Converting Data Into Actionable Insights: Open Data, Open Code, and Open Education

Katy Börner

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

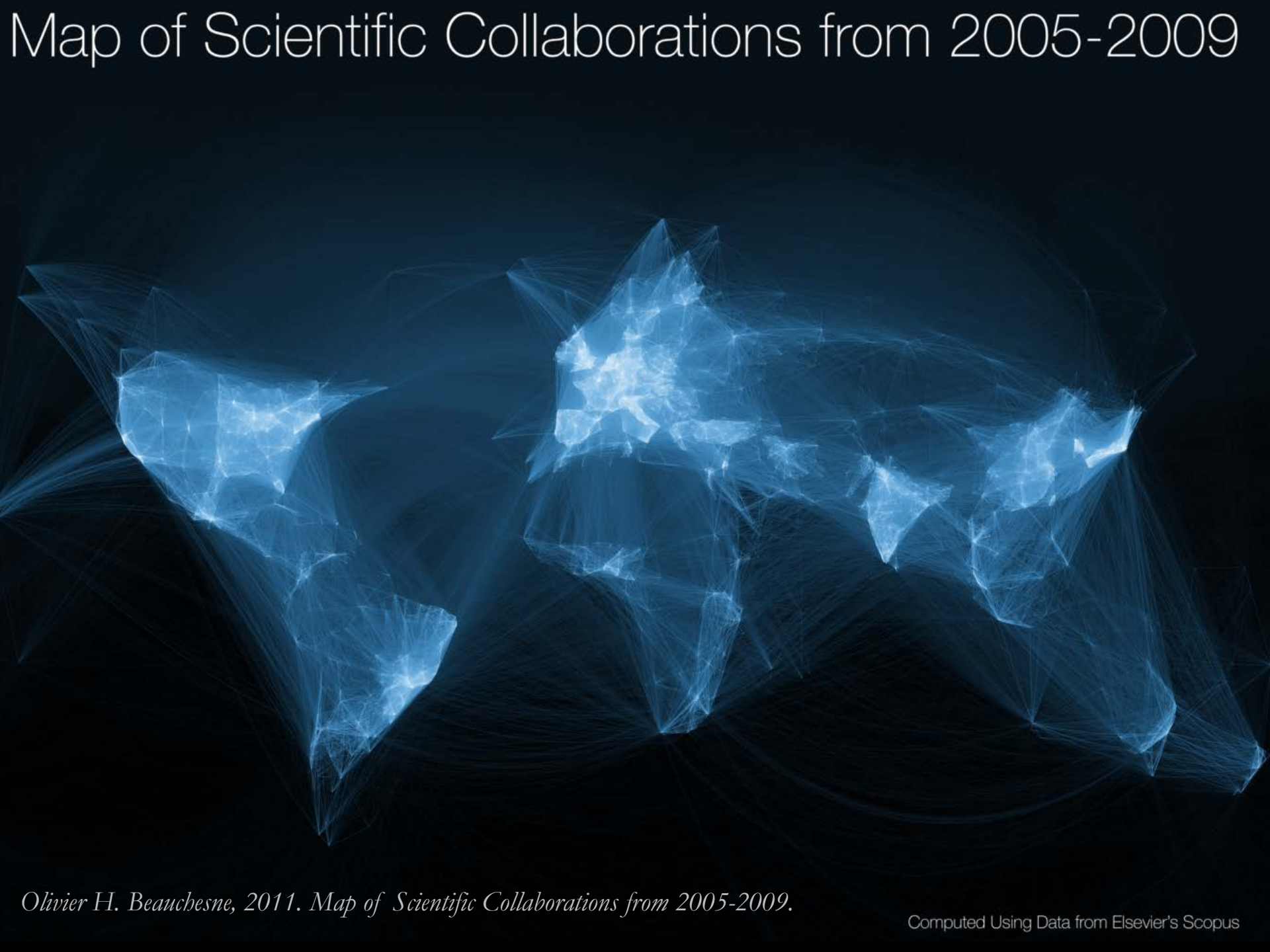
Commerce Data Advisory Council (CDAC) Meeting
National Oceanic and Atmospheric Administration (NOAA)
David Skaggs Research Center, Boulder, CO

October 29-30, 2015

Olivier H. Beauchesne, 2011. Map of Scientific Collaborations from 2005-2009.

Computed Using Data from Elsevier's Scopus

Map of Scientific Collaborations from 2005-2009



Olivier H. Beauchesne, 2011. Map of Scientific Collaborations from 2005-2009.

Open Data

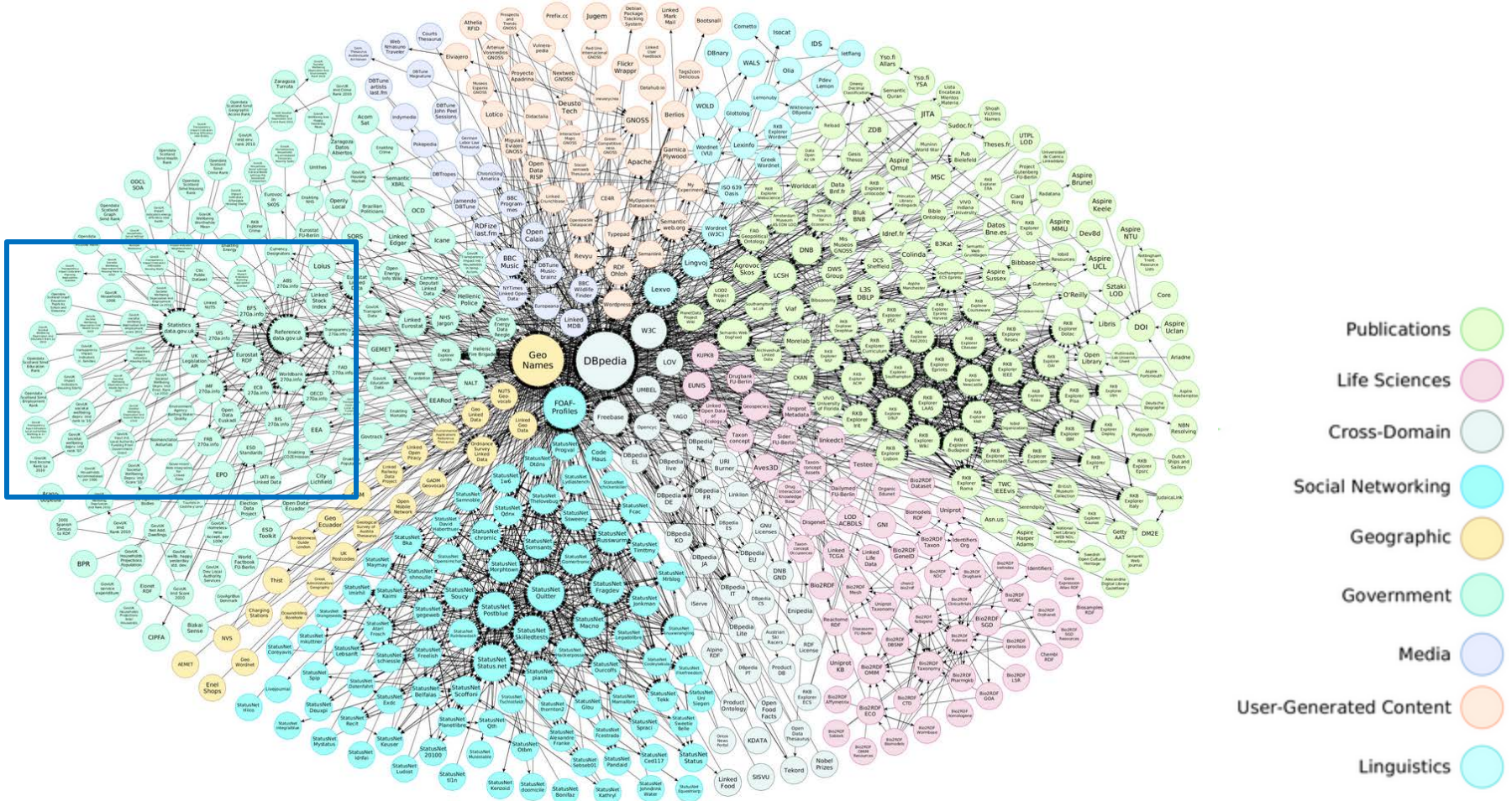
Data - economic, scientific, environmental, patent, geospatial

Core CDAC Mission:

- data management practices
- common, open data standards
- policy issues related to privacy, latency, and consistency
- effective models for public-private partnership
- external uses of Commerce data
- methods to build new feedback loops between the Department and data users.

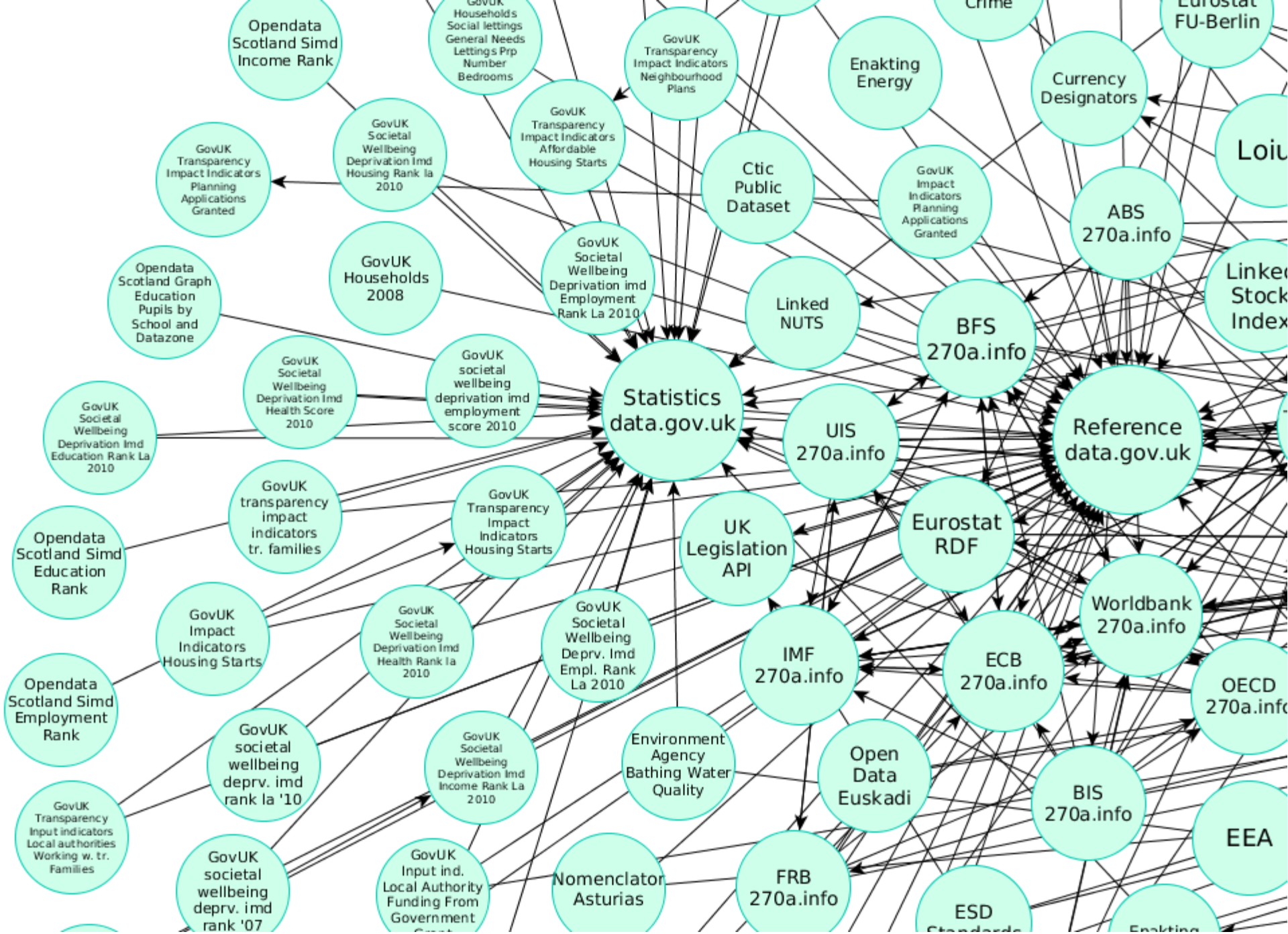
http://www.esa.doc.gov/content/cdac-mission-goals

Data - economic, scientific, environmental, patent, geospatial

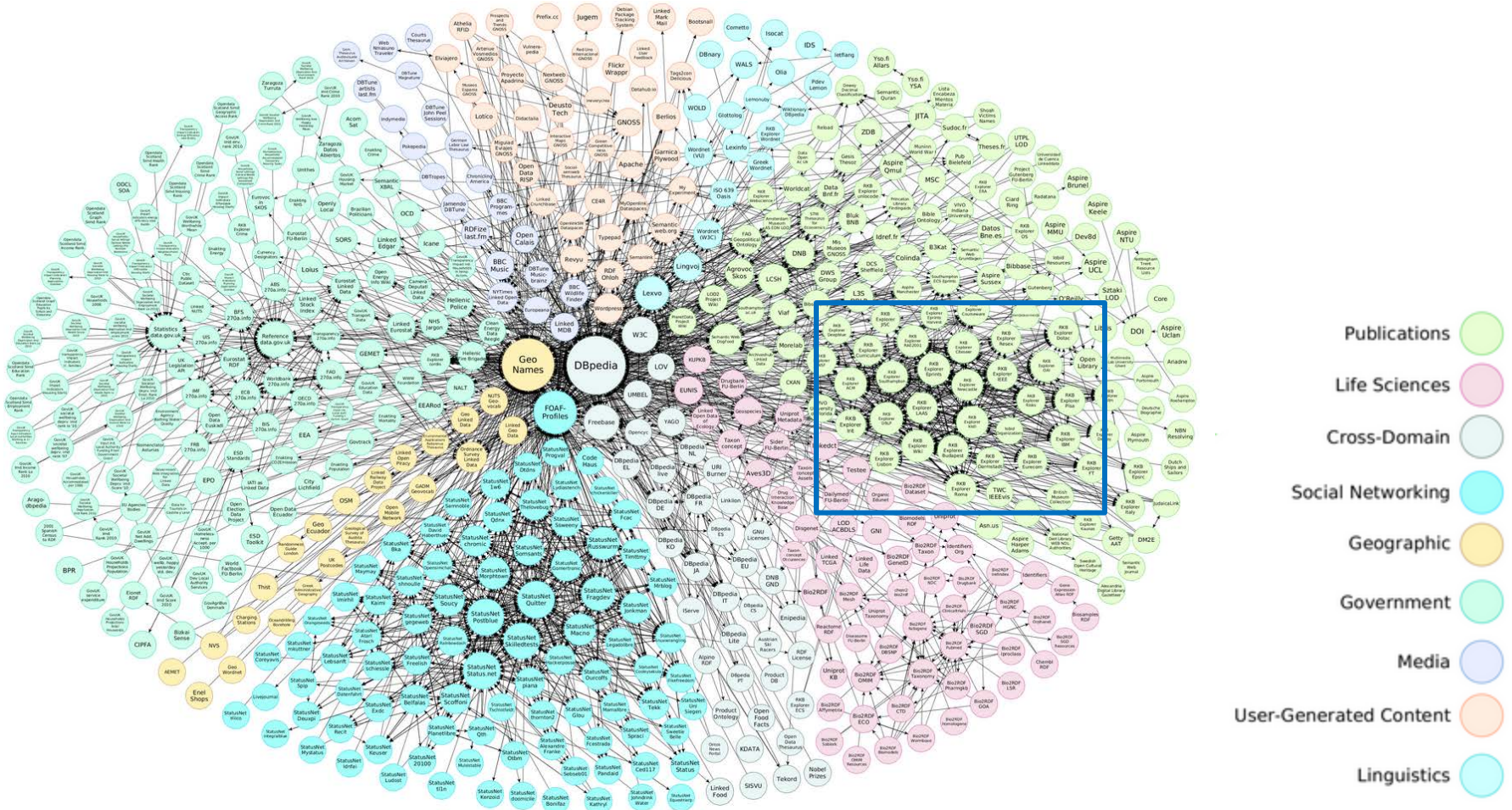


Linked Open Data Cloud Diagram 2014

http://data.dws.informatik.uni-mannheim.de/lodcloud/2014/

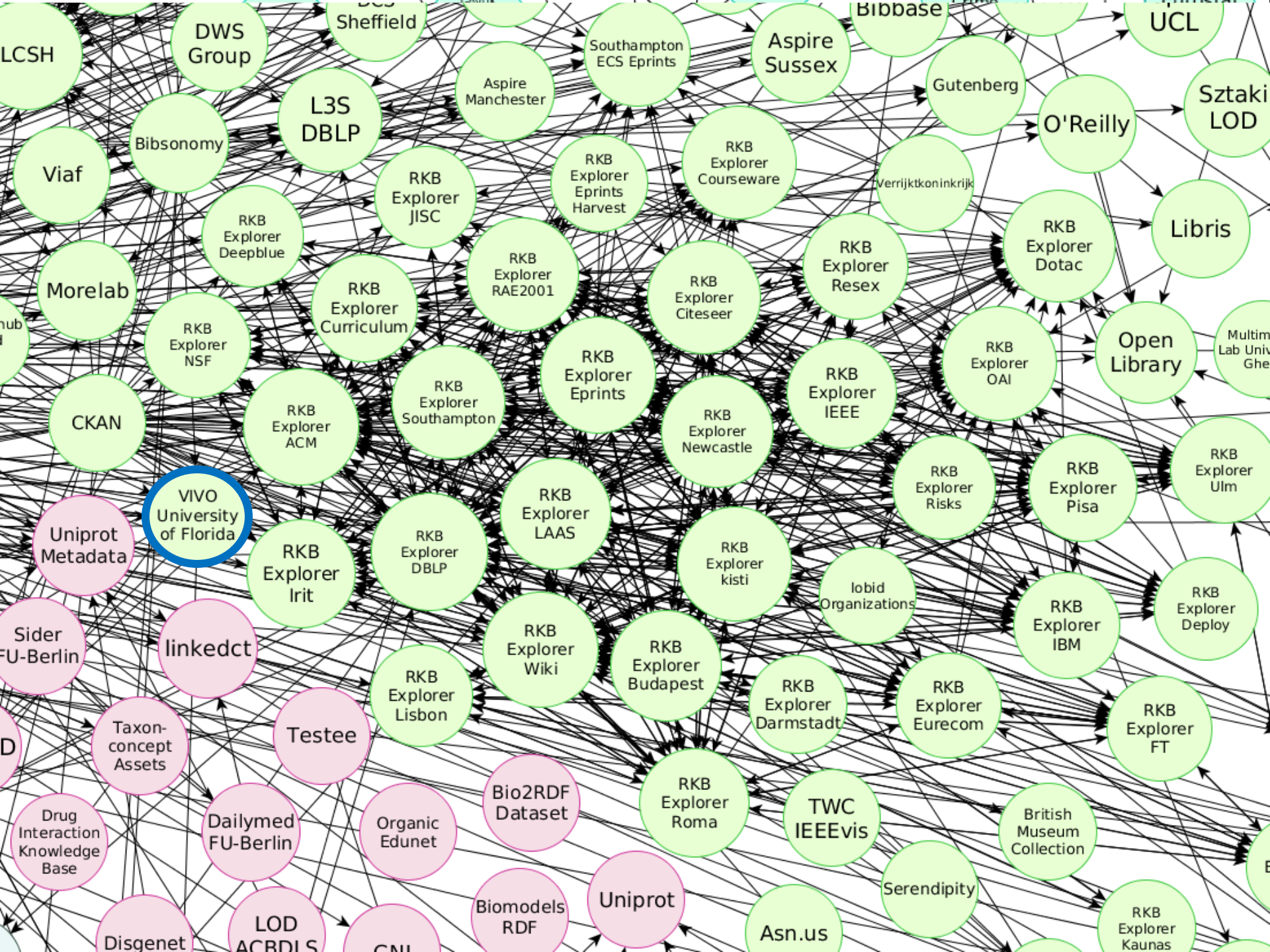

Opendata Scotland Simd Income Rank
GovUK Households Social lettings General Needs Lettings Prp Number Bedrooms
GovUK Transparency Impact Indicators Neighbourhood Plans
Enakting Energy
Currency Designators
Eurostat FU-Berlin
GovUK Transparency Impact Indicators Planning Applications Granted
GovUK Societal Wellbeing Deprivation Imd Housing Rank la 2010
GovUK Transparency Impact Indicators Affordable Housing Starts
Ctic Public Dataset
GovUK Impact Indicators Planning Applications Granted
ABS 270a.info
Opendata Scotland Graph Education Pupils by School and Datazone
GovUK Households 2008
GovUK Societal Wellbeing Deprivation imd Employment Rank La 2010
Linked NUTS
BFS 270a.info
GovUK Societal Wellbeing Deprivation Imd Health Score 2010
GovUK societal wellbeing deprivation imd employment score 2010
Statistics data.gov.uk
UIS 270a.info
Reference data.gov.uk
GovUK Societal Wellbeing Deprivation Imd Education Rank La 2010
GovUK transparency impact indicators tr. families
GovUK Transparency Impact Indicators Housing Starts
UK Legislation API
Eurostat RDF
Opendata Scotland Simd Education Rank
GovUK Impact Indicators Housing Starts
GovUK Societal Wellbeing Deprivation Imd Health Rank la 2010
GovUK Societal Wellbeing Deprv. Imd Empl. Rank La 2010
IMF 270a.info
ECB 270a.info
Worldbank 270a.info
OECD 270a.info
Opendata Scotland Simd Employment Rank
GovUK societal wellbeing deprv. imd rank la '10
GovUK Societal Wellbeing Deprivation Imd Income Rank La 2010
Environment Agency Bathing Water Quality
Open Data Euskadi
BIS 270a.info
EEA
GovUK Transparency Input indicators Local authorities Working w. tr. Families
GovUK societal wellbeing deprv. imd rank '07
GovUK Input ind. Local Authority Funding From Government
Nomenclator Asturias
FRB 270a.info
ESD

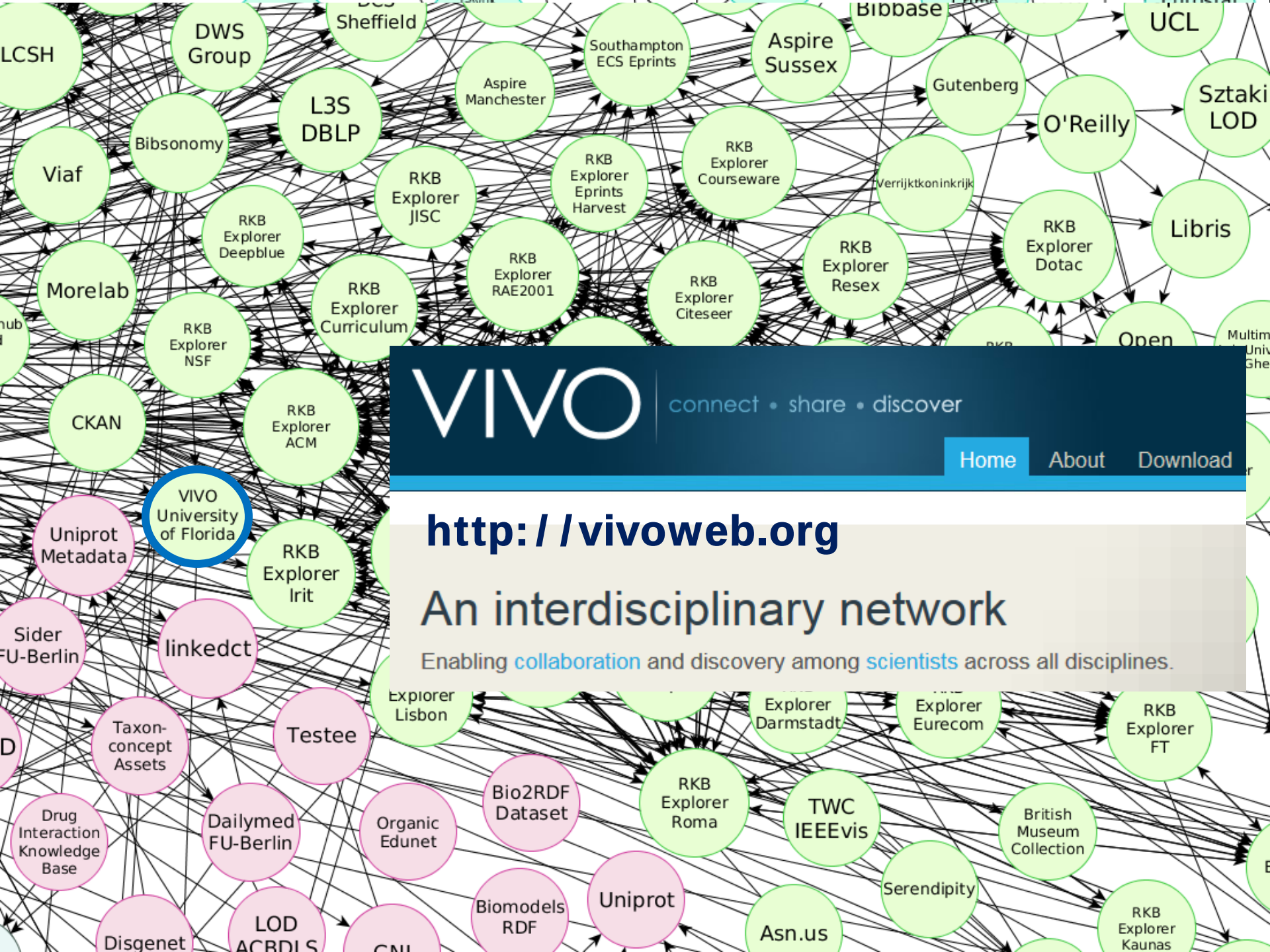
Data – Expertise



Linked Open Data Cloud Diagram 2014

http://data.dws.informatik.uni-mannheim.de/lodcloud/2014/


LCSH
DWS Group
Sheffield
Southampton ECS Eprints
Aspire Sussex
Bibbase
UCL
Aspire Manchester
L3S DBLP
Gutenberg
O'Reilly
Sztaki LOD
Bibsonomy
Viaf
RKB Explorer JISC
RKB Explorer Eprints Harvest
RKB Explorer Courseware
Verrijktkoninkrijk
Libris
RKB Explorer Deepblue
RKB Explorer Dotac
Morelab
RKB Explorer RAE2001
RKB Explorer Resex
RKB Explorer Curriculum
RKB Explorer Citeseer
RKB Explorer NSF
RKB Explorer OAI
Open Library
RKB Explorer Eprints
RKB Explorer IEEE
CKAN
RKB Explorer ACM
RKB Explorer Southampton
RKB Explorer Newcastle
RKB Explorer Risks
RKB Explorer Pisa
RKB Explorer Ulm
VIVO University of Florida
Uniprot Metadata
RKB Explorer LAAS
RKB Explorer Irit
RKB Explorer DBLP
RKB Explorer kisti
lobid Organizations
RKB Explorer IBM
RKB Explorer Deploy
Sider FU-Berlin
linkedct
RKB Explorer Wiki
RKB Explorer Budapest
RKB Explorer Lisbon
RKB Explorer Darmstadt
RKB Explorer Eurecom
RKB Explorer FT
Taxon-concept Assets
Testee
Bio2RDF Dataset
RKB Explorer Roma
TWC IEEEvis
British Museum Collection
Drug Interaction Knowledge Base
Dailymed FU-Berlin
Organic Edunet
Serendipity
Biomodels RDF
Uniprot
Asn.us
RKB Explorer Kaunas
Disgenet
LOD ACBDLS


LCSH
DWS Group
Sheffield
Southampton ECS Eprints
Aspire Sussex
Bibbase
UCL
Aspire Manchester
L3S DBLP
Gutenberg
O'Reilly
Sztaki LOD
Bibsonomy
Viaf
RKB Explorer JISC
RKB Explorer Eprints Harvest
RKB Explorer Courseware
Verrijktkoninkrijk
RKB Explorer Deepblue
RKB Explorer Dotac
Libris
Morelab
RKB Explorer RAE2001
RKB Explorer Resex
RKB Explorer Citeseer
RKB Explorer Curriculum
RKB Explorer NSF
CKAN
RKB Explorer ACM
VIVO University of Florida
Uniprot Metadata
RKB Explorer Irit
Sider FU-Berlin
linkedct
Taxon-concept Assets
Testee
Explorer Lisbon
Explorer Darmstadt
Explorer Eurecom
RKB Explorer FT
Drug Interaction Knowledge Base
Dailymed FU-Berlin
Organic Edunet
Bio2RDF Dataset
RKB Explorer Roma
TWC IEEEvis
British Museum Collection
Serendipity
Disgenet
LOD ACBDLS
Biomodels RDF
Uniprot
Asn.us
RKB Explorer Kaunas
VIVO
connect • share • discover
Home
About
Download
http://vivoweb.org
An interdisciplinary network
Enabling collaboration and discovery among scientists across all disciplines.

http://nrn.cns.iu.edu

Open Data Portals in Europe

Open Data Portals as part of the Europe 2020 Initiative

http://ec.europa.eu/digital-agenda/en/open-data-portals

European Union Open Data Portal

http://open-data.europa.eu/en/data/

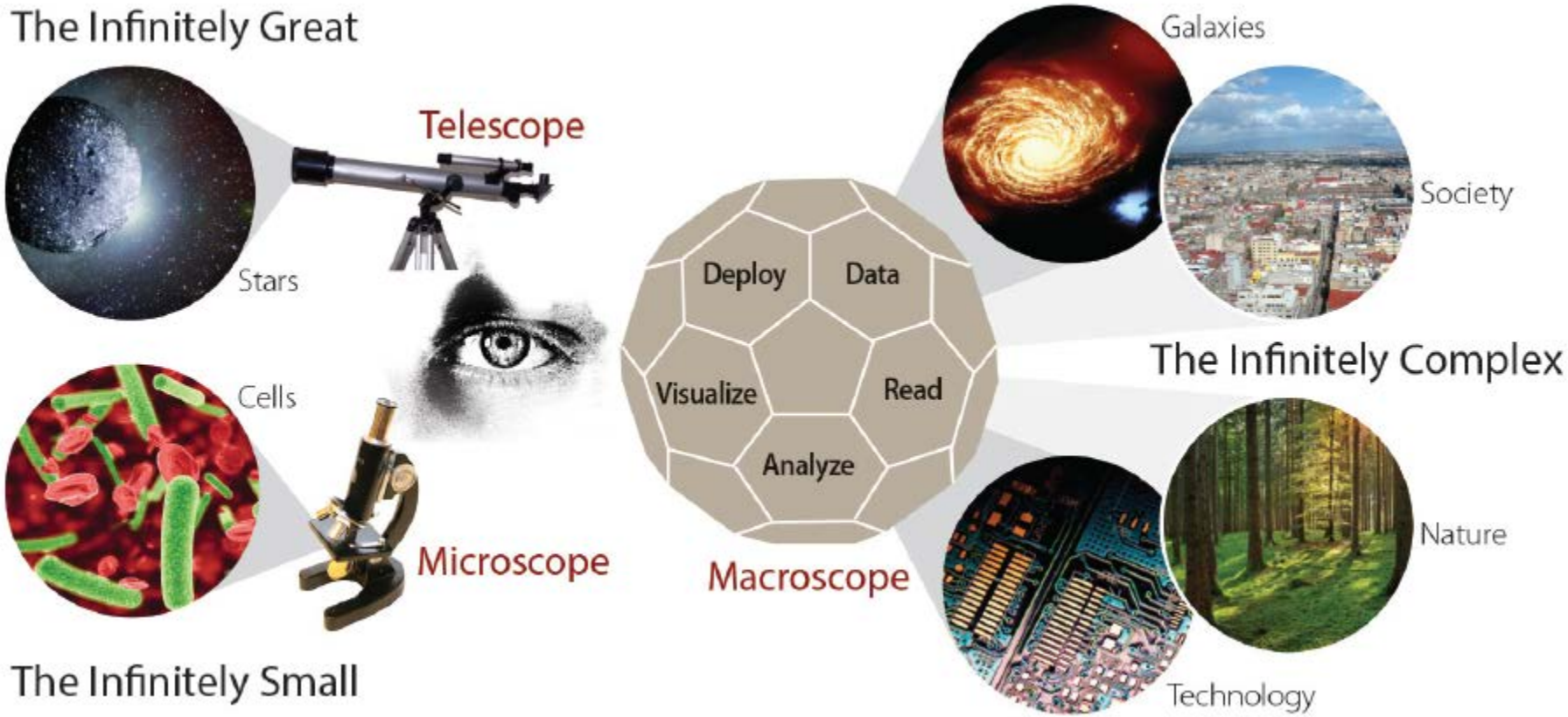
European cloud of public administration data and services

http://open-dai.eu

Open Code

Code – analysis, simulation, visualization

Microscopes, Telescopes, Macroscopes Plug-and-Play Macroscopes



Code – analysis, simulation, visualization

Plug-and-Play Macroscopes

Algorithm Developers

Programmers

Algorithm + Workflow Users

Alg

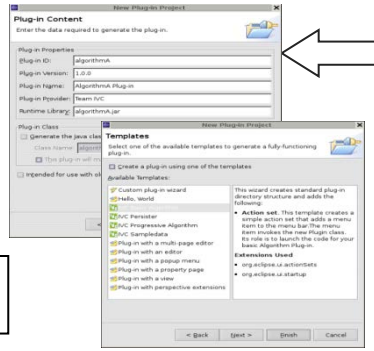
Alg

Alg

Tool

Tool

CIShell Wizards

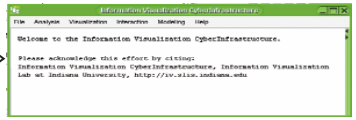


CIShell

CIShell cyberinfrastructure shell

OSGi Alliance

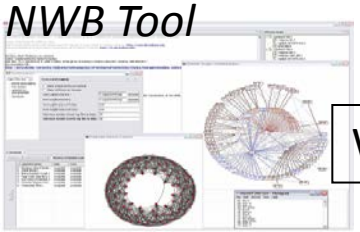
Sci2 Tool



Workflow

Workflow

NWB Tool

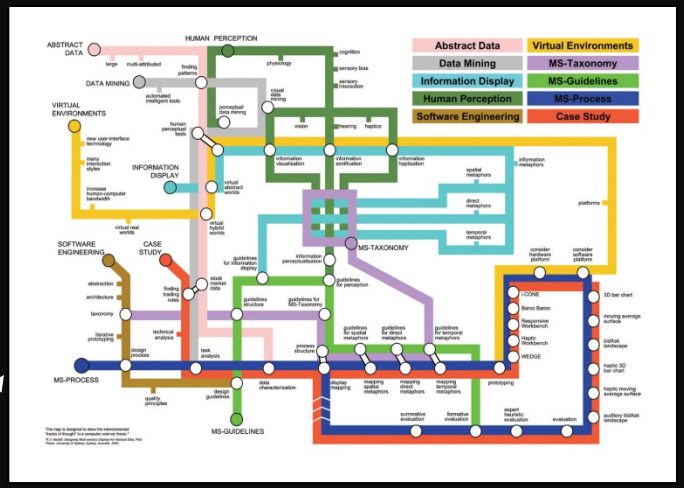


Workflow

Workflow

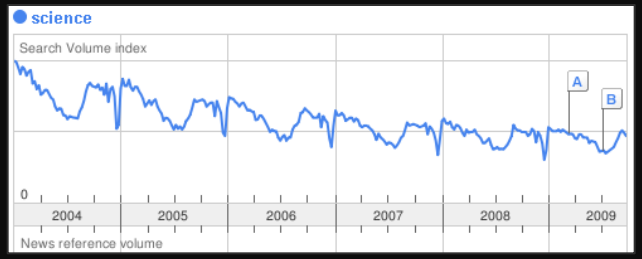
Terra bytes of data

Macroscopes



Find your way

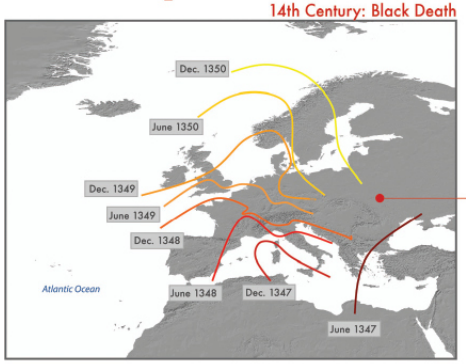
Find collaborators, friends



Identify trends

Impact of Air Travel on Global Spread of Infectious Diseases

14th Century: Black Death

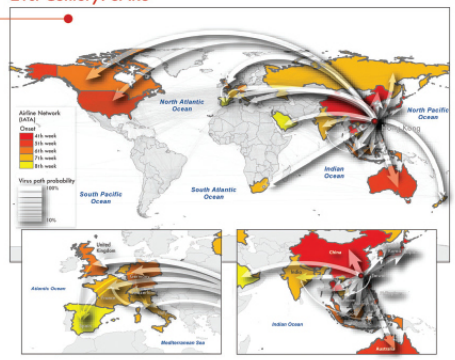


Epidemic spreading pattern changed dramatically after the development of modern transportation systems.

In pre-industrial times disease spread was mainly a spatial diffusion phenomenon. During the spread of Black Death in the 14th century Europe, only few traveling means were available and typical trips were limited to relatively short distances on the time scale of one day. Historical studies confirm that the disease diffused smoothly generating an epidemic front traveling as a continuous wave through the continent at an approximate velocity of 200-400 miles per year.

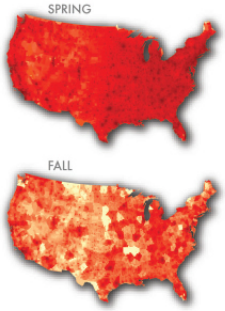
21st Century: SARS

The SARS outbreak on the other hand was characterized by a patched and heterogeneous spatio-temporal pattern mainly due to the air transportation network identified as the major channel of epidemic diffusion and ability to connect far apart regions in a short time period. The SARS maps are obtained with a data-driven stochastic computational model aimed at the study of the SARS epidemic pattern and analysis of the accuracy of the model's predictions. Simulation results describe a spatio-temporal evolution of the disease (color coded countries) in agreement with the historical data. Analysis on the robustness of the model's forecasts leads to the emergence and identification of epidemic pathways as the most probable routes of propagation of the disease. Only few preferential channels are selected (arrows; width indicates the probability of propagation along that path) out of the huge number of possible paths the infection could take by following the complex nature of airline connections (light grey, source: IATA).



Forecasts of the Next Pandemic Influenza

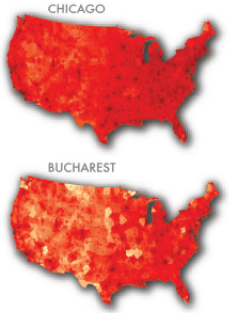
Seasonal



Forecasts are obtained with a stochastic computational model which explicitly incorporates data on worldwide air travel and detailed census data to simulate the global spread of an influenza pandemic.

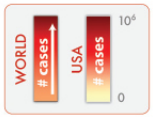
The modeling approach considers infection dynamics (i.e., virus transmission, onset of symptoms, infectiousness, recovery, etc.) among individuals living in urban areas around the world, and assumes that individuals are allowed to travel fom one city to another by means of the airline transportation network.

Geographical

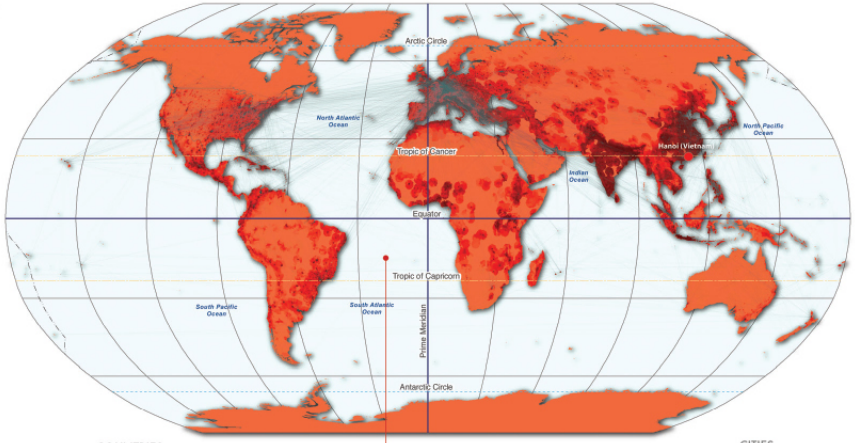


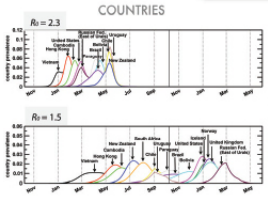
Numerical simulations provide results for the temporal and geographic evolution of the pandemic influenza in 3,100 urban areas located in 220 different countries. The model allows to study different spreading scenarios, characterized by different initial outbreak conditions, both **geographical** and **seasonal**.

The central map represents the cumulative number of cases in the world after the first year from the start of a pandemic influenza with R0=1.9 originating in Hanoi (Vietnam) in the Spring.



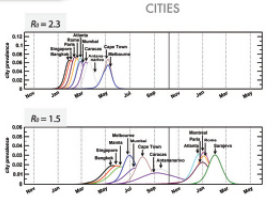
The US maps focus on the situation in the US after one year, and show the effect of changes in the original scenario analyzed. Different color coding is used for the sake of visualization.





R0=1.5 **R0=1.9** **R0=2.3**

Time evolution of a pandemic starting in Hanoi (Vietnam) in the Fall in the no intervention scenario. Profiles of the fraction of infectious individuals in time (prevalence) are shown for some representative countries (left) and cities (right). Two different values of the reproductive number are considered: R0=1.5, consistently with the values shown for the US map (top right), and R0=2.3, in order to provide the comparison with faster spreading.

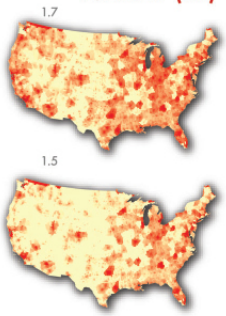


The model inlcudes the worldwide air transportation network (source: IATA) composed of 3,100 airports in 220 countries and E=17,182 direct connections, each of them associated to the corresponding passenger flow. This dataset accounts for 99% of the worldwide traffic and is complemented by the census data of each large metropolitan area served by the corresponding airport.

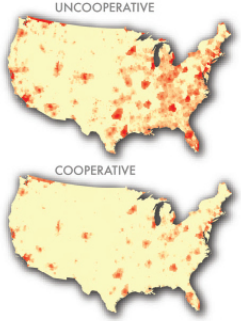
Additional spreading scenarios can be obtained by modeling different levels of infectiousness of the virus, as expressed in terms of the **reproductive number R0**, representing the average number of infections generated by a sick person in a fully susceptible population.

Intervention strategies modeling the use of antiviral drugs can be considered. Two scenarios are compared:
an uncooperative strategy in which countries only use their own stockpiles, and a cooperative intervention which envisions a limited worldwide sharing of the resources.

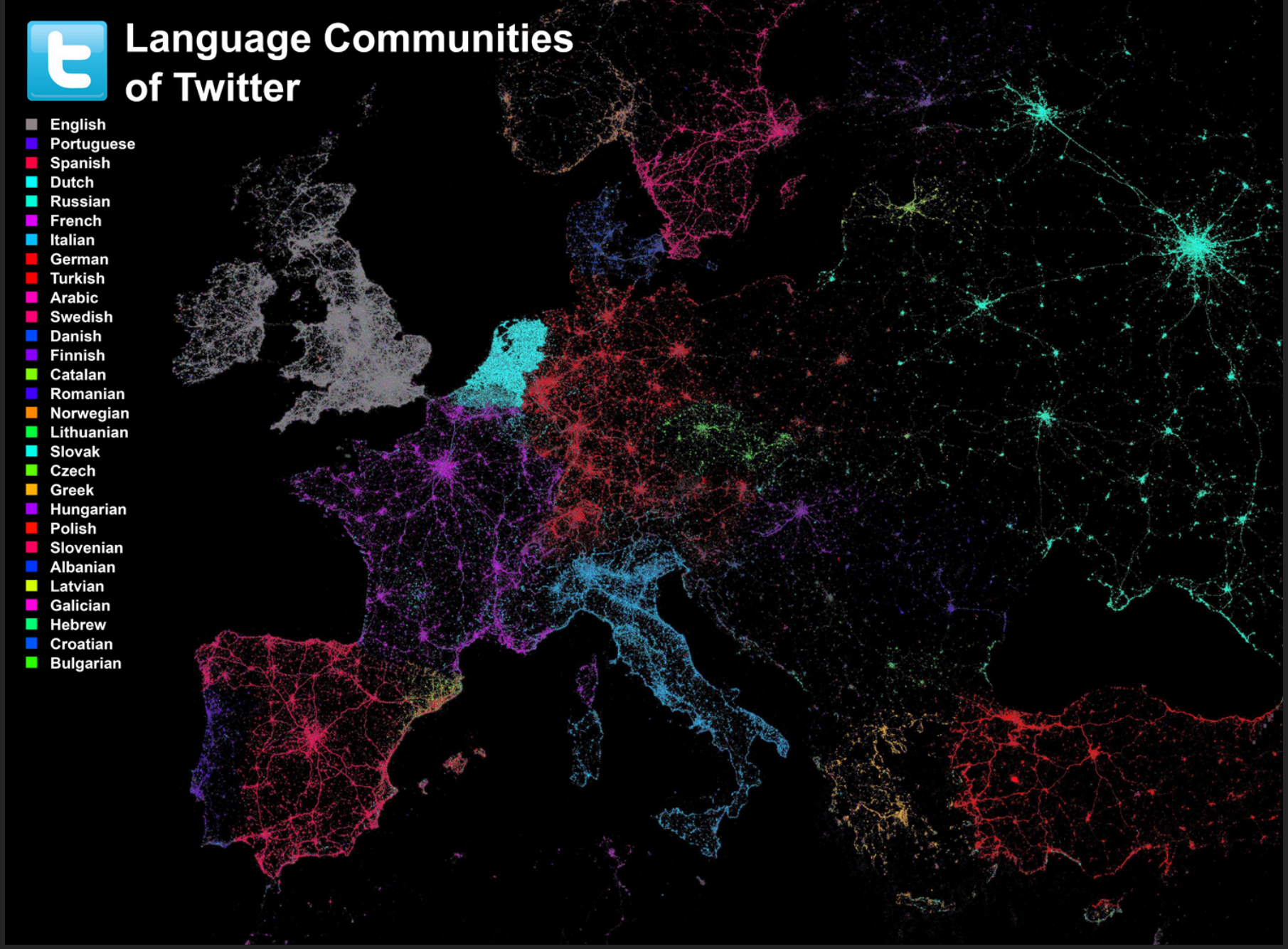
Reproductive Number (R0)



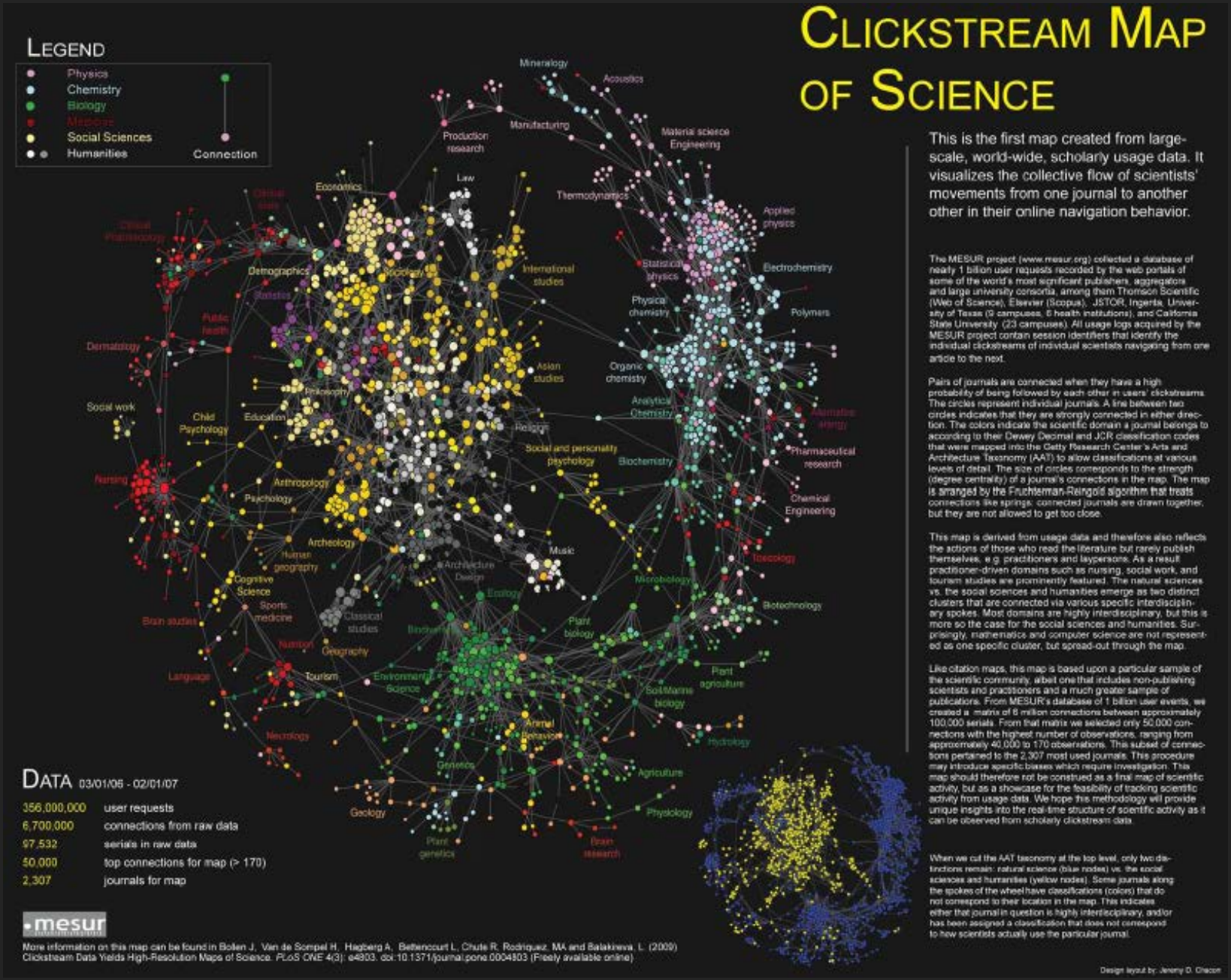
Intervention



Impact of Air Travel on Global Spread of Infectious Diseases - Vittoria Colizza, Alessandro Vespignani - 2007



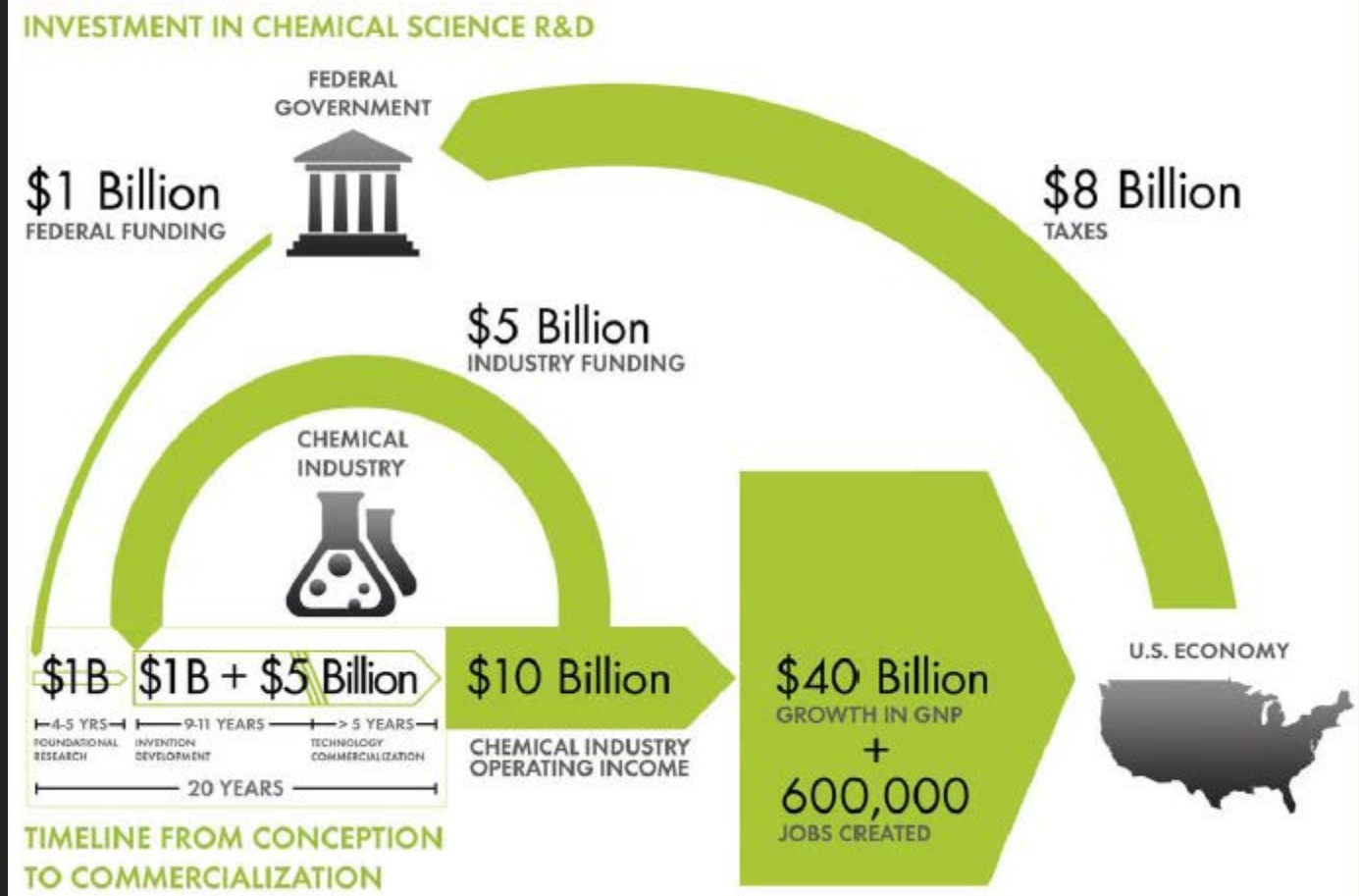
Language Communities of Twitter - Eric Fischer - 2012



Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva. 2008. A Clickstream Map of Science.

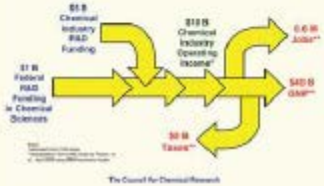
Chemical Research & Development
Powers the U.S. Innovation Engine

Macroeconomic Implications of Public and Private R&D Investments in Chemical Sciences



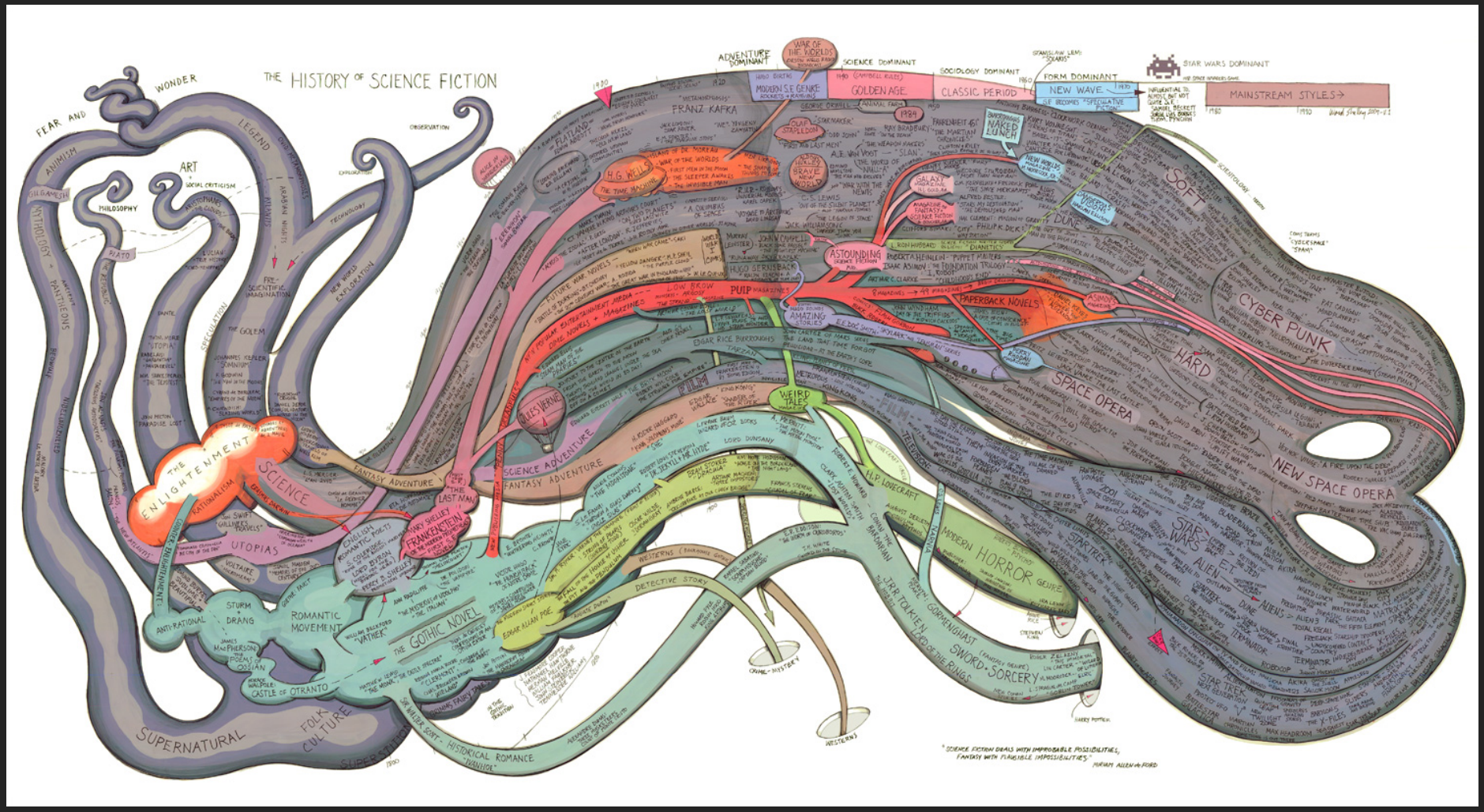
The Council for Chemical Research (CCR)

has provided the U.S. Congress and government policy makers with important results regarding the impact of Federal Research & Development (R&D) investments on U.S. innovation and global competitiveness through its commissioned 5-year two phase study. To take full advantage of typically brief access to policy makers, CCR developed the graphic below as a communication tool that distills the complex data produced by these studies in direct, concise and clear terms.

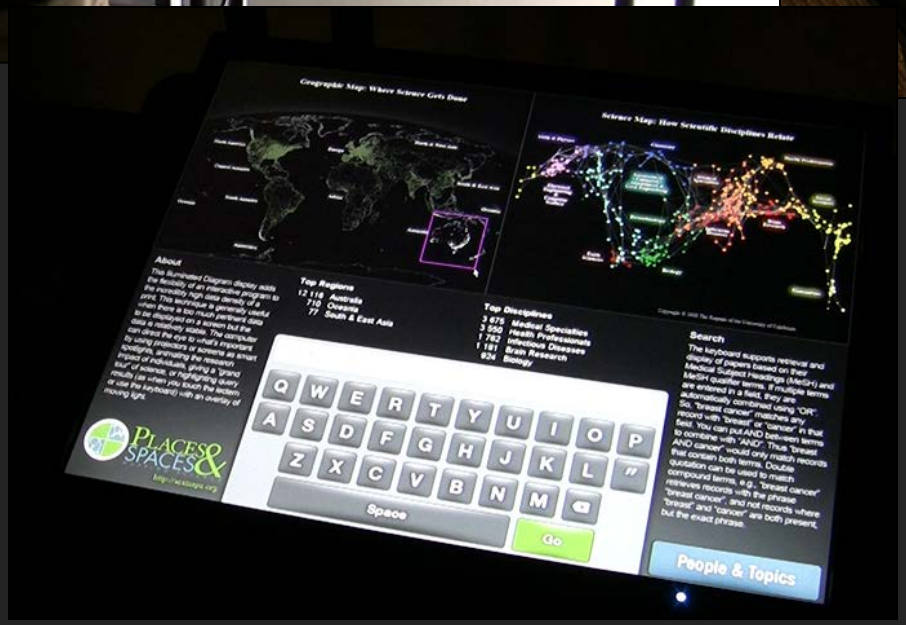


The design shows that an input of $1B in federal investment, leveraged by $5B industry investment, brings new technologies to market and results in $10B of operating income for the chemical industry, $40B growth in the Gross National Product (GNP) and further impacts the US economy by generating approximately 600,000 jobs, along with a return of $8B in taxes. Additional details, also reported in the CCR studies, are depicted in the map to the left. This map clearly shows the two R&D investment cycles; the shorter industry investment at the innovation stage to commercialization cycle; and the longer federal investment cycle which begins in basic research and culminates in national economic and job growth along with the increase tax base that in turn is available for investment in basic research.

Council for Chemical Research. 2009. Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research.



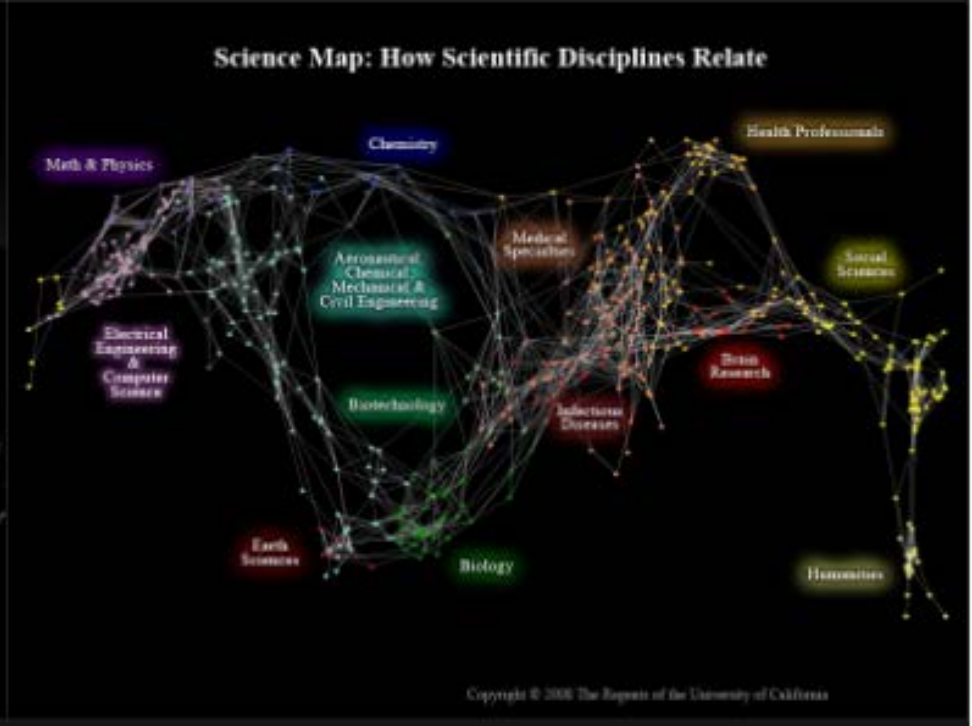
Ward Shelley . 2011. History of Science Fiction.



Illuminated Diagram Display

on display at the Smithsonian in DC.

http://scimaps.org/exhibit_info/#ID



About

This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.

Top Five Continents

North America - 4,000 records
South & East Asia - 3,589
Australia - 2,431
Africa - 2,208
South America - 1,562

Top Five Scientific Disciplines

Math & Physics - 4,000 records
Health Professionals - 3,589
Social Sciences - 2,431
Aeronautical, Chemical, Mechanical & Civil Engineering - 2,208
Humanities - 1,562

Search

The keyboard supports retrieval and display of papers based on their Medical Subject Headings (MeSH) and MeSH qualifier terms. If multiple terms are entered in a field, they are automatically combined using "OR". So, "breast cancer" matches any record with "breast" or "cancer" in that field. You can put AND between terms to combine with "AND". Thus "breast AND cancer" would only match records that contain both terms. Double quotation can be used to match compound terms, e.g., "breast cancer" retrieves records with the phrase "breast cancer", and not records where "breast" and "cancer" are both present, but the exact phrase.

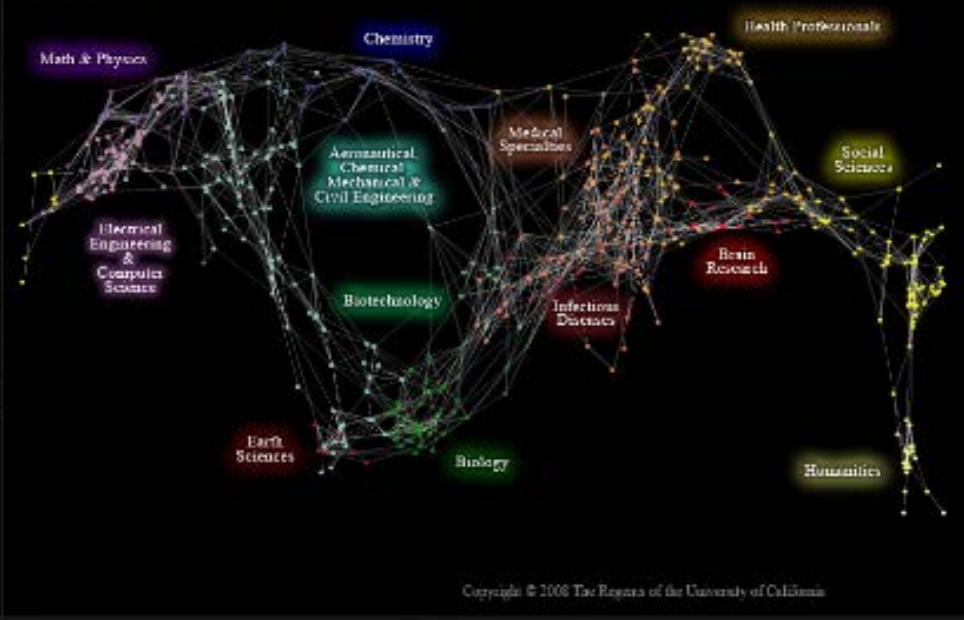
PLACES & SPACES

http://scimaps.org

People & Topics

Geographic Map: Where Science Gets Done

Science Map: How Scientific Disciplines Relate



Copyright © 2008 The Regents of the University of California

About

This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.

Elinor Ostrom - Nobel Prize in Economic Sciences 2009

Born: 7 August 1933, New York, NY, USA

Affiliation at the time of the award: Indiana University, Bloomington, IN, USA, Arizona State University, Tempe, AZ, USA

Prize motivation: "for her analysis of economic governance, especially the commons"

Field: Economic governance

Contribution: Challenged the conventional wisdom by demonstrating how local property can be successfully managed by local commons without any regulation by central authorities or privatization.

Interact

Select any location on the *Geographic Map* location (by brushing your finger over an area on the lectern's touch screen) and topics studied in that area will highlight on the *Science Map*: the brighter a topic glows, the more papers on that topic originated in the selected area. Converstely, touching a scientific area in the *Science Map* illuminates places on the *Geographic Map* where that topic is studied. People and topic buttons support the exploration of publication output by selected Noble laureates and particular lines of research using MEDLINE data from 2000-2009.

Cancer	Cloning	HIV	Robert G. Edwards	Roger D. Kornberg	Elinor Ostrom
Obesity	Quality of Life	Smoking	Stanley B. Prusiner	Ahmed H. Zewail	View All

Keyword Search

PLACES & SPACES
http://scimaps.org

Open Education

Education – K-12, college, grad., informal, professional training

"Expedition Zukunft" German science train visiting 62 cities in 7 months. 12 coaches, 300 m long.

North Carolina State's Immersion Theater

Hidalgo, et al., 2007.

See also The Product Space map in http://scimaps.org

Places & Spaces Exhibit at the David J. Sencer CDC Museum, Atlanta, GA
January 25-June 17, 2016

Information Visualization MOOC 2015

INDIANA UNIVERSITY CNS

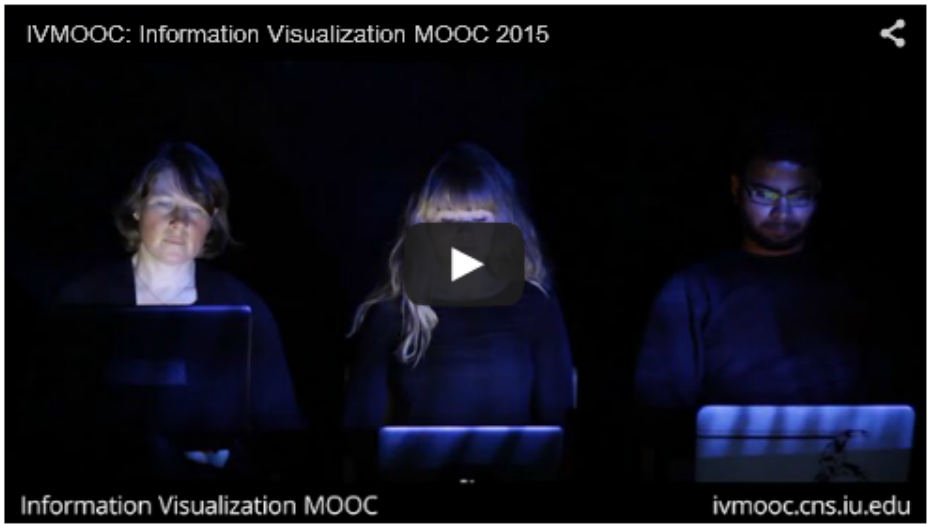
Overview

This course provides an overview about the state of the art in information visualization. It teaches the process of producing effective visualizations that take the needs of users into account.

The course can be taken for three Indiana University credits as part of the Online Data Science Program, as part of the Information and Library Science M.S. program, and as part of the online Data Science M.S. Program offered by the School of Informatics and Computing. Students seeking enrollment information should contact Rhonda Spencer at 812-855-2018, ilsmain@indiana.edu or datasci@indiana.edu.

Among other topics, the course covers:

- Data analysis algorithms that enable extraction of patterns and trends in data
- Major temporal, geospatial, topical, and network visualization techniques
- Discussions of systems that drive research and development.



Register for Course

Already registered? Click here to go to the course.
Forgot your password? Click here to reset it.

Register for free at http://ivmooc.cns.iu.edu. Class restarts January 12, 2016.

50 courses available

By start date

- Recently started or starting soon (12)
- Just Announced (2)
- Courses in Progress (4)
- Future courses (14)
- Self Paced (9)
- Finished courses (20)

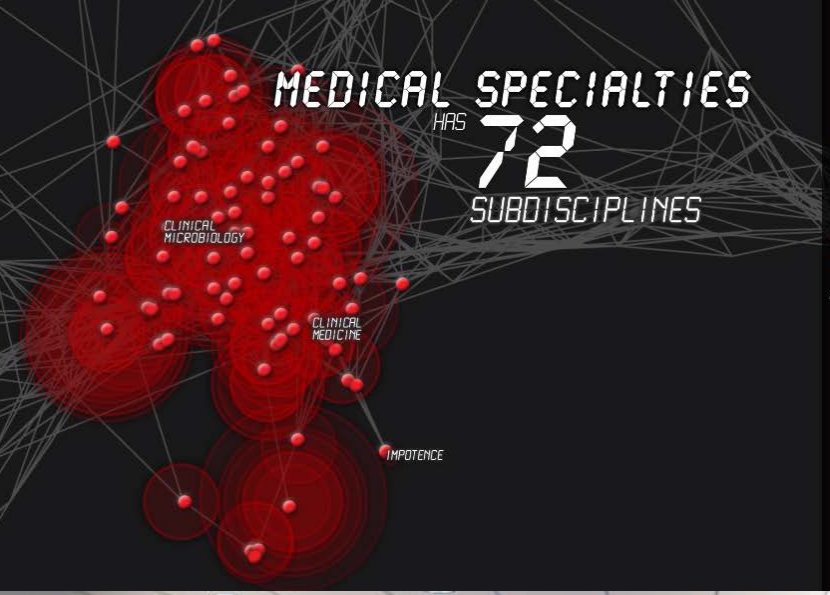
By subject

- Computer Science (23)
- Health & Medicine (5)
- Mathematics (4)
- Business & Management (6)

Add	Course Name	Start Date	Rating
Ad	Introduction to Programming for the Visual Arts with p5.js University of California, Los Angeles via **Kadenze**	4th Nov, 2015	★★★★★
+	Data Visualization University of Illinois at Urbana-Champaign via **Coursera** EARN A CREDENTIAL Part of Data Mining Specialization	20th Jul, 2015	★★★½ 11
+	Information Visualization Indiana University via **Independent**	13th Jan, 2015	0
+	Data Management and Visualization Wesleyan University via **Coursera** EARN A CREDENTIAL Part of Data Analysis and Interpretation Specialization	26th Oct, 2015	★★★★★ 1
+	Data Visualization and D3.js via **Udacity**	Self paced	★★★ 1
+	Communicating Results: Visualization, Ethics, Reproducibility University of Washington via **Coursera** EARN A CREDENTIAL Part of Data Science at Scale Specialization	1st Nov, 2015	0
+	Data Visualization and Communication with Tableau Duke University via **Coursera** EARN A CREDENTIAL Part of Excel to MySQL: Analytic Techniques for Business Specialization	1st Nov, 2015	0

https://www.class-central.com/search?q=visualization

Science Forecast S1:E1, 2015



References

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

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1

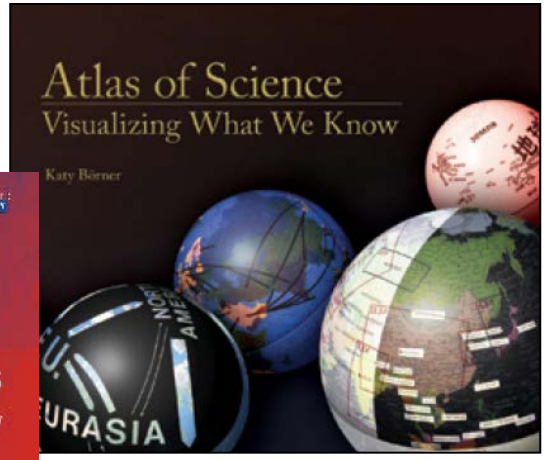
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

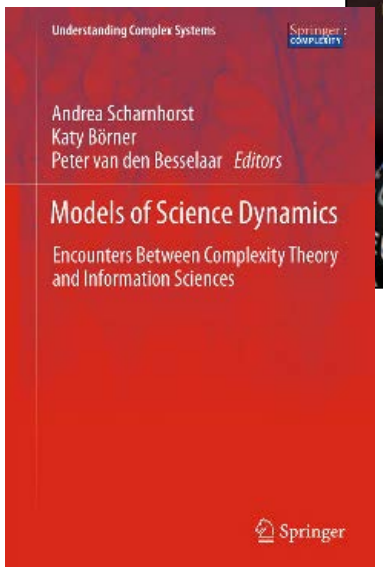
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

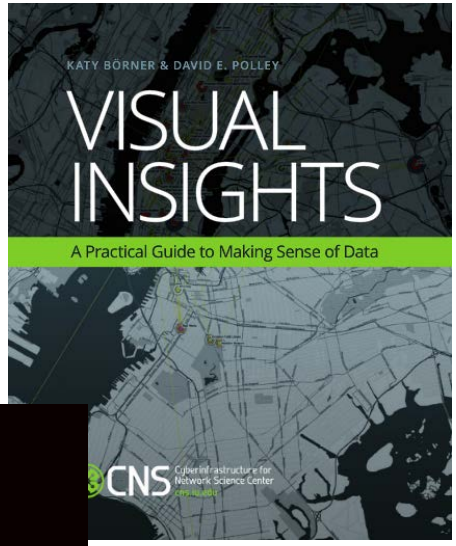
Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

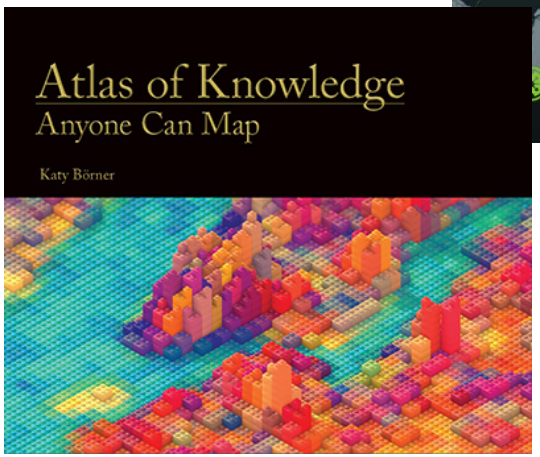
Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

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









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

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