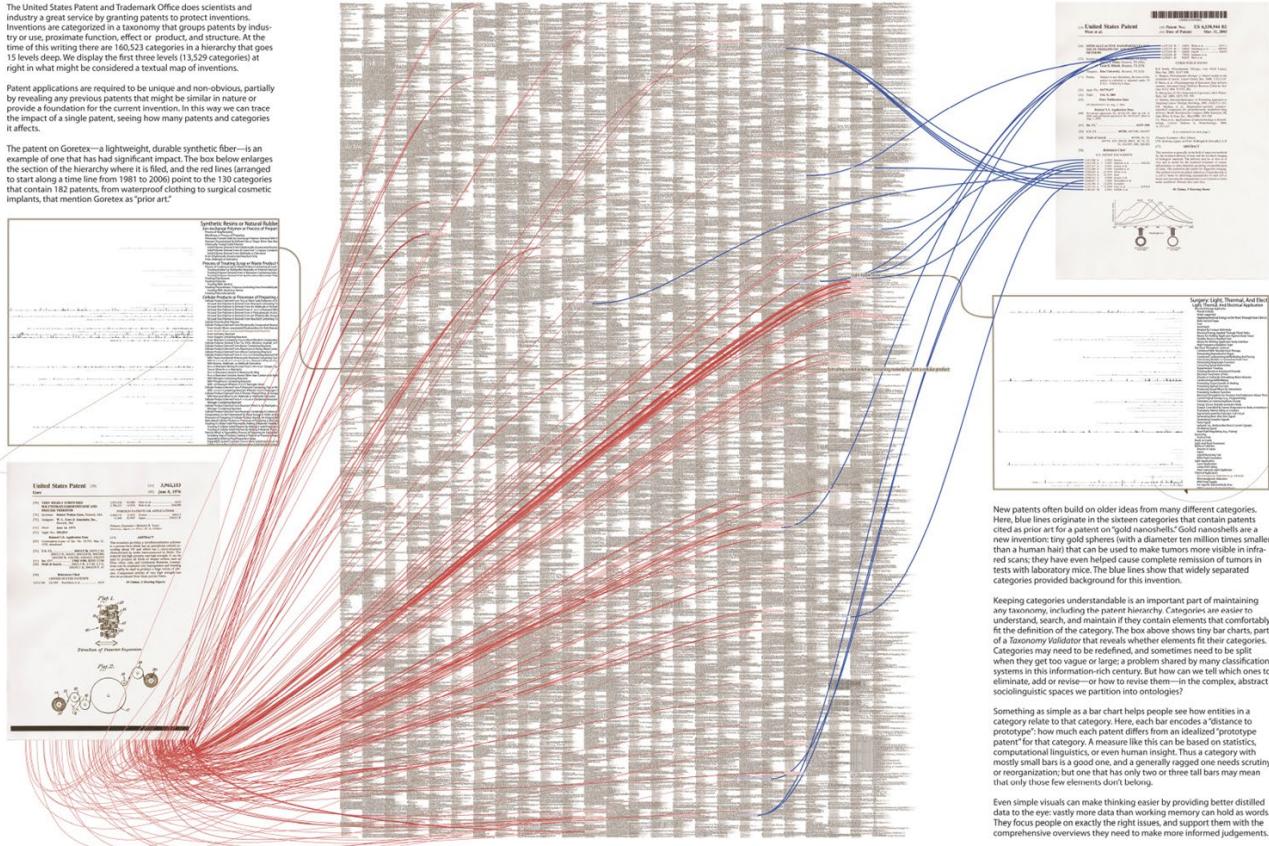
Patent applications are required to be unique and non-obvious, partially revealing prior art that is similar to the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

The patent on Goretex—a lightweight, durable synthetic fiber—is an example of a highly cited prior art patent. Below is a small enlargement of the section of the hierarchy where it is filed, and the red lines mapped to start along a time line from 1981 to 2000 point to the 130 categories that contain 182 patients, from waterproof clothing to surgical cosmetic implants, that mention Goretex as "prior art."



The US Patent Hierarchy

Prior Art

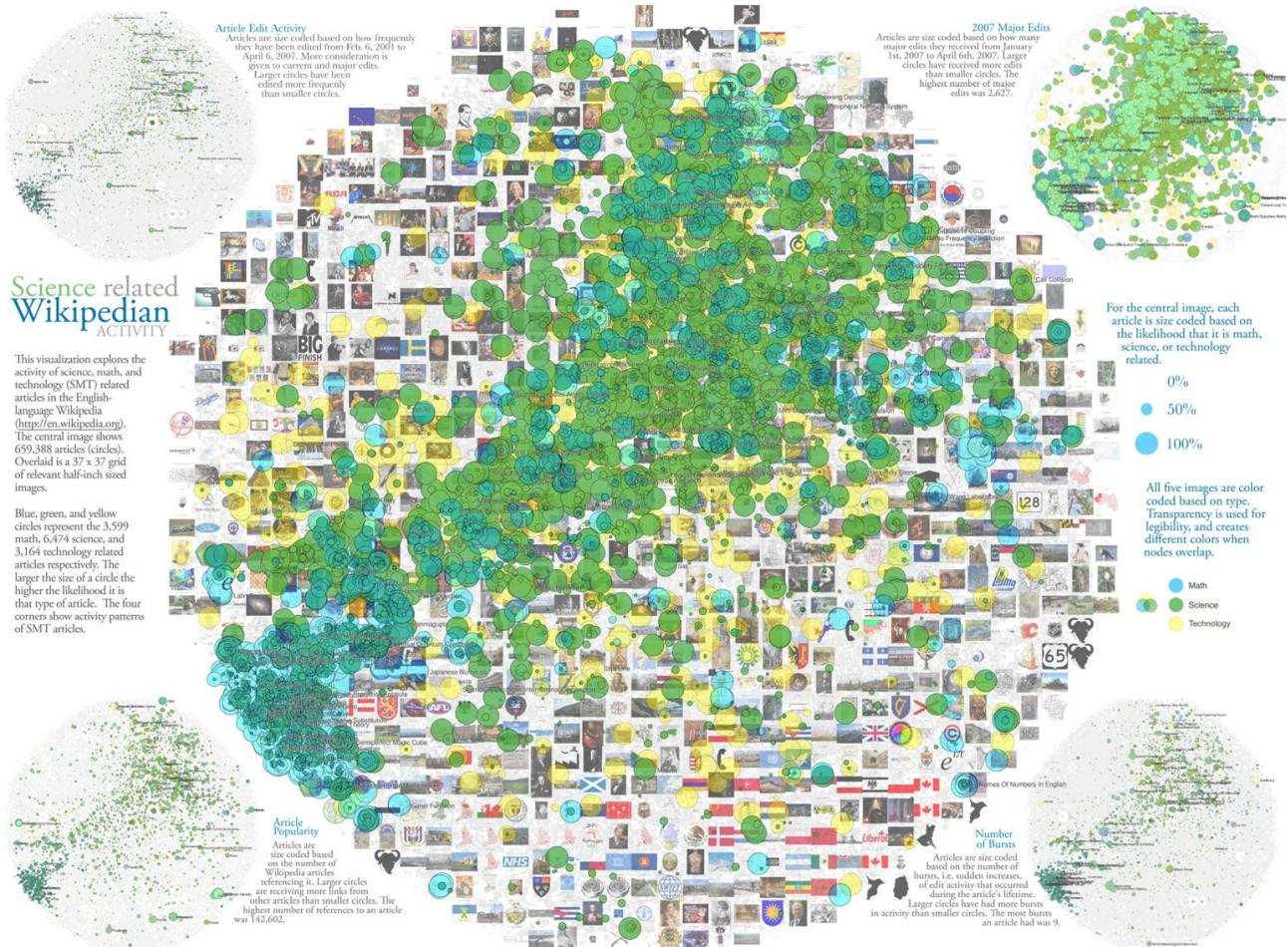


New patents often build on older ideas from many different categories. Here, blue lines originate in the sixteen categories that contain patents cited as prior art for a patent on "gortex". These categories are now used in many products (sphincters with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans; they have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories produce backlinks to this invention.

Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements that comfortably fit the definition of the category. The box above shows tiny bar charts, part of a *Distance Validation*, that reveals whether elements fit their categories. Categories may need to be refined and sometimes need to be split when they get too vague or large: a problem shared by many classification systems in this information-rich century. But how can we tell which ones to eliminate, add or revise—or how to revise them—in the complex, abstract space of categories?

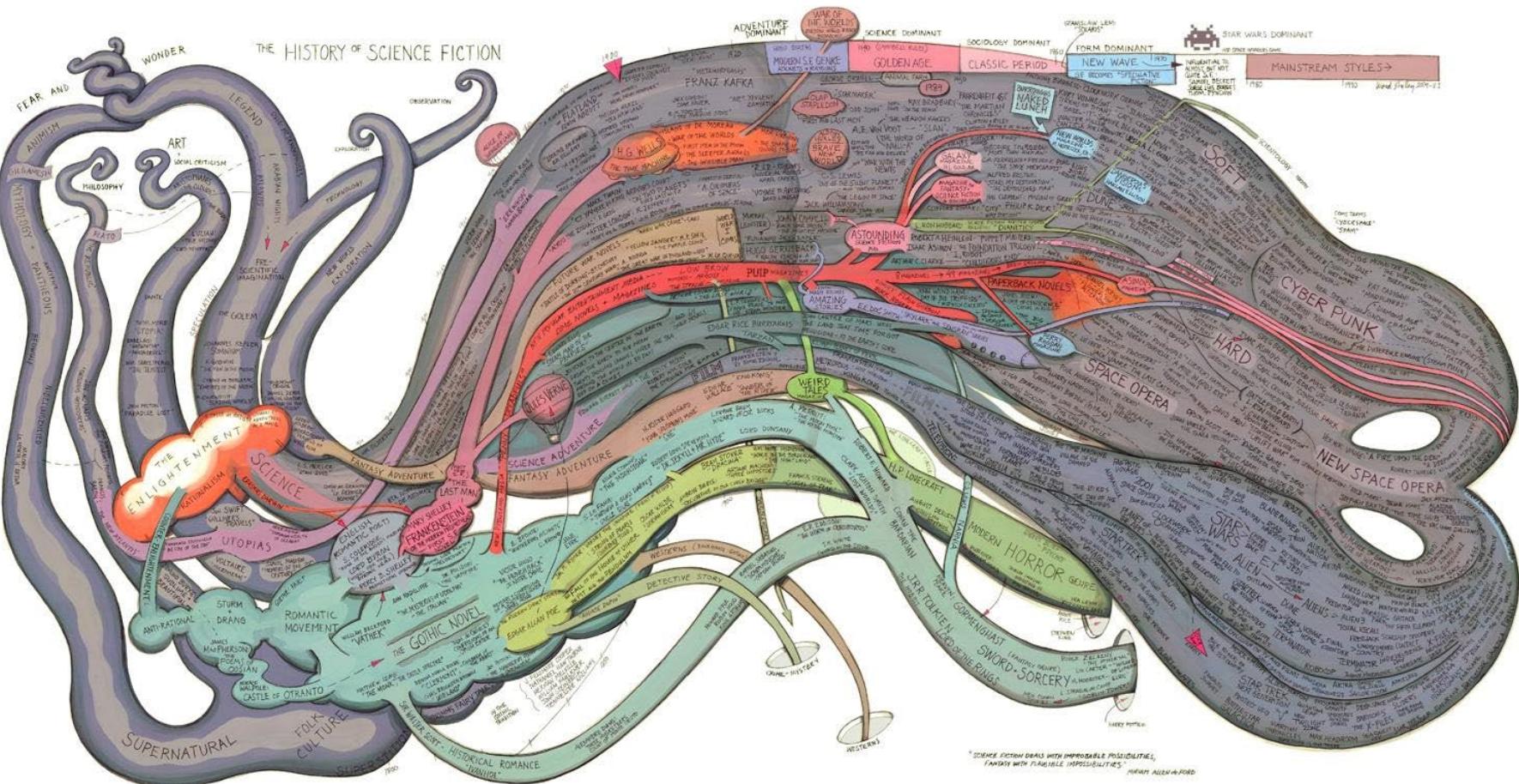
Something as simple as a bar chart helps people see how entities in a category relate to that category. Here, each bar encodes a "distance to prototype": how much each patient differs from an idealized "prototype patent" for that category. A measure like this can be based on statistics, common sense, or even gut instinct. Thus, a category with mostly small bars is a good one, and a general rule is one needs scutiny or reorganization; but one that has only two or three tall bars may mean that only those few elements don't belong.

Even simple visuals can make thinking easier by providing better distilled data to the eye: vastly more data than working memory can hold as words. They focus people on exactly the right issues, and support them with the comprehensive overviews they need to make more informed judgements.



III.8 Science-Related Wikipedian Activity

- Bruce W. Herr II, Todd M. Holloway, Elisha F. Hardy, Katy Börner, and Kevin Boyack - 2007



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



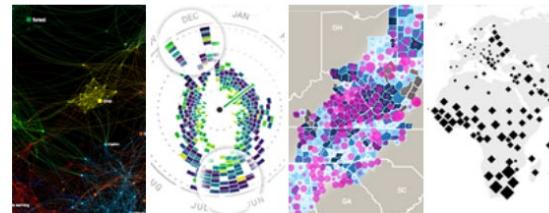
Iteration XIII (2017)

Macroscopes for Playing with Scale



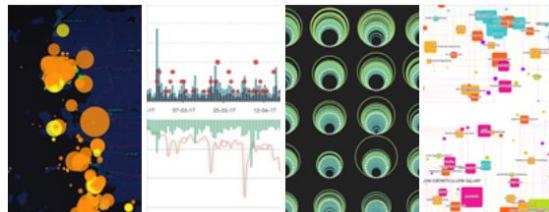
Iteration XIV (2018)

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





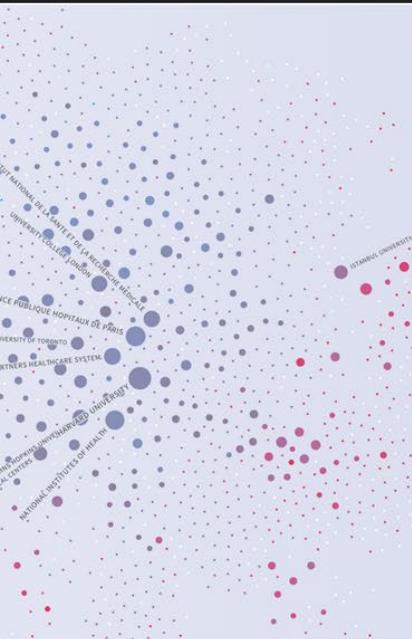
Smelly Maps

Charting urban smellscapes



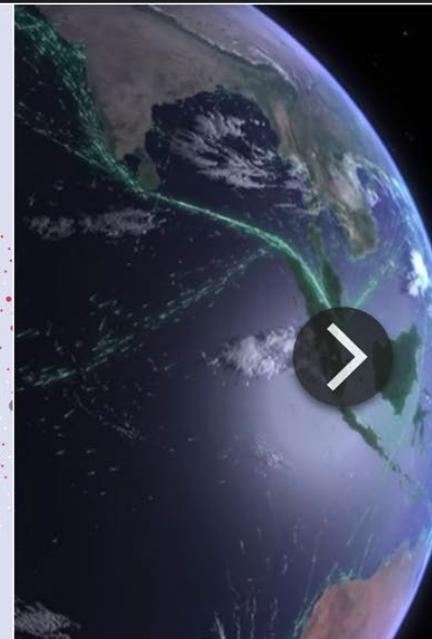
HathiTrust

Storehouse of knowledge



Excellence Networks

Publish or perish together

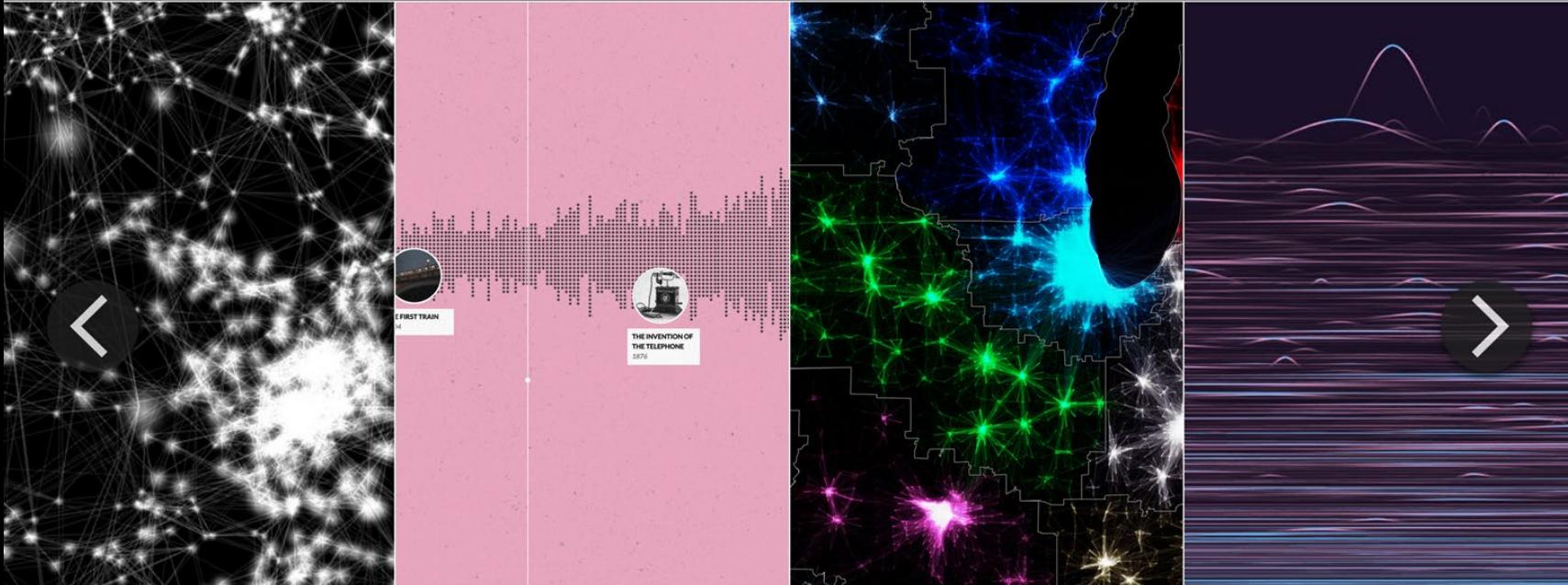


FleetMon Explorer

Tracking the seven seas

SMELLY
MAPS





The Cosmic Web
And the network behind it

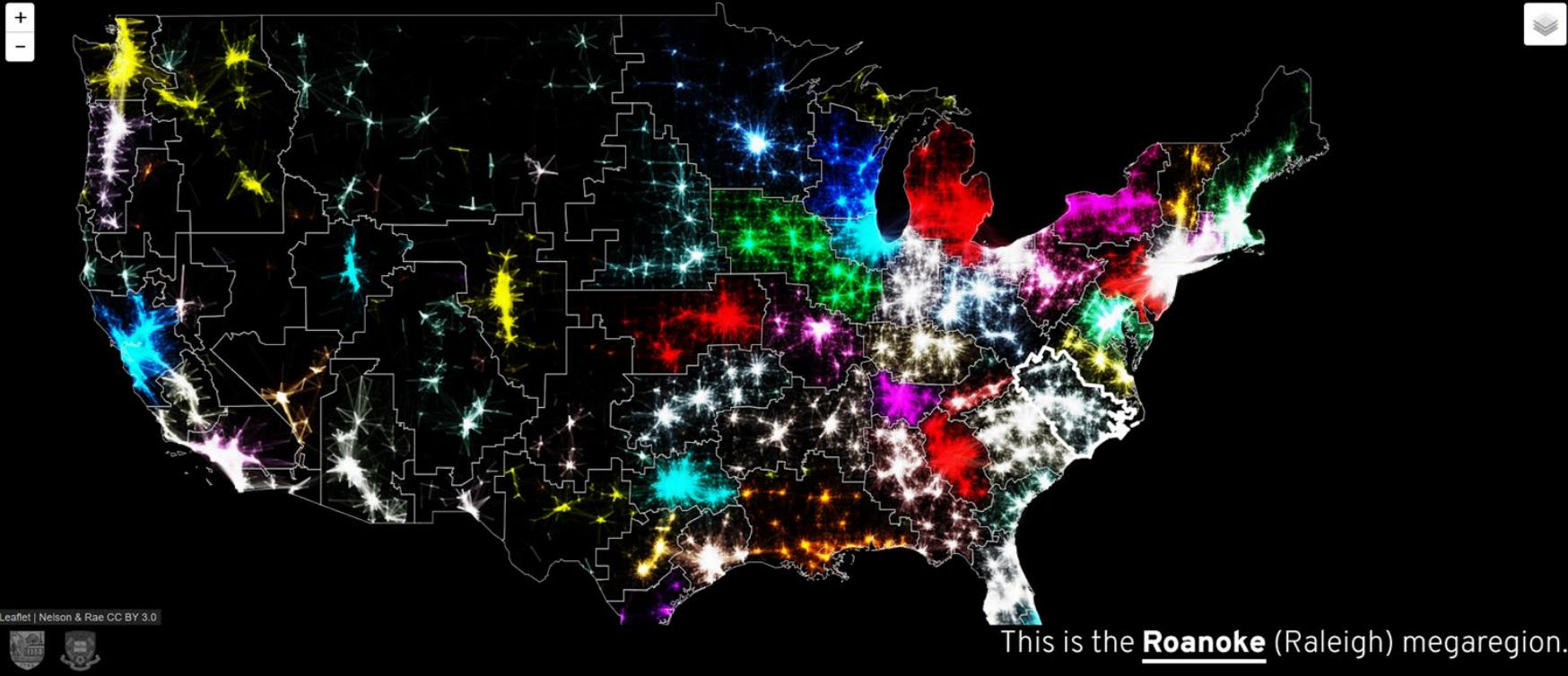
Histograms
An interactive timeline

Megaregions of the US
Mapping commuter patterns

Science Paths
The random impact rule

THE MEGAREGIONS OF THE US

Explore the new geography of commuter connections in the US.
Tap to identify regions. Tap and hold to see a single location's commuteshed.



Leaflet | Nelson & Rae CC BY 3.0





How Do We Compare?

Using metrics for global good



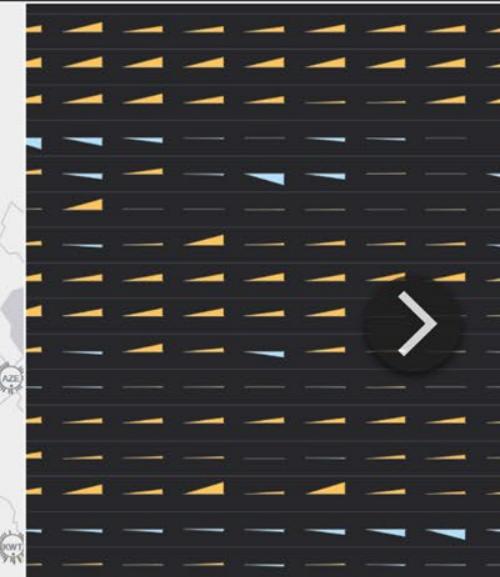
River Runner

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



2nd Decade (2015-2024)

Macroscopes



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



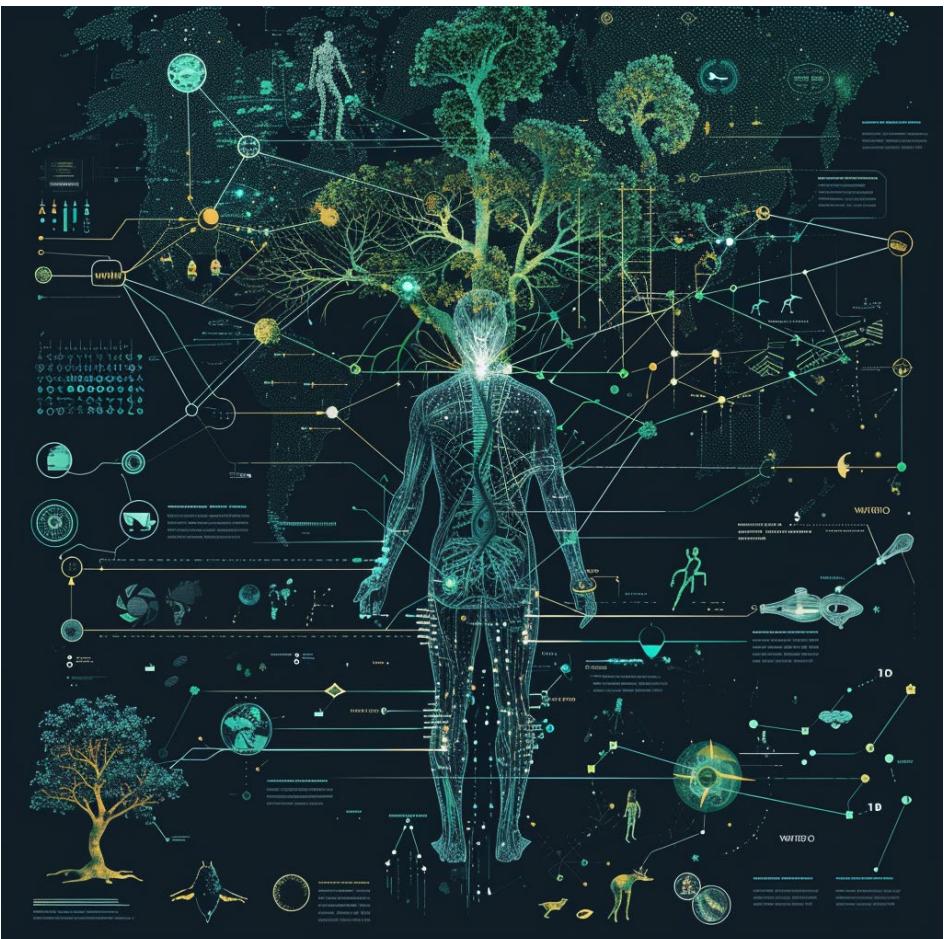
PRESS ITEMS

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

<http://scimaps.org>

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

<https://scimaps.org>



wermind using <https://www.midjourney.com>

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 Leibniz Institute for Regional Geography, Leipzig University, Germany

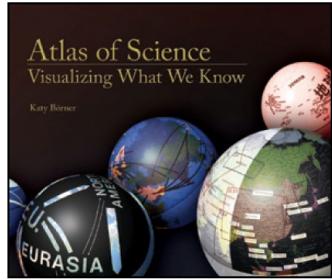


Benjamin Wiederkehr
Founding Partner and Managing Director of [Interactive Things](#) in Zürich, Switzerland

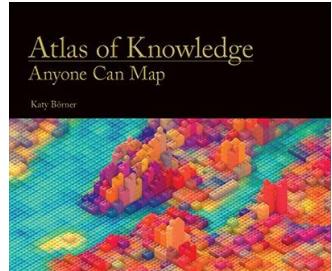


Lev Manovich
Professor, [The Graduate Center](#), City University of New York; Director, [Software Studies Initiative](#) (big data, digital humanities, visualization)

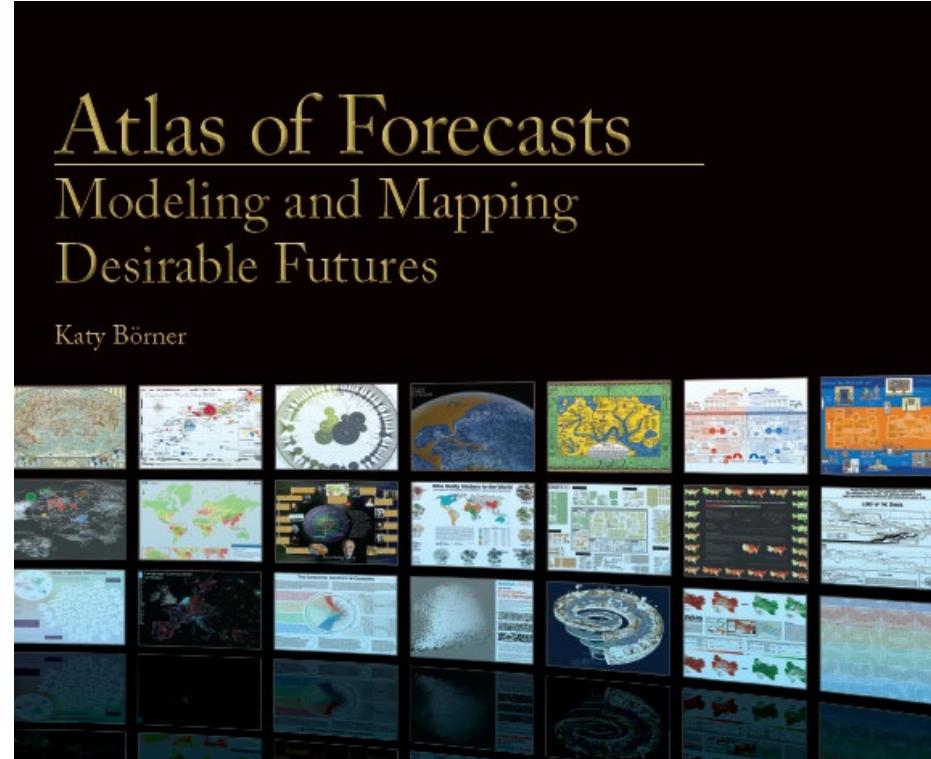
Atlas Trilogy



2010



2015



2021

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



About 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.



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](#)

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

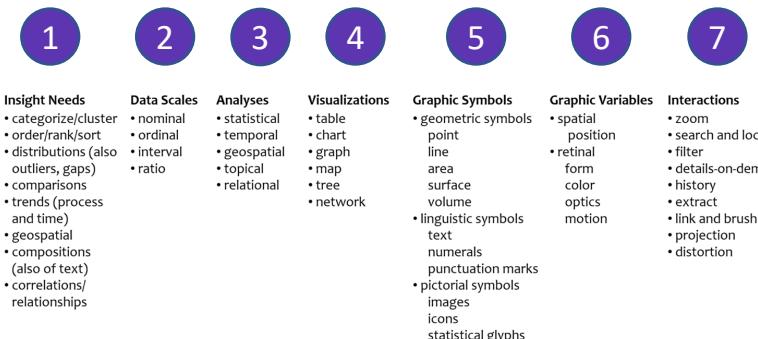
- The initial DVL-FW was developed via an extensive literature review.
- The resulting DVL-FW typology, process model, exercises, and assessments were then tested in the *Information Visualization* course taught for more than 17 years at Indiana University. More than 8,500 students enrolled in the IVMOOC 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:

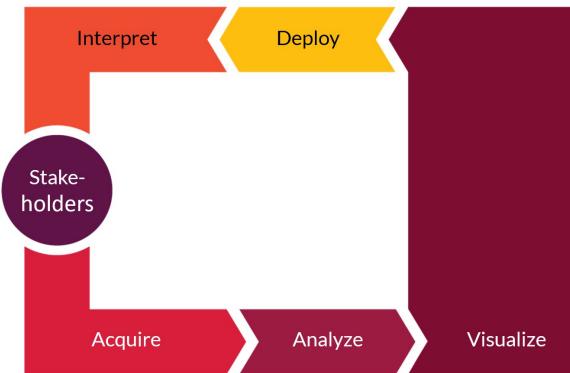
DVL Typology

Defines 7 types with 4-17 members each.



DVL Workflow Process

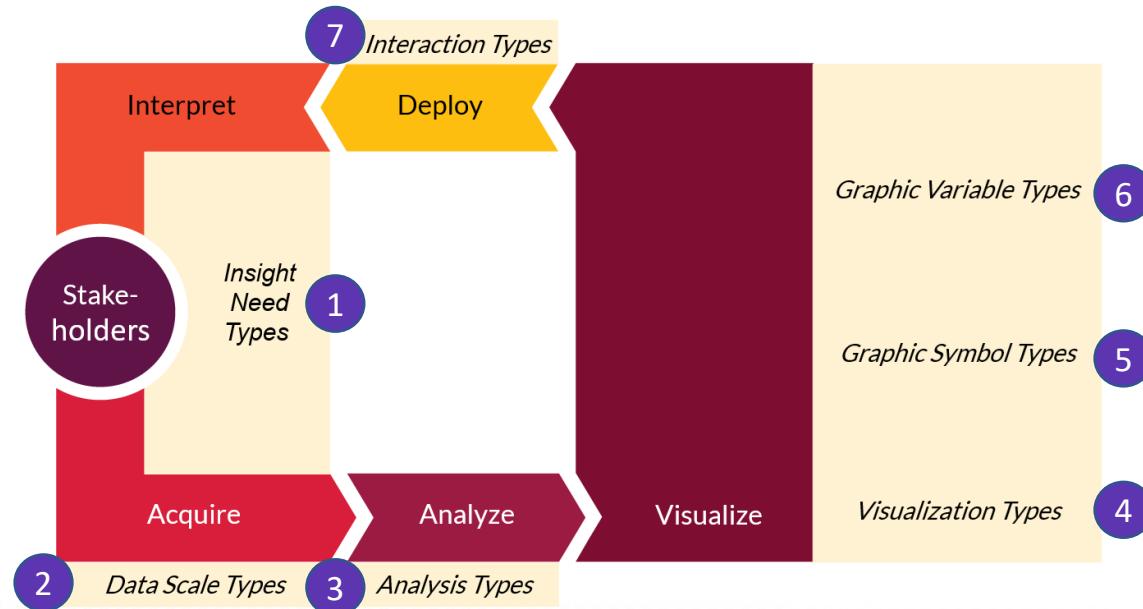
Defines 5 steps required to render data into insights.

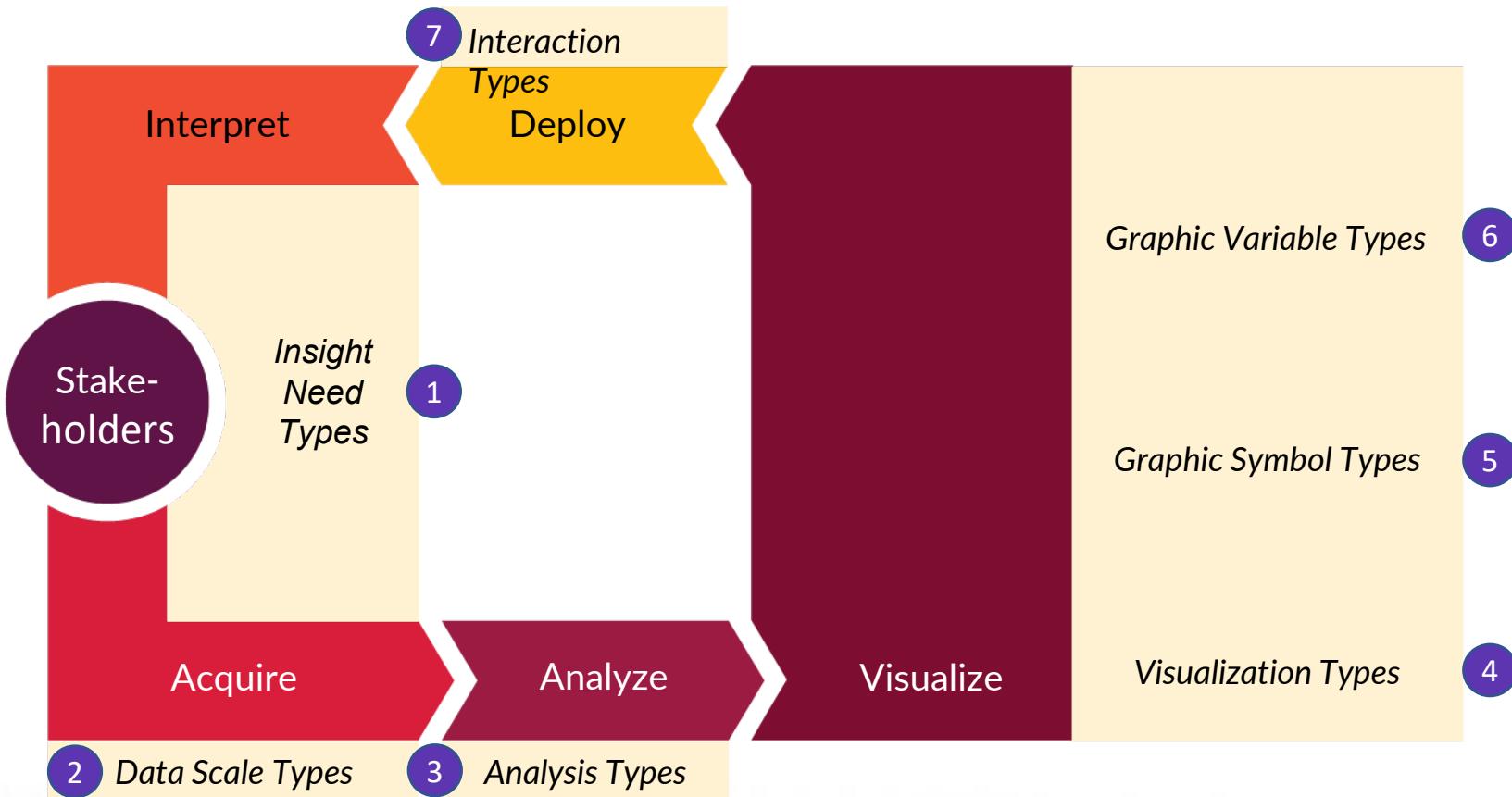


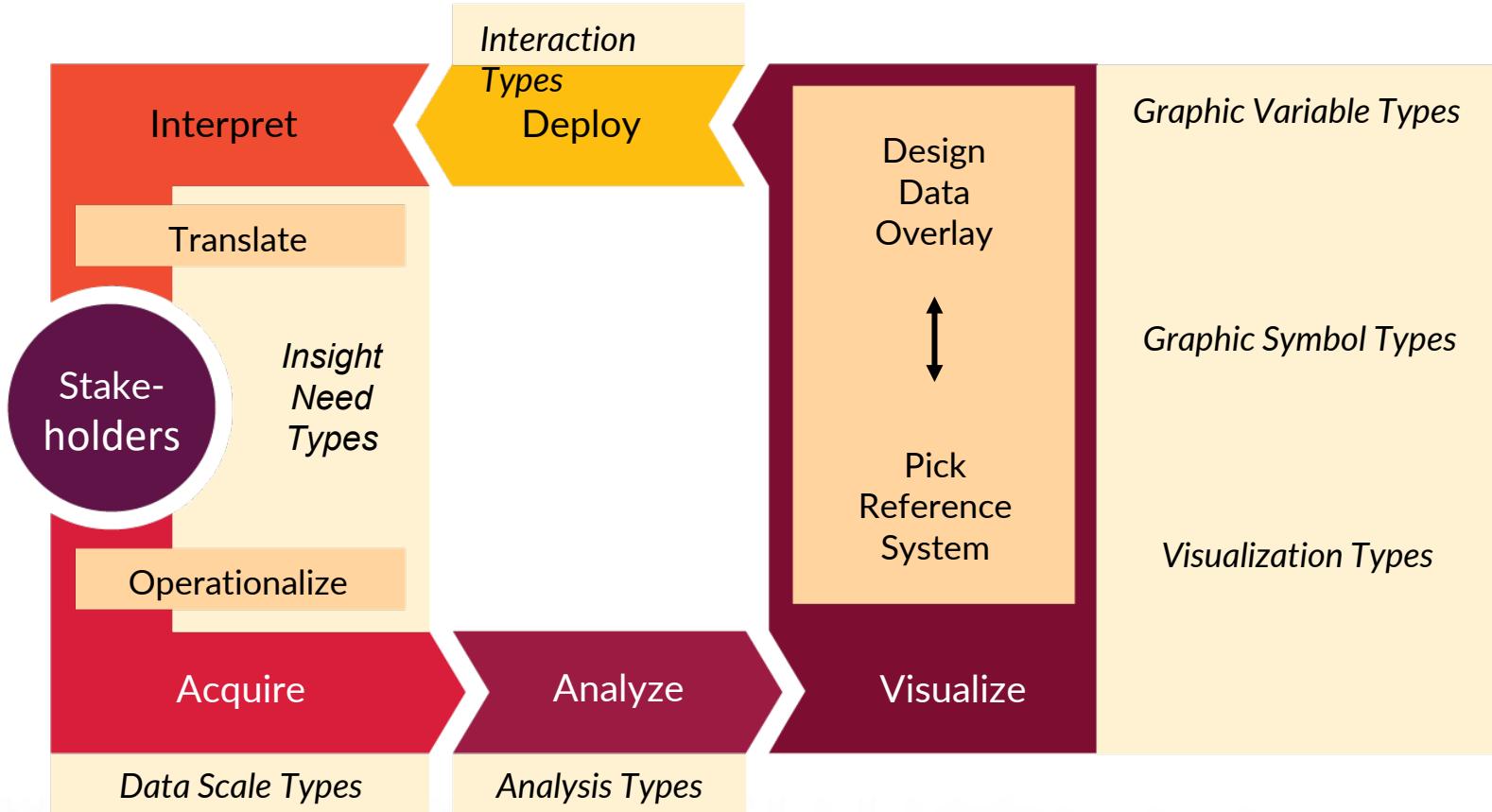
Data Visualization Literacy Framework (DVL-FW)

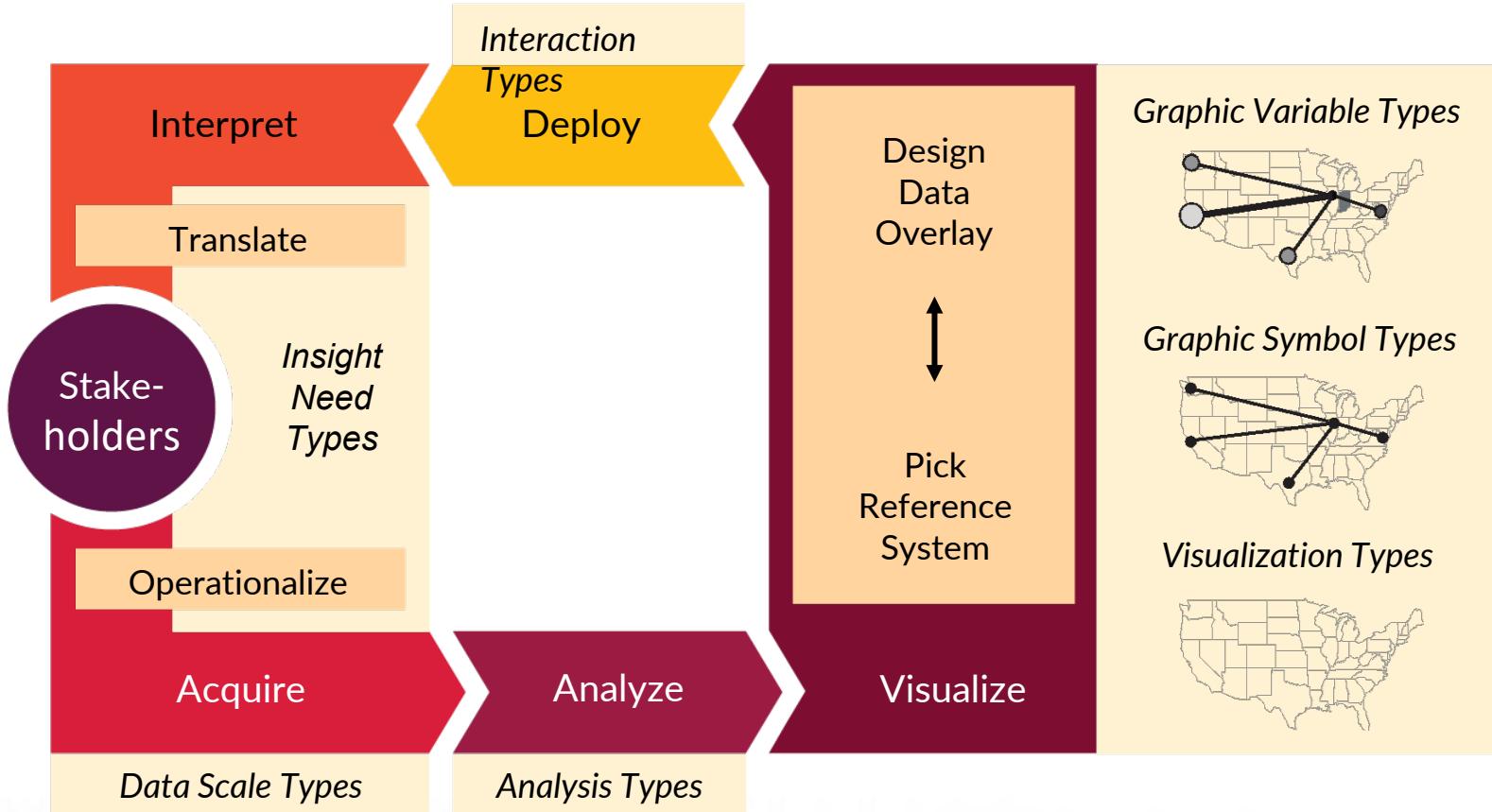
Consists of two parts **that are interlinked**:

**DVL Typology +
DVL Workflow
Process**









Data Visualization Literacy Framework (DVL-FW)

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

The screenshot shows the Make-A-Vis (MAV) application interface. On the left, the 'Data' tab is selected, displaying two sections: 'ISI Publications' and 'Journals'. The 'ISI Publications' section shows a table with columns: Title, Authors, Journal, Year, #Cites, and a total record count of 562. The 'Journals' section shows a table with columns: Name, #Papers, #Cites, First Year, Last Year, and a total record count of 562. In the center, a 'Make Visualization' panel is open, showing four visualization types: Scatter Graph, Temporal Bar Graph (selected), Geomap, and Scimap. A purple circle with the number 4 is centered over the 'Temporal Bar Graph' button. Below this panel are three dropdown menus: 'Select Graphic Symbol Type(s)', 'Select Graphic Variable Types', and 'Done'. A purple circle with the number 5 is centered over the 'Select Graphic Symbol Type(s)' dropdown, and a purple circle with the number 6 is centered over the 'Select Graphic Variable Types' dropdown. On the right, a 'Temporal Bar Graph' visualization is displayed, showing the growth of various fields over time from 1998 to 2017. The graph has a pink header bar. A purple circle with the number 7 is centered over the graph area.

Typology of the Data Visualization Literacy Framework

1	2	3	4	5	6	7
Insight Needs <ul style="list-style-type: none"> • categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships 	Data Scales <ul style="list-style-type: none"> • nominal • ordinal • interval • ratio 	Analyses <ul style="list-style-type: none"> • statistical • temporal • geospatial • topical • relational 	Visualizations <ul style="list-style-type: none"> • table • chart • graph • map • tree • network 	Graphic Symbols <ul style="list-style-type: none"> • geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs 	Graphic Variables <ul style="list-style-type: none"> • spatial position • retinal form color optics motion 	Interactions <ul style="list-style-type: none"> • 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
<ul style="list-style-type: none"> • categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships 	<ul style="list-style-type: none"> • nominal • ordinal • interval • ratio 	<ul style="list-style-type: none"> • statistical • temporal • geospatial • topical • relational 	<ul style="list-style-type: none"> • table • chart • graph • map • tree • network 	<ul style="list-style-type: none"> • geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs 	<ul style="list-style-type: none"> • spatial position • retinal form color optics motion 	<ul style="list-style-type: none"> • 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

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

Data Scales

- nominal
- ordinal
- interval
- ratio

Analyses

- statistical
- temporal
- geospatial
- topical
- relational

Visualizations

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

Graphic Symbols

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

Graphic Variables

- spatial position
- retinal form color optics motion

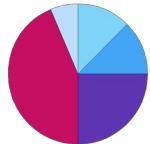
Interactions

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

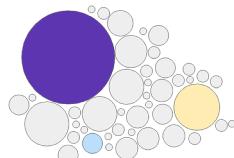
Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 30-31.

Visualization Types

Chart

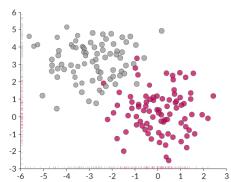


Pie Chart

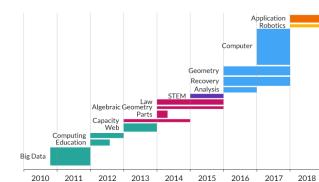


Bubble Chart

Graph

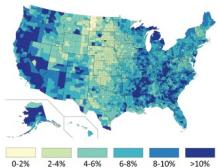


Scatter Graph

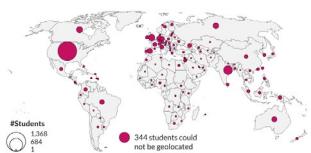


Temporal Bar Graph

Map

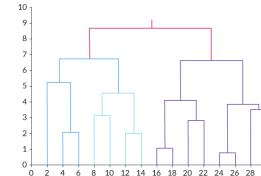


Choropleth Map

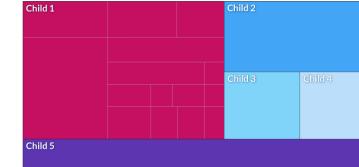


Proportional Symbol Map

Tree

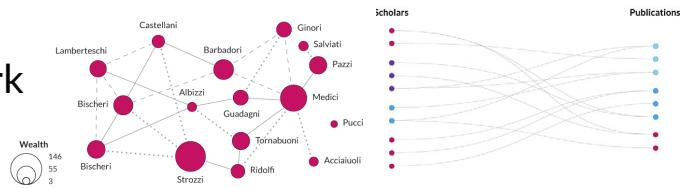


Dendrogram



Tree Map

Network



Force-Directed Network Layout

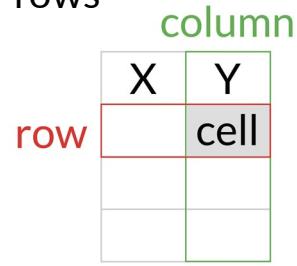


Bimodal Network Layout

Visualize: Reference Systems

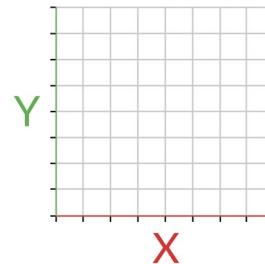
Table

columns by
rows



Graph

x-y
coordinates



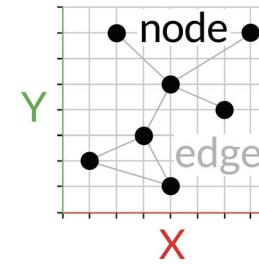
Map

latitude/
longitude



Network

local
similarity

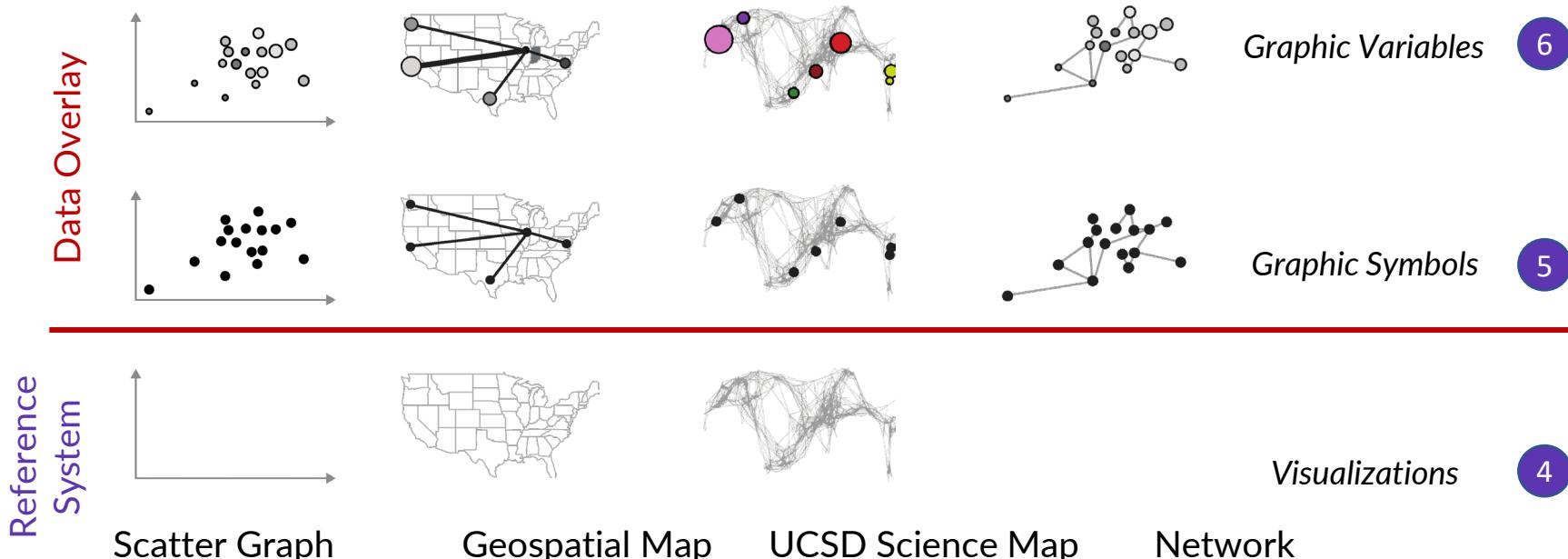


4

Visualization Types

- table
- chart
- graph
- map
- network layout

Visualize: Reference Systems, Graphic Symbols and Variables



Typology of the Data Visualization Literacy Framework

Insight Needs	Data Scales	Analyses	Visualizations	Graphic Symbols	Graphic Variables	Interactions
<ul style="list-style-type: none"> categorize/cluster order/rank/sort distributions (also outliers, gaps) comparisons trends (process and time) geospatial compositions (also of text) correlations/relationships 	<ul style="list-style-type: none"> nominal ordinal interval ratio 	<ul style="list-style-type: none"> statistical temporal geospatial topical relational 	<ul style="list-style-type: none"> table chart graph map tree network 	<ul style="list-style-type: none"> geometric symbols point line area surface volume linguistic symbols text numerals punctuation marks pictorial symbols images icons statistical glyphs 	<ul style="list-style-type: none"> spatial position retinal form color optics motion 	<ul style="list-style-type: none"> 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

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

6

Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 34-35.

Graphic Variable Types

Position: x, y; possibly z

Form:

- Size
- Shape
- Rotation (Orientation)

Color:

- Value (Lightness)



- Hue (Tint)



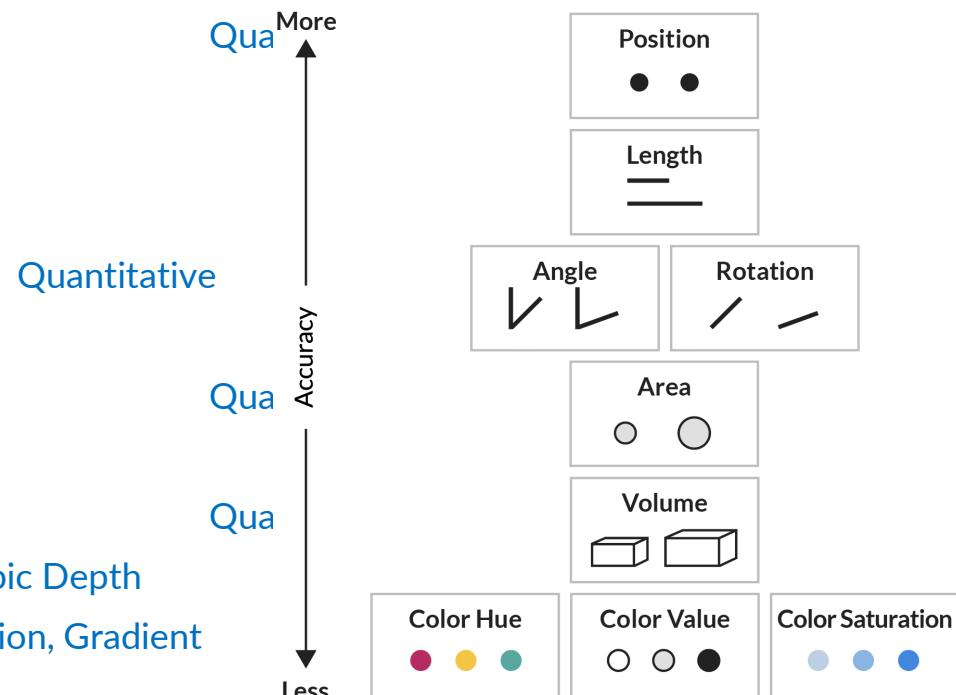
- Saturation (Intensity)



Optics: Blur, Transparency, Shading, Stereoscopic Depth

Texture: Spacing, Granularity, Pattern, Orientation, Gradient

Motion: Speed, Velocity, Rhythm



Graphic Symbol Types

Graphic Variable Types

		Geometric Symbols		Linguistic Symbols	Pictorial Symbols
		Point	Line		
Spatial Position	X Y				
Form	Size	• • •		Text Text Text	☺ ☺ ☺
	Shape	● ▲ ■	⋮ ⋮ ⋮	Text Text Text	☺ ☺ ☺
Color	Value	••••••		Text Text Text	
	Hue	••••••		Text Text Text	
Retinal	Saturation	••••••		Text Text Text	
Texture	Granularity				
	Pattern				
Motion Optics	Blur	••••••		Text Text Text	☺ ☺ ☺
	Speed	→→→→→	→→→→→	⑦→ ⑦→ ⑦→	☺→☺→☺→

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



Qualitative

Also called:
Categorical Attributes
Identity Channels

Quantitative

Also called:
Ordered Attributes
Magnitude Channels

Graphic Variable Types Versus Graphic Symbol Types

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

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

[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

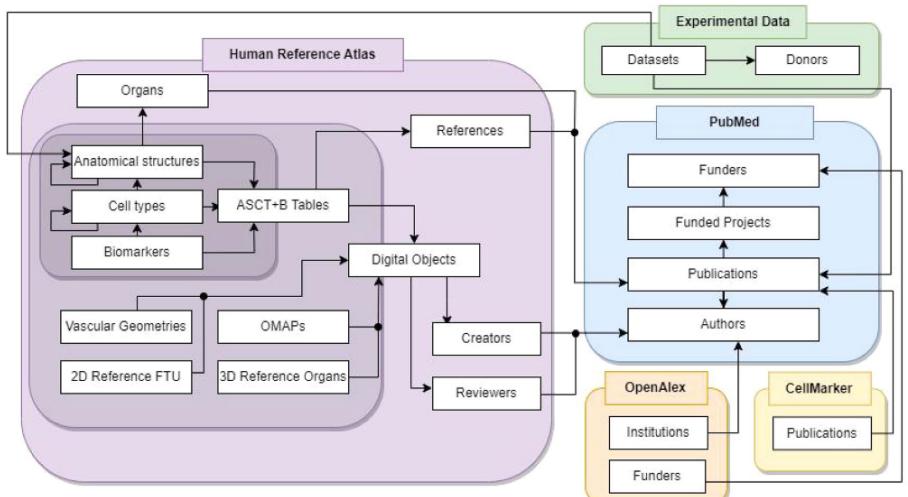


Fig. 1 Overview of the HRALit database. HRA data types are linked to experimental data and to PubMed data using HuBMAP IDs of the HRA digital objects; experimental dataset IDs, donor IDs, and publication DOIs; PubMed publication PMIDs, author IDs, funding IDs, institution IDs, and funder IDs.

Kong, Y., Börner, K. 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>.

scientific data

OPEN
DATA DESCRIPTOR

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

Yongxin Kong^{1,2} & Katy Börner^{1,3}

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). 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 organs¹. 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 DOI 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)², Cellular Senescence Network (SenNet)³, Kidney Precision Medicine Project (KPM)^{4,5}, GenitoUrinary Developmental Molecular Anatomy Project (GUDMAP)⁶, the Genotype-Tissue Expression (GTEx)⁷, or CZ CELxGENE⁸. 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

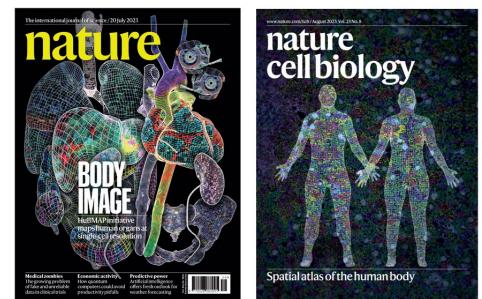
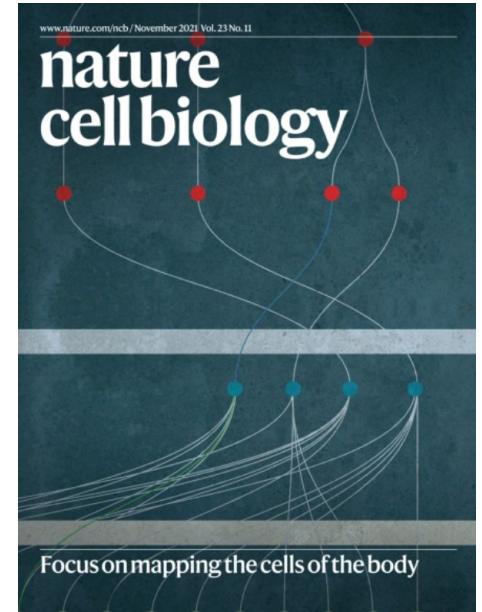
¹Department of Intelligent Systems Engineering, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, 47408, USA. ²School of Information Management, Sun Yat-sen University, Guangzhou, 510006, China. e-mail: yokong@iup.edu; katy@iup.edu

Human Reference Atlas (HRA)

The Human Reference Atlas (HRA)

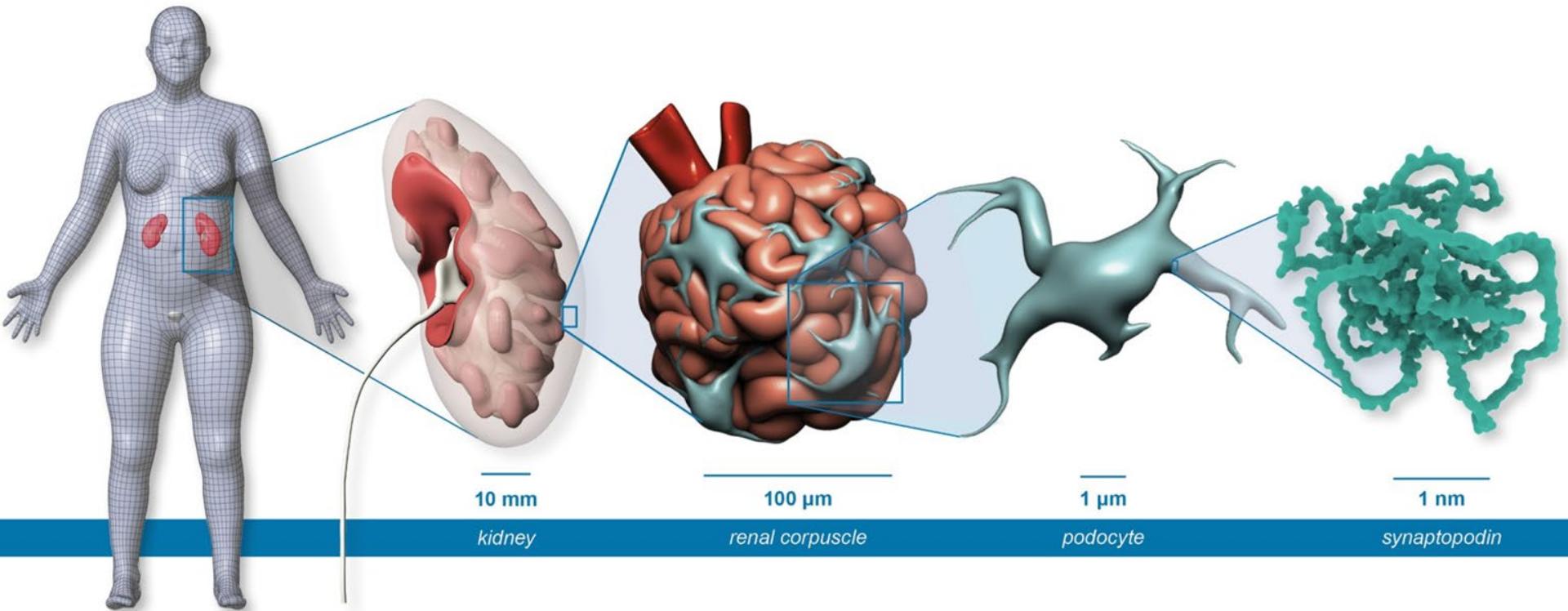
1. defines the 3D space and shape of anatomical structures and cell types that are of biomedical relevance plus the biomarkers used to characterize them. Anatomical structures, cell types and biomarkers are validated and represented in/added to ontologies (Uberon/FMA, CL, HGNC).
2. defines how new datasets 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.

The screenshot shows the bioRxiv preprint server interface for the "Human BioMolecular Atlas Program (HuBMAP): 3D Human Reference Atlas Construction and Usage" paper. At the top, the CSHL logo and the bioRxiv logo are visible, along with navigation links for HOME, SUBMIT, FAQ, BLOG, ALERTS / RSS, RESOURCES, ABOUT, and CHANNELS. A search bar is present, with an "Advanced Search" link below it. The main content area displays the title, authors (Katy Börner, Philip D. Blood, Jonathan C. Silverstein, Matthew Ruffalo, Sarah A. Teichmann, Gloria Pryhuber, Ravi Misra, Jeffrey Purkerson, Jean Fan, John W. Hickey, Gasmira Molla, Chuan Xu, Yun Zhang, Griffin Weber, Yashvardhan Jain, Daniel Quaaroni, Yongxin Kong, HRA Team, Andreas Bueckle, Bruce W. Herr II), and the DOI (<https://doi.org/10.1101/2024.03.27.587041>). Below the abstract, there are social sharing icons (Facebook, Twitter, LinkedIn, etc.) and a "Post" button. The "Abstract" section is expanded, showing the full text: "The Human BioMolecular Atlas Program (HuBMAP) aims to construct a reference 3D structural, cellular, and molecular atlas of the healthy adult human body. The HuBMAP Data Portal (<https://portal.hubmapconsortium.org>) serves experimental datasets and supports data processing, search, filtering, and visualization. The Human Reference Atlas (HRA) Portal (<https://humanatlas.io>) provides open access to atlas data, code, procedures, and instructional materials. Experts from more than 20 consortia are collaborating to construct the HRA's Common Coordinate Framework (CCF), knowledge graphs, and tools that describe the multiscale structure of the human body (from organs and tissues down to cells, genes, and biomarkers) and to use the HRA to understand changes that occur at each of these levels with aging, disease, and other". To the right of the abstract, there are links for "Download PDF", "Print/Save Options", "Supplementary Material", "Data/Code", "Revision Summary", and "Email", "Share", "Citation Tools", and "Get QR code". A red banner at the bottom right promotes COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv. On the far right, there are sections for "Subject Area" (Bioinformatics), "Subject Areas" (Animal Behavior and Cognition, Biochemistry, Bioengineering, Bioinformatics, Biophysics), and "All Articles".

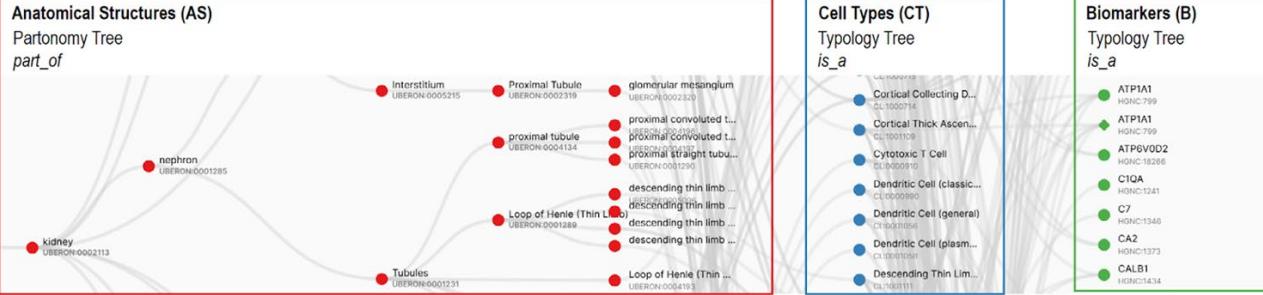
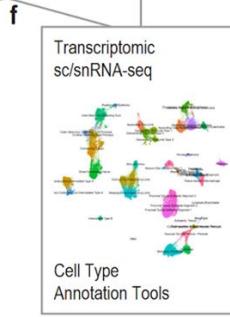
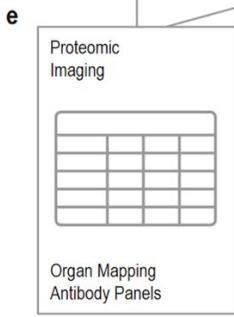
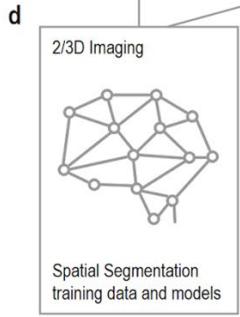
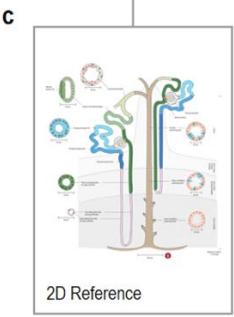
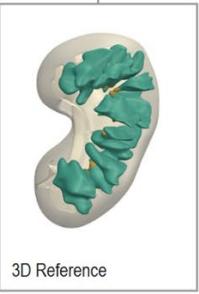
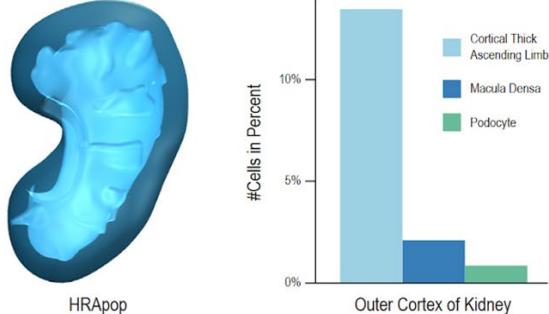
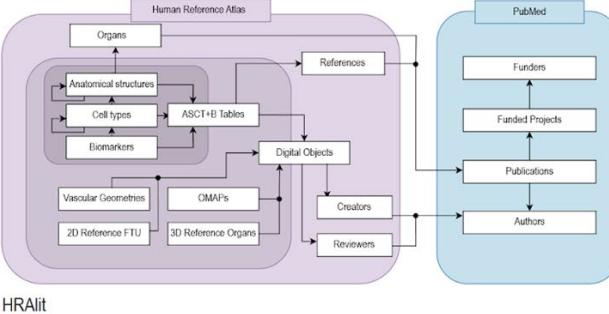
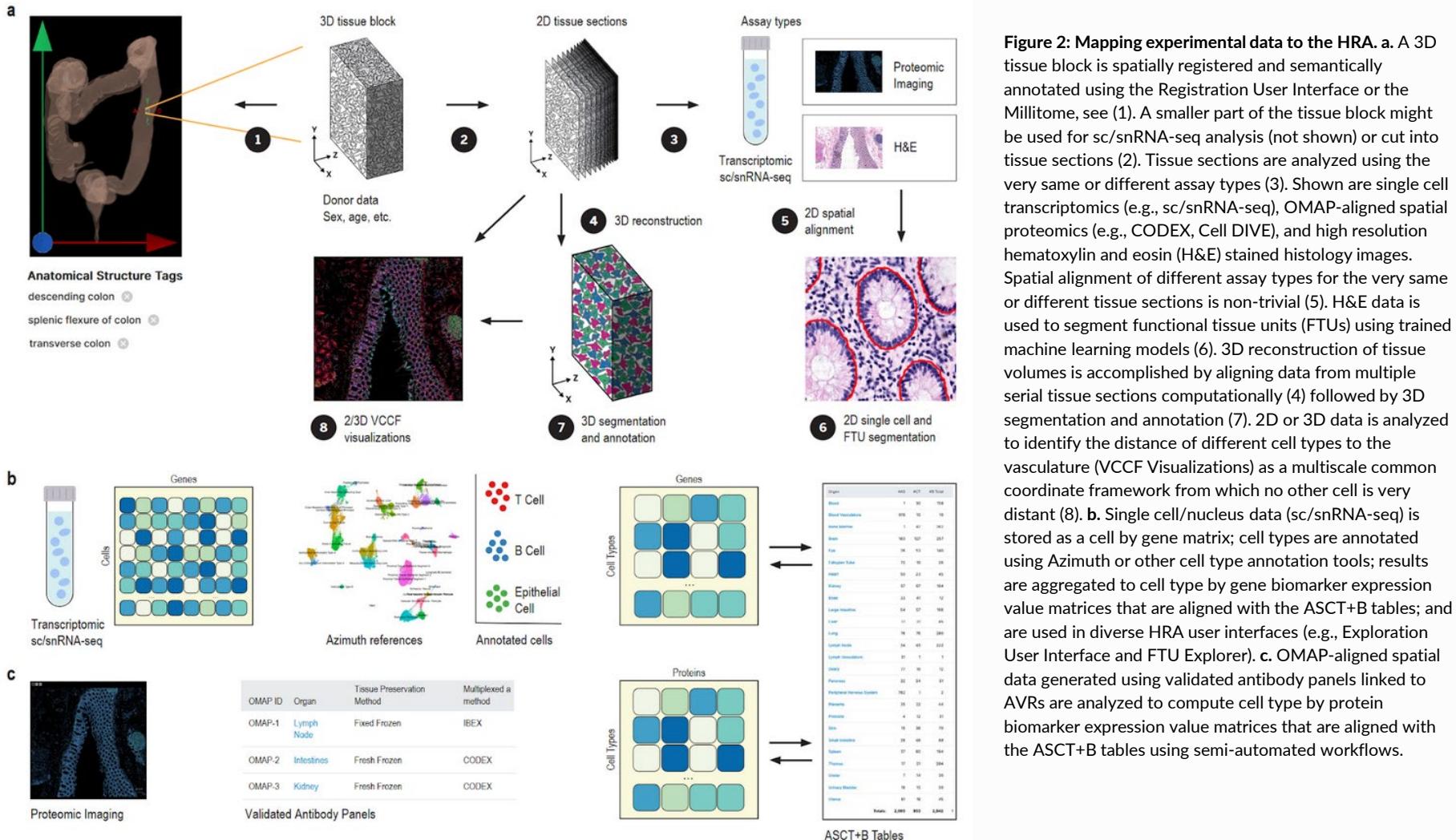
a**b****g****h**

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, proteiforms, 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).



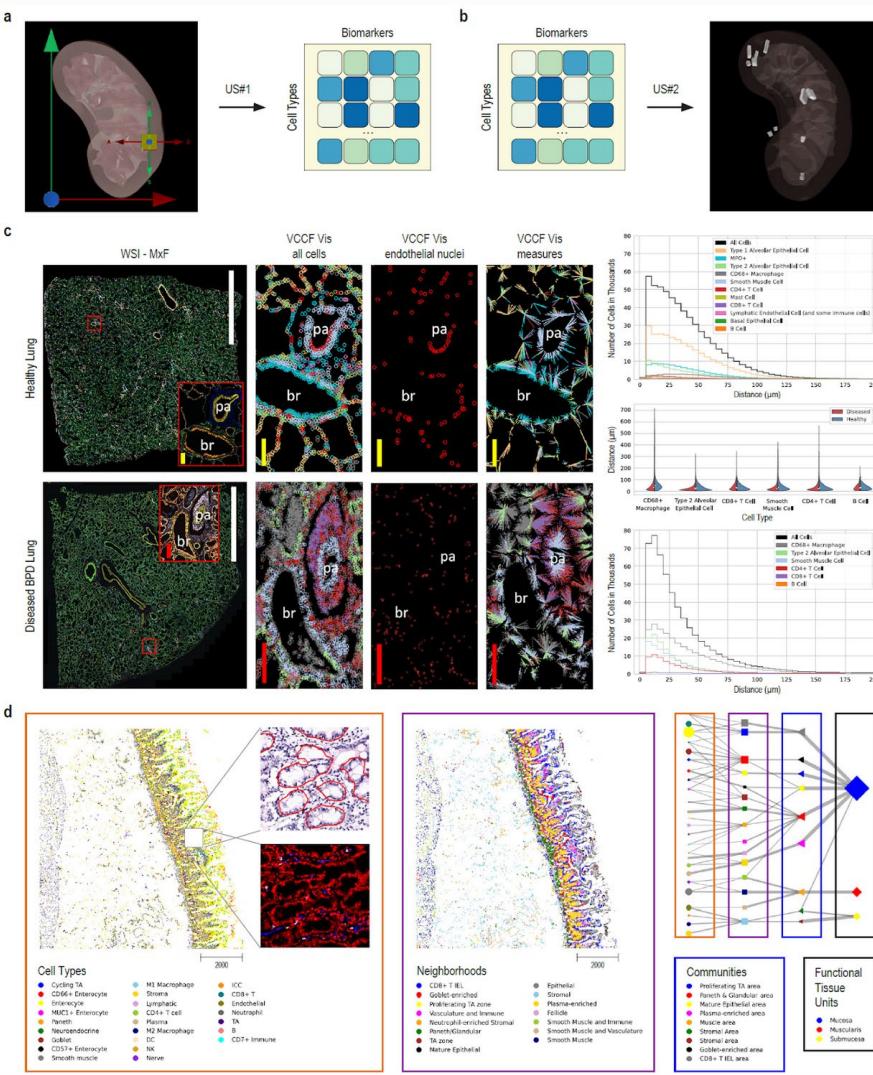


Figure 3: Human Reference Atlas Usage. **a.** User story #1 (US#1) lets a user define a 3D volume inside of the HRA reference body using the RUI and it predicts cell type populations and mean expression values for cell types in that volume. **b.** User story #2 (US#2) reads cell type population data 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 μ m, yellow: 100 μ m. 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>

CIFAR MacMillan Multiscale Human

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.

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.

Founded

2023

Supporters

The MacMillan Family Foundation

CIFAR Contact

[Amy Cook](#)

<https://events.cifar.ca/website/78533/venue/>

<https://cifar.ca/research-programs/cifar-macmillan-multiscale-human/>

Q&A