Page one of this issue is a departure from our usual cover article. A theme like research data management is so strongly visual that I felt we could convey more of a story by sharing a glimpse of some of the images within. Research data just may be the hot topic of 2013. Funding agencies are increasingly requiring that it be shared and accessible, and institutions are challenged to meet those requirements. The material itself can run the gamut from data sets to information about the researchers themselves. Libraries are taking a central role in the archiving, management and reuse of this data. To help frame some of the key issues, I suggest you start with David Marques’ article on page 2. From there, you can dive more deeply into insight from individuals and institutions on the leading edge of various aspects of research data management. 

Colleen DeLory
Editor, Library Connect Newsletter

c.delory@elsevier.com
Research data driving new services

By David Marques, SVP Research Data Services, Elsevier, Seattle, WA, USA

Big or small data?

Research data has always been at the core of much scientific research, though the primary conduit of scientific communication has been the peer-reviewed journal article. The article summarizes, synthesizes and interprets the raw data, places the data in the context of theory and hypotheses and mechanisms, and takes a point of view on the data. However, it does not provide sufficient details on the data to facilitate integration within larger data contexts, or alternative synthesis and interpretation.

The era of big data launched with advances in technology power and analytic software, accelerated in part by a seminal conference and book on the future of scientific communication.1 Big data is a fast-growing trend, resulting in great demand for open data programs2 and influential studies highlighting the problems and challenges with the current informal data practices.3,4 In response, the research community has made open publishing of research data a core part of scientific research and communication. Many have argued that the value of the journal article will decrease4 as the value of research data archives increases over the next few years, and recommendations for what needs to be done abound.3,5,6,7,8

The research community is increasingly making raw and summarized research data available for preservation and use by other researchers, both by linking to publications and placing directly into open repositories. At present, researchers in most scientific disciplines — genomics, astronomy and physics are exceptions — make little research data available to other scientists, for reasons including lack of credit, lack of distribution control, and fear that others will point out key insights that they overlooked. Another important factor is a lack of the effort and the informatics expertise required to standardize and normalize the data, and to add sufficient provenance and the descriptive metadata required for domain-specific data repositories.

More researchers participate in loading data sets into repositories that require little metadata or informatics support. These repositories do not integrate data sets on the same topic, are not easily discoverable, and provide little opportunity for analytics. Nevertheless, funding bodies and government organizations are calling for massive increases in the sharing and availability of research data.

Figure 1: Research data flow and services

---

1. Capture
2. Normalize nomenclature and data
3. Add provenance, other metadata
4. Peer review
5. Crawl, index, aggregate
6. Join, analyze
7. Derived data sets
8. Retrieval and linking
9. Compare, contrast, merge
10. Insight, discovery

---

[Figure 1: Research data flow and services]

---

One key difficulty in building sustainable services around research data is that scientific research is built largely on tiny niches of research — a “long tail effect” — with many thousands of small data sets. The problem is better thought of as small data, but the need is just as real.

**Research data flow and services**

Figure 1 illustrates the general flow of research data; this model has been built up from various reports of the research and education community. It also highlights work that could be done to disclose and increase the value of research data.

**The problem is better thought of as small data, but the need is just as real.**

The first step (archive) is required, and a number of self-serve options are already available. All others are optional and discipline-dependent.

These services could include:

1. Deposit in preservation archive and/or register in appropriate database with citable DOIs (with researcher controlling distribution and timing). Where possible, the data should be deposited in discipline-specific data repositories to maximize value to the research community in terms of discovery, analytics, comparisons and pattern detection.
2. Normalize data (align to taxonomies, link to existing entities, scale values to reference models where appropriate); anonymize where needed
3. Perform standard checks for accuracy and typographical errors (e.g., materials properties, where there is evidence of the problem)
4. Annotate with standard descriptive metadata. This critical step makes the data discoverable, and can require effort in the data-capture phase
5. Annotate with provenance metadata and descriptions (very difficult in some disciplines)
6. Obtain peer review where appropriate and requested
7. Track and report downloads, usage and citations to:
   - Scientist
   - Institution
   - Funding agencies
8. Link between data sets and papers
9. Enhance discoverability (mostly nomenclature, but also registrations or search services as needed) such that visualizations, analytics, special reports, and transformations across data sets are possible
10. When appropriate, create solutions for specific tasks, such as diagnostics or design

Elsevier has much to contribute in all of these areas and has a unique depth of insight provided by its activities in publishing (journals, books) and institutional performance reporting (SciVal® suite of products).

Refining key services through pilots

We started this year with several pilots, working with research institutions and data repository owners (e.g., Columbia University, Duke University, Carnegie-Mellon University, and University College London) to increase the flow of shared, open research data. Together we are working out the processes and resources needed to most efficiently share more research data in the target disciplines. These initiatives will only be undertaken in collaboration — never unilaterally — with the academic and research communities. After the pilots end, we will examine funding models to make the process sustainable, and to explore the need for other services, e.g., establishing new databases for data that have no good home today (such as neuro-imaging), or providing analytics based on domain informatics expertise.

The two most important principles for these services are:

1. Data must be open and shared, with distribution controlled by the creator of the data (when possible).
2. The model must be derived in collaboration with the research community and funding agencies, not driven by Elsevier or any publisher.

Publishing research data is a labor-intensive job, requiring special data science, process management, annotation and informatics skills. Institutions and labs will determine the best allocation of resources to meet these needs among their own researchers, librarians and other internal staff, and external services. Organizations at the cusp of establishing a research data management program (increasingly demanded by funding bodies) may want advice on jump-starting and formalizing the program. Those further down the road may be interested in introducing additional expertise, efficiency and rigor.

If you would like to be notified when Elsevier has further news about these research data initiatives, e-mail: ResearchData@elsevier.com LC

**References**

At the University of St Andrews we have had an integrated research information infrastructure since 2006. The overall architecture has remained unchanged, with a current research information system (CRIS) providing tools for managers and researchers to access all research-related institutional data from corporate systems such as Human Resources, Student Records, Research Grants and Finance. In addition the CRIS stores research outputs, outcomes, impacts and activities either via harvesting from third-party sources, such as Scopus and Web of Science, or via manual data entry by researchers.

The technology has been updated over the years, with an in-house CRIS being replaced by the market leader, Pure CRIS, in 2010. From the outset the CRIS has been integrated with our open access (OA) institutional repository running in the DSpace platform. The CRIS is the single, “golden” data source for the research publication metadata and, where a full-text version can be made OA, these metadata are pushed through to the institutional repository (IR) together with the full text. All workflow on copyright clearance and embargo periods is done in the CRIS. Thus the IR acts as a genuine repository of openly accessible documents.

Information management principles

The CRIS-IR is a prime example of the successful practical application of the principles of good information management:

- Data are entered once, as close to source as possible, and reused.
- Data stewards keep control of the data within their domain of expertise.
- Data are available only to those who need it, in the format needed and when needed.
- Data standards such as CERIF and existing data sources such as Web of Knowledge and Scopus are used.

This approach has also resulted in close cooperation across several university functions, principally the Research Policy Office, the Library and IT Services. This strong sense of co-ownership of the CRIS-IR infrastructure has resulted in clear and coordinated communication to researchers about services and tools available and an effective mechanism for gathering feedback to improve these services. It has also ensured a very strong technical and support infrastructure to build on new services, such as research data management.

Figure 1: Existing research information architecture – indicating future link to data sets
TBU Library makes research outputs more visible and more useful

By Ivan Masár, IT Department, Tomas Bata University, Zlín, Czech Republic

In 2011, the library of Tomas Bata University in Zlín began to promote the university’s research outputs. This is a critical activity in the Czech Republic because research outputs — certain articles, conference papers and patents — influence the Ministry of Education’s evaluation of the university and subsequent budget allocation.

One fundamental part of this effort was creating a separate institutional repository used solely to showcase these high-quality outputs. We started by collecting all the past bibliographic records and importing them into the repository built on the DSpace platform with the then-new Discovery interface. The metadata was manually checked and enriched to achieve the highest possible quality.

In the spirit of open access, one of our long-term goals has been to collect as many preprints as possible. In the meantime, we still wanted to make the database useful even without full texts, so one of the first services we integrated was the SFX OpenURL resolver. For the majority of records, this allows our users to find the full text in any of the article databases we have licensed. Next, we plan to eliminate the extra click and include this information directly on the article page.

A list of related articles provided by the bX service is available directly on the item page.

Another feature our users appreciate is a machine-generated citation in the new ČSN ISO 690:2011 format used in the Czech Republic. As a service to our researchers, we include the number of citations from both Scopus and Web of Science, and link directly to the list of citing articles.

To help users decide which journals are worth publishing in, we display their impact factor (and link to a chart of IF development in time), SNIP and SJR metrics.

One of the newer services is usage statistics along with a location of downloads on the world map.

Thanks to the integration of SHERPA/RoMEO, the journals’ open access friendliness is also easily visible. One of the newer services is usage statistics along with a location of downloads on the world map. In the future, we will continue to improve the usability of our repository and stay on the bleeding edge of technology. LC

 babies@k.utb.cz

Extending the CRIS-IR

In the UK, research data management is high on the agenda of funders and, therefore, institutions. A new joint OA policy from the UK’s seven Research Councils (RCUK) goes into effect April 1, 2013. It will not only require the article to be published in an RCUK OA-compliant journal, but also specifies that it “must include ... a statement on how the underlying research materials — such as data, samples or models — can be assessed.”

In addition, data management plans or similar are required for all RCUK grant applications. The Engineering and Physical Sciences Research Council in particular has set a May 1, 2015, deadline for institutions to have the policies, processes, infrastructure and tools in place to satisfy the main principles “that publicly funded research data should generally be made as widely and freely available as possible in a timely and responsible manner; and, secondly, that the research process should not be damaged by the inappropriate release of such data.”

At first these requirements appear contradictory; the Research Councils are working with universities and other stakeholders to provide further clarification. However, whatever the interpretation of the words, the practical consequences are far reaching — not least because universities need to know what research data its researchers generate (who funded it, formats used, quantities, how sensitive it is, where it is, whether it needs to be stored, and so on). The minimum requirement is a catalog of research data sets, and that is where our existing CRIS-IR infrastructure comes in. We can link our research data sets to the information on people, organizations, projects, funding, outputs, impacts and activities that we already have in one place. We are still at the early stages of determining where the research data are and should be stored, how much we need to keep, and for how long — but we are at least in the position to build a catalog of the data within the existing research information environment.

CRIS-IR or IR-CRIS: Who cares?

I am often asked to talk about CRIS-IR and where the responsibility for such systems should lie within an institution — or perhaps more controversially, “Do we need an IR if we have a CRIS?” or vice versa. My answer: Stop thinking about systems and think instead about services. What people, processes, tools and standards are available internally or externally that can best deliver the necessary services now and in the future? At St Andrews we have certainly benefited from this approach and concentrated our efforts into delivering joined-up services to our researchers and research managers, whatever the system. LC

akc@st-andrews.ac.uk

References

1 http://www.atira.dk/en/pure/
2 Common European Research Information Format, see www.euroCRIS.org
3 http://thomsonreuters.com/products_services/science/science__products/a-z/si_web_of_knowledge/
4 http://www.info.sciverse.com/scopus
Paul Albert of Weill Cornell Medical College (WCMC) describes his institution’s experience in implementing VIVO. Albert joined WCMC in 2007 as Digital Services Librarian and is currently the Assistant Director of Research and Digital Services. As the VIVO product manager, he assigns and prioritizes tasks, acquires data, promotes VIVO usage, and communicates with stakeholders.

**Library Connect: Would you tell us about VIVO?**

**Paul Albert:** VIVO is an institution-driven, semantic tool for accessing authoritative data about faculty and researchers. VIVO is open source with its own ontology. When institutions map their data in a semantic way, it enables coordination without cooperation.

Data in VIVO is not necessarily authoritative, but often is ingested from authoritative internal sources such as human resources, grants, demographics, classes, and faculty affairs, or authoritative external sources such as PubMed or NIH RePORTER. When data is updated in those sources, VIVO presents the most current data.

**How did WCMC become involved in VIVO?**

The National Institutes of Health put out a request for proposal for a researcher profile system to reduce redundancies and improve cooperation among researchers. Working in collaboration with our New York City-based Clinical and Translational Science Center (CTSC), WCMC was asked to participate in this grant. Our existing researcher profile system was becoming a bit long in the tooth, so we enthusiastically agreed.

**What role has the WCMC library played in implementing VIVO?**

At WCMC, the library is part of Information Technologies & Services, though in the case of larger institutions, the library may have its own IT department. Libraries have a history of caring about data quality as well as the needs of their users. Because the library can play the role of visionary in how the data is used, it was natural in our case for us to take a leadership role.

**What kind of human resources did it take to implement VIVO?**

At any time, we average the equivalent of approximately three full-time positions working on the project. The most frequent contributors are a full-time programmer and a half-time product manager. Others involved include an information/data architect, principal investigator, project executive (someone who gets the landscape of all the source systems — he’d be able to answer “What system has a middle initial?” in a nuanced way), a server/maintenance person, and a user interface specialist.

**What are your data sources?**

That’s a question best answered visually. The flowchart in Figure 1 depicts our data sources.
In addition to data from WCMC we’re negotiating for access to data from other institutions associated with the CTSC such as Memorial Sloan-Kettering Cancer Center and Hospital for Special Surgery.

Libraries have a history of caring about data quality as well as the needs of their users.

**How has WCMC benefited from implementing VIVO?**

Our top administrators expect to make decisions based on evidence. Because VIVO aggregates data from across institutional silos, we can easily access and parse that data. For example, we provide a monthly report of Weill Cornell-authored articles appearing in journals of a certain ranking, or determine which researchers have published the most in a given list of journals in the last five years.

We aim to provide a 100 percent complete, accurate, and automatically populated list of all a given researcher’s journal articles from PubMed. Toward that end, our end user feedback has been very positive thus far.

Also, we take data from VIVO, particularly publications, and ingest it into a custom Drupal installation, which allows users to perform advanced visualizations.

**What lessons did you learn from implementing VIVO at WCMC?**

As early as possible, think about how you might take advantage of the data in VIVO to address your institution’s real-world needs: generating biosketches, making grant recommendations, reporting on faculty publication patterns, automatically filling out forms. Often, your best customers don’t articulate what they want. There was very little clamouring for iPhones before they came out, because people didn’t know they wanted them! Using your VIVO data creates a feedback loop, which improves its accuracy, not to mention stokes enthusiasm among your key stakeholders.

It makes sense to have someone administering your project, such as a Chief Information Officer or Vice Chancellor of Research, who can ask for data and get a timely response. This is a person who can also allocate the staff resources necessary to tackle deep-rooted data problems. At WCMC, a given faculty member might have a phone number listed in up to seven sources!

You need to be able to sum up VIVO in a compelling elevator pitch such as “authoritative researcher metadata in one place.” Finally, I have found that strength in numbers can be important, e.g., walking into a meeting with three to four colleagues on my side of the table. Make sure to have your colleagues nod vigorously and say, “I totally agree” at key points in the presentation.

**Library Connect – India**

These images reflect only a small portion of the collaborative happenings Elsevier hosts and co-sponsors around the world with our customers and partners. Our goal — to continually reach out to you to improve products and services, garner new ideas to add value to research, teaching and learning processes, and to explore and promote best practices.

Library technology expert Marshall Breeding (seated center) presented on the librarian’s evolving role at a Library Connect event in Bhubaneswar, India, on Feb. 15, 2013.

**University Forum – Sri Lanka**

The University Grants Commission – Sri Lanka and Elsevier partnered to host a University Forum in Colombo on Feb. 28, 2013. More than 40 university librarians, LIS professionals and university leaders from all over the country attended to discuss establishing a digital library in Sri Lanka.
Research data are valuable and ubiquitous. The permanent access to research data is a challenge for all stakeholders in the scientific community. The long-term preservation of and principle of open access to research data offer broad opportunities for the scientific community. Vitaly important, however, is to guarantee that research data sets can be found and cited.

More and more universities and research centers and their scientific communities are starting to build research data repositories allowing permanent access to data sets in a trustworthy environment. Due to disciplinary requirements, the landscape of data repositories is rather heterogeneous. Thus it is sometimes difficult for researchers, funding bodies, publishers and scholarly institutions to select appropriate repositories for archiving, retrieving or citing research data.

Recognizing these challenges and needs, the School of Library and Information Science at Humboldt University Berlin, GFZ German Research Centre for Geosciences – Library and Information Services, and the Library of the Karlsruhe Institute of Technology (KIT) partnered to form re3data.org. The goal of re3data.org is to establish a global registry of research data repositories.

The long-term preservation of and principle of open access to research data offer broad opportunities for the scientific community.

The registry will cover research data repositories from different academic disciplines, and will be a source of information on the permanent storage of and access to data sets for researchers, funding bodies, publishers and scholarly institutions. In the course of this mission, re3data.org aims to promote a culture of sharing, increased access and better visibility of research data. In the first phase of the project, in 2012, the re3data.org partners:

- Conceptualized and implemented a Web-based registry of research data repositories
- Defined selection criteria of research data repositories
- Visualized descriptive criteria using icons
- Developed and implemented the first extensive metadata schema to describe research data repositories

To advance these issues we signed a memorandum of understanding with DataCite in February 2012.

The alpha version of the registry has been online since December 2012. The launch of version 1.0 is planned for spring 2013. Metadata that has been entered from existing sources will be reviewed and complemented if necessary. This will help to ensure the quality of the description of a given repository; it does not make a statement on the quality of the research data contained in the repository.

We have planned the following tasks in 2013 to complete the project:

- Develop interfaces for data import and export
- Support online workflows for repository managers and reviewers
- Foster contacts with repository managers and research data communities

The partners are actively involved in the German Initiative for Network Information (DINI). DINI is committed to improving the information and communication services in higher education institutions by providing the necessary information infrastructures locally and nationally.

The partners intend to establish re3data.org as a reliable service within the emerging infrastructure for research data management. re3data.org is funded by the German Research Foundation DFG.

Further information can be found at: www.re3data.org

References


Library marketing blog

http://libraryconnect.elsevier.com/blogs/

In a recent Publishers Weekly article, Brian Kenney asks if the library brand, once bound to books, is undergoing a transformation. The answer is yes, and has been for awhile. And if that’s the case, then does library marketing need to undergo a transformation as well? As someone who has spent her career in marketing, I am wondering how librarians are feeling about their current marketing efforts.

In 2013, Library Connect will be devoting its blog to the joys and perils of effectively marketing the library. To get the ball rolling, I’ll start with a request: Tell me about the biggest challenge your library is facing in the current year. Then in future blogs, I’ll provide (and recruit other library marketing experts’ input) ideas, suggestions, and solutions to help address those challenges from a marketing perspective.

Andrea Cowan

a.cowan@elsevier.com
Taking care of research data: Outcome of a seminar on managing data integrity in research

By Henk van den Hoogen, Programme Manager Research Support; Strategic Coordinator CRIS/Repository, University Library, Maastricht University, Maastricht, The Netherlands

Last October the Maastricht University Library organized a seminar on data management. The most striking issues were the widely accepted urge to act, the movements in academic conduct, and the lack of support to assist researchers, at least the new ones, in managing their data carefully.

Rob Grim, one of the speakers, outlined several drivers of change in carefully handling research data. In the Netherlands, we have recently had some widely reported fraudulent practices that incited the VSNU (Association of Dutch Universities) and KNAW (Royal Dutch Society of Science) to address the Dutch code of scientific behavior. The KNAW committee Schuyt concluded that new or adjusted rules were not needed, but that raised awareness of the importance of careful and proper handling of research data is essential.

A second driver of change is the important funding stakeholders, like EU and national organizations for scientific research (NWO in the Netherlands) who have explicitly stated that research output that results from their funding should be made available to the world.

Organizations like the royal societies of science (Dutch and British) are also advocates of a (more) open science and a call for action to shift away from a research culture where data is viewed as the private preserve. Therefore changes have to be made not only in the institutional strategy and policies for academic conduct, but also in data infrastructures and intellectual property rights.

When we talk about a code of conduct for data management, we must tackle issues like care, reliability, checkability, impartiality and independence. In addition, Stan van Hoesel, Maastricht’s Associate Dean of Research at the School of Business and Economics, suggested types of data to consider:

- Self-generated data (surveys, experiments)
- Open access data (publicly available data from central bureaus of statistics, government, etc.)
- Data with restricted access (commercially available, confidential data)

Van Hoesel also linked the processing of research data to the principles of a code of conduct:

- Generating data: Careful generation of data plus a well-documented setup are cornerstones of proper data management.
- Data manipulation: For the reliability of data and their sources, data sets must have a careful description of how original data were processed.
- Data analysis: Use statistical methods accurately and be aware of the principle that scientific research must be impartial and independent from organizations that have a special interest in the research outcome (for instance, funding by semi-commercial parties).

A second driver of change is the important funding stakeholders... who have explicitly stated that research output that results from their funding should be made available to the world.

It is also important to identify research data as the most valuable asset for researchers. That is why the School of Business and Economics of Maastricht University is thinking about a semi-open data policy: raw, self-generated data are stored at the faculty level with restricted access and a careful description of the generation process, and made available for 10–15 years after generation (in case of problems). Manipulated data should be stored at a central university level, also well documented and in principle available for colleagues to use.

In more generic terms, scientific research (and the underlying data) should be open in order to:

- Validate research results (accountability)
- Prevent sloppy and fraudulent research
- Facilitate reuse and sharing of data

The seminar taught us also that young, new researchers do not know how to manage their research data. They are eager to receive the proper support in the life cycle for research data (storing, sharing and publishing). Libraries have ample opportunities to support their local research community by offering dedicated research data management support services. See, for instance, the outcome of investigations of the Association of European Research Libraries (LIBER) working group on research data management.

References
1 Seminar “Managing your research data”; Maastricht, 22 October 2013
3 LIBER working group on e-science/research data management: “Ten recommendations for libraries to get started with research data management,” July 2012
Where research goes, so go research libraries and scholarly communication

By Catherine Murray-Rust, Vice Provost for Learning Excellence and Dean of Libraries, Georgia Institute of Technology, Atlanta, Georgia, USA

In early January, an article titled “The End of History Illusion” was published in Science. The researchers, Quoidbach, Gilbert and Wilson, define the end of history illusion as the tendency of humans to underestimate the magnitude of future change.

What are we — those of us whose professional purpose is to provide support for research — doing to anticipate and influence the future of research? How do we overcome the “end of history” illusion and face unwanted possibilities as well as desirable ones?

We can conduct marketing studies, assessments and evaluations with the goal of improving services in the short and medium term. These studies only go so far in encouraging our thinking beyond this budget cycle, this cohort of Ph.D. candidates, this building project. To successfully plan for the future with a longer time horizon, we need to know more about the world of academic disciplines and the behavior of researchers themselves at various stages in their careers.

We know that there is a deep interest among doctoral candidates and early-career faculty in wide dissemination of their research work and the building of their personal “brand” — to use a word that probably would horrify most of them. Although some aspire to be research rock stars and celebrities, most know they will need at least some sort of public presence for their careers and work to thrive. They know that they need more than a page on a departmental website to be successful in the world of global research partnerships and increasing competition for awards. They are beginning to use social media and more sophisticated research profile systems because they need to be visible to attract support.

Universities are anxious to build their research brand as well. They want their faculty to be on fact-finding panels and the major networks, explaining everything from how solar energy works to how music affects the brain.

If more will be required of researchers, what more will be required of us in support of their work? The Association of Research Libraries (ARL) 2030 Scenarios project focused on engaging researchers, as well as librarians and publishers, in a discussion about how their future selves would operate in four research climates. It addresses questions such as: What role will universities have in research? What role, if any, will research libraries play? Can we imagine an end to university research?

The library’s role in implementing MOOCs

By Judith C. Russell, Dean of University Libraries, University of Florida, Gainesville, FL, USA

This article presents one topic, MOOCs, covered within Judith Russell’s presentation at Elsevier’s 2013 Digital Libraries Symposium.

Academic libraries are continually evolving to adapt to changing methods of research, teaching and learning. This evolution includes creating or acquiring tools and resources (primarily digital ones) to serve students and faculty. While this responsibility is not new, the pace of change has accelerated, while resources have been constrained by the current economic situation.

The challenge to the status quo often comes to us (or at us), not from us. The university establishes new programs, expands online and distance-education courses, or forms strategic alliances. A recent example at the University of Florida is its decision to join Coursera, one of the major providers of massive open online courses (MOOCs).

MOOC participants do not need to be a registered student at the university offering the course, and they do not pay tuition or other fees. The requirement for open-source digital textbooks and other supporting materials creates new opportunities for academic libraries to work with faculty. Although these resources and services may be free to the students who are taking MOOCs, they are not without cost to our universities or our libraries.

The University of Florida has already seen substantial enrollment in its first five MOOCs, and is developing additional courses. The libraries are partnering with faculty to assist them with embedded or linked information resources in their MOOC course material. This is such a significant paradigm shift that the faculty have not yet fully understood how their rights and responsibilities differ when they offer these materials to students who are not enrolled at our universities and therefore may not be eligible to access our licensed content.

How often have we sighed (or groaned) at the statement that everything is free on the Internet? Our efficiency at delivering information has made the differences between using paid licensed content or accessing open source or other free content nearly imperceptible to our students and faculty. The analogy to managing this content through electronic course reserves or embedding it in our course management systems where access is limited to enrolled students is certainly clear, but the number of users obtaining access is radically different. Clearly these...
Open access: Current developments, future perspectives

By Alicia Wise, Director of Universal Access, Elsevier, Oxford, UK

Open access (OA) is high on the political agenda not only in the US, but also in many other parts of the world. Many of you are aware of the OA policy recommendations of the Finch Report in the UK. Elsevier supported the results of the report and how it shifted the conversation from the need for an OA policy to when and how a managed transition to OA would be achieved. When I talk about OA, it is often to consider the implications for researchers: choosing where to publish, affording article publishing charges, or considering new policies and mandates. Speaking at the ALA Midwinter Meeting provided a great opportunity to think about these changes from a librarian’s perspective.

Although the cost of the system as a whole remains the same, the way that the cost is shouldered is shifting from reader-side to author-side. Rather than the library carrying the weight of the cost load alone, research funders and research budgets will begin to help. Multiple budgets across funders, campuses and departments may need to be combined. Intensive internal coordination will be required to ensure institutions can enable and support new OA mandates.

With global investment in research increasing at approximately 4 percent per year, the number of researchers employed and articles published increases in kind. Therefore, budgets will remain under pressure.

You may be surprised to read that publishers feel this pressure too. Global investment in R&D fuels more research, which in turn produces more research articles. Elsevier’s submissions increased 9 percent during 2012, although the number of articles accepted for publication grew by only 4 percent. We must do more each year to ensure the quality and integrity of the scientific record, and with our average journal price increases at well below 9 percent, that means annual efficiency gains. Librarians will understand how challenging this is for any organization.

We also continue to make significant investments in our back-office systems to scale our OA services. Authors now have the choice to publish OA in 1,500 hybrid and 30 OA journals. We also have a green OA program.

A question from the Digital Libraries Symposium is representative of the issues now arising: What is Elsevier doing to ensure that OA content in hybrid journals is discoverable by institutions that do not subscribe to that title? My colleagues inform me that they are discussing the issue with an array of vendors to find a solution.

References
Library Connect shares the experience of two librarians in promoting and implementing ORCID at their institutions. With ORCID, short for the Open Researcher and Contributor ID repository, researchers can register online for a unique identifier (ID). This facilitates full and correct attribution of their research activities and outputs.

A community-based effort, ORCID is free for individual researchers to use. Organizations, including funders, research organizations and publishers, may pay a membership fee to achieve greater integration with the ORCID system. Elsevier is a founding member organization and participates on the board of directors. (See Quick Click below for information on importing Scopus data to ORCID via the Scopus2ORCID tool.)

How have you promoted ORCID registration to researchers at your institution?

Linda Galloway, Syracuse University, New York:
I try to include information about ORCID registration every time I talk with faculty (and graduate students) about tracking their citation metrics. I explain that if they use inconsistent naming conventions in their publications, and/or have worked at several institutions, it becomes more difficult to calculate accurate citation counts. I mention a young, prolific faculty member who spelled her first name three different ways early in her career and how difficult it was to find all her publications. I also explain how difficult it is to disambiguate authors who have common names.

I have also publicized ORCID and provided the registration link in the past three monthly newsletters that I send to faculty in my departments (Chemistry, Biology and Forensic Science) here at Syracuse University. The ability to link to funding activities may also appeal to researchers.

In addition, we describe and link to ORCID in our Citation Metrics Research Guide.

Pat Loria, University of Southern Queensland, Australia: ORCID registration will be promoted on our new 2013 library website using the following text:

**Author ID**

Open Researcher and Contributor ID (ORCID) is an open, non-profit organization that maintains an international registry of unique researcher identifiers and a method of linking research activities to those identifiers. You can link to your other identifiers (such as Scopus Author ID or ResearcherID or LinkedIn). And you can include your ORCID ID on your webpage, when you submit publications, apply for grants, and in any research workflow to ensure you get credit for your work. Register for an ORCID ID.

We are actively promoting ORCID to faculty librarians, who are in turn promoting to academic staff members. We are also investigating the possibility of incorporating ORCID IDs into our institutional repository in 2013.

Scopus2ORCID Tool
Help Scopus users in your organization save time with a new tool to help with creating a full ORCID profile by importing an author’s Scopus Author Identifier and associated list of publications.

The Scopus to ORCID wizard demonstrates the steps to follow for finding the correct profile(s) in Scopus and then checking the publications associated with the profile(s). Once each profile has been checked and any necessary corrections made in the wizard, the data can be imported. The ORCID ID for each author will then be updated automatically as Scopus content changes.

Note: Your institution’s authors may have more than one Scopus profile. In such cases, complete the wizard and supply the author’s e-mail address. The Scopus Author Feedback Team will contact the author to help correct the profile.


**Read the TrainingDesk Tip Sheet** for an easy-to-follow guide on creating an ORCID record from a Scopus Author ID. [http://trainingdesk.elsevier.com/scopus-orcid-tipsheet](http://trainingdesk.elsevier.com/scopus-orcid-tipsheet)
The Journal of Academic Librarianship focuses on open access with its January issue

January 2013 articles are available to all in sample issue on ScienceDirect

By Wyoma vanDuinkerken and Wendi Arant Kaspar, Associate Professors, Texas A&M University and Co-Editors of The Journal of Academic Librarianship

With the first issue of 2013, The Journal of Academic Librarianship focused on the open access (OA) debate. As new editors, we hoped that this issue would communicate the diversity of opinions and experiences that the topic merits. Those invited to contribute were not only librarians but publishers, policymakers and academic professionals in a variety of disciplines. The majority of the contributors supported the OA movement, but there was variance in terms of how OA is defined, what it should look like, and the ideal model for implementing and sustaining it.

We achieved part of our goal of bringing JAL readers a balanced view of the OA debate.

OA is a noble goal for scholarship, based upon free information for all, furthering research and scholarship, and dissemination of information. However, some fundamental aspects of OA are often overlooked. A perfect example is provided by one of the contributors. He refers to a compelling term: “digital advantage,” an elegant term for a messy sociopolitical issue. It illuminates a number of assumptions made about OA and scholarly communication and is used to address the dynamics of expanding OA into developing countries.

Another article discussed the concern over universities mandating graduate students to place their Electronic Theses and Dissertations (ETDs) into institutional repositories, losing their copyrights.

We were also fortunate to have Mark J. McCabe, Christopher M. Snyder and Anna Fagin discuss OA vs. traditional journal pricing. David J. Solomon examined the past 20 years of the digital distribution of academic journals and their impact on scholarly communication, while Alicia Wise presented a commercial publisher’s perspective from Elsevier. Several authors contributed to a broad spectrum of perspectives on OA outside the United States, including Saskia Woutersen-Windhouwer, Dehua Hu, Aijing Luo, Haizia Liu, Sarika Sawant, Rajiv Nariani, Sandra Miguel, Paola C. Bongiovani, Nancy D. Gómez, Gema Bueno-de-la-Fuente and Williams E. Nwagwu.

We hope that this JAL issue will continue to cultivate and amplify a thoughtful and open conversation about OA among the scholarly community of libraries, researchers, faculty, publishers, and their representative associations. We encourage letters to the editor about the articles in the OA issue, and we will publish feedback and commentary throughout 2013 to continue the dialogue.

Colleen DeLory << page 12

How do you describe the key benefits of ORCID to faculty and researchers?

Pat: The disambiguation of author identities will improve attribution and reporting on publication outputs, while decreasing staff time in the management of author identity.

Linda: I begin by suggesting that it is important to take control of one’s online scholarly identity. Authors want the correct information attributed to them and should develop a cohesive online presence. ORCID can help link all forms of scholarly output and it seems the promise of linking all this information is a key benefit.

The fact that so many prestigious publishers and institutions are members of the ORCID initiative definitely piques interest and adds authority. Cooperation with other key identifiers is integral to ORCID’s success.

Do researchers register themselves, or does the library play a role in the registration process?

Pat: One faculty librarian has been setting up some ORCID profiles on behalf of her academics. Other faculties have not yet actively started engaging with the registration process.

Linda: At Syracuse, we make faculty and researchers aware of the benefits of ORCID and advise faculty to register themselves.

Are there any barriers to adoption?

Linda: There is some hesitation about yet another tool that requires registration and collects personal information. Unless one really delves deeply into citation analyses, it is difficult to understand the importance of easily and accurately attributing scholarly output.

Are there additional benefits to ORCID participation beyond the individual researcher?

Pat: We are hoping to use ORCID as another source of data for academic research output, as part of an aggregated search, so that we have as full a picture as possible of research output for internal and external reporting.

Linda: Eventually ORCID could help link all types of data associated with a scholar. The code is open source, and third parties are invited to develop applications. Linking more ephemeral forms of output, such as times an article has been tweeted, to a particular author via the code is open source, and third parties are invited to develop applications. Linking more ephemeral forms of output, such as times an article has been tweeted, to a particular author via a unique identifier has promise.

Are you a scholar who uses ORCID?

Pat: I use ORCID for my own scholarly identity, linking my publication outputs to my ORCID record. I believe that linking more ephemeral forms of output, such as times an article has been tweeted, to a particular author via a unique identifier has promise.

Linda: There is some hesitation about yet another tool that requires registration and collects personal information. Unless one really delves deeply into citation analyses, it is difficult to understand the importance of easily and accurately attributing scholarly output.
Data revival: Old data becoming new

By Maliaca Oxnam, Associate Librarian, University of Arizona, Tucson, Arizona, USA

When people talk about data management, they are often referring to the shiny, exciting new data that is promised with a newly funded grant project. Sometimes they mean a nearly completed research project where they are interested in data stewardship; less often they are talking about preserving the data on some form of nearly outdated media requested by the researcher. Rarely, however, are people talking about data sets captured in print as part of the corpus of reports and grey literature on our library shelves.

Advances in technology and in the mass digitization of sources of grey literature, such as government documents that contain data sets, mean that researchers can more easily discover, acquire, extract, reformat and re-analyze data from past experiments. The concept of replication — sharing an experiment’s data and the description of its exact processes so that others can replicate and test the discoveries — has always been at the heart of the scientific method. With the advent of mass digitization efforts, this newly available “hidden” data may be a gold mine for researchers, especially in areas of research where data cannot be easily reproduced or where researchers may want to validate past findings using new techniques and analyses.

The Technical Report Archive & Image Library, also known as TRAIL, contains digitized versions of scientific and technical reports of research performed by and for the US federal government. Many of the more than 42,000 reports available in TRAIL contain data sets that may be useful to researchers today. As researchers discover, extract and reformat this data, one would hope they will be willing to redeposit their new data files for others to use.

In a competitive research market, “data sharing” may sound threatening. While more policies and mandates are enacted to aim to prevent data withholding, maybe a demonstration of the value of data sharing using a body of research for which there is less of a sense of ownership can help to demonstrate the greater scientific value of the whole and help data sharing become more accepted. In other words, as data management as a whole matures, and researchers become more open and willing to share data, these revitalized data will serve as a great test bed for demonstrating how data sharing, data archiving and data stewardship are beneficial to the research process.

Catherine Murray-Rust

programs? Can we imagine an end to research libraries? Can we imagine an end to commercial scholarly publishing?

We should be facilitating and encouraging conversations about the future with the people who will live and work in that future. We have come a long way from the days when librarians and publishers and other communities that support researchers decided what was best without dialogue. We have become far better at listening and learning, especially about short-term and medium-term needs and wants.

What remains is one of the more difficult human challenges: to imagine the future, with the help of formal scenarios or not. In that manner, we can encourage our users and colleagues to avoid the “history stops today” thinking that comes so naturally to all of us.

Judith C. Russell

resources cannot be searched and retrieved through the open Web, so they will not be available for broad public access. It is hard to imagine people signing up for dozens of MOOCs to obtain access to an undetermined (until they are enrolled) set of licensed content, but it is certainly something that will generate discussion and perhaps require adjustment in our licenses.

We recently passed the 150th anniversary of the Morrill Act, the 1862 law signed by Abraham Lincoln that created the land grant university system and made higher education accessible to more people. I think Lincoln would be amazed by MOOCs, and embrace them enthusiastically because of how they provide educational opportunities to people throughout the United States and the world. Let’s hope that a sustainable funding model emerges to deal with the significant costs of developing and delivering “free” MOOCs.

Recent changes to SNIP & SJR metrics

The multidisciplinary landscape of journal evaluation

In recent years, computational advances have contributed to acceleration in the field of bibliometrics. While the journal evaluation landscape was previously somewhat characterized by a scarcity of measures, now many journal metrics are available, providing a varied and more integral picture of journal impact. Librarians may find these useful to compare journals in various systematic ways.

Scopus features two citation indicators to measure a journal’s impact: SNIP (Source Normalised Impact per Paper) and SJR (SCImago Journal Rank). These indicators use the citation data captured in the Scopus database to reveal two aspects of a journal’s impact:

- **SNIP** takes into account the field in which a journal operates, smoothing differences between field-specific properties such as the number of citations per paper, the amount of indexed literature, and the speed of the publication process.
- **SJR** takes into account the prestige of the citing journal: Citations are weighted depending on whether they come from a journal with a high or low SJR.

These two indicators use a three-year window, are freely available on the Web and are calculated for all journals indexed in the Scopus database. The metrics have article type consistency: Only citations to and from scholarly papers are considered.

In October 2012, changes were introduced to both metrics to make them more intuitive and easily understandable. Following these improvements, the values are now computed and released once a year in the summer.

**SNIP: How does it work?**

SNIP was developed by Henk Moed, who was then part of the CWTS bibliometrics group at the University of Leiden. It is a ratio, with a numerator and a denominator. SNIP’s numerator gives a journal’s raw impact per paper (RIP). This is simply the average number of citations received in a particular year by papers published in the journal during the three preceding years.

SNIP’s denominator is the Database Citation Potential (DCP). Because there are large differences in the frequency with which authors cite papers between various scientific subfields, the DCP indicates a journal’s citation potential in the subject field it covers.

SNIP is RIP divided by DCP.

As of October 2012, the following changes apply:

- A different averaging procedure is used to calculate the denominator, to reduce the impact of outliers.
- A correction factor now weights citations from journals with low numbers of references.
- The new calculation results in a SNIP average score for all journals in Scopus to approximately equal one.

“SNIP allows the impact of journals to be compared across fields in a fair way,” comments Ludo Waltman, a researcher at the Centre for Science and Technology Studies of Leiden University, “and has been updated following the most recent insights in the fields of bibliometrics and scientometrics. The recent changes ensure the most balanced treatment of journals from different fields, with minimal implications for users.”

**SJR: SCImago Journal Rank**

SJR was developed by the SCImago research group from the University of Granada, dedicated to information analysis, representation and retrieval by means of visualization techniques. SJR looks at the prestige of a journal, as indicated by considering the sources of citations to it, rather than its popularity as measured simply by counting all citations equally. Each citation received by a journal is assigned a weight based on the SJR of the citing journal.

As of October 2012, the following changes apply:

- A heavier weighting of the more prestigious citations that come from within or closely related fields
- A compensating factor to overcome the decrease of prestige scores over time as the number of journals increases
- A more readily understandable scoring scale with an average of one LC

---

References

Research Data Symposium #RDS2013

Visit Columbia University’s RDS Web page for a variety of resources available from a Research Data Symposium held on Feb. 27, 2013.

Program
The Symposium covered managing and curating research data and a variety of other research outputs via four panels of speakers focused on:

- Plan and collect
- Assure, describe and preserve
- Integrate and analyze
- Discover, share and impact

Discover
The Symposium webcast and links to information on the poster session are available at: http://conferences.cdrs.columbia.edu/rds/index.php/rds/

Symposium Hosts
- Columbia University Libraries/Information Services
- Columbia’s Institute for Data Sciences and Engineering
- Columbia’s Libraries/Information Services’ Center for Digital Research and Scholarship
- Elsevier