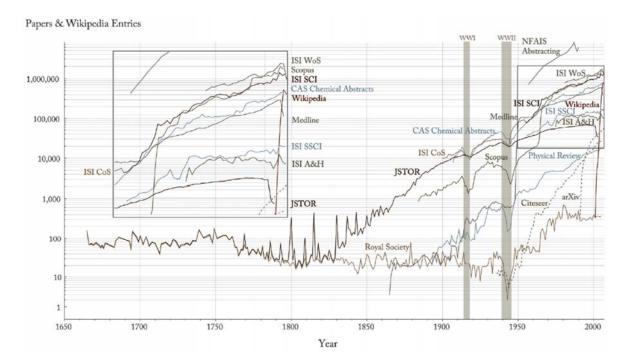


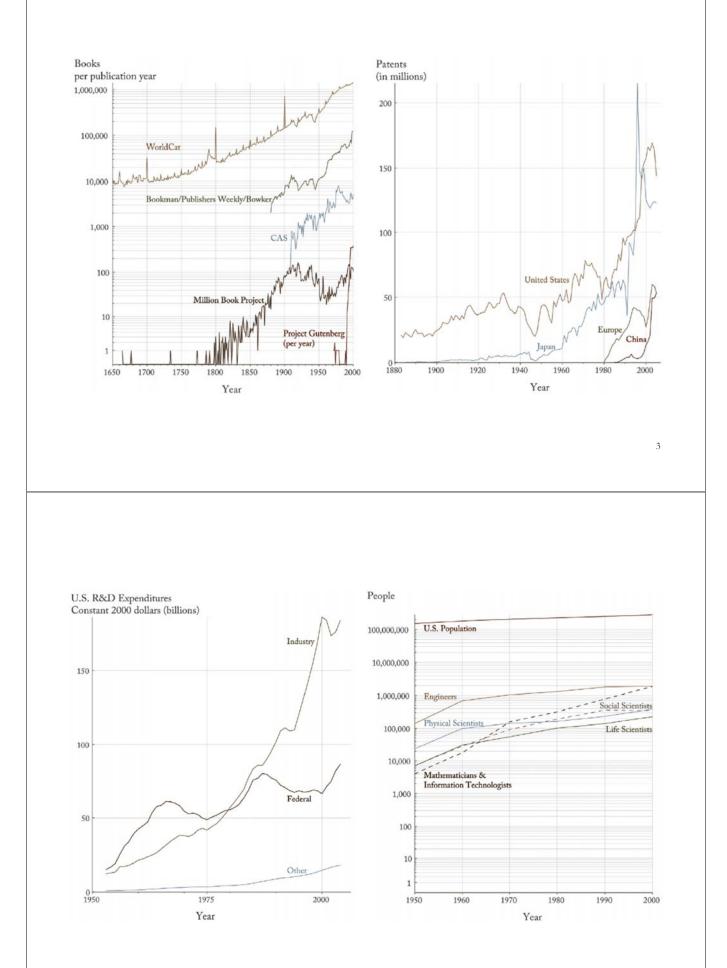
katy@indiana.edu | http://cns.iu.edu

GIS Day at Indiana University, Bloomington, IN November 16, 2011

IRAS

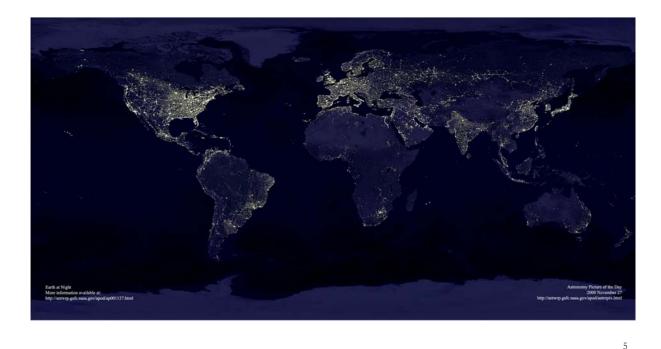
The Rise of Science and Technology





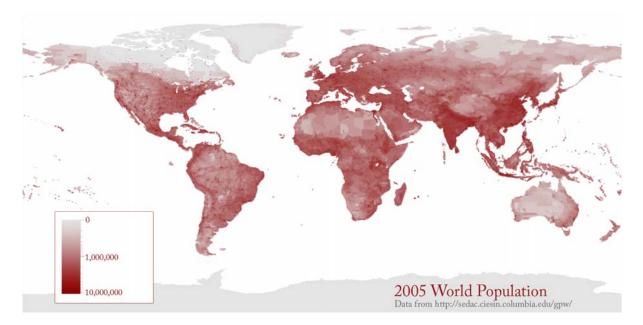
2000 Night on Earth

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.



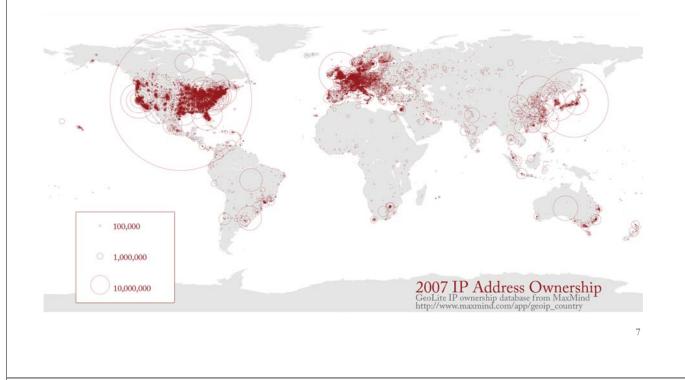
2005 World Population

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).



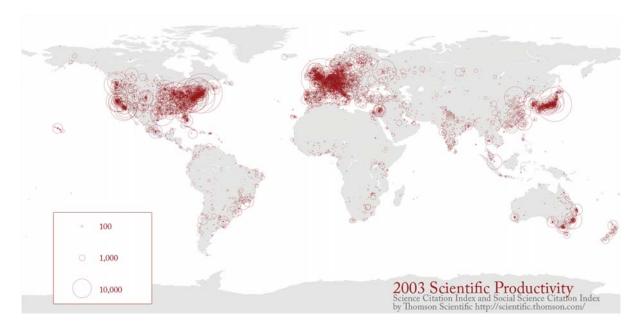
2007 IP Address Ownership

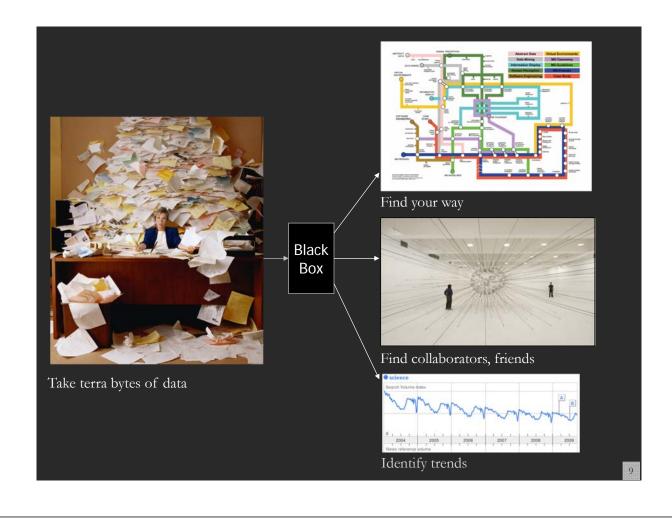
This map shows IP address ownership by location. Each owner is represented by a circle and the area size of the circle corresponds to the number of IP addresses owned. The larges circle denotes MIT's holdings of an entire class A subnet, which equates to 16,581,375 IP addresses. The countries that own 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.





Early Maps of the World

VERSUS

Early Maps of Science



3D

Physically-based Accuracy is measurable Trade-offs have more to do with granularity 2-D projections are very accurate at local levels Centuries of experience **Geo-maps can be a template for other data**



n-D Abstract space Accuracy is difficult Trade-offs indirectly affect accuracy 2-D projections neglect a great deal of data Decades of experience Science maps can be a template for other data

Foreword

•••

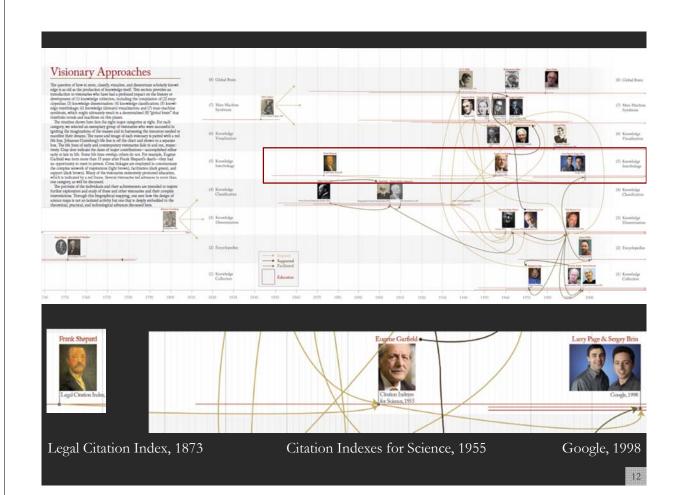
...

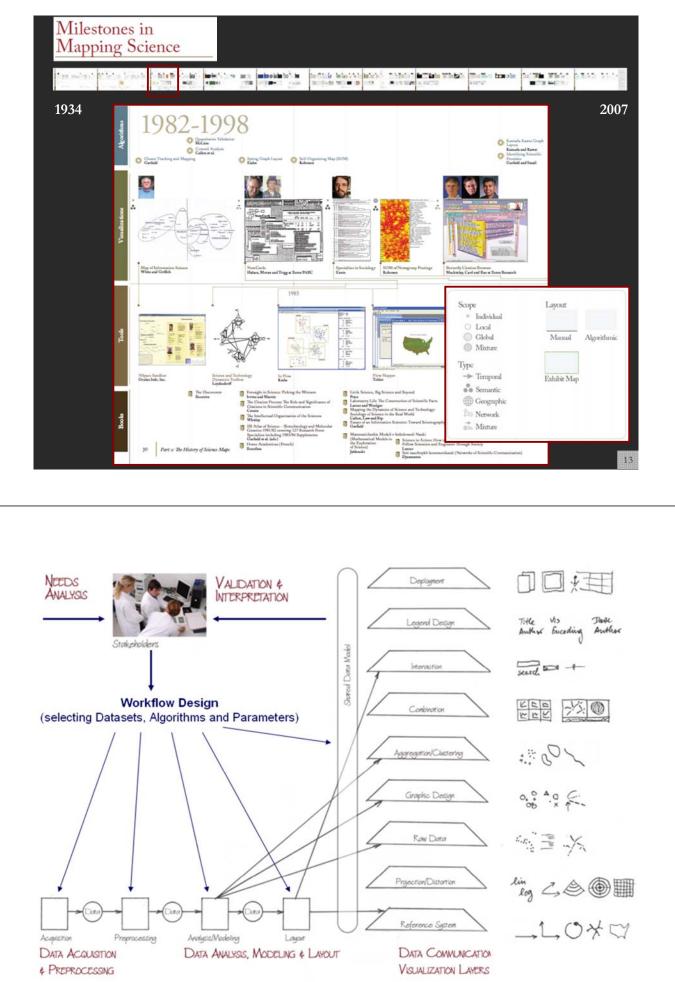
The explorers whose work is represented in the pages of this rich and fascinating volume face challenges far more daunting. First, the world they strive to represent is an abstract and intellectual one, not a physical reality that can be imaged from space, surveyed on the ground, and depicted in miniature on a map. The interrelationships among the land-marks of this <u>abstract world</u> are real, but they are not easily represented in the simple, straightforward ways that one can convey the distances between, say, three cities.

Second, there is no equivalent in the cartography of science to the <u>standards and conventions</u> upon which we mappers of the physical world comfortably depend. There's no agreed-upon notion of north-asup, of systems of latitude and longitude, of symbols, scale, and projection. Mapping the world of science requires the invention of a brand-new geography. Not only that, but the new geography then needs to be represented visually using colors, lines, and symbols for which no conventions exist. Third, the world that is being mapped in this book is changing at a dizzying rate. It's a fact of twenty-first-century science that whole realms of inquiry bloom into existence almost overnight, creating new places and spaces in ways that are alien to "normal" cartography. It is as if entire continents and archipelagoes were to constantly erupt on the roiling surface of a map even as that map was being drawn for the first time.

•••

Allen Caroll Chief Cartographer National Geographic Society





Temporal Analysis

A comportal Aratalysis Science overwar sins. Autibate values of schol-arly entities and their diverse aggregations increase and decrease at different nates years of request with dif-ferent latesty-rates to internal and external events. Temporal analysis sins to identify the mature of phenomena represented by a sequence of observa-tions such as patterns, trends, seasonality, outliers, and bursts of activity.

Data

A time series is a sequence of events or observations that are ordered in time. Time-series data can be continuous (there is an observation at every instant of time; see figure to the right) or discrete (obser-vations exist for regularly or irregularly spaced intervals). Temporal aggregations—over journal volumes, years, or decades—are common.

Algorithms

Augurization Proquently, some form of filtering is applied to reduce noise and make patterns more saliser. Smoothing (averaging using a smoothing window of a cortain widd) and cares approximation might be applied. The number of scholarly records is often

Topical Analysis

The topic coverage and topical similarity of basic and aggregate units of science (authors or institu-tions) can be derived from the units associated with them (papers, patents, or grants).

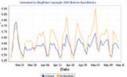
Data

ntic coverage of a unit of science The topic or se can be derived from the text associated with it. Topical aggregations (for example, over journal volumes, scientific disciplines, or institutions)

Algorithms

Topic analysis extracts the set of unique words or word profiles and their frequency from a text corpus, Stop words, such as "the" and "of," are removed. Summing can be applied. Coword analysis identifies the number of times two words are used in the title, keyword set, abstract, or full text of a paper. The space of en-occurring words can be mapped, providing a unique view of the topic rage of a data set (see page 66, Exemplification). Similarly, units of science can be grouped according to the number of words they have in common. Salun's term frequency inverse document

plotted to get a first idea of the temporal distribuprotons or get a new task of the horses in testal values from of a data and it. It might he showes in testal values or as a percentage of those. One may find out how long a scholarly entity was actively, how old it was at a cortain point, what growth, haterey to peak, or ducay rate it has, what correlations with other time series schet, or what trends are observable. Data models such as the least squares reedel (available in most statistical software packages) are applied to bost fit a selected function to a data set and to determine if the trend is significant. Kleinburg's burst detaction algorithm is commonly applied to identify words that have experienced a stalders change in frequency of occurrence. frequency of occur



fromatory (TFIDP) is a statistical measure used to evaluate the importance of a word in a corpus. The importance increases proportionally to the number of times a word appears in the paper but is officit by the frequency of the word in the corpus. Dimensionality reductions techniques (see table on opposite page) are commonly used to project

high-dimens nal information spaces (for example, ingo and all unique papers multiplied by their unique terms) into a low, typically two-dimen-sional space. The SOM map below shows the topic landscape

of geography abstracts; see page 102. In Terms of Geography.



Geographic Analysis

Geographic analysis aims to answer the question o where something happens and what impact it has on of on neighboring areas

Data

Geographic analysis requires spatial attribute val-ues or geolocations for authors and their papers, extracted from affiliation data or spatial positions of nodes, generated from layout algorithms. Geographic data can be continuous (each record has a specific position) or discrete (a position or area axists for sats of records, like the number of paper per country). Spatial aggregations (for example, merging data via positi code, countries, states, ics, and continunts) are common (see page 66, Exemplification).

Algorithms

Cartographic generalization refers to the process of abstraction. This includes (1) graphic generaliza-tion: the emplification, enlargensare, displacement, merging, or selection of entities without enhancet or effect to their symbology and (2) concep-symbolization: the merging, selection, and tual symbolic

Network Analysis

The study of networks aims to increase our under standing of natural and manmade networks. It builds on social network analysis, physics, informa-tion science, hibitometrics, scientemetrics, infor-metrics, webomstrics, communication theory, soci-ology of science, and several other disciplines.

Data

Authors, institutions, and countries, as well as words, papers, journals, patents, and funding, are represented as nodes and their complex interrelations as talges (see Part 3: Toward a Science of Science/Conceptualizing Science: Basic Anatomy of Science). Nodes and edges can have timestamped attributes.

Algorithms

Diverse algorithms exist to calculate specific node, edge, and network properties (see "Network Science" review). Node properties include degree cantrality, betweenness centrality, or hab and authority scores. Edge properties include darabil-ity, raciprocity, intensity (weak or strong), density (how many potential edges in a network actually exist), reachability (how many steps it takes to go

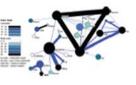
nbolization of entities, including enh (such as representing high-density areas with a city symbol).

Geometric generalization aims to solve the conflict between the number of visualized featu the size of symbols, and the size of the display surface. Cartographers dealt with this conflict intuitively in part until researchers like Friedrich Töpfer attempted to solve them with quantifiable.

Flow maps use line thickness and direction to show the number of tangible or intangible entities that diffuse over a geographic location or ecimes space (see CAS coasthor network, below, and page 158, 113 Years of Physical Review).



from one "end" of a network to another), ity (whether a network has a "conter" point), qual ity (reliability or certainty), and strongth. Network properties refer to the number of nodes and edges, network density, average path length, clustering coefficient, and distributions from which general properties such as "small-world," "scale-free," or "hierarchical" can be derived. Identifying major communities via community detection algorithm and calculating the "backbone" of a network via pathfinder network scaling or maximum flow algo rithms helps to communicate and make sense of large-scale networks. See the coauthor network of information visualization researchers have



15

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 knowled

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 application serving as useful navigational tools.

The Power of Maps features four cartographic map earliest global maps of our world by Ptolemy, an earl Johannes Janssonius, an early map of the whole worl 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 r visual index of our collective science and technology

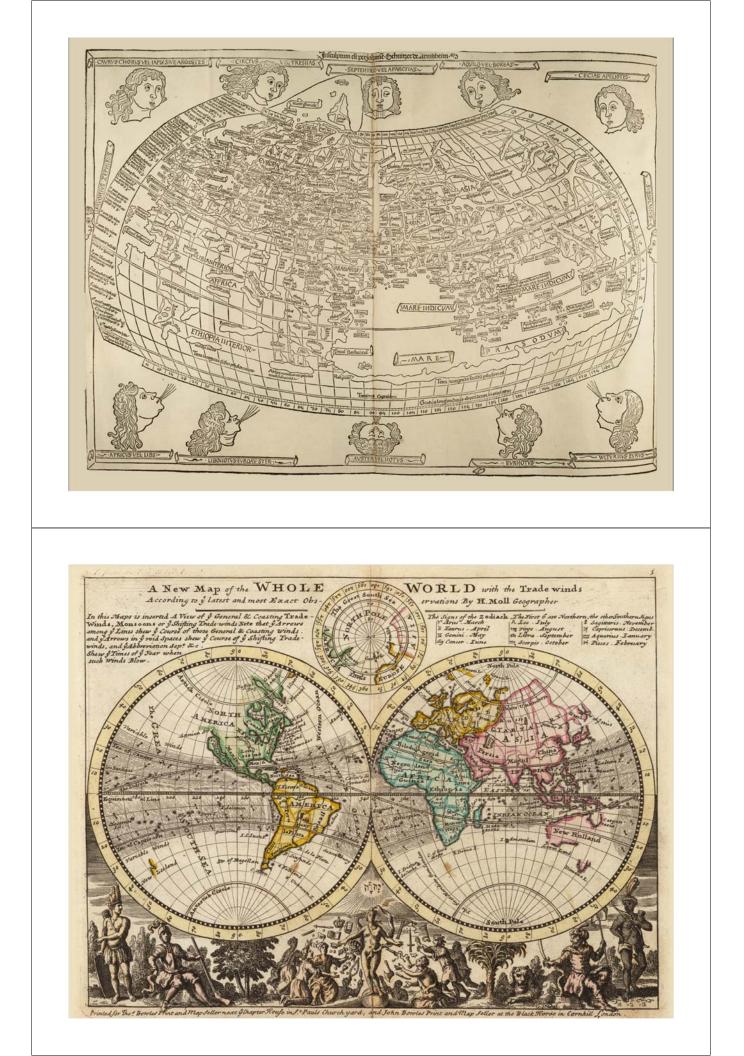
Note that the makers of the early cartographic ma ing presses, while the makers of the first maps of sci

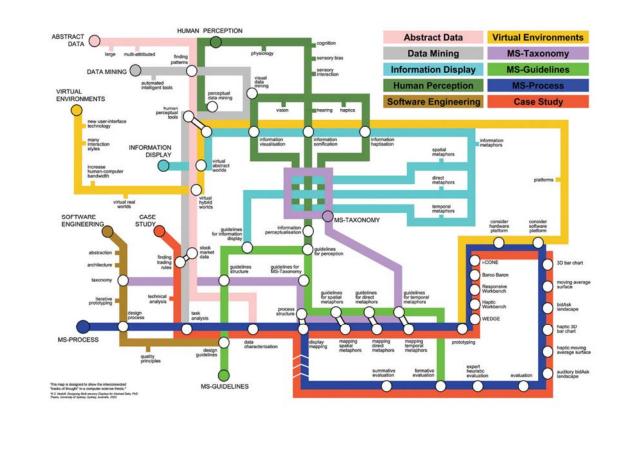


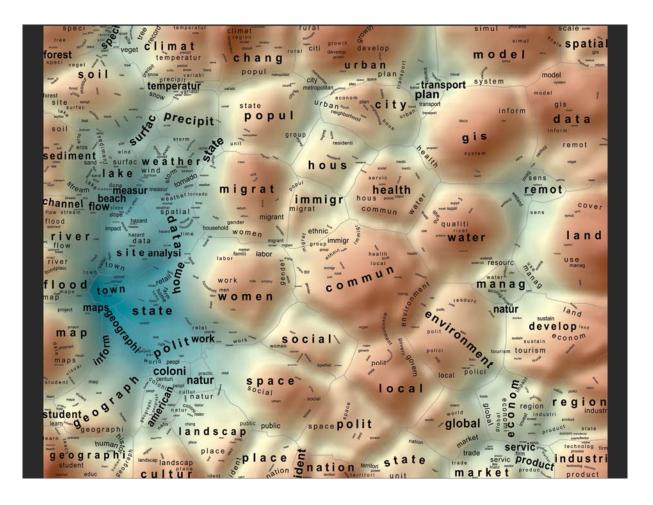
rec it i

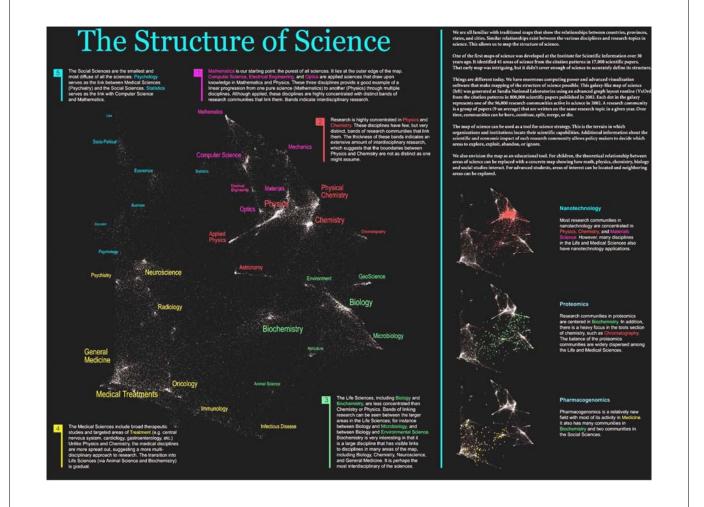












Second Iteration of Exhibit (2006): The Power of Reference Systems

Four Existing Reference Systems Versus Six Potential Reference Systems

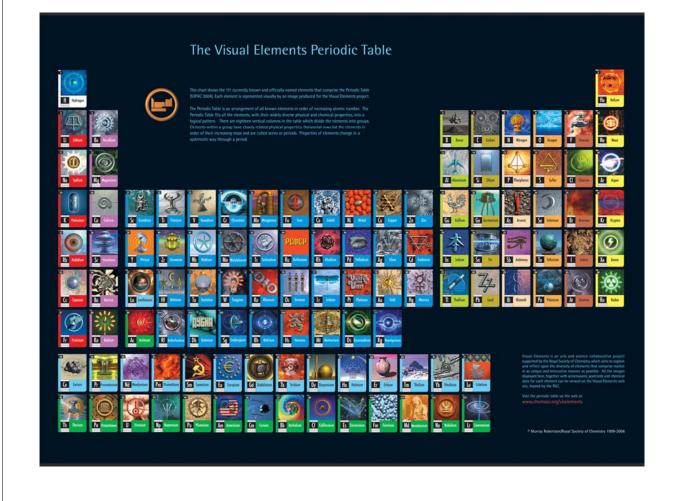
This iteration aims to inspire discussion about a common reference system for all existing scholarly knowledge. Throughout history, scientists have battled to agree on standardized reference systems for their respective fields of research. These standards are invaluable for indexing, storing, accessing,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

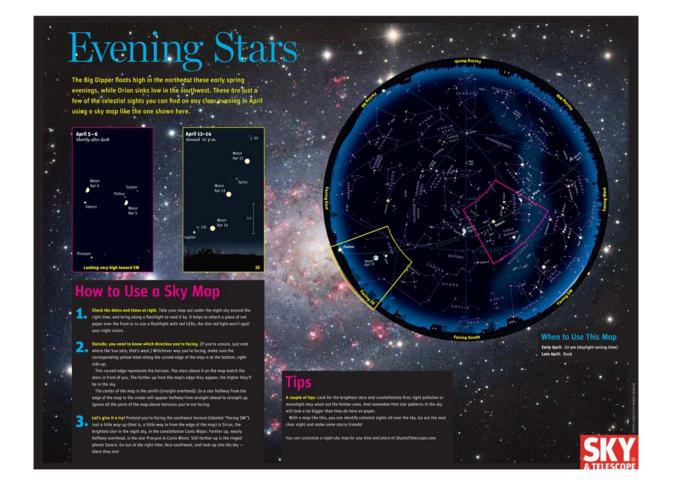
Results include the description of the electromag odic table of elements, geographic projections, and systems, shown here. Note that the geographic may from paper to geographic information systems (GIS for public use and consumption.

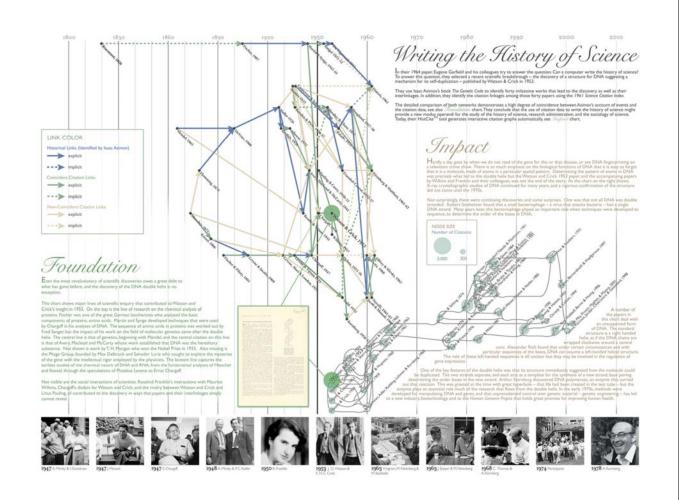
In comparison to these four existing systems are systems for scholarly knowledge. Each reference sy sional timeline and the geographic system to the se used to identify the location of an author, paper, pa tory or contribution.



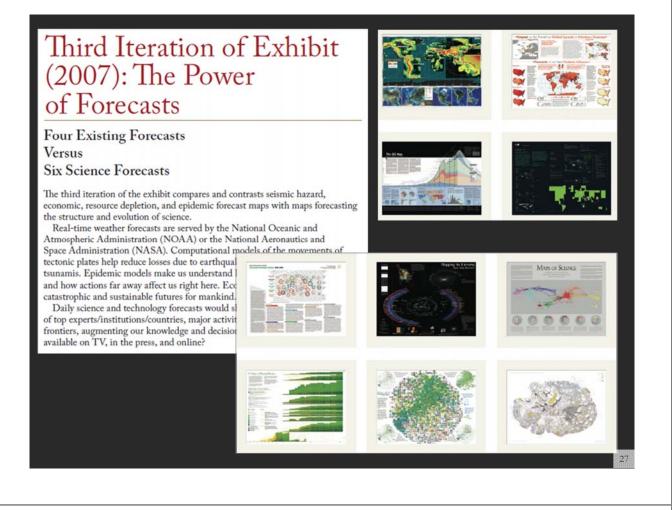


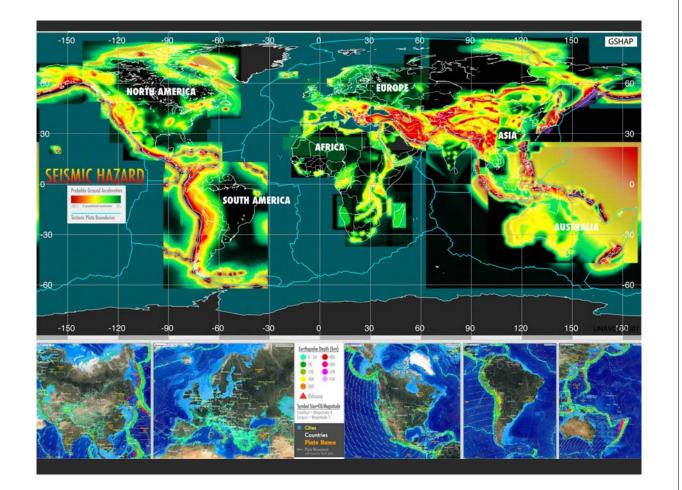


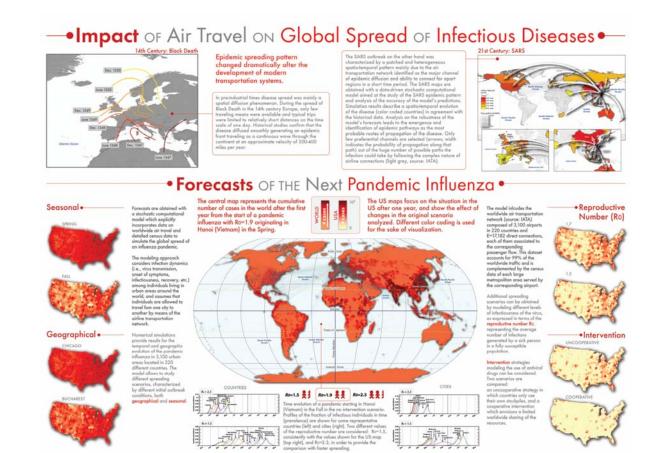




L Science can be thought of as containing themes and paradigms. Themes are areas of current research, while paradigms comprise the dominant tool sets and existing knowledge that are used by today's researchers. This map shows 776 major paradigms in science along with the dominant relationships between henes paradigms. Paradigms are shown as circles; strong relationships between paradigms are indicated by the lines connecting the circles. The map was created by recursively clustering the 820,000 papers referenced most often in 2003. Clustering at each leved was done using Y&Ord, a force-directed graph layout routine. These papers formed 33,000 clusters, 6,100 bigher-level clusters, and finally the 776 paradigms. Although each paradigm contains, on average, 1,000 papers, some are larger and some are smaller, as shown by different sized circles on the map. The ring-like structure that is formed by scientific paradigms The ring-like structure that is formed by scientific paradigms is very robust. We find similar structures for different years, and for maps generated from scientific journals. "The Structure of Science", a galaxy map shown in the first iteration of Places is Spaces, is a map based on clustering of scientific journals, with superimposition of papers on the journal structure, whereas this map was gene-rated directly from highly-cited papers. "The Structure of Science" shows current science in a disciplinary context, while this map can show the breadth of disciplines that contribute to single paradigms. Matha Mechanics Because of the robust nature of the str Because of the robust nature of the structure of science and its paradigms, we have placed our 776 scientific paradigms within a reference sys-tem containing 12 radial slices and 6 rings. This liows the position of each paradigm to be codified and available for lookup; for instance *Fluid Mechanics* paradigms are in grid B3. sychiat Ecology We have also calculated and displayed the vitality of each paradigm. *Vitality* is a measure of the speed at which a group of researchers reaches consensus about major improvements. Paradigms are major. The white circles represent usually takes years to reach consensus about which improvements are major. The white circles represent communities where consensus is reached relatively slowly. This is a common phenomenon in the social sciences, ecological sciences, computer sciences, and mathematics disciplines. The red circles represent communities of researchers where common in physics, chemistry, biochemistry, and many medical disciplines. Very dark circles juch as those in Astrophysics, L5-6) resent communities where consensus is reached extremely quickly. We have also calculated and displayed the vi Circle size = Para Vitality Scale Average or li High Very high The map of scientific paradigms and its reference system can be H E used for multiple purposes. Countries, industries, companies, universities and individual researchers can all locate themselves within the map, eithe Disea and individual researchers can all **locate themserves wrunn the map**, encore as single points, or as a specific collection of paradigms. Various metrics, such a vitality, can be overlayed on this reference system to highlight specific impacts Science education and personal discovery can also be enhanced by linking stories an facts to the map that highlight scientific history, current advances and relationship between scientific paradigms G







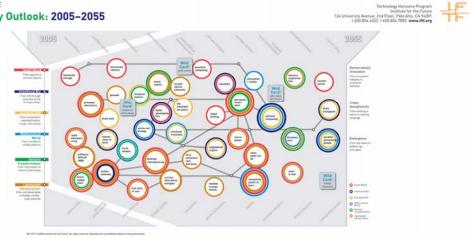
INSTITUTE FOR THE FUTURE Science & Technology Outlook: 2005-2055

HAP THEMES

Intentional Biology

anginee Still, wy elegant 0

Small World

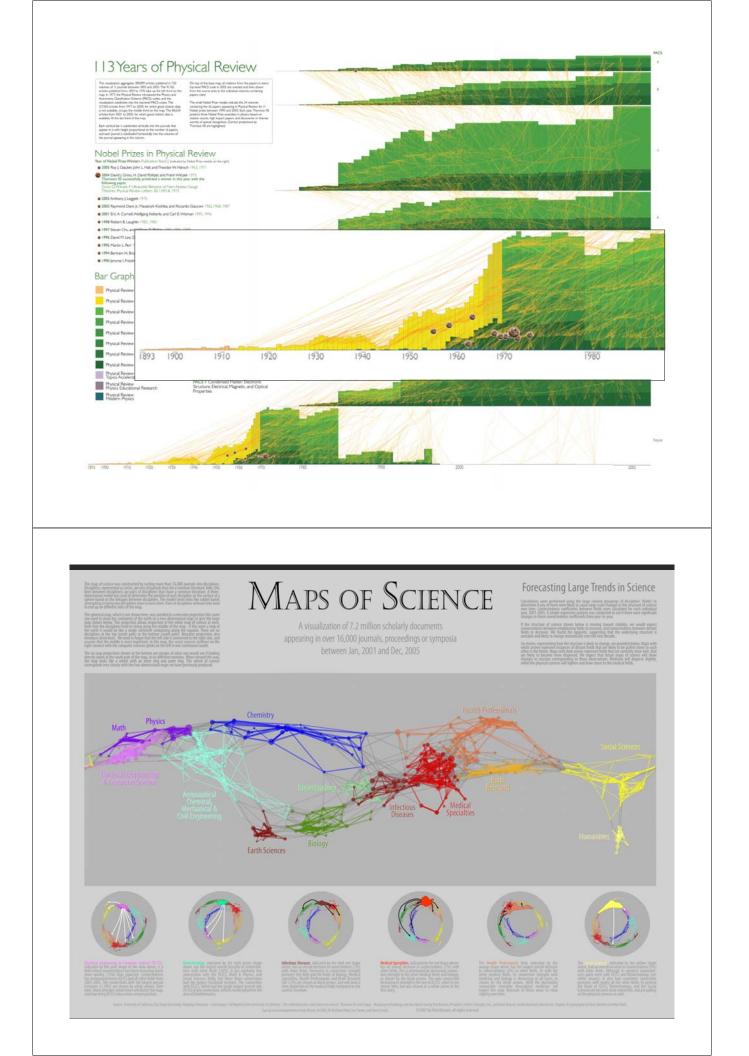


ry Transformation

124 U

0..... the 20th cent

OTransdiscip



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 science

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

The Hands-On Scient stand science from abc color drawings. Childi placing images of majo appropriate places on to of various countries fo patents. Shape of Sciet The Video of the Exhi Public Library (NYPI NYPL officials, who c



Illuminated Diagram Display

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

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.

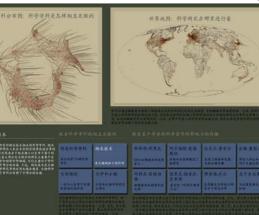




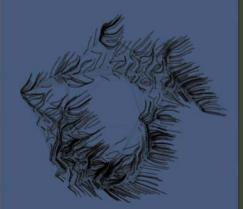


Large-scale, high resolution prints illuminated via projector or screen.

⁻ Interactive touch panel.



TOPIC MAP: HOW SCIENTIFIC PARADIGMS RELATE





You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

Nanotechnology

This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower right.

All Topics Sweep through all 776 scientific paradigms	Nanotechnology Science on the tiny scale of molecules	Francis H. C. CRICK Co-discovered DNAs double helix	Albert EINSTEIN Revitalized physics with Relativity theories	Michael E. FISHER Models critical phase transitions of matter	Susan T. FISKE Connects perception and stereotypes	
Sustainability The science behind our long-term hopes	Biology & Chemistry The interface between these two vital fields	Joshua LEDERBERG Pioneer in bacterial genetic mechanisms	Derek J. de Solla PRICE Known as the "Father of Scientometrics"	Richard N. ZARE Uses laser chemistry in molecular dynamics	About this display People & organizations that helped create it	
We sweep slowly through adjoining related topics lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.		genetic mechanisms of scientometrics molecular dynamics that hered create it A single person's spreading influence is shown as a series of four snopshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights every- thing that cites that original work. Note that this first-generation impact extends to far more topics than did the original work. The third shapshot lights science that cites the second; and the fourth lights science that cites the third.				





You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

Nanotechnology

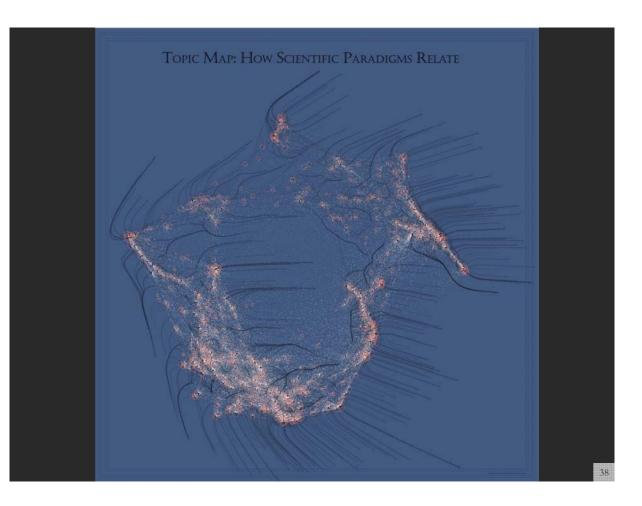
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower right.

All Topics Sweep through all 776 scientific paradigms	Nanotechnology Science on the tiny scale of molecules	Francis H. C. CRICK Co-discovered DNA's double helix	Albert EINSTEIN Revitalized physics with Relativity theories	Michael E. FISHER Models critical phase transitions of matter	Susan T. FISKE Connects perception and stereotypes
Sustainability The science behind our long-term hopes	Biology & Chemistry The interface between these two vital fields	Joshua LEDERBERG Pioneer in bacterial genetic mechanisms	Derek J. de Solla PRICE Known as the "Father of Scientometrics"	Richard N. ZARE Uses laser chemistry in molecular dynamics	About this display People & organizations that helped create it
lighting up the places in t topic. You may select a su	adjoining related topics, he world that study each bset of the topics that deal g subjects by touching it.	places relating to that pe thing that cites that orig	ng influence is shown as a s erson's papers — papers that inal work. Note that this firs e third shapshot lights scier	are still highly cited today. t-generation impact exten	The second lights every- ds to far more topics than

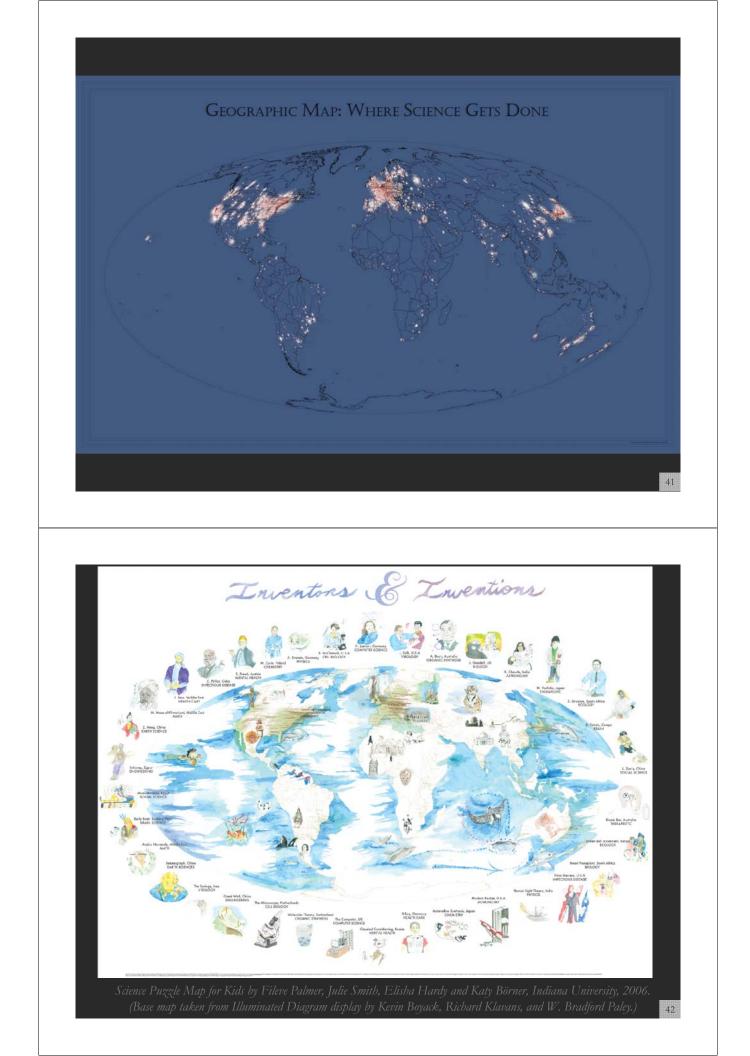
OEDGRAFFIIC WIAF, WHERE OCIENCE OEIS I

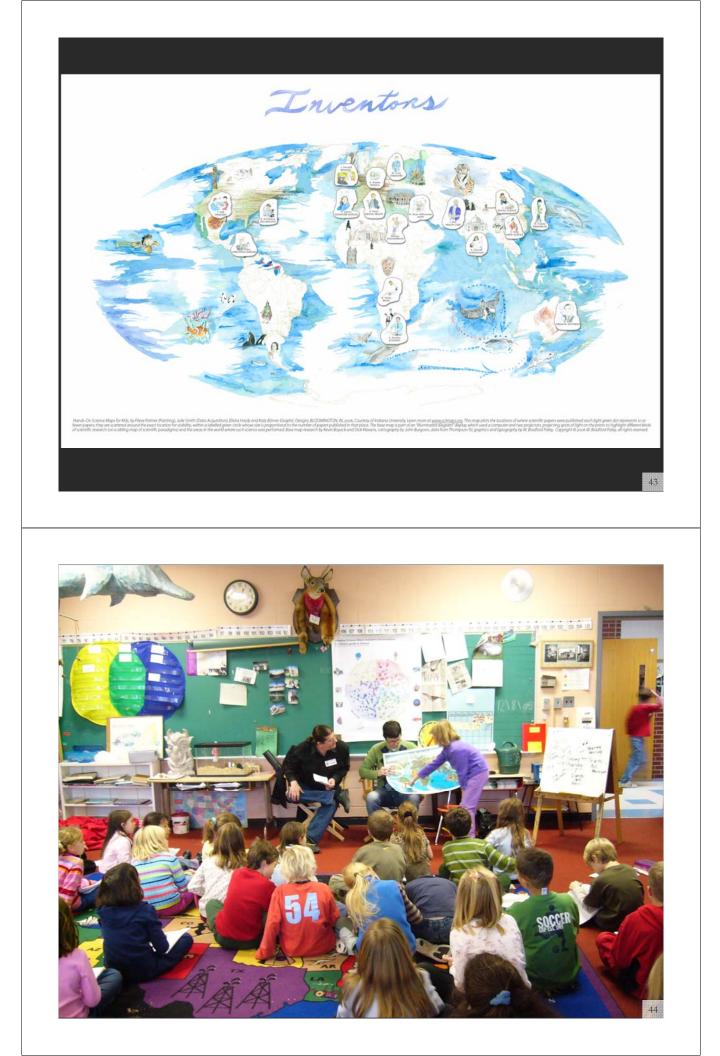


Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months 12 coaches, 300 m long Opening was on April 23rd, 2009 by German Chancellor Merkel <u>http://www.expedition-zukunft.de</u>









Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org/

ORDER



* 2.5		-		12	
<u>x, 200</u>				T B-SHR	
Science	Maps f	or Scien	ce Polio	cy Make	rs (2009)
	4	12		00	
	SIGNAR	<u>a</u>			

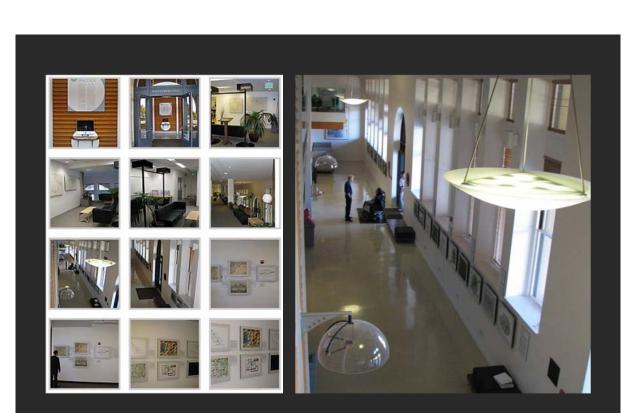
Science Maps for Economic Decision Makers (2008)

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

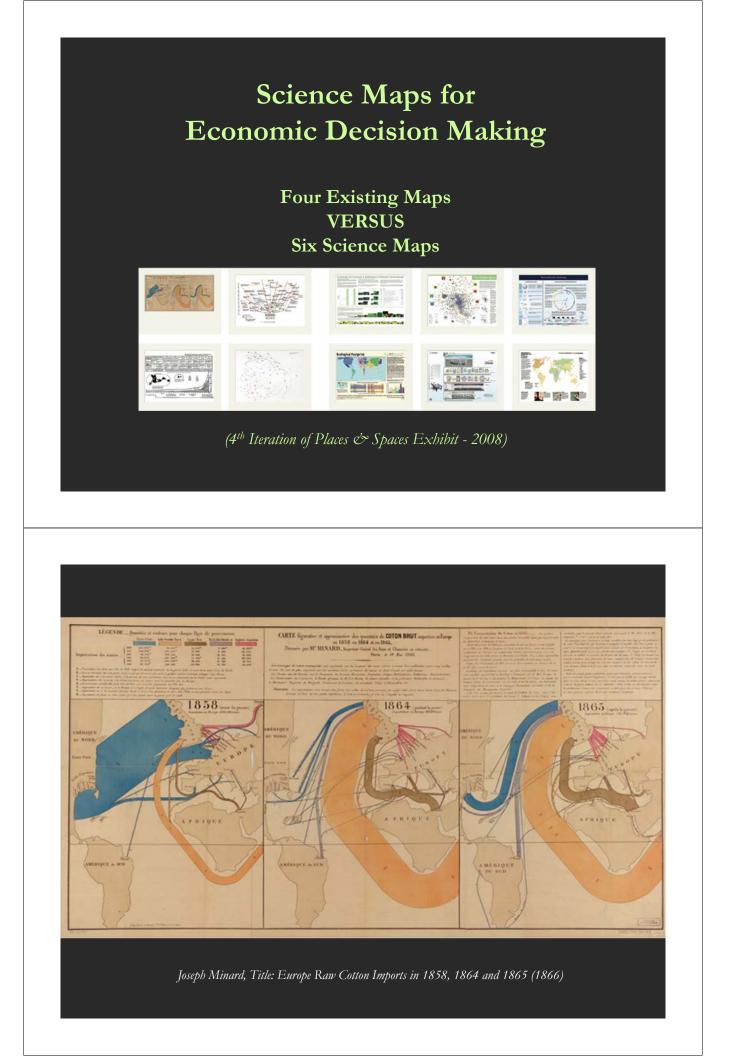


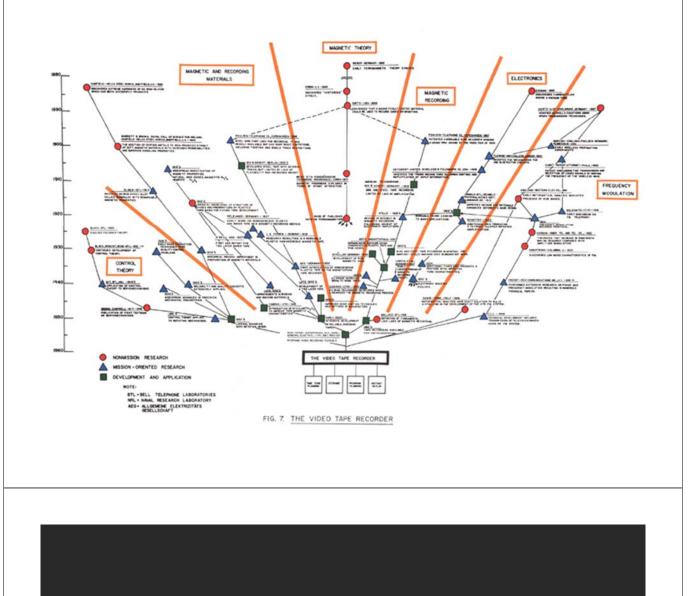


Debut of 5th Iteration of Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University, <u>http://mediax.stanford.edu</u>, <u>http://scaleindependentthought.typepad.com/photos/scimaps</u>



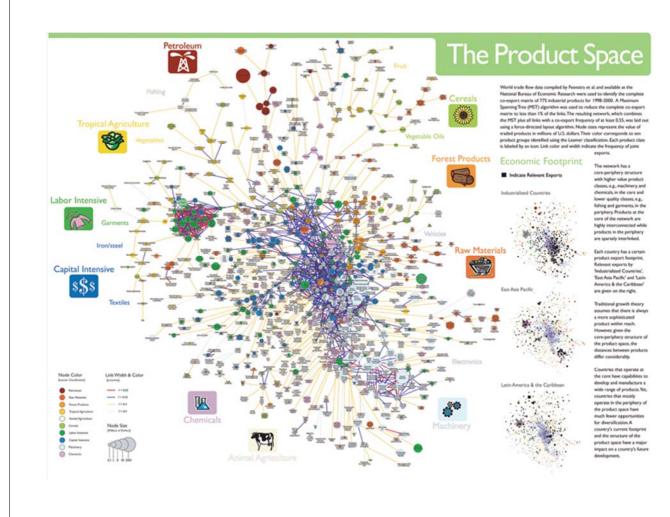






What insight needs to economic decision makers have?

What data views are most useful?



Happiness Depends on Various Factors

Cocial scientists are starting to include relative happiness with hard data on conomic status, health, and other factors as they assess quality of life, they rely on surveys of "subjective well-being"—box good people feel about their lives. A wordf map of one" happiness index" shows many, but not all, wealthy northern or "happiness index" shows many, but not all, wealthy northern countries faring well. Residents of sub-Saharan Africa and the former particularly low levels of contentinent.

Any attempt to measure happiness will fall short-each life is a series of joys, straggles, and sorrows, and satisfaction can depend as much on outlook as on circumstances. Averages obscure the happy in straughing nations, as rd outloos a... Averages obscure the happy moments in straggling nations, as well as people who suffer from poor health, poverty, or discrimination in countries that rank high. Still, happiness indices can help researchers move beyond simple economics as they track progress—or backsliding—over

MEASURING THE

ed fro 273 (Denmark and 5 to a low of 100 (Burt

G WELL-BEING



Hippy Content Orhappy No-data



VEALTH



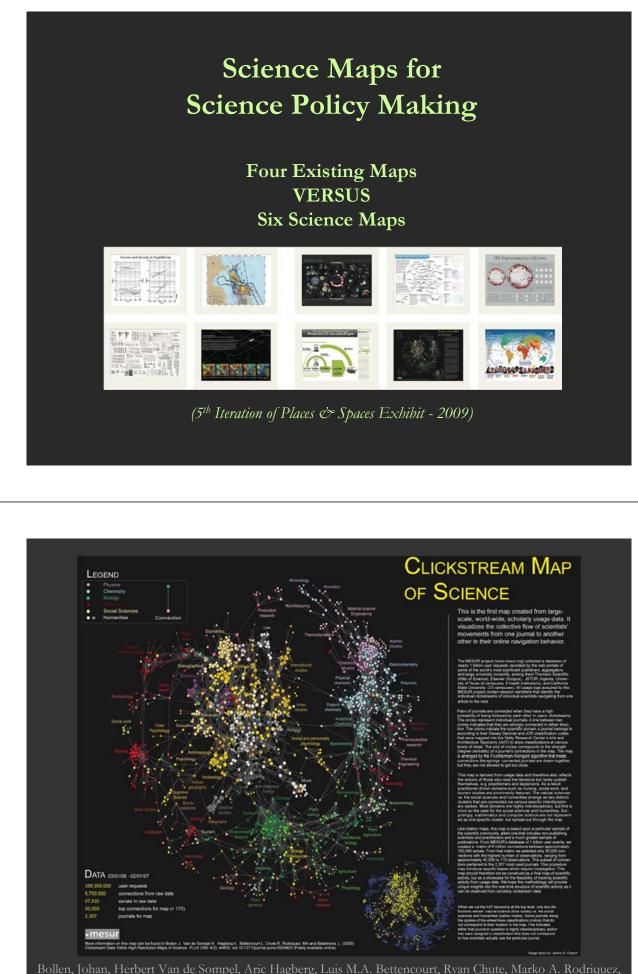
"It's time we admitted there's more to life than money."

RANKING THE WORLD'S HAPPIEST PLACES Northern Europe, North and several wealthy co, make the list, but so do proposed in pation

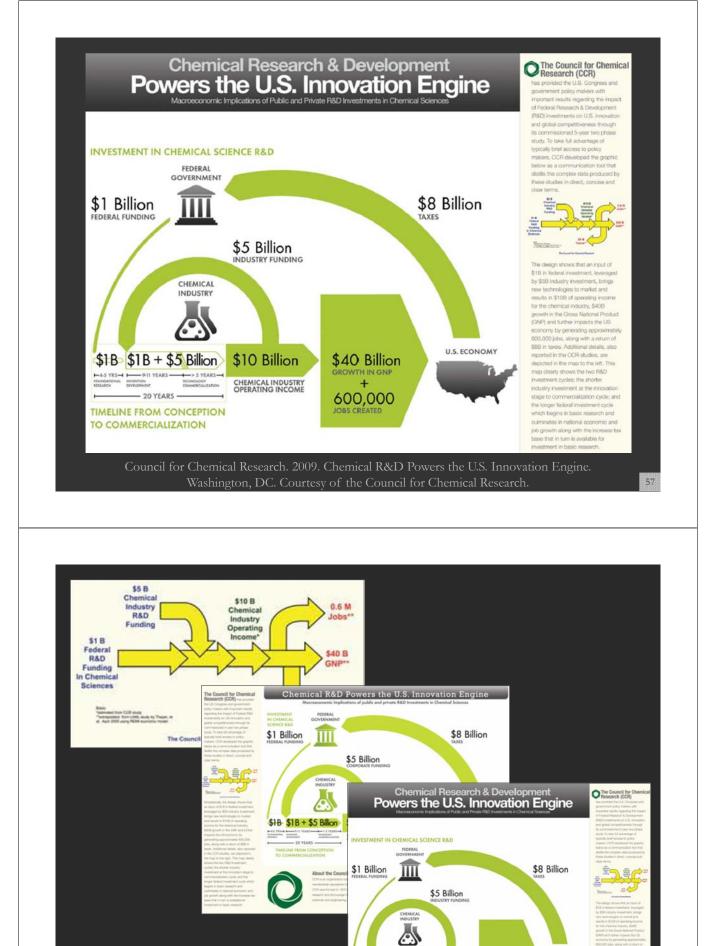
DENMARK SWITZERLAND 2 AUSTRIA ICELAND BAHAMAS FINLAND SWEDEN BHUTAN BRUNEI CANADA IRELAND LUXEMBO

5 COSTA RICA MALTA NETHERLANDS

ANTIGUA AND BARBUDA MALAYSIA NEW ZEALAND NORWAY SEYOHELES ST, KITTS AND NEWS UNITED ATAGE EMIRATES UNITED STATES UNITED STATES UNITED STATES VARUATU VENEZUELA



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. 56



\$1B \$1B + \$5 Billion \$10 Billion

CHEMICAL INDUSTRY OPERATING INCOME

H-L2 TE-4 H-S1 TEAS - 1 TEAS

- 20 YEARS

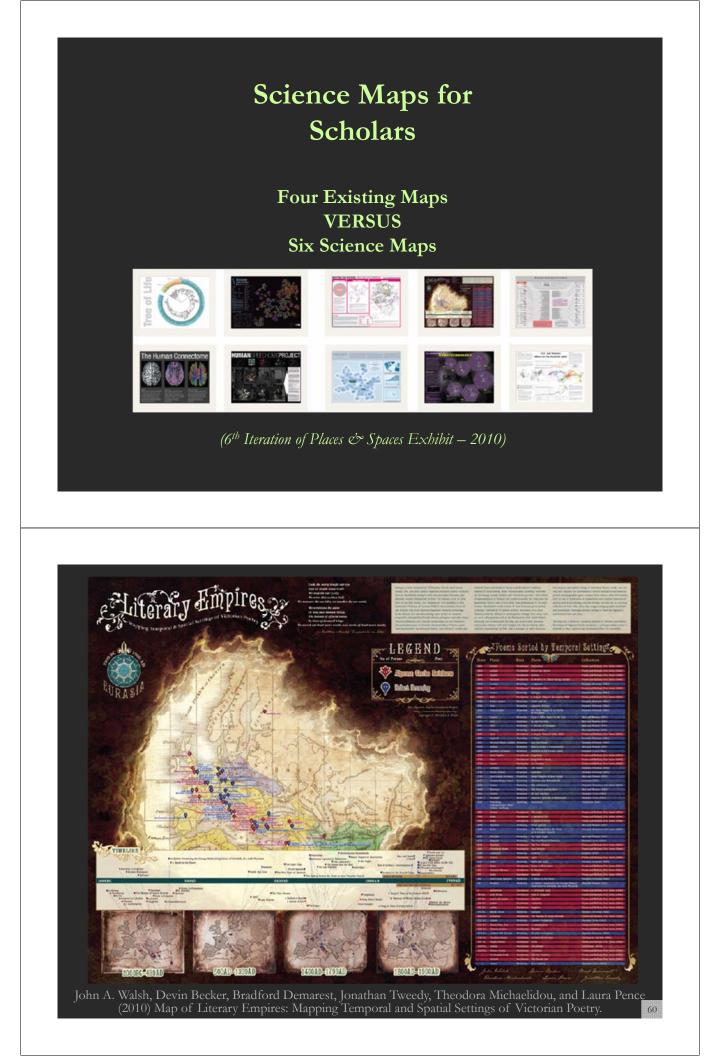
TO CON

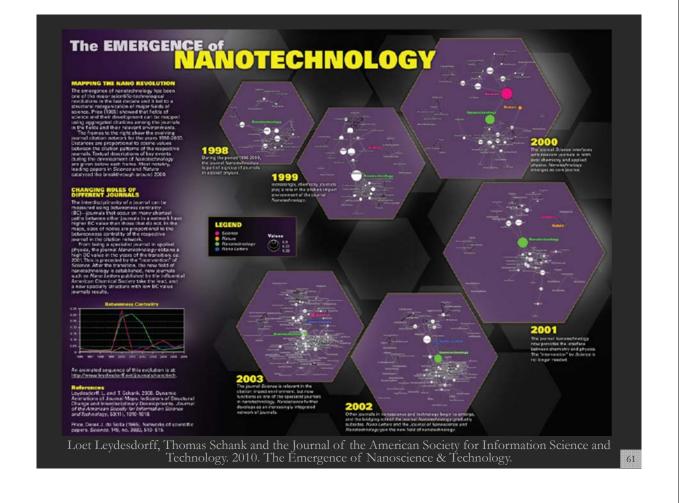
MELINE FROM CONCEPTION

U.S. ECC

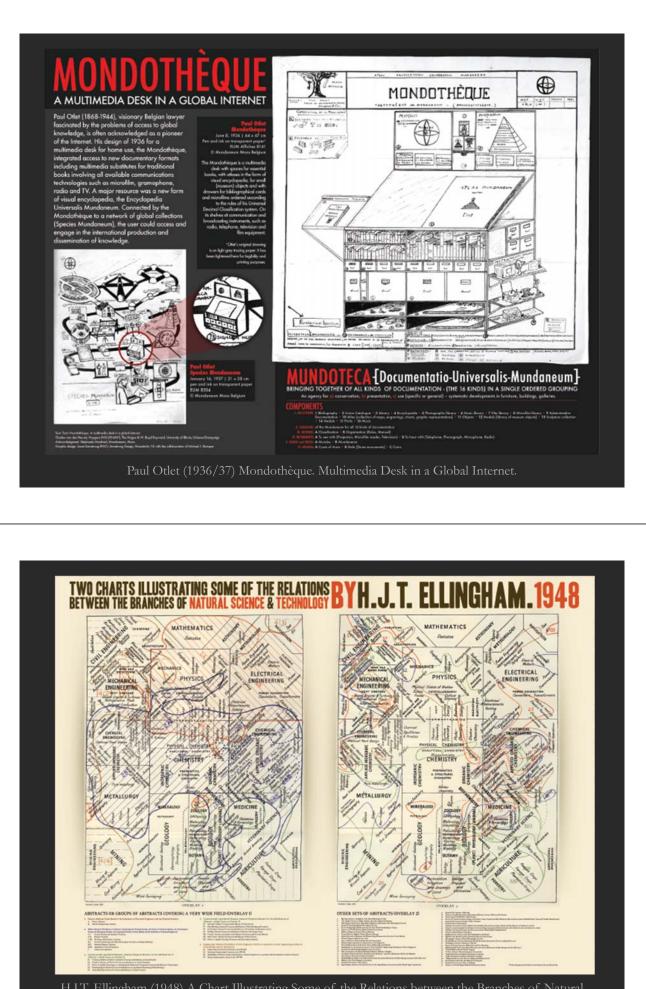
\$40 Billion

600,000

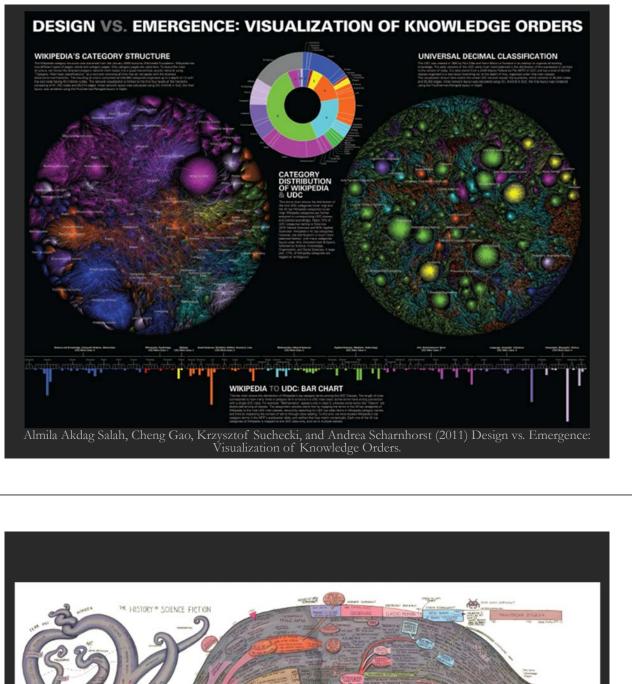


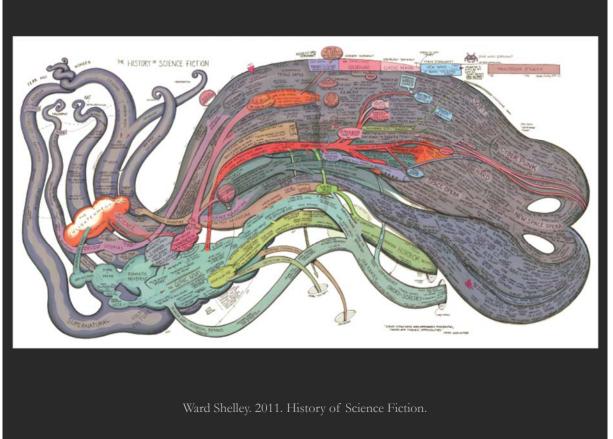


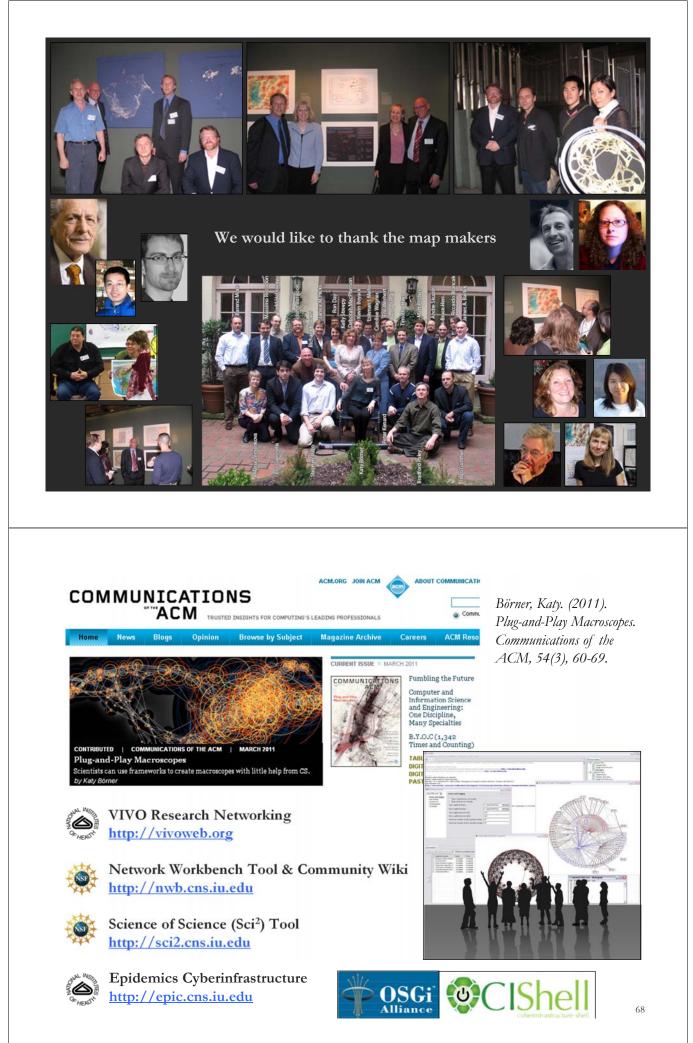
Science Maps as Visual Interfaces to **Digital Libraries** Four Existing Maps **VERSUS** Six Science Maps VII.3 VII.5 VII.7 VII.9 🔳 (...) 🖬 VII.2 VII.4 VII.6 VII.8 **VII.10** (7th Iteration of Places & Spaces Exhibit – 2011)



H.J.T. Ellingham (1948) A Chart Illustrating Some of the Relations between the Branches of Natural Science and Technology.







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

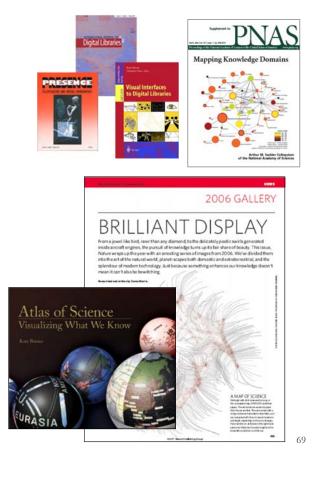
Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. Proceedings of the National Academy of Sciences of the United States of America, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1/

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: <u>http://www.facebook.com/cnscenter</u> Mapping Science Exhibit Facebook: <u>http://www.facebook.com/mappingscience</u>