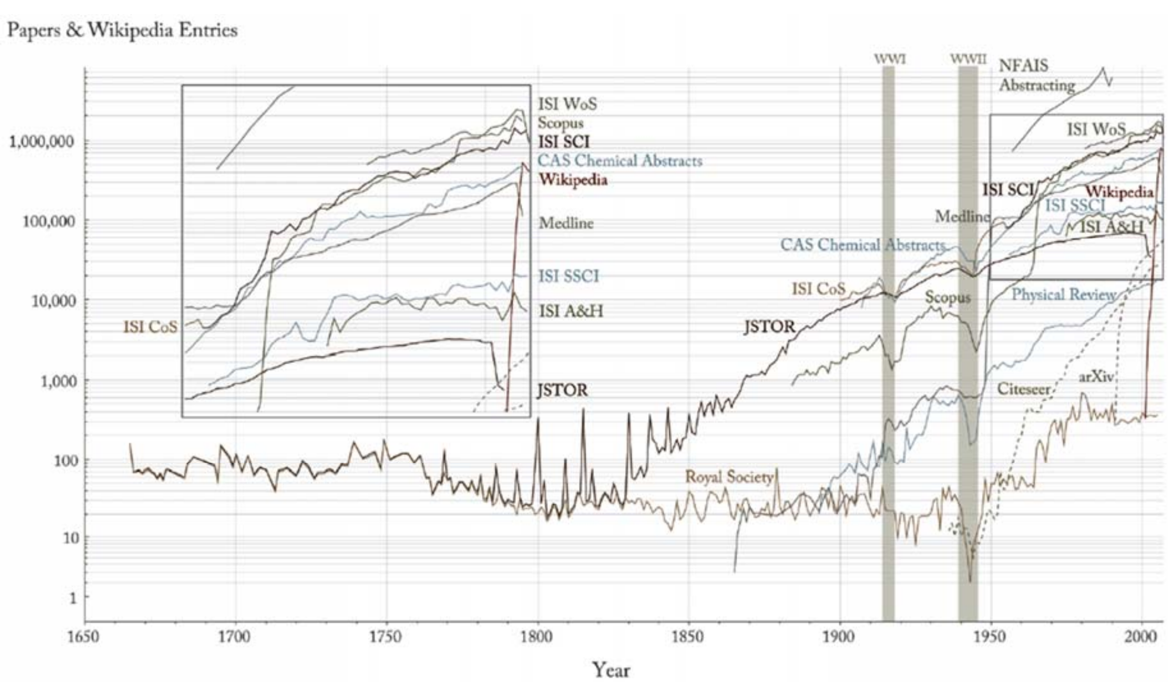
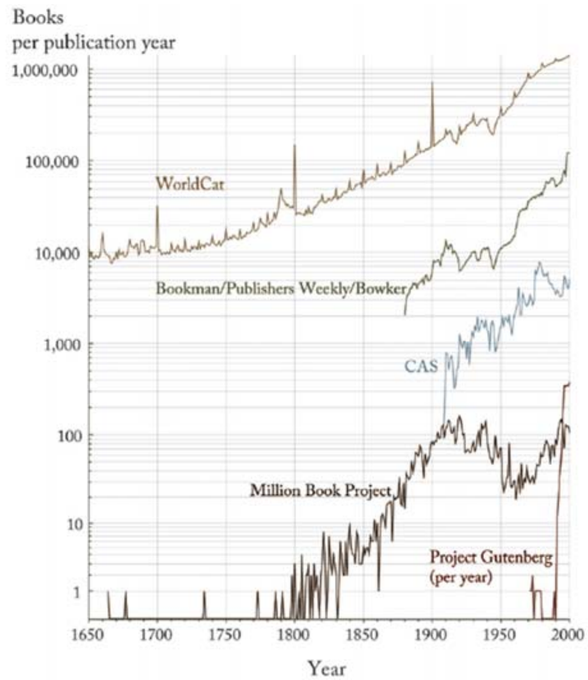
Envisioning Scholarly Data

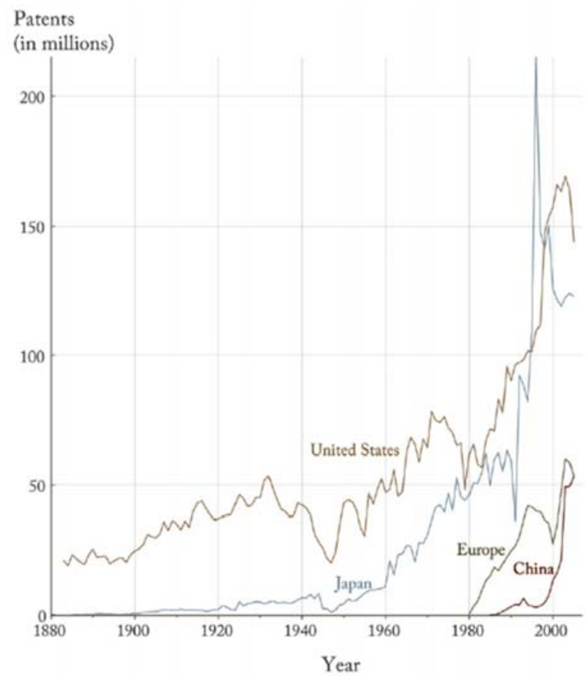
Katy Börner
CNS, SLIS, IU, Bloomington, IN
katy@indiana.edu | http://cns.iu.edu

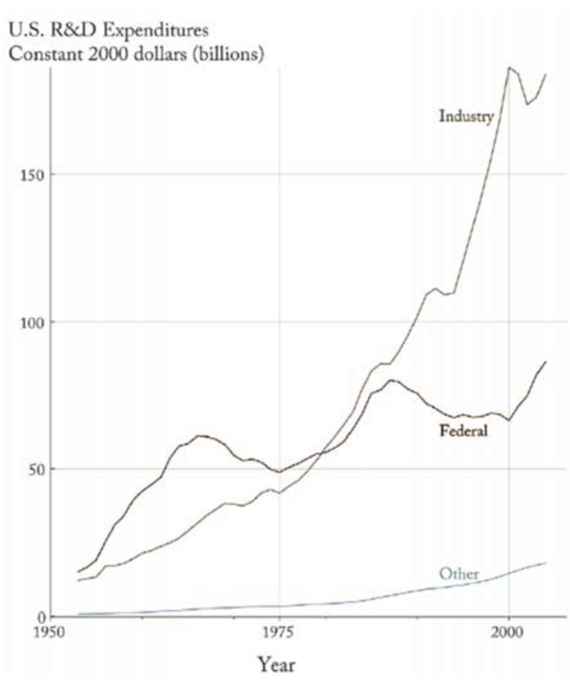
Compatible Data Initiative Meeting
Fordham University-Lincoln Center Campus, NYC
September 23, 2011

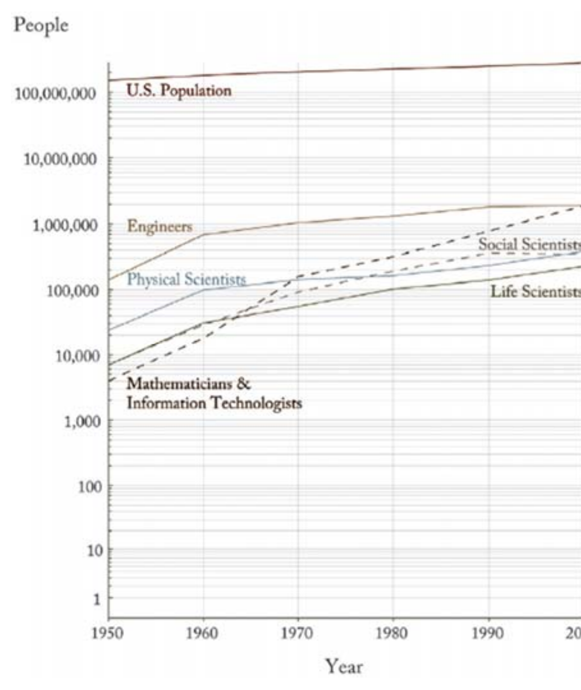
The Rise of Science and Technology

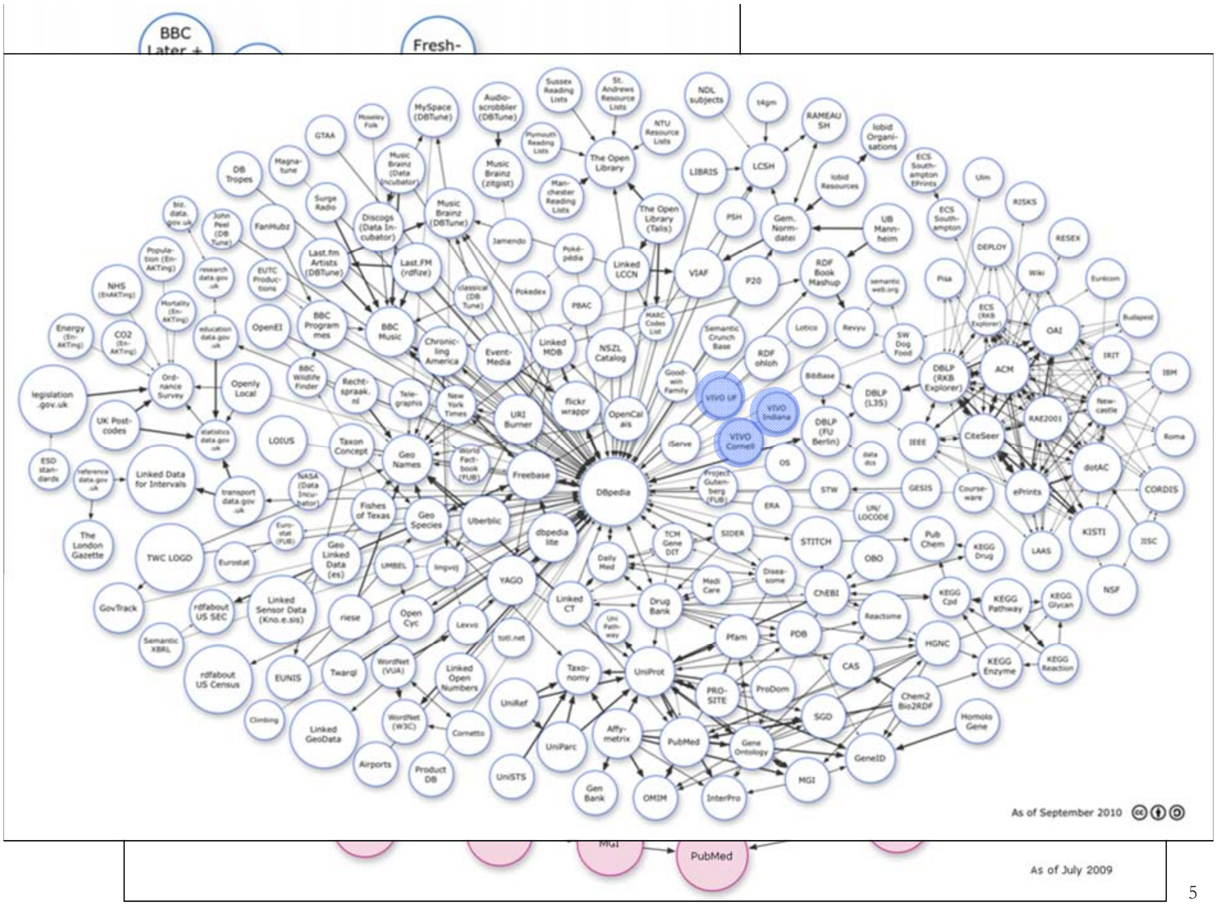



Books
per publication year
1,000,000
100,000
10,000
1,000
100
10
1
WorldCat
Bookman/Publishers Weekly/Bowker
CAS
Million Book Project
Project Gutenberg
(per year)
1650
1700
1750
1800
1850
1900
1950
2000
Year


Patents
(in millions)
200
150
100
50
0
United States
Japan
Europe
China
1880
1900
1920
1940
1960
1980
2000
Year

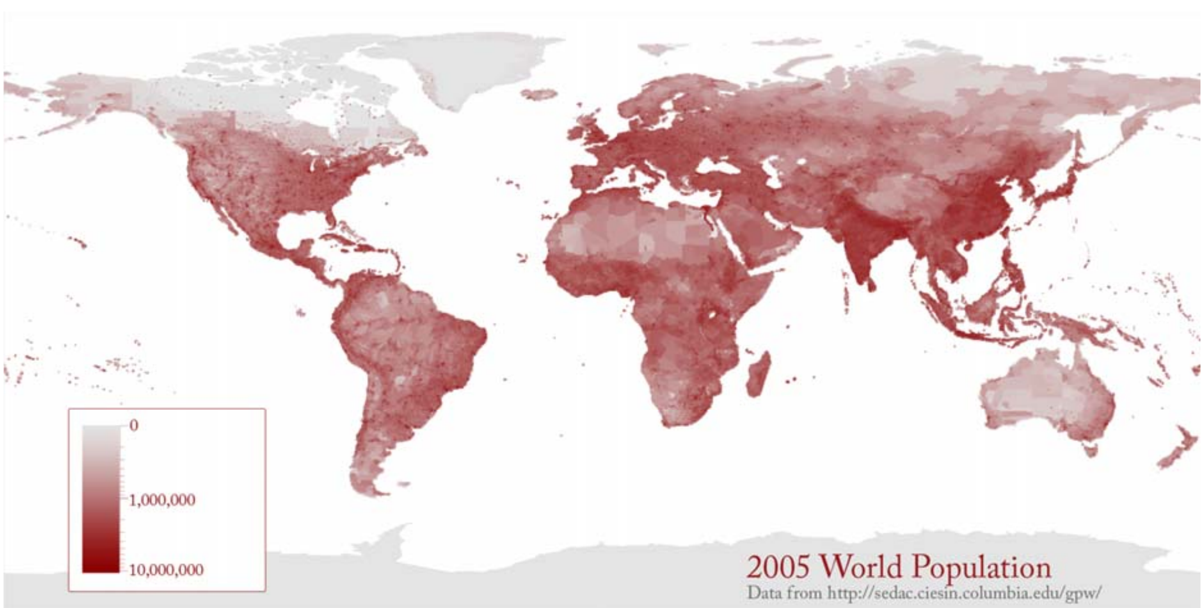

U.S. R&D Expenditures
Constant 2000 dollars (billions)
150
100
50
0
Industry
Federal
Other
1950
1975
2000
Year


People
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
100
10
1
U.S. Population
Engineers
Social Scientists
Physical Scientists
Life Scientists
Mathematicians &
Information Technologists
1950
1960
1970
1980
1990
2000
Year



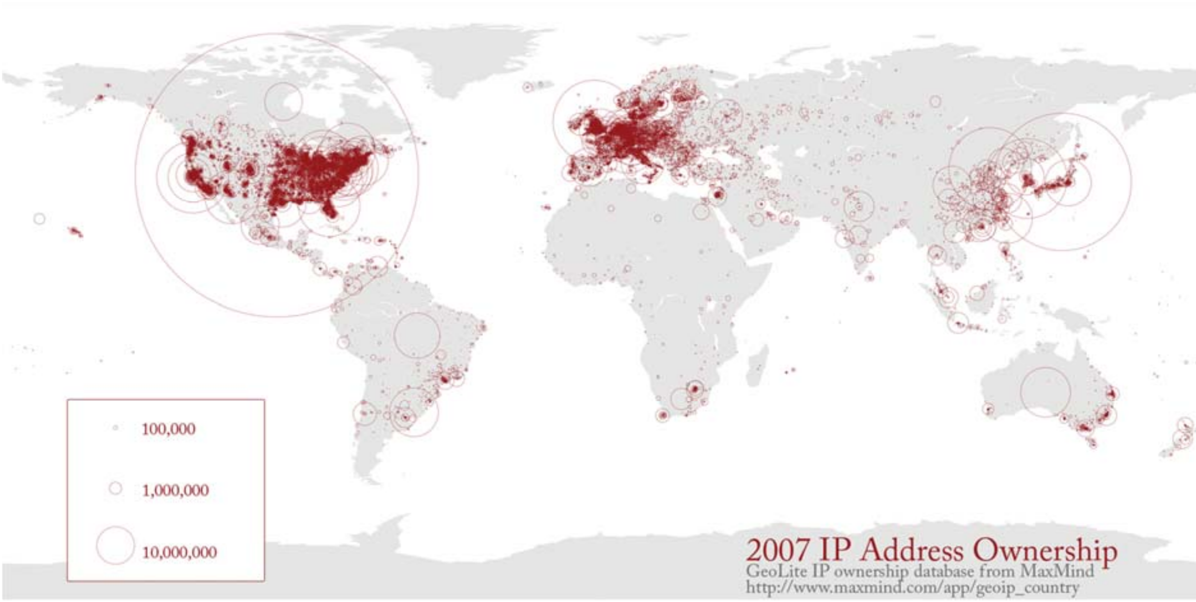
2005 World Population

The population map uses a quarter degree box resolution. Boxes with zero people are given in white. Darker shades of red indicate higher population counts per box using a logarithmic interpolation. The highest density boxes appear in Mumbai, with 11,687,850 people in the quarter degree block, Calcutta (10,816,010), and Shanghai (8,628,088).



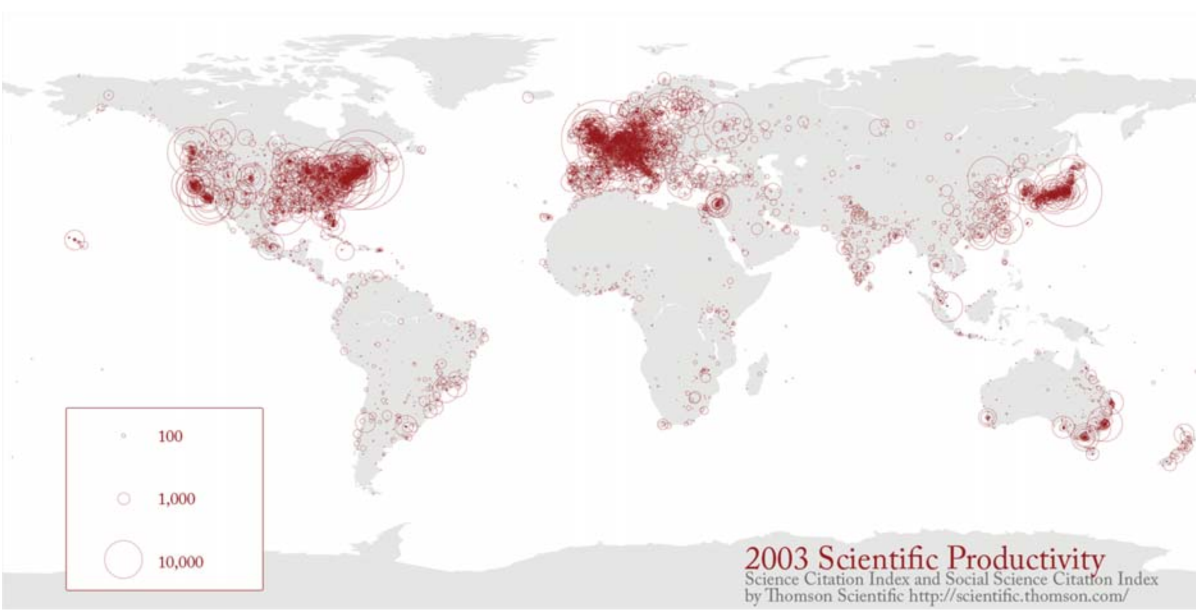
2007 IP Address Ownership

This map shows IP address ownership by location. Each owner is represented by a circle and the area size of the circle corresponds to the number of IP addresses owned. The larges circle denotes MIT's holdings of an entire class A subnet, which equates to 16,581,375 IP addresses. The countries that own the most IP addresses are US (560 million), Japan (130 million), Great Britain (47 million).



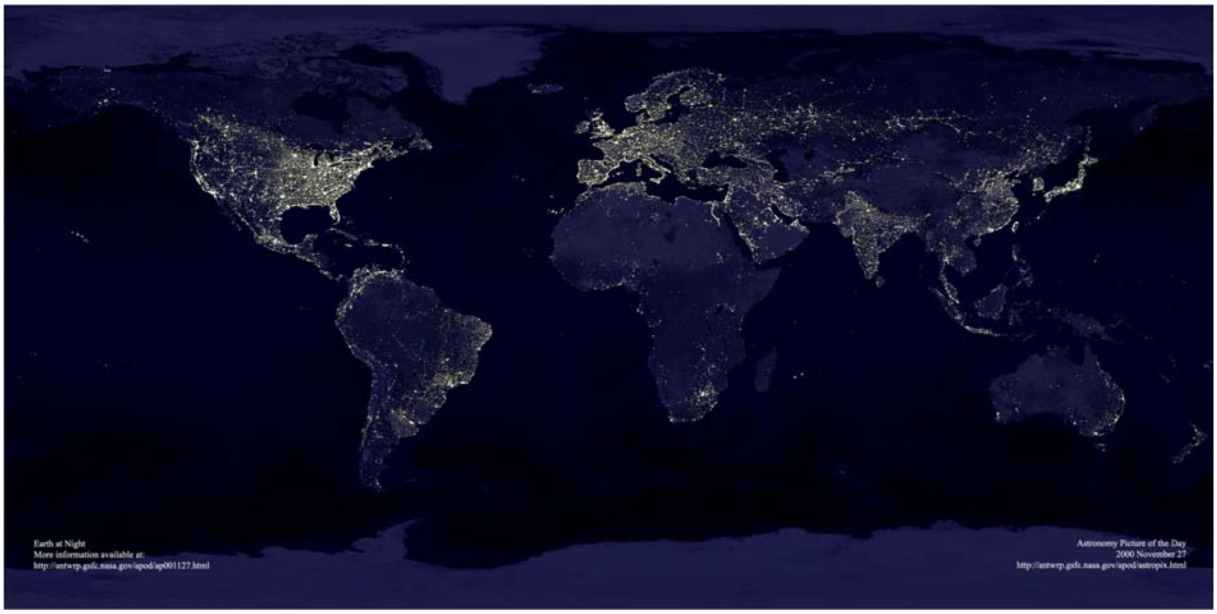
2003 Scientific Productivity

Shown is where science is performed today. Each circle indicates a geographic location at which scholarly papers are published. The larger the circle the more papers are produced. Boston, MA, London, England, and New York, NY are the top three paper production areas. Note the strong resemblance with the Night on Earth and the IP Ownership maps and the striking differences to the world population map.

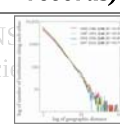
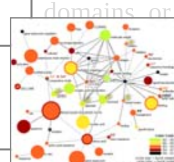
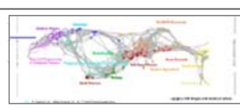
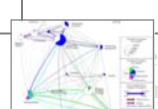
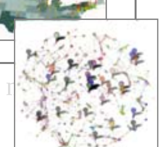
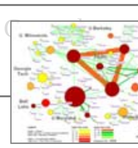
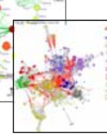
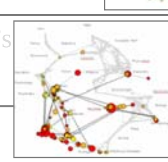


2000 Night on Earth

This image shows city lights at night. It was composed from hundreds of pictures made by orbiting satellites. The seaboards of Europe, the eastern United States, and Japan are particularly well lit. Many cities exist near rivers or oceans so that goods can be exchanged cheaply by boat. The central parts of South America, Africa, Asia, and Australia are rather dark despite their high population density, see map to the left.



Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	*Meso/Local (101–10,000 records)*	*Macro/Global (10,000 < records)*
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NS SA, all of sci 
Temporal Analysis (When)	Funding portfolio of one individual	 ic bursts f PNAS	113 Years of P Research 
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a s intellectual l 	PNAS 
Topical Analysis (What)	 s.	 research	VxOrd/Topic r NIH funding 
Network Analysis (With Whom?)	NSF work of one 	C k  	NIH's cy 

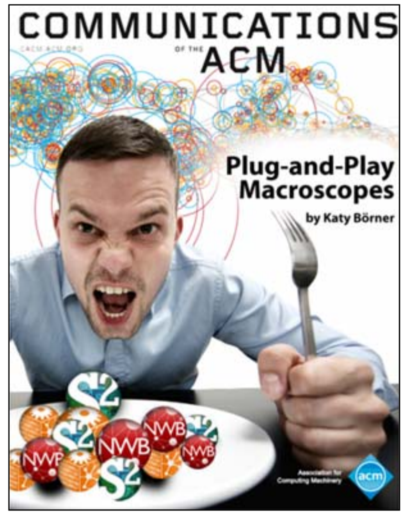
Plug-and-Play Macroscopes

While **microscopes** and **telescopes** are physical instruments, **macroscopes** resemble continuously changing bundles of software plug-ins.

Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube. Assembling custom tools is as quick as compiling your custom music collection.

They provide a **common standard** for

- the design of **modular, compatible algorithm and tool plug-ins**
- that can be **easily combined into scientific workflows**, and
- **packaged as custom tools.**

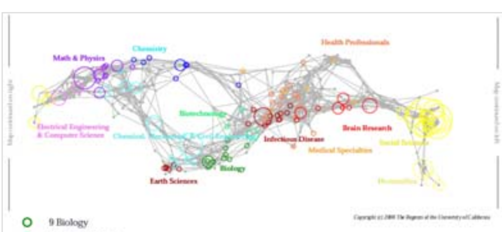


Börner, Katy. (2011). Plug-and-Play Macroscopes. *Communications of the ACM, 54(3), 60-69.*
Video is at http://www.scivee.tv/node/27704

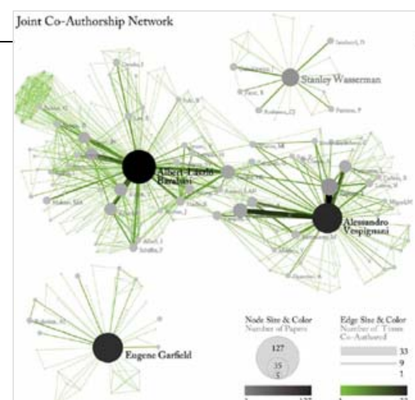
Sci² Tool – "Open Code for S&T Assessment"

http://sci2.cns.iu.edu

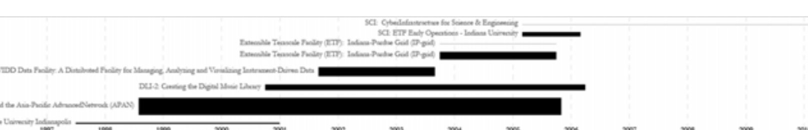
OSGi/CIShell powered tool with NWB plugins and many new scientometrics and visualizations plugins.



Sci Maps



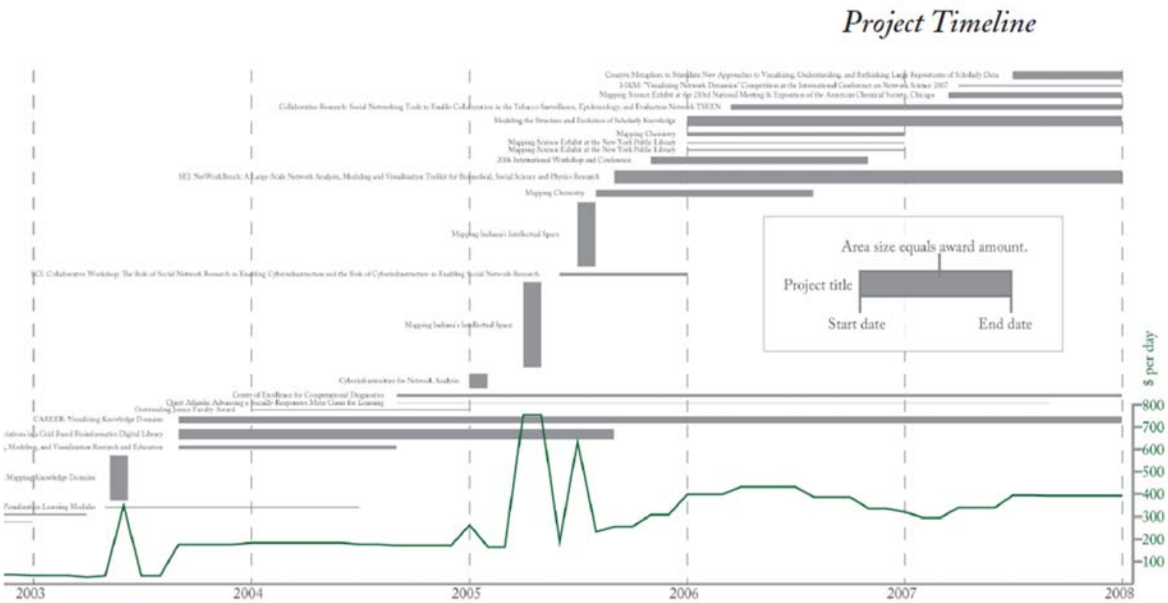
GUESS Network Vis



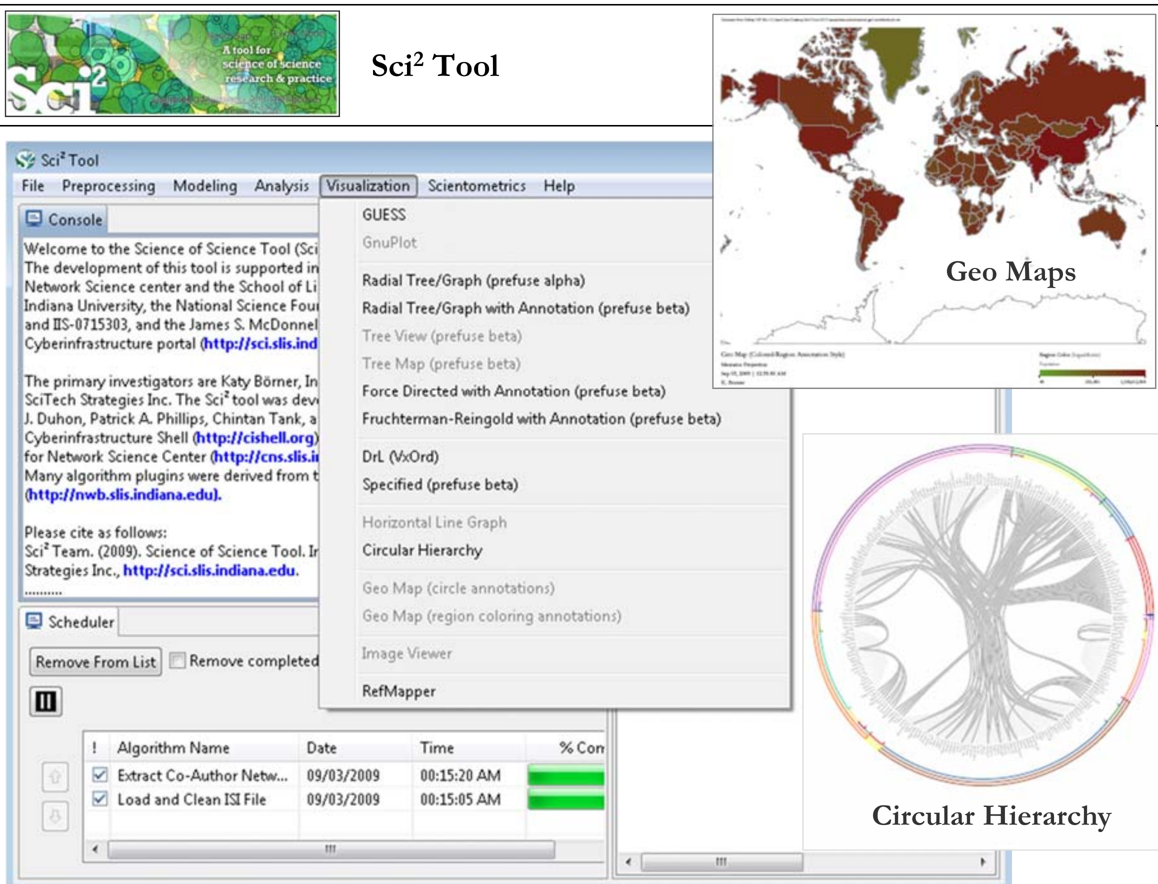
Horizontal Bar Graphs

Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Duhon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool. Proceedings of ISSI 2009: 12th International Conference on Scientometrics and Informetrics, Rio de Janeiro, Brazil, July 14-17 . Vol. 2, pp. 619-630.

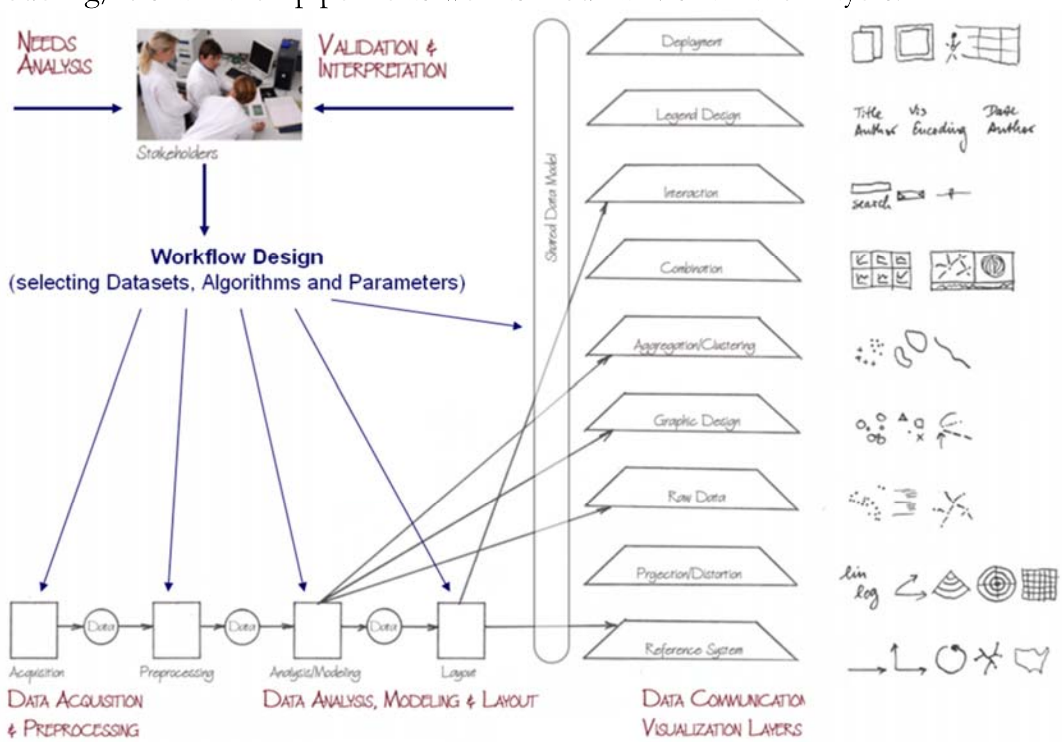
Timeline Visualization: Example



Sci² Tool

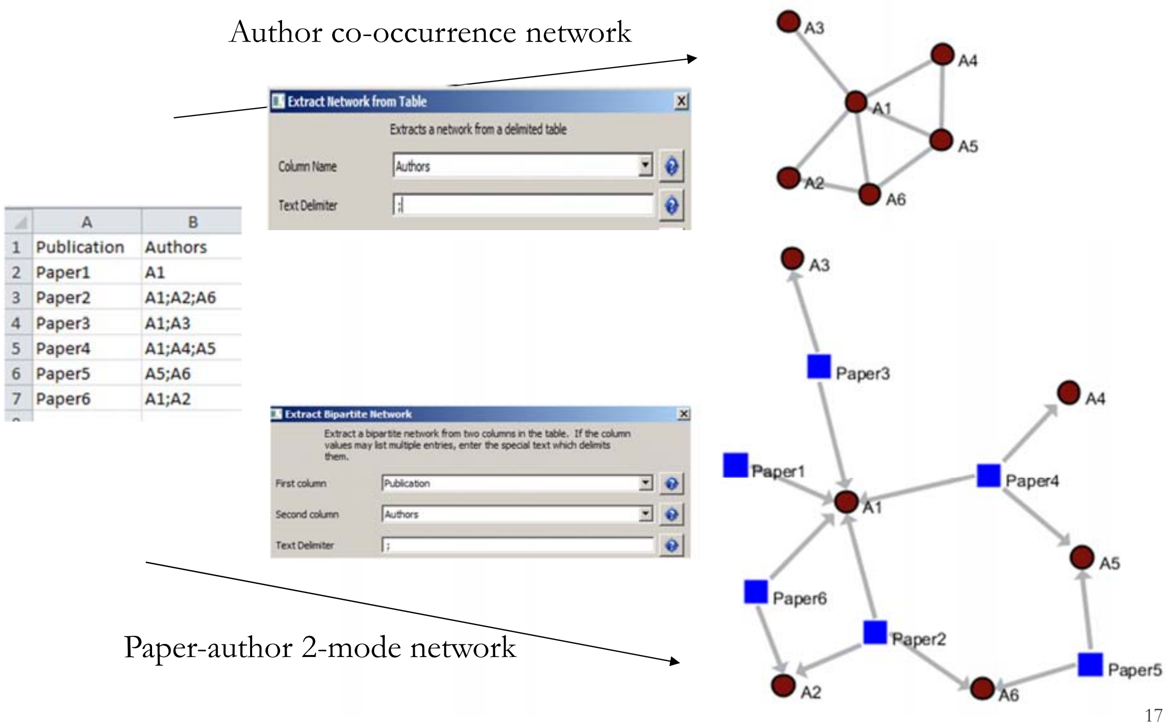


Needs-Driven Workflow Design using a modular data acquisition/analysis/ modeling/ visualization pipeline as well as modular visualization layers.

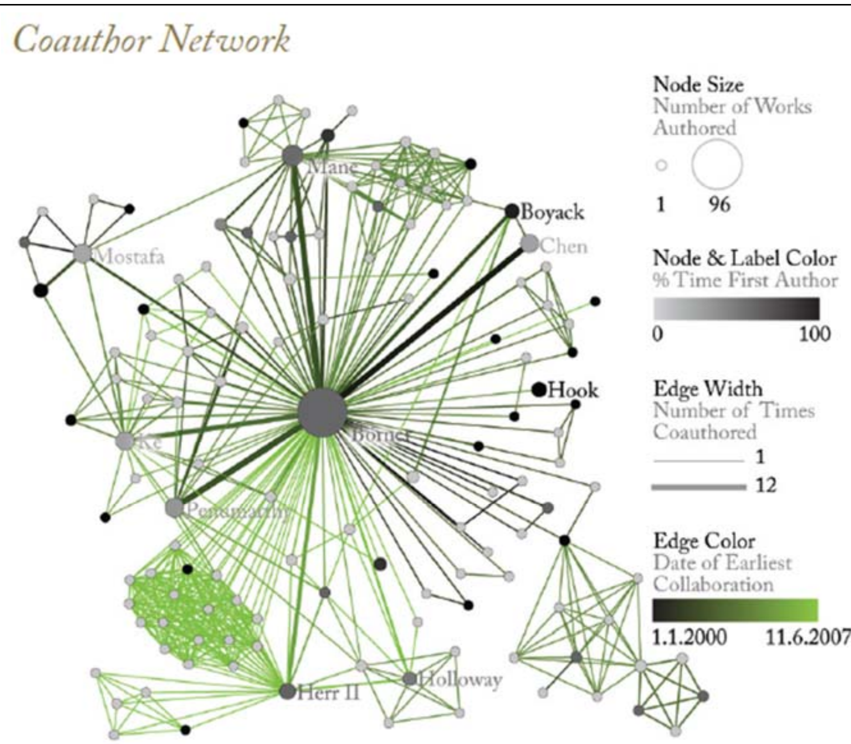


Börner, Katy (2010) Atlas of Science. MIT Press.

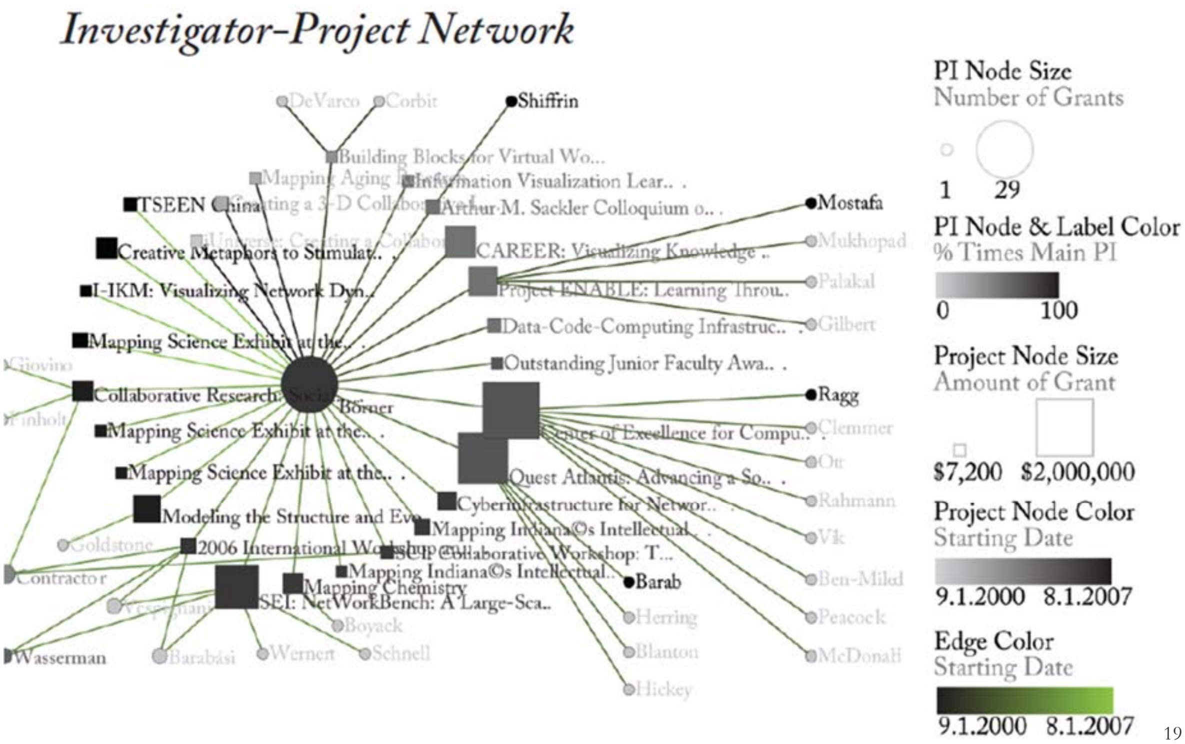
Network Extraction: Examples



Network Visualization: Example



Network Visualization: Example



Sci2 Tool Adoption

The Sci^2 Tool is used by NSF, NIH, USDA, and private foundations.

Upcoming Tutorial: Sci2: A Tool of Science of Science Research and Practice

Instructor:	Dr. Katy Börner, Indiana University
Time/Date:	8:30a-11:30a on Oct 17, 2011
Place:	Room II-555 in NSF's Stafford Place II Conference Center, 4121 Wilson Boulevard, Arlington, Virginia 22230, USA
Audience:	This tutorial is designed for researchers, practitioners, program staff from federal agencies interested to use advanced data mining algorithms and visualizations in their work and daily decision making.
Cost:	Free. Registration by Oct 10, 2011 required.
Register:	Please use http://www.surveymonkey.com/s/MVC8LWW to register by Oct 10, 2012. NSF will issue visitor badges.

First Iteration of Exhibit (2005): The Power of Maps

Four Early Maps of Our World
Versus
Six Early Maps of Science

The first exhibit iteration on *The Power of Maps* demonstrates how maps help us to understand, navigate, and manage both physical places and abstract knowledge spaces.

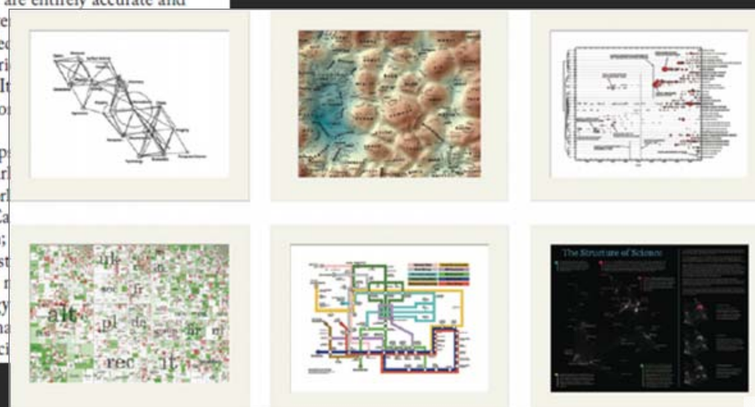
Early maps of our planet were certainly neither complete nor perfect, yet they proved invaluable for explorers. As keys to navigation, exploration, and communication, maps helped explorers find promising new lands while avoiding sea monsters.

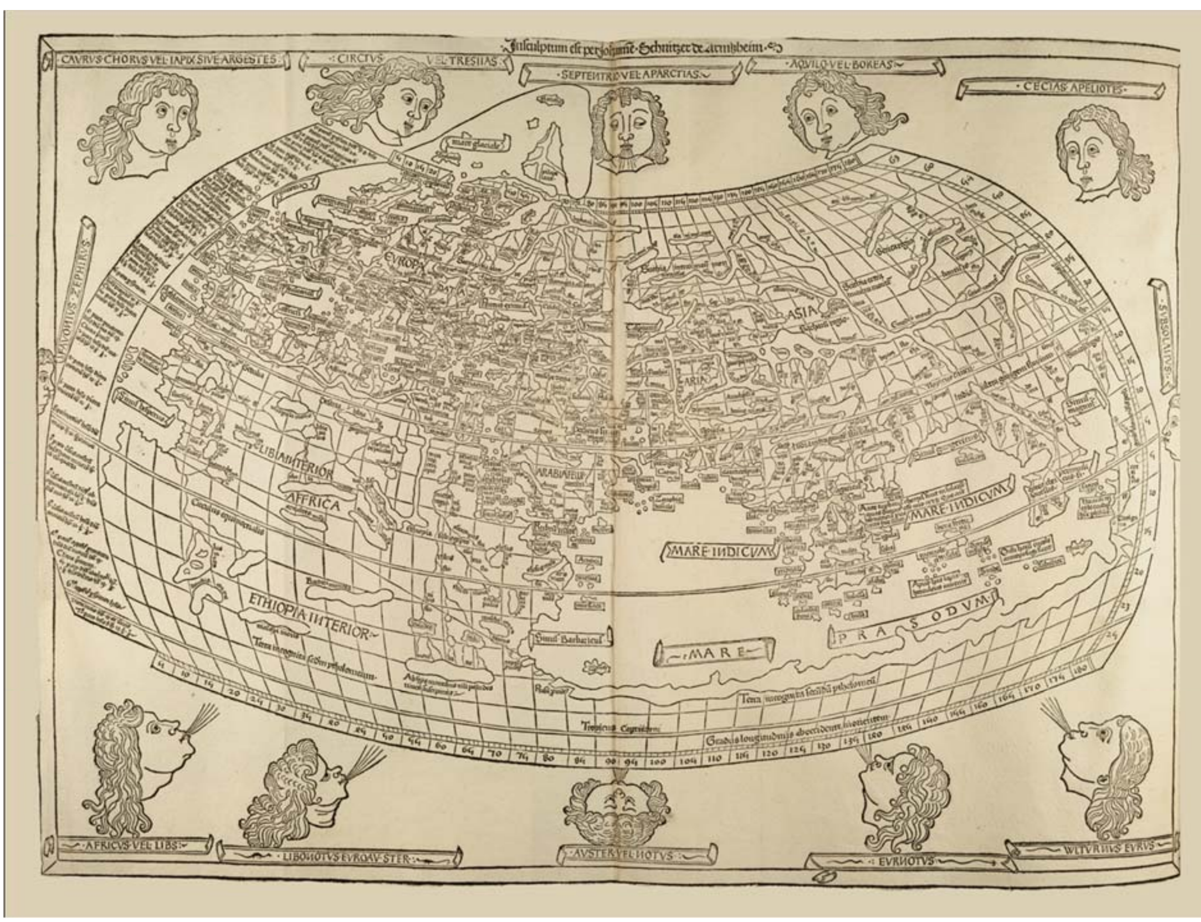
Maps of science today are based on limited knowledge and therefore imperfect. In order to generate comprehensive maps that are entirely accurate and reliable, we must first have proper coverage and inte… multidisciplinary, and multimedia scholarly knowle…

The first pictures of Earth from space were experi… mative of their perceptions of life and the cosmos. It… science will increase our appreciation and applicatio… serving as useful navigational tools.

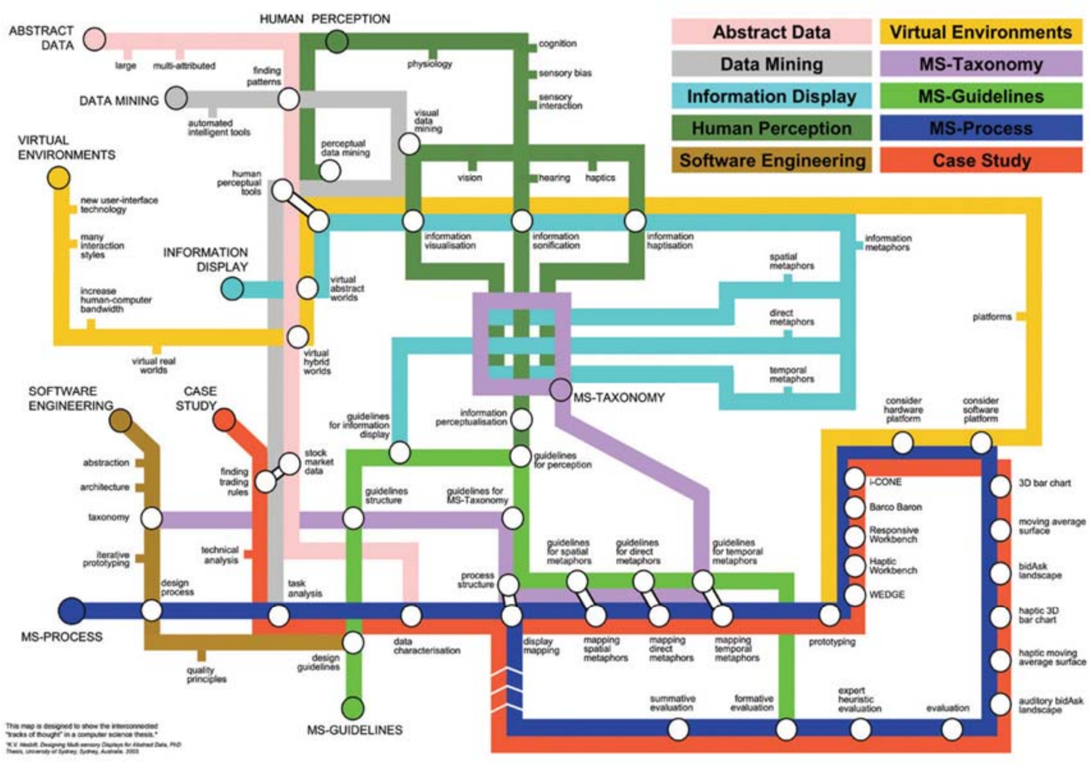
The Power of Maps features four cartographic map… earliest global maps of our world by Ptolemy, an ear… Johannes Janssonius, an early map of the whole wor… early statistical graph by Charles Joseph Minard. Ea… employs a different metaphor: a node-link diagram; … ing map rendered using geographic information syst… a crossmap; and a galaxy view. Which metaphor is … visual index of our collective science and technology…

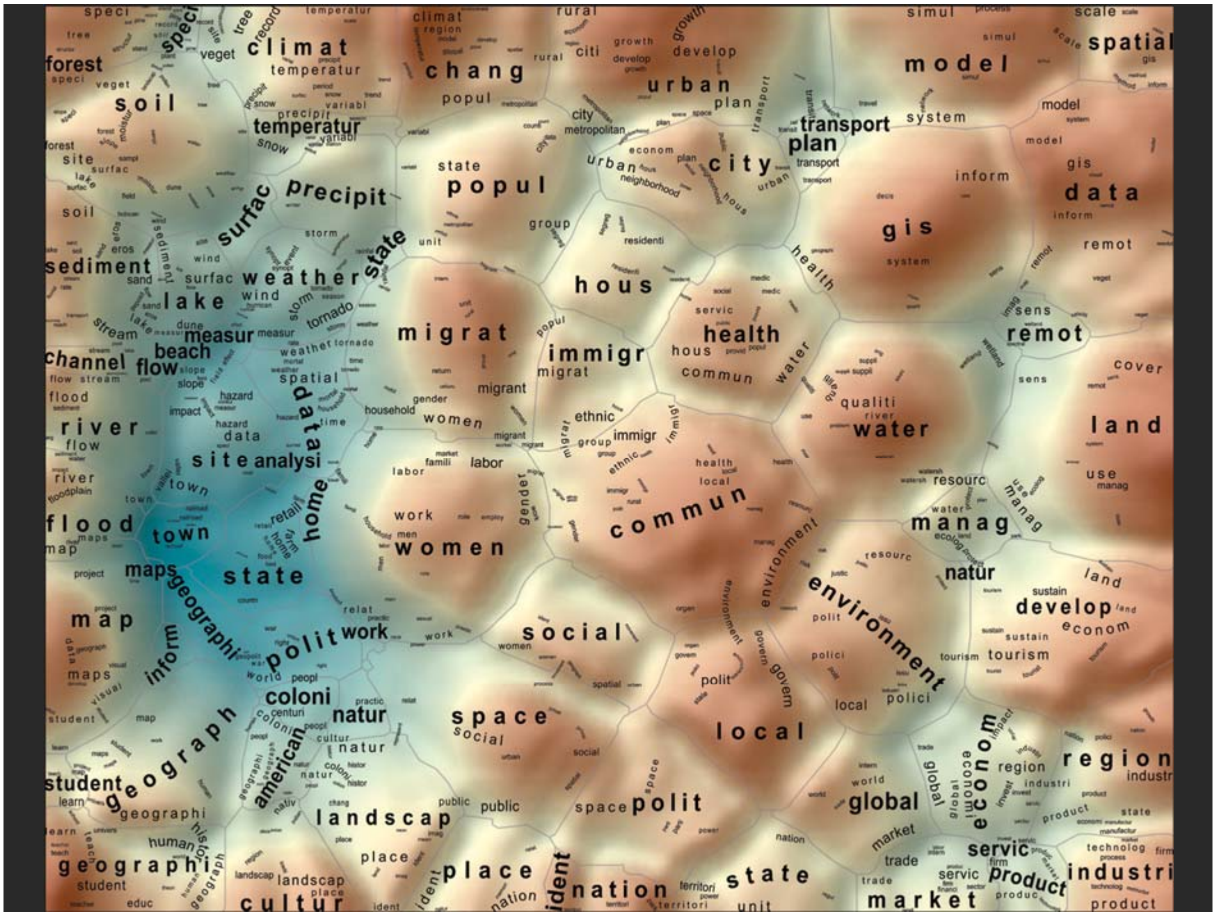
Note that the makers of the early cartographic ma… ing presses, while the makers of the first maps of sci…






A New Map of the WHOLE WORLD with the Trade winds
According to ye latest and most Exact Observations By H. Moll Geographer
In this Maps is inserted A View of ye General & Coasting Trade-Winds, Monsoons or ye Shifting Trade-winds Note that ye Arrows among ye Lines shew ye Course of those General & Coasting Winds. and ye Arrows in ye void Spaces shew ye Course of ye Shifting Trade-winds, and ye Abbreviation Sept. &c. Shew ye Times of ye Year when such Winds Blow.
The Signs of the Zodiack The First 6 are Northern, the other Southern Signs
Aries . March
Taurus . April
Gemini . May
Cancer . June
Leo . July
Virgo . August
Libra . September
Scorpio . October
Sagittaris . November
Capricornus . Decemb.
Aquarius . January
Pisces . February
NORTH POLE
EUROPE
NORTH AMERICA
AMERICA
SOUTH SEA
Western Ocean
AFRICA
ASIA
INDIAN OCEAN
New Holland
North Pole
South Pole
Printed for Thos. Bowles Print and Map seller next ye Chapter House in St. Pauls Church yard, and John Bowles Print and Map seller at the Black Horse in Cornhill London.


Abstract Data
Data Mining
Information Display
Human Perception
Software Engineering
Virtual Environments
MS-Taxonomy
MS-Guidelines
MS-Process
Case Study
ABSTRACT DATA
large
multi-attributed
finding patterns
HUMAN PERCEPTION
physiology
cognition
sensory bias
sensory interaction
DATA MINING
automated intelligent tools
visual data mining
perceptual data mining
human perceptual tools
VIRTUAL ENVIRONMENTS
new user-interface technology
many interaction styles
increase human-computer bandwidth
virtual real worlds
vision
hearing
haptics
information visualisation
information sonification
information haptisation
information metaphors
spatial metaphors
direct metaphors
temporal metaphors
platforms
INFORMATION DISPLAY
virtual abstract worlds
virtual hybrid worlds
MS-TAXONOMY
SOFTWARE ENGINEERING
CASE STUDY
abstraction
architecture
taxonomy
iterative prototyping
finding trading rules
stock market data
technical analysis
guidelines for information display
information perceptualisation
guidelines for perception
guidelines structure
guidelines for MS-Taxonomy
guidelines for spatial metaphors
guidelines for direct metaphors
guidelines for temporal metaphors
consider hardware platform
consider software platform
i-CONE
Barco Baron
Responsive Workbench
Haptic Workbench
WEDGE
3D bar chart
moving average surface
bidAsk landscape
haptic 3D bar chart
haptic moving average surface
auditory bidAsk landscape
design process
task analysis
process structure
MS-PROCESS
quality principles
design guidelines
data characterisation
display mapping
mapping spatial metaphors
mapping direct metaphors
mapping temporal metaphors
prototyping
summative evaluation
formative evaluation
expert heuristic evaluation
evaluation
MS-GUIDELINES
This map is designed to show the interconnected "tracks of thought" in a computer science thesis.*
*K.V. Nesbitt, Designing Multi-sensory Displays for Abstract Data, PhD Thesis, University of Sydney, Sydney, Australia. 2003

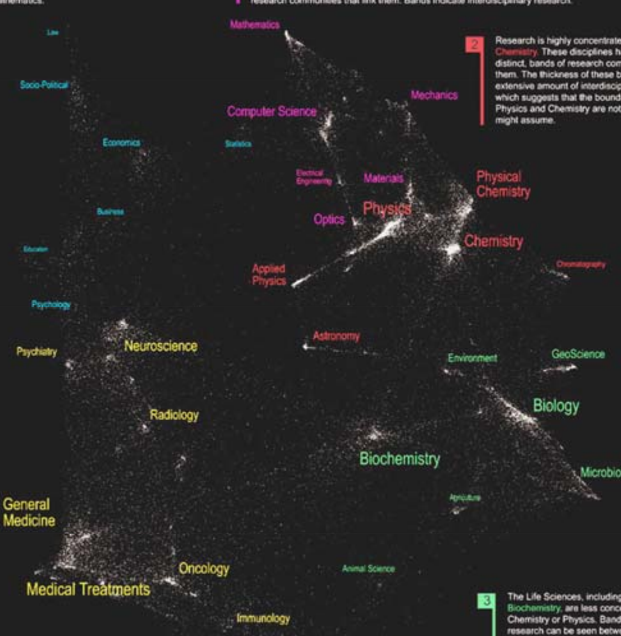


The Structure of Science

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

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

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



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

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

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.

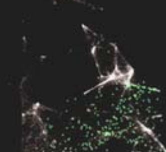
Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd from the citation 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.

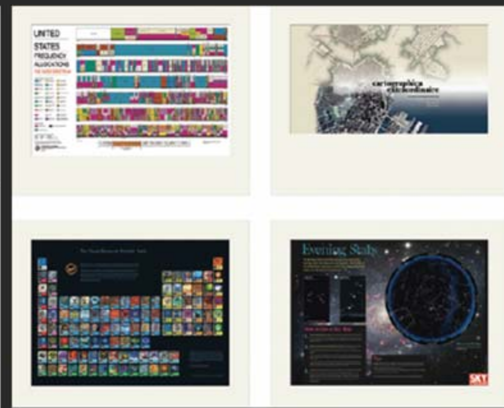
Second Iteration of Exhibit (2006): The Power of Reference Systems

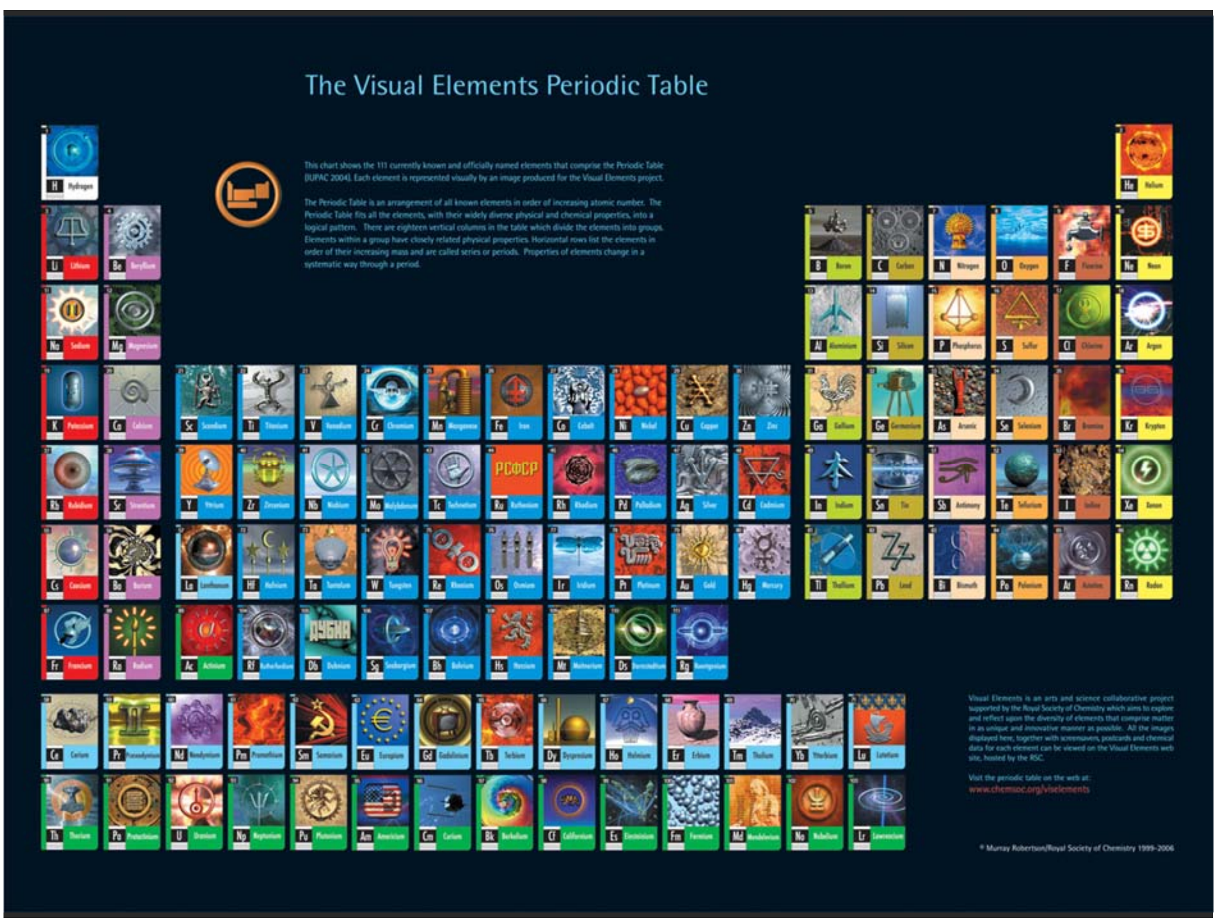
Four Existing Reference Systems
Versus
Six Potential Reference Systems

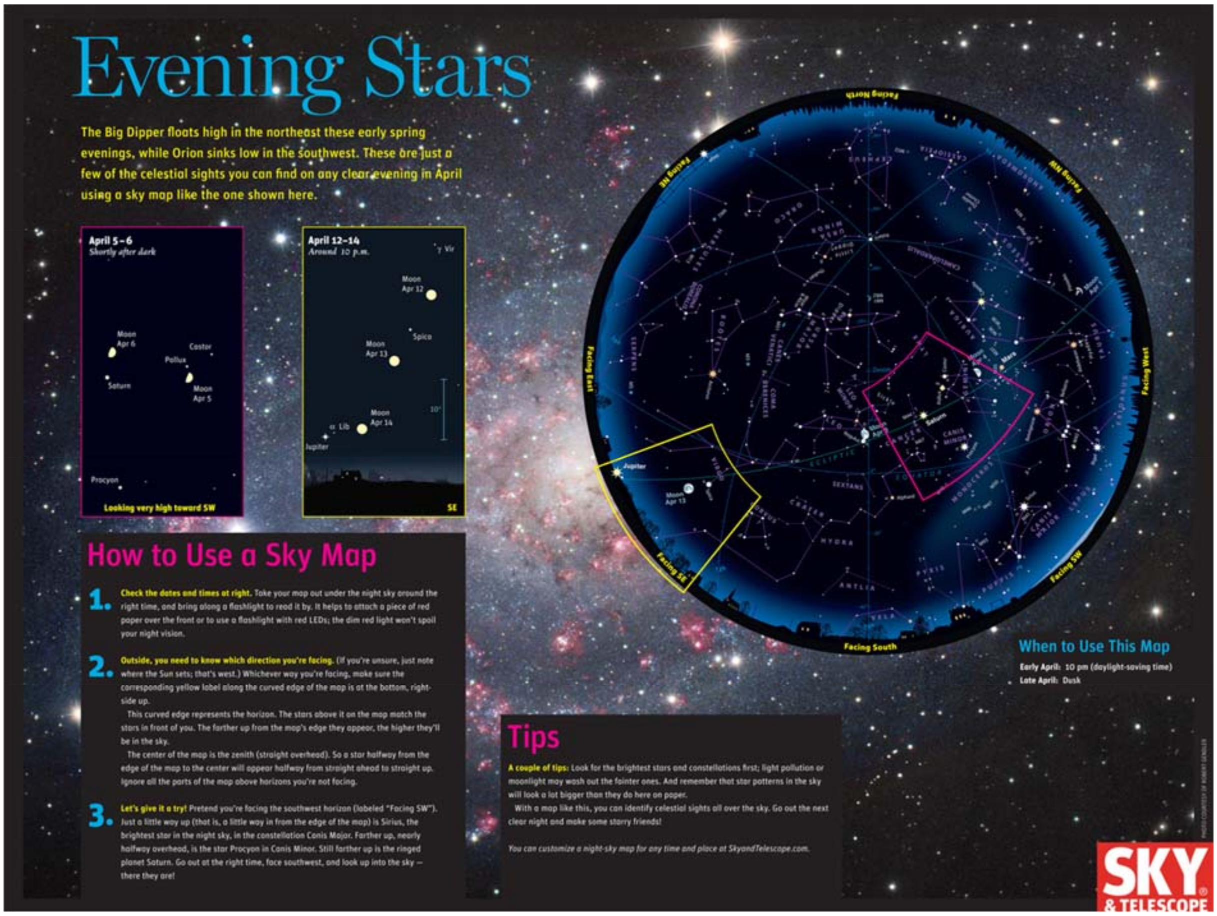
This iteration aims to inspire discussion about a common reference system for all existing scholarly knowledge. Throughout history, scientists have battled to agree on standardized reference systems for their respective fields of research. These standards are invaluable for indexing, storing, accessing, and managing scientific data efficiently.

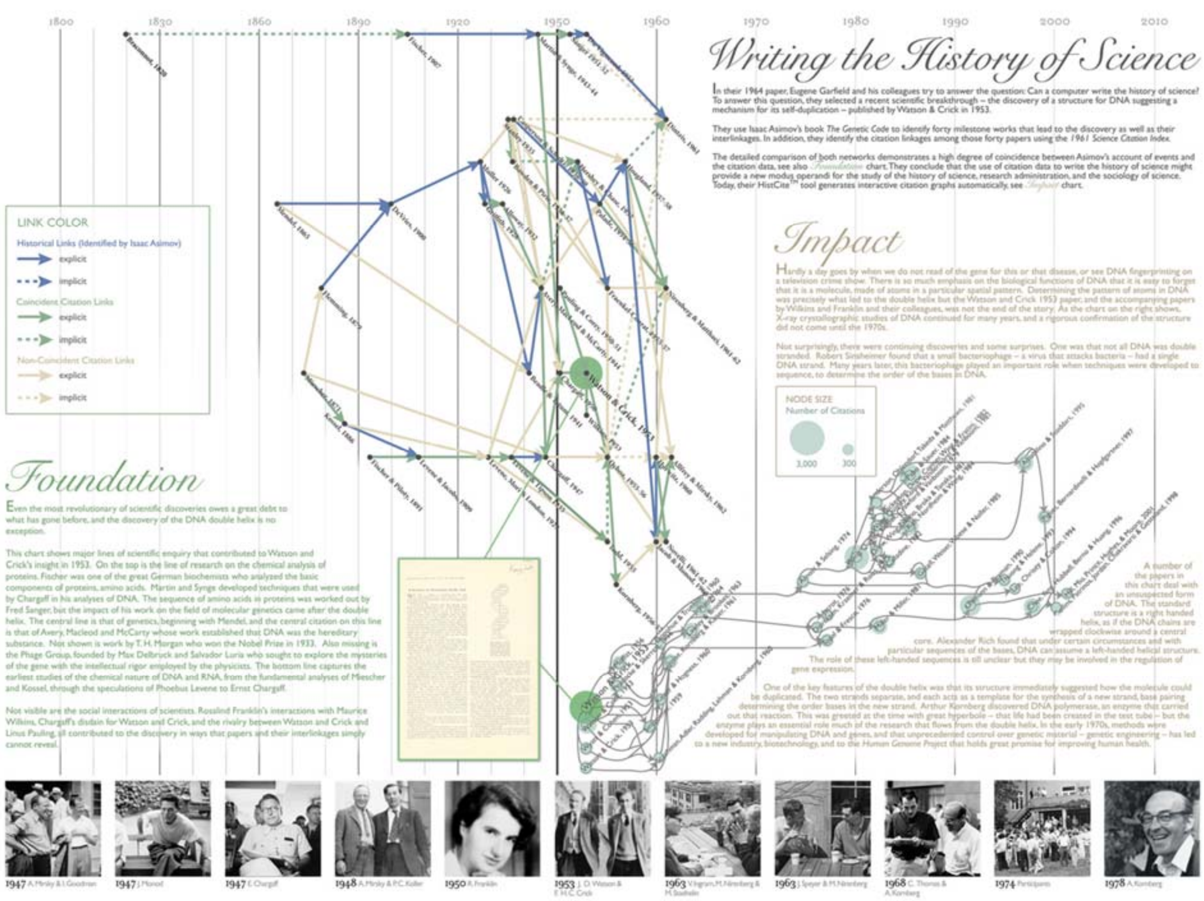
Results include the description of the electromag odic table of elements, geographic projections, and systems, shown here. Note that the geographic map from paper to geographic information systems (GIS for public use and consumption.

In comparison to these four existing systems are systems for scholarly knowledge. Each reference sy sional timeline and the geographic system to the se used to identify the location of an author, paper, pa tory or contribution.






Evening Stars
The Big Dipper floats high in the northeast these early spring evenings, while Orion sinks low in the southwest. These are just a few of the celestial sights you can find on any clear evening in April using a sky map like the one shown here.
April 5–6
Shortly after dark
Looking very high toward SW
April 12–14
Around 10 p.m.
SE
How to Use a Sky Map
1. Check the dates and times at right. Take your map out under the night sky around the right time, and bring along a flashlight to read it by. It helps to attach a piece of red paper over the front or to use a flashlight with red LEDs; the dim red light won't spoil your night vision.
2. Outside, you need to know which direction you're facing. (If you're unsure, just note where the Sun sets; that's west.) Whichever way you're facing, make sure the corresponding yellow label along the curved edge of the map is at the bottom, right-side up.
This curved edge represents the horizon. The stars above it on the map match the stars in front of you. The farther up from the map's edge they appear, the higher they'll be in the sky.
The center of the map is the zenith (straight overhead). So a star halfway from the edge of the map to the center will appear halfway from straight ahead to straight up. Ignore all the parts of the map above horizons you're not facing.
3. Let's give it a try! Pretend you're facing the southwest horizon (labeled "Facing SW"). Just a little way up (that is, a little way in from the edge of the map) is Sirius, the brightest star in the night sky, in the constellation Canis Major. Farther up, nearly halfway overhead, is the star Procyon in Canis Minor. Still farther up is the ringed planet Saturn. Go out at the right time, face southwest, and look up into the sky — there they are!
Tips
A couple of tips: Look for the brightest stars and constellations first; light pollution or moonlight may wash out the fainter ones. And remember that star patterns in the sky will look a lot bigger than they do here on paper.
With a map like this, you can identify celestial sights all over the sky. Go out the next clear night and make some starry friends!
You can customize a night-sky map for any time and place at SkyandTelescope.com.
Facing North
Facing NW
Facing West
Facing SW
Facing South
Facing SE
Facing East
When to Use This Map
Early April: 10 pm (daylight-saving time)
Late April: Dusk
SKY & TELESCOPE


1800
1830
1860
1890
1920
1950
1960
1970
1980
1990
2000
2010
Writing the History of Science
In their 1964 paper, Eugene Garfield and his colleagues try to answer the question: Can a computer write the history of science? To answer this question, they selected a recent scientific breakthrough – the discovery of a structure for DNA suggesting a mechanism for its self-duplication – published by Watson & Crick in 1953.
They use Isaac Asimov's book The Genetic Code to identify forty milestone works that lead to the discovery as well as their interlinkages. In addition, they identify the citation linkages among those forty papers using the 1961 Science Citation Index.
The detailed comparison of both networks demonstrates a high degree of coincidence between Asimov's account of events and the citation data, see also Foundation chart. They conclude that the use of citation data to write the history of science might provide a new modus operandi for the study of the history of science, research administration, and the sociology of science. Today, their HistCite™ tool generates interactive citation graphs automatically, see Impact chart.
LINK COLOR
Historical Links (Identified by Isaac Asimov)
explicit
implicit
Coincident Citation Links
explicit
implicit
Non-Coincident Citation Links
explicit
implicit
Impact
Hardly a day goes by when we do not read of the gene for this or that disease, or see DNA fingerprinting on a television crime show. There is so much emphasis on the biological functions of DNA that it is easy to forget that it is a molecule, made of atoms in a particular spatial pattern. Determining the pattern of atoms in DNA was precisely what led to the double helix but the Watson and Crick 1953 paper, and the accompanying papers by Wilkins and Franklin and their colleagues, was not the end of the story. As the chart on the right shows, X-ray crystallographic studies of DNA continued for many years, and a rigorous confirmation of the structure did not come until the 1970s.
Not surprisingly, there were continuing discoveries and some surprises. One was that not all DNA was double stranded. Robert Sinsheimer found that a small bacteriophage – a virus that attacks bacteria – had a single DNA strand. Many years later, this bacteriophage played an important role when techniques were developed to sequence, to determine the order of the bases in DNA.
NODE SIZE
Number of Citations
3,000
300
A number of the papers in this chart deal with an unsuspected form of DNA. The standard structure is a right handed helix, as if the DNA chains are wrapped clockwise around a central core. Alexander Rich found that under certain circumstances and with particular sequences of the bases, DNA can assume a left-handed helical structure. The role of these left-handed sequences is till unclear but they may be involved in the regulation of gene expression.
One of the key features of the double helix was that its structure immediately suggested how the molecule could be duplicated. The two strands separate, and each acts as a template for the synthesis of a new strand, base pairing determining the order bases in the new strand. Arthur Kornberg discovered DNA polymerase, an enzyme that carried out that reaction. This was greeted at the time with great hyperbole – that life had been created in the test tube – but the enzyme plays an essential role much of the research that flows from the double helix. In the early 1970s, methods were developed for manipulating DNA and genes, and that unprecedented control over genetic material – genetic engineering – has led to a new industry, biotechnology, and to the Human Genome Project that holds great promise for improving human health.
Foundation
Even the most revolutionary of scientific discoveries owes a great debt to what has gone before, and the discovery of the DNA double helix is no exception.
This chart shows major lines of scientific enquiry that contributed to Watson and Crick's insight in 1953. On the top is the line of research on the chemical analysis of proteins. Fischer was one of the great German biochemists who analyzed the basic components of proteins, amino acids. Martin and Synge developed techniques that were used by Chargaff in his analyses of DNA. The sequence of amino acids in proteins was worked out by Fred Sanger, but the impact of his work on the field of molecular genetics came after the double helix. The central line is that of genetics, beginning with Mendel, and the central citation on this line is that of Avery, Macleod and McCarty whose work established that DNA was the hereditary substance. Not shown is work by T. H. Morgan who won the Nobel Prize in 1933. Also missing is the Phage Group, founded by Max Delbruck and Salvador Luria who sought to explore the mysteries of the gene with the intellectual rigor employed by the physicists. The bottom line captures the earliest studies of the chemical nature of DNA and RNA, from the fundamental analyses of Miescher and Kossel, through the speculations of Phoebus Levene to Ernst Chargaff.
Not visible are the social interactions of scientists. Rosalind Franklin's interactions with Maurice Wilkins, Chargaff's disdain for Watson and Crick, and the rivalry between Watson and Crick and Linus Pauling, all contributed to the discovery in ways that papers and their interlinkages simply cannot reveal.
1947
1947
1947
1948
1950
1953
1963
1963
1968
1974
1978

US Patent Hierarchy

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 can get as deep as 15 levels. 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 by revealing any previous patents that might be similar in nature or provide a foundation for 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 one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1981 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical cosmetic implants, that mention Goretex as prior art.

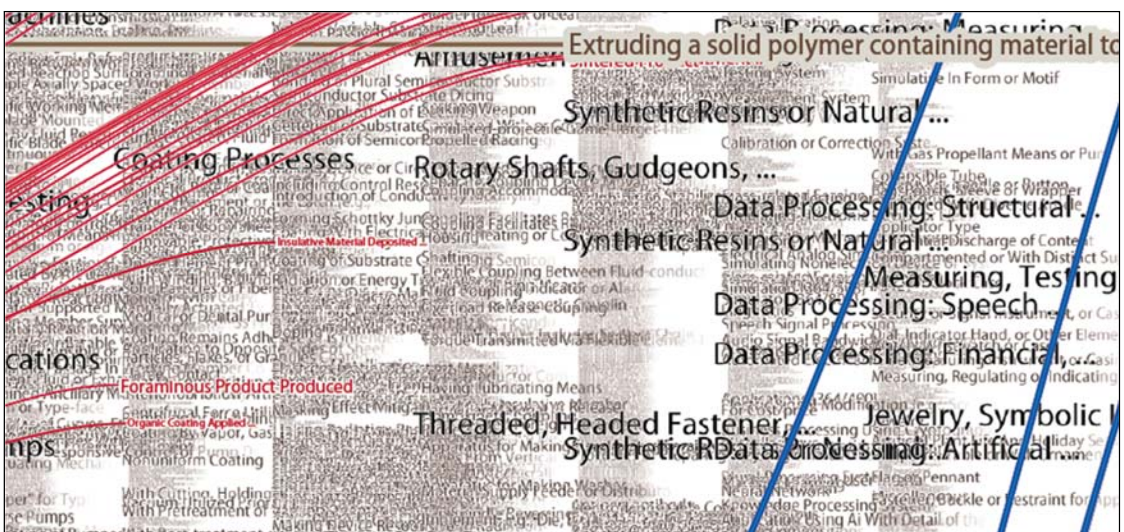
Prior Art

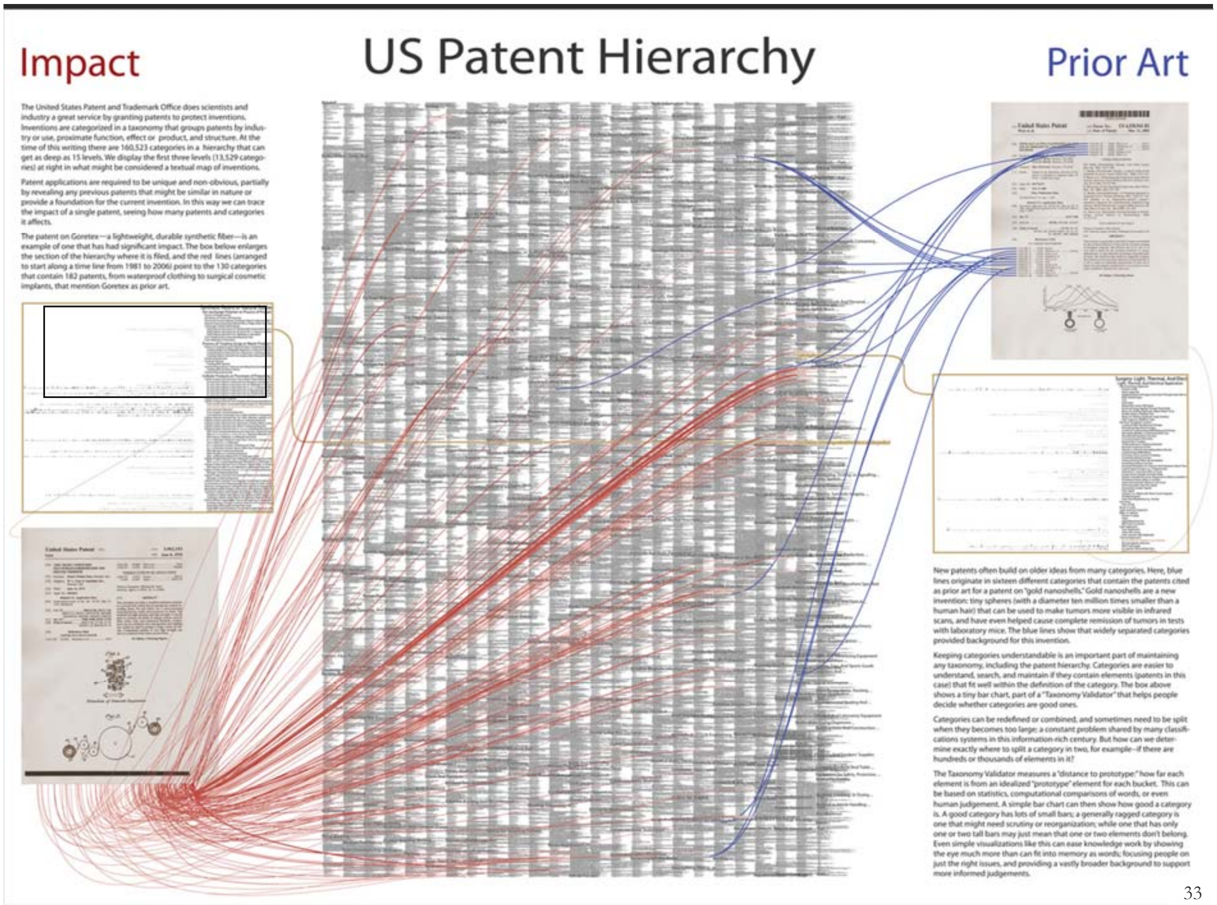
New patents often build on older ideas from many categories. Here, blue lines originate in sixteen different categories that contain the patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans, and have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.

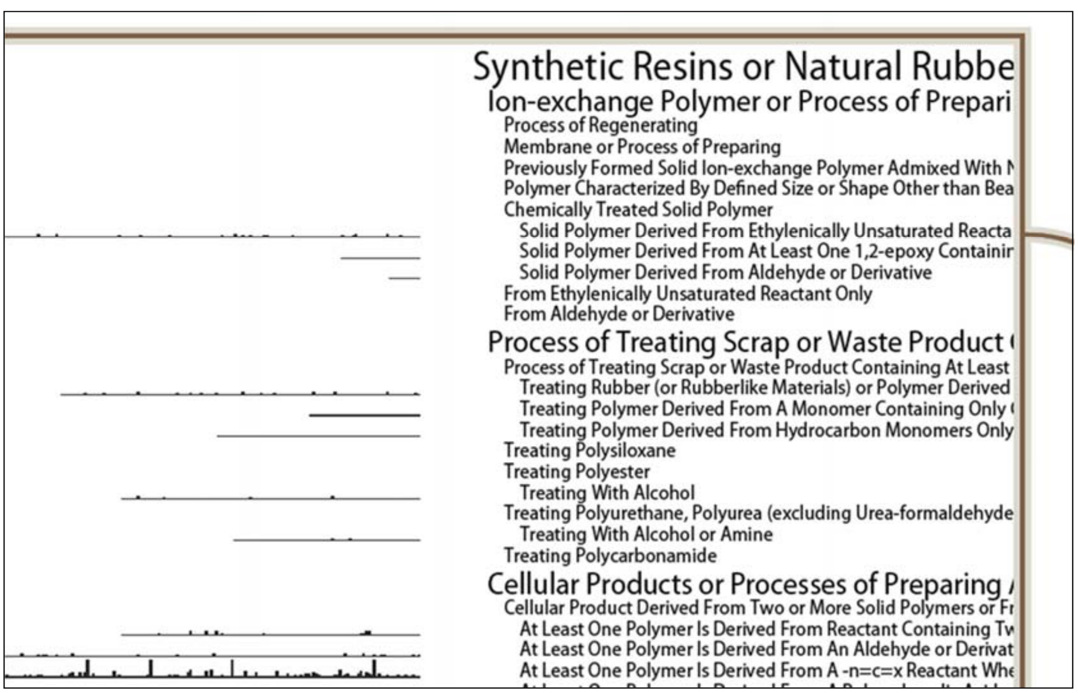
Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements (patents in this case) that fit well within the definition of the category. The box above shows a tiny bar chart, part of a "Taxonomy Validator" that helps people decide whether categories are good ones.

Categories can be redefined or combined, and sometimes need to be split when they becomes too large; a constant problem shared by many classifications systems in this information-rich century. But how can we determine exactly where to split a category in two, for example—if there are hundreds or thousands of elements in it?

The Taxonomy Validator measures a "distance to prototype:" how far each element is from an idealized "prototype" element for each bucket. This can be based on statistics, computational comparisons of words, or even human judgement. A simple bar chart can then show how good a category is. A good category has lots of small bars; a generally ragged category is one that might need scrutiny or reorganization; while one that has only one or two tall bars may just mean that one or two elements don't belong. Even simple visualizations like this can ease knowledge work by showing the eye much more than can fit into memory as words; focusing people on just the right issues, and providing a vastly broader background to support more informed judgements.




Impact
US Patent Hierarchy
Prior Art
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 can get as deep as 15 levels. 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 by revealing any previous patents that might be similar in nature or provide a foundation for 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 one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1981 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical cosmetic implants, that mention Goretex as prior art.
New patents often build on older ideas from many categories. Here, blue lines originate in sixteen different categories that contain the patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans, and have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for 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 (patents in this case) that fit well within the definition of the category. The box above shows a tiny bar chart, part of a "Taxonomy Validator" that helps people decide whether categories are good ones.
Categories can be redefined or combined, and sometimes need to be split when they becomes too large; a constant problem shared by many classifications systems in this information-rich century. But how can we determine exactly where to split a category in two, for example–if there are hundreds or thousands of elements in it?
The Taxonomy Validator measures a "distance to prototype:" how far each element is from an idealized "prototype" element for each bucket. This can be based on statistics, computational comparisons of words, or even human judgement. A simple bar chart can then show how good a category is. A good category has lots of small bars; a generally ragged category is one that might need scrutiny or reorganization; while one that has only one or two tall bars may just mean that one or two elements don't belong. Even simple visualizations like this can ease knowledge work by showing the eye much more than can fit into memory as words; focusing people on just the right issues, and providing a vastly broader background to support more informed judgements.
33


Synthetic Resins or Natural Rubbe
Ion-exchange Polymer or Process of Prepari
Process of Regenerating
Membrane or Process of Preparing
Previously Formed Solid Ion-exchange Polymer Admixed With N
Polymer Characterized By Defined Size or Shape Other than Bea
Chemically Treated Solid Polymer
Solid Polymer Derived From Ethylenically Unsaturated Reacta
Solid Polymer Derived From At Least One 1,2-epoxy Containir
Solid Polymer Derived From Aldehyde or Derivative
From Ethylenically Unsaturated Reactant Only
From Aldehyde or Derivative
Process of Treating Scrap or Waste Product
Process of Treating Scrap or Waste Product Containing At Least
Treating Rubber (or Rubberlike Materials) or Polymer Derived
Treating Polymer Derived From A Monomer Containing Only
Treating Polymer Derived From Hydrocarbon Monomers Only
Treating Polysiloxane
Treating Polyester
Treating With Alcohol
Treating Polyurethane, Polyurea (excluding Urea-formaldehyde
Treating With Alcohol or Amine
Treating Polycarbonamide
Cellular Products or Processes of Preparing
Cellular Product Derived From Two or More Solid Polymers or Fr
At Least One Polymer Is Derived From Reactant Containing Tw
At Least One Polymer Is Derived From An Aldehyde or Derivat
At Least One Polymer Is Derived From A -n=c=x Reactant Whe

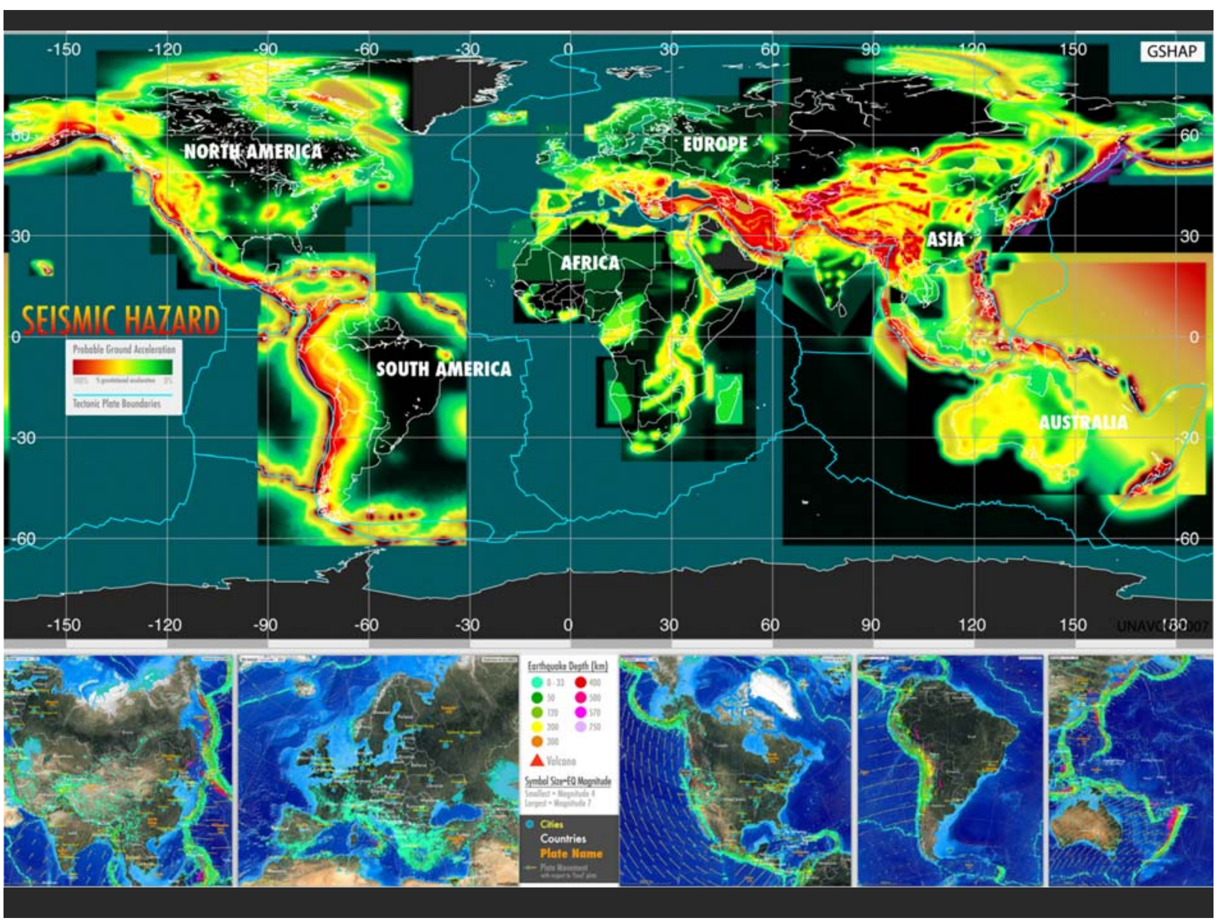
Third Iteration of Exhibit (2007): The Power of Forecasts

Four Existing Forecasts
Versus
Six Science Forecasts

The third iteration of the exhibit compares and contrasts seismic hazard, economic, resource depletion, and epidemic forecast maps with maps forecasting the structure and evolution of science.

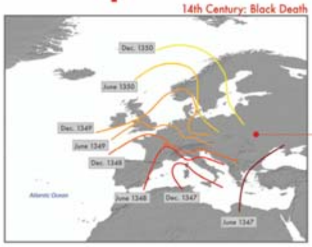
Real-time weather forecasts are served by the National Oceanic and Atmospheric Administration (NOAA) or the National Aeronautics and Space Administration (NASA). Computational models of the movements of tectonic plates help reduce losses due to earthqual tsunamis. Epidemic models make us understand and how actions far away affect us right here. Ec catastrophic and sustainable futures for mankind.

Daily science and technology forecasts would s of top experts/institutions/countries, major activit frontiers, augmenting our knowledge and decisio available on TV, in the press, and online?



Impact OF Air Travel ON Global Spread OF Infectious Diseases

14th Century: Black Death

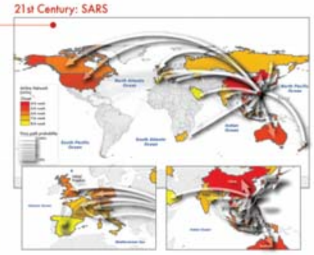


Epidemic spreading pattern changed dramatically after the development of modern transportation systems.

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.

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).

21st Century: SARS



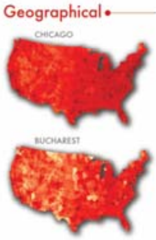
Forecasts OF THE Next Pandemic Influenza

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.

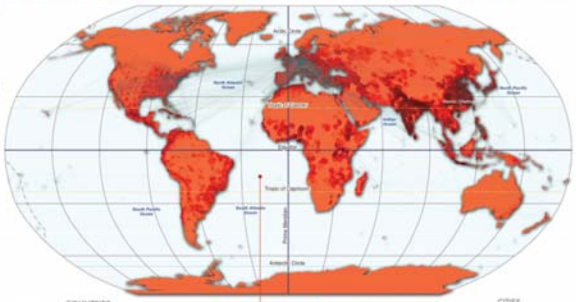
Geographical

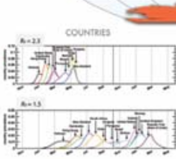


Numerical simulations provide results for the temporal and geographic evolution of the pandemic influenza in 3,100 urban areas located in 220 different countries. The model allows to study different spreading scenarios, characterized by different initial outbreak conditions, both geographical and seasonal.

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.

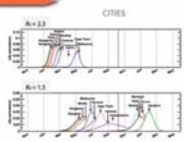
The US maps focus on the situation in the US after one year, and show the effect of changes in the original scenario analyzed. Different color coding is used for the sake of visualization.





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.



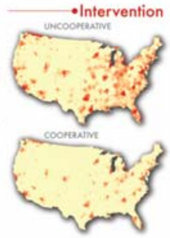
The model inlcudes the worldwide air transportation network (source: IATA) composed of 3,100 airports in 220 countries and E=17,182 direct connections, each of them associated to the corresponding passenger flow. This dataset accounts for 99% of the worldwide traffic and is complemented by the census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

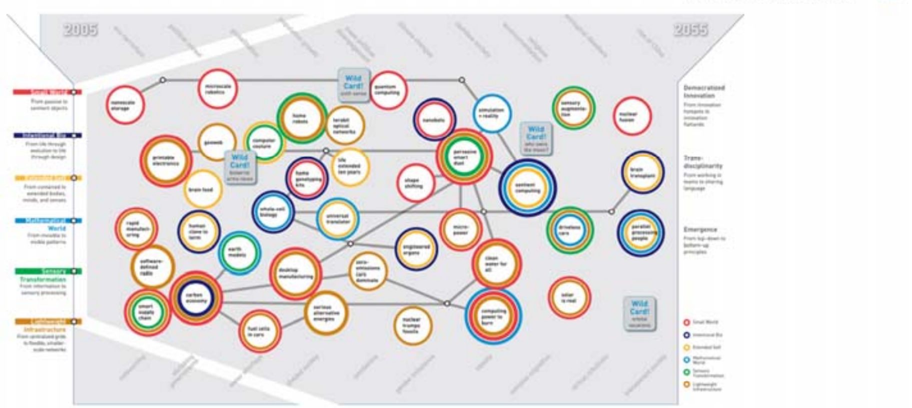
Intervention



INSTITUTE FOR THE FUTURE

Science & Technology Outlook: 2005–2055

Technology Horizons Program
Institute for the Future
124 University Avenue, 2nd Floor, Palo Alto, CA 94301
t 650.854.6322 f 650.854.7850 www.iftf.org



A map is a tool for navigating an unknown terrain. In the case of this map, Science & Technology Outlook: 2005–2055, the terrain we're navigating is the uncharted territory of science and technology (S&T) in the next 50 years. However, the map of the future is not a tool for prediction or, for that matter, the product of predictions. Nor is it comparable to modern navigation techniques in which we rely on a shrinking number of strong signals, like GPS coordinates, to show the right path. Rather, it's more akin to classical low-tech navigational techniques with their reliance on an array of weak signals such as wind direction, the look and feel of the water, and the shape of cloud formations. Taken together, these signals often prove more useful for navigation than high-tech methods because, in addition to aiding travelers in selecting the "right" path, the signals contextualize information and reveal interdependencies and connections between seemingly unrelated events, thus enriching our understanding of the landscape. That's precisely the intention of this map of the future of S&T—to give the reader a deeper contextual understanding of the landscape and to point to the intricacies and interdependencies between trends.

While developing the map, the Institute for the Future (IFTF) team listened for and connected a variety of weak signals, including those generated during interviews and workshop conversations involving more than 100 eminent U.K. and U.S. experts in S&T—academicians, policymakers, journalists, and corporate researchers. The IFTF team also compiled a database of outlooks on developments that are likely to impact the full range of S&T disciplines and practice areas over the next 50 years. We also relied on IFTF's 40 years of experience in forecasting S&T developments to create the map and an accompanying set of S&T Perspectives that discuss issues emerging on the S&T horizon and are important for organizations, policymakers, and society-at-large to understand.

On this map, six themes are woven together across the 50-year horizon, often resulting in important breakthroughs. These are supported by key technologies, innovations, and discoveries. In addition to the six themes, three meta-themes—democratized innovation, transdisciplinarity, and emergence—will overlay the future S&T landscape influencing how we think about, learn about, and practice science. Finally, S&T trends won't operate in a vacuum. Wider social, demographic, political, economic, and environmental trends will both influence S&T trends and will be influenced by them. Some of these wider trends surround the map to remind us of the larger picture.

MAP THEMES

Small World

After 20 years of basic research and development at the 100-nanometer scale, the importance of nanotechnology as a source of innovations and new capabilities in everything from materials science to medicine is already well-understood. Three trends, however, will define how nanotechnology will unfold, and what impacts it will have. First, nanotechnology is not a single field with a coherent intellectual program; it's an opportunistic hybrid, shaped by a combination of fundamental research questions, promising technical applications, and venture and state capital. Second, nanotechnology is moving away from the original vision of small-scale mechanical engineering—in which assemblers build mechanical systems from individual atoms—toward one in which molecular biology and biochemistry contribute essential tools (such as proteins that build nanowires). Finally, nanotechnology will also serve as a model for transdisciplinary science. It will support both fundamental research and commercially oriented innovation; and it will be conducted not within the boundaries of conventional academic or corporate research departments, but in institutional and social milieux that emphasize heterogeneity.

Intentional Biology

For 3.6 billion years, evolution has governed biology on this planet. But today, Mother Nature has a collaborator. Inexpensive tools to read and rewrite the genetic code of life will bootstrap our ability to manipulate biology from the bottom up. We'll not only genetically re-engineer existing life but actually create new life forms with purpose. Still, we will not be blind to what nature has to teach us. Evolution's elegant engineering at the smallest scales will be a rich source of inspiration as we build the bio-nanotechnology of the next 50 years.

Extended Self

In the next 50 years, we will be faced with broad opportunities to remake our minds and bodies in profoundly different ways. Advances in biotechnology, brain science, information technology, and robotics will result in an array of methods to dramatically alter, enhance, and extend the mental and physical hand that nature has dealt us. Wielding these tools on ourselves, humans will begin to define a variety of different "transhumanist" paths—that is, ways of being and living that extend beyond what we today consider natural for our species. In the very long term, following these paths could someday lead to an evolutionary leap for humanity.

Mathematical World

The ability to process, manipulate, and ultimately understand patterns in enormous amounts of data will allow decoding of previously mysterious processes in everything from biological to social systems. Scientists are learning that at the core of many biological phenomena—reproduction, growth, repair, and others—are computational processes that can be decoded and simulated. Using techniques of combinatorial science to uncover such patterns—whether these are physical, biological, or social—will likely occupy an increasing share of computing cycles in the next 50 years. Such massive computation will also make simulation widespread. Computer simulation will be used not only to help make decisions about large complex scientific and social problems but also to help individuals make better choices in their daily lives.

Sensory Transformation

In the next ten years, physical objects, places, and even human beings themselves will increasingly become embedded with computational devices that can sense, understand, and act upon their environment. They will be able to react to contextual clues about the physical, social, and even emotional state of people and things in their surroundings. As a result, increasing demands will be placed on our visual, auditory, and other sensory abilities. Information previously encoded as text and numbers will be displayed in richer sensory formats—as graphics, pictures, patterns, sounds, smells, and tactile experiences. This enriched sensory environment will coincide with major breakthroughs in our understanding of the brain—in how we process sensory information and connect various sensory functions. Humans will become much more sophisticated in their ability to understand, create, and manage sensory information and ability to perform such tasks will become keys to success.

Lightweight Infrastructure

A confluence of new materials and distributed intelligence is pointing the way toward a new kind of infrastructure that will dramatically reshape the economics of moving people, goods, energy, and information. From the molecular level to the macro-economic level, these new infrastructure designs will emphasize smaller, smarter, more independent components. These components will be organized into more efficient, more flexible, and more secure ways than the capital-intensive networks of the 20th century. These lightweight infrastructures have the potential to boost emerging economies, improve social connectivity, mitigate the environmental impacts of rapid global urbanization, and offer new future paths in energy.

META-THEMES

Democratized Innovation

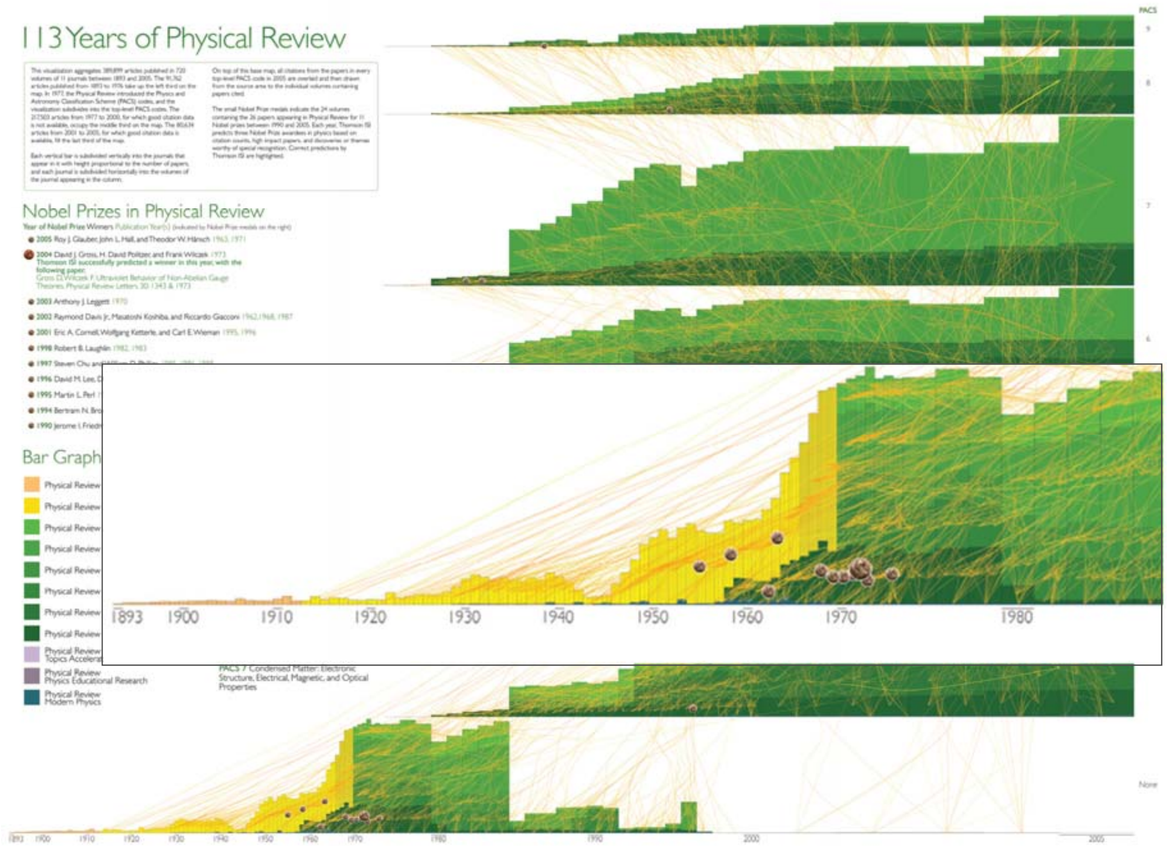
Before the 20th century, many of the greatest scientific discoveries and technical inventions were made by amateur scientists and independent inventors. In the last 100 years, a professional class of scientists and engineers, supported by universities, industry, and the state, pushed amateurs aside as a creative force. At the national scale, the capital-intensive character of scientific research made world-class research the property of prosperous advanced nations. In the new century, a number of trends and technologies will lower the barriers to participation in science and technology again, both for individuals and for emerging countries. The result with be a renaissance of the serious amateurs, the growth of new scientific and technical centres of excellence in developing countries, and a more global distribution of world-class scientists and technologists.

Transdisciplinarity

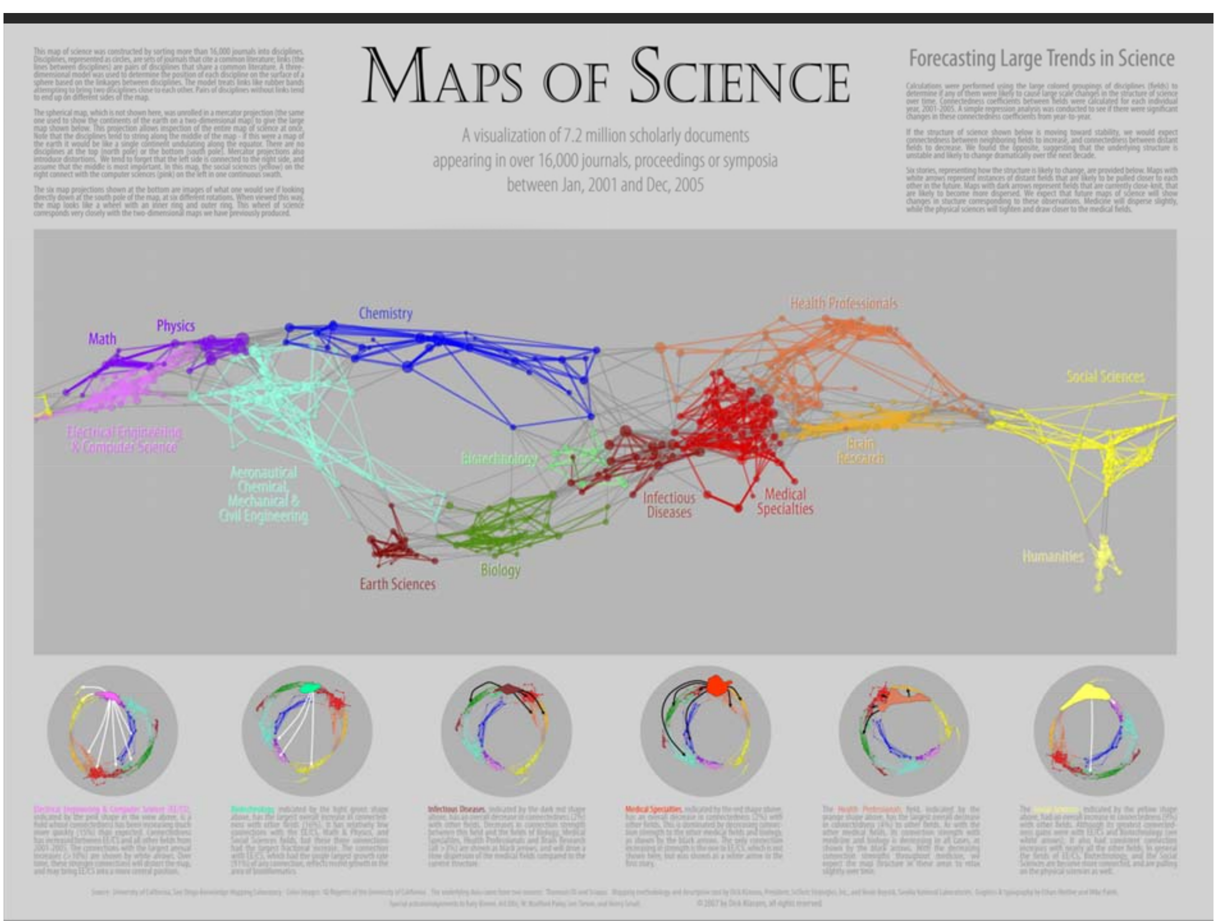
In the last two centuries, natural philosophy and natural history fractured into the now-familiar disciplines of physics, chemistry, biology, and so on. The sciences evolved into their current form in response to intellectual and professional opportunities, philanthropic priorities, and economic and state needs. Through most of the 20th century, the growth of the sciences, and academic and career pressures, encouraged ever-greater specialization. In the coming decades, transdisciplinary research will become an imperative. According to Howard Rheingold, a prominent forecaster and author, "transdisciplinarity goes beyond bringing together researchers from different disciplines to work in multidisciplinary teams. It means educating researchers who can speak languages of multiple disciplines—biologists who have understanding of mathematics, mathematicians who understand biology."

Emergence

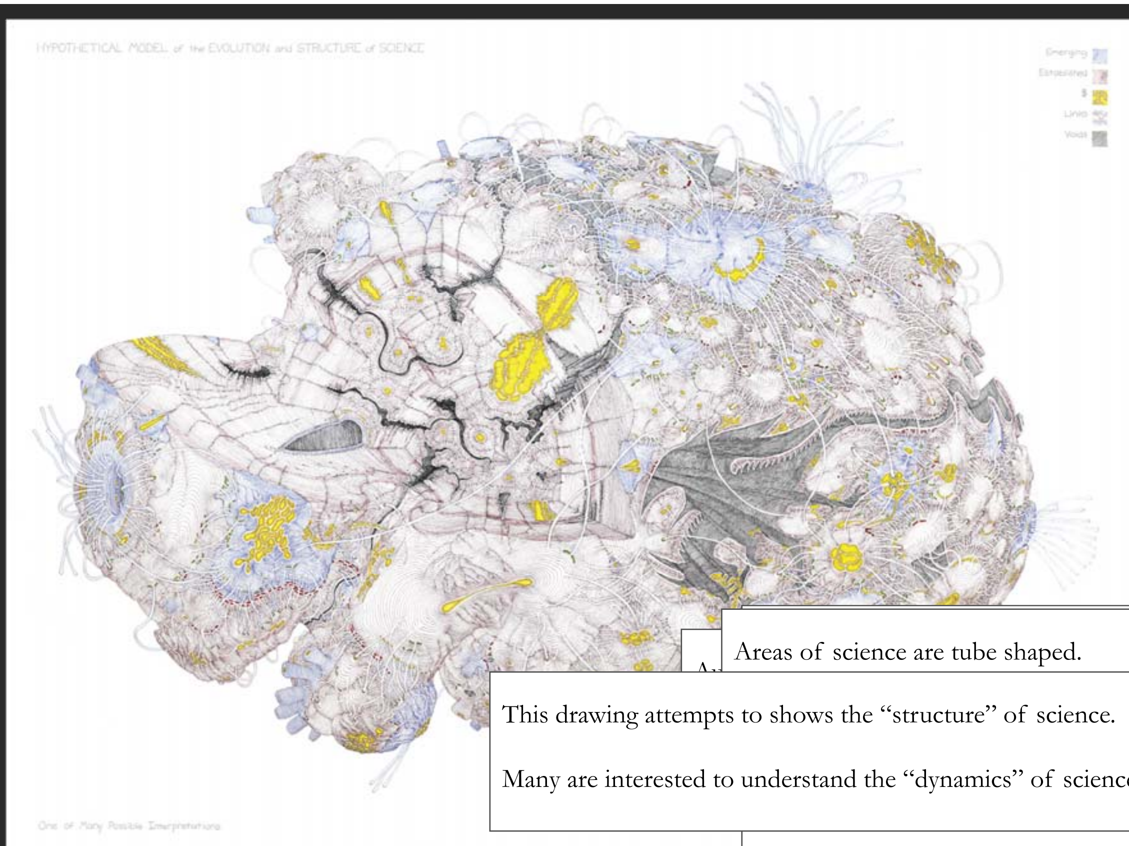
The phenomenon of self-organizing swarms that generate complex behavior by following simple rules—will likely become an important research area, and an important model for understanding how the natural world works and how artificial worlds can be designed. Emergent phenomena have been observed across a variety of natural phenomena, from physics to biology to sociology. The concept has broad appeal due to the diversity of fields and problems to which it can be applied. It is proving useful for making sense of a very wide range of phenomena. Meanwhile, emergence can be modeled using relatively simple computational tools, although these models often require substantial processing power. More generally, it is a richly suggestive as a way of thinking about designing complex, robust technological systems. Finally, emergence is an accessible and vivid a metaphor for understanding nature. Just as classical physics profited from popular treatments of Newtonian mechanics, so too will scientific study and technical reproductions of emergent phenomena likely draw benefits from the popularization of its underlying concepts.



113 Years of Physical Review
Nobel Prizes in Physical Review
Bar Graph
PACS
1893
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2005
None



Maps of Science
A visualization of 7.2 million scholarly documents appearing in over 16,000 journals, proceedings or symposia between Jan, 2001 and Dec, 2005
Forecasting Large Trends in Science
Math
Physics
Chemistry
Health Professionals
Social Sciences
Electrical Engineering & Computer Science
Aeronautical Chemical, Mechanical & Civil Engineering
Biotechnology
Brain Research
Infectious Diseases
Medical Specialties
Earth Sciences
Biology
Humanities



Areas of science are tube shaped.

This drawing attempts to shows the "structure" of science.

Many are interested to understand the "dynamics" of science.

Hypothetical Model of the Evolution of Science - Daniel Zeller - 2007

Additional Elements of the Exhibit

Certainly science maps and data graphs work to engage viewers intellectually—but can they also capture the imagination, as did the early maps of the world? Is it possible to involve viewers in a more dynamic way that heightens both their awareness and appreciation of data, information, and knowledge? What can be learned from theater, movies, and art exhibits—as well as science displays—to improve the ability of science maps to entertain while educating, to inspire while being true to facts, and
ers to engage in scienc

Additional exhibit e
and interact with scien
exceptional high data
and a map of today's s
drives a touch panel di
the touch panel display
on any given topic are
given geographic locat

The *Hands-On Scien*
stand science from abo
color drawings. Childr
placing images of maj
appropriate places on t
of various countries fo
patents. **Shape of Scie**
The Video of the Exhi
Public Library (NYPI
NYPL officials, who

Illuminated Diagram Display

W. Bradford Paley, Kevin W. Boyack, Richard Kalvans, and Katy Börner (2007) Mapping, Illuminating, and Interacting with Science. SIGGRAPH 2007.

Large-scale, high resolution prints illuminated via projector or screen.

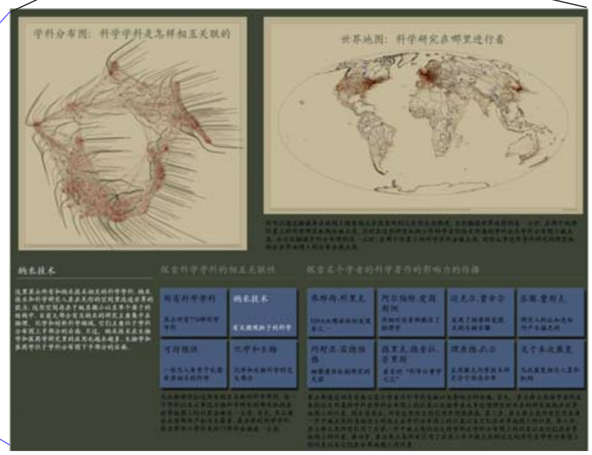
Interactive touch panel.

Questions:

- Who is doing research on what topic and where?
- What is the 'footprint' of interdisciplinary research fields?
- What impact have scientists?

Contributions:

- Interactive, high resolution interface to access and make sense of data about scholarly activity.



Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months 12 coaches, 300 m long
Opening was on April 23rd, 2009 by German Chancellor Merkel
http://www.expedition-zukunft.de

Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org/

The Power of Maps (2005)

Science Maps for Economic Decision Makers (2008)

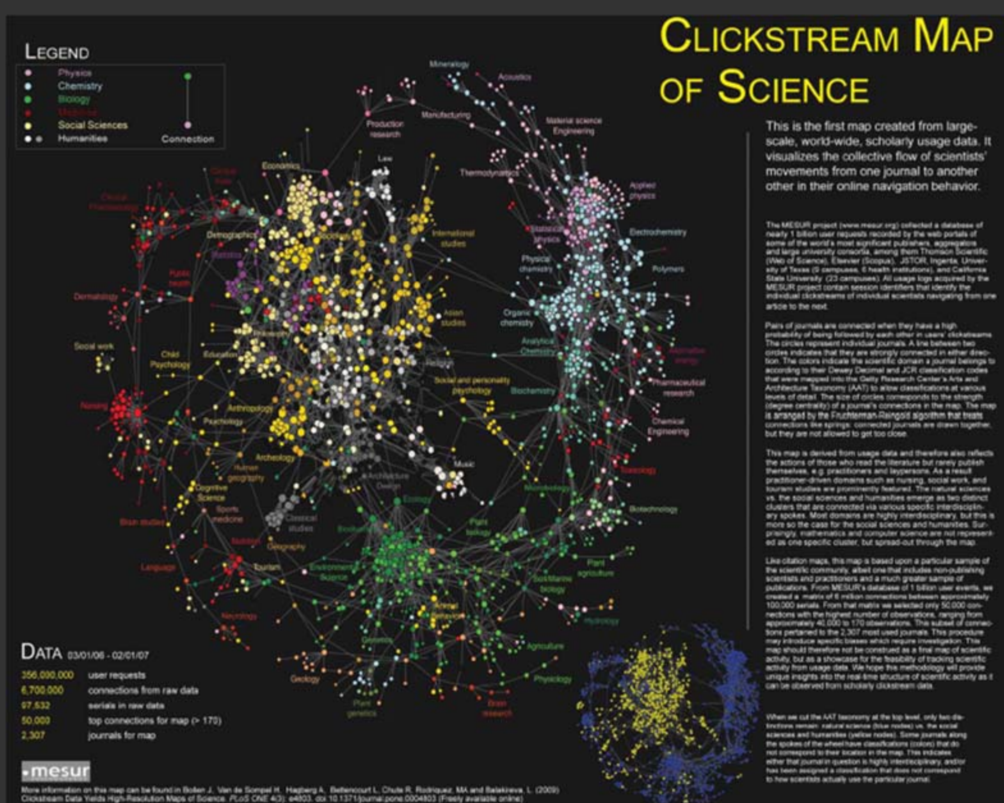
The Power of Reference Systems (2006)

Science Maps for Science Policy Makers (2009)

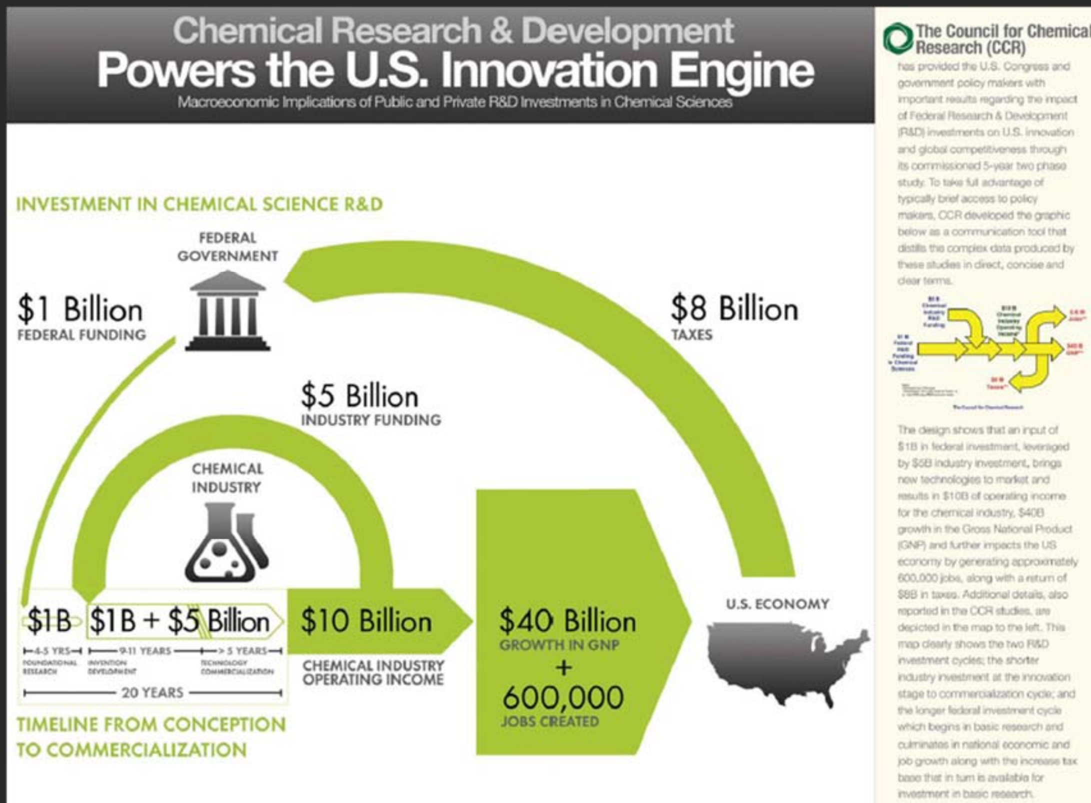
The Power of Forecasts (2007)

Science Maps for Scholars (2010)
Science Maps as Visual Interfaces to Digital Libraries (2011)
Science Maps for Kids (2012)
Science Forecasts (2013)
How to Lie with Science Maps (2014)

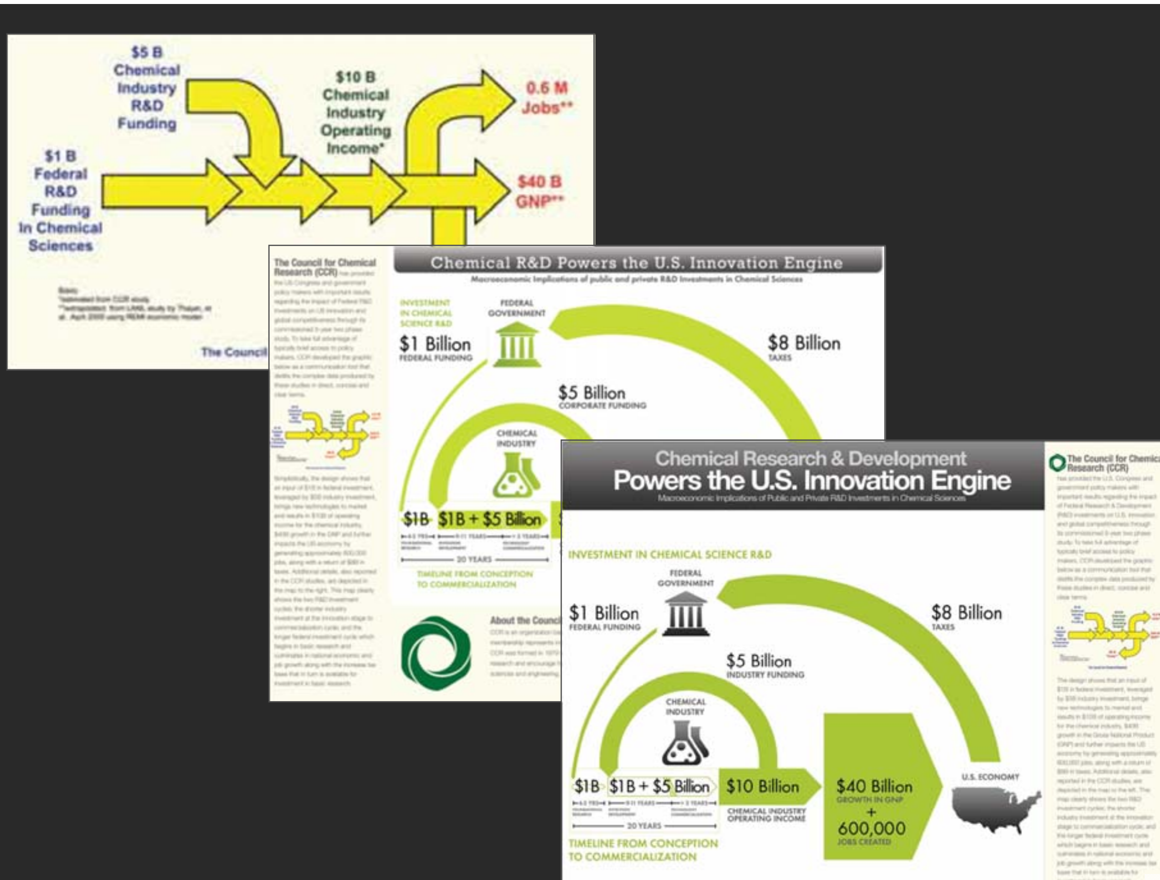
Exhibit has been shown in 72 venues on four continents. Currently at
- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA
- Center of Advanced European Studies and Research, Bonn, Germany
- Science Train, Germany
- Cultural Dimensions of Innovation, UCD Conference, Dublin, Ireland

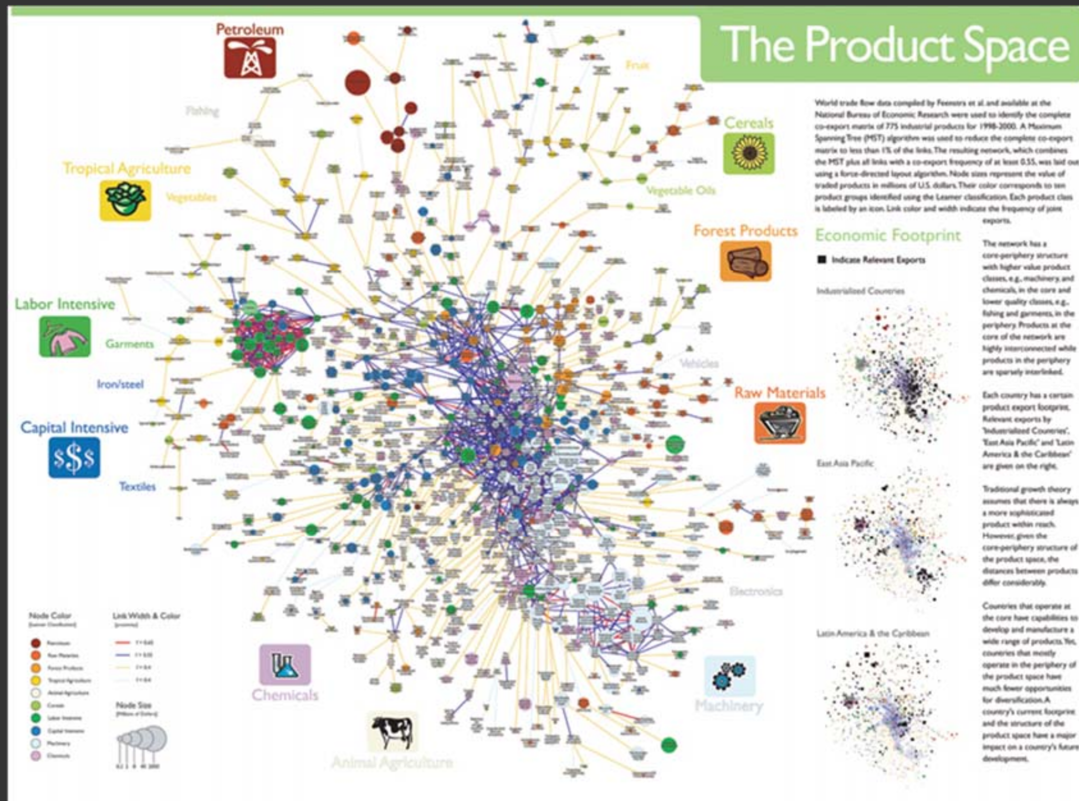


Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva. 2008. A Clickstream Map of Science.



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

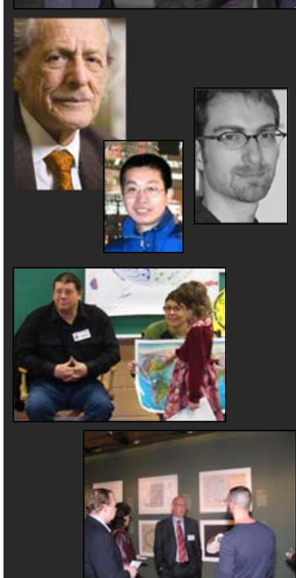




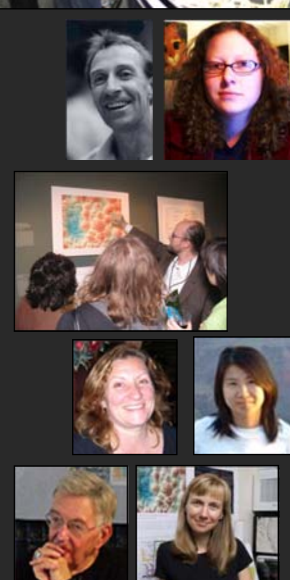
Cesar A. Hidalgo, Bailey Klinger, Albert-László Barabási , Ricardo Hausmann . 2007. The Product Space

Adrian White and the National Geographic EarthPulse Team. 2008.
A Global Projection of Subjective Well-being

Debut of 5th Iteration of Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University, *http://mediax.stanford.edu*, *http://scaleindependentthought.typepad.com/photos/scimaps*

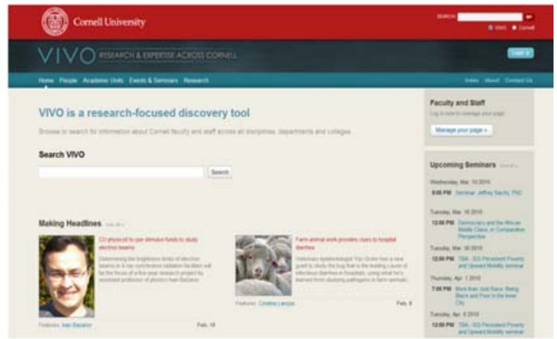


We would like to thank the map makers



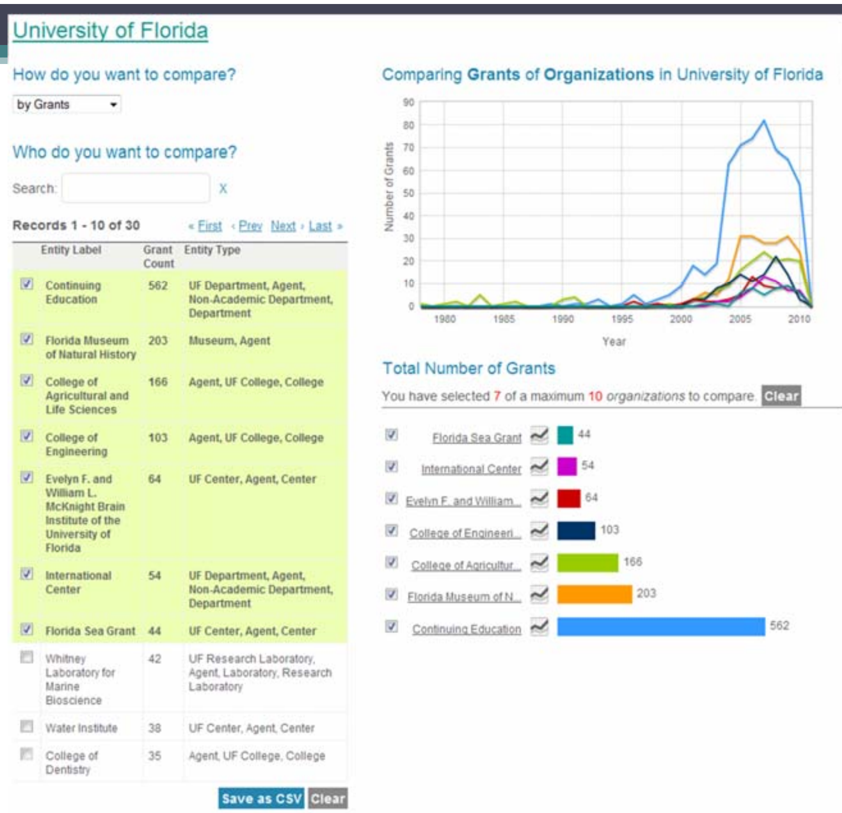
VIVO: A Semantic Approach to Creating a National Network of Researchers (http://vivoweb.org)

- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.
- Simplify reporting tasks, e.g., generate biosketch, department report.

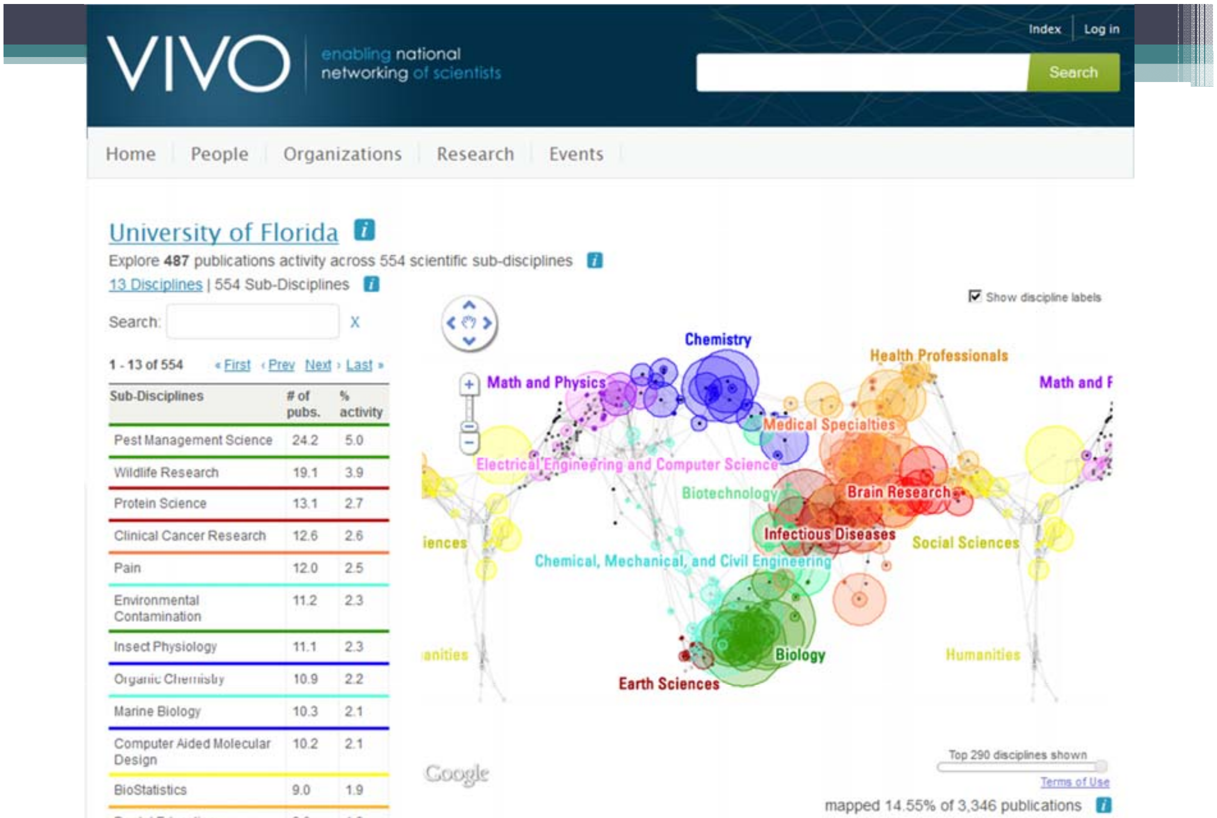


Funded by $12 million NIH award.

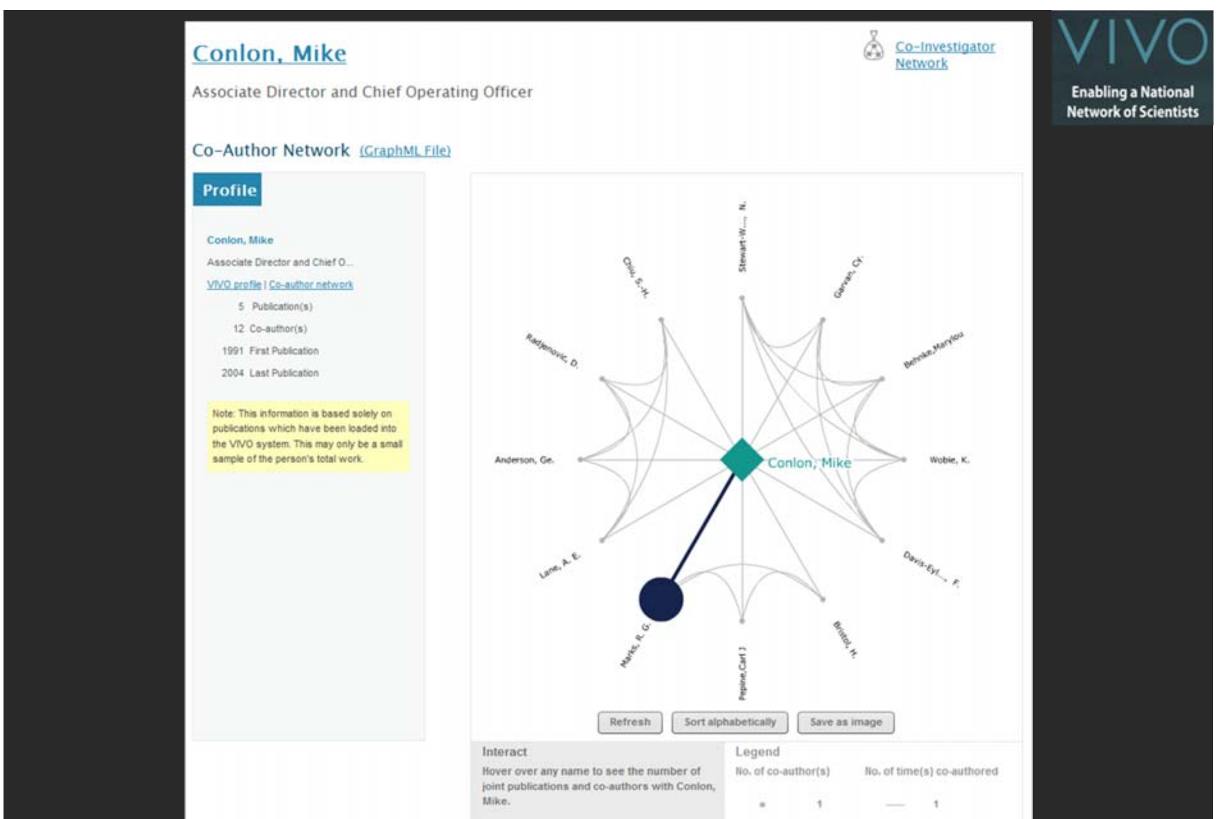
Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Chris Case, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinest, Yang Li, Paula Markes, Sara Russell Gonzalez, Alexander Rockwell, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raum, Brenda Stevens, Alicia Turner, Stephen Williams. **Indiana University**: Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhon, Jon Dunn, Micah Linnemeier, Nianli Ma, Robert McDonald, Barbara Ann O'Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine:** Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **The Scripps Research Institute:** Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaeleen Trimarchi. **Washington University, St. Louis**: Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. **Weill Cornell Medical College:** Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cheriff, Oscar Cruz, Dan Dickinson, Chris Huang, Itay Klaz, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Jesse Turner, Vinay Varughese.



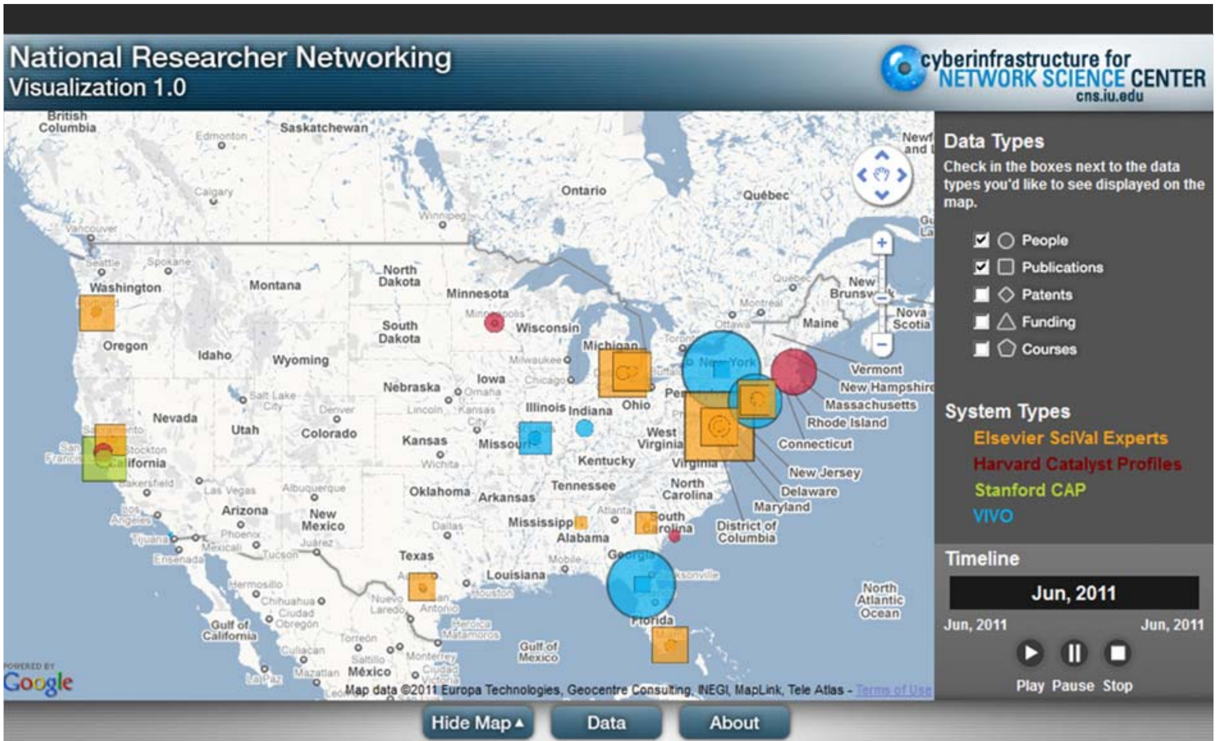
Temporal Analysis (When) Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level



Topical Analysis (What) Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)



Network Analysis (With Whom?) Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?



http://nrn.cns.iu.edu

Geospatial Analysis (Where) Where is what science performed by whom? Science is global and needs to be studied globally.

References

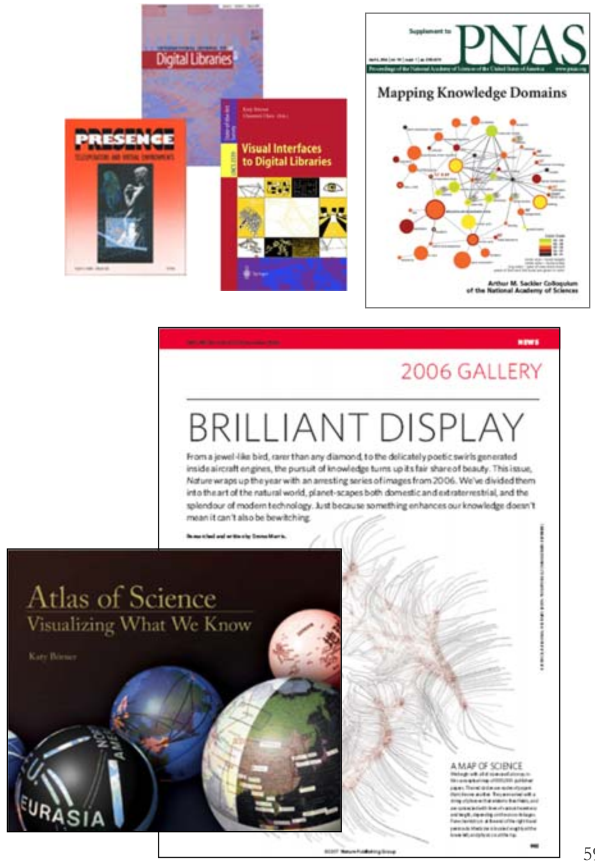
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255. http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf

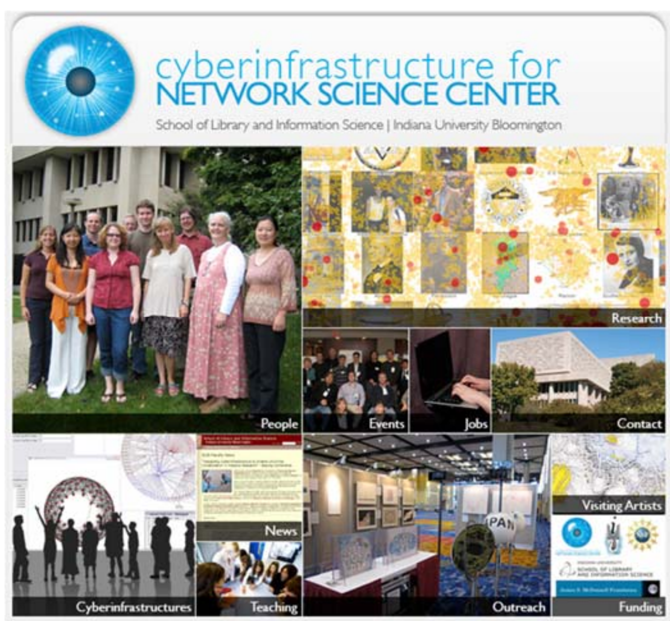
Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1/

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science.** In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc., Volume 41, Chapter 12, pp. 537-607.
http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

Börner, Katy (2010) **Atlas of Science**. MIT Press. http://scimaps.org/atlas

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2011) **Models of Science Dynamics**. Springer Verlag.





All papers, maps, tools, talks, press are linked from http://cns.iu.edu

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