Science Engagement: Best Practices & Scalability in the 21st Century

This report investigates the relatively new role of science engagement, including terminology, ideal characteristics of science engagers, understanding the "science work cycle," measurements of success, useful modalities, and how to scale science engagement.

As more scientists, researchers and university faculty use "big data," higher capacity networks are increasingly in demand to access computation, storage and collaboration. This is especially true of the NSF-funded Pacific Research Platform's (PRP) principal investigators and staff, who are attempting to understand how to scale up from a West Coast initiative—overseen by UC San Diego and UC Berkeley — into a nation-wide effort, the National Research Platform (NRP).

Yet, most domain scientists, who find they are generating increasing amounts of data, whether carrying out research in genomics, climatology or seismology, aren't taking full advantage of the resources available to them. "Sneaker net" continues: transporting hard drives or thumb-drives between facilities.

Domain scientists know they need to adapt, but aren't typically network savvy. Sometimes they have tried to optimize their use of the network, but when they encountered obstacles, they moved on to other priorities. In any case, they are not particularly keen to take their focus away from their work to delve into a field as intimidating to the uninitiated as networking. It has been said that in the university setting, faculty often prefer to do things the way they learned when they were graduate students.

It is equally daunting from the other side of the equation: network engineering. The focus of network engineers is to keep the pipes open and running smoothly, and troubleshoot when they aren't. Trying to understand the workflows or processes of diverse scientific research projects has not been expected of them in classic work environments. "Network engineering for science and for daily use, are fundamentally different things," notes **Jeffrey Weekley,** PRP campus champion and director of cyber-infrastructure and research computing at the University of California at Merced.

This report explores aspects of science engagement to help those in universities and research institutions who are expanding their networks. Simply stated: science engagers are going to be integral participants in a successful NRP, and the lessons learned so far can be applied more broadly.

The background interviews for this report initially occurred in August of 2017 at a workshop in Bozeman, Montana entitled *First National Research Platform Workshop,* co-sponsored by the PRP, CENIC and Internet2. Additional interviews with sciences engagers around the country occurred through early 2018.

Terminology

In order for domain scientists to take full advantage of the networks available — or even to know they exist — and for network engineers to work successfully with them, a new role is emerging: the *science engager*, sometimes also known as the *research engager*. This individual, or team, not only brings people from the engineering and science domains together, but also assists in optimizing network designs for the purposes of science. Ideally, the science engager goes beyond simply exposing researchers to Using Networks 101, but illustrates how the science itself can be improved in new and unexpected ways — through collaboration, through access to larger data sets, or through creative ways of illustrating approaches that haven't yet been considered.

The term *science engagers* appears to have first been developed at the U.S. Department of Energy/National Science Foundation's Open Science Grid Consortium and its Web Science Portals, then elaborated upon by DOE's Office of Science's ESnet (Energy Sciences Network – http://www.es.net/) as part of its "requirements review workshops" with domain scientists and science programs. Originally, ESnet had used the terms *user outreach* and *science outreach*, but for the past few years or so they have been using the term *science engagement*.

In other large institutions, the same concept has been developing over time, but with different designations: *collaborators*, *research facilitators* (as at several University of California campuses), *consultants* (at the University Corporation for Atmospheric Research), or *campus champions* (in NSF's XSEDE program).

What to call this new role is one focus of a relatively new NSF-funded national organization: the Campus Research Computing Consortium (CaRCC.org/), which aims to professionalize the cyber-infrastructure (CI) workforce. According to **Patrick Schmitz**, associate director of Research IT focusing on architecture and strategy at UC Berkeley, some terms being used by the CaRCC include *research engagers*, *education facilitators* or *facilitators.*

Schmitz also notes that at a large organization such as UC Berkeley, his mandate includes supporting social science, journalism, arts and humanity faculty and researchers, as well as upper-level university management, so the title of *"science" engager* can be considered somewhat limiting, in that regard.

If science (or research) engagement is part of a role, the title might be folded into a larger title, such as *cyberinfrastructure engineer* (CIE), as it is for **Shawfeng Dong** at the University of California at Santa Cruz. In other instances, science engagement is part of the mandate at the highest levels, such as with the University of Michigan's **Eric Boyd**, whose title is director of advanced computing technology services, yet includes science engagement.

So, whether called *science engager* by name as a singular role, or some other title as part of a larger research IT oversight mission, it's a new and evolving function to bridge the chasm that all too often exists between scientists who want to focus on their science, and network engineers who want to help them. It's also a new career path for computer science majors who love science, and others in data-intensive research with high-level people-to-people communication skills.

The science engagement role is an emerging part of an ecosystem of other computer professionals and network engineers, as well as tools, services and resources, which work together to support scientists as they conduct their research. Other pieces of the puzzle may include, for example, computational researchers, as well as such tools as workflow software; algorithms for modeling or doing analytics on the data; ID management procedures for operators and instruments; privacy and access-level approaches; and cyber-security considerations across multiple domains. It's also important to realize that the role of science engagement continues to evolve, and development of the National Research Platform can play a role in its evolution.

What does science engagement look like?

According to **Eric Boyd**, director of research networks in advanced research computing technology services at the University of Michigan, the most common need of the faculty or researchers who contact his group (after network upgrades) is to interact with the network "exactly the way we used it before." Typically, users don't know how to articulate this, which is where a science engager comes in. University users must also understand that although the network engineers might try to solve their specific problems, the campus-wide network needs to adhere to a consistent protocol rather than creating individual solutions for each department. The intermediary science engager is often a more effective messenger than the network engineer.

Thomas Cheatham, director of research computing at the University of Utah and director of the Center for High-Performance Computing, notes that researchers contact his group when they "get into trouble," or when data transfer is too slow over the networks. Cheatham, who in addition to his other role is a professor of medicinal chemistry, also focuses on educating his users about Science DMZ (http://fasterdata.es.net/science-dmz/), and about the privacy issues surrounding restricted information, such as data governed by the Health Insurance Portability and Accountability Act (HIPAA).

"Even more important than the actual technology are the *facilitators* who can enable researchers to effectively utilize advanced cyber-infrastructure," Cheatham says. "As books and librarians were critical to support research in the past, now the essential elements are research computing and cyber-infrastructure: virtual and physical computing, storage, data management and networking resources to support collaborative research. Research computing is the new university library."

Without science engagers, networking engineers don't always know if they're solving the right problems, and therefore, being successful. **Steve Huter**, director of the Network Startup Resource Center at the University of Oregon, who works extensively overseas, found that international medical research teams from the U.S. National Institutes of Health (NIH) and European universities were able to work more efficiently at Makerere University in Uganda "because the science engager was able to clarify the network scaling needs during a major upgrade."

Another example of how science engagers can assist is provided by **Lauren Rotman** of ESnet, whose team helped a Michigan-based hospital researcher work with a beamline scientist at Advanced Light Source, a Dept. of Energy National Laboratory managed by UC Berkeley. The researcher was trying to use a particular type of x-ray to study pediatric brittle bone disease. As with supercomputers, machines at this synchrotron facility are available to scientists for set periods of time. Typically, the data would then be shipped off to a supercomputer back at the researcher's lab or designated facility. But, what if there had been some error in the instrument calibration, or the angle at which the bone had been placed in the x-ray hatch? In this instance, the science engager was able to set up a real-time network connection to a high-performance computing environment to provide a pre-visualization of the x-ray, helping the researcher to maximize success in his short window of time on the instrument.

In another case, **Jeff Weekley,** of UC Merced's Office of Information Technology, was able to assist a cell biologist beyond mere network efficiency to markedly improve the scientific results. The scientist at his facility has an $800,000 microscope used to look at live tissue samples. The task was to try to find tissue with a particular genetic marker that had been dyed red. It was time sensitive; if it took too long, cells would die. There was no network drop in his lab, which he shared with mice that could escape through openings. The microscope had been set up so that the data was downloaded to a hard drive, which would be hand carried to a workstation in another room, where the scientist would sit and wait for the images to load. Then, another hard drive would be swapped in with more images. The microscope was continually generating images, but the scientist couldn't look at them in real time, thus potentially missing something as disks were being swapped in and out. Wearing his research consultant hat, Weekley explained to the scientist the boundary between streaming workflows and file-based workflows. The solution was to get the microscope on the network, and have it write to a high-availability drive. Then, they applied a type of GPU-based computer vision workflow to the task, so that images without the red pixels would be tossed out. "The biologist wasn't even aware that this was a possibility," Weekley explains.

Weekley emphasizes that working with domain scientists is best done in incremental stages. "We don't take their tools away and say "ours our better." We try to inject a little technology into what they're already doing to make it more efficient, scalable and productive — always keeping in mind what their goals are," he notes. So, in the example of the cell biologist, above, he started with a small data set and replicated the process on a network. When that worked, he added machine vision to see if they could approximate the data coming off the microscope the same way it had before adding the network. Then, they worked with a more complex data set. By the time he implemented the full strategy, the biologist was comfortable with the process — and any unexpected results were readily accepted as not due to a flaw in the design.

Who make the best science engagers?

Most of those interviewed for this report agree that scientists are more ideally suited to science engagement than IT personnel. As **William E. Johnston**, former director of ESnet at Lawrence Berkeley National Laboratory, puts it, "They have to know the issues involved in computing and networking, but also have an abiding interest in science itself."

In addition to a science background, key traits for engagers include:

- a baseline facility with technology (computation, data, storage, software) without necessarily having to be engineers,
- an understanding of how scientific research is conducted,
- willingness to solve problems as part of a group (not only on one's own),
- excellent communications skills (both written and verbal),
- superb "active" listening skills,
- willingness to document what has been heard (as needed),
- being a quick study at understanding problems in which one is not an expert,
- ability to recognize (and leverage/scale) patterns,
- willingness to be humble & open-minded ("I want to help you and learn what you need, I'm not the big expert who knows everything"),
- customer service-oriented,
- having vision and leadership skills,
- being outgoing and sociable,
- curiosity and inquisitiveness,
- creativity ("isn't afraid of dreaming a bit & imagining a different way"),
- ability to learn — and to enjoy it,
- adaptability,
- patience, and
- not doing science engagement for the money!

Ideally, science engagers are either actively doing their own research, or are used to supporting other researchers. According to High-Performance Network Manager **Marla Meehl** of the University Corporation for Atmospheric Research (UCAR), science consultants have to have enough science background and credentials that they will be respected and listened to. “Scientists like to talk to other scientists,” she says. The facilitator has “to appear trustworthy and genuinely interested. It doesn’t work if you’re just doing lip service and checking a box.”

Sometimes science engagers are asked relatively straightforward questions, for example, “How can we store this data we’re creating?” This can be answered directly, but UC Berkeley’s **Chris Hoffman** notes that it’s better to be able to look broadly even at seemingly basic problems, asking such questions as:

- What is the entire workflow?
- What landscape of technologies are out there?
- What are the real issue and what are the opportunities?

At the same time, one has to be thinking about how much time users have for science engagement, and whether other members of the CIE or IT team with specific skill sets need to be called in. This also means having the ability to create relationships with other service providers and, acting as the bridge between them and the user. “It requires research engagers with perspective: the analytic skills to be able to step back and see the big picture,” says Hoffman, whose area of focus at UC Berkeley is research data management (RDM) and informatics. “Basically, you’re looking for holistic thinkers with analytical skills.”

Sometimes these skills are not available in a single person, so Hoffman and his colleagues try to pair science engagers into a team — one, with great people skills and another, with more technical expertise. “It’s a unique person who combines all the traits you’re looking for,” he says.

According to **Shawfeng Dong** at UC Santa Cruz, “there is a shortage of talent that can understand the scientific field *and* provide CI help. Some individuals are ‘people’ people; some are ‘technology’ people. The right kinds of people for science facilitation are right in middle. They like both technology/science, and also enjoy interacting with people.” Dong himself fits that profile: he is an astrophysicist and astronomer, and also holds the title of cyberinfrastructure engineer in UCSC’s Information Technology Division for Research and Faculty Partnerships. Moreover, he wears a third hat as an adjunct faculty member in the applied mathematics department, where he teaches a course on high-performance computing. All these roles make him ideally suited for science engagement.

Another such multidisciplinary science engager is **Heidi Morgan,** director emeritus of the Center for Internet Augmented Research & Assessment (CIARA) at Florida International University and currently senior computer scientist at the Information Sciences Institute at the University of Southern California. CIARA’s science engagement mission is noted on its Web page: “… creating a new generation of scientists and engineers who are capable of fully integrating IT into the whole educational, professional, and creative process of diverse disciplines.” Morgan exemplifies this synthesis: she holds a masters’ degree in English, and a PhD in information communication technology from the Erasmus Research Institute of Management (ERIM) in the Netherlands. “I use my research skills to engage science communities, and to understand their interests,” she says. “To engage that community, common ground must be established. There must be trust — not in the sense of cyber-security, but rather social network trust.”

Because good science engagers are hard to find, **Camille Crittenden**, deputy director of the Center for Information Technology Research in the Interest of Society (CITRIS) and co-PI of PRP, is actively working with student teams at UC Berkeley to train future engagement personnel. "There are so many talented, ambitious students who can learn about problem solving and network engineering, and we can help them in terms of workforce development," she says. CITRIS, which is one of the University of California's multi-campus institutes for science and innovation and also a PRP/NRP partner, draws on the school's Undergraduate Research Apprentice Program (URAP) to recruit multi-disciplinary, diverse student teams to create methods and tools for tracking local network performance as it relates to the Pacific Research Platform. Students from archaeology and anthropology are also being recruited to work with those creating visualization platforms to preserve and share valuable cultural heritage resources that are at risk of deterioration, whether from erosion or armed conflict.

Patrick Schmitz, at UC Berkeley, mentions that the Research IT group he heads runs internships not only with graduate students, but undergraduates, even in the early years of their college careers, who are paired with consultants. "We're creating undergraduate fellowships for freshman and sophomores who will be ready to walk into labs when they are juniors and seniors," says Schmitz. "They get engagement exposure as part of a team." Not only computer science or electrical engineering students are being recruited, but also those in computationally intensive science domains. Ideally, they look for students with double majors in science and some form of communication, such as English or journalism.

Because the role science of research engagers is so new and remains undefined — or defined differently — the Campus Research Computing Consortium is investigating how to create an H.R.-type classification. This will provide a career path and clear skill metrics against which budding engagers can be professionally measured.

Understanding the "science work cycle"

William E. Johnston, formerly of ESnet, was part of a team that developed protocols for the science engagement role. First and foremost, he says, science engagement has to be approached in terms of what process of science is being conducted, and where it's headed. Network engineers don't need to be engaged in the actual science itself, but will want to be concerned about how the data is handled.

To know how to build out the network for that goal, ESnet developed what Johnston calls the "characteristics of the communications infrastructure." The first questions to ask are:

- What is the process of the science involved?
- How is research conducted on a day-to-day basis?
- What is the extent of the community: just principal scientists and their graduate students, or a larger user group?

There are different models for how science is conducted, which is important for science engagers to understand so they can figure out which areas are impeded by data communication, says Johnston.

Science communities have differing needs and data structures. The Large Hadron Collider (LHC) project entails one enormous data source fanning out to many thousands of users. The genomic community deals with amounts of data comparable in size to the LHC, but the data is generated in a distributed fashion and, therefore, is handled differently. NIH's GenBank, for example, receives direct submissions from various stand-alone laboratories, as well as submissions from large-scale sequencing centers. Berkeley Lab's Advanced Light Source serves 2,000 researchers. All of these data dissemination patterns result in different network problems, and thus different approaches to engagement.

After knowing the model of science being done, ESnet determined five basic questions to help the science engager, particularly if a network build-out is envisioned:

1. How are the data to be generated?
2. Where will the data be stored?
3. Where is its user community?
4. Where will the computing on that data occur?
5. What changes are expected in the science process timeline, or use of facilities or instruments — in the next two years; the next two-to-five years, or beyond five years?

A variant to these questions is relevant in situations where a scientific instrument is generating the data:

1. What kinds of instruments are being built?
2. Where are they to be located? (Not usually near the scientist.)
3. Is the instrument to be interacted with while in operation? Are these interactions time-constrained or open-ended? Are the interactions done with a human in the loop, and if so, what are those time considerations — data to human, human to data?
4. Where will the computing occur? Supercomputer center? University IT center?
5. Will data analysis be local or will there be a subscription involved to an external computing facility, e.g., a commercial cloud?
6. Where are the data generated to be stored? (Not usually nearby.)
7. Where are the user populations located? How will the greater scientific community use the data?
8. Are subsidiary communities going to access the data as a repository resource, e.g., unfunded, as a service? In the case of DOE, for example, the computers and network systems of external users might have different characteristics than the DOE-funded community. Are third-party contractors involved, and if so, what are their licensing requirements?

Measuring science engagement success

Meaningful engagement varies from case to case, but one approach to keeping track of encounters is by using customer relationship management tools. These can range from simple Excel spreadsheets, which might only record who's calling and their requests, to slightly more expensive turnkey, out-of-the box solutions. More sophisticated tools from such companies as Salesforce can run analytics or identify how many open “trouble tickets” remain unresolved. **Lauren Rotman’s** group at ESnet uses a ticketing system called ServiceNow to track progress in supporting and consulting with science collaborations. The team also seeks to understand which science domains are trending in inquiries. “In a way, these statistics end up showing you which communities are in pain due to the growth in data they’re seeing,” she says. And that helps with planning science engagements with scientific communities.

Associate Directors **Patrick Schmitz**, **Chris Hoffman** and their colleagues have identified three kinds of metrics to measure success within UC Berkeley’s Research IT groups:

1. Performance metrics:
    a. inward looking,
    b. how effectively using computing and network resources,
    c. how long taking to respond to requests,
    d. are we closing tickets (used by managers to calculate efficiency),
    e. are we helping a broad range of people or departments?

2. Metrics of activities, translated into a meaningful story: important to stakeholders such as the executive tier or academic leadership to understand what services are being offered and how they can be funded.

    For example:
    a. What is the number of users and how many cycles have they consumed in the research computing environment?
    b. What have the number of consultations been, within the university and across the network of partnerships?

        That data might then be analyzed further to determine how many consultations were done for a particular college department, or if outside the university, for researchers in the various National Science Foundation domains?

3. Impact metrics: correlation of the users of services to key academic metrics, such as publications generated, grants applied for and won, classes supported, or recruitment and retention of faculty and Ph.D. students. To obtain this information, Schmitz’s group runs an annual survey asking about these metrics, as well as general questions about the use of, and satisfaction with, Research IT services. (They also separately track references to publications, and ask researchers to acknowledge the department in their papers.)

    Impact metrics can be used in ways not immediately obvious. For example, they can help senior management recruit high-level researchers, not only by demonstrating the kinds of research IT services available, but also by specifically citing past successes obtained through the collaboration of scientists and IT staff. Statements

> made by research facilitators such as, "We could work with you to get a grant to upgrade your lab," is certain to focus the mind! In this sense, research engagement not only enhances science, but has a marketing, and ultimately, funding function, as well. This is especially important in these times of budget cutbacks.

Schmitz also unofficially keeps track of what he calls the "squishiest of metrics" — "golden quotes" — both through the survey and during engagement follow up. Examples of golden quotes are: "Without you, I couldn't have done my research." Or, "Other universities weren't using this software, and you got it running. The fact that you got it working made a difference." He hires interns to interview users and create a profile: How did you use our services?

Science engagement modalities

One of the common complaints voiced by heads of university IT departments or research laboratories is that certain populations they serve remain unaware of all the services available. It is for this reason that UC Berkeley's Schmitz puts so much emphasis on communication and outreach. Among his strategies:

• Target the IT staff within departments: "Can I get five minutes in front of your faculty meeting or new grad student orientation?"

• Generate a lot of writing: In order to develop staff communication skills, each staff member has a performance goal to write six articles per year for the UC IT blog. These can be new technologies or Research Profiles, stories about research successes and how central IT has contributed to them. Stories are published online as blog entries or in a newsfeed, as well as in print newsletters. "We find that if a faculty member asks what we do, we can point them to the Research Profiles to show them what their peers are doing and saying."

<u>Workshops & Seminars</u>

Workshops are much more useful than one-on-one communication, and are probably the best way to scale the science engagement process. Even larger institutions such as DOE's ESnet, with its four dedicated science engagers, can't meet the needs of all those seeking help. According to **Lauren Rotman**, ESnet science engagement group lead in Washington, D.C., they run workshops to build "an army of engineers and outreach practitioners" who can return to their university campuses with the knowledge of how, for example, to build, Science DMZs (secure, high-performance computer sub-networks) or deploy perfSONAR (a tool to monitor network performance). Other workshops will be for domain scientists only, who explain their research: the instruments, sensors or tools they're using and where they hope to take their research in the future.

In either instance, her group follows up with surveys to see if there are outstanding questions or areas of confusion.

According to **Marla Meehl** of the Computational and Information Systems Laboratory at the University Corporation for Atmospheric Research (UCAR), they do targeted trainings when they want to instill best security practices, or when they get a new supercomputer. These trainings are supplemented by a variety of modalities such as Webinars, blogs and newsletters. Since UCAR's constituency is international, they record presentations to be viewed online.

Meehl also finds the annual reports published by ESnet on various scientific topics, such as energy or climate sciences, to be useful in science engagement. Large group meetings for particular science domains are offered at UCAR in which they discuss the future research:

- How will data requirements change?
- Will research teams be local or international?
- What are the game changers that will put stress on the cyber-infrastructure two-, three- or five-years out?

She finds the most effective method is to have an early network-adopter domain scientist as part of a seminar to serve as the science engager and encourage his or her fellow scientists to increase their use of networks. "If there's a peer-to-peer exchange with another scientist, even one not in their field, the seminar attendees are more likely to listen to the benefits of using networks for collaboration, than if the presentation is made by an IT person."

"The easier and more transparent you can make it, the more likely they are to adopt it," she says. "It's more painless that way."

Steve Huter, director of the University of Oregon's Network Startup Resource Center, states that they hold science engagement workshops piggy-backed onto multi-day events focused on particular scientific topics. For example, the National Center for Atmospheric Research (NCAR) held a meeting in Boulder, Colorado on weather and climate in Africa. "At the tail-end," he notes, "we brought in our network people and took advantage of the fact that scientists were already together working there. At that point, the domain scientists were ripe for listening. That's why we seek out partners like NCAR."

UC Berkeley's Research IT team runs a biweekly "Reading Group," which explores such topics as data management, workflows or specific tools. Background material is distributed in advance to a wide mailing list. At the event, which typically draws 30 to 40 people, a 20-minute presentation is followed by a discussion. The demographics of the attendee population varies greatly from event to event. Academics or researchers offer suggestions to their peers about how to solve a problem that they themselves have just learned about, which is considered the most ideal form of communication.

At UC Merced, **Jeff Weekley** runs a 1½-hour high-performance computing walk-in clinic on Fridays for graduate students or faculty who want to use the cyber-infrastructure on campus. Here they can get expert help from the staff or mentoring from peers. "The need for the three pillars of HPC — computation, storage, networking, and visualization — is almost unbounded. More and more disciplines are converging on HPC that require advanced CI, even disciplines that never before relied on computational sciences or big data."

Chris Hoffman of UC Berkeley recommends holding workshops "where researchers are already used to going," rather than in IT conference rooms or the campus library. For example, his science engagement teams will often give presentations on "publishing your data" or "copyright issues for data" at the Academic Innovation Studio on campus, where faculty are already comfortable going for educational services, or the Berkeley Institute for Data Science, a hub for nurturing data-intensive science.

<u>Creating online resources</u>

• Wiki: UCSD's **Shawfeng Dong** notes that not everyone has the budget to sponsor workshops, and one-on-one interactions are not always scalable. So Dong has found hosting an informational Wiki is a good method for outreach as a science facilitator. He provides tutorials, "how-to's," and answers common questions at: https://pleiades.ucsc.edu/hyades/

• Blog: **Thomas Cheatham** of the University of Utah's University IT department, is the co-PI of the NSF-funded ACI-REF (Advanced CyberInfrastructure-Research and Education Facilitators). Its mission is to work with the "academic missing middle" of ACI users, "those scholars and faculty members who traditionally have not benefited from the power of massively scaled cluster computing but who recognize that their research requires access to more compute power than can be provided by their desktop machines," according to the ACI-REF Website. He maintains a "best practices" blog at: https://aciref.org/category/experience/

• Web-based resources & Webinars: **Eric Boyd** at the University of Michigan also uses the Web for scalable communication. His group puts up network diagrams for his faculty researchers, but instead of the diagrams network engineers typically see, these show only the kind of information relevant to the researchers, such as the speed to different commercial cloud resources. Boyd also publishes case studies for best practices online, extrapolating from frequent problems that have come up in one-on-one sessions. "Every time we find a solution, we generate slides to use at workshops, chalk talks, Web videos. That way, we can scale our resources to help more people."

ESnet and the University of Oregon's **Steve Huter,** among others, have frequently recorded technical education forums that are archived online for both scientists and network personnel to understand the basic concepts.

From PRP to NRP

Patrick Schmitz of UC Berkeley's Research IT group has created an informal far-flung digital humanities research community of institutions around the country, including Indiana University, Rutgers, Clemson and other UCs such as UCLA, among others. He calls it a "network of partners," who leverage one another's knowledge and solutions through phone consultations or email. In this manner, research engagers in one institution aid scientists or faculty elsewhere. For example, there might not be enough staff focusing on research data management solutions for the humanities at any single institution.

Another example is how to deal with what's called "human subject data"—common in social science research—in terms of privacy regulations around clinical data, when other researchers want to share it. The answer might be in how to design access, sharing protocols or workflows that maintain privacy obligations. Or, it might be to show how the sensitive data can be secured on the computer, even if it's on the network (when the licensor has dictated "no connectivity").

As **Thomas Cheatham** of the University of Utah notes, "NRP needs to address comprehensive identity and access management (IAM). If we are going to share data, especially restricted data, we need to know that there are appropriate security and controls on both sides, potentially business associate agreements (or trust relationships), and ultimately, federated ID. What this means is that rather than having separate logins or accounts at each place, ideally collaborators can trust our IDs, and we theirs. In other words, a login that reveals exactly who you are, what permissions/restrictions on data access are appropriate, etc."

Jeff Weekley at UC Merced acknowledges that scaling PRP to the NRP remains a challenge. And he suggests that upcoming conferences create a multi-disciplinary, multi-organizational, multi-faceted project focusing on scaling science engagement. "It would help if we tried solving the problem by giving some context to it," he says. "We know how to do this: We've done it before. Everyone has 4K on their TVs in the living room: We changed the marketplace. And, we could do the same for research networks."

Acknowledgements

The PRP is supported by the NSF CC*DNI DIBBS program under Cooperative Agreement ACI-1541349; additional NSF workshop support was provided under award #1540112. Seed funding and ongoing support have been provided by the University of California Office of the President, the Corporation for Education Network Initiatives in California (CENIC), and the Qualcomm Institute (QI), the UCSD Division of the California Institute for Telecommunications and Information Technology (Calit2), and the Center for Information Technology Research in the Interest of Society (CITRIS) and the Banatao Institute.

The author thanks the following people for their time and insights offered during interviews:

Eric Boyd, University of Michigan
Thomas Cheatham, University of Utah
Camille Crittenden, University of California, Berkeley (CITRIS)
Tom DeFanti, University of California, San Diego
Shawfeng Dong, University of California, Santa Cruz
Chris Hoffman, University of California, Berkeley
Steve Huter, University of Oregon
William H. Johnston, ESnet, Lawrence Berkeley National Laboratory (retired)
Marla Meehl, University Corporation for Atmospheric Research (UCAR)
Heidi Morgan, University of Southern California
Lauren Rotman, ESnet
Patrick Schmitz, University of California, Berkeley
Larry Smarr, University of California, San Diego
Jeffrey Weekley, University of California, Merced