SPECIAL SECTION

Research Data Access & Preservation

17] RDAP13 Summit: Introduction
by Susan Wells Parham, Elizabeth Rolando and Jennifer Doty,
Special Section Guest Editors

19] Partnerships Between Institutional Repositories, Domain Repositories and Publishers
by Gail Steinhart

23] The Relevance of Research Data Sharing and Reuse Studies
by Nicholas M. Weber

27] Tracking Citations and Altmetrics for Research Data: Challenges and Opportunities
by Stacy Konkiel

33] The Research Data Alliance: Implementing the Technology, Practice and Connections
of a Data Infrastructure
by Mark A. Parsons and Francine Berman

37] The DCC’s Institutional Engagements: Raising Research Data Management Capacity
in UK Higher Education
by Martin Donnelly

41] RDAP13 Poster Session Summary
by Jennifer Doty

FEATURE

44] Reflections on Our Future
by Douglas C. Engelbart
Research data and preservation (RDAP) is an area presently receiving intense focus from information professionals. This issue features extensive coverage of the ASIS&T RDAP Summit 2013 assembled very ably by guest editors, Susan Wells Parham, Elizabeth Rolando and Jennifer Doty. Panel organizers, speakers and attendees discuss the Summit and many aspects of the RDAP challenge. Areas covered include repositories, curation, partnerships, infrastructure, data use and reuse and data citation and metrics.

From an ASIS&T sponsored event, we next turn to ASIS&T as an organization, including

- ASIS&T President Andrew Dillon’s reflections on what ASIS&T and similar organizations must do to stay relevant and the late Douglas Engelbart’s thoughts on our future.
- Specific examples of ASIS&T’s current evolution as Mei-Mei Wu and Diane Sonnenwald report on the results of an International Relations Committee survey on the feasibility of holding Annual Meetings outside North America and Emil Levine covers progress in the formation of an Asian Pacific Chapter.

We are, indeed, reminded of the Association’s past and future in remembering Engelbart. In an obituary Gary Marchionini recalls Engelbart, the inventor of the mouse, as the “father of interactive computing” and the inspiration of a generation of information scientists who work in human-computer interaction and related fields to enhance human capabilities. Engelbart was honored by ASIS&T at our 1996 Annual Meeting at which he participated in a panel on the future of the Association (then the American Society for Information Science). Fortunately his remarks were reported in the Bulletin at the time, and we reprint them in this issue, where they still resonate as ASIS&T looks for new paths forward.
As I write this column we are preparing for our annual ASIS&T Board of Directors retreat where the members gather for almost two days of discussion and planning. It might seem old-fashioned in the age of distributed working and collaborative meeting tools that we come together as a group in person, but in my view the face-to-face gathering is essential for our work. While basic examination and commentary on routine proposals and committee reports can be productively handled remotely, the retreat is an opportunity for the leadership to talk openly about the issues facing us as an association. Through co-location in a hotel or meeting space (this time generously donated by the School of Information Studies at McGill), members can focus without distraction on the important matters and gain a better sense of the collective view of issues than can be easily enabled through conference calls.

This year, in addition to the usual standing reports, we are allowing open space each day for discussion of the most pressing matters facing ASIS&T. Without revealing the precise agenda, I view all the discussion topics as fundamentally addressing the mission of ASIS&T in the coming years. Whether it is the experience of membership or the organizational structure of the society, all our conversations will explore where ASIS&T is heading.

EDITOR’S SUMMARY
As the ASIS&T Board of Directors prepares for the annual planning retreat, critical issues face the organization. Despite the centrality of information in the digital age and the focus, mission and international scope of ASIS&T, the decline in membership numbers since 2000 raises concerns about the association’s future and economic sustainability. To survive, professional associations face demands to streamline operations and justify services. The structure and processes of ASIS&T and its effectiveness at serving the professional community must be open to scrutiny. All members are encouraged to offer their thoughts on what works, what drives involvement and how to add value to reinforce ASIS&T’s position as the primary association for information professionals.

KEYWORDS
professional associations
strategic planning
If you read this column regularly (and that is a big “if” in my mind, given how few regular columns, presidential or otherwise, I read myself) you will know that ASIS&T is not a growing society, at least in terms of membership. In fact, the long-term trend (since 2000) shows a slow, steady decline in membership that raises important concerns about our long-term future. While it might be fun to imagine a smaller ASIS&T where each member knows the others by first name, such a model is unsustainable economically. We’re not there yet but the data are unarguable, and we need to address the matter now, not in a few years time.

That we face such issues when everything else information-related is the center of attention is, at first blush, somewhat odd. I do not see a better-positioned scholarly and professional society out there than ASIS&T, and the recent name change to reflect growing international interest surely strengthens our claim to hold the pole-position. Yet one must ask, given the data, is there still a need in today’s world for professional associations such as ours? In Race for Relevance, management consultants Coerver and Byers (2011) argue that the days of traditional professional associations are numbered, and they propose a radical makeover for survival. Their approach requires greatly streamlined boards, significantly empowered CEOs, robust technical infrastructures and rationalized services aimed directly at a clearly targeted demographic. Like so much management work I read, there is more than the occasional superficial analysis behind some claims, but the major points seem well taken, at least from my vantage point.

I have my own views on our structure and the rather complicated processes we appear to follow for even the most routine of actions, and I will be keen to hear from my fellow board members what they believe we should do. I also wonder, in a push to be taken seriously as a scholarly association, whether we might have overlooked the needs of the professional community. Surely we can serve both. I am really keen to hear directly from members themselves about how they see value in ASIS&T and what keeps them involved or lessens their interest in membership. The term of a president is short, and drastic measures are not required, but this is an ongoing set of discussions that will be perhaps uncomfortable for some but are necessary for our future. If you want to have a say, don’t wait for a survey or the business meeting at the Annual Meeting. I’ve argued, and will continue to argue, that we need a better means of information sharing, be it a forum or other, for ongoing constructive dialog among members. Until that time, I encourage you to make your views known to me or any other member of the board directly. Over the years, ASIS&T has become my primary professional association, and I am committed to ensuring its longevity. I suspect you feel the same. The challenge then is for us to find the best means of adding value to our ASIS&T experience, and I suspect if we can achieve this goal, the future will take care of itself.
Exciting Plans Continue for the 2013 ASIS&T Annual Meeting

Plans are firming up nicely for Beyond the Cloud: Rethinking Information Boundaries, the 2013 ASIS&T Annual Meeting taking place in Montreal, November 1-6. As this issue of the Bulletin is readied for publication, conference chairs France Bouthillier, McGill University, and Boryung Ju, Louisiana State University, report that they and their committee chairs have made preliminary selections for all program slots. They previously announced that Jorge Garcia, a senior business intelligence and data management analyst for Technical Evaluation Centers in Montreal, will deliver the keynote address at the meeting.

Pre-conference Workshops and Seminars

To accompany the technical program, an interesting array of pre-conference sessions has been selected for the meeting. Full details will be available in the preliminary program to be mailed shortly and on the ASIS&T website. But make plans now to participate in one or more of these sessions:

Friday, November 1

- PIM 2013: Breaking Out to More Practical Progress in Personal Information Management
- Metrics 2013: Workshop on Informetric and Scientometric Research (SIG/MET)

Saturday, November 2

- 9th Annual Social Informatics Research Symposium (SIG/SI)
- SIG/CR Workshop: Big Data, Linked Data: Classification Research at the Junction
- Eye-tracking for Information Science Researchers and Practitioners

Time to Vote for ASIS&T Officers and Directors

As it is every summer, it’s election season once again for the Association for Information Science and Technology. Electronic balloting is now underway for the election of a president-elect, treasurer and two directors-at-large. In addition, members are invited to cast their ballot for or against a bylaws change proposed by the Board of Directors. Voting ends on August 31, 2013. If you have not received an email containing your special user name and password for this election, contact ASIS&T headquarters immediately for instructions.

Newly elected officers and directors will take their seats on the Board of Directors at the conclusion of the ASIS&T Annual Meeting in Montreal.

Candidates for president-elect for 2014 are Sandra Hirsh and Prudence Dalrymple. Treasurer candidates are Vicki L. Gregory and Lisl Zach. Four candidates vying for two open director seats are Lauren Harrison, Isto Huvila, Mei-Mei Wu and June Abbas.

Candidates for President-elect

SANDRA HIRSH

Sandra Hirsh is professor and director of the School of Library and Information Science at San Jose State University. Prior to joining SJSU in 2010, Hirsh held a
number of corporate jobs, including senior user experience manager at Microsoft and director of the information research program at HP Labs. Before the corporate stints, Hirsh was on the faculty at the University of Arizona. Her research interests focus on information-seeking behavior and understanding the information needs of a broad spectrum of users. Hirsh has served ASIS&T in a number of capacities, including as chair of the current Information Professionals Task Force and on conference organizing and program planning committees. Hirsh was serving as faculty advisor to the SJSU student chapter of ASIS&T when it was named Student Chapter-of-the-Year in 2012.

**PRUDENCE DALRYMPLE**

Prudence (Pru) Dalrymple is research and teaching professor at Drexel University and director of the Institute for Healthcare Informatics at the iSchool at Drexel University, where she works closely with the Colleges of Medicine and Nursing and Health Professions and the Schools of Public Health and Biomedical Engineering. After service as a medical librarian, she joined the faculty at the University of Illinois Urbana-Champaign and was dean of the Graduate School of Library and Information Science at Dominican University. Within ASIS&T, she has twice served on the JASIST Best Paper Award jury, chaired the Education Committee and participated in the leadership of various SIGs and task forces. She served as director-at-large from 2009-12. She is also the longest serving member of the Information Professionals Task Force. She recently edited a special issue of the Bulletin on health informatics and has presented at numerous ASIS&T meetings.

**Treasurer**

**VICKI L. GREGORY** is a professor at the School of Information, University of South Florida. Since joining ASIS&T in 1984, she has served as president of the Florida Chapter and, on the national level, has been a member of SIG/LAN, which she also chaired, and SIG/DL. She has held the elected positions of both deputy director and director of SIG Cabinet. For the last six years she has been ASIS&T treasurer; previously she was a member of the Budget and Finance Committee, pursuant to which service she developed an understanding of the strengths and weaknesses of the Society’s finances and as well as the ASIS&T budgeting process. During the last several years she has worked with ASIS&T management to developing a prudent investment program for the organization.

**PRUDENCE DALRYMPLE** is senior scientist in pharmaceutical research and development informatics at Hoffmann-La Roche Inc., where she has been employed since 1979. Lauren has served in many global information scientist positions, including manager of library and information science and most recently as part of Hoffmann-La Roche’s newly created Translational Clinical Research Center. Lauren is also passionate about creating a new...
generation of information scientists so she serves as an adjunct professor at the Palmer School of Library & Information. Lauren has been active in ASIS&T since 1991. On the chapter level, she served as program chair, chapter chair and immediate past chair of the NJ Chapter. She also served as chair of SIG/MED for three terms. On the national level Lauren has served on the Leadership Committee and the ASIS&T Lecture Series Award Jury.

**ISTO HUVILA**

Isto Huvila is senior lecturer at the School of Business and Economics, Åbo Akademi University in Finland. He also works as an associate professor and project director at Uppsala University in Sweden. Isto’s research interests include information and knowledge management, information work, knowledge organization, documentation and social and participatory information practices. The context of his research ranges from archaeology and cultural heritage, archives, libraries and museums to social media, virtual worlds and corporate and public organizations. Isto’s past chair of the ASIS&T European Chapter. He has also served on the planning committee of the 2013 SIG/USE research symposium, as a chair of the second ASIS&T European Workshop, technical program chair of the first ASIS&T European Workshop and as the secretary of the ASIS&T European Student Chapter.

**MEI-MEI WU**

Mei-Mei Wu is professor and founding chair of the Graduate Institute of Library & Information Studies at the National Taiwan Normal University (NTNU) in Taipei, Taiwan. She has served as chair of the ASIS&T International Relations Committee and chair of the ASIS&T Taipei Chapter. She has organized numerous international conferences and chaired multiple sessions/tracks in conferences. She is the co-founder of Chinese Information Literacy Association (CILA), founded in 2003. She also serves as the editor-in-chief of the *Journal of Library and Information Science*, an open access journal published by NTNU. She has led nationwide information literacy teaching and research projects and won online teaching awards. Her research publications appear in both Chinese and English, in monographs, book chapters, journal articles and conference papers.

**JUNE ABBAS**

June Abbas is professor in the School of Library and Information Studies (SLIS) at the University of Oklahoma, Norman campus. She taught in the Department of Library and Information Studies at the State University of New York in Buffalo before joining moving to Oklahoma. She also held professional positions in public and special libraries. Her research focuses on the development of user-centered digital libraries, institutional repositories and other knowledge organization structures. She conducts research on youth and their use of technology, and the intersection between information behavior, information retrieval, and structures for organizing knowledge. She has also served as project manager on eight digital library projects and on task forces to develop institutional repositories. Her recently published book *Structures for Organizing Knowledge: Exploring Taxonomies, Ontologies and Other Schema* was nominated for ASIS&T Book of the Year in 2011.

Additional candidate information, including expanded biographical data and position statements, is available at the ASIS&T website at [http://asis.org/elections/](http://asis.org/elections/).

**Bylaws Amendment**

A proposed amendment to the ASIS&T Bylaws is also included on the ballot. Full details about the proposal are included in the June/July 2013 issue of the *Bulletin of the Association of Information Science and Technology* at the ASIS&T website.

---

The International Calendar of Information Science Conferences ([icisc.neasist.org/](http://icisc.neasist.org/)) is a nonprofit collaboration between the Special Interest Group / International Information Issues (SIG/II) and the European (ASIST/EC) and New England (NEASIST) chapters of the Association for Information Science and Technology, with the additional support of Haworth Press.
Claire McInerney is the new acting dean of the School of Communication and Information, Rutgers University, succeeding Jorge Reina Schement, who was appointed Rutgers’ first vice president of diversity and inclusion. Dr. McInerney joined the SC&I faculty in 2000 and most recently served as associate dean for curriculum and assessment. She also serves on the editorial board of *JASIST*, the *Journal of the Association for Information Science & Technology*.

Heidi Julien, most recently a professor at the University of Alabama, is joining the Department of Library and Information Studies, University of Buffalo, in mid-August. She will serve as chair and professor in the Department of Library and Information Studies. Her research interests focus on information behavior, information literacy and education for information studies. She has served as chair of ASIS&T SIG/USE.

**An Asian-Pacific ASIS&T Chapter?**

by Emil Levine

Twenty-two participants attended a roundtable discussion explaining the Association for Information Science and Technology and considering the formation of an Asian-Pacific ASIS&T chapter. The session was held at the 5th International Conference on Asia-Pacific Library and Information Education and Practice (http://aliep2013.com) July 10-12, 2013, in Khon Kaen City, Thailand.

Dr. Yin Leng Theng, associate professor, Division of Information Studies, Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, chaired the roundtable. She was joined by former ASIS&T president Candy Schwartz, Graduate School of Library and Information Science, Simmons College, and a keynote speaker at the conference, and Emil Levine, former chair and advisor of the European Chapter, who made the presentation.

ASIS&T promoted the roundtable by sponsoring a “light buffet” for the attendees. Dr. Kulthida Tuamsuk, dean, Faculty of Humanities and Social Sciences, Khon Kahn University, Khon Kaen, Thailand, graduate of Simmons College and chair of the organizing committee for the conference, also assisted with the coordination of the roundtable.

The event was an outgrowth of the ASIS&T European Chapter’s 2012 annual meeting held in conjunction with the Libraries in the Digital Age (LIDA) conference in Zadar, Croatia, in June 2012. At this meeting, attended by Levine, Dr. Theng and Dr. Makiko Miwa, The Open University of Japan, a discussion began on creation of an Asian-Pacific Chapter (see http://www.asis.org/Bulletin/Aug-12/AugSep12_Levine.html).

Levine outlined the structure and program of ASIS&T, how to create a chapter and student chapter, ASIS&T publications and activities, and costs of membership.
Attendee Comments

Yin Leng Theng interacted with the attendees to determine their ideas of the opportunities and challenges in creating an ASIS&T Asian Pacific Chapter. This interaction elicited the following brief discussions, conclusions and actions:

- Although JASIST is the top journal in the field, it may be too “technical” and “difficult” for many potential members. There is a need for a mid-range journal.
- The Taipei Chapter has not been so active but will increase its activity.
- Through ASIS&T, students can take advantage of the networks, meet famous people and start student chapters.
- Taipei, Bangkok and Singapore will promote student chapters.
- Former president Schwartz stated that capacity building and relevant formulation of information policies are essential.
- Indonesia noted their attendees needed to know more and that there is a lack of awareness about ASIS&T.
- Malaysia considered that one might increase awareness through academic staff and LIS schools.
- Social media should be exploited in the creation of an Asian-Pacific Chapter.

Dr. Songphan Choemprayong, a graduate of the School of Library and Information Science, University of North Carolina, Chapel Hill and lecturer in the Department of Library Science, Chulalongkorn University, Bangkok, agreed to serve as a coordinator for creation of an Asian-Pacific Chapter.

Christopher Khoo Soo Guan, also from Wee Kim Wee School, Singapore, and Hao-ren Ke, chair of the ASIS&T Taipei Chapter, also attended the roundtable and contributed to the discussion.

See Attendee Comments at above right for more from the roundtable. Additional photos of the roundtable and information on A-LIEP 2013 are posted at http://aliep2013.com/.
To explore the feasibility of holding an annual meeting outside North America, the ASIS&T International Relations Committee (IRC) conducted a survey from September 18 to October 6 in 2012. The purpose of the survey was to determine the likelihood that ASIS&T members would attend an ASIS&T Annual Meeting and/or other events outside North America over the next five years. This report summarizes the main findings from the survey.

Main Findings

Validity of findings: Self-Selective Sampling Fits the Membership Population in Terms of the Geographic Distribution. A total of 247 responses were received. About two-thirds of all respondents are from North America (70.45%), followed by Europe (15.79%). The 2010 membership data indicated that members from North America represented 87.6%, followed by Europe at 6.4% and Asia at 3.5%. Although 2012 the membership count is not available, the percentage matches 2010 pattern (See Table 1). Compared with the 2010

![Table 1. Geographical location of respondents (N = 247)](image)

Note: Misc means that the respondent does not select any country.

Mei-Mei Wu is chair of the ASIS&T International Relations Committee. She is a professor in the Graduate Institute of Library and Information Studies at the National Taiwan Norman University (NTNU), Taipei, Taiwan, ROC. She can be reached at meiwu<at>ntnu.edu.

Diane H. Sonnenwald is immediate past president of ASIS&T. She is professor and chair of information and library studies at UCD in Dublin, Ireland. She can be reached at diane.sonnenwald<at>ucd.ie.
Inside ASIS&T

Membership data provided by ASIS&T headquarters, the survey sample was a reasonably good representation of the overall membership.

Preference for having an annual meeting in Europe rather than in the Asia-Pacific (AP) Region. The first part of the survey questions asked about the likelihood of attending ASIS&T Annual Meetings outside of North America. Figure 1 (question 1a) shows that 49% of the respondents indicated they were “very likely” or “likely” to attend an annual meeting in Europe, with only 21% stating they are “very unlikely” to attend an annual meeting in Europe. Figure 1 (question 1b) shows that 28% stated they are “very likely” or “likely” to attend an annual meeting in the Asia-Pacific (AP) region, and 40% are “very unlikely” to attend an annual meeting in AP. The data shows a stronger preference of having an annual meeting in Europe rather than in AP.

Respondents commented that holding an annual meeting outside North America could increase our membership. They wrote:

An excellent idea that would consolidate ASIS&T as a truly international association. For decades people from outside of the USA have made great efforts to attend – and at great expense.

If ASIS&T is going to grow as an organization, it needs to expand beyond the United States to attract and sustain vibrant member communities all over the globe.

Meetings outside North America would demonstrate ASIS&T’s openness to countries other than the United States.

Respondents commented that holding an annual meeting outside North America is also important for our discipline:

I do want to hear international perspectives on information science; they are absolutely vital to the progress of any field of research.

Greater involvement with our colleagues outside of North America is desirable. Definitely needed in this global environment.

Respondents raised concerns regarding U.S. green cards and visas and permission from some U.S. funding agencies and universities. It was noted that junior faculty waiting for green cards in the United States could not attend an annual meeting outside the United States. On the other hand, respondents also mentioned how difficult it can be to get a visa to attend the annual meeting in the United States. Two comments illustrate these points:

For our international junior faculty members who are awaiting green cards this could be a show stopper.

Right now it is very difficult to get an American visa although I got my doctorate degree from Case Western Reserve U, so I prefer if the conference is conducted outside United States.
Timing of the conference was another issue raised by respondents. A personal benefit to holding a conference outside North America is the opportunity to visit other places. However, holding the annual meeting during the U.S. academic fall semester limits opportunities to combine sightseeing with annual meeting attendance.

Would be most likely to go to Europe or an Asian/Pacific country if I were interested in combining it with travel in that area.

Many of our people have the wherewithal to pay to see an exotic place.

I would love to attend an ASIS&T Annual Meeting outside [the United States]... but the travel time in the middle of a semester would make it impossible.

The current annual meeting is during our university course time. A Europe annual meeting would be able to schedule in a better way, as the venue would be closer.

[From a respondent in Europe]

An issue raised in the past concerns participation in the administrative/business components of ASIS&T that are held during the annual meeting. If members could not attend an annual meeting outside North America, how would this impact the administrative/business components? It should be noted that currently this is an issue for international members. A respondent suggested a solution that could enhance the administrative components of all annual meetings:

The administrative components of annual meeting, e.g., business meeting, Chapter Assembly and SIG Cabinet, SIG business meetings) should be separated from the physical meeting when outside North America taking advantage of electronic communications.

Cost Analysis. As expected a concern expressed by respondents focuses on costs to attend an annual meeting. Many respondents from North America said that the costs to attend today are too high. Two respondents explained:

It is a hardship for me to even travel inside the United States. I had to really work hard to have the money to attend ASIS&T 2012.

My organization is not supporting any travel funding for the foreseeable future, so my chance of attending Annual Meeting even within North America is slim.

Thus no matter where annual meetings are held, some respondents will not be financially able to attend.

As shown in Figure 2 (question 4a) 51% of the respondents spend more than $500 to travel to the annual meeting. The average cost of travel is $672. Sixty-one percent reported spending more than $500 for lodging, with an average lodging cost of $673. When food and other costs are considered, an average cost of $1795 was reported. Recall, this cost does not include the annual meeting registration fee.
Table 2 provides an overview of all the related costs by geographic region. Not surprisingly since all annual meetings have been held in North America, the cost for attending is lowest for attendees from that region. But even so the cost to attend (excluding annual meeting registration fees) for attendees from North America is $1475.

As mentioned earlier, the average cost to attend an annual meeting is $1795. About 37.4% of this amount is for airfare ($672), another 37.5% is hotel costs ($673). Food (13.4%) and other costs (11.6%) account for another 25% of total cost. Attendees from North and South America spent 24.2% and 33.3% of average total cost on airfare, whereas attendees from Europe spent 42.8%. Attendees from other areas spent more than 50% on airfare.

Respondents mentioned their willingness to travel to an annual meeting outside North America provided costs are reasonable. They wrote:

As long as the venue is cost efficient.

If it were done in Western Europe in a location that was easy to get to from the United States, I’d be okay with it, providing other costs are kept down.

From the questionnaire data it appears that reasonable costs would be in the $1500-$1800 range for members in North America. Hotels are generally less expensive in Europe, and this may help reduce the registration fee as well as other costs because many hotels and universities in Europe charge groups less to use their facilities than hotels in North America. In addition many European universities offer inexpensive dorm rooms/suites for visitors. More rooms are usually available when the university is not in session, but usually a number of rooms are available even when the university is in session. With sufficient long-range planning as is usually done with regard to the annual meeting location these rooms could be reserved for annual meeting attendees.

There was also concerned expressed with respect to the financial ability of students to attend an annual meeting outside North America:

I do worry about the ability of students to attend the meeting if held outside of the United States.

It will be a good thing. This may enable ASIS&T to be better known outside North America.

<table>
<thead>
<tr>
<th>TABLE 2. Estimated costs by geographic locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airfare</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Africa</td>
</tr>
<tr>
<td>Asia</td>
</tr>
<tr>
<td>Australasia</td>
</tr>
<tr>
<td>Europe</td>
</tr>
<tr>
<td>Middle East</td>
</tr>
<tr>
<td>North America</td>
</tr>
<tr>
<td>South America</td>
</tr>
<tr>
<td>*No country identified</td>
</tr>
</tbody>
</table>

*7 respondents did not list their countries in the survey.
Many countries in Europe offer some financial support to international conferences that are held in their countries. This financial support could be used to support student attendance at the annual meeting. In addition student travel support could be sought from national funding agencies.

Respondents commented that holding the annual meeting periodically, e.g., every three to four years, outside North America would be a fair approach.

Holding them occasionally outside of North America would be good. Perhaps having every third conference outside of North America would be a good way to introduce this concept.

Maybe we can have the Annual Meeting outside of North America every three to five years?

What percentage of our membership is coming from overseas? Is it enough to warrant this “fair trade” or perhaps holding the annual meeting internationally once every four years or some other rotation?

Co-Located/Joint Conferences Outside North America Favored. Sixty-nine percent of the survey respondents said they were neutral, likely or very likely to attend an annual meeting held outside North America when the annual meeting was co-located with another conference (Figure 3, question 2a).
could co-locate with. Conferences mentioned most frequently were IFLA, SLA and ISIC. Each was mentioned eight times. The next most frequently mentioned conference was CoLIS (mentioned six times), and ACM SIGIR (mentioned four times). Given these responses it appears there is no strong consensus regarding which conference ASIS&T should consider co-locating with. This may be a reflection of the unique multidisciplinary perspective of ASIS&T.

**Recommended Frequency of Co-Located Conferences.** Sixty-two percent suggested ASIS&T should hold a co-located/joint conference every two to three years (Figure 5).

**Recommendations**

The survey results suggest an annual meeting could be held in Europe, ideally in conjunction with another conference every few years. The following action plans are suggested:

1. In addition to the conferences mentioned by respondents, perhaps an option could be to hold the Annual Meeting in conjunction with EuroIA (www.euroia.org), a conference ASIS&T already runs. This option could also help bridge the researcher-practitioner gap that exists in our discipline.

2. ASIS&T could extend a call to the membership asking for proposals with respect to holding an annual meeting in Europe. Each proposal should include cost estimates for travel and for local hotel and food costs as well as other items specified by the Executive Director. In addition any funding and/or sponsorship opportunities should be identified in each proposal. Many cities in Europe have non-profit organizations that can prepare such proposals.

3. Alternatively ASIS&T headquarters can develop a proposal that identifies one or two cost-effective locations in Europe for an annual meeting.

**Acknowledgements**

We would like to thank the 247 respondents who participated in the survey. We also bid our appreciation to the Board members for providing precious suggestions for the survey design and Dick Hill for the keen support for running the survey. Vivienne Houghton’s genuine contribution in preparing and running the survey and Dr. Ying-Hsang Liu’s assistance in data analysis are most appreciated.

![Figure 5. Period of co-located conference](image-url)
Douglas C. Engelbart, a pioneer in the design of interactive computer environments and the inventor of the computer mouse, died at his California home in early July at the age of 88. Born in 1925 in Portland, Oregon, Engelbart was midway through his college studies at Oregon State College, near the end of World War II, when he was drafted into the United States Navy. While serving as a radar technician in the Philippines, it is said that he was in the library one day on a small island when he happened upon Vannevar Bush’s article “As We May Think.” Engelbart was fascinated with the idea of Bush’s information retrieval system called Memex. He made it his life’s work.

In 1996, ASIS&T honored Engelbart with a special award for long-term contributions to the advancement of information science and technology, particularly in recognizing “enhancement of public access to information and discovery of mechanisms for improved transfer and utilization of knowledge.” At the same ASIS&T Annual Meeting at which Engelbart received that award, he was one of seven participants on a special panel entitled Reflections on Our Future. Other panelists were moderator Chuck Davis, Indiana University; Candy Schwartz, Simmons College; Gary Marchionini, University of Maryland; Belver Griffith, Drexel University; Clifford Lynch, University of California; and Eugene Garfield, Institute for Scientific Information. The session featured these visionaries, researchers and academics looking at the state-of-the-art of information science at that time within the context of what it portended for the future of society, as well as for the information profession. An edited transcript of Engelbart’s comments is included as a feature article in this issue of the Bulletin, beginning on page 44.

Gary Marchionini wrote the following words upon hearing of Engelbart’s death.

Some of you might have heard that Doug Engelbart died on July 2. Engelbart was the father of interactive computing and the intellectual inspiration for my work over the past 30 years. In the post-WWII period he was the first to realize that computing could “augment the intellect,” and he began working to demonstrate how to achieve this vision. His 1962 paper “Augmenting Human Intellect: A Conceptual Framework” was always required reading in my HCI seminar because not only does he lay out his vision for improving the human condition through computing but he presents ideas that led to word processing and collaborative work. Engelbart is best known for inventing the mouse and for his demo at the Joint Computer Conference in 1968 that demonstrated online networking (NLS) that included hypertext, word processing and the mouse. Through his bootstrapping institute he aimed to augment our collective intellect through collaborative technologies. Although he recognized the importance of collaborative technologies for all organizations, he was particularly interested in developing ways that professional societies, community groups and other socially responsible organizations could leverage computing to improve their impact and effect. At the 1996 ASIS&T Annual Meeting he made a plea for ASIS&T and other professional societies to increase collaboration to achieve common scholarly goals, a theme that resonated with the strong digital library movement underway at that time.

His influence on information science is paramount because he gave us a practical (tractable) path toward an ideal. Engelbart stands with Shannon, Bush and Simon as inspirations for information science. He inspired a generation of information scientists who work in HCI, HCIR and CSCW to empower and augment human capabilities. Most of all, Doug Engelbart was a kind and humble human being who would take the time to talk to young scholars, who accepted the trials as well as the blessings of life and who inspired many of us to keep our focus on why we invest so much in information and technology – to help people live productive and meaningful lives.
**RDAP13 Summit: Introduction**

by Susan Wells Parham, Elizabeth Rolando and Jennifer Doty, Special Section Guest Editors

**EDITOR’S SUMMARY**

This special section of the *Bulletin* explores issues in research data access and preservation (RDAP) and samples the program highlights and attendees’ reflections of the ASIS&T RDAP13 Summit, the 4th annual meeting, held in Baltimore, Maryland, on April 4-5, 2013. Contributed papers address the value of partnering with publishers and repositories as a way to address the complexity of data curation and explore the issue of data use and reuse from a data curator’s perspective. One article discusses the use, benefits and challenges of altmetrics to track citations of research data. Others tackle the need to develop a standard data infrastructure and describe efforts to promote research data management across institutions of higher education in the United Kingdom. Last, a sampling of the posters presented at RDAP13 reflects the scope of the topic, current progress and advances to come.

**KEYWORDS**

data curation  
altmetrics  
research data sets  
information infrastructure  
information reuse  
standards  
digital repositories

---

**The 4th Annual ASIS&T RDAP Summit was held in Baltimore on April 4-5, 2013. In this special section of the *Bulletin*, we have asked panel organizers, panel speakers, lightning talk presenters and Summit attendees to discuss their experiences at the Summit and to reflect on the remaining challenges and opportunities with respect to research data access and preservation. These authors provide program highlights, while further delving into issues important to the research data curation community. We’ve also included links to presentations and a selection of posters on slideshare, as well as information on how to learn about next year’s RDAP14 Summit. Special thanks go to the RDAP13 Summit program committee, panel organizers and program chairs Stacy Konkiel and Jordon Andrade.**

**Gail Steinhart**, co-chair of the panel *Partnerships Between Institutional Repositories, Domain Repositories, and Publishers*, reflects on the advantages of building partnerships to address the complexities of data curation, noting the specific success of panelists providing direct access to tools and services that support researchers. **Nic Weber**, organizer of the panel on *Data Use & Re-Use*, discusses highlights and emergent research themes and issues a challenge to the information science community to continue striving to adequately define the research area of “data use and reuse.” **Stacy Konkiel**, chair of the *Data Citation and Altmetrics* panel, provides perspective on recent developments and continuing challenges for accurate citation and tracking of research data.

**Mark Parsons**, panelist on *Global Scientific Data Infrastructure*, and his co-author Fran Berman describe in depth the efforts of the Research Data Alliance to address the data infrastructure challenge, and lightning talk presenter **Martin Donnelly** writes about the Digital Curation Centre’s efforts to improve institutional research data management in United Kingdom higher education, expanding on his talk given at the Summit. Finally, **Jennifer Doty** summarizes the poster session, spotlighting some of the unique projects and findings presented at this year’s Summit.
ASIS&T SIG/DL (Special Interest Group/Digital Libraries) sponsored the *Linked Data & Metadata* panel and an afternoon break for attendees. For additional information on the RDAP13 Summit and for plans for the 2014 Summit, please refer to the following resources:

- Slides from presentations and a selection of posters from the Summit are available on slideshare: [www.slideshare.net/asist_org/tag/rdap13](http://www.slideshare.net/asist_org/tag/rdap13)
- The full RDAP13 program, with additional information on Summit themes and presentations: [www.asis.org/rdap/program/](http://www.asis.org/rdap/program/)

Join the RDAP community’s on-going conversation and stay updated on RDAP14 event information by signing up for our email listserv at [http://mail.asis.org/mailman/listinfo/rdap](http://mail.asis.org/mailman/listinfo/rdap)

Congratulations to incoming Summit chair Andrew Johnson. We hope to see many of you in 2014 at the 5th Annual Summit in San Diego, California.
Partnerships Between Institutional Repositories, Domain Repositories and Publishers

by Gail Steinhart

Numerous stakeholders have an interest in ensuring that digital research data are made accessible and preserved for the long term. Stakeholders include researchers themselves, their institutions, their discipline-based communities and repositories, publishers and the general public. These stakeholders have diverse perspectives and goals but also share a common interest in advancing discovery and preserving and providing access to the scholarly record. Curating the record is a complex undertaking, with many tasks, tools and roles. Creating partnerships among stakeholders is emerging as one way to address the complexity of the tasks at hand. Partnerships can be explicit, with agreements to work together toward a common goal, or implicit, as when an organization works to develop tools and resources to meet its own needs while also addressing those of a larger community.

The three speakers on this panel, Jared Lyle of the Inter-university Consortium for Political and Social Research (ICPSR), John Kunze of the California Digital Library (CDL) and Amy Nurnberger of Columbia University Library, identified several challenges related to curating research data. For institutional repositories, as well as domain repositories and publishers, these challenges can include persuading researchers to expend the time and effort required to submit their data to a trusted repository; organizing, documenting and enhancing research data to make it usable by others and preservable over the long term; dealing with a wide range of media types and file formats; working with data from a broad range of disciplines and providing incentives and measuring impact in order to reward researchers for sharing data. Overcoming these challenges and making research data widely available supports validation and replication of research and facilitates new discoveries. The partnerships described by the three panelists all offer some distinct advantages to working independently to achieve these goals.
Partnerships Between Domain Repositories and Institutional Repositories

Jared Lyle described ICPSR’s work to engage institutional repositories and data librarians to collect and curate research data [1]. Inspired by the work of Green and Gutmann [2] and with funding from the Institute for Museum and Library Services, ICPSR has been making available its considerable expertise and curation resources to local curators. Work by ICPSR [3] showed that while in spirit researchers are willing to share their data, in practice they are less inclined to do so, with lack of time, resources and expertise presenting considerable barriers. Nevertheless, the significant value of shared data was apparent, with many more secondary publications (publications authored by individuals not associated with the core research team) resulting when data are shared than when they are not. A related study, examining the fate of data from NSF- and NIH-funded projects, showed that about 20% of projects had archived their data, about 50% had un-archived copies and nearly 25% of projects could no longer retrieve or locate their data [4]. Taken together, these studies suggest an opportunity for local data curators if they engage with researchers earlier in the research process, recruiting datasets into their institutional repositories and/or domain repositories such as ICPSR and finding ways to demonstrate the value of shared data. Lyle pointed out that by intervening earlier and by making the data curation process more personal, local curators might be more successful in recruiting data for deposit than those at more remote data archives.

While local curators may have easier access to researchers and their data, domain repositories tend to have greater and more advanced curatorial resources. Working with a handful of institutional partners to collect and curate several historical demographic data collections, ICPSR is experimenting with making available to their partners tools such as QualAnon (an anonymization tool for human subjects data) [5] and providing access to their data processing pipeline. Some challenges they and their partners have encountered along the way include the need to select and appraise candidate data collections. Institutional partners find themselves in the position of having to sift through content that investigators had not necessarily intended to deposit, often with little initial information or guidance from the data owners. Some of these collections are stored on obsolete media or are in unreadable file formats. While institutional partners do enjoy a local advantage in terms of access to researchers and their data, researchers still have difficulty finding the time to work with curators to interpret, organize and document their datasets.

In spite of these challenges, ICPSR identified a number of productive roles it can play to facilitate data collection and curation at partner institutions. From a survey of repository managers and curators, ICPSR found that help with media recovery, format migration, data recovery, cost estimation, tools for metadata creation, policy review and confidential data dissemination are all useful for institutional partners. ICPSR can also serve as a “community wayfinder” by continuing to develop resources such as the Guide to Social Science Data Preparation and Archiving [6].

Partnerships Between Publishers and Institutional Repositories

Amy Nurnberger, Columbia University Library, described the library’s work with the Public Library of Science (PLoS) and the Ecological Society of America (ESA) to better understand the behavior and requirements of authors with respect to sharing and depositing data related to their publications. She noted that the library and publishers share common goals, including a commitment to serving the scholarly community to advancing scholarship through publication and data sharing to supporting robust connections between publications and their underlying datasets and to ensuring that authors and data owners are properly credited for their work. She noted also that libraries and publishers face the same challenges and questions: locating and recovering datasets, understanding how data are reused and understanding and overcoming barriers to data sharing.

Implicit Partnerships: Serving the Curation Community at Large

In the spirit of offering tools and capacity to a larger community, John Kunze described some of the work of the California Digital Library and how it benefits researchers and the data curation community at large. Kunze argued that libraries are well positioned to do this work, given the broad
range of stakeholders, libraries’ status as neutral entities and their experience preserving the scholarly record. Nevertheless, data present some new and different challenges in comparison to journal articles, the more traditional form of research publication, particularly when it comes to incentives for researchers to publish their datasets.

CDL’s tools target different stages of the research process. At or even preceding the data collection stage, the DMP Tool \[7\] simplifies the creation of data management plans by providing a series of templates customized according to research funder and by partner institutions to meet the specific needs of their researchers. DataUP \[8\] works with Microsoft Excel to help researchers create metadata for spreadsheet-based datasets, obtain a DOI and deposit directly to a repository. To facilitate data publication, the EZID service \[9\] supplies researchers with a permanent identifier (currently DOIs – digital object identifiers – and ARKs – Archival Resource Keys – are supported), even in advance of publication, making it easier to link datasets and related publications. The Merritt repository \[10\] (a University of California instance is available to UC system researchers, and the software itself will soon be available for use by other institutions) and ONEShare repository (a special instance of Merritt, linked to DataUP and available for use by anyone) make it possible for researchers to store and share their datasets. Finally, to make data publication attractive to researchers, CDL is a founding member of DataCite \[11\], a consortium which aims to support data citation, discovery and reuse. Kunze noted that there are other important activities in the data citation area, including Thomson Reuters’ Data Citation Index and the development of alternative approaches to measuring impact such as altmetrics and ImpactStory.

**Common Themes**

**Everyone is busy.** Lack of time and resources is consistently challenging to researchers, even when they are supportive of sharing data. Partnerships that relieve researchers of some of the burden of preparing data for sharing and archiving can help to address this problem. ICPSR’s practice of enhancing data selected for their collection is one such example, when it can be extended to include institutional partners, as it has been in cases where partners are granted access to ICPSR’s data processing pipeline. Placing curatorial tools directly into researchers’ workspaces is another promising strategy. CDL’s DataUP is an excellent example of providing curation tools in a familiar and widely used environment.

**Local service providers have an advantage.** Both Lyle and Nurnberger asserted that local data curators have an advantage in working with researchers. When working locally, the process can be more personal and can also proceed iteratively. University research offices are also increasingly taking an interest in data retention and in helping researchers meet the requirements of funders, which can help local curators gain traction with researchers. While it may be too soon to tell whether researchers prefer to deposit data in their institutional repositories, in domain repositories (where more specialized tools may be available and where there may be added value in having datasets co-located with similar content) or with publishers, local contacts to assist with the work of curating research data are helpful. Coupling the greater access to researchers that local contacts such as librarians have with the considerable expertise of institutions such as ICPSR and CDL holds great promise for addressing data curation challenges.

**Heterogeneous data pose challenges.** The numerous file and media formats encountered by curators, as well as researchers’ idiosyncratic approaches to managing their own data, make it difficult to establish consistent processes and workflows for dealing with research data and/or require curators to set a lower bar for preparing data for archiving. It’s not just institutional repository managers that face this challenge: ICPSR and CDL have faced it as well. ICPSR is currently working to integrate a video collection into its data archive that will quadruple the size of its entire collection and that presents some challenges with respect to preservation quality, as well as possible confidentiality and disclosure issues. CDL, in managing repositories for the University of California system, has decided to make their repositories format agnostic, while doing their best to encourage researchers to adopt preservation-friendly file formats.

**We need better tools.** ICPSR and CDL currently serve the curation community admirably by providing a variety of tools for use by repository
managers, curators and researchers; however, the discussion took a lively turn when Kunze pointed to a class of repositories that have largely been absent from the discussion: services such as FigShare and SlideShare. These services are very popular among researchers, cross disciplinary and very easy to use, leading him to dub them a new category of “low-barrier repositories.” There was some discussion as to whether the research library community and their partners should attempt to develop similar services and some concern over the sometimes unclear business and preservation models of these currently available services. The discussion also touched on the need for better metrics and tools to help researchers demonstrate the impact of data, and by implication, increasing the importance and recognition of these activities with respect to tenure and promotion.

Overall, it appears some of the curation community’s most productive approaches and strategies are emerging from stakeholder partnerships, with some impressive successes to date. A solution developed by one organization can be widely adopted and applied by others, avoiding duplication of effort to solve common problems. Partnerships can also exploit local contacts and local knowledge, facilitating relationships across institutional boundaries.

Resources Mentioned in the Article

[9] EZID: http://n2t.net/ezid
The Relevance of Research Data Sharing and Reuse Studies
by Nicholas M. Weber

The hyperbole of “big data” [1] and the surprising backlash for “small data” advocacy [2] have been well documented in popular journalism, but research data are definitely experiencing a cultural moment. Much of this excitement turns on the potential that increased accessibility, interoperability and computing power can offer to the exploration of loosely related datasets (for example, tweets and the fluctuations of a stock index like NASDAQ) [3].

Research data’s cultural moment should also be one for the field of information science. Our field has traditionally studied some of the most difficult problems in the use of large-scale information resources, including the meaningful organization, access, management and storage of scholarly products in all of their formats and encodings. But, thus far, our field has struggled to make its expertise in this area well understood, and more importantly, we’ve been slow to demonstrate the relevancy of our work to the vital issues that we face as an intellectual community and, more importantly, as a society [4].

Our struggles stem in part from the fact that this space is already crowded with sociologists, economists, computer scientists and statisticians, to name a few of the disciplines involved. These disciplines all play an increasingly important and insightful role in building information systems, developing standards and creating services to support the meaningful use and preservation of research data. While each of these fields faces similar dilemmas with respect to meaningful engagement with research data, information science should be well equipped to handle systems-based problems. In short, we must better apply what we’ve traditionally known about citation behavior, document retrieval and information seeking to a data-intensive paradigm, while simultaneously avoiding generic simplifications such as “publications are just like datasets.”

Another part of our struggle is that we have a poor conception of our
problem space – as a colleague more elegantly put it, complications in data sharing and reuse are often due to the fact that we study “poorly bounded” digital objects [5]. At the most basic level, we have no idea how much data actually exist within an institution, department or even research group, let alone the entire enterprise of science [6]. This point is an especially important one for our field to address earnestly; it is tempting to indignantly accept headlines like “75% of research data is never made openly available” [7], but how much or how little of research data are “made available” is a fundamentally unknowable number. Irresponsible statements such as this headline hurt our cause much more than they help. We have no idea how much research data are produced or stored, let alone how much of it is shared. The goal of information science shouldn’t even include the search for such numbers. Instead, we should seek ways to meaningfully define our research subject so that we can make reliable statements about what is knowable – a perfect example is a recent study indicating the citation advantage for publications with archived, openly accessible data [8].

**RDAP’s Promising Results**

Over its short period of existence, the Research Data Access & Preservation Summit has become a crucible for data sharing and reuse studies. It provides a forum for those working in research institutions to share early results and take stock of or note gaps in our current understanding of these two related issues. Two particularly promising areas of research in sharing and data reuse were presented this year at RDAP – I highlight them both because their preliminary results are exciting and to combat any negativity that might be inferred from my earlier comments.

**Data Management Planning.** The first area is the rapid and unbridled success of research data management planning. Most formalized data management planning services and tools have emerged only as a result of requirements from funding agencies for grant applicants to explicitly document how research data will be stored and made available for future use. As these mandates were handed down, many institutions were quick to adopt templates and develop tools to help grant applicants submit competitive data management plans. Emerging from the data management planning work is a thread of evaluation research that looks at how these policies have shaped grant applicants’ behaviors and points out the sometimes subtle gap between the expectations of funding agencies and potential best practices for institutions supporting basic research activities. The RDAP13 lightning talks presented by Katherine G. Akers and Jennifer Doty, Heather Coates and Martin Donnelly all address this research question.

The RDAP community has been especially active in this sphere, and this year a quarter of all accepted Summit submissions had data management in the title or subject. Some of these presentations included general overviews, while others offered a more detailed look at data management for a more narrowly defined field, such as Konkie’s lightning round talk, “Bootstrapping Library Data Management Services for Epidemiology.”

As part of the panel Data Use and Reuse – Sharing Open Data Success Stories, Renata Curty shared preliminary results from a survey study of successful NSF grant applicants. The approach taken in this pilot study was sample awardees that recently created data management plans for NSF – asking about attitudes and opinions on both the process of creating and the prospect of sharing the data produced in the course of their funded research projects. Though this survey was largely a proof of concept study, its results can serve a dual purpose. On the one hand it helps in the gathering of valuable demographic data so that the RDAP community can better understand questions regarding who shares or reuses which data; on the other, it provides valuable insight as to what is difficult about this process and what can be improved in the near future.

This study is a promising first attempt to look at how sharing and reuse are affected by data management planning policies from the perspective of data producers. We should also think about operationalizing this type of study across different funding agencies and awardee types – as there are many public policy implications to this type of work. To put this work in some perspective, the U.S. government contributes about 59% ($32.6 billion) of the $54.9 billion in annual spending on research and development in higher education. Of that $32.6 billion, six agencies (Departments of Agriculture, Defense and Energy, NASA, NIH, NSF) provide 97% of all research and development funding [9], and NSF has by far the smallest...
budget of the six. Understanding differences among awardees of NIH, NSF or DoD is an important next step to understanding sharing and reuse in a broader research data context.

**Data Sharers, Data Reusers.** A second exciting thread of research focuses more closely on the behavior of researchers in sharing and reusing data, often from a practice-based perspective. This vein of research is similar to traditional use and user studies in information science, but also includes issues of appraisal, valuation and quality that cut across archival science and information systems literature more broadly.

Ixchel Faniel presented novel findings from the DIPIR (Dissemination Information Packages for Information Reuse) project. In this iteration, she largely focused on data quality indicators for reuse satisfaction among quantitative social scientists. Early analysis of an Inter-university Consortium for Political and Social Science Research (ICPSR) survey indicated that, not surprisingly, data documentation, as well as completeness and accessibility, were important in determining data quality; however, survey results also indicated that producer reputation was not a significant factor in reuse satisfaction. Of particular value in this analysis is Faniel et al.’s framework of data quality indicators. This framework drew upon diverse information systems literature, and it will be interesting to see how it evolves as it is reused and tested by the data curation community in future applications.

Similarly, Dharma Akmon offered early analysis from a dissertation that tackles large questions about how time-scales and accessibility affect perceptions of value in research data. The implications of this work may affect not only systems development, but also our understanding of data practices more generally. Akmon’s field work and observations of stream ecologists emphasized an evaluation of data in an instrumental capacity, which resonates with Bernd Frohmann’s plea for information science to more seriously consider material practices in user studies [10]. This type of ethnographic work plays a mapping role similar to the surveys discussed by Curty and Faniel, but we get a narrower, yet richer account of how data use and reuse might be better supported in field campaigns. This perspective is especially valuable as data curators begin to embed themselves in research centers and individual small science laboratories.

Slides from the Summit for many of these talks are available on slideshare at [www.slideshare.net/asist_org/tag/rdap13](http://www.slideshare.net/asist_org/tag/rdap13).

**Future Directions for RDAP Research**

At a time of sequestered budgets and shrinking economies, our field should be capable of providing policymakers and legislators with reliable information about the value of research data from one domain to the next. As the RDAP community begins to coalesce and clearly articulate its research agenda, it is similarly worth considering whom the audience of this research might include. As studies of data sharing and reuse expand in scope and sophistication, so, too, will their impact. While ASIS&T may not have the capacity to lobby Capital Hill on our behalf, the research of this community increasingly plays an evaluative role for agency mandates and funding at a federal level, so it seems natural that we should point our research results, however preliminary, in that direction.
Resources Mentioned in the Article


Tracking Citations and Altmetrics for Research Data: Challenges and Opportunities
by Stacy Konkiel

EDITOR’S SUMMARY
Methods for determining research quality have long been debated but with little lasting agreement on standards, leading to the emergence of alternative metrics. Altmetrics are a useful supplement to traditional citation metrics, reflecting a variety of measurement points that give different perspectives on how a dataset is used and by whom. A positive development is the integration of a number of research datasets into the ISI Data Citation Index, making datasets searchable and linking them to published articles. Yet access to data resources and tracking the resulting altmetrics depend on specific qualities of the datasets and the systems where they are archived. Though research on altmetrics use is growing, the lack of standardization across datasets and system architecture undermines its generalizability. Without some standards, stakeholders’ adoption of altmetrics will be limited.

KEYWORDS
altmetrics
research data sets
citation analysis
standardization
access to resources

Stacy Konkiel is science data management librarian at Indiana University. She can be reached at skonkiel<at>indiana.edu.

The recently announced San Francisco Declaration on Research Assessment [1], which calls for the abandonment of the journal impact factor as a means to determine the quality of research, highlights how important and contested the measurement of scholarly impact has become. Measuring impact for research data is also complicated. Data citation itself is not yet a standard practice [2, 3], and there is no authoritative agreement on how and when data should be cited [4]. Altmetrics, which track scholarship’s usage on the social and scholarly web, comprise a nebulous group of metrics that use an ever-shifting list of web services’ APIs as a source of their data [5]. As with data citations, standards do not yet exist to record or report the impact of different types of altmetrics. In light of these challenges, a panel was convened at the ASIS&T Research Data Access & Preservation Summit 2013 (RDAP13) to discuss new developments in exactly how researchers track the impact of data.

Overview of Data Metrics

Though discussions of data citation practices have occurred since the 1980s, it is in recent years that domain specialists, scientometricians and data curators have attempted to define standards for the citation of data and other data-related metrics. The closest the field has come to defining a standard is establishing DataCite [6], an organization that registers permanent identifiers (PIDs) for data and indexes associated metadata for discovery.

Such standards were the subject of the National Academies’ Board on Research Data and Information workshop, “For Attribution – Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop” (2012), a full report of which is available at the National Academies Press website [7]. Various stakeholders, including
Researchers, librarians and publishers, put forth their positions on what attribution for data should look like (citation versus varied metrics), what functions it should serve (attribution, showing provenance or defining the impact of researchers overall), how its infrastructure should behave (characteristics of host repositories, executable papers or linked data) and which communities are responsible for its development and implementation (libraries, publishers, data centers or researchers). No single position or suite of recommendations emerged from the meeting nor from a similar meeting, “Bridging Data Lifecycles: Tracking Data Use via Data Citations Data Workshop” [8], held earlier that year.

Other researchers are tackling the problem of tracking impact with a bottom-up approach. The Data Usage Index (DUI) has been proposed for the field of biodiversity, based on a variety of metrics culled from the Global Biodiversity Information Facility (GBIF) repository [4]. The authors call for a move beyond data citations, which mimic the citation of traditional publications, primarily because existing metrics do not “recognize all players involved in the life cycle of those data from collection to publication” nor are they yet standardized. Based on usage logs from the GBIF servers, Ingwersen and Chavan conceptualized a set of measures that are either “absolute” or “relative”: number of searched records, download frequencies, number of datasets, download densities and number of searches, to name a few of the 14 metrics. These measures are intended to show value to the researcher and to be used to demonstrate impact in a manner analogous to other altmetrics. While the study has implications for further development of related DUls, the authors acknowledge that their index is specific to the GBIF repository and therefore not generalizable to all research data repositories.

These limitations are the starting point for the study, “The Product and System Specificities of Measuring Impact: Indicators of Use in Research Data Archives,” presented at the 2013 International Digital Curation Conference [9]. The overall aim of the group’s research is to develop a suite of metrics that can expose the value that data curators add to a dataset, which in itself is an intriguing concept. The researchers’ conceptual framework is especially interesting in that they acknowledge that so-called “specificities” of systems and products – that is, the various sociotechnical factors that influence a system’s or an organization’s design and development – have more to do with the value of metrics that can be extracted than external factors.

Data curation work is related to both system and product specificities. It is reliant on a system’s specificities – architecture and arrangement that dictate how the user can interact with an archive – in that such specificities have an influence on metrics like number of search hits and number unique users who can discover the content. Equally important are product specificities, which are “the qualities and properties of the datasets themselves – their file structure, format and size – that affect the way a user can interact with the archive in consuming and discovering data” [9]. Though the researchers do not go into detail about the effects of particular data-curation activities (such as describing data using metadata standards and controlled vocabularies, reorganizing data for understandability and consumption) on data metrics, the area is tantalizingly open for further study.

Another major study in this area that addresses the various metrics, stakeholders and infrastructure considerations from a 10,000-foot view is the report “The Value of Research Data: Metrics for Datasets from a Cultural and Technical Point of View” accessible at the Knowledge Exchange website [10]. The authors give a rich overview of the challenges and opportunities that lie in capturing metrics for data and report on stakeholder views of the viability of the currently available metrics. Chief among the challenges are culture and infrastructure.

The authors posit that researchers have little reason to value data metrics (including citations) as yet, since they are not considered as valuable as citations to traditional publications. They also have little reason to adopt practices that will enable data metrics to be easily tracked, such as standardized citations for data or the assignment of permanent identifiers such as DOIs (digital object identifiers) to datasets, because the technical infrastructure currently does not support such practices for the most part. This presents a chicken-or-egg conundrum for those developing infrastructure for data citation, which is currently suboptimal, as there does not yet seem to be a need for such an infrastructure, given the lack of interest from researchers. Results from stakeholder interviews and environmental scans inform much of their report.
At RDAP13, Kathleen Fear, University of Michigan; Elizabeth Moss, ICPSR; and Heather Piwowar, ImpactStory/Duke University, presented their work and research related to measuring the impact of data. While all three researchers agreed that data citation is a good way to measure scholarly impact, they also shared their ideas on how to capture a fuller picture of the impact of data, including how the data has been reused and by which audiences.

The Impact of Data Reuse: A Pilot Study of Five Measures

Fear, a PhD candidate at University of Michigan, began the panel by sharing her research into the many ways that citations and usage statistics such as downloads can be used to track various degrees of impact for social science datasets [11]. The impacts boil down into five categories: data reuse, quality of publications that reuse data, diversity of publications that reuse data, size of network stemming from a single dataset and number of unique individuals who download a dataset.

The measurement for the number of times the data has been reused is analogous to how many times a dataset has been cited. While most datasets in Fear’s sample had never been cited, many were cited two to 10 times over the course of their lifecycle, with some receiving as many as 30 citations in journal articles. Fear measured the quality and diversity of publications that cite (reuse) the data by determining the citation rates for articles that cite the datasets and the breadth of publications. She noted that reuse rates can be affected by the publications in which a dataset was cited and also by disciplinary differences.

By counting the number of unique individuals who download a dataset, repositories can make general estimates of the data’s popular impact. However, we cannot be sure if downloads mean that the dataset has been used in any way, just as we cannot be sure that downloads of journal articles guarantee a paper has been read [9].

The final metric, the size of publication network that stems from a single dataset, is still being researched. The other measures are, interestingly, for the most part all interrelated. Fear found that data reuse counts had little to do with unique downloaders or the data’s secondary impact.

The results of Fear’s study are interesting, but are they generalizable to all data and data repositories? In our current environment, the answer is, “No.” In working with social science datasets culled from the Inter-university Consortium for Political and Social Research (ICPSR), Fear was able to track reuse using the repository’s Bibliography of Data-Related Literature (which is described in more detail below). The bibliography is, by necessity, a manually curated list; data citation standards have not yet been fully developed or implemented in a way that can automate the tracking for all data held in the ICPSR.

However, in a future where data is cited as strictly as prior publications are cited, one could imagine that Fear’s measures of impact take on great importance. Data potentially could have a much broader impact than publications, because they are open to interpretation and analysis: different communities often repurpose data in many different ways with many different results. Determining the scope and quality of that impact could speak volumes about the quality and utility of the data itself.

Viable Data Citation: Expanding the Impact of Social Science Research

ICPSR has done much in the years since its launch to track the citations for data stored in its repository via its Bibliography of Data-Related Literature – a manually curated list of more than 60,000 articles that are based in whole or part on findings culled from ICPSR data. In her presentation, the bibliography’s chief architect, Elizabeth Moss, stressed the importance of cultivating a culture of data citation: “Impact can be better measured if data use is readily discernible.” [12]

Impact is broken down by ICPSR to help understand who uses the data and to what effect. There are certain measures that ICPSR’s own website tracks easily: download statistics, unique sessions and users and the names of ICPSR member institutions where downloads of datasets occur. These metrics track who uses the data, while the Bibliography of Data-Related Literature more broadly tracks the data’s impact in the literature.

ICPSR has engineered some aspects of its repository to encourage citation of both data and related publications, as well as to support different uses by its various audiences. Within the bibliography, literature is searchable
and exportable to reference manager programs. Item records for publications link back to related datasets. This tool can be used in teaching students how to conduct and document their own research, helping researchers perform literature reviews, allowing researchers and funders to track how data is used and enabling reporters and policymakers to see both statistics and the related reports [12]. Digital object identifiers (DOIs) are also issued for data, both at the collection and the study level, with links resolving to the web page with the richest metadata that can help users understand the dataset. These system specificities likely have an effect on how the data is cited and on the other metrics that are collected, as described in the previous paragraph.

Despite ICPSR’s efforts to encourage good citation practices, Moss finds that data is rarely explicitly referred to in the literature or discoverable within academic databases. Often, ICPSR staff must comb through articles’ methods descriptions and figures to uncover the original dataset a project might be based upon. Most academic databases do not index data – it is simply out of their scope – and current full-text search capabilities are not sophisticated enough for the nuanced search techniques that are currently required to uncover references to datasets. Moss’s current strategy to overcome these challenges is to combine text-mining scripts with Google Alerts, which can alert Moss whenever a dataset’s creator is mentioned or its DOI is referenced.

ICPSR’s recent partnership with the Institute for Scientific Information’s (ISI) new Data Citation Index (DCI) initiative aims to address some of these issues by integrating its datasets and the Bibliography of Data-Related Literature, as well as many other repositories’ data and related citations, into the DCI database. Within the DCI, datasets are fully searchable and are treated as research objects that are on par with journal articles, conference proceedings and other traditional outputs. The database search functionality for the DCI as well as related databases like the Web of Knowledge is being converted to meet the needs of those searching for data. As a result, articles can be more easily linked to data, leading to increased data discovery, which is itself a reward for data citation and also rewards those who make their data easily citable – all these benefits from a search interface that many researchers are already using to find emerging research.

Moss concluded by explaining how ICPSR helps “build a culture of viable data citation to improve measures of impact” by providing principal investigators and users with citations, metrics and DOIs for data. Moss encouraged the audience to join groups and attend conferences to advocate for viable data citation practices, including DataCite, iASSIST and the Research Data Alliance. She also advocated that journal editors, domain repositories and funders work together to support repositories and change publishing practices, by requiring authors to better steward and clearly cite the data that underpins their studies.

No More Waiting! Tools that Work Today to Reveal Dataset Use”

Heather Piwowar, co-founder of the altmetrics service ImpactStory, discussed the responsibility of librarians, metrics providers and data scientists to go beyond citations when considering the impact of dataset reuse [13]. Altmetrics can track many types of engagement (views, saves, discussions, formal references and recommendations) that many different types of user groups (researchers, teachers, students, policy makers and practitioners) can have with a single dataset. Those are characterized as “impact flavors,” and tools such as ImpactStory, Altmetric.com and Plum Analytics are well suited to help aggregate and display them.

Piwowar laid out three ways in which the community can help encourage more diverse research metrics for dataset reuse: by exposing more metrics, supporting more types of engagement with datasets and lobbying and negotiating for Open Access to research.

Taking the ICPSR and its metrics as an example, Piwowar argued that content providers not only should provide information on dataset usage (downloads and pageviews for descriptive information), but also other rich metrics such as institutions from which a dataset was downloaded and classifications of unique users (into categories such as graduate students, undergraduates, university staff or faculty). However, many repositories do not expose any metrics, especially at the dataset level [10]. It is the responsibility of data curators and repository administrators to expose such metrics.

Secondly, datasets are complicated research products. The scholarly community has not yet figured out an efficient or standardized way to support
peer-reviewed data publications. It follows that scholarly social media sites like Faculty of 1000 or Mendeley would have difficulty addressing datasets and their usage. Piwowar called upon service providers – and altmetrics service providers – to report metrics for all types of engagement with data.

Finally, Piwowar advocated for advocacy itself, as it relates to data metrics. As data curators, librarians, researchers and university administrators, Piwowar argued that it is our duty to lobby and negotiate for open access to research, including open-text mining of articles, open data from repositories and open metrics from aggregators.

Piwowar’s last point led to a general discussion of whether repositories like ICPSR should allow commercial, toll-access services such as the DCI to index their metadata, much of which is the result of manual curation. Moss proposed the idea that any exposure to data, whether via the open web or a service like the DCI, is beneficial to the data creator and end user alike.

Piwowar, as the founder of a service that relies on open APIs to report metrics, acknowledged that toll-access services and closed APIs inhibit both the ability of end-users to find datasets and platforms such as hers to track their impact.

**Summary**

Data citations are just one metric that can be tracked to determine the impact of datasets made available through repositories. Altmetrics and usage statistics can determine the impact of data and publications beyond the academy and are useful supplements to citations. The technical infrastructure of repositories and the characteristics of the datasets stored in them can sometimes dictate which metrics can be applied to fully evaluate the impact of data. No metrics can be fully implemented until certain standards, such as DOI usage or commonly agreed-upon best practices for data citation, are widely adopted. Even then, manual intervention to link data to publications and other research outputs may be necessary, making the role of repository staff and librarians ever more essential.

**Resources Mentioned in the Article**

Resources Mentioned in the Article, cont.


The Research Data Alliance: Implementing the Technology, Practice and Connections of a Data Infrastructure
by Mark A. Parsons and Francine Berman

EDITOR’S SUMMARY
With research in demand for innumerable applications to society’s challenges, the Research Data Alliance (RDA) was formed to create and implement a unified data infrastructure. Established in March 2013, the alliance is directed toward establishing a framework to serve social and technical research needs at the local and global levels. The RDA brings together a community of stakeholders to guide the organization as it evolves and align participants’ diverse perspectives, striving for consensus, harmonization and an open community. Its structure features organizational and technical advisory boards, interest groups, shorter term working groups and an elected council to steer the alliance. With 700 members representing the academic, public and private sectors across 44 countries, the RDA seeks support from researchers, data managers, funding agencies and governments to promote the strategic and effective use of research data.

KEYWORDS
information infrastructure
research data sets
collaboration
information associations

All of society’s grand challenges – be it addressing rapid climate change, curing cancer and other diseases, providing food and water for more than seven billion people, understanding the origins of the universe or the mind – all of them require diverse and sometimes very large data to be shared and integrated across cultures, scales and technologies. This task requires a new form and new conception of infrastructure. The Research Data Alliance (RDA) is creating and implementing this new data infrastructure. It is building the connections that make data work across social and technical barriers.

RDA launched in March 2013 as an international alliance of researchers, data scientists and organizations to build these connections and the infrastructure to accelerate data-driven innovation. RDA facilitates research data sharing, use, reuse, discoverability and standards harmonization through the development and adoption of technologies, policy, practice, standards and other deliverables. Our vision is researchers around the world sharing and using research data without barriers.

The Challenge of Creating a Data Infrastructure

Infrastructure is hard to define. We don’t usually recognize it until it’s gone. When it’s there and functioning, we tend to take it for granted – we flip the switch and the lights come on. It is only when it stops working that the full complexity of the infrastructure is revealed. As such, it can be very difficult, if not impossible, to define new emerging infrastructure.

We may be able to plan out a vision of the physical infrastructure, but things never play out according to plan. Inevitable tensions and complex dynamics emerge. For example, it was one thing to design an orderly interstate highway system with even-numbered roads running east-west, odd-numbered roads running north-south and three-digit-numbered roads as connectors and bypasses. It was quite another thing to deal with how interstates altered U.S.
culture by promoting the growth of suburbs and fast-food chains while undermining other, often more efficient, transportation schemes.

Star and Ruhleder take a holistic and realistic view of infrastructure [1]. They argue convincingly that infrastructure is better viewed as a body of complex social, technical and sociotechnical relationships. They describe eight attributes that characterize infrastructure and ask not what is infrastructure but rather when is infrastructure. “Infrastructure is something that emerges for people in practice, connected to activities and structures” (p.112). Edwards et al take this view a step further in their report, Understanding Infrastructure: Dynamics, Tensions, and Design [2]. They show how we can arrange these eight attributes on technical/social and global/local axes (Figure 1.). The point is that we should not view the creation of infrastructure as a technical (or social) problem, but rather that we should decide whether we need social or technical, global or local solutions (or both) to infrastructural problems. It is in this dynamic space, dancing between the social and technical and between the global and local, that RDA works and tries to make an impact.

Edwards et al and the general field of infrastructure studies (for example, Bowker et al [3]) show that there are multiple things happening at once when we are designing a data infrastructure. We simultaneously need to enable distributed collaborative work, engineer changes in the organization of research, such as rewarding data sharing, and enable interdisciplinary collaboration and data sharing. This work is largely social and organizational, but system designers often do not recognize these dimensions of their work practice. A central conclusion of Edwards et al is that the necessary data infrastructure will not be built from the center with a single design. Instead, it will be built from the ground up and in modular units.

Edwards et al further argue that infrastructures become “ubiquitous, accessible, reliable and transparent” as they mature (p. i). They go through a staged evolution characterized initially by deliberate design of targeted technical services. Then technology transfer and adaptations lead to variations on the original design and the development of competing systems. Finally, a consolidation process occurs, in which systems link into networks and networks link into internetworks. This last phase is the crucial, make-or-break phase of infrastructure development. Arguably, the data infrastructure is beginning to enter this phase. It is also at this level where infrastructural tensions and conflicts are often most pronounced. Researchers can have both very proprietary and highly variable attitudes toward data. How and why research data are collected can be very personal, while the dataset itself can also serve as a public good. Data may be physical samples or output from a model. They may be numeric or textual, quantitative or qualitative, an intermediate outcome or a final product. Aligning and realigning these many perspectives is a major challenge and a primary function of RDA.

The RDA Solution
This theoretical examination of how infrastructure is created is illuminating, but it only takes us so far. How does it manifest itself into actual implementation? That is the challenge of RDA.

We take a community-based approach that seeks to create the modular units of infrastructure that interconnect over time, and we do it in an environment guided by common principles that encourage harmonization. We also recognize that the final consolidation phase of infrastructure creation is typically characterized by gateways, brokers or intermediaries that allow dissimilar systems to interconnect. These gateways are not only technologies, but are often a technical solution combined with one or more social choices.
like the adoption or adaptation of a standard [2]. We, therefore, focus on adoption and implementation of the tools, code, best practices, standards and so forth that are created.

RDA Organization

The overall RDA organizational structure is outlined in Figure 2. The goal is to create the sort of adaptive, responsive environment that helps us avoid the negative path dependence that new infrastructures often encounter and address the critical unresolved problems that impede overall progress.

RDA membership is open to anyone who subscribes to the RDA Guiding Principles:

- Openness – Membership is open. RDA community meetings and processes are open, and the deliverables of RDA working groups will be publicly disseminated.
- Consensus – RDA advances by achieving consensus among its membership. RDA processes and procedures include appropriate mechanisms to resolve conflicts.
- Balance – RDA seeks to promote balanced representation of its membership and stakeholder communities.
- Harmonization – RDA works to achieve harmonization across data standards, policies, technologies, infrastructure and communities.

- Community-driven – RDA is a public, community-driven body comprising volunteer members and organizations.
- Non-profit – RDA does not promote, endorse or sell commercial products, technologies or services.

Data, research and technology organizations may also join RDA as organizational members with voting privileges on an Organizational Advisory Board or as more informal affiliates.

RDA members form short-term, very highly focused working groups that make up the heart of RDA. Working groups conduct very specific, 12-18 month efforts that implement specific tools, code, best practices, standards and so forth at multiple institutions. Furthermore, to encourage harmonization and balanced choices, working groups produce detailed case statements that undergo extensive community and technical reviews before they begin. The short timeframe demands a focused, modular approach.

It takes time and community conversation to define the specific modular work conducted by working groups, especially given their short duration. Therefore, RDA also recognizes interest groups that have a broader scope and longer life. Interest groups work to define common issues and interests that ultimately lead to the creation of more focused working groups. Creating an interest group is a simple process of preparing a short charter.

A 12-member Technical Advisory Board, elected by the membership, is responsible for the overall technical direction of RDA. They also review working group case statements and ensure that the proposed approaches are appropriate, technically viable and not just the pet project of a few people.

An elected council of nine senior, well-respected leaders – or statespersons – maintains the overall vision for RDA. The council reviews working group case statements and public comments on the case statements. They provide an objective overview that gauges overall community consensus and ensures the proposed working group adheres to RDA principles and has a solid adoption plan demonstrating that they are making appropriate social as well as technical choices. Council also reviews interest group charters and ensures that they are within scope of RDA and not conflicting with other efforts.

The globally distributed Secretariat provides coordination and administrative leadership. They help coordinate the operations of the council,
boards and working and interest groups, provide overall communications and generally act as the face of RDA.

The RDA Colloquium (RDAC), consisting of representatives from governmental funding agencies, acts as an international steering committee that has provided initial funding support and organizational support for RDA.

A final, critical component is the RDA plenary. Plenary meetings are held every six months and allow the community to come together to conduct business and make progress on their plans and deliverables. The plenary is a forum to show the work of RDA, receive feedback from the broader research and policy communities and to continue to define and build relationships with initiatives and organizations that share the RDA vision.

Current Status and Plans

In this initial start-up phase, the organization and efforts of the RDA have been sponsored by the Research Data Alliance Colloquium (RDAC). RDAC currently consists of government agencies from the United States, the European Commission and Australia. It is likely to expand beyond those countries. RDAC appointed an initial organizing group and is appointing the initial council (currently six members and expanding to nine), but future council members will be elected by the membership as the terms of current members end. The organizing group has spun off several volunteer task forces to organize the rest of the RDA structure in collaboration with the Secretariat and council. The council has appointed an interim Technical Advisory Board of six people. The remaining six members of the board will be elected by the membership later this year. Thereafter, one-third of the board will be replaced in annual elections. A task force, collaborating with RDAC and council, is working to define organizational membership and the corresponding Organizational Advisory Board. A proposal should be out for community comment later this summer.

Despite the nascent structure of RDA, it is growing quickly. Roughly 700 members from 44 countries span all sectors, with 61% of the participants from academia, 11% from the public sector, 21% from the private sector and 8% undetermined. Several working groups have been recognized, and more than a dozen interest groups have formed. The number of working and interest groups continues to grow rapidly. As is their nature, working groups are focused and may be even a bit esoteric. Current working groups are working on specific details like agreeing on core foundational terminology, registering different types of data and identifiers and defining common computer actionable rules. Current interest groups are, of course, more broad-ranging; they address diverse topics including community engagement, legal issues around data sharing, repository certification, agricultural data interoperability and more. A current and complete list can be found at http://rd-alliance.org.

RDA officially launched in March 2013 at the first plenary in Gothenburg, Sweden. Some 240 participants from 31 countries attended. The second plenary will be in Washington, D.C. September 16-18, 2013, and attendance may top 400. The third plenary is likely to be in Europe again. A bidding process is being established for nations and organizations to host further plenaries.

Conclusion

The Research Data Alliance has emerged at a critical time when society is facing many complex problems requiring new and creative uses of diverse data. RDA has garnered the attention of senior researchers, data practitioners, funding agencies and government ministers not only for its timely emergence, but also for its pragmatic approach rooted in theory but focused on getting things done. It is an exciting time to be engaged in information science and technology. We encourage you to join RDA and get involved with its working groups and other structures.

Resources Mentioned in the Article


The DCC’s Institutional Engagements:
Raising Research Data Management Capacity in UK Higher Education

by Martin Donnelly

The Digital Curation Centre (DCC) is a United Kingdom-supported service and center of excellence for digital preservation and data management founded in 2004. This article gives an overview of the DCC’s Higher Education Funding Council for England (HEFCE) funded institutional engagement program (2011-2013), wherein we worked intensively with over 20 UK higher education institutions (HEIs) to identify, describe and raise levels of awareness and capability in institutional research data management. The set of engagements involved HEIs from all constituent countries of the United Kingdom and covered a range of institutional types, from ancient, research-intensive universities to newer universities seeking to raise their research profiles to specialist institutions with a focus in particular subject areas such as economics or art.

The Institutional Engagement Program: An Overview

Between Spring 2011 and Summer 2013, the DCC undertook a significant outreach program designed to assist a cohort of individual universities in the development of their research data management capabilities. Funded by the HEFCE’s Universities Modernization Fund, in sympathy with the DCC’s core funding from Jisc, this became known as the Institutional Engagement (IE) Programme.

This work shared as its bedrock the view subsequently expressed in the Royal Society’s Science as an Open Enterprise report ([1], p.8) that “a shift away from a research culture in which data is viewed as a private preserve” is essential to achieving improvements in the exploitation of research. Such cultural change requires the involvement and collaboration of numerous stakeholder groups and the joining together of what could be thought of as work silos. The work does not stop at the gates of the university, though: research funders, charities, government agencies and learned societies are all also involved.
Context

A number of factors influenced this work, including the increase in data-related requirements and expectations of research funders, publishers and the government. At the highest level, the UK government, via a white paper published in December 2011, has exerted increased pressure on their seven public funding councils to implement and enforce the data management policies and expectations that are currently in place.

Over and above this white paper, the Engineering and Physical Sciences Research Council (EPSRC) – the largest of the research councils in terms of monies disbursed – has informed all research institutions that they are each expected to prepare a data management roadmap by May 2012 and to have a functioning research data infrastructure in place by May 2015, galvanizing many senior managers, researchers and research support staff to take prompt action.

Other funders, notably the Medical Research Council and Economic and Social Research Council, have tightened up their existing regulations, and carried out quality checking exercises on the data management plans received at the application stage [2]. All of these developments point at a trend towards increasing regulation and requirements and expectations of higher quality data management.

The First IE Program

Cohort. At the very beginning of the program, we were able to help the limited number of universities that were interested in participating and met the basic criteria for the program, but – as word got around – demand to participate increased, and by the mid-point we had a waiting list of institutions keen to work with us. These institutions are now being contacted with a view to taking part in a follow-up program beginning in Summer 2013.

The first cohort comprised a variety of institutions, ranging from ancient, research-intensive universities such as Glasgow and Edinburgh to newer (post-1992) universities with smaller research portfolios to specialist institutions such as the London School of Economics and the University of the Arts London. We sought to work with at least one institution from each of the four countries that make up the United Kingdom: England, Scotland, Wales and Northern Ireland.

We initially calculated that the available funding would enable 18 full engagements, with an allocation of 60 days of DCC effort provided free for each, drawn from a mixture of staff based at each of the three DCC sites: Edinburgh, Bath and Glasgow. In practice, however, given the pace of the individual engagements, which in some cases were becoming extended over a lengthy period, we were able to extend this assistance to a few more institutions based on their particular circumstances.

Process. Each engagement was tailored to meet the specific needs of the institution in question, with the DCC providing in-depth and concentrated support to a community of institutions already known to be underprepared, in general, to face the growing challenge of research data management.

Following initial contact and agreement from a senior champion, the DCC and the institution agreed on the key contact people to manage the process from both sides. In most cases, task forces or working groups were assembled, and in some instances a cycle of business process redesign was put in motion. Three key perspectives were encompassed: that of the research practitioners themselves; support staff from research offices, libraries and IT departments; and the senior managers with their hands on the purse strings. (This situation was reminiscent of Cornell’s “three-legged stool” model, wherein technology, resources and organization need to be roughly equal in order to keep the chair upright [3].)

It quickly became apparent that the engagements needed authority, clarity and drive if they were to be successful. Without the ongoing backing of a senior figure, research support staff struggled to motivate or engage researchers. Without dedicated resourcing, this work often took a backseat to other pressing concerns, not least of which was the Research Excellence
Framework exercise, the semi-regular mechanism by which the UK government determines how much funding to provide to each university. This multi-stakeholder endeavor led to the creation of often unfamiliar collaborations and assemblages, as central services became responsible for issues that had previously been completely left to the researchers to handle, and, similarly, researchers became obliged to work more closely with administrative departments that they had been more accustomed to keeping at arm’s length.

**Activities Undertaken.** When developing work plans for each engagement, we sought to focus on desirability, achievability and sustainability. We found that early-stage funder analysis was useful, determining which funders were most important to the institution in question – both now and in the strategic future – and identifying the requirements that the institution would have to meet in order to safeguard this income stream.

The approaches taken to developing these new, data-centric competencies covered a range of contexts, aims, methodologies and processes. Work involved in the engagements ranged from developing research data management roadmaps and policies to identifying training and support needs to trialling and customizing tools such as DMPonline, the Data Asset Framework (DAF) and CARDIO to integrate within the existing technical infrastructures.

Each of the work programs was different, but we found that three issues predominated: planning, storage and training. Six of the seven state-owned Research Councils-UK (RCUK) funders require data management plans (or an equivalent) at the application stage, and one, the Natural Environment Research Council (NERC), requires two versions: a minimal plan at application and a jointly-agreed upon plan once funding is granted. The DCC’s DMPonline data management planning tool was therefore popular with the institutions, enabling the creation of tailored and branded versions that incorporated boilerplate responses, if desired, with institutional plan templates, as well as helpful links to local web pages, training materials, support contacts.

On the storage side, universities frequently came up against the problem of providing sufficient capacity for research data, while also competing with commercial offerings such as Microsoft Skydrive, as well as tools to enable sharing with collaborators such as Dropbox. Universities also struggled to provide advice on the selection of storage options such as onsite repository, national services or cloud services and advice on how to integrate relevant storage-focused initiatives, such as the Jisc-funded DataStage and DataShare projects.

The first of the EPSRC data-related expectations, which underpins any working data infrastructure, is an appropriate level of awareness and accompanying training resources, so that colleagues...
know what is expected of them. It is often said that it is neither necessary nor desirable for every researcher to become an expert in all aspects of research data management, but rather they need a decent grounding in basic principles (such as keeping a regular backup of anything significant and making sure it is not held in the same building as the master copy) and to be comfortable in seeking advice or help from the colleagues best placed to provide it. These are largely human infrastructure issues, and frequent communication is necessary to ensure they stay afloat in people’s minds.

**Lessons Learned.** At the time of writing (June 2013) we are finishing off the few engagements from the first cohort and beginning the next batch. The lessons and examples of the first cohort will be shared with the community via a synthesis report that is currently in preparation, and details will be available via [www.dcc.ac.uk/community/institutional-engagements](http://www.dcc.ac.uk/community/institutional-engagements) soon.

The chief interim observations are:

- Universities are generally in the early stages of scoping research data management (RDM) needs and obtaining benchmarks.
- Few have the effective components for RDM services or infrastructure in place at the moment.
- Management tend to be more concerned with meeting funder expectations than with a desire for sustainable infrastructure.
- There is confusion over the sequence for producing strategies, plans and policies.
- The gulf between early adopters and late entrants is widening.

**The Future for These Engagements**

With around 140 higher educational institutions in the United Kingdom and only a little over 20 involved in the first cohort, there is, of course, much work still to be done and more demand than the DCC can currently meet. We have, therefore, realigned our program to reflect the widening spectrum of need, offering specific curation techniques to enable infrastructure development, together with sociotechnical support, from advocacy to skills reengineering to organizational repositioning.

DCC staff will continue to be engaged as expert hands-on consultants delivering specific tasks, but the one-size-fits-all approach of 60 days has been done away with; instead, universities will make requests for specific, defined work packages which will be agreed upon by the DCC and a senior member of the university. Upon satisfactory completion of one piece of work, we can move on to another at the same institution, provided it offers a clear benefit to the institution or the wider research community.

Finally, HEIs will be required to demonstrate commitment to maintain the engagements, and if progress is not being made we have more frequent stop-go points to enable us to provide support where it is most needed and will be most valuable.

**Acknowledgements**

Parts of this article, including the diagrams, are drawn from a paper delivered by DCC associate director Graham Pryor at the 8th International Digital Curation Conference in Amsterdam in January 2013. The full paper is due for publication in the *International Journal of Digital Curation*. I am grateful to him and to my colleague Jonathan Rans for casting an eye over the penultimate draft of this article.

**Resources Mentioned in the Article**

2. A special Research Data Management Forum event held in April 2013 brought together a number of representatives of the UK research funders to enable them to go on the record about what aspects of research data management infrastructure (human and technical) they were and were not happy to fund via their grants. See Graham Pryor’s blogpost, *A conversation with the funders*. Retrieved June 13, 2013, from [www.dcc.ac.uk/blog/conversation-funders](http://www.dcc.ac.uk/blog/conversation-funders).
RDAP13 Poster Session Summary
by Jennifer Doty

EDITOR'S SUMMARY
Posters presented at the 2013 Research Data Access & Preservation Summit (RDAP) spanned an array of topics of current interest for research data managers. Helping doctoral candidates archive their research data was the focus of one, stressing the need to provide support early in the researchers’ process. Another described a multi-format data literacy program geared to graduate level researchers. The presentation on one university’s “DataDay” described a training workshop designed to help librarians understand the research data lifecycle and pass on essential points to researchers. Another team of presenters offered a methodical approach to establishing the provenance of research datasets that later appear in other contexts. Several posters addressed data management policies and practices at the institutional level, with one offering model policy language and guidelines and another analyzing current and likely future trends and needs to support archiving research data. The full collection of posters reveals the breadth of issues, progress to date and advances to come for the area.

KEYWORDS
research data sets
data curation
policy
standards

Jennifer Doty is a data management specialist with the Emory University Libraries. Her primary focus is developing services and support for the management and curation of research data at Emory. She can be reached at jennifer.doty<at>emory.edu.

Research Data Access & Preservation

With presenters covering topics ranging from provenance to preservation to publication of data, the RDAP13 Poster Session provided a wealth of information on recent trends in research data management. Posters ran the gamut with new explorations into systems and support for accessing, preserving and sharing research data.

The poster presented by David Fearon and Betsy Gunea of the Johns Hopkins University Data Management Services team demonstrated the challenges inherent in supporting previously published data. In a time when many librarians and information professionals are exploring methods to engage with faculty and student researchers, JHU presents a compelling model for working with PhD candidates to archive data related to their publications. The experience posed a challenge in providing support to document and prepare data underlying already published results and reaffirmed the importance of working with researchers earlier in the data collection process to effectively create project metadata.

Data information literacy for graduate students was the focus of the poster from the IMLS-sponsored project represented by librarians from Cornell, Purdue and the Universities of Minnesota and Oregon. By focusing their lens on researchers at the graduate level, the project has developed a comparative model for packaging data literacy instruction in a variety of formats (for example, online course, in-person seminar or workshop, embedded librarian) and applying it in a range of situations with students in scientific disciplines. The team adopted an integrated and tailored approach, and one of their most interesting findings is the usefulness of focusing on the mechanics of data management and the local, immediate needs of a specific research group.

From instructional programming for graduate students, we turn to an
example of participant-driven training for information professionals. The librarians at the University of Colorado Boulder presented a comprehensive overview of their first ever “DataDay” training designed for subject librarians. The program development began with participant input, which resulted in creation of an interactive one-day workshop that incorporated hands-on exercises, panel presentations and informal discussions. Pre- and post-training assessments revealed that workshop participants self-reported having higher levels of understanding of the research data lifecycle and greater confidence in their abilities to incorporate assistance with research data management into their established roles as subject librarians. The success of the initial session is attributed in part to the direct involvement of the participants in designing the workshop training, and future plans include additional assessment of the librarians’ training needs to strengthen this iterative process of acquiring knowledge about research data support.

Institutional data management policies have not yet been consistently implemented or reliably enforced at many higher education institutions. In the wake of the February 2013 Office of Science and Technology (OSTP) memo regarding federally funded research outputs, however, those with plans to raise the issue with campus stakeholders would do well to consult the Association of Southeastern Research Libraries (ASERL) Research Data Coordinating Committee presentation on a “Model for Developing Data Management Policy Language.” A collaborative effort from librarians at several ASERL member institutions, the process included an environmental scan of developments in data management support. The final outcome was model language for establishing institutional policies with clear guidelines to support researchers as they try to meet existing and potential funding agency mandates, as well as institutional expectations, for managing research data.

The challenge of establishing provenance and ensuring reproducibility of research for items lacking adequate metadata was explored using a representative figure from the National Climate Assessment (NCA), which serves as the base for the Global Change Information System. Justin Goldstein et al from the U.S. Global Change Research Program, stepped through their sample process for tracing the lineage of a particular image in the NCA, displayed through three different representations. The complexity of dealing with representations and publications that pull data from multiple sources, especially when not all the data are archived or accessible, can be expected to pose an ongoing challenge for many systems. This measured and methodical approach, firmly grounded in best practices for diagramming the provenance of research data, provides a useful example in delineating the origins of digital objects.

The Cornell University Library used a two-pronged approach to postulate present and future trends for depositing research data in institutional repositories. Wendy Kozlowski et al analyzed usage of eCommons, the Cornell University institutional repository (IR), for existing deposits of items designated as datasets (currently a small fraction of the repository’s holdings)
and conducted interviews with researchers to identify the highest priority data management features and functionality. Their findings were consistent with prior assessments of researchers’ perspectives and include the desire for functions such as citation support, discoverability, versioning, self-service submission and linkages between publications and related datasets. Future developments for eCommons are expected to address those functions that received the highest ranking from researchers. Soliciting input from the users, current and future, to prioritize additional improvement is a commendable approach and an encouraging trend for IR administration.

On the classification and metadata front, the Biodiversity Heritage Library presented a new program using open scientific name data to more completely describe items in their collection. In collaboration with the uBio initiative to develop a comprehensive catalog of biological names, this application of open metadata elements to the digitized books has enabled BHL to organize its collection in a more meaningful manner for biologists and zoologists [1]. It also serves as a useful model for other online collections of digital publications and data to emulate when seeking to incorporate open, collaborative metadata schemas to classify items.

This overview highlights just a handful of the informative and engaging posters on display at the RDAP13 Summit. The variety of projects and research demonstrated in one session bodes well for the advancement of research data access and preservation.

For a complete list of the posters presented at RDAP13 please see the program at www.asis.org/rdap/program/#post.

Slides from a selection of posters (as well as keynote, panel and lightning talk presentations) from the Summit are available on slideshare at www.slideshare.net/asist_org/tag/rdap13.

Resources Mentioned in the Article

Reflections on Our Future
by Douglas C. Engelbart

Editor's Note: Douglas C. Engelbart, computer and technology visionary whose death in early July is noted in Inside ASIS&T, participated on a panel at the 1996 ASIS&T Annual Meeting at which he and other panelists discussed the state-of-the-art of information science at the time in the context of what it portends for the future. Other panelists were moderator Chuck Davis, Indiana University; Candy Schwartz, Simmons College; Gary Marchionini, University of Maryland; Belver Griffith, Drexel University; Clifford Lynch, University of California; and Eugene Garfield, Institute for Scientific Information. The following is an edited transcript of Engelbart's comments that appeared in the Bulletin of the American Society for Information Science, December/January 1997. Note that our organization's name was the American Society for Information Science (ASIS) at the time these remarks were delivered.

Well, thank you. I have an interesting association with ASIS. The very first professional paper I published was with ADI [forerunner of ASIS]. I guess it was about 1959, and related very much what Gary was saying. The theme of it was, “Hey, you librarian people, etc., the producers of your information, that world’s going to change a lot with the computerization of its generation. And the world of your users is going to change a lot. And you can’t sit there in the middle of that thinking you’re going to just be monitoring the same kind of freight and storage as you used to.” It was a fun paper to write, but I think there was one review and it focused on my example of edge-notched cards as a primitive example.

Then in 1969, we were lucky enough to be able to put on for ASIS a live presentation and a video-projection of what we were doing and proposing. On that system we were showing hypermedia interaction, viewing people from our laboratory, their faces would come up and they would be showing what they’re doing and we could interact on the screen. I was trying to give a picture of what we viewed was going to be the future. We’d done that a year earlier at a computer conference, and those were, for us, just big hopes. But, they made a little bit of a blip and then nothing. Similar things in the coming years really began to teach me about the term paradigm. And that is, what is the prevailing way in which people perceive their world and the future. And if that paradigm, which often is narrow enough just to be practical about coping with the days in the world we work, if that paradigm doesn’t expand far enough to take care of the imminent future, then your ability to plan for that future and deal with it is very much inhibited.
Well, digital technology has been changing so fast that the way in which our society has dealt with its evolving paradigms just isn’t up to being able to include even the near future. And now it’s not even being able to include what’s here, technically. So it became clear to me that that was a dominant problem in the world.

Another relationship with this conference that really struck me is the term complexity. What triggered me into taking this very divergent career path in 1951 was realizing the world’s problems are becoming more and more complex and more urgent, and they need dealing with – the serious ones need dealing with collectively. And man’s collective capability to deal with complexity and urgency wasn’t evolving and maturing to keep pace with the complexity and urgency of the problems and the challenges. Helping and contributing in any way to improve that capability was and is something that I could invest my career in. So that’s what I did explicitly. Quitting my job, taking my new bride, going up to Berkeley to study computers, etc. That picture’s been dominating me for all the years since.

And as I experienced more and more, the paradigm issue and problem became clear to me in the 50s. Because I was interested in computers becoming cheaper and faster and more available, I did a study on how the scaling down of electronic components would probably be done. And that introduced me to a world of people studying the impact of scale change on environments. And everyone sort of knows that a scale model of an airplane is not likely to fly, because at different scales physically, things operate differently. The design of you people would be very nice for the five- to six-foot model, but if you were fifty-foot models, your design would just not work. You couldn’t even stand up. If you were the size of a mosquito, you could probably flap your hands and maybe fly. You get surprises when scale changes beyond a certain point, change beyond where your intuition would guess.

So when I looked at this tool-system, human-system thing, I said, “The scale on the computer, the digital stuff, is just going to erupt, and if that happens the adaptation of the human-system side is going to have all kinds of surprises.” And that just stayed with me all this time. Trying to tell the world about these surprises doesn’t do you much good, because trying to tell people the paradigm isn’t right, just doesn’t work. It’s sort of like becoming a political radical. People see you coming and they move or change the subject.

So we were very lucky to get money from ARPA and others, starting in ’63 and ’64, and we actually could build a system. I said, “Hey, one of the surprises that’s very important from my point of view is, if you’re going to be more collectively capable, we want to improve the intellectual capability of harnessing your brain, especially collectively.” So I looked and thought, “What’s that going to be like? And I considered externalizing your concepts. Oh, great! You do it on hard copy, but what’s the advantage you will get in putting it into a computer?” It’ll be easier to get there and the computer structure could actually model the conceptual structures in your mind.

The first thing we built had hyper-linkage, structure, all kinds of optional viewing for once you’re in there moving around and collaborative capabilities. We had all that working in 1970 when I volunteered, because I was in the ARPA community, to be the online information center for the network. So they made me the second node. And what happened is we had that hypertext stuff all available for usage, and hypertext electronic mail, that could interlink, etc. And we could not get any of the other participants in the ARPA community to use it. Everybody just said, “That editor is horribly complex. We [especially the AI people] will make our own editors.” But, it’s more than an editor. By the mid-70s we actually got judged to be going very much in the wrong direction and lost our research money.
The only way to survive with the system we had, which was supporting customers at the time, was to be moved out to the commercial world. The whole organization got sold to McDonnell Douglas, which was building an information systems group. We insisted that we work and use the system ourselves, so our source code is hyper-structured and inter-linkable and on and on. We had our own library system, in which you submit something and it stays available and can be cited from any of the 20 servers around the country, and it has a name of a journal and its journal number, and there’s cataloguing and you’re guaranteed to get back what you authored, at that time. And you could interlink to any object in there. So, it proved by experience an immensely valuable thing. But all of that died because the paradigm couldn’t accept it. If IBM, Hewlett Packard, McDonnell Douglas and DEC weren’t doing it, it’s suspect. It just became clear that only experiences can shift paradigms; it doesn’t do to tell people – you’ve got to give experience.

So we worked up this system that we called bootstrapping, and you guys in ASIS are part of the scene. We at the Bootstrap Institute say the world has one category of people who are operating and another category of activity that’s improving the capability to do that work. So we called the first part the “A” activity and the next part the “B.” The “B” is that which is busy trying to improve how capable you can be at “A.” Because we have significantly more challenges coming, we must get a more effective “B” going to cope with that change. To improve the capability for doing “B,” you obviously have to add a “C” to improve your capability to improve. We already have a lot of “C” activities. You guys are embedded in it right now. Well, the “C” world is usually a community of people operating. Can you fix it so you can augment communities? Of course, that’s what we really pushed in the collaborative, distributed work by being on that ARPA network for some important work. So, what we’re telling people today is, “Look. It’ll be very important for the ‘C’ activity people to say, ‘Hey, we better start doing what we’re preaching.’”

So why isn’t ASIS really busy augmenting its own capability, because its way of working in the future is going to be a lot different. And if it stays the way it is now, pretty soon we’ll have an empty room, because who wants to go to all the trouble of travelling to get together. While once in a while it’s important to have that personal bonding, more and more the importance of what ASIS can do, can be done a lot more effectively if you really get a network going.

We find that we really love going out to organizations. It’s a very different thing approaching and saying, “Let me tell you what you ought to do.” They look at you and say, “Don’t you do heavy knowledge work?” “Oh yeah, but I’m here to tell you ...” “No, why don’t you show me how you’re doing it?” So you as individuals and ASIS as an organization are going to be in that position to go out there and tell the world what they ought to do and they’ll say, “Ha! If ASIS is really going to do that, why isn’t it an example of that different institutional way of working?”

In the Bootstrap Institute we’re really trying to get that word out. We’re saying, essentially, “team work is very different, so let’s try to get an open sort of environment of teams of teams.” And a very important category of that is professional and trade associations. So, we can knock on your door and say, “Why don’t you guys make an interface to this Bootstrapping thing?”

We’re getting some government and corporate deals going, and it’s the challenge. Either that or figure out for yourselves how you’re going to do it. But our kind of bootstrapping proposes having a central thing that societies can belong to that would help teach societies how better to become professional societies by merging your experience and being a society in itself. So that is the central kind of bootstrapping I see.

You guys would be an extremely important segment of that, because you’re a rarity among all of the already existing “C” communities in the nature of what you’re dealing with. And you can shift a little bit away from information toward knowledge as the core. The center of all our future activities will be a dynamic depository of knowledge packages, based in the document. The new document of the future is going to be the knowledge-carrying thing that evolves and directs. And it’s going to need some really careful thinking about what the standard form of that future document is going to be. It’s clear that it’s going to be hyper, and it’s going to need some other things to become a really effective generic, widespread, inter-operable, knowledge package system. So, that’s part of the challenge.