



## **Workshop Report Series**

# Information, Interaction, and Influence

Research Information Technologies and their Role in Advancing Science

**Robert Mitchum**, Computation Institute, University of Chicago

**Amy Brand**, Digital Science

**Caitlin Trasande**, Digital Science

**JULY 2014**

## About Digital Science

**Digital Science** is a technology company serving the needs of scientific research. We offer a range of scientific technology and content solutions that help make scientific research more efficient. Whether at the bench or in a research setting, our products help to simplify workflows and change the way science is done. We believe passionately that tomorrow's research will be different - and better - than today's.

Visit [www.digital-science.com](http://www.digital-science.com)

This report has been published by Digital Science, a business division of Macmillan Science and Education.

The Macmillan Building, 4 Crinan Street, London, N1 9XW  
[info@digital-science.com](mailto:info@digital-science.com)

Copyright © 2014 Digital Science

# Information, Interaction, and Influence

## Executive Summary

For as long as there has been science, there have been data. Researchers collect observations and information about the world around them, and use theory and statistical methods to extract knowledge. In the 21st century, more and more fields grow increasingly quantitative and digitized, with larger and more complex datasets driving discovery in the natural sciences, social sciences, and humanities. But one area where the potential of data remains unfulfilled is the process of science itself -- how research is planned, funded, disseminated, and tracked.

A possible solution may be found in research information technologies, such as research profiling, data and publication management, and networking systems. To draw attention to these emerging tools and discuss their potential and challenges, [Digital Science](#) and the [Computation Institute](#) (a joint initiative of the [University of Chicago](#) and [Argonne National Laboratory](#)) organized the [Information, Interaction, and Influence](#) conference, held at the University of Chicago on May 19th and 20th.

Researchers, entrepreneurs, software developers, foundation representatives, administrators, and IT professionals from the United States, England, and Australia attended the workshop and shared their experience building and using research information technology. Panels and talks also discussed barriers to wider implementation of these new research tools, the experience of commercializing software and services in this space, how these tools can benefit scientists and institutions, and early examples of how this technology facilitates improved scientific collaboration and understanding of academic influence and impact.

Recurring themes at the workshop included, 1) the growing demand for open access to research articles and open sharing of data and software used to generate the findings; 2) alternative metrics of scientific impact that capture newer forms of scientific communication, such as published data, software, videos, social media, talks, and news articles; 3) the increasing use of metrics to help funding agencies and administrators choose and track researchers; 4) the creation of smarter profiling systems that promote collaboration within and between institutions; 5) the challenges faced by scientists transitioning between academia and entrepreneurship; 6) the cultural challenges of multidisciplinary research across fields with different perspectives and incentives on the use and sharing of data; 7) how to encourage wider implementation of research information technology, and how to design tools that best address scientist's needs.

Conclusions and recommendations from the workshop include:

- Reach out to fields not traditionally steeped in computational and quantitative methods, so that their needs are also served by research information technology
- Conduct more research on whether and how scholars use these tools, and further understand the needs of scientists that such technologies can address.
- Improve outreach to graduate students, postdoctoral scholars, and early-career scientists; populations who have the greatest need for networking and may benefit most from these systems.
- Work with institutions to better understand and facilitate entrepreneurship as the lines between academic and commercial work grow increasingly fluid.
- Create more forums that bring together participants from different spheres: research, administration, libraries, information technology, entrepreneurship, academic publishing, and funding.

## Workshop Background

Digital Science is a technology company that develops and funds software for scientists to make discovery more efficient. The company invests in promising startups and small businesses -- many founded by former academics and researchers -- producing research technologies, such as [Altmetric](#), [figshare](#), and [ReadCube](#).

The Computation Institute (CI) is a joint initiative of the University of Chicago and Argonne National Laboratory, founded in 1999 to realize the potential of computation to advance and accelerate discovery in all fields of research. CI research centers apply advanced data and computation approaches to topics such as climate change, the design and management of cities, supernovae and other astronomical phenomena, the creation of knowledge, and biology, chemistry and medicine. The CI also develops new tools to facilitate research, including the data transfer and management service [Globus](#), the parallel programming language [Swift](#), and the new profile system for the University of Chicago.

With a shared interest in the future of research information technology, Digital Science and the Computation Institute united to co-organize a workshop on the benefits and challenges of these novel scientific tools. [Amy Brand](#) and [Caitlin Trasande](#) of Digital Science came together with [Tanu Malik](#) and [Eamon Duede](#) of the Computation Institute and planned the workshop for May 19th and 20th, 2014 at the University of Chicago, giving it the name Information, Interaction, and Influence.

Together, the organizers drafted the following description/agenda for the workshop:

*This workshop will explore the role that research information technologies – including research profiling, management, and networking systems – play in improving scientific collaboration and our understanding of academic influence. The key objective is to surface the challenges that exist in conceptualizing and designing systems for managing research and scholarly information, and to better understand the vast potential of these systems. The workshop aims to break down barriers – barriers among academic and administrative silos within the university, and barriers between commercial and university-based efforts to develop tools that improve the management and discoverability of networked research information.*

**Systems that harvest, manage, and network research information increasingly serve as knowledge warehouses for the research enterprise.** *Given their potential to capture some of the complex institutional and social dynamics of science and scholarship, these systems are of active interest to funders, investigators, scholars, clinicians, academic administrators, community partners, and entrepreneurs. Of particular interest are emerging systems to mine the hidden knowledge networks embedded in the scholarly literature, and systems to track new measures of impact and influence in science.*

*As the academic publishing landscape continues to evolve and authors increasingly turn to non-traditional media for the dissemination of their research, the cloud is becoming an equal partner in enabling scientific discovery. Publishing a paper is no longer sufficient; underlying data and the computational results used to produce the paper are equally vital, and reproducibility is as important as the publication itself.*

To address these points, organizers raised several preliminary questions regarding research information and data management:

- How do we find the most impactful research and researchers?
- How do we identify experts to review manuscripts and funding proposals?
- How do we keep up with significant advances in our chosen fields?
- How do we search for work or authors of interest?
- What is the relationship between internal-facing administrative systems for the management of research information and external-facing profile and discovery tools?

- What are the incentives for researchers to use research networking platforms?
- How can we track the complete record of scholarship for an individual or an institution, going beyond articles and books to include research data and software?

More than 100 people attended the workshop over the two days, representing 23 academic and government institutions, 13 companies, and three foundations.

## Introductory Remarks

[Ian Foster](#), Director, Computation Institute

Foster kicked off the conference with a quote from legendary physicist Richard Feynman: “Science is a way of trying not to fool yourself.” The scientific method is based around testing theories with data, Foster said, and revising those theories when the observations proves them wrong.

While this rigorous approach is standard in nearly every field of research, it is rarely applied to the practice of science itself -- largely due to a lack of data. But as science and knowledge are increasingly digitized, there is increased opportunity to study and improve the scientific process through information systems, computational techniques such as machine learning, and other methods.

As an example, Foster pointed to computer scientist [Don Swanson's concept of “undiscovered public knowledge.”](#) the idea that important scientific connections are often overlooked because critical pieces of information are siloed into separate disciplines and publications. In 1986, by manually searching literature databases, Swanson discovered that fish oil was a promising treatment for Raynaud's disease. Today, that process could be automated and expanded to millions of publications using various computational techniques.

However, Foster reminded that challenges remain in the development of these tools. Information systems are often designed around administrative priorities, but can also be used to advance research, if used properly. **As with all software, it's important to create scalable, usable, and sustainable information systems to insure a long life and wide implementation.** And the larger these datasets grow, the more sociological, technological, and geopolitical obstacles arise. Foster emphasized the need for partnerships between academic and private sectors to make sure digitally-mediated research is increasingly transparent for mutually beneficial purposes.

# How Science is Different: Digitizing for Discovery

[Victoria Stodden](#), Assistant Professor of Statistics, Columbia University

Stodden agreed that now was an important “tipping point” for the digitization of science, raising urgent issues of transparency, public participation, and technology. Certain models outside the sphere of science, such as the open source movement in software and Creative Commons for media, may offer some guidance as science confronts this important shift. But these solutions are not perfectly suited for the challenges faced by increasingly digitized and open science, Stodden said.

Science in all fields is increasingly computational, but scientists have yet to reach consensus on standard practices for sharing the data and software used in computational research. If computation is now thought of as the third branch (or the fourth paradigm) of science, it must develop standards as robust as those used for hundreds of years in the the branches of theory and experimentation. Central to both classic branches of science are reproducibility and verifiability, the ability of other scientists to recreate and re-test the original work.

But in computational studies using large, often closed data sets and proprietary or homemade software, methods sections are often insufficient. Stodden quoted [Jon Claerbout](#) (paraphrased by [David Donoho](#)) on the topic of “really reproducible research,” saying:

*“An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete... set of instructions [and data] which generated the figures.”*

This standard of reproducibility inspired the creation of tools to help scientists manage their computational research workflows and communicate their work to the broader community. Tools such as [Research Compendia](#), [figshare](#), [iPython](#), [Galaxy](#), and [knitR](#) help computational scientists disseminate their work, track their workflow and research environments, and publish articles embedded with code and software. Unfortunately, while many different disciplines — such as math, psychology, geoscience, computer science, and law — are moving towards new ways of publishing more reproducible research, there is minimal communication between those disciplines on the solutions they discover.

This shift within science is accompanied by growing external pressure from the government and the public to make scientific findings and data more accessible. An [executive order from the White House in 2013](#) directed federal funding agencies to develop plans for public access to data and publications — a plan which still awaits implementation. But the order sparked discussion about open data, transparency, and treating data as a digital scholarly object. Besides the philosophical debate, there are also technical concerns about building the infrastructure and methods needed for scientists to publish and curate data and code.

Legal issues also threaten the openness of science; as Stodden said, copyright is “orthogonal” to science. While [a 1991 U.S. Supreme Court ruling](#) that people cannot copyright “raw facts” implies that data is not subject to these restrictions, software is more legally complicated. Stodden suggested that science look to the open source software movement, which allows a community of programmers to modify, improve, and adapt code, or the [Creative Commons](#) licenses developed by Larry Lessig to allow open sharing of photos, movies, and sound with creator-defined restrictions. Stodden proposed a new “Reproducible Research Standard” of licenses that scientists can use to control the use of their published data and software.

One important payoff of creating a more digitized, transparent, and open science is a dramatic expansion of the field’s boundaries. The swelling movements of crowdsourcing and citizen science can engage with more than just data collection, Stodden said, learning how to synthesize and curate data and use advanced

analytics to produce results. Another side effect of this increased participation may be improved public trust in science, informing debates about evidence-based policy, medicine, and other areas. It's important to facilitate this outreach in a way that creates trust instead of exploiting gaps in knowledge, and communicates that science is an active exploration and search for truth, Stodden said.

## *Conclusions*

- Computational science must develop tools and standards for publishing code, software and methods that allow for true reproducibility.
- Government and public pressure for open access to publications and data must drive new platforms and methods for scientists to disseminate these research objects alongside findings.

But bottom-up practices driven by the scientific community will also be essential to broad acceptance of new standards for publishing and reproducibility.

- Guidance on the legal issues of open science can be found in open source software and media movements; new “Reproducible Research” licenses are one possible solution to balance access with ownership. Fields that have a longer tradition of data-heavy research, such as astronomy or physics, may also offer guidance to disciplines just now moving into this space.
- If done right, open science will facilitate public participation in research and may also increase public trust in the scientific process.

## Founder Stories

Moderator: [Robert Rosenberg](#), Director of Entrepreneurship Program, Polsky Center for Entrepreneurship and Innovation, University of Chicago Booth School of Business

[Daniel Hook](#), Co-Founder of Symplectic & Director of Research Metrics, Digital Science

[Stephen Leicht](#), Managing Director and Co-Founder of UberResearch

[Euan Adie](#), Founder and CEO of Altmetric

[Robert Lowe](#), CEO of Wellspring Worldwide

[David Beiser](#), Associate Professor of Medicine, University of Chicago & Co-Founder of Qualia Health

Much of the innovation in research information technology currently comes from outside of academia, developed by startups and established private companies. But many of the founders of these enterprises have prior experience as academic researchers, either as graduate students, postdoctoral scholars, or faculty. That blend of scientific and business experience creates a unique perspective on the tools that researchers, funding agencies, and administrators need the most. But panel moderator Robert Rosenberg was also interested in hearing about the challenges that each founder faced as they made the transition from the research sector to the private sector.

Some of the panelists made the leap after coming up with a gem of an idea as practicing researchers. Euan Adie, founder and CEO of [Altmetric](#), was a bioinformatician who wondered why the in-depth research blogs he read were not counted alongside scientific publications in measures of impact. David Beiser, an assistant professor of medicine at the University of Chicago, saw the potential of new data streams from wireless home health devices, mobile computing, and environmental sensors to inform a new definition of health and change the practice of medicine.

But the panelists reported difficulties in building a business while continuing their academic work. [Symplectic](#) was founded by Daniel Hook and three officemates initially to build a content management system for a faculty member. However, they found that the university did not initially trust a product made by its graduate students. Beiser voiced concerns about how his superiors and peers would judge the non-traditional work of starting a business, instead of more typical research pursuits. The long hours and late phone calls of entrepreneurs also take their toll on people trying to juggle both careers, many panelists said.

Conversely, there were also advantages to launching a business in an academic environment. The campus setting is full of intelligent, motivated people who like to work on challenging problems, Adie said. Beiser cited the “random collisions” on campus that helped him develop his idea into a business, such as the [New Venture Challenge](#) at the University of Chicago Booth School of Business, and the [Chicago Innovation Exchange](#), a new, shared workspace for university and community startups. Robert Lowe of [Wellspring Worldwide](#) cited the low success rate of most entrepreneurial efforts, but said that a motivated academic can develop their idea on the side and see if it is gaining traction before jumping full-time into a startup.

From another perspective, the research world is also a promising customer base for new products, many panelists said. [UberResearch](#), co-founded by panelist Stephen Leicht, helps funding agencies analyze and categorize their portfolio of researchers to streamline and improve reporting and decision-making. Their recent work has focused on bringing the types of complex analytics that were once only available to large funders to the many smaller organizations that previously could not afford the service. Rob Lowe, a former economics professor at Carnegie Mellon University, co-founded Wellspring Worldwide to help research and development institutions with their “knowledge supply chain” -- how they manage their patents and discoveries and bring them to market.



Panelists provided important advice for would-be founders moving from academia to the private sector. Many cited the importance of designing for customers and end users -- an unfamiliar concept for scientists used to fulfilling the milestones of a grant and publishing papers.

*“In industry, we want everything to be validated by a customer,” Adie said. “For a project to be successful, you need buy-in from the people on the ground rather than the people who ‘know best.’”*

Focus groups, improved user interface, and sales strategies could be the difference between a successful and failed business. The panelists also emphasized that entrepreneurs should pursue ambitious ideas that make the challenges worthwhile; as Lowe said, “if it isn’t risky or uncertain, you haven’t picked the right thing to market.”

## **Conclusions**

- There are challenges to founding a company within a university, including time constraints, lack of trust and respect from superiors and peers, and difficulty reaching funding and research goals.
- However, an academic setting also provides benefits for startups, such as entrepreneurship programs and competitions, and a rich talent pool for potential collaborators.
- Universities and other research institutions are receptive but challenging markets for new technologies, particularly those that increase efficiency on arduous or expensive tasks.
- For-profit pursuits prioritize the customer experience -- often an unfamiliar and difficult shift for academics used to fulfilling funder and discipline expectations.

# Research Networking and Profile Platforms: Design, Technology and Adoption of Networking Tools

Moderator: [Tanu Malik](#), Research Associate, Computation Institute

[Griffin Weber](#), Head of the Knowledge Discovery & Management Group, Center for Biomedical Informatics, Harvard Medical School

[Bill Barnett](#), Director, Science Community Tools, Indiana University

[Bart Trawick](#), Literature Resources Lead, National Center for Biotechnology Information

[Leslie Yuan](#), Director, Virtual Home, UCSF School of Medicine

[Simon Porter](#), University Administration and Support, University of Melbourne

Silos are one of the primary obstacles to scientific progress. Too often, scientists stay within familiar circles, collaborating only with people they know and trust inside their own discipline. But a solution to this research isolationism may come from an unappreciated tool -- the online profile. Just as social media sites created professional and personal relationships across long distances, new profile platforms offer scientists fresh opportunities for research networking, finding potential collaborators both within and beyond their institution. These profile systems also enable new metrics for administrators to assess their workforce, improved public and media engagement with scientists, and a richer format for researchers to display their accomplishments.

These technologies are another dimension of treating science as an object of study, said panel moderator Tanu Malik. They hope to address the related objectives of organizing and curating the increasingly diverse and digitized forms of scientific communication, while also providing a platform for analysis of the relationships within science. The computational challenges are large, but inspiring widespread implementation of these systems by researchers and institutions may be even more difficult.

Still, early successes were on display during the panel. A leading platform is [Harvard's Profiles](#), an open source tool described by Griffin Weber. Built using the ontology system [VIVO](#), Profiles was built for the monumental task of creating a common directory for Harvard's biomedical researchers -- some 20,000 faculty over 33 health centers. UCSF adapted Harvard's platform to build [a new profile system](#) for their biomedical faculty and staff, while the University of Melbourne put together their "[Find An Expert](#)" system for their entire faculty, also using VIVO.

All three representatives of these schools emphasized the importance of early buy-in from the subjects of the profile system itself: the researchers. While initial information can be gathered from HR resources or, in the case of Australia, a national scientific reporting policy, sustainability and participation from scientists and administrators is required to improve the worth of the system. As Leslie Yuan of UCSF said, in a twist on the film *Field of Dreams*, "If you build it, they will not come." Selling the product requires providing data, analyses, and customizability that attract users.

For Yuan's team, which mostly consists of people from outside of academia, that involved sending a yearly e-mail of personalized statistics to each researcher in the system, including information about how many people viewed their profile, and developing an "on-ramping" protocol with HR for new faculty and staff. The UCSF system also built in search engine optimization so their profiles would appear among the top three Google searches for each researcher, raising their visibility to the public and media. Expanded functionality of these profiling systems -- allowing researchers to post "non-traditional" content such as videos, social media, and slides from talks -- also attracted more participation.

On a deeper level, these systems provide new information about researcher relationships -- or potential relationships -- through advanced metrics and visualizations. The Harvard Profile platform automatically uses a researcher's publications to build a rich, detailed history of their specialties and areas of research,

even building timelines to show how those focuses change over the subject's career. Advanced search capabilities draw upon these categories to make it easier for users to find researchers in their topic of interest, even when that topic is one that spans disciplines, such as disaster management. The hope is that these features will create more serendipitous connections, increasing collaboration within and without the institution.

*“We believe that the maximum number of research opportunities are possible when we can maximize the number of people discovering or engaging with our research,” said Simon Porter.*

While these systems have worked well within a single discipline, panelists warned of the obstacles facing the expansion of profiling platforms across an entire campus. Different fields have different priorities, requiring system developers to find new ways of importing the right information. While life scientists care most about publications, other fields may want their profile to feature accomplishments such as discovering asteroids or arguing cases before the Supreme Court, Weber said, challenging developers to find new data sources for automated profile input. Departments often want their profiles to match the look and feel of departmental websites, requiring different skins for different areas. And early efforts to link profiles across multiple institutions, such as [Direct2Experts](#), create an additional degree of difficulty, as universities want to keep some of their data private and in-house.

Bill Barnett of Indiana University also sounded a note of caution that there remains very little hard information on how these new profile systems are used -- or not used -- by researchers. He suggested more research on whether these systems do in fact increase collaboration or expand research partnerships beyond traditional silos. Some fields may also be faster to embrace these systems than others; facilitating collaboration may be a natural fit for biomedical researchers, but more solitary fields such as humanities and social sciences may be less likely to see the value of enhanced profiles.

A different, top-down variant of a profile platform was presented by Bart Trawick of the [National Center for Biotechnology Information](#). While most researchers know the NCBI for hosting the publication database PubMed, they have also created a dashboard with multiple tools for helping researchers easily handle the requirements of their federal grants. [My Bibliography](#) allows scientists to attach publications to specific grants, providing information to funding agencies about return on investment and compliance with open access policies. [SciENcv](#) is an automated tool that generates the proper bio sketch for NIH grant applications, a service that the NCBI plans to expand to fulfil the different requirements asked for by each federal agency. In a government setting, these concessions to the unique needs of different agencies is essential for the success of these profile tools, Trawick said.

## Conclusions

- New, enhanced profile systems have the potential to break down scientific silos, increase collaboration, and assist the study of science as an object and the creation of knowledge.
- For a new profile platform to thrive, researcher buy-in is essential. Systems at Harvard, UCSF, and the University of Melbourne have proven their worth with new functionality, metrics, and customization options for researchers and administrators.
- Expansion of these profile systems beyond a single department, division, or institution can be very difficult, with different fields and schools wanting different features, unique branding, and data security.
- Despite early successes, more research is needed to determine how and if scientists are using these profile systems, and if there are differences in uptake by discipline.
- Federal agencies are also exploring profile systems and tools to help researchers navi-

gate enormous libraries of publications and stay compliant with the requirements of their grants. But more open APIs are needed to capture researcher information automatically and minimize slow and tedious data entry.

# The Administrative Perspective: Managing and Disseminating Academic Research Information

Moderator: [Oren Sreebny](#), Senior Director for Emerging Technologies & Communications, University of Chicago Information Technology Services

[Mitra Dutta](#), Vice Chancellor for Research, University of Illinois at Chicago

[Elisabeth Long](#), Associate University Librarian for Digital Services University of Chicago

[Jeremy Manier](#), Director of the News Office, University of Chicago

Many of the new research information technologies are useful as well for parties beyond the researcher. Administrators, librarians, IT staffs, and communicators can also find novel uses for tools that organize and curate large amounts of research data, removing inefficiencies and annoyances in their own work and providing new ways of supporting researchers in theirs. Just as science grows increasingly digitized, these fields too are in the midst of dramatic technological changes, some of which will make it easier to participate in the process of disseminating, archiving, and promoting scientific research.

Mitra Dutta of the [University of Illinois at Chicago](#) said her two primary goals as an administrator are reducing the administrative burden on researchers, and keeping track of the research landscape at the institution. Profile systems would directly address both of these priorities, as well as helping administrators easily find the right researchers to recruit or organize into a multidisciplinary effort and track faculty productivity -- an often sensitive subject. But while profile systems are attractive, their expense may also be beyond the budget of some institutions, leading Dutta to look at open source solutions. At UIC, the implementation of simpler tools such as a PI dashboard for researchers to easily track grant status provide some administrative relief.

Not long ago, university IT departments ran academic computing centers where faculty and students visited to conduct their computational research. But Oren Sreebny said that the more distributed and data-heavy environment of today's science has IT staffs re-examining how they can best help researchers in their work.

*"The IT environment of academic research today is more distributed and data heavy," Sreebny said. "One way that we can help is in thinking about how to enable those collaborations that happen across the institution and across the globe."*

In addition to current goals of reducing administrative burden and enabling collaboration, the impending open access policies of federal and state governments may create an important for these departments, Sreebny said. By building data repositories and managing metadata for datasets, IT departments can help make sure research information is compliant with new standards and widely available beyond campus.

Libraries have always captured and preserved research outputs, traditionally in the form of publications, monographs, and personal notes. But in the faster, digital world, they are confronted with a dramatically wider range of new forms to archive, including datasets, software, talks, and websites. One challenge, Elisabeth Long said, is the lack of uniform, unique IDs for individual researchers and digital objects, though initiatives such as researcher ID system [ORCID](#) and national systems in Australia, Japan, and the Netherlands are promising. Libraries are also increasingly exploring machine-assisted classification for the flood of digital materials coming their way, given that traditional (yet effective) procedures of human-driven categorization cannot scale up to meet demand.

News and communications officers are charged with disseminating important research findings beyond the scientific community to the general public and media. But like libraries, the methods for accomplishing

this task are dramatically different from just a decade ago, said Jeremy Manier, evolving from a world of faxed press releases to websites, social media, and multimedia. Manier said the mission of his office has also changed from publicity to facilitation of global conversations and collaborations, steering his staff towards new channels such as MOOCs and webcasts. In many modern formats, the success of a story is often driven not so much by the science within as by narrative, often a missing element in many of the current generation of research disseminating tools.

Each panelist emphasized that no single department can handle all of these changes and initiatives alone, and that institutions must bring all stakeholders to the table to plan and implement new research information technologies. Building new data repositories, for instance, will require input from researchers, IT professionals, librarians, legal staff, and other relevant parties. **Working together on these issues will raise the probability of wide acceptance and user satisfaction, and prevent redundancy and wasted resources.**

### *Conclusions*

- Administrators want to reduce paperwork burden on faculty and monitor institutional research. New profile systems will be useful for recruiting, assembling multidisciplinary initiatives, and measuring faculty productivity, but are currently too expensive for some institutions.
- IT departments are re-examining their role under forthcoming open access policies, and considering leadership in building campus-wide data repositories, creating consistent metadata, and ensuring wide distribution.
- Libraries have always captured and preserved research outputs, but the definition of these materials is rapidly expanding in the digital age. There is an urgent need for machine-assisted categorization tools and better ID systems for people and objects.
- News offices are using new channels such as social media, MOOCs, and webcasts to broaden public reach and facilitate collaboration and conversation. Narrative is important in the success of a news story, is it also relevant to research information tools?
- Interconnectedness and communication between these various institutional sectors and others will be important in the successful implementation of new profiles, data repositories, metadata, and privacy policies.

# The Postdoctoral Perspective

Moderator: [Rebecca Bryant](#), Director of Community at ORCID

[Sean McConnell](#), Postdoctoral Scholar and University of Chicago PDA co-President

[Irene Gallego Romero](#), Postdoctoral Scholar and University of Chicago PDA co-President

[Erin Thomas](#), Gender Diversity Specialist, Argonne National Laboratory

[Daniel Spiess](#), Assistant Director, International Activities and Postdoctoral Affairs

[Kristine Henne](#), Postdoc Program Coordinator, Argonne National Laboratory

Graduate students, postdoctoral scholars, and new faculty are often overlooked demographics when discussing the world of research. However, these **early career scientists may actually be the population best served by and most accepting of new research information technologies**. Young researchers are more likely to use social media, disseminate research findings through non-traditional formats, and are typically more receptive to open data and publishing in open access journals. In order to succeed, research technology tools must serve these groups as well as more established scholars.

In the last panel of the workshop's first day, Rebecca Bryant from the open researcher ID system [ORCID](#) moderated a panel of postdoctoral researchers and administrative staff that work with them at Argonne National Laboratory and the University of Chicago. Bryant said that postdocs can sometimes be an "invisible" population, so much so that there are no hard statistics even on how many work at U.S. institutions. The short-term duration of a postdoctoral position leads to long hours, high career pressure, and isolation -- even from other researchers and resources at the same institution.

Sean McConnell and Irene Gallego Romero, co-presidents of the [University of Chicago Biological Sciences Division Postdoctoral Association](#), said they use online tools such as [ResearchGate](#), [myNCBI](#), and [Twitter](#) to network, keep up with their field, manage and find relevant publications, and demonstrate non-journal article accomplishments such as presentations and datasets. Gallego Romero said that connectivity is largely no longer an issue for her, and she would now like to see more tools built to improve productivity, such as making it easier to submit datasets to the NIH. Both said there also remains unmet demand for gaining more teaching experience and training, and building collaborations through face-to-face interaction -- needs that are unlikely to be addressed with digital tools.

Erin Thomas and Kristine Henne of [Argonne](#) cautioned that not every postdoctoral scholar was social media savvy, particularly within minority and international populations. To help these groups avoid isolation, establishing institutional and cross-institutional programs that provide opportunities for networking, finding a mentor, and other forms of support may be more effective than new technologies. Postdocs from all backgrounds and fields could also use more career guidance, said Daniel Spiess of the University of Chicago, whether it take the form of more information about non-academic careers, connecting young researchers with their school's alumni network, or helping them improve their "soft skills" of communicating research.

One place where research information technology may benefit young researchers and the institutions who employ them are new metrics that better evaluate researcher productivity and the impact of their work. These improved measures will give postdocs a new way to prove their worth during the job hunt, Henne said, and also help institutions track the performance of their postdocs, either after they leave for a new job or are retained as permanent hires.

## *Conclusions*

- Postdoctoral scholars, graduate students, and new faculty are an often “invisible” demographic within the research world that new research technologies should address and serve.
- While many young researchers are adept at using online tools for networking, managing publications, and promoting their work, there are also many who remain isolated, particularly minority and international researchers.
- Many of the needs for postdocs and other young scientists are difficult targets for digital tools: teaching experience, face-to-face interaction to meet and build trust with potential collaborators, interpersonal skills that will help find new career opportunities.
- New metrics and productivity tools can help postdocs accomplish more in their short time with an institution and prove their worth to potential future employers.



# Digital Science Portfolio Demos

Moderator: [Caitlin Trasande](#), Head of Research Policy, Digital Science

[Dan Valen](#), Product Sales Manager, figshare

[Daniel Hook](#), CoFounder of Symplectic; Director of Research Metrics, Digital Science

[Stephen Leicht](#), Managing Director and CoFounder of UberResearch

[Euan Adie](#), Founder and CEO of Altmetric

[Adrian Stanley](#), Vice President of Global Business Development, ReadCube

The second day of the workshop began with brief demonstrations of products supported by Digital Science. Many of these technologies addressed topics that drove much of the first day's discussions, such as the shift towards open access in science, the needs of researchers, funders, and administrators for new impact and productivity metrics, and the challenge of managing the rapidly expanding flow of journal publications.

[Symplectic Elements](#): A rich source of faculty information for reporting and showcasing research. Elements is a system that can capture information about faculty researchers at an institution for use in reporting, compliance with open access policies, tracking publication impact, and highlighting interesting research through profiles or public relations. The platform uses as much automated data as possible -- fed from [Web of Science](#), [Scopus](#), institutional human resources departments, and federal grants -- to reduce the paperwork burden for both researchers and administrators.

[Altmetric](#): Providing a broader view of research impact that better reflects how research is used and talked about in the increasingly digital scientific world. Currently, Altmetric primarily works with publishers such as Nature, Science, and JAMA to help them provide detailed data to authors about the spread of their work, which researchers can then use in grant and faculty promotion processes. Altmetric also provides free metrics to institutional librarians and IT departments, and is now building a system for institutions to monitor social media, news articles, patents, policy documents, and other non-traditional outlets for research impact.

[ReadCube](#): Software to help researchers stay on top of relevant literature despite the flood of new articles and journals published every day. Formed originally to address the needs of computer science and genetics researchers, ReadCube now provides search and management functions to many disciplines, as well as publishers and institutions to close the "collections gap," where libraries may not have subscriptions to all of the journals their users require. A citation plugin helps scientists write papers with full references in Microsoft Word, and a pro version provides cloud storage and sync with mobile devices.

[figshare](#): A cloud-based research management tool for scientists to manage research objects such as datasets and figures collaboratively. Portals created by researchers can be either private or public, and can be used as public repositories by institutions. The numbers of shares and downloads from public portals can also be tracked, providing users and institutions with another metric for the dissemination of their research.

[UberResearch](#): Helping funding agencies and organizations manage and evaluate their portfolios through classification and analytics. Without running enormous and slow audits, funders often struggle to keep detailed data on the grants they hand out, much less the investments made by other scientific funding organizations. UberResearch organizes funder data and allows users to compare their portfolio with other funders, tracking specific terms about research areas (such as atherosclerosis or diagnostic radiology). The company also recently launched a partnership with the ORCID ID system to link funding to researchers automatically, for even easier tracking and reporting.

## *Conclusions*

- Demonstrations of Digital Science products sample the range of research information technologies currently available to researchers, publishers, funding agencies, and administrators
- Current products focus on gathering information on researchers and research for institutional reporting and funding agency self-evaluation, alternative metrics for scientific impact, online repositories that enable collaboration and sharing, and digital publication management.

# Information Intensive Research Initiatives at the University of Chicago

Moderator: [Ian Foster](#), director, Computation Institute

[Samuel Volchenboun](#), Director and Associate Chief Research Informatics Officer for Translational Research, University of Chicago Medicine

[Alison Brizius](#), Executive Director, Center on Robust Decision Making for Climate and Energy Policy

[Michael Wilde](#), Senior Fellow, Computation Institute

[Kyle Chard](#), Senior Research Specialist, Globus

[James A. Evans](#), Director, Knowledge Lab

[Allison Heath](#), Open Science Data Cloud, Institute for Genomics and Systems Biology

Computation is now an essential approach for nearly every research field, from medicine and biology to the humanities, art, and social sciences. As data analytics and high-performance computing spread out from classically computational disciplines such as physics and astronomy, the specific technical and cultural challenges that face each area of study emerge. In turn, these new ways of using computation inform the development of new tools that address field-specific needs or are more flexible and adaptable for researchers from different backgrounds.

At the University of Chicago, several units assist researchers by providing the resources, tools, and expertise to integrate computation into their work. The [Computation Institute](#), the [Research Computing Center](#), the [Center for Research Informatics \(CRI\)](#), the [Institute for Genomics & Systems Biology \(IGSB\)](#), and other groups help new and experienced researchers capitalize upon the potential of computational methods. Representatives from many of these units appeared on a panel to discuss the current state of conducting information-intensive research with new technologies and building new technologies for information-intensive research.

All of the research efforts described on the panel were fueled by the recent explosion of digital information useful for scientific and medical investigation. The [Center for Robust Decision Making on Climate and Energy Policy \(RDCEP\)](#) combines the massive datasets generated by advanced climate models run on supercomputers with the data from models of other sectors climate change will likely impact, such as agriculture and the economy. The Knowledge Lab works with enormous pools of text, image, and sensor data from journal publications, patents, grant proposals and other sources to trace the origin and transformation of knowledge over time. The CRI and IGSB utilize the flood of genomic and electronic medical record data to pursue new discoveries in biology and medicine.

But behind these exciting new strains of data-rich research lie several technical, cultural, and legal challenges, panelists said. The CRI data warehouse contains sensitive patient data, necessitating compliance with HIPAA patient privacy standards, approval from the institutional review board, and the creation of detailed governance polices before it could be launched. **Collaborations that cross research boundaries, such as the 8 institutions and 19 disciplines included within RDCEP, can face significant disconnects between researchers in their willingness to share data.** [Knowledge Lab](#) projects that work with copyrighted information, such as articles from journal publishers, must convince content owners of the mutual benefits of utilizing their data.

Technical challenges include integration of datasets that may be in dramatically different formats, working with and sharing terabytes and petabytes of data, and publishing computation-heavy research in a reproducible manner. Panelists described several technologies they have developed to address these needs, many harnessing cloud computing resources to overcome obstacles that can slow and limit research.

RDCEP developed the [FACE-IT platform](#) to improve access to climate data and software, enabling researchers to focus on analysis instead of tedious data gathering and management. The [Globus](#) project creates cloud-based data management services that allow researchers to transfer, share, and publish large datasets, facilitating long-distance collaborations. [Swift](#), a high-level programming language, lowers the entry barriers to faster parallel computing while also organizing the inputs and outputs of scientific computing for easier and cleaner publication of methods. The [Open Science Data Cloud](#) puts enormous datasets into the cloud for easy access by researchers without the advanced computing resources to compare or analyze terabytes or petabytes of data in genomics, earth science, and other fields.

Developing these tools is only half the battle, panelists said. **While tech-savvy researchers and fields may immediately see the benefits of these tools and quickly implement them into their workflow, reaching less computationally-experienced researchers is difficult.** Michael Wilde of Swift encouraged “deep embedding” within these fields, working with research groups to understand what software tools they already use and figuring out how to enhance them, not replace them. James Evans of Knowledge Lab said that his group has organized meetings that bring together researchers in the humanities and social sciences with computer scientists and software developers, to brainstorm both ideas and computational solutions. Both Globus and Open Science Data Cloud have worked hard on the user experience of their tools, listening to their user community and improving their interface to make it as simple as possible.

Aside from this in-depth outreach, panelists recommended other cultural shifts that would promote the uptake of research information technologies. Samuel Volchenbom of the CRI said that his team doesn't merely provide an automated service to users, but works with them throughout their project, from conceiving the experiment to running the analysis to preparing the journal article reporting the results. This approach helps users consider the data warehouse as less of a “tool where you flip a switch,” he said, but more as an integral component of the research process.

The panel also echoed Victoria Stodden's keynote speech in advocating for improved reproducibility in computational science. Alison Heath of OSDC recommended the sharing of virtual machine images which can contain all of the data and software used to conduct a published analysis, so that other researchers can easily repeat the same specific methods. To encourage the sharing of data and code, Evans proposed a public “leaderboard” that can be used to track and incentivize scientists that are properly following open science standards.

## Conclusions

- Computational methods are now applied to a wide range of research topics. However, some legal and cultural challenges remain unresolved, such as different attitudes towards data sharing within multidisciplinary collaborations and the protection of patient or copyrighted data.
- University of Chicago researchers are developing technologies to address some of the core technical challenges of data-driven research, such as the sharing and publication of data, analyzing large datasets, and creating more transparent reporting of research methods.
- Spreading computational tools to less tech-savvy researchers and disciplines requires deep embedding and conversations between computer scientists and domain experts to understand their needs and explain the benefit of new technologies. User experience and customer service are also important for the sustained usage of tools.
- New attitudes about the role of these technologies can be promoted by helping researchers throughout their process, not just providing a static tool, and by publicly celebrating researchers who practice good open science behaviors of publishing clean data and software.

## Funding Perspectives

Moderator: [James Evans](#), Director, Knowledge Lab, Sr. Fellow, Computation Institute, Associate Professor, Sociology, University of Chicago

[Josh Greenberg](#), Director, Digital Information Technology Program, Alfred P. Sloan Foundation

[Stephen Fitzmier](#), Director, Planning & Evaluation, John Templeton Foundation

[Rebecca Rosen](#), Senior Researcher, American Institutes for Research

[E.J. Reedy](#), Director in Research and Policy, Ewing Marion Kauffman Foundation

Data and computation aren't just changing the way that scientists do research, they're also changing the methods of the agencies and foundations that fund that research. In the final session of the workshop, representatives from three foundations and a non-profit organization that provides research and evaluation services for funding agencies gave a "peek behind the curtain," as moderator James Evans put it, on how data-driven metrics inform their decision-making.

Organizations that fund science increasingly seek to update their traditional processes with new data-driven ways to select scientists to fund, monitor how those projects proceed, and track the impact of their funded work long past the end of the grant. According to the panelists, they're interested in questions such as where to direct their attention in finding promising science to fund, how to capture the best people, how well their funding portfolio matches their overall strategy, how their funding activities compare to other agencies and foundations, and the long-term impact of funded research and tools -- not just on science, but on business and policy as well.

For many of these questions, older metrics such as citation rate provide insufficient answers. So **funders are increasingly looking at new and alternative metrics that better suit their goals**, the panelists said. E.J. Reddy of the [Kaufmann Foundation](#) said tracking formal publications is almost irrelevant to their mission of supporting entrepreneurship and real-world impact; they need metrics that track more immediate results and the influence of funded events such as consortia and conferences. For grants that fund interdisciplinary projects, groups such as the Sloan Foundation and Templeton Foundation are interested in knowing how well these efforts spawn new collaborations and scientific advances across fields during and after the grant.

While organizations may already have methods for answering these questions, data-driven methods can make those evaluations faster, more efficient, and more accurate. The [Center for the Science of Science and Innovation Policy \(CSSIP\)](#) at the [American Institutes for Research](#) helps funders better use their data, working with them on 6-to-9-month projects that construct a data infrastructure -- "an archeological dig" that links their funded projects to other data sources, said Rebecca Rosen. Importantly, they also train people within the funding organization to expand and scale up the infrastructure after CSSIP has left.

One benefit of the CSSIP intervention is that organizations can compare their investments to other groups around the world, to see how they fit into the global landscape of research funding. **Data sharing across public and private foundations and agencies remains a sensitive subject, but a side effect of these new metrics and practices may be increased transparency and collaboration for the greater good of research**, the panelists agreed. As Stephen Fitzmier of the [Templeton Foundation](#) said, "We can agree on the positive change we are trying to create, so how can we work at this together?"

## *Conclusions*

- Funding agencies seek new ways to use data and metrics to evaluate their portfolio, find promising new researchers and projects, and assess the long-term impact of funded research beyond the lifetime of a grant.
- Foundations and agencies would like to see alternative metrics that reflect the full spectrum of how research, technologies, meetings, and other funded activities influence the scientific landscape.
- These new data-driven approaches may also enable more collaboration and interaction between foundations and agencies with common goals.
- Groups that help funders make sense of their data have found success in short projects that build a new, scalable data infrastructure combined with training to create internal expertise in using data more efficiently in their decision-making and evaluation.

## Concluding Observations

[Daniel Hook](#), CoFounder of Symplectic; Director of Research Metrics, Digital Science

To close the workshop, Daniel Hook reiterated **the need for corporations and institutions to come together in forums such as these to create research information technologies that will attract the most users and meet the needs of many different groups.** However, Hook cautioned that the conversation should not be limited to science, but should expand to include scholars of other disciplines who can equally benefit from new data and computation tools for research. While the challenges may be different from field to field, the approaches discussed at the workshop in the context of life sciences, natural sciences, and social sciences may be equally effective in other areas for addressing obstacles and realizing the potential of data-driven research. “I challenge you to think about how our message...can be inclusive of our colleagues in different areas,” Hook said.



Work smart. Discover more.

Our tools and services help at every stage of the research cycle.

