Michael J Stamper, Chin Hua Kong, Nianli Ma, Angela M Zoss and Katy Börner | Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University, Bloomington, IN
mstamper@indiana.edu; kongch@indiana.edu; nianma@indiana.edu; amzoss@indiana.edu; katy@indiana.edu

MAPSustain

Visualising biomass and biofuel research

Abstract
This paper discusses the visual and interaction design of an online website that supports the interactive exploration of research on biomass and biofuel research. The dataset used covers 13,506 publication records from MEDLINE, the Department of Energy (DOE) and Thomson Reuter's Web of Science (ISI), as well as the United States Patent and Trademark Office (USPTO) patents, and funding by the National Science Foundation (NSF), the National Institutes of Health (NIH), and the US Department of Agriculture (USDA) for the years 1965 to 2010 (see details in Kong et al., forthcoming). The analyses and visualisations reveal where research and development (R&D) takes place in the United States and also in topic space. Results aim to guide researchers gaining an overview of relevant expertise and research, industry representatives interested in identifying potential collaborators and competitors, or agencies interested in supporting or promoting a specific area of sustainability research.

Keywords
geospatial, Google Maps, knowledge mapping, network analysis, sustainability, visualisation

Introduction

The number of scientific publications and patents is increasing continuously; the expertise profiles of researchers are changing over time, making it difficult to identify potential collaborators and synergies. Highly interdisciplinary research areas such as sustainability research might draw and combine approaches and technologies from many different domains making it even harder to keep track, manage, understand and build on existing works and efforts. Financial support might come from different funding institutions, for example the DOE and USDA funded the development of technologies and efficient processes to produce biofuels, bioenergy, and other bio-based products with more than $33 million dollars in 2010 (US Department of Energy, 2010), yet it is desirable to coordinate efforts.

The MAPSustain project presented here extends prior work by Zoss et al. (2010) and aims to help researchers, industry, funders and the general public to navigate information in an easy-to-use online interactive map outlining funding awards, publications produced and patents granted in the fields of biomass and biofuel research. It is also the goal of MAPSustain to foster collaborations and increase the sharing of knowledge on similar research projects. The remainder of the paper is organised as follows. The next section explains the inner workings of MAPSustain, including data preparation and system architecture; subsequently, the MAPSustain interface design and interactivity are presented; finally, the paper concludes with a discussion of related work and an outlook.

MAPSustain data

MAPSustain is drawing upon a dataset that covers 13,506 records on 'biomass' and 'biofuel' research and technology (R&T) retrieved from different publication, patent and funding datasets spanning the years from 1965 to 2010. The three categories come from seven sources:

1 Funding category, consists of NIH, NSF and USDA sources;

Figure 1
MAPSustain geomap interface at http://mapsustain.cns.iu.edu

2 Publications category, consists of DOE, MEDLINE and ISI sources; and

3 Patents, consists of USPTO data.

The different data categories were geocoded and science coded (see details in Kong et al., forthcoming), and overlaid on a geospatial map and a map of science. Figure 1 shows a snapshot of the interface with the map on the top left, a search and filter panel below, and a text panel on the right. The latter has four different tabs that provide more information on the maps and geo/science location used, details on search results, information on the data used, and background information on the project itself.

A user of the site has the choice of using either a geospatial map of the US or the UCSD Map of Science (University of California San Diego). The UCSD map is the product of a large study by researchers at the University of California San Diego using 7.2 million papers and over 16,000 separate journals, proceedings and series from Thomson Scientific and Scopus over the five-year period from 2001 to 2005 (SciTech Strategies Inc., 2010). The researchers used citations between the papers and journals to cluster journals into small groups of highly related journals. The map of science is composed of 554 individual nodes that are connected by links if their respective journal sets are related. Each cluster node is labelled by the content area, shared by the journals in the cluster and colour-coded by its overarching scientific domain. Both maps use Google's map API technology.

The geospatial map shows where research and development (R&D) efforts are conducted. The UCSD Map of Science topic map is used to communicate what topics are covered, for example how interdisciplinary R&D efforts are on a certain topic.

The MAPSustain icons

MAPSustain uses three basic geometric shapes to represent the three categories of information overlaid on both the map of science and the geospatial map. Funding is represented by orange triangles, publications are represented by green squares, and patent information is represented by blue diamonds.

The shapes of the icons were influenced by the principles of Gestalt (Koffka, 1935). Five visual perception concepts seem to be particularly relevant: *similarity*, *continuation*, *proximity* and *figure/ground* and are frequently used in the practice of graphic design (Behrens, 1998). When developing the interface and running through how both the geographic and map of science would be seen and used for MAPSustain, it became apparent that the icons needed to be some of the most basic and non-complex of shapes due to the way that the information on the maps would be rendered and displayed at both the zoomed in and zoomed out states. Specifically, the concepts of *similarity*, *proximity* and *figure/ground* were used in the icon design.

Since the map of science already has circles on it in the form of nodes, circles to represent a category were ruled out in favour of other shapes that would stand out from the maps when they are over-laid on either the geomap or the map of science. The shapes chosen: triangles, squares and diamonds, are at once recognisable as markers for information over both maps.

These shapes allow the user to access and see 'easily and comprehensively' (Keller and Tergan, 2005: 12) the areas of science that have generated more or less funding, publications produced, or patents granted on the map of science, or using the geographic map, the regions of the United States where funding is going or where it has gone, publications produced by researchers, and who has been granted patents for their research.

How the icons function

When the MAPSustain web page is loaded, the geographic Google map is automatically loaded with the most general zoom level and all check boxes in the user control panel selected. Placing the cursor over one of the icons located within a state, a tool-tip window will appear and display either the publications, patents or funding records for that state, depending on which icon the cursor is over. The icons are size-coded, so when they are displayed on a map that is zoomed out, a user can see areas that either have more or less activity and can compare areas of more or less research. This is helpful for someone doing research and is interested to see who is studying and publishing, or how many records of funding are being granted for research and who is receiving the funding in particular geographical areas.

The map of science panel is similar to the geographic map panel, except that users are able to visually distinguish the areas of science where most research is being performed. As on the geomap, when the page is first loaded, the map is presented at the most general zoom level and all boxes in the user control area are checked, giving the user an instant rendering of what research is being performed in what major areas of science. The icons are size-coded on this map like on the geomap, to give the user a visual idea of where and what amount of research is being performed.

An option that the map of science has that the geographic maps does not, is that users have the choice of either seeing the nodes in colour, or in grayscale. At the macro-level, having this option does not make a lot of difference in term of legibility. But, when a user zooms into the map, the 13 disciplinary research areas on the UCSD map of science split into 554 sub-disciplines and the option to view the map in plain black and white helps users more easily distinguish the icons from the nodes when they are layered on top of one another. For example, when a user zooms into the *Medical Specialties*, *Chemistry* or *Biotechnology* sections of the map in the colour-mode, the icons become less apparent. There are two main factors that make this effect happen. The first one is the use of size coding for breaking apart the main groupings of information when the map is viewed from a zoomed out stage, which causes the icons to decrease in size and get 'lost' when they are positioned over a node of the same or similar colour. The second is due to the particular colours chosen for the icons, since because the map uses 13 different colours to distinguish the disciplines of the nodes, it was impossible to select icon colours that were completely dissimilar to the colours already present in the map.

How the MAPSustain interface works

The MAPSustain interface is laid out in two columns for both the geomap and map of science. The first column on the left is the main area that a user interacts with. This column is made up of the Google map API on the top, and right below it are the main categories (Funding, Publications and Patents) of the datasets that a user can use to display desired information in the map area. Below those main choices a user can break down what they'd like to see further in order to get a more detailed view of desired information, having the ability to display amounts, citations, counts and the option to search between a set range of years, and the option to search for keywords.

The column on the right is reserved for displaying textual information, either about MAPSustain itself under the 'maps', 'data', or 'about' tabs, or hyperlinks that correspond to an icon that is clicked on in either the geographic or science map.

Related work

Many examples exist of online maps that support the interactive exploration of complex datasets. Among them are:

- USAID's HealthMap, an interactive map that uses Google Map Javascript API code (HealthMap, 2011). HealthMap is a non-profit project that gathers reports and data from multiple sources to provide a comprehensive, geographically organised picture of what is happening in health and development worldwide, see Figure 2.

Figure 2 HealthMap interface

- Mapping the Measure of America, developed by the American Human Development Project, explores and displays data to indicate populations' overall health, education, and standards of living by measuring the average income for each state, see Figure 3. These numbers are combined, resulting in an overall score on a scale of 1 to 10. Both the maps and interface use a Flash-based design (American Human Development Project of the Social Science Research Council, 2010).

Most of these existing maps use the Google Maps Javascript API or Adobe® Flash as the main development software for the interface and

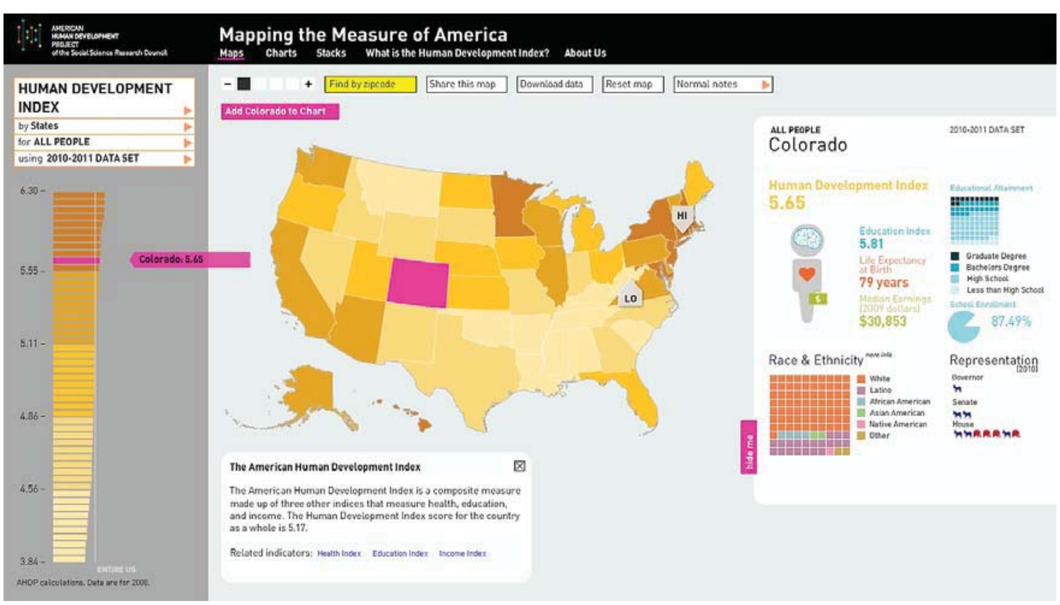

Figure 3 Mapping the Measure of America interface

interaction design and for displaying the data. The current MAPSustain service uses the Google Maps Javascript API code due to its flexibility.

Outlook

MAPSustain not only helps people see the where, what, and who about funding, publications, and patents concerning biofuel and biomass research, but it also provides an avenue to connect and collaborate. By having the information displayed visually over a geographic map, one can easily see the areas of the United States that are getting the most funding, what areas are producing the most publications, and where patents have been granted. If you choose to see more, click an icon and more information will be displayed in the right column. Clicking on the hyperlinks in that column will take a person to a corresponding website (depending on what is selected in the check boxes) that will display even more information about the particular searched subject. Listed within the web page that is brought up, will most likely be the name and affiliation of the researcher(s) who have either received funds, published something relevant, or was/were granted a patent. By knowing this information, one can follow up with the researchers, ask questions, and possibly collaborate on future projects.

Further development in terms of how data is gathered and rendered is needed to make the design of MAPSustain more efficient. One of the drawbacks of the Google maps API is that every time a user zooms into or out of the map, the data needs to be re-aggregated and re-rendered, which can be slow at times, depending on which browser is used and the particular area that is being viewed. In addition, we would like to optimise the legend design so that the smallest icon is more easily legible. Currently, the icon size is correctly calculated and displayed yet further work is needed to automatically generate easy-to-read legends.

Acknowledgements

This project received a great deal of support from members of the Cyberinfrastructure for Network Science Center at the School of Library and Information Science at Indiana University. The work was funded in part by the James S McDonnell Foundation, the National Science Foundation under award CBET-0831636, and the National Institutes of Health under award R21DA024259.

References

Alpert SR (2003) Abstraction in concept map and coupled outline knowledge representations. *Knowledge Creation Diffusion Utilization* 14(1): 31–49.

Alpert SR and Grueneberg K (2000) 'Concept mapping with multimedia on the web', *Journal of Educational Multimedia and Hypermedia* 9(4): 313–331.

American Human Development Project of the Social Science Research Council (2010) *Mapping the Measure of America*. Available at: http://www.measureofamerica.org/maps/ (accessed 5 February 2011).

Behrens RR (1998) Art, design, and Gestalt theory. *Leonardo* 31(4): 229–303. Available at: http://www.jstor.org/stable/1576669 (accessed 10 January 2011).

HealthMap (2011) *Global Health, Local Information*. Available at: http://healthmap.org/en/ (accessed 5 February 2011).

Keller T and Tergan S-O (2005) Visualizing knowledge and information: An introduction. *Knowledge and Information Visualization* 3426(4): 1–23. Available at: http://www.springerlink.com/index/MHNBDL1G9C6W2YCQ.pdf (accessed 11 January 2011).

Koffka K (1935) *The Principles of Gestalt Psychology*. London: Lund Humphries.

Kong CH, Ma N, Zoss AM, Stamper MJ and Börner K (forthcoming) Analyzing and mapping biomass and biofuel science and technology.

SciTech Strategies Inc. (2010) *Maps of Science.* Available at: http://mapofscience.com/wdau.html (accessed 1 July 2011).

US Department of Energy (2010) *DOE, USDA Announce Funding for Biomass Research and Development Initiative*. Available at: http://www.energy.gov/news/8949.htm (accessed 10 January 2011).

Ware C (2004) *Information Visualization: Perception for Design*. San Francisco, CA: Morgan Kaufmann.

Zoss A, Conover M and Börner K (2010) Where are the academic jobs? Interactive exploration of job advertisements in geospatial and topical space. *Third International Conference on Social Computing, Behavioral Modeling and Prediction*, Bethesda, MD, pp. 238–247.