places & spaces

Cartography of the Physical and the Abstract

On display in the Mezzanine Space
Natcher Conference Center, Bethesda, MD
February 6, 2006

Motivation and Goals

Humanity's knowledge and our means to share it are increasing at an accelerating rate. Yet, our perceptual and cognitive abilities stay nearly constant. We are expected to know more works than we could possibly read and understand in a hundred lifetimes. As a consequence, experts become highly specialized and isolated. Science continues to fragment, to duplicate, and to re-invent itself.

Today, we use search engines to access all of humanity's knowledge and expertise. Search engines retrieve facts from a growing sea of information. However, how big is this sea? How can we efficiently navigate to the useful islands of knowledge? How is knowledge interlinked on a global scale? In which areas is it worth investing resources? We don't know.

It is not the first time mankind has faced this type of question. However, it is the first time that there is an opportunity to coordinate efforts across cultures and disciplines to provide answers.

Cartographic maps of physical places have guided mankind's explorations for centuries. They enabled the discovery of new worlds while also marking territories inhabited by unknown monsters. Without maps, we would be lost.

Domain maps of abstract semantic spaces aim to serve today's explorers navigating the world of science. These maps are generated through a scientific analysis of large-scale scholarly datasets in an effort to connect and make sense of the bits and pieces of knowledge they contain. They can be used to objectively identify major research areas, experts, institutions, collections, grants, papers, journals, ideas, etc. in a domain of interest. Local maps provide overviews of a specific area: its homogeneity, import-export factors, and relative speed. They allow one to track the emergence, evolution, and disappearance of topics and help to identify the most promising areas of research.

The Places & Spaces exhibit has been created to demonstrate the power of maps. It has two components: the physical part supports the close inspection of high quality reproductions for display at conferences and education centers. It is meant to inspire cross-disciplinary discussion on how to best track and communicate human activity and scientific progress on a global scale. The online counterpart at http://vw.indiana.edu/places&spaces/ provides links to a selected series of maps and their makers along with detailed explanations of why these maps work. It also has the schedule of physical showings.

An initial theme of this exhibit is to compare and contrast first maps of our entire planet with the first maps of all of science as we know it. Come see with your own eyes the extent to which maps can be employed to help make sense of the flood of information we are confronted with and how domain maps can be used to locate complex and beautiful information.

Places & Spaces was originally created in April 2005 for "Mapping Humanity's Knowledge and Expertise in the Digital Domain" as part of the 101st Annual Meeting of the Association of American Geographers in Denver, Colorado. It will continue to expand and explore map making in its many dimensions and purposes over time.

For further information please contact Katy Börner, Indiana University, katy@indiana.edu & Deborah MacPherson, Accuracy&Aesthetics, deborah.macpherson@cox.net.

This exhibit was supported in part by a NSF CAREER grant under IIS-0238261.

Displayed Maps

Cosmographia World Map; by Ptolemy; Ulm, Germany; 1482
Courtesy of the James Ford Bell Library, University of Minnesota, Minneapolis, Minnesota

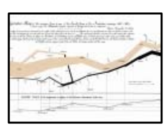
Claudius Ptolemy lived in Alexandria Egypt from 85 to 165 AD and was the first to establish the elements and form of scientific cartography. Ptolemy was interested in the whole world, not just places where people lived. He was curious about the dynamic relationships between the earth and the sun, the earth and the moon, and the cause and effect of climate. Hence the faces depicting the prevailing winds surrounding this map. His greatest works include the Almagest, Geographia, and Cosmographia. The Almagest and Euclid's Elements share the distinction of being the oldest scientific texts still in use today. Geographia has eight parts; the last contains detailed instructions for constructing maps of the world. His mathematical proofs that the earth is a sphere are still accepted in spite of the fact that Ptolemy incorrectly believed the earth was a fixed point in the center of a universe revolving around it daily. Ptolemy rejected the theory that the earth rotated around its own axis as fundamentally absurd! All of these works were substantially lost during the Middle Ages. This map is taken from an edition of Cosmographia published in Ulm, Germany soon after the great works were rediscovered during the Renaissance. The 1482 Ulm edition is noticeably different than previous Italian editions because it was printed from carved wood blocks rather than copperplate engravings. This map shows Africa as an extended southern land and the Indian Ocean as inland water.

Nova Anglia, Novvm Belgivm et Virginia; by Johannes Jansson; Amsterdam, Holland;1642 *Courtesy of the Library of Congress, Geography and Map Division, Washington D.C.*

The first maps of the 'new world' were made by Native Americans on skins, bark, rock, and the earth itself. Johannes Jansson's maps were some of the first to be created in Europe. Maps of the new world from this time typically show the east coast and were created in Holland, England, and France. A few maps became prototypes, their general layout was repeated over and over from one mapmaker to another. The 1612 map of Virginia by John Smith and the 1651 map of New England by Johannes Jansson are two that were extensively copied. This map is based on De Laet's rare map of 1630 and is actually the 3rd state of his 1636 Nova Anglia Novvm Belgivm et Virginia which was very influential as it showed the current Dutch holdings all the way from New England to Virginia. Of particular importance to historians is the inclusion of the place names of "N. Amsterdam" (New York) and "Manbattes" (Manhattan). This map was among the first to show New Amsterdam founded less than 20 years before in 1626, and is widely considered to be the first map with the name "Massachusets". The increased interest shown after 1600 by Europeans in the colonization of North America is concisely shown in this map and partially because of this map.

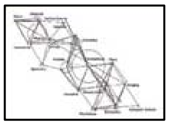
A New Map of the Whole World with the Trade Winds According To the Latest and Most Exact Observations; by Herman Moll; London, England; 1736 *Courtesy of the David Rumsey Map Collection, Cartography Associates*

At the turn of the 17th century, Herman Moll was the most famous map publisher in England. He was the first cartographer to create an elegant map with the correct shape of England. His style is a unique combination of blunt clear lettering with time consuming detail and embellishment out of the way, yet an important part of every map. Moll's maps often include innovative details such as roads and distances between towns; and symbols, for example swords to mark the scenes of famous battles, and other notable events in human history. Moll was charismatic and had interesting friends including Daniel Defoe, Jonathan Swift (he provided maps for Robinson Caruso and for Gulliver's Travels), explorers William Dampier and Woodes Rogers, and the scientist Robert Hooke. He expected excellence from everyone and would attack mapmakers that republish maps by others simply by replacing the title with something new or pretentious without knowing if these map were accurate, or even worse, potentially fatal by showing depths of water or sands that did not exist. This map is a hand colored engraved double hemisphere map of the whole world featuring California as an island, a popular misconception at the time. The continents are represented by the twelve allegorical figures embellishing the bottom along with a series of plants native to each of these places and a lion observing it all. The long note across the top discusses the trade winds which are indicated throughout the map by arrows.



Napoleon's March to Moscow; by Charles Joseph Minard; France; 1869. *Courtesy of Edward Tufte, Graphics Press, Cheshire, Connecticut*

Charles Joseph Minard was a prominent French engineer who lived from 1781 to 1870. He learned to read and write when he was only four years old. He liked to study streams and physics, his profession was to design and oversee civil engineering projects for example bridges, railways, canals, locks, and dams. Figurative maps and graphic tables were among his favorite studies. He did not care for dry and complicated columns of statistical data. He felt this kind of analysis and discussion required a great sustained mental effort, so he substituted images mathematically proportioned so that the first glance could "take it all in". His thematic maps let viewers understand information without fatigue and often lead to unforeseen comparisons. The combination of data map and time-series shown was created near the end of his life. It is probably the best statistical graphic ever drawn. It portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the Polish-Russian border, the thick band shows the size of the army at each position. The path of Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales. A total of six variables are shown in one comprehensive portrait.



1996 Map of Science: A Network Representation of the 43 Fourth Level Clusters Based on Data from the 1996 Science Citation Index; by Henry Small; Philadelphia, USA; 1999.
Courtesy of Henry Small, Thomson Scientific, Philadelphia, Pennsylvania

Dr. Henry Small has pioneered diverse approaches that make it possible to map the structure of science, year-by-year, based on scholarly publications and independent of existing disciplinary categories. In the early seventies, he and Belver Griffith introduced co-citation analysis as an instrument to study the structure and dynamics of science. Today, co-citation analysis is not only used as a research tool for science studies but also as a commercial tool for science policy. Shown is the very first map of all of science. It was generated based on 36,720 multidisciplinary documents for a 15-year sampling window from 1981 through 1995. A combination of fractional citation counting and co-citation clustering via multidimensional scaling was used to extract four nested levels of clustering via single and complete linkage. Major disciplines of science emerge from a bottom-up aggregation of highly cited papers. They are displayed in two dimensions using an order-dependent, geometric triangulation process that produces a unified ordination of a hierarchical arrangement of documents. Each circle contains a map of similar construction at a lower level of aggregation. Circle size corresponds to the number of distinct citations received by documents in each cluster. Links among circles represent aggregate document co-citations. The original map facilitated the interactive exploration of this nested hierarchy of scientific disciplines.

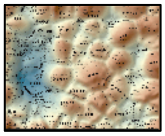
Treemap View of 2004 Usenet Returnees, by Marc Smith. 2005. *Courtesy of the Community Technologies Group, Microsoft Research. Redmond, Washington*

Dr. Marc Smith is a Research Sociologist leading the Community Technologies Group at Microsoft. In the mid-nineties, he started to develop cross-post maps of newsgroup data to measure and map the social structure of the Usene, a large online social space. At Microsoft, he has directed the setup of Netscan for the tracking, analysis, and visualization of multiple terabytes of Usenet data. Netscan is one of the most powerful social network analysis infrastructures in existence today. The visualization portrays the activity of 189,144 newsgroups with 257,442,374 postings in 2004. It uses the treemap layout originally introduced by Ben Shneiderman. Each newsgroup is represented by a square. The size of each square corresponds to the number of people who posted at least twice in 2004. Color coding is used to show the increase or decrease in the number of posters compared to the 2003 data: red indicates fewer postings compared to 2003; green squares denote newsgroups which attracted more posters over time. Each square is labeled and given as a literal hierarchy, i.e., rec.pets “contains” rec.pets.cats. The map gives an efficient overview of the structure and dynamics of the Usenet. The growth of certain newsgroups, e.g., the alt.binaries groups (in the bottom left) and the decline of the comp. groups (in the middle right) can be seen at a glance.

Ph.D. Thesis Map; by Keith V. Nesbitt, Newcastle, Australia, 2004 (© 2004 IEEE).
Courtesy of IEEE and Keith V. Nesbitt, Charles Sturt University, Australia

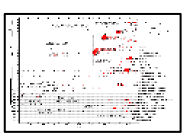
Dr. Keith V. Nesbitt is a lecturer in the School of Information Technology at Charles Sturt University in Bathurst, Australia. The hand-drawn map shows the main interconnecting ideas that run through his Ph.D. thesis. It uses his supervisor's favorite visualization: the London metro map. However, instead of depicting metro lines, it shows “tracks of abstract thought”. Each separate track is given in a different color. Related ideas correspond to nominal (unordered) category stations along that track. Overlapping ideas between tracks are shown as connected stations. The thesis concerns the design of multi-sensory displays of abstract data and introduces three new ideas called the MS-Taxonomy, the MS-Guidelines and the MS-Process. These ideas are demonstrated in a Case Study using stock market data. The motivation for the work is the Data-Mining of Abstract Data. The thesis incorporates ideas from the fields of Virtual Environments, Information Display, Software Engineering and Human Perception. The familiarity of metro maps makes it an easy diagram for readers to interpret.

The space in which the tracks are laid out has no meaning and so it is possible to read the map in any direction. However, there is a cultural bias for the tracks to be followed from left to right and top to bottom.



In Terms of Geography; by André Skupin; New Orleans LA, USA; 2005. *Courtesy of André Skupin, University of New Orleans*

Dr. André Skupin is an Assistant Professor in the Department of Geography at the University of New Orleans. Dr. Skupin's research interests are focused on geographic visualization, cartographic generalization, data mining, and information visualization. This map is a visualization derived from over 22,000 abstracts submitted to the Annual Meetings of the Association of American Geographers during a ten-year period between 1993 and 2002. The aim of this visualization is to give a broad overview of the variety of topics addressed by the geographic community. This is done through a combination of geographic metaphors, cartographic design, and GIS technology. The methodology used to create this map is centered around the representation of each document as an n-dimensional vector of terms. These vectors are used to construct a neural network model of the geographic knowledge domain using the self-organizing map (SOM) method. The neural network is then transformed into two types of information presented here: (1) a landscape in which elevation indicates the degree to which a single, focused topic is addressed; and (2) multi-level text labels associated with regions in the visualization. Final rendering was executed in standard GIS software. André Skupin's work is partially supported by the Louisiana Board of Regents Support Fund, grant #LEQSF (2002-05)-RD-A-34.



Timeline of 60 Years of Anthrax Research Literature; by Steven A Morris; Stillwater; 2005 *Courtesy of Steven A. Morris, Oklahoma State University, Stillwater, Oklahoma*

Dr. Steven A. Morris is a Research Associate with the Department of Electrical and Computer Engineering at Oklahoma State University, Stillwater, Oklahoma. His research is focused on the development of visualization methods and analytical techniques to study collections of scientific papers for technology forecasting and management. This is a timeline map of journal papers about anthrax research. The raw data for this map is a list of references cited by each paper in the collection, which is processed to find the common references cited by each pair of papers, and that information is used to cluster papers into groups using the statistical technique of *agglomerative hierarchical clustering*. Groups of related papers are plotted in horizontal tracks by publication date. A clustering tree on the left shows the structure of the topics, while the topic labels are on the right. Important events in anthrax research are noted on the plot. Importance is measured by the number of times a paper has been cited. Note the emergence of new topics at the bottom of the plot, showing the response of the research community to the anthrax postal bioterror attacks in late 2001. The map uses the visual perception principle of *continuance* to communicate temporal relations among the papers and topic groups*; proximity* communicates the relation of topics, while *size* communicates importance of individual papers. *Hierarchical sequencing* is used to show the overall structure of topics in the anthrax research specialty.

The Structure of Science; by Kevin W. Boyack in Albuquerque, NM and Richard Klavans, Berwyn, PA USA; 2005. *Courtesy of Kevin Boyack, Sandia National Lab and Richard Klavans, SciTech Strategies, Inc.*

Dr. Kevin Boyack is in the Computation, Computers, and Mathematics Center at Sandia National Laboratories using Sandia's VxInsight® knowledge visualization tool. He is interested in semantics, augmented cognition, and the application of mathematical tools to information spaces. The galaxy-like map of science was created using a multi-step process 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 2002. Over time, communities can be born, continue, split, merge, or die. The galaxy map was generated by calculating a disciplinary map based on journals then overlaying research community positions. The disciplinary map was calculated using 7300 journals that formed 671 clusters calculated using the VxOrd layout algorithm. Research community memberships were calculated from coupling coefficients between papers, VxOrd for layout, and a modified single-link clustering routine. Research community positions were then calculated based on the distribution. Thus, communities with all papers in biochemistry journals show up in the biochemistry section of the galaxy map, while communities evenly split between biochemistry and chemistry journals show up midway between the biochemistry and chemistry areas of the map. This is the most comprehensive literature map ever generated.