BULLETIN
of the Association for Information Science and Technology

asis&t

DECEMBER/JANUARY 2015

Volume 41, Number 2
ISSN: 1550-8366

See Page 6

SPECIAL SECTION

Mapping Science

DEPARTMENTS

COLUMN

NEXT PAGE >

by Irene L. Travis

IRENE L. TRAVIS
Editor
Bulletin of the Association for Information Science and Technology
Bulletin<at>asis.org

Visualization techniques greatly enhance the comprehension and impact of data. To promote visualization techniques, the *Places & Spaces: Mapping Science* exhibit has been selecting and displaying 10 outstanding maps of science each year since 2005. In 2014 they reached their goal of displaying 100 outstanding maps from 236 map makers. In this issue of the *Bulletin*, curators Katy Börner and Todd Theriault with Kevin Boyack present five of these maps, which are described and discussed by their creators. Types of data mined and combined in these maps include bibliometric and geo-spatial as well as data that is indicative of social priorities such as the activities of not-for-profits. From showing aspects of the invention and development of a particular material (graphene) to modeling distance in terms of travel times in the Roman world, these fascinating maps combine data sources in unique and interesting ways to give us insights otherwise unobtainable.

Visualization techniques depend on large databases, whose curation and use are the core objectives of the ASIS&T research data access and preservation (RDAP) community. In this month's RDAP Review, Wendy Hagenmaier, Dana M. Lamparello, Karen S. Baker, Janina Mueller and Stewart Varner look at the impact of this activity on the archival community.

ASIS&T itself is moving forward on various fronts. Our social media manager, Diane Rasmussen Pennington, discusses the results of her recent survey on social media use in Inside ASIS&T, while our new president, Sandra Hirsch, details plans for action in three areas on her President's Page: strategic planning, membership and promotion of the value of information professionals.

Our next issue will be devoted to the 2014 Annual Meeting in Seattle, which had a large number of pre-conference workshops in addition to its very fine program. It was wonderful to see so many of you there, and I look forward to helping share the experience with all our members. ■

by Sandy Hirsh

SANDY HIRSH
2015 ASIS&T President
Director and professor
School of Information
San José State University
sandy.hirsh<at>sjsu.edu

EDITOR'S SUMMARY
Plans to mobilize the goals of Association president Sandy Hirsh have been put into place with feedback from focus groups and 2014 Annual Meeting attendees and the formation of several task forces. The primary goal is to develop a strategic plan for the Association, ensuring that resources align with the group's mission and members' needs. As the second goal, member attraction, engagement and retention efforts will promote membership among practitioners as a new membership category recognizing applied information science, as well as students and international members. The 2016 Annual Meeting planned for Copenhagen is an expression of the Association's global view. The third goal is for ASIS&T to advocate for the information professions, expanding understanding of professionals' diverse activities and value. ASIS&T members are encouraged to participate in these initiatives designed to strengthen the Association.

KEYWORDS
Association for Information Science and Technology
strategic planning
information professionals
product and service providers
students
international aspects
cultural diversity

Attending the ASIS&T Annual Meeting in Seattle was energizing, and I am excited about the opportunities to work with ASIS&T members this year. I have three major goals as ASIS&T President:

1. Develop a strategic plan for the Association;
2. Identify ways to attract, engage and retain members, especially practitioners, students and international members
3. Advocate for the information professions.

Goal 1: Develop a Strategic Plan for the Association

With the ongoing changes to our field, the range of choices that people have to engage professionally and the importance of charting a clear future for the Association, one of my major initiatives as ASIS&T president this year is strategic planning. Specifically, the strategic planning process will include the following activities:

- Review ASIS&T's mission, vision and goals and translate these into clear actions that can be reported on.
- Get input from ASIS&T members to ensure the Association is meeting membership needs.
- Develop a plan of action, with goals and objectives, that will provide strategic direction for the Association and will position the Association for long-term success.
- Assess resources to better align the Association with strategic goals and objectives.

All of ASIS&T's members are extremely important, and we want to ensure that the Association is meeting all of our members' needs.

The strategic planning initiative got off to a strong start at the Annual Meeting in Seattle. We ran six focus group interview sessions with 50 participants. We also solicited conference attendee feedback on four questions posted on poster boards. We are already learning a great deal from the members who participated in these activities at the Annual Meeting.

I am very pleased that Michael Leach has agreed to chair a new Strategic Planning Task Force and will work closely with the ASIS&T Board to drive the strategic planning process throughout the year. The goal is to develop the strategic plan by next year's Annual Meeting.

There will be more opportunities over the next several months to provide input, and I hope to engage the full spectrum of ASIS&T members in the strategic planning process. There will be additional focus group sessions (including virtual ones), as well as virtual town hall meetings where we will review and discuss draft versions of the strategic plan. I want to encourage you to participate actively in this process. I will be posting periodic updates on the ASIS&T blog and in the *Bulletin* so members can keep up with what is going on with the strategic planning initiative.

Goal 2: Identify Ways to Attract, Engage and Retain Members, Especially Practitioners, Students and International Members

Members are the key to ASIS&T's ongoing success as an association. All of ASIS&T's members are extremely important, and we want to ensure that the Association is meeting all of our members' needs. This year, we will be paying special attention to three membership categories: practitioners, students and international members.

Practitioners: Over the years, fewer information practitioners have chosen ASIS&T as their home. We have an opportunity to explore why this is and take steps to address this gap. The Membership Committee will work on defining new membership categories, in particular, one called *Professionals* that will help practitioners recognize that ASIS&T is for them. The ASIS&T Board is also interested in attracting more practitioners to participate in next year's Annual Meeting. The theme of next year's Annual Meeting in St. Louis is "Information Science with Impact: Research in and for the Community." This theme bridges theoretical and applied research and will be a good start toward striking the right balance to address the varied research interests of ASIS&T membership.

Students: Students are the future of the association. To pay special attention to their interests and needs, I have created a Student Task Force that will be lead by Eric Meyers. The charge for this new task force is to recommend, develop and implement programs and activities that create more meaningful experiences for student members and recent graduates of undergraduate, masters' and doctoral programs.

International: As an international association, it is exciting to look forward to 2016 when ASIS&T will hold its first Annual Meeting outside of North America in Copenhagen. This meeting is a good start toward welcoming the full range of ASIS&T members, wherever they happen to live around the world. To ensure that all members feel welcome on the ASIS&T website, the International Relations Committee, chaired by Diane Sonnenwald and Mei-Me Wu, will review the new ASIS&T website to ensure the content on the site reflects our global association's membership.

I have also created the Outreach and Engagement Task Force, lead by Dirk Lewandowski, to identify, recommend and implement ways to collaborate, engage and participate in the association through in-person and virtual activities. The objective is to ensure members can participate in and feel connected to ASIS&T – even if they can't attend Annual Meetings – and to provide additional ways to interact with the Association throughout the year. This year, Dirk will work with the International Relations Committee to focus on new ways to reach out and engage with members across our international association.

Goal 3: Advocate for the Information Professions

As the leading association for information science, representing information professionals from around the world, ASIS&T can help broaden people's understanding about what information professionals do, who they are and what their value is. The Information Professionals Task Force will create a web presence – building on an earlier prototype developed by Marcia Bates, Prudence Dalrymple and me – that highlights the diversity and value of the information professions and increases awareness and interest among the general public, employers, students and other information professionals. Specifically, the goals are to advocate for and market the information professions, consolidate a professional identify and benefit ASIS&T by building awareness of the Association. I am very pleased that Nancy Roderer, the original chair of the Information Professionals Task Force, has agreed to co-chair this effort with Diane Rasmussen Pennington, who has served as the chair of the Web Presence Task Force. I think their combined skills and expertise will lead to a tremendous end product that will benefit the information professions as a whole, as well as ASIS&T.

In sum, we have a very busy and active year planned! I look forward to working with ASIS&T members to build on the strong foundation we already have and to help define the ongoing success of the Association. These activities and initiatives, taken together, should put ASIS&T on a strong path toward deeper member engagement and participation. I hope you will participate actively in the process. ■

Inside ASIS&T

ASIS&T ANNUAL MEETING COVERAGE

Looking Back; Looking Ahead

In keeping with our annual tradition, the bulk of our reporting on the 2014 ASIS&T Annual Meeting – including both photographic and substantive coverage – will be included in the February/March 2015 issue of the *Bulletin*. In this issue, you'll find a list of the winners of the 2014 ASIS&T Annual Awards for which more details will be provided in the next issue.

When one Annual Meeting concludes, planning for the next is already well underway. Committees are hard at work seeking submissions for all aspects of the technical program, suggestions for pre- and post-conference undertakings and ideas for social events and new ways to engage attendees in networking and team-building activities. Mark your calendars for November 6-10, 2015, and plan to join the festivities at the Hyatt Regency in St. Louis, Missouri.

The theme for the 2015 Annual Meeting is *Information Science with Impact: Research in and for the Community*. As always, the conference committees welcome contributions from all areas of information science and technology. For more information about submission ideas, formats and deadlines, please visit the ASIS&T website at www.asist.org.

78TH ASIS&T ANNUAL MEETING
November 6-10, 2015
St. Louis, MO USA

INFORMATION SCIENCE WITH IMPACT: RESEARCH IN AND FOR THE COMMUNITY

Conference Committees

Lisa Given, Charles Sturt University, conference chair
Brian Detlor, McMaster University, and **Hazel Hall**, Edinburgh Napier University, papers co-chairs
Heather O'Brien, University of British Columbia, and **Alison Brettle**, University of Salford, panels co-chairs
Lynn Westbrook, University of Texas at Austin, and **Michael Khoo**, Drexel University, posters co-chairs
Richard Hill, ASIS&T executive director, workshops/tutorials, local arrangements, logistics ■

2014 ASIS&T Award Winners

AWARD OF MERIT: **Marjorie M.K. Hlava**, Access Innovations, Inc.
WATSON DAVIS AWARD: **Vicki Gregory**, University of South Florida
RESEARCH AWARD: **Diane Kelly**, University of North Carolina at Chapel Hill
THOMSON REUTERS OUTSTANDING INFORMATION SCIENCE TEACHER AWARD: **Michelle Kazmer**, Florida State University
BEST *JASIST* PAPER AWARD: **Suzan Verberne**, **Maarten van der Heijden**, **Max Hinne**, **Maya Sappelli**, **Saskia Koldijk**, **Eduard Hoenkamp** and **Wessel Kraaij**, *Reliability and Validity of Query Intent Assessments*
PRATT SEVERN BEST STUDENT RESEARCH PAPER AWARD: **Curt Arledge**, University of North Carolina at Chapel Hill, *Filled-in vs. Outline Icons: The Impact of the Icon*
BEST INFORMATION SCIENCE BOOK AWARDS: **Robert J. Glushko**, *The Discipline of Organizing*, The MIT Press, and **Karine Nahon** and **Jeff Hemsley**, *Going Viral*, Polity
THOMSON REUTERS DOCTORAL DISSERTATION PROPOSAL SCHOLARSHIP: **Tiffany Chao**, University of Illinois at Urbana-Champaign, *Methods Metadata: Curating Scientific Research Data for Reuse*
PROQUEST DOCTORAL DISSERTATION AWARD: **Amelia Acker**, *Born Networked Records: A History of the Short Message Service Format*
NEW LEADERS AWARD: **Agnes Mainka**, **Maric Kramer**, **Anne Pepitone**, **Jeremy L. McLaughlin**, **Karen Miller**, **Emily Vardell**, **Devon Greyson** and **Stephann Makri**
CHAPTER-OF-THE-YEAR AWARD: **New England Chapter of ASIS&T** (**NEASIST**)
STUDENT CHAPTER-OF-THE-YEAR AWARD: **Simmons College**
SIG-OF-THE-YEAR AWARD: **SIG/Digital Libraries** (**SIG/DL**)
SIG MEMBER-OF-THE-YEAR AWARD: **Abebe Rorissa**, SIG/III
SIG PUBLICATION-OF-THE-YEAR AWARD: **SIG/Social Informatics** (**SIG/SI**), *Social Informatics: Past, Present and Future*

SIG/III Announces Winners in International Paper Contest

Winners of the 13th annual International Paper Contest sponsored by **ASIS&T Special Interest Group/International Information Issues** (**SIG/III**) are **Bhakti Gala**, India for *Blogs, Bloggers and Scholarly Publications*, first place, and **Muhammad Javed Iqbal**, Pakistan, for *Scholars' Perceptions about HEC Library Resources and Services: A Quantitative Study of User Satisfaction*, second place.

The principal authors of the winning papers receive a two-year individual membership in ASIS&T, as well as financial support to attend the ASIS&T Annual Meeting. Winning papers are also considered for publication in the *International Information and Library Review*.

Jurors for the 2014 contest were **Maqsood Shaheen**, chair, and **Fatih Oguz** and **Alma Rivera**.

ASIS&T's Social Media Presence: Where Do We Go from Here?

by Diane Rasmussen Pennington

EDITOR'S SUMMARY

As part of ASIS&T's renewed attention to social media, the author extended her professional expertise with the communication channel to serve as the Association's social media manager, joining five contributors in examining members' practices and possible future directions. The Association maintains multiple social media outlets, including a listserv, accounts on Facebook, Twitter and LinkedIn, and a blog, and uses third party tools for metrics and analytics. An August 2014 survey garnered responses from 76 individuals, 58% of whom were members. Respondents clearly favored Facebook, followed by LinkedIn and Twitter, and used the services for networking and sharing resources and organizational announcements. Engaging topics are most likely to motivate participation. Responses revealed clear differences of opinion about ASIS&T's online offerings, while 26% were unaware of ASIS&T social media outlets. With the overhauled ASIS&T website, social content is prominently displayed and easy to engage with. Further strategic planning by Association president Sandy Hirsh will guide future social media activity. The author offers assistance to SIGs, chapters and committees to develop their social media presence and invites members to contribute to The ASIS&T Blog and mention ASIS&T in tweets and in Facebook and LinkedIn comments.

KEYWORDS

Association for Information Science and Technology

strategic planning

social web

computer mediated communications

surveys

It is an honor to serve as the new social media manager for the Association for Information Science and Technology. I've taught courses in social media for libraries and other organizations since 2007, I do research related to social media and I edited a book called *Social Media for Academics: A Practical Guide*. Despite my experience, I never imagined I would have the opportunity to run the social media efforts for my primary professional association. I promise to continue putting forth my best efforts in this important role for our Association.

This article will outline where we are with our social media efforts as of October 2014. It will also suggest directions about where we can go next based on data and observations I've gathered since I took over this role. I wanted to write this article because I believe strongly in transparent leadership; additionally, I know that many other members are knowledgeable about social media, so I think it is appropriate to keep everyone informed.

Where Are We?

The team: Joining me in my role as social media manager position is a group of contributors who have a great deal to offer us. As a result of the Board-led "Engagement across Boundaries" initiative, we have secured five social-media savvy contributors who are working on their own ASIS&T social media initiatives and posting to our accounts as they see fit. These social media contributors are Laura Christopherson, Kate Dillon, Tamara Heck, Sara Mooney and Adam Worrall. The Board-approved Social Media Advisory Committee consists of me as chair, Naresh Agarwal and Heather

Diane Rasmussen Pennington is ASIS&T social media manager and an assistant professor at Ashford University. She can be reached by email at diane<at>asis.org or follow her on Twitter @infogamerist.

Pfeiffer; we worked especially closely together when we first implemented the initiative.

Listserv: For many years, ASIS&T has had a listserv called asis-l. In earlier days, it was perhaps more of a discussion forum, but it has evolved into a list for announcements. Subscribers get information about calls for papers, conferences, ASIS&T-sponsored events and so on. These announcements are useful for our professional advancement, but it doesn't really serve us well for discussions anymore.

Facebook: Our Facebook page has just over 750 "likes." Facebook page managers can collect a variety of other data about the page, including how many people are reached with each post and how many "engagements" are garnered on each post, such as the number of "likes," comments and shares.

Twitter: @asist_org has close to 4,000 Twitter followers. We gain new followers almost every day, and certain tweets get a respectable number of retweets. @asist_org gets mentions in certain contexts from other accounts, such as when a SIG or chapter is making an announcement or when a member has a question for ASIS&T.

LinkedIn: The LinkedIn group has almost 2,200 members. This group doesn't see as much interaction as our other accounts, but the diversity of the group's members is striking, including people in industry and people from many countries.

Blog: I launched The ASIS&T Blog in July 2014. As of October 2014, it has had around 2,000 unique views since its inception. I hope to grow this blog presence substantially over the next few months because it can be a powerful but informal tool for distributing association news and discussing topics related to information science.

Third-party tools: The contributors and I are currently experimenting with Hootsuite and Buffer to see what works best for coordinating our combined efforts. Hootsuite, Buffer and ReadyGraph are all proving helpful in creating analytics for gaining insight into our social media metrics.

Social Media Survey: Results

These numbers about our followers and what they're doing with our tools are useful, but they only tell part of the story. When I first took this position, one of my primary concerns was whether ASIS&T's social media efforts were meeting the needs of its followers. I heard anecdotal concerns from Association leaders that our individual platforms don't coordinate with each other well. I also noticed that there was not much actual engagement from followers, so how could content be presented in ways that would increase two-way activity? I decided to launch a survey to see what people thought about our social media efforts and what we could do to better meet their needs.

Adam Worrall, ASIS&T social media contributor, played a substantial role in developing the survey. I invited people to participate using our existing social media channels, and the survey was available for approximately the last two weeks of August 2014. I received 76 completed responses. A summary of the results follows.

Which social media services do you use? 88% use Facebook and approximately 70% use LinkedIn and Twitter. Less than half use listservs; around 30% chose Google+, ResearchGate and academia.edu. Write-in responses included last.fm, Instagram, Tumblr, Pinterest, Mendeley, Path, Flickr and xing.de.

How often do you use specific social media services? Around 70% of Facebook users check it multiple times a day, and about 40% of Twitter users check it just as often. Activity frequency drops significantly for the other services listed.

Which social media service do you use most frequently? 57% said Facebook, 24% selected Twitter, 11% chose listservs and 6% said LinkedIn. When asked why that service is used most frequently, participants provided a range of open-ended responses:

- Facebook: more of their connections are there than on any other service; to keep up with friends and family as well as (to a lesser extent) colleagues
- Twitter: brevity of posts makes it easy to keep up; greater control over what is presented in the news feed; useful for news gathering
- Listservs: convenience, content, habit, human-human interaction rather than human-machine interaction
- A smaller number of participants discussed preferences for ResearchGate, LinkedIn and Xing due to their professional advantages.

Others chose Tumblr for its variety of content, Google+ for file sharing and Instagram because it is less personal than other choices.

- One senior participant provided a detailed response: "Facebook is more versatile and friendly. It even duplicates some of what the other services offer. Twitter is too confining and requires too many infantile iconic representations. The others mentioned are of less interest to me now that I'm retired. My consulting activities can be satisfied by other means."

Why do you use social media in your personal life? Most often, participants use it in their personal lives to keep in touch with friends and family as well as share links, photos and videos. Discussions were less commonly chosen.

Why do you use social media in your professional life? The most popular responses included networking, sharing resourcesand sharing organizational announcements. Discussions, photo and video sharing, and life update sharing were not as popular.

What is your primary motivation for using social media? By far, the most popular open-ended response related to keeping in touch with family, friends and/or colleagues. Slightly less common motivations included consuming information and sharing information; these two were distinct and popular categories. Less frequently noted motivations included enjoying the speed and convenience of communication, accomplishing collaborative work and participating in discussion.

What motivates you to engage in back-and-forth interaction or discussion via social media? Overwhelmingly, participants said that the topic encourages them to participate more than anything else. Other common reasons included the enjoyment of participating in high-quality discussion (respectful, high-level, etc.), connecting with others and for completing professional tasks. One participant responded, "No alternative since many folks today don't know how to use paper, pen, envelope and stamp."

Do you belong to ASIS&T SIGs, chapters or committees? 87% belong to one or more SIGs, 26% belong to a student chapter, 26% belong to a regional chapter and 20% participate in one or more committees. Of these, 37% follow ASIS&T-wide social media venues and at least one SIG, chapter or committee.

Which ASIS&T social media venues do you visit, and how often do you visit them?

- 21% only follow ASIS&T-wide channels, and 26% don't follow any of our social media outlets.
- About half of those who follow both ASIS&T-wide presences *and* SIG/chapter/committee sites visit them once a week, while others in this category read them either once a month or rarely. From most popular to least popular, they selected listservs, Facebook, Twitter, The ASIS&T Blog and LinkedIn.
- Of those who follow only the ASIS&T-wide presences, 31% read them multiple times a day, 31% read them once per week and 31% visit them rarely. From most popular to least popular, they selected Twitter, Facebook, LinkedIn, The ASIS&T Blog and listservs.
- Some participants indicated that they do not follow our social media accounts. The most frequent reason was because they didn't know they existed. Less frequently, they don't feel the content is relevant to their field or their interests. One person noted, "I'm not a social media person. I don't go on Facebook or Twitter. I hate the permanent exhibition of one's life. It's just another gadget that distracts our attention from the more important things. Emails and mailing lists work fine for me."

Do you feel that ASIS&T social media presence is coherent? Please explain why or why not. We received a wide range of answers to this question. Some felt it was coherent; others said it was not. Some people like the type and amount of content that is provided, but other participants want more variety and more information. Respondents raised questions around the meaning of the term coherent; some saw it as visual identity, others interpreted it as providing the same material on all sites, and so on.

What one thing would you like to see ASIS&T do more of in its use of social media? Some participants were happy with the current type and frequency of activity. Others had various suggestions: it should be more engaging, create more community, appear less formal, include more content in type and frequency, post jobs and event announcements, highlight things in the field and share activities of its various groups.

What one thing would you like to see ASIS&T do less of in its use of social media? Many people did not provide substantial responses to this question. The very few that did answer provided a wide range of ideas: be less formal, reduce repetition and appear less scattered.

Where do you live? 79% of participants live in North America, 16% are in Europe, 3% live in Asia and 1.5% live in Oceania/Australia.

Are you an ASIS&T member? 58% of respondents are members, 28% have never been a member, 9% used to be a member and 4% are unsure.

Do you serve as an officer with an ASIS&T SIG, chapter or committee? 57% were not current officers. The others indicated a range of officer positions; the most popularly chosen positions were membership and past chair.

Social Media Survey: Discussion

Not surprisingly, the most common reasons that participants use social media include connections with other people. Sharing and receiving information on topics of interest are important motivators to this group as well.

The results also reveal differences in opinion, especially as things relate to ASIS&T's online offerings. Some people are happy with relatively more traditional email and listserv communications, while other participants want more social media posts and increased engagement.

We can see from these informally presented results that while many people use social media for similar reasons, they also express a diverse group of desires. This finding brings up interesting questions for our social media practices in the future.

We can see from these informally presented results that while many people use social media for similar reasons, they also express a diverse group of desires. This finding brings up interesting questions for our social media practices in the future. Not very many people indicated that they participate in discussions online, but interesting topics do motivate some people to participate; therefore, is it practical to spend too much time providing items for discussion? Should we focus more on facilitating personal and professional connections among members? These questions have yet to be answered, but we do have some firm plans moving forward.

Where Do We Go From Here?

It stands to reason that a specialized non-profit organization without a full-time social media staff member will not grow as quickly or be as active as a large corporation. For example, Target's Facebook page has gained over 23,000,000 "likes," and Walmart has over 591,000 Twitter followers. That said, the contributors and I are dedicated to developing ASIS&T's online presences as fully as possible within our relatively limited time and resource constraints. Also as with any organization, we do not work in isolation. Fortunately, we have a number of things working in our favor as we move forward.

From a technological perspective, the new ASIS&T website integrates social content much better than the old one. At any time, an individual can visit the home page and see our most recent tweet. Links to the ASIS&T-wide social media venues are displayed prominently at the top of the page. The ASIS&T Blog is more closely integrated with the new site, since everything is all based on the same WordPress installation. SIGs and chapters will have an easier time updating their blogs as well due to the provided templates.

President Sandy Hirsh's strategic planning initiative will also help inform future social media practices. The direction will become clearer over the next year.

Every one of our venues continues to gain followers almost daily, so that tells me we are doing at least some things right. While it's clear from the survey results that not all of our members care about our social media presences, the best thing we can do is focus on those who do want to read and engage with our content. ASIS&T has so much to offer as an organization; our publications, our conferences and our members provide endless inspiration for younger members and intellectual outlets for long-standing members. In practice, it's my job to lead efforts toward offering that content in ways that make it accessible to anyone who is interested in the places where they already live online.

The social media contributors and I are available to support you and your organizational needs. If your SIG, chapter or committee would like some assistance starting or revitalizing its social media presence, please contact me. If you have ideas about how to make our virtual communities more vibrant, let me know about that as well. At the same time, the potential for social media success resides with every member. If you would like to write a post for The ASIS&T Blog, please email me, and I will create an account for anyone who is interested. If you tweet interesting news items or ideas related to information science, mention @asist_org in your post. Feel free to share stories or ideas on our Facebook page or LinkedIn group as well.

See you online!

Please join us on our social media channels

Twitter: @asist_org
Facebook: www.facebook.com/asist.org
LinkedIn:
www.linkedin.com/groups/Association-Information-Science-Technology-44332
Blog: www.asist.org/SocialMedia ■

For further information

Rasmussen Pennington, D. (August 28, 2014). *ASIS&T's social media presence: Channels, communications and challenges* [Video file]. Retrieved from www.asis.org/Conferences/webinars/Webinar-8-28-2014_Social_Media_Presence.html

Rasmussen Neal, D. (Ed.) (2012). *Social media for academics: A practical guide.* Oxford: Chandos.

Mapping Science Introduction: Past, Present and Future

by Katy Börner, Todd N. Theriault and Kevin W. Boyack

EDITOR'S SUMMARY

From early cartography to modern science maps, visual presentations facilitate understanding of large amounts of data. A traveling exhibit entitled *Places and Spaces: Mapping Science* has presented outstanding maps illustrating different designs and applications since 2005. The 10th year of the exhibit focuses on the future of science mapping and features five maps described in this special section of the *Bulletin*. Topics include the history of physics and key contributors, the development of the Internet and the structure of fields and topics in science and technology. Each emerged from latent relationships among elements in large volumes of data, made clear through visualization in an easily understandable format. Given high quality data, processing tools, design and analysis expertise and research funding, science mapping can be expected to expand in application and usefulness. Key challenges include insufficient numbers of experts, lack of sophisticated tools, low literacy in data visualization and absence of design standards.

KEYWORDS

data maps
scientific and technical information
electronic visualization
scientometrics
trends

Katy Börner is the Victor H. Yngve Professor of Information Science in the Department of Information and Library Science, School of Informatics and Computing, and founding director of the Cyberinfrastructure for Network Science Center at Indiana University. She and Todd N. Theriault are the curators of *Mapping Science*, the exhibit from which this material comes. They can be reached at kboyack<at>mapofscience.com and ttheriau<at>indiana.edu.

Kevin W. Boyack is a researcher at SciTech Strategies, Inc. He can be reached at kboyack<at>mapofscience.com.

For centuries, visual depictions of data and of the world have been used to understand and communicate information in novel ways. A prime example is cartographic maps of earth and water that have guided human exploration. They have marked the border between the known and the unknown, firing the imagination and fueling the desire for new knowledge and new explorations. Over time, geographic maps have become more detailed, more accurate and more sophisticated. Despite these advances, the thirst for exploration and discovery, along with the need for maps to guide our travels, remains undiminished.

Today, our opportunities for discovery reside less in physical places than in abstract information spaces that are ever growing, ever changing. Indeed, *big data* has become a buzzword that crosscuts all fields. But while search engines can retrieve facts from oceans of data, they cannot answer larger questions about those data as a whole: How big is this ocean? How can we navigate to the useful islands of knowledge and wisdom? How are knowledge and expertise interlinked on a global scale? In which areas is it worth investing time, effort, resources and compassion? Advances in science and technology are increasingly dependent on effective navigation and management of information spaces. But how do we make sense of all the data, information, knowledge and expertise that is relevant for our daily decision-making?

The field of information science is well situated to make contributions to all fields that are awash in data and information. Coping with this flood is where science maps and, more generally, visualizations can help. They serve as visual interfaces to immense collections of data, depicting myriad objects in ways that allow us to effectively discern apparent outliers, clusters and trends. This issue of the *Bulletin of the Association for Information Science and Technology* focuses on science mapping. It presents five short papers that provide detail on recent maps from the *Places & Spaces: Mapping Science* exhibit. While the exhibit aims to introduce science mapping techniques to the general public and to experts across diverse disciplines for educational, scientific and practical purposes, this issue of the *Bulletin* aims to make the purposes, methods and lessons learned from science mapping more transparent to the information science community. Although detailed recipes for creating visualizations are not presented here, our hope is that the suite of examples given in the following papers will provide a sense of the types and scope of analyses that can be done using science mapping and visualization techniques. Perhaps these authors and their works will inspire students and practitioners in information science to make greater use of these tools [1].

Places & Spaces: Mapping Science Exhibit – 100 Maps from 236 Mapmakers in 10 Years

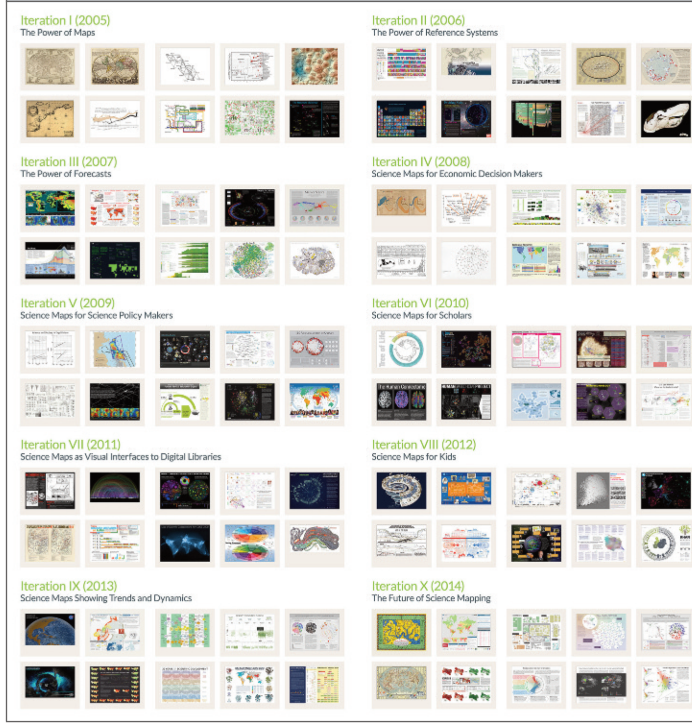
The *Places and Spaces: Mapping Science* exhibit began in 2005 with an initial set of 10 maps from 12 mapmakers (see Figure 1, left) and a 10-year goal and plan to curate 10 exemplary maps each year (through competitive selection), ultimately resulting in an exhibit with 100 maps. This goal has been reached; the 10th annual iteration of the exhibit is now complete and the exhibit features 100 maps from 236 mapmakers that render science and technology data into actionable insights (see exhibit display at the University of Miami in Figure 1, left, and thumbnail versions of all maps in Figure 2). Each iteration of the exhibit has had a different focus, and the resulting maps provide a very rich set of exemplars of the design and uses of science mapping and visualization. Although only five maps are discussed in the following papers, detail on all 100 maps is available at http://scimaps.org. Background on the processes used in many of these maps and visualizations are also described in [2, 3].

How Are Science Maps Used?

The last 10 years have seen a Cambrian explosion of science maps. While a timeline of major maps from the 1930s to 2007 has been compiled previously [2], it would be difficult if not impossible to map all the diverse developments across the sciences since 2008. The *Mapping Science* exhibit promotes map-making excellence, and the different themed iterations that introduce science maps for different stakeholders – from science policy makers, industry leaders and scholars to kids and the general public – have introduced the value and utility of topic maps to many.

FIGURE 1. First public display of exhibit maps at the 101st Annual Meeting of the Association of American Geographers in Denver, Colorado, April 5-9, 2005 (left), and display of 100 maps and public lecture at the Jorge M. Perez Architecture Center, University of Miami in Miami, Florida, September 4–December 11, 2014 (right); photograph is the courtesy of University of Miami Communications.

FIGURE 2. 100 maps in 10 Iterations over 10 years.



The 10th iteration of the exhibit focuses on the future of science mapping. It showcases some of the best science maps in existence and reviews existing challenges and opportunities. The exhibit maps are large, high-resolution graphics – they are 24" x 30" (76 x 61 cm) and thus cannot be properly reproduced in a standard-size scholarly journal. All of the maps aim to help us understand and communicate complex S&T data. Many of the maps represent advances in data federation, mining and visualization approaches. Ultimately, each map tells a story. The best of these maps tell stories that can be understood by both experts and novices.

The 10 maps in the 10th iteration address a variety of topics and are used to tell stories about a variety of questions. One key feature of many maps is that they provide a broad context within which specific points can be addressed. Here are several examples:

- A map of physics (http://scimaps.org/mapdetail/being_a_map_of_physi_171) shows key points and people in its historical development from the 6th century to the present. The broad perspective not only enables a view of the entire field, but also allows key theoretical starting points, streams of thoughts and key figures to be seen within the context of the whole.

- Exploring a highly detailed map of the Internet can easily be consuming. Such a map (http://scimaps.org/mapdetail/map_of_the_internet_172) features old and new continents of companies, websites, services and social media. This map confirms relationships that we may already know about while also surprising us with unexpected connections. The best maps present familiar knowledge in a way that is consistent with our understanding and then build off that familiar knowledge to suggest things that we did not know. Indeed, hypothesis generation is one of the key uses of science maps.

- Maps of sets of documents, such as scientific papers or patents (see papers by Klavans & Boyack and map at http://scimaps.org/mapdetail/exploring_ the_relati_180; Kay et al. and map at http://scimaps.org/mapdetail/mapping_ graphene_sci_179) or of grants (http://scimaps.org/mapdetail/visual_ funding_portf_178), are often used to show the structure of science and technology and the relationships between fields and topics. Large-scale maps of papers, patents or grants can then be used as basemaps or templates upon which other information can be overlaid. For example, the maps of Kay et al. (http://scimaps.org/mapdetail/mapping_graphene_sci_179) show that while the science of graphene is relatively focused in materials science, its application space as measured by patents is much broader.

Science maps often take advantage of geographic maps and geographic information systems to situate scientific data within the geographic contexts that are familiar to us all. For example, geographic maps can be used to show the potential spread of pandemic threats (http://scimaps.org/mapdetail/predict_healthmap_173) and demonstrate how different interventions can change the timing and breadth of that spread. They can also be used in concert with a flowmap metaphor to illustrate the impact of different variables on travel times in both current and historical contexts (see the paper by Meeks and map at http://scimaps.org/mapdetail/orbis_174).

Maps can also be used to uncover relationships between a multitude of different object types. For instance, maps can show the relationships between words and their usage (see the paper by Bertin et al. and map at http://scimaps.org/mapdetail/the_linguistic_conte_177) or between organizations of different types (see the paper by Oberg et al. and map at http://scimaps.org/mapdetail/interstitial_organiz_181), thus revealing patterns that can help us to better understand our world. Maps may also work on multiple levels – from the micro (individual) level to the macro (population) level – revealing patterns and trends across and between multiple scales. An example is the money map by Randall Munroe (http://www.scimaps.org/mapdetail/money_176) that shows how economic resources are used and distributed at the one dollar to trillion dollar levels.

This issue includes maps by **Meeks, Kay et al.**, **Bertin et al.**, **Klavans & Boyack** and **Korff**. Ideally, a close examination of different maps will empower and inspire individuals to tell their very own stories about the enormous complexity, sheer beauty, practical utility and societal value of science and technology.

Opportunities and Challenges

The design of actionable science maps requires four critical ingredients: high-quality data; algorithms and tools to process this data; expertise to design valid workflows and to interpret results; and funding to pay for personnel, data storage, compute power and so forth. These four ingredients are discussed below.

In the last 10 years, the size of datasets, their coverage (for example, book data can now be used to map scholarly activities of arts and humanities scholars) and interlinkage (for example, http://linkeddata.org interconnects many previously isolated data silos, while the International Researcher Network at http://nrn.cns.iu.edu allows anyone to browse publication, teaching and funding data from institutions around the globe) have increased enormously. Thanks to the open access movement, many high quality datasets are becoming available online free of charge (such as MEDLINE publications at http://nlm.nih.gov). Increasingly, data covering other areas of scholarship and creative activity are becoming available, including datasets of genes, proteins, diseases, films and film characters, music, photos and social media data.

The number, sophistication and scalability of data processing, mining, modeling and visualization algorithms have improved substantially as well over the past decade. Many different open-source libraries, tools and services exist to perform temporal, geospatial, topical and network studies. Several of the tools come with extensive online documentation on how to run expert-validated workflows (that is, sets of algorithms that are executed in a well-defined sequence with specific parameter values) so that users can visualize their own data using advanced workflows. While many government agencies, researchers and practitioners are not able (or willing) to share their data, they can now apply the very same workflows and compare results across institutional and disciplinary boundaries.

The number of experts able to advance data mining and visualization research and development has grown enormously but not as fast as the need to render data into actionable insights. The McKinsey Global Institute (2011) forecasts a 50-60% gap between the supply and demand of people with deep analytical talent and projects. They project that by 2018 the United States may experience "a potential shortfall of 1.5 million data-savvy managers and analysts" [4]. Domain expertise is needed to validate and interpret S&T maps and to tell the stories that make their content relevant for many. The design of truly actionable science maps frequently requires a close collaboration between data analysts and domain experts.

Funding and attention for the science of science research has increased enormously over the last 10 years. The National Science Foundation's Science of Science and Innovation Policy (SciSIP) program has co-funded 108 projects since 2009 alone, with a total award amount of about $58 million dollars. The recent 1 billion euro FutureICT flagship project proposal brought together hundreds of the best scientists from Europe and elsewhere to design a data-simulation-visualization platform that can help accelerate science, technology and innovation (http://www.futurict.eu). In a time of tightened budgets, federal and personal funds need to be spent more wisely, and accountability for the impact of investments becomes more important. We note, however, that visualizations can and should be a part of projects funded across all of science and not just by projects aligned with policymaking.

While access to the four critical ingredients is getting easier, there are a number of serious challenges that need to be addressed to make science maps truly useful for different stakeholders. First, insight needs and datasets are growing faster than both the number of experts and the scalability and sophistication of algorithms and tools. Varied algorithms exist for the study of heterogeneous networks (networks with multiple node types and link types such as authors and papers with author-author, paper-paper and author-paper links). However, fewer approaches exist to study problems that are based in multi-level networks (for example, the impact of a certain nationwide policy decision on the career trajectories of individual researchers).

Second, data visualization literacy, the ability to read and make visualizations, is rather low. A recent study of 900 youth and adult visitors across five science museums in the United States revealed that many people cannot interpret basic data visualizations, such as scatter plots or geospatial maps with data overlays. Asked to read more advanced visualizations such as network layouts, many had no understanding of or vocabulary to refer to key features such as high-degree nodes or network clusters. Additional studies showed that the composition tasks – asking visitors to compose visualizations from multiple data layers – led to a higher level of understanding than asking them to decompose visualizations. This finding is one of several reasons why Phase II of *Mapping Science* (coming soon) will focus on interactive visualizations and invite submission of tools that empower users to render data into insights. That is, future iterations of the exhibit will not only invite a general audience to view and read S&T maps but also to make S&T maps. It is our hope that the ability to see the impact of different data cleaning, analysis and visualization algorithms will help to increase the data visualization literacy of users. The maps and tools will be made available in formal education in schools and informal encounters in (science) museums, libraries and other public spaces.

Third, few guidelines and even fewer standards exist for the design of science maps (but also for science metrics or science classifications). Just like early maps of the world did not place north on top, used different color schemas and did not show latitude and longitude information, today's maps of science come in all shapes and forms. However, many science maps seem to have the very same general structure – a comparison of 20 maps of science generated from different datasets manually or using diverse algorithms showed a very high level of correspondence [5]. The UCSD science map and classification system [6] is now used across 10 different tools and online services, and the color schema for the different scientific disciplines has been adopted for several other maps. Work on multi-level classification systems and a map of science that also covers book and proceedings data is in progress. Ultimately, it seems highly desirable to have free datasets, tools and workflows in place that anyone can plug-and-play to replicate and advance the generation of science base maps and data overlays. Different stakeholders and application domains might require different base maps. However, concordances (also called *alignments* or *mappings*) between classification system and maps will make it possible to visualize the very same dataset on different maps, enabling comparisons and human subject studies that can increase our understanding of what visualization to use when.

Additional opportunities and challenges as well as future works are discussed in the papers presented in this special section of the *Bulletin*. ■

Resources Mentioned in the Article

[1] Börner, K., & Polley, D. E. (2014). *Visual insights: A practical guide to making sense of data.* Cambridge, MA: The MIT Press.

[2] Börner, K. (2010). *Atlas of science: Visualizing what we know.* Cambridge, MA: The MIT Press.

[3] Börner, K. (2015). *Atlas of knowledge: Anyone can map.* Cambridge, MA: The MIT Press.

[4] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011). *Big data: The next frontier for innovation, competitiveness and productivity.* New York: McKinsey Global Institute.

[5] Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology, 60*(3), 455-476.

[6] Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., . . .& Boyack, K. W. (2012). Design and update of a classification system: The UCSD map of science. *PLOS One, 7* (7), e39464.

The Design and Implementation of ORBIS: The Stanford Geospatial Network Model of the Roman World

by Elijah Meeks

EDITOR'S SUMMARY

ORBIS is a geospatial model of the Roman world representing the network of cities and travel routes that enabled movement across the Roman Empire. It is an example of neogeography, use of geographic information systems and mapmaking techniques by non-experts. ORBIS was created using data from both primary sources and computational geography simulations about travel, wind and sea patterns, seasonal access, costs and other considerations to plot realistic transport networks. Through the development process, server infrastructure and user interface upgrades supported the volume of data, user needs and extended functionalities, such as revealing the most efficient routes between points. The result provides scholarly information on life and travel in the Roman era. ORBIS illustrates how network and hierarchical data can be used to visualize complex, multilayered geospatial information.

KEYWORDS

data maps

geospatial information

history

geography

geographic information systems

electronic visualization

Elijah Meeks is the lead developer at the Center for Interdisciplinary Digital Research in the Stanford University Library and can be reached at emeeks<at>stanford.edu.

Creating a network model of the Roman world presented an exciting opportunity to implement techniques for the study of the past that had only previously been seen in computational geography departments and as interactive toy demos. Such a visualization would allow scholars the capacity to explore how the Roman world may appear differently from different perspectives within the system, highlighting a sense of distance and space that was inherently topological and perspectival.

The role of neogeography in disseminating radical cartography and GIS methods outside of academic geography departments cannot be overstated. Radical cartography refers to challenging the implied neutrality of geographic systems and functions, such as map projection. Neogeography, in contrast, refers to the use of geographic information systems (GIS) and map-making by journalists, artists and designers outside of the professional community of cartographers and the academic community of geographers. Both of these developments factored into the creation of ORBIS, a network model that represents the movement of goods and people in the Roman world.

In 2011, Walter Scheidel, classics professor at Stanford University, was exposed to the concept of a distance cartogram via the work of designer Tom Carden [1]. Carden, well-entrenched in the field of information visualization and web mapping, had produced a geospatial information visualization of the London subway system that showed the routes using a traditional geographic projection but then distorted the geographic location of each subway stop and line to reflect the time taken from one stop to another. Cartograms are not new in academic geography circles, but the accessibility of this relatively simple information visualization was what allowed someone in Scheidel's position to understand and use such a representation, which was more in line with the manner in which Romans would have historically viewed space.

How ORBIS Was Made

To create such a map is relatively easy for modern transportation networks, where the components and costs can be accessed using modern application programming interfaces (APIs) and GIS applications. For the Roman world, the map would require significant effort to create the necessary data and data models. The tracks of Roman roads were gathered from the Barrington atlas [2], the river routes by which goods and people flowed in Roman times were taken from modern river courses [3] and the cost in time for the movement of peoples and the cost to ship goods were gathered from primary sources such as the Antonine Itinerary and Diocletian's Edict on Costs [4]. But for sea travel, which is of primary importance for understanding the Roman transportation network, the speed and course of sea routes needed to be simulated based on the performance of historical sea vessels and calibrated with the scant evidence available. This sea model relied on network path-finding calculations utilizing modern wind and sea patterns, which match well with historically known environmental conditions [5].

Upon completion of this network, the second technical hurdle was making it accessible in a manner that allowed scholars to explore the effect of varying priorities, times of year and accessibility of travel modes (sea, road and river). By leveraging a robust spatial database linked to a web map that could run in any browser, ORBIS provided just such access. But the computational cost of calculating one-to-many pathfinding meant that the initial version of ORBIS could only show a few pre-calculated distortions of the Roman world using distance cartograms. It did, however, allow any user to calculate any one route from one site to another within the system using any variation on the parameters of the model. As a result, even though ORBIS was designed for dynamic representations of the entire world, it proved most popular as a Google map for the Roman Empire. It allows scholars and the general public to see the cost in time and money to move from one part of the Roman world to another based on different priorities and times of year.

A significant upgrade to both the server infrastructure and user interface of ORBIS in 2014 (ORBIS v2) finally allowed users to not only compute cost from one site to another using any parameters but also from one site to many. This improvement afforded the capacity to represent distance in time or money using not only the distance cartogram method (Figure 1) but also by grouping sites of similar cost into regions and representing the results using isochronal (contours indicating cost in time) and isophoretic (contours indicating cost in expense to ship goods or individuals) maps (Figure 2).

FIGURE 1. A distance cartogram distorts geography from a central point, in this case Constantinople, such that the angular position of other points is the same as it is in the base geographic projection, but the distance of points reflects their distance in cost. In this case, all points (both sites and points that determine the shape of the polylines representing routes) are the same angle as they would be from Constantinople in a Mercator projection, but the distance reflects the cost to reach each point in days.

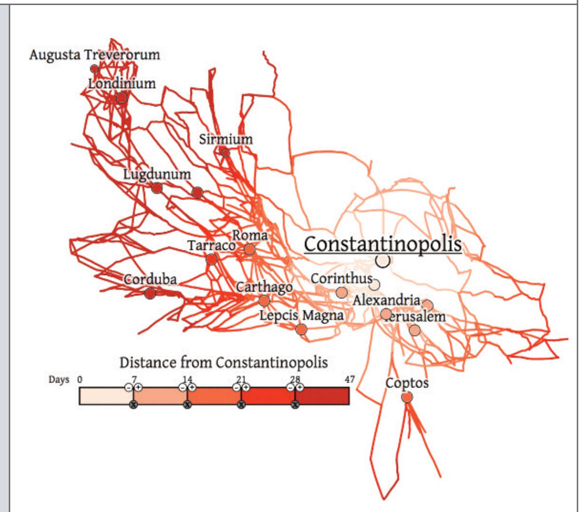


FIGURE 2. Two maps using contours to indicate cost to reach parts of the Roman world from Constantinople. The isophoretic map on the left uses those contours to indicate expense to ship goods whereas the isochronal map, right, uses those contours to indicate time to reach a location. Regions are derived from Voronoi polygons, which are joined together to form single polygons for a region where the cost falls into a user-defined band.

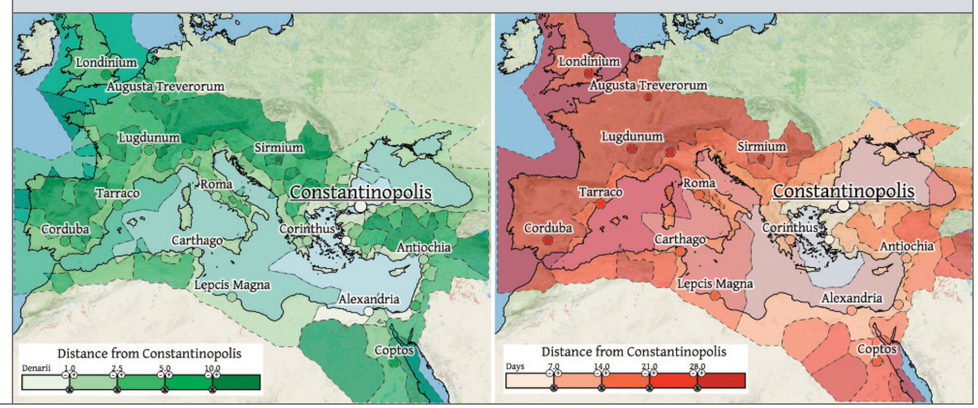


FIGURE 3. Clustering aggregates sites based on the cost to reach that site calculated by the user. Here the model is divided between five calculated centers and a sixth category, frontier, in dark blue, indicating areas where the cost to reach is within 10% of the cost from more than one site.

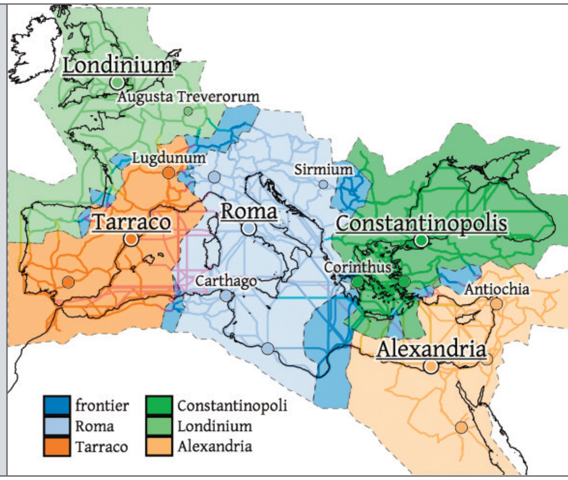
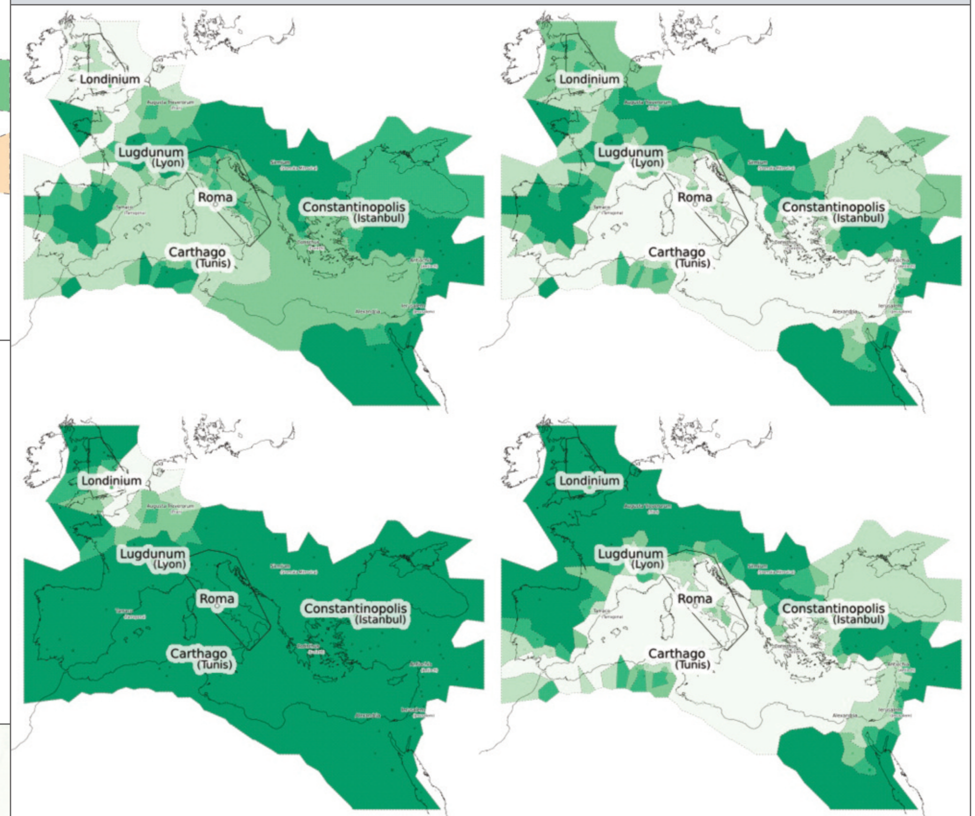


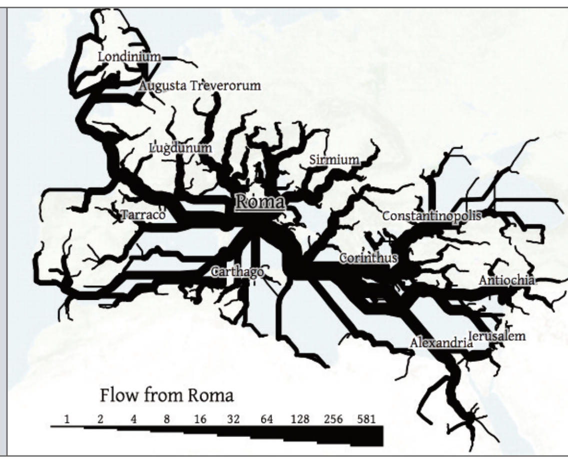
FIGURE 5. Isophoretic maps of London, left, and Rome, right, in summer (top) and winter (bottom) highlight the temporal contingency of the spatial experience of the Roman world from London that is not nearly so prominent from Rome.



Two new pieces of functionality are also provided in ORBIS v2: clustering and flow maps. Clustering (Figure 3) provides users with the ability to create simple maps of the Roman world wherein sites or regions are shown that are closer to one center than another. (Note: The custom algorithm used to determine clusters in ORBIS assigns sites to groups based on the closest calculated center, with the capacity to designate sites within a user-defined threshold to a frontier cluster if the cost to reach a site plus the threshold value is within the cost to reach the site from multiple centers.)

Flow maps (Figure 4) utilize the network statistic of edge betweenness centrality to highlight the main arterial routes by calculating the most efficient routes from one site to the rest of the model and aggregating shared routes onto segments. This is not a typical definition of flow, which relies on demographics of movement or the results of flow models that simulate such demographics, but seems to align well with the traditional definition, although this relationship has not yet been studied.

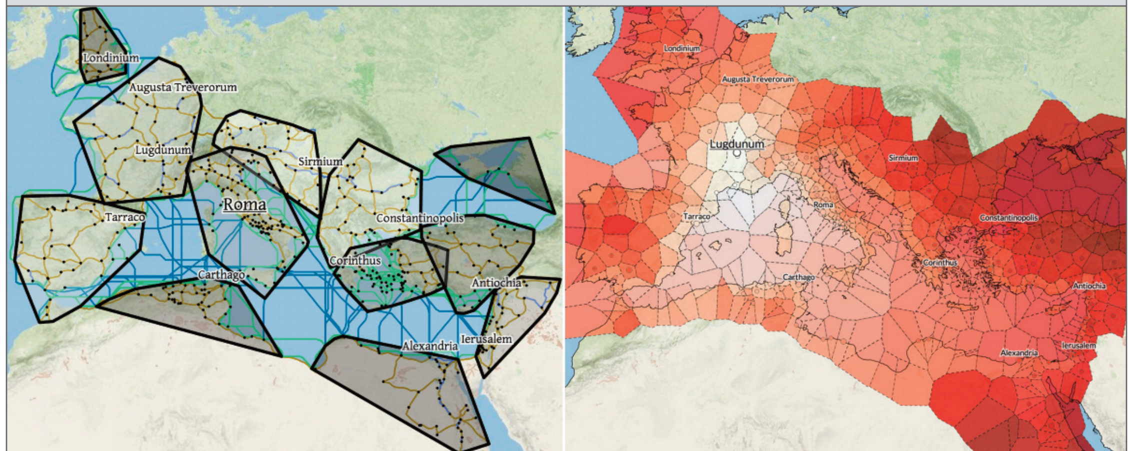
FIGURE 4. Flow maps calculate routes to or from a single site and aggregate the results by segment, known as edge betweenness. This raw value is represented to the user with line thickness.



In ORBIS v1 and especially in ORBIS v2, the application provides a representation of the Roman world as one where the shape of the world is dependent on priority (for example, the cheapest or fastest route), time of year (Atlantic coastal travel is restricted during the winter, as are mountain passes) and the perspective on the model. (By virtue of their location in the model, some sites, such as Rome, experience little change, whereas others, like London, have a highly contingent place in the network.) This distinction is demonstrated in Figure 5.

The dramatic difference between modern travel times and costs in comparison to Roman-era times and costs is apparent even to the lay user, as reflected in numerous references to ORBIS in social media. For scholars, the topography of cost, both in individual routes and the aggregate, has proved useful for the purpose of better informing historical climate models or the movement of specialized materials such as marble.

FIGURE 6. Derived areas from point data in ORBIS take two forms. Convex hulls (left) are used to define regions that roughly correspond to Roman provinces and are amenable to deformation in cartograms. Voronoi polygons (right) are derived from sites and then merged within quantized bands to define zones for isochronal and isophoretic maps.



Creating an Application like ORBIS

The sophistication of ORBIS, which is among the most complex pieces of geospatial information visualization on the web, makes it difficult to replicate. Some similarly exemplary projects are Earth, which shows global wind and current patterns (http://earth.nullschool.net/) and the geospatial information visualization experiments of Jason Davies (http://www.jasondavies.com/maps/). The ORBIS codebase, while open-source (https://github.com/emeeks/orbis_v2), is poorly documented and specific to this single application. However, some of the principles in place that proved so successful, such as using topologically aware geospatial features to create regions or the calculation and representation of system-wide travel costs, have been made available in better documented and more accessible libraries such as d3.carto (https://github.com/emeeks/d3-carto-map) and orbis-in-a-box. (Note: The only current examples of orbis-in-a-box are these experiments with 19C United States postal routes: http://bl.ocks.org/emeeks/8edaa27a121dc2a227ec; http://bl.ocks.org/emeeks/b8da1d56fd9c21244fdd)

The strength of ORBIS is its existence as a network model – an explicit system that represents and annotates the connection between sites. This network model influenced the creation of Kindred Britain (http://kindred.stanford.edu), which uses many of the same network cost principles to understand not physical distance but genealogical distance between British cultural elites.

The Future of Science and Technology Mapping

ORBIS represents a transition away from web mapping toward geospatial information visualization. The distinction, arbitrary and declared, is not only the focus of applications such as the attempt to use maps as a view into a model, but also in the integration of techniques such as convex hulls, Voronoi tesselation and cartograms into the map interface (Figure 6).

Traditional web mapping relies on a leveraging of layers to allow users to compare and contrast the distribution of spatial phenomena. Geospatial information visualization requires more sophisticated network and hierarchical data to provide radically different views into spatialized phenomena that can look like a traditional web map in one mode and much more like abstract information visualization in other modes. By necessity, this technique requires practitioners who are familiar not only with spatial analysis but also with network analysis and information visualization. There is not yet as strong a community of practice for geospatial information visualization as there is for web mapping. ORBIS takes advantage of the D3 information visualization library (www.d3js.org), which provides robust capacity not only for traditional data visualization but also for network visualization and geospatial visualization. In contrast, web mapping libraries like Leaflet (http://leafletjs.com/) focus on providing convenient representation of geospatial data and require significant effort to integrate with cutting-edge network and data visualization.

Outlook

Google maps taught many how to read and use geospatial maps online. This novel map reading literacy can be used to integrate more abstract representations of space and place growth. These more sophisticated representations require more sophisticated material. If one only has simple data in point of polygon form, then that is all that can be shown on a map. In contrast, building a system operationalizes the relationship between elements in that system and therefore formalizes claims in a way that can intimidate academic and journalistic practitioners. It is also simply easier to acquire and deploy simple spatial datasets, whereas models require more effort to create. But if a system like the model underlying ORBIS is available, then a map can represent views into that system using the methods developed and explored in computational geography, as well as methods used for information visualization in other domains. ■

Resources Mentioned in the Article

[1] Carden, T. (July 9, 2006). Travel time tube map. Retrieved from http://www.tom-carden.co.uk/2005/12/31/travel-time-tube-map/

[2] Talbert, R. J. A. (ed.). (2000). *Barrington atlas of the Greek and Roman world* {Vols.1-3]. Princeton, NJ: Princeton University Press.

[3] From *Natural Earth*: http://www.naturalearthdata.com/

[4] Graser, E.R. (1940), A text and translation of the *Edict of Diocletian*. In T. Frank (Ed.). *An economic survey of ancient Rome [Vol. 5: Rome and Italy of the Empire* (1st ed.) [pp. 307-421]]. Baltimore: Johns Hopkins Press.

[5] ORBIS: The Stanford Geospatial Model of the Roman World: http://orbis.stanford.edu/orbis2012/#building - § Geospatial Technology

Special Section

Mapping Graphene Science and Development: Focused Research with Multiple Application Areas

by Luciano Kay, Alan L. Porter, Jan Youtie, Ismael Rafols and Nils Newman

Mapping Science

EDITOR'S SUMMARY

Focusing on graphene, a material made from carbon atoms, researchers constructed a map illustrating the material in the context of scientific developments and patent activity over time. Two basemaps reflecting general science and patent activity were created, in addition to two overlay maps reflecting scientific publication and patenting activities specific to graphene. The science and patent overlay maps together display how the material was discovered and used between 2000 and 2013. The maps reveal 222 Web of Science categories and 466 International Patent Classification categories and identify 18 macro-disciplines and 35 macro-patent categories. Graphene research is shown to focus on material sciences and closely related fields such as engineering and physical chemistry, while graphene inventions cover technologies from catalysis and semiconductors to the pharmaceutical arena. The node size, density of relationships and colors of the map give additional information on research and development efforts and relationships among subset topics. The process employed text mining, visualization and graphics software tools in addition to a mapping kit available from the authors.

KEYWORDS

data maps
patents
research and development
scientific and technical information
electronic visualization
scientometrics

This map illustrates the particular features of research and development (R&D) of graphene through improved science and patent basemaps and data overlays. Graphene is a two-dimensional material made out of a single layer of carbon atoms. Its special features – including great strength, yet lightness, and electrical and thermal conductivities – were recognized in the 2010 Nobel Prize in physics. This map, created for the 10th iteration of *Places & Spaces*, represents a unique opportunity to demonstrate the power of science and patent mapping in understanding scientific and technological development activities with an interesting case study. We combine science maps developed by some of us with colleagues [1] with our new patent mapping method [2]. We also show that, compared to other patent maps that also use International Patent Classification (IPC) technology categories [3, 4], our approach reveals technological relationships that are not well captured by the IPC system.

Basemaps can show the structure of the landscapes of science such as all scientific publications in a time period, patenting activity such as all patents in a time period and the relative position of scientific and technological fields. They are the backbone of the overlay map. Overlay maps represent case studies or concrete examples that allow learning through comparison.

Luciano Kay is postdoctoral scholar at the Center for Nanotechnology in Society, University of California, Santa Barbara, and can be reached at luciano.kay<at>ucsb.edu.

Alan L. Porter is professor emeritus at the School of Public Policy, Georgia Institute of Technology, and director of R&D, Search Technology, Inc., and can be reached at alan.porter<at>isye.gatech.edu.

Jan Youtie is a director at the Enterprise Innovation Institute and adjunct at the School of Public Policy, Georgia Institute of Technology, and can be reached at jan.youtie<at>innovate.gatech.edu.

Ismael Rafols is affiliated with Ingenio (CSIC-UPV), Universitat Politècnica de València, València, Spain and SPRU – Science and Technology Policy Research, University of Sussex, Brighton, England, e-mail: i.rafols<at>ingenio.upv.es.

Nils Newman is president of Intelligent Information Services Corporation, Atlanta, and can be reached at newman<at>iisco.com.

FIGURE 1. Science basemap [1]

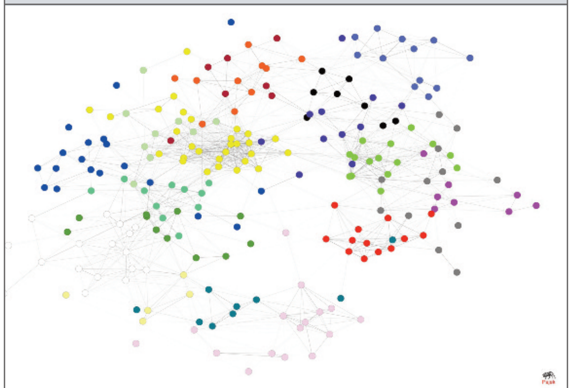
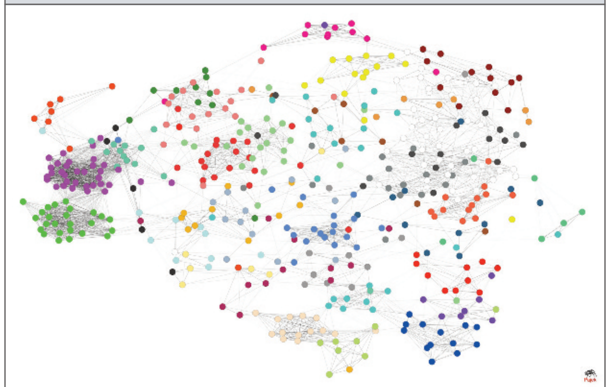


FIGURE 2. Patent basemap [4]



18 macro-disciplines (groups of subject categories) and 35 macro-patent categories (technology groups) using factor analysis. Our approach to put together multiple IPC class levels in patent visualization (by re-grouping classes according to number of records) compensates for IPC classification issues by (1) disaggregating IPC categories and (2) reforming them based on citing-to-cited reference patterns. The number of groups is chosen to increase the face validity of the maps. Colors represent 18 science macro-disciplines and 35 macro-patent categories (there is no relation among the colors of each map). Lines represent relationships between disciplines and groups. Only the strongest relationships are shown (threshold set at median value). Darker lines represent stronger relationships. The position of nodes results from visualizing the overlays with Pajek using the Kamada Kawai layout. The closer the nodes, the stronger the relationship between them.

Creating a Basemap

To make these overlay maps of graphene publications and patents, we use recently developed science and patent mapping methods and a number of software tools such as VantagePoint (www.thevantagepoint.com) for text-mining, Pajek (http://pajek.imfm.si/doku.php?id=pajek) for network visualization and analysis, and Adobe Photoshop for graphics creation. Our science basemap [1] draws on data from Thomson Reuter's Journal Citation Reports of the Science Citation Index and Social Science Citations Index of 2009, including 61 million citation instances (Figure 1). Our patent basemap [2] draws on data from the European Patent Office's PATSTAT patents database from 2000 to 2006, including more than 760,000 patent records (Figure 2). The relative stability of the IPC7 classification system in this time period made it an appropriate time span for the first version of the basemap.

Both science and patent maps result from network visualizations of cosine similarity matrices that reflect scientific publication citation patterns and citing-to-cited relationships of IPC classes, respectively. In total, 222 Web of Science categories are represented and 466 IPC-based categories are created using sized groupings of mixed hierarchical composition (IPC classes under a 1,000-record threshold are regrouped). We also identified

Creating Overlays

We then create overlays for graphene scientific publication and patenting activities (Figure 3 and Figure 4, respectively). These are based on 26,381 publication and 6,457 patent records found between 2000 and 2013 using

FIGURE 3. Graphene data overlay on science basemap. Node size represents share of publication records in each subject category

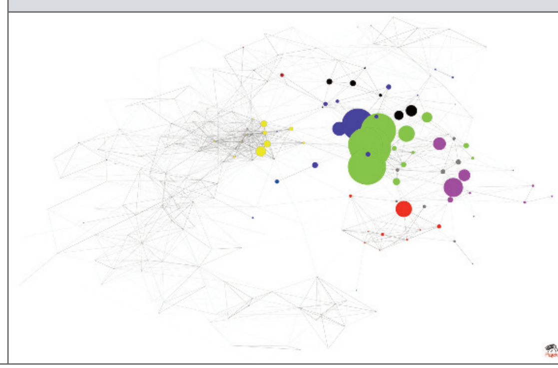
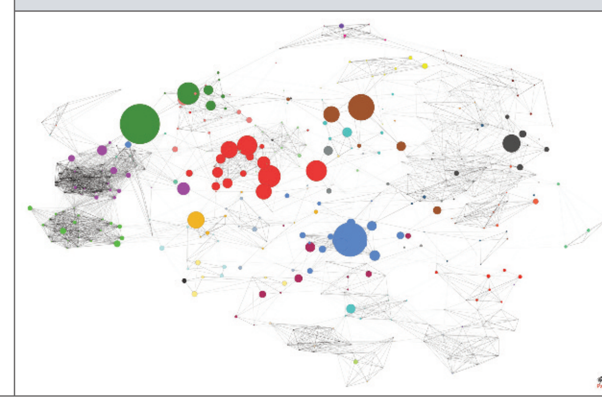


FIGURE 4. Graphene data overlay on patent basemap. Node size represents share of patent records in each IPC category.



the keyword-based "graphene" search query on data sources similar to those used in our basemaps. Overlay nodes represent Web of Science category groups (science map) and IPC-based technology groups (patent map) that have some level of activity in that time period. Their size represents the number of scientific-publication or patent records in the datasets.

Learning from Maps

This kind of approach to visually locate bodies of research within the sciences and areas of technology development can help users to understand innovation dynamics and to support business and policy decision-making. Our science and patent overlay maps show the unfolding of graphene discovery and exploitation between 2000 and 2013. Interestingly, graphene research and development follows a funnel pattern where scholarly publications are primarily focused on one area of science, and patenting activity is widespread across multiple application areas. Graphene research focuses on material sciences and closely related fields such as engineering and physical chemistry. There is very little activity in other fields such as biology. Graphene inventions, on the other hand, cover a wider variety of technologies with focus on catalysis and separation, semiconductor parts and devices, batteries and capacitors, and the chemical properties of graphene. Other inventive activity focuses on a small number of applications related to computing and pharmaceutical uses such as drug delivery.

Through these maps, novice users can learn about complex processes that occur in science and technology with concrete application examples that present large amounts of data in an easy-to-understand visual format. Map features, such as the density of relationships, the size of nodes (i.e., disciplines and categories) and colors, tell the story and visually represent key points or takeaways in each analysis. Such visual representation of scientific and patenting activity can help novice users grasp the extent and diversity of academic or corporate R&D and their relationship with invention endeavors. Science and patent mapping of data subsets (for example, graphene with batteries; particular organizations' contributions; and/or comparison of time slices) can help explore particular relationships.

Facilitating replication and the creation of similar maps based on different data has been among the top design criteria applied in the development of our patent mapping method. Therefore, we produced a set of basic tools or a mapping kit (available upon request from the authors) that allows creating maps in a few steps. Our mapping kit includes source files that represent the structure of the basemaps and thesauri files that represent scientific publication and IPC-based category definitions and enable creation of overlay maps using software such as VantagePoint and Pajek. People can use this kind of tool to do analysis to support competitive intelligence, corporate R&D and policy decision-making. For instance, companies can create overlay maps to visualize competitors' portfolios and anticipate competitive threats. Companies can also visualize their own patent portfolios to make decisions about their R&D and licensing strategies. Moreover, policy makers can use these kinds of tools to, for instance, analyze the science and technology profile of regions and design programs that support local companies and innovation. More generally, researchers, policy makers and R&D managers can find value in patent maps to assess where and how knowledge migrates and integrates across application areas. One caveat is that the maps need to balance the level of granularity between offering so much detail as to prohibit visualization of the overall system and insufficient detail to allow drilling down into particular scientific and/or technological areas. We balance detail and visualization by using a bottom threshold. This threshold culls out patent groupings that attract relatively few numbers of patents and enables our map to focus on those patent groupings with consistent and sizeable numbers of patents. Currently shown maps are very coarse grained and offer a bird's-eye perspective. They locate patents in areas, but cannot provide details of the trajectories.

Future Outlook

We are currently working on the development of an iteration of our method to create global maps/patent map overlays, which will lead to new maps and case studies. The new IPC8 patent map will capture some 1,000,000 PATSTAT records for EPO patents from 2007, and it increases the threshold for inclusion of IPC categories from 1,000 to 1,500, to deal with a comparable magnitude of resulting categories. As before, we grab the cited patent records and extract Cited IPC information. We obtain 434 IPC-based categories and we group them into 35 usable macro categories through factor analysis (accounting for some half of the variance) plus a set of categories that do not load well on any factors. We have constructed both 434-node and 35-node maps. We expect to use this new patent-mapping iteration to develop new case studies – overlays – for specific technological fields. A high-resolution version of this map at is available at http://scimaps.org/mapdetail/mapping_graphene_sci_179.

Acknowledgments

This research was undertaken largely at Georgia Tech drawing on support from the National Science Foundation (NSF) through the Center for Nanotechnology in Society (Arizona State University; Award No. 0531194); and NSF Award No. 1064146 ("Revealing Innovation Pathways: Hybrid Science Maps for Technology Assessment and Foresight"). Part of this research was also undertaken in collaboration with the Science Policy Research Unit at the University of Sussex and the Center for Nanotechnology in Society, University of California, Santa Barbara (NSF Awards No. 0938099 and No. 0531184). The findings and observations contained in this work are those of the authors and do not necessarily reflect the views of NSF. ■

Resources Mentioned in the Article

[1] Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Overlay science maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology, 61*(9), 871– 1887. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/asi.21368/abstract

[2] Kay, L., Newman, N., Youtie, J., Porter, A. L., & Rafols, I. (2014). Patent overlay mapping: Visualizing technological distance. *Journal of the American Society for Information Science and Technology*. Early View available at http://onlinelibrary.wiley.com/doi/10.1002/asi.23146/abstract. doi: 10.1002/asi.23146:

[3] Hinze, S., Reiss, T., & Schmoch, U. (1997). *Statistical analysis on the distance between fields of technology*. Innovation Systems and European Integration (ISE), Targeted Socio-Economic Research Program, 4th Framework Program of the European Commission (DGXII), Karlsruhe, Germany.

[4] Leydesdorff, L., Kushnir, D., & Rafols, I. (In press). Interactive overlay maps for U.S. patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics*. Retrieved from http://arxiv.org/pdf/1210.6456.pdf

Special Section

Mapping the Linguistic Context of Citations

by Marc Bertin, Iana Atanassova, Vincent Larivière and Yves Gingras

Mapping Science

EDITOR'S SUMMARY

Scientific papers are routinely structured in sections for introduction, methods, research and discussion, a standard since the 1970s. Citations originating within each section serve different purposes and can be meaningfully classified according to position, shedding light on an author's purpose for the citation. Furthermore, words near the citations in the various sections differ, providing the basis for lexical and semantic analysis of citation contexts. Approximately 50,000 scientific papers from seven PLOS journals published between 2009 and 2012 were analyzed for citation use within the identifiable document structure and for verbs used in the context of the citations. Frequencies of verbs in the four section types demonstrate the predominant use of certain words by section. Introduction sections showed greater variety of verbs, while a more limited range of verbs was seen in Methods sections. The lexical distribution process may be applied to other contexts supporting text processing based on XML format.

KEYWORDS

bibliographic citations
document structure
linguistic analysis
verbs
contextual information
intent

Marc Bertin is a postdoctoral fellow at the Centre Interuniversitaire de Recherche sur La Science et La Technologie at the Université du Québec à Montréal, Canada. He can be reached at bertin.marc<at>gmail.com.

Iana Atanassova is an assistant professor at Centre Tesniere at the Université de Franche-Comté, France. She can be reached at iana.atanassova <at>univ-fcomte.fr.

Vincent Larivière holds the Canada Research Chair on the Transformations of Scholarly Communication at the Université de Montréal, Canada. He can be reached at vincent.lariviere<at>umontreal.ca.

Yves Gingras is a professor in the département d'histoire at the Université du Québec à Montréal, Canada. He can be reached at gingras.yves<at>uqam.ca.

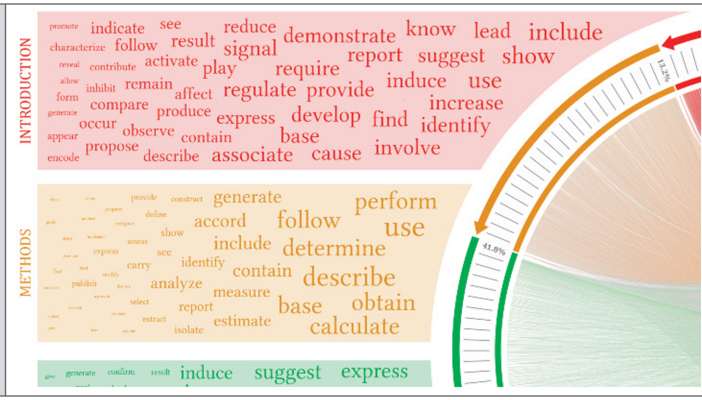
The IMRaD structure of scientific papers (Introduction, Methods, Results and Discussion) was adopted by most journals since the mid-20th century and became standard in the 1970s. It was introduced to facilitate the reading of publications and to provide faster access to information by standardizing the argumentative structure of articles. The IMRaD sequence provides an outline for scientific writing, dividing the articles into four sections, each one having a specific rhetorical function.

While citations have been studied extensively from a quantitative point of view, our approach is motivated by the need to examine more closely the position of citation references at the level of sentences and observe the trends related to the ways authors cite previous works in different sections of a publication. In this study, we research how authors use citations in the different sections of scientific articles that follow the IMRaD structure. We consider that verbs found in sentences containing citations are an important indicator of the purpose of citations and of the reasons behind citing a given document. Results of the study are communicated using a visualization that shows the relations between verbs used around references and the structure of scientific papers.

Figure 1 presents a cutout of the overall map. The numbers inside the circular representation correspond to the average positions of section boundaries observed in the articles: 13.2% for the end of the Introduction and 41.8% for the end of the Methods section. By showing the differences in the frequencies of verbs that appear in the four section types of the IMRaD structure, this map communicates that the way an author cites a reference depends highly on the position in the IMRaD structure. This study is an important step towards the lexical and semantic analysis of citation contexts. The results indicate that the rhetorical structure of scientific articles determines the positions of references and their relation to the

CONTENTS
< PREVIOUS PAGE
NEXT PAGE >
NEXT ARTICLE >

FIGURE 1. Most frequent verbs in the Introduction and Methods sections



article's study. The verbs that we have identified can further be used as clues for the categorization of the relations between authors and for the analysis of citation networks.

Creating a Map of Citation Contexts

The dataset used in this study comprises the entire set of PLOS (Public Library of Science) documents over the period 2009-2012. It contains about 50,000 scientific papers in seven different journals. Most of the articles in the corpus (about 87%) are in the domains of biology and medicine, and the rest of the articles cover a wide range of subject areas such as computer and information sciences, physics and social sciences. The articles are published in Open Access in XML (data harvested from http://www.plos.org in October 2012) using the Journal Article Tag Suite (JATS). This standard is an application of NISO Z39.96-2012 and JATS is a continuation of the NLM Archiving and Interchange DTD (http://jats.nlm.nih.gov). Table 1 contains general statistics on the corpus.

Automatic processing of the data proceeded as follows: Firstly, in order to link each citation context with its position in the IMRaD structure, we identified the different sections of the papers (Introduction, Methods, Results and Discussion) and their sizes in terms of number of sentences. In order to do so, we processed the section titles using a set of regular expressions specifically designed for this task. After analyzing the density of citations along the four main section types, we showed that the distribution of references along the text progression is essentially invariant across the different PLOS journals [1]. This first result showed that a strong relation exists between the IMRaD structure and the use of citations in scientific writing.

Secondly, we extracted verb occurrences found in citation contexts. Verbs were identified using the Stanford POS-tagger [2]. As a result, we obtained the frequency of each verb in the four section types, as well as the links between their occurrences and the positions in the text. These data were visualized using the CIRCOS tool [3], written in Perl, which allows the visualization of information in a circular layout. The data related to the positions of verbs along the text progression were converted into plain text configuration files for the CIRCOS tool. Taking into consideration the verbs that appear in all four sections, we have obtained a set of 1,807 unique verbs. About 500 of them account for 90% of all citation contexts [4]. For the visualization, we have used the top 87 most frequent verbs that account for about 60% of all citation contexts.

Visualization Details and Observations

The graphical representation of the links between verbs in citation contexts and positions in the rhetorical sections can be viewed from two

TABLE 1. General statistics on the PLOS corpus

Journal	Number of articles	Avg number of sentences per article	Avg number of citation contexts per article
PLOS Biology	2,965	141.77	54.63
PLOS Computational Biology	2,107	242.00	87.49
PLOS Genetics	2,560	218.80	91.09
PLOS Medicine	2,228	95.98	39.62
PLOS Neglected Tropical Diseases	1,366	157.44	67.57
PLOS Pathogens	2,354	216.91	93.41
PLOS ONE	33,782	177.90	74.10
All PLOS journals	**47,362**	**178.19**	**73.55**

TABLE 2. Top 10 most frequent verbs in the four section types

Rank	Introduction	Methods	Results	Discussion
1	show	use	use	show
2	use	perform	show	suggest
3	include	follow	find	use
4	suggest	obtain	report	report
5	identify	generate	observe	find
6	find	base	suggest	include
7	require	determine	identify	observe
8	associate	contain	express	require
9	involve	calculate	see	associate
10	lead	carry	include	involve

different perspectives. On the one hand, it characterizes the four section types through the sets of verbs that are most prominent in each section. For example, we can observe that the Introduction section uses citations in a different way than the Methods section, as the former shows a great diversity in the verbs that are used while in the latter a small number of verbs (*use*, *describe*, *obtain*, *calculate*, *perform*) stands out as having very high relative frequencies in citation contexts. On the other hand, each verb can be observed independently and analyzed with respect to the sections in which it appears. For example, the verb *show* occurs frequently in all sections except for the Methods section, and the verb *describe* is most frequent in the Methods and Discussion sections and occurs rarely in the other two sections. Table 2 shows the ranked lists of the top 10 most frequent verbs in the four section types. This result confirms that authors use different verbs to introduce citations according to the position in the rhetorical structure and these positions are therefore an important factor for the analysis of citation acts.

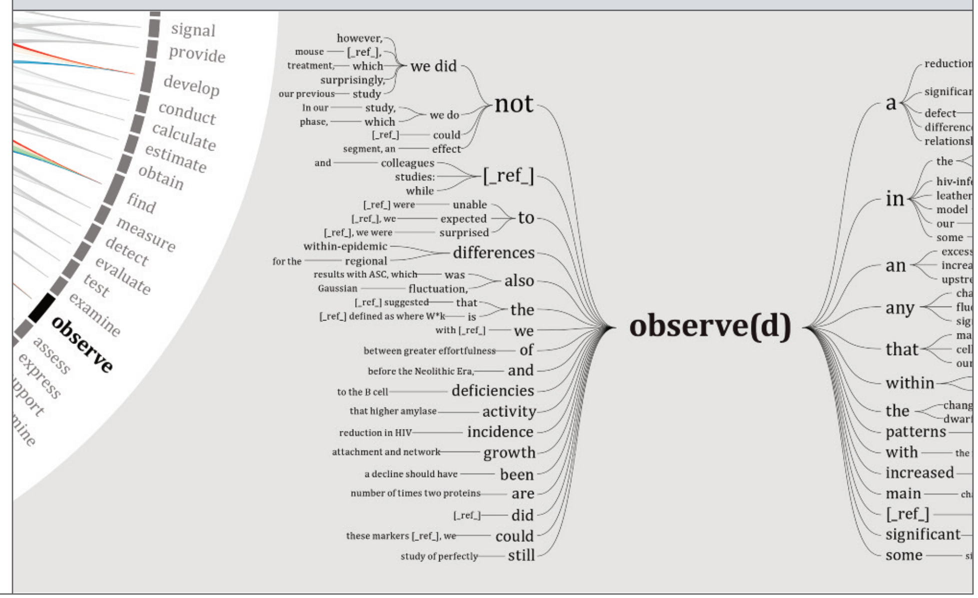
Figure 2 represents one of the most frequent verbs with its left context using a textual tree. These contexts were extracted from the corpus that we studied and the tree shows some of the patterns of the use of this verb. Such representations can be used to analyze the different semantic values of the verb according to the contexts in which it appears. For example, the first type of context contains "not" and is used by scientists to indicate observations that differ from previous work.

From the point of view of bibliometrics, this map shows clearly that the section structure of research papers is an important element to consider as a classifier for citation context analysis. These results confirm the hypothesis that citations play different roles according to their position in the rhetorical structure of scientific articles. The study of citation through the frequency of verbs used in sentences is a first step towards a better categorization of citations.

Outlook

Our approach is designed around open-source tools such as Stanford POS-tagger and CIRCOS. While we focus on verbs in citation contexts, other linguistic phenomena can be studied using a similar approach, such as sentiment analysis or lexical distributions. However, the availability of the corpus in XML format is an important factor for automatic text processing

FIGURE 2. One of the most frequent verbs and its left context



because it gives access to the articles' content in structured full text. Our processing relies on XML-parsing tools that we specifically developed for the JATS schema.

Our current research involves developing textual navigation interfaces using R-Shiny in order to explore different visualizations of texts. We are also working on the processing of large textual datasets in order to produce the tools for text navigation and analysis that could be used by different communities in Sociology, Semantic Web or natural language processing. ■

Resources Mentioned in the Article

[1] Bertin, M., Atanassova, I., Larivière, V., & Gingras, Y. (In press). The invariant distribution of references in scientific articles. *Journal of the Association for Information Science and Technology.*

[2] Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing, 12*, 44–49.

[3] Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research, 19* (9), 1639-45. doi: 10.1101/gr.092759.109

[4] Bertin, M., & Atanassova, I. (2014). A study of lexical distribution in citation contexts through the IMRaD standard. *Proceedings of the First Workshop on Bibliometric-Enhanced Information Retrieval co-located with the 36th European Conference on Information Retrieval (BIR@ECIR 2014)*, 5-12. (CEUR Workshop Proceedings, 1143). Retrieved from http://ceur-ws.org/Vol-1143/paper1.pdf

Exploring the Relationships Between a Map of Altruism and a Map of Science

by Richard Klavans and Kevin W. Boyack

EDITOR'S SUMMARY

The science mapping community offers insights and points to trends in scientific inquiry by revealing connections among publications, authors, terminology and citation patterns. The authors applied the process of creating science maps to topics compiled by GuideStar to reveal altruistic motives driving nonprofit organizations (NPOs). From data mined from nearly four million web pages from 125,000 NPO websites, they created a map of the altruistic motive space. Topic modeling from words on the web pages generated 1,000 topics that were manually screened to yield 357 topics related to altruistic motives and their interrelationships. Nearly 100,000 NPOs were positioned by topic on an overlay map, and map labels capturing altruistic motives were created by human analysis. The product was compared with a map of science based on Elsevier's Scopus database. The maps of altruism and science show the strongest link between the altruistic trait of caring and scientific research in medicine, health and neuroscience. Links are also demonstrated between other scientific areas and altruistic beliefs underlying environment, innovation and policy.

KEYWORDS

data maps

nonprofit sector

topic models

link analysis

scientific and technical information

scientometrics

Richard Klavans and Kevin Boyack are researchers at SciTech Strategies, Inc., and can be reached at [rklavans, kboyack] <at> mapofscience.com.

A scientific community has developed over the past 30 years on the topic of "science mapping" – modeling and visualizing the structure and evolution of scientific inquiry by analyzing features associated with scientific outputs such as articles, their authors, the words used in these articles or article citation patterns. We have found that compelling visual representations (that is, accurate global maps of science) are very useful for understanding current structures and trends along with how science has evolved.

We have also come to believe that forecasting future changes requires an understanding of the motives underlying scientific inquiry. Why do people become scientists, engineers or researchers? How and why do they choose the topics on which they work? How are these motivations reflected in the mission statements of funding and funded institutions? What is the overall set of motives that inform research, and how are these distributed? These are difficult questions to answer. Complete answers to these questions are obviously multidimensional and far beyond the scope of a simple study.

While we were beginning to ponder and discuss these questions, one of us had a serendipitous meeting with the president of GuideStar USA, a non-profit organization (NPO) that compiles data on NPOs registered in the United States. We wondered whether it might be possible to generate a map of altruism (a visualization of the topics that NPOs pursue) using the same tools that create science maps. Could we somehow understand the altruistic motive space associated with NPOs? Could we create a "map" of that institutional motive space? If so, might there be a relationship between altruistic motive space and the evolving world of scientific inquiry?

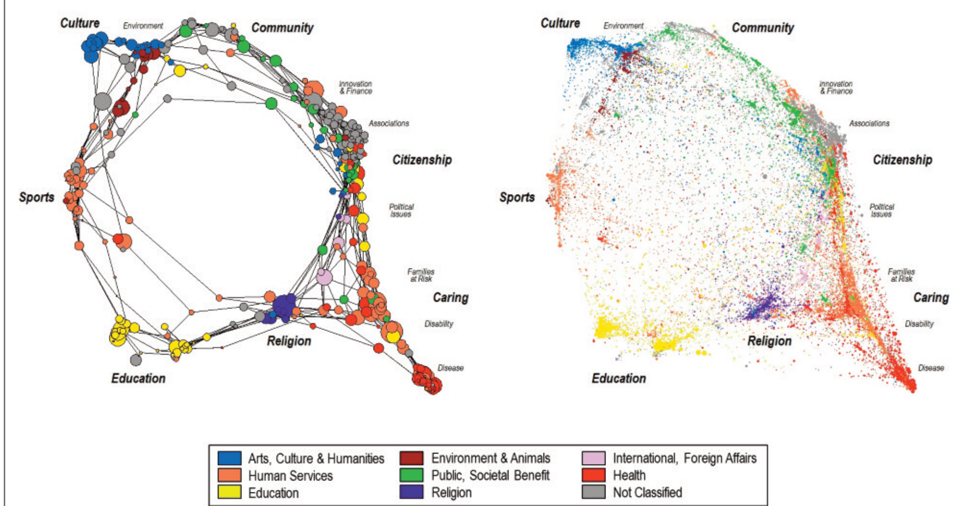
Mapping Altruistic Motives

GuideStar graciously agreed to provide us with data – we received financial data for over 2 million NPOs registered in the United States in 2011. Over 170,000 of these records also contained web addresses for the organizations. Using the assumption that the websites of these NPOs would say something about their motives, causes or missions, we mined nearly 4 million unique web pages from the 125,000 NPO web sites (72%) that were currently active. This provided the raw textual content from which we hoped to gain an understanding about "motive space."

The process we used to create a map of altruistic motive space from these data is described in great detail in [1]. Briefly, the steps were as follows:

- A topic model was created with the MALLET toolkit [2], using the text from the 4,000,000 web pages as input. Topic modeling creates sets of abstract topics that are represented as bins of related words. By examining the words associated with a topic, one can typically understand what that topic is about.
- The 1000 topics that were created by the topic model were manually screened, and topics that were not focused on motives were removed from the list. For example, topics associated with non-English pronouns, names, geography, histories, web protocols and how to donate were all removed from consideration. After screening, 357 of the original 1000 topics remained; these are the topics that are related to altruistic motives.
- One of the features of a topic model is that the distribution of the original input (in this case, 4,000,000 web pages) over topics is calculated. Thus, we knew which NPOs were associated with each topic. Many NPOs were fractionally associated with multiple topics. We calculated a topic-topic similarity based on the distribution of NPOs over topics. These topic-topic similarity values were then used as input to the OpenOrd (or DrL) algorithm [3], which generates an x,y layout (a map) of the input objects based on their similarities. The resulting map is shown as the left panel in Figure 1 and represents the overall motive space associated with the NPOs. It shows high-level sets of motives and the relationships among them.
- A second version of the map was created, this one showing the positions of nearly 100,000 NPOs overlaid on the topic map. For NPOs that were fractionally assigned to multiple topics, their positions were triangulated using fourth roots so that they would be placed very close to their dominant topics in most cases. This map is shown as the right panel in Figure 1. Note that in this map it appears as if there are "trails" of NPOs between topics in different parts of the map. These represent NPOs that are linking diverse motives.

FIGURE 1. Map of altruistic motives. On the left, 357 motive-related topics [1]; right, 100,000 NPOs positioned on the space identified by the topics.



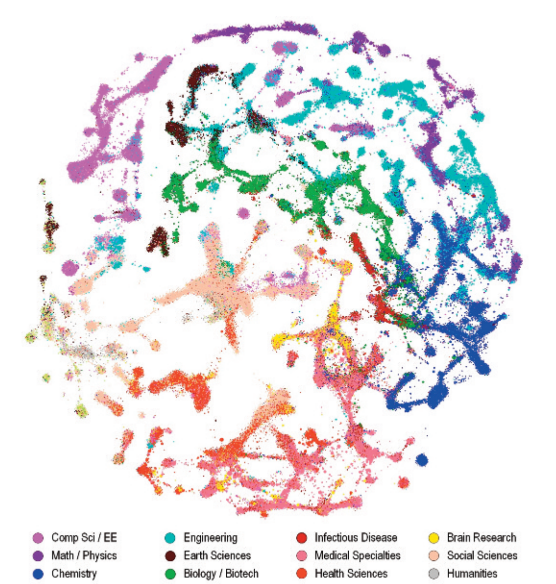
Labeling of the map was done manually by examining the topics and NPOs that were found in each part of the map. This map of altruism reveals the motives around which our donation of time and money are based: culture, sports, education, religion, community, citizenship and caring. Coloring of topics and NPOs in Figure 1 was based on the National Taxonomy of Exempt Entities (NTEE) codes associated with the NPOs [1]. High-level groupings of these codes are shown in the legend of Figure 1. Some of the NTEE codes are specifically associated with scientific research. We will return to this observation later.

Mapping Science

To start to explore the relationships between motives and scientific topics and to enable a visual representation of those relationships, we employ our most recent map of science. This map was created from the metadata associated with 43,000,000 scientific documents from Elsevier's Scopus database. The detailed process and explanation of this map are available in [4]. A brief explanation of the two-step process used is as follows:

- Documents were grouped into 156,000 clusters by direct citation using the recently introduced modularity-based clustering algorithm of Waltman & van Eck [5]. In a direct citation-based map, documents typically end up in the same cluster with the documents they cite. Thus, the clusters give a historical view of research in highly detailed topics.
- The titles and abstracts associated with the documents in each cluster were combined, thus giving a textual (word) profile for each cluster. Similarities between pairs of clusters were calculated using these word profiles with the bm25 metric. These cluster-cluster similarity values were then used as input to the OpenOrd algorithm [3], generating a map of the clusters based on their similarities. The resulting map is shown in Figure 2.

FIGURE 2. Map of science based on 43 million Scopus documents [4]



Although the map of altruistic motives and map of science were created using different methods, their creation was similar in the following regard. First, clusters of content were created from very large corpuses. For the altruism map, clusters (called *topics*) of words were created using topic modeling. For the science map, clusters of documents were created using direct citation. Second, those clusters were organized into a visual picture or map using a graph layout algorithm using similarities between the clusters. This is a general process that can be used for nearly every type of mapping exercise.

Relationships Between Altruism and Science

Most people consider altruism and science to be unrelated; the data suggest that financial relationships are relatively weak. The funding of scientific inquiry from NPOs in the United States amounts to roughly $30 billion annually [1]. This amount is only 1% of the revenue stream of U.S. NPOs ($2.7 trillion) and represents less than 7% of U.S. R&D investments ($450 billion in 2011).

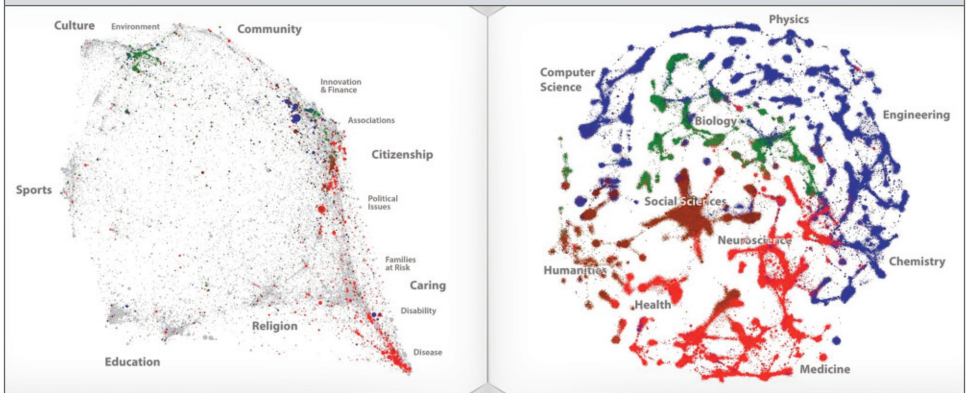
A much stronger relationship appears when we realize that national commitments to science may have their roots in altruistic motives. From an historical perspective, it is easy to show how the post-WWII commitment to science has its roots in altruistic motives. For instance, Vannevar Bush's influential description of science as an "Endless Frontier" [6] focused specifically on how science could be used to cure disease and encourage economic growth. Historical and current concerns about security fill out this historical landscape.

We show relationships between altruistic motives and science based on this historical perspective using four colors in Figure 3, which is a cutout of the map included in the *Places & Spaces: Mapping Science* exhibit (http://scimaps.org/mapdetail/exploring_the_relati_180). NPOs that specifically fund scientific research have been colored in the altruism map based on their NTEE codes, while the corresponding scientific areas have been colored in the science map.

Red is used to link the altruistic motives associated with caring (disease, disability and families at risk) with the scientific research in medicine, health and neuroscience. This connection represents what is perhaps the strongest link between altruism and science. NPOs associated with these motives contribute a significant amount of funding for medical research.

Blue and brown are primarily associated with citizenship. Blue areas link altruistic beliefs that innovation and economic growth are important national

FIGURE 3. Maps of altruism (left) and science (right) highlighted using common color coding. Red: caring (medical sciences); blue: innovation (physical sciences); brown: policy (social sciences); green: environment (life sciences)



goals with corresponding investments and advances in the physical sciences, engineering and computer science that have addressed these goals. Brown areas are mostly associated with the political issues that form the heart of citizenship. These topics are irresolvable issues that every nation faces – issues around whether we will have food, water, breathable air, security and meaningful employment. They are central issues to a national commitment to quality of life. Increasing our understanding about social sciences and the humanities is an integral part of addressing these altruistic motives.

Green represents a concern for the environment. From an altruistic perspective, environmental research is tied to both culture (reflecting a strong commitment to parks and open spaces) and citizenship (for example, sustainability). The areas of science addressing these environmental motives are mostly associated with biology and the earth sciences. Innovations that provide sustainable economic growth are also supported by some engineering disciplines.

Conclusion

Maps of topic spaces, such as those shown here, can be very useful in providing a context for both the generation and exploration of hypotheses. By exploring and comparing the two maps (and their underlying data) presented here, we have shown that there are plausible links between altruistic motives and science. We consider this topic to be a fruitful one for future research.

We also note that maps, data and analyses such as these have their limitations. For example, the altruism map is limited to data from the United States, and the two maps shown are static rather than dynamically evolving or interactive. Thus, we know very little about national differences in altruistic motives, how altruistic motives change and how these changes might fundamentally shift future scientific priorities. These areas are for future inquiry.

A high-resolution version of this map at is available at http://scimaps.org/mapdetail/exploring_the_relati_180. ■

Resources Mentioned in the Article

[1] Klavans, R., & Boyack, K. W. (2014). Mapping altruism. *Journal of Informetrics,* 8, 431-447. Retrieved from www.sciencedirect.com/science/article/pii/S1751157714000285

[2] McCallum, A.K. (2002). MALLET: A machine learning for language toolkit. Retrieved from http://mallet.cs.umass.edu

[3] Martin, S., et al. (2011). OpenOrd: An open-source toolbox for large graph layout. *Proceedings of SPIE - The International Society for Optical Engineering*, 7868, art #786806: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=731088

[4] Boyack, K.W. & R. Klavans (2014). Including non-source items in a large-scale map of science: What difference does it make? *Journal of Informetrics*, 8, 569-580: www.sciencedirect.com/science/article/pii/S1751157714000406

[5] Waltman, L. & Van Eck, N.J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378-2392: http://onlinelibrary.wiley.com/doi/10.1002/asi.22748/abstract

[6] Bush, V. (1945). Science: The endless frontier. *Transactions of the Kansas Academy of Science*, 48, 231-264: www.nsf.gov/od/lpa/nsf50/vbush1945.htm

Interstitial Organizations as Conversational Bridges

by Valeska P. Korff, Achim Oberg and Walter W. Powell

EDITOR'S SUMMARY

A novel approach to evaluating the impact of nonprofit organizations is proposed, combining social network and linguistic analysis. The authors examined data from nonprofit organizations' websites and site hyperlinks to other organizations. They identified 369 sites of organizations that measure social impact and captured inbound and outbound weblinks to construct a relational structure. Keywords from the sites were categorized as reflecting scientific, civil society or managerial domains and located the entity in a triangular semantic space. While most organizations fell into one of the three communities, some were in an interstitial space spanning these domains. The interstitial organizations used a mix of terminology, were densely interconnected and connected extensively to organizations across domains, but there were few direct connections among the three domains. The resulting map integrates cultural and relational dimensions and reveals hidden patterns and clusters. The approach can be used with other social systems combining rich text with relational data.

KEYWORDS

nonprofit sector
domain analysis
network analysis
linguistic analysis
link analysis
data maps
scientometrics

Valeska P. Korff is a junior professor in the Faculty for Economics and Social Sciences, University of Potsdam, Potsdam, Germany, and can be reached at vkorff<at>uni-potsdam.de.

Achim Oberg is a research fellow at the Institute for Small and Medium Sized Enterprises, University of Mannheim, Germany. All visuals in this paper are courtesy of Achim. He can be reached at oberg<at>ifm.uni-mannheim.de.

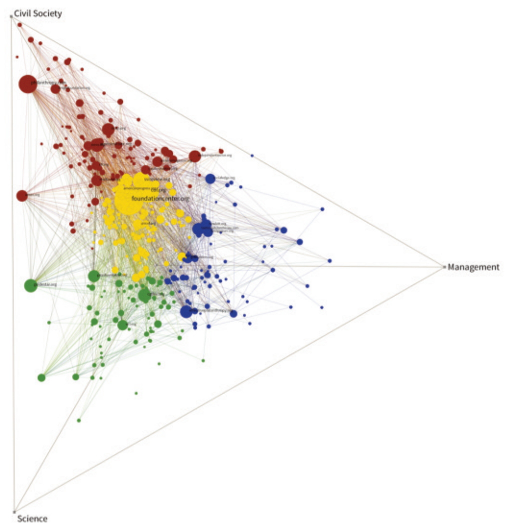
Walter W. Powell is faculty co-director of the Stanford Center on Philanthropy and Civil Society and professor of education, sociology, engineering and business at Stanford University. He can be reached at woodyp<at>stanford.edu.

Rationalization is a major social trend, entailing the spread of scientific practices of evaluation into many domains of modern life. Writing a century ago, German sociologist Max Weber referred to this process as "the disenchantment of the world." In a famous lecture at the University of Munich in 1918, he noted that this means that in principle one could, if one wished, master all things by calculation. Our map is an effort to capture how evaluation and measurement are influencing the contemporary social sector. We analyze how scientific and managerial discourses are combining with civic ideals in the context of efforts to assess the impact of nonprofit organizations. The debate over evaluation has brought together a plethora of organizations – nonprofits, foundations, consultancies, national and transnational government bodies, select corporations, academic centers, even blogs and social movements – in an effort to render nonprofit performance measurable. We visualize these conversations through an analysis of shared weblinks and the language reflected on webpages.

Visualizing Relations *and* Language

Two differing approaches have typically guided research on the structure and content of organizational communities. Network analysis emphasizes the role of social ties; in contrast, discourse analysis highlights the meanings of cultural codes. In our study, the ties that connect organizations and the language that unites or distinguishes them are combined. Our map, depicted in Figure 1, fuses these approaches into a visualization of relational *and* discursive dimensions. This simultaneous representation reveals how linguistic patterns and relational ties contribute to the formation of a new narrative of performance assessment. The map provides a basis for theorization of how a community develops a new language, in our case a metrics of evaluation. Further, the map permits comparison with other cases and enables examination of the role of

FIGURE 1. Relational and discursive dimensions of nonprofit performance evaluation



particular organizations in the debate on nonprofit performance evaluation. We regard our map as a *social x-ray* that captures, in a systematic and comparable manner, the configuration of a social system.

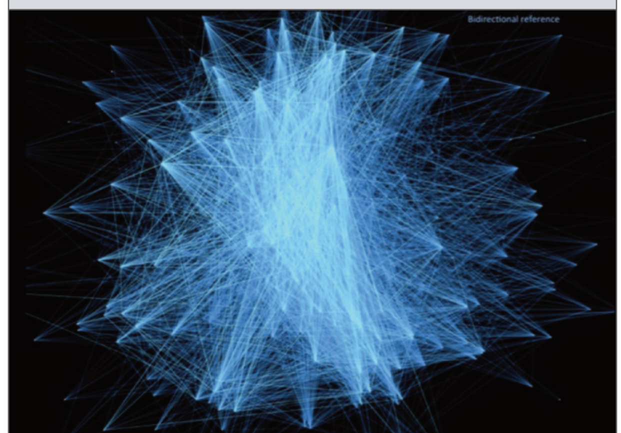
The map derives from a research project analyzing the relational structures and discourse between organizations involved in nonprofit performance evaluation (Powell, et al. [1]; also see http://pacscenter.stanford.edu/overview/research/powell). The data were obtained from organizations' websites, which reflect how organizations present themselves to a broad range of publics. Websites are increasingly a primary point of access and communication for many organizations. In addition, hyperlinks to other entities form a representational relation that sends a message to the public about the association between two parties. In this respect, hyperlinks are similar to bibliometric networks and friendship networks on social media. In focusing on websites, or more specifically URLs, we analyze organizations' positions in an evolving discourse and a web-link network.

Drawing on the Wealth of Online Data

Web-crawler and web-scraper technologies, developed by Oberg and Schoellhorn at the University of Mannheim, enable us to select websites for inclusion in the analysis and then to automatically compile their informational content. The webcrawler tracks all hyperlinks to other websites, identifying those frequently referenced as possible sample members. Starting with a list of 36 organizations suggested by experts as critically involved in the discourse on nonprofit performance evaluation, we used an iterative snowball sampling approach. Two iterations produced a total of 1,394 websites, from which we pruned generic websites such as the *New York Times*. This process produced a sample of 369 websites, representing entities actively involved as creators, carriers and consumers of nonprofit performance evaluation practices, recognized and referenced by others within this domain. By recording all weblinks through which organizations reference to and are referenced by others, we are able to draw the relational structure and produce the network graph in Figure 2.

The web scraper collects the complete text – which may include tens of thousands of pages – of each website in the sample. This textual data is stored in a special search index that we analyze for linguistic orientation. Keywords are socially prominent terms that indicate a user's cultural orientation. We identified keywords related to nonprofit performance evaluation by consulting five sources that provide glossaries for the sector: 3ie Impact, Roberts Enterprise Development Fund, Charities Evaluation Service, Innovation Network and TRASI – Tools and Resources for Assessing Social Impact. These concept guides were culled and matched with an analysis of our sample websites and checked with practitioners to generate a list of keywords used to discuss evaluation efforts. This approach produced a list of 105 keywords, which were categorized as deriving from either a scientific, civil society or managerial discourse. For each entity, we counted the number of occurrences of keywords on its website.

FIGURE 2. Relational structure of nonprofit performance evaluation

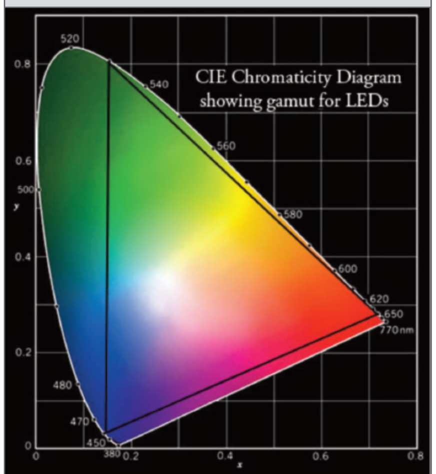


Bulletin of the Association for Information Science and Technology – December/January 2015 – Volume 41, Number 2

Depicting the Linguistic Topography

Representations of three-dimensional color spaces provided the inspiration for an illustration of how organizations combine different discourse components on their websites. The corners of a chromatic triangle are purely red, green or blue, respectively; each point in between the corners represents a specific color blend ratio. The center of the triangle, for example, is constituted by an equal amount of 33.3% of each color.

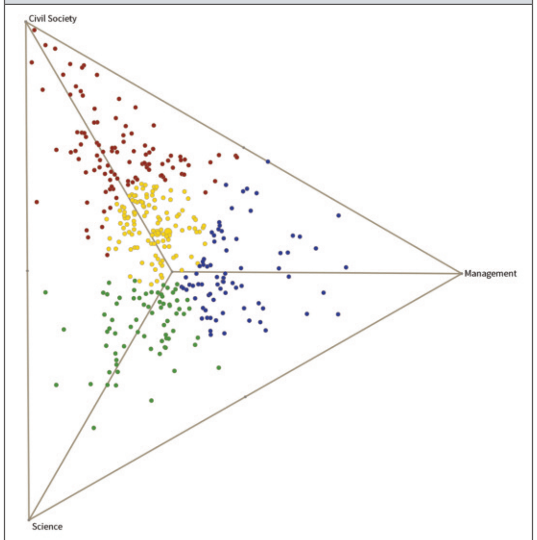
FIGURE 3. Chromatic triangle



Following this approach, based on a ratio among the three discourses, we calculated each entity's discursive position as a point in an equilateral triangle, a semantic space spanned by the three axes of civil society, science and management.

Depending on an organization's location on this triangle, we identify it as belonging to one of four communities. Membership is determined by vocabulary used on webpages. Organizations in the civil society community (red) emphasize "social change," "participation" and "justice." In the science community (green), we find extensive use of such terms as "data," "survey" and "framework," whereas the management community (blue) includes organizations that frequently mention "performance," "efficiency" and "outcomes." At the intersection of these three discourses, identified by calculating the median for each axis, there is a fourth group of organizations whose use of keywords defies categorization into the original communities. Members of this interstitial community straddle domains and recombine diverse concepts into an amalgamated discourse.

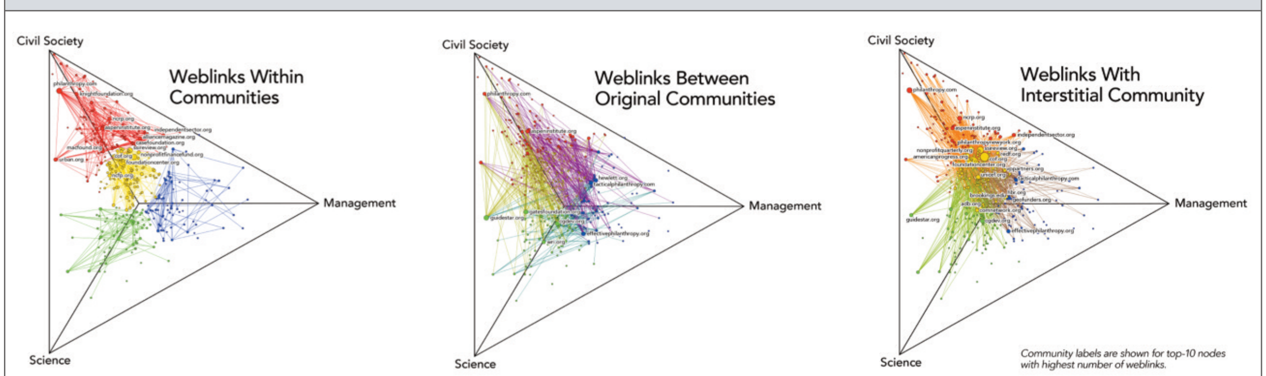
FIGURE 4. Linguistic topography of nonprofit performance evaluation



Interstitial Organizations as Conversational Bridges

The relevance of the interstice becomes evident when network ties are layered on the linguistic topography. This simultaneous representation of the relational and discursive dimensions shows that interstitial organizations not only blend discourses, but also connect organizations across domains. When comparing different types of ties, we observe that a) the interstitial community is more densely internally connected than the other communities; b) connections among the three communities are relatively sparse; and c) organizations in the interstitial community have extensive ties to all communities. Taken together, the visualization in Figure 5 underscores that interstitial organizations recombine specialist terms into an amalgamated discourse that is new and yet familiar, and they sit astride

FIGURE 5. Weblinks within and between discursive communities



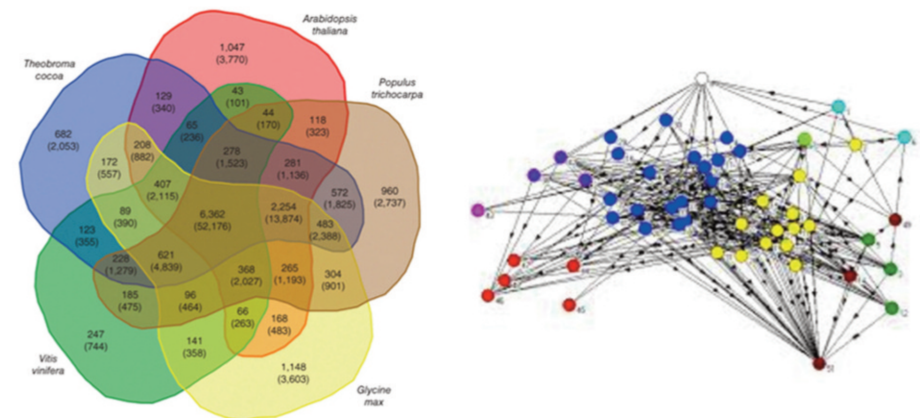
extensive channels through which this narrative is communicated. Organizations in the interstitial community are in both relational and cultural respects at the center of the debate on social impact, becoming conversational bridges among the domains of science, management and civil society.

Map-Reading and Map-Making

The map holds considerable interest for practitioners of nonprofit performance evaluation. The labels identify sample organizations, which gives those familiar with the sector an immediate impression of where they stand relative to their fellow participants. The integration of network and discourse dimensions in the visualization reveals complex network structures that capture an ongoing social transformation. As a contribution to network science, we depict how cultural cues condition how relations are forged, the invitations that are accepted and the relations that endure. More generally, our approach can be applied to study various transformations, both small and large, within organizational fields or other social systems. Data may come in various formats – film or restaurant reviews, initial public offerings, press releases – virtually any source of rich text that could be supplemented with relational data. As an analytical instrument, our visualization is not exclusive to the empirical case of nonprofit performance evaluation but invites broad scholarly application.

A central question for creating a linguistic topography on which ties can be layered is the number of cultural dimensions or discourses to integrate. In our map, we positioned organizations according to their relative use of keywords by locating them on a triangular plane which reflected both the relative content of their websites – how much associational, scientific and managerial language respectively – as well as the resulting cultural distances to organizations exhibiting a different composition. Visualization is, however, not limited to three dimensions. Multidimensional Venn diagrams and multidimensional scaling, respectively, can accommodate a larger number of possible dimensions. For example, Argout [2] uses a Venn diagram to depict the distribution of shared gene families among five plant species, whereas Luczkovich and Johnson [3] apply a multidimensional scaling plot to visualize the relations among 10 groups in a food web

FIGURE 6. Examples of multidimensional visualizations: Argout's Venn diagram [2] and Luczkovich and Johnsons' multi-dimensional scaling plot [3]



Multidimensional Venn diagramm (Argout et al., 2011)

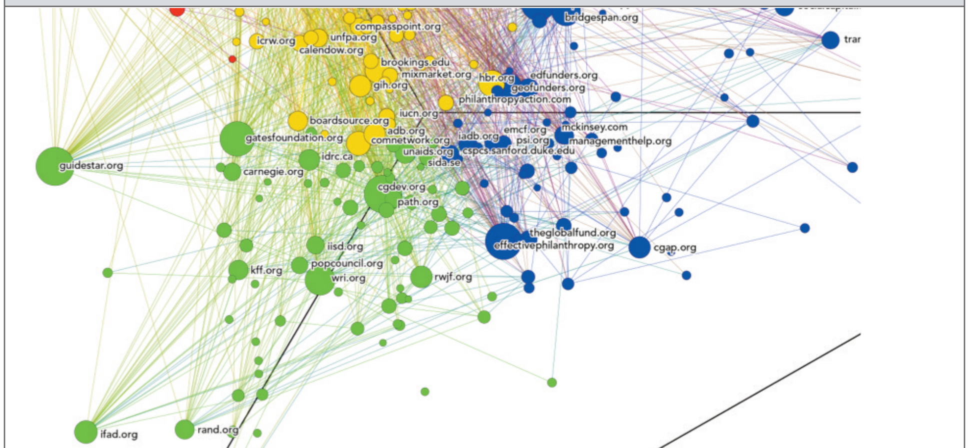
Multidimensional scaling plot (Luczkovich and Johnson, 2005)

network. Such expansion does, however, entail a tradeoff between the two approaches. Whereas Venn diagrams depict the blended content, they do not precisely reflect distances between nodes if the distances are in a space with more than two dimensions. In contrast, multidimensional scaling preserves distances, but reflects less information about composition.

Where the Action Is

The integration of cultural and relational dimensions into one visualization reveals many previously elusive patterns and phenomena. By layering weblinks onto a linguistic topography, we are able to identify cultural areas in which new ties form and clusters emerge. In the case of nonprofit performance evaluation, we find the interstice to be a hotbed in which organizations of diverse backgrounds intermingle and recombine traditional concepts into novel approaches. Such spaces, however, do not necessarily occur only in interstitial zones. It is entirely possible that dense areas of interaction could emerge within an original community or at the borderlands between two domains rather than in an interstice among all

FIGURE 7. Close-up of the borderland between science and management



three. Their absence, in turn, can also be of relevance. Figure 7 highlights that there are few connections – either in language or weblinks – between the managerial and scientific communities, suggesting there is not a general colonization of civil society by a collective scientific and managerial discourse. Instead of a common language of rationalization, the science and management communities speak independently and influences from all three domains come together in the interstitial zone.

The social x-ray as a snapshot allows us to identify spaces where culture and connectivity combine, drawing our attention to potential areas that could be centers of transformation, nuclei for the emergence of new fields or areas of imminent segregation. To know their ultimate fate – whether they form a new structure, change the existing one or dissolve again – the analysis can be extended over time. Once the sample and keyword list has been established, website content (or other textual data) and relations can be recorded at different time points. In doing so, the currently static map transforms into a dynamic visualization – an animated feature exhibiting shifts in structure and meaning. ■

Link to related work: http://pacscenter.stanford.edu/overview/research/powell
Online Version: www.know-your-field.net/metrics

Resources Mentioned in the Article

[1] Powell, W.W., A. Oberg, V.P. Korff, C. Oelberger, & K. Kloos. (2014). Neither contestation nor convergence. The proto-institutionalization of nonprofit performance metrics. Paper presented at the American Sociological Association Annual Meeting, San Francisco, CA.

[2] Argout, X., J. Salse, J.-M. Aury, M. J. Guiltinan, G. Droc, J. Gouzy, . . . & C. Lanau. (2011). The genome of Theobroma cacao. *Nature Genetics, 43*, 101–108.

[3] Luczkovich, J. and J. C. Johnson. (2004). Dynamic Visualization of Estuarine Food Webs. Presentation at the Atlantic Estuarine Research Society Meeting.

Archivist! Data Librarian! Asset Manager! Do the Differences Really Matter?

Reflecting on Breakout Discussions at the Society of American Archivists 2014 Annual Meeting

by Wendy Hagenmaier, Dana M. Lamparello, Karen S. Baker, Janina Mueller and Stewart Varner

EDITOR'S SUMMARY

Prompted by proliferating titles for those charged with managing digital data, archivists, embedded data managers, data librarians and data users explored terminology at the 2014 Annual Meeting of the Society of American Archivists. Digital data creation may originate with a submission to a repository, copies ingested at different locations or even reassembly of existing data. Contrary to conventional concepts for archives, data is not complete but may represent one version, a level in a process or point in a workflow. Data must be accessible throughout versions and stages and often across a network of locations. Additional discussions focused on data ownership and responsibility for stewardship, the need for a common vocabulary to support interoperability by managers in varied roles, and the reincarnation of data as it is reused. The session made evident the critical need to reframe communications among those involved with data management to overcome barriers rising from vocabulary differences.

KEYWORDS

terminology
archivists
librarians
digital libraries
archives
librarianship
document management
data set management

Research data management librarian; digital asset manager; archivist/digital data specialist; born-digital processor; curation archivist; data curator – all of these positions have come online recently to address the explosive growth of digital data. What they all have in common – despite their varying titles – is digital data management. So why the variety of titles? Are we using different names for the same work? We developed this session at the Society of American Archivists 2014 Annual Meeting [1] as a series of breakout discussion groups around the themes of data creation, access and reuse to highlight areas in which data management roles overlap. The panelists represented four archetypal roles commonly encountered in the digital data world – archivist, embedded data manager, data librarian and data user – but were motivated by a desire to break down barriers extant among such positions and to explore the diversity of data and information needs in practice. Our session was intended as a first step in collectively developing a common conceptual understanding of semantics and roles to bridge disparate professional communities, including the archives community, research data management community, digital curation community and digital humanities community, among others. This column is a brief overview of the session, but we encourage you to join the continuing discussion by commenting here: goo.gl/yCRHqG

Data Creation

Digital data creation is a complex phenomenon encountered by the archetypes in differing circumstances. Researchers, for instance, create data in the field and laboratory. Yet dataset creation may also be recognized as occurring at the moment of repository submission. Or given the reproducibility of digital data, copies of the same dataset may reside in several different repositories where each views ingestion as a moment of creation. And finally, data creation may be identified during assembly of a new collection from pre-existing data.

Wendy Hagenmaier, Georgia Institute of Technology, can be reached at wendy.hagenmaier<at>library.gatech.edu; Dana M. Lamparello, Chicago History Museum, can be reached at lamparello<at>chicagohistory.org; Karen S. Baker, University of Illinois at Urbana-Champaign, can be reached at ksbaker2<at>illinois.edu; Janina Mueller, Harvard Design School, can be reached at jmueller<at>gsd.harvard.edu; and Stewart Varner, University of North Carolina at Chapel Hill, can be reached at svarner<at>email.unc.edu.

Across the archetypal roles, two concepts represent a shift in traditional archival thinking: awareness of data versions and data levels. With data versioning, the "doneness" of data is no longer taken for granted. Rather than producing final data records, research practices may result in the creation of many versions of a dataset due to corrections or upgrades. There also may be a multitude of data products resulting from transformations of the data. Data creators have developed data levels, distinguishing raw, refined, derived and interpreted data. Each level represents a well-defined step in the processing, transformation or presentation of data. The concepts of data versions and levels are foreign to many archivists and digital preservationists.

Data Access

Currently, each archetype is grappling with several challenges that complicate access to digital data, including the unprecedented quantity of data and the legal tangle of intellectual property issues endemic to such volume. Adding to the complexity is the evolutionary nature of digital records: as aforementioned, data are no longer "done" or definitively inactive; our roles must shift from providing access to discrete records to providing access to interactions and dynamic entities. Other current challenges include the seemingly impossible task of architecting systems that anticipate future users' access needs and provide meaningful access to levels of data through sustainable repository workflows. In our breakout groups, we discussed how providing creators with early organized access to data during the creation process is our entrée to encouraging their participation in curation efforts. And to ensure future access, we identified the need to become involved in data creation, policymaking and system building.

In order to address these challenges and transitions, ideal data access should involve network models – networks of discovery, data stewardship institutions and data professionals. The definition of access needs reshaping, as it will no longer mean access to one-stop-shop institutional or disciplinary repositories that must contain all records, but rather access to a network of linked information sources. Portable, replicable, linkable data (and metadata) mean that data will be accessible in different places and discovery will unfold in myriad locations. Digital data will become a renewable resource living in an ecosystem of repositories.

Additionally, a need exists to reevaluate how we define ownership (that is, not always as an exclusive right) and stewardship (not purely institution-based but repository-based, inclusive of national, global, inter-institutional, intra-institutional and community efforts). This emerging network model takes alliances to a new level, one that may be uncomfortable for traditional archives and libraries. It is a model that requires additional infrastructure and a change in mindset to create interoperability among institutions. Above all, the network model demands a common vocabulary to empower each archetype to collaborate much more closely than ever before. And yet, despite significant shared priorities and responsibilities, data managers are not always doing the same work; that is, specialization of roles can be very important. But our specialized roles must be interoperable and speak a common language.

Data Reuse

Understanding how and at what points data are created and ensuring their accessibility by redefining traditional concepts of access are the first steps to ensuring their future use. Emerging technologies and proliferating modes of scholarly discourse are greatly expanding potential uses and creating new kinds of data users driven to the archives in search of raw materials for new projects. Scientific data may be used to replicate the results of an experiment or they could form the basis of an entirely new line of inquiry. Digital humanists looking for linguistic or social patterns across centuries of newspaper articles can mine digitized text collections. Historical social networks may be reconstructed and visualized based on information extracted from archived correspondences.

With the energy and excitement surrounding open data and the digital tools used to wrangle it, new methods and techniques will continue to emerge, and researchers will continue to find new ways to put archived data to work. Because we cannot anticipate every possible use, managers of digital data must maintain active and open lines of communication with end users and work together to push the limits of our collections. Of course, such collaborations will produce new data that will need to be accessible and ready for reuse, thereby establishing a new data creator, a new point of data creation and a new data level.

Conclusions and Invitation to Join the Conversation

The major takeaways from our discussions highlighted the following:

1) the iterative and cyclical nature of digital data work and the flexibility required for ever-evolving records;
2) the need for a common vocabulary to bridge professional divides and enable interoperability among specialized data roles, including reconsideration of key, traditionally archival concepts; and
3) a glaring need to reframe communication among digital data managers, creators and users as a natural part of data workflow, rather than treating it as a difficult crossing-boundaries effort.

Additionally, we recognize each archetype came from a fairly well-funded academic institution, which strongly impacts available resources and expectations in digital data arenas. We plan to include more environments, archetypal roles and types of data as we continue these discussions. Again, we encourage data managers from all types of environments and perspectives – including you – to join the discussion: goo.gl/yCRHqG. ■

Resources Mentioned in the Article

[1] Archivist! Data Librarian! Asset Manager! Do the Differences Really Matter? (August 16, 2014). Session 708, Society of American Archivists 2014 Annual Meeting. Description available at http://sched.co/1qX7PB6