The Research Data Alliance: globally co-ordinated action against barriers to data publishing and sharing

Andrew TRELOAR
Research Data Alliance and Australian National Data Service

ABSTRACT. *This article discusses the drivers behind the formation of the Research Data Alliance (RDA), its current state, the lessons learned from its first full year of operation, and its anticipated impact on data publishing and sharing. One of the pressing challenges in data infrastructure (taken here to include issues relating to hardware, software and content format, as well as human actors) is how best to enable data interoperability across boundaries. This is particularly critical as the world deals with bigger and more complex problems that require data and insights from a range of disciplines. The RDA has been set up to enable more data to be shared across barriers to address these challenges. It does this through focused Working Groups and Interest Groups, formed of experts from around the world, and drawing from the academic, industry, and government sectors.*

Andrew Treloar

Context

The amount of activity dealing with the importance of data to research has increased perceptibly over the last five years. This includes conferences specifically focused on research data issues,[1–3] data-focused tracks at discipline conferences (too many to cite), national reports,[4,5] funder requirements,[6–10] special issues of journals,[11,12] and new journals altogether. Those wishing to read further about some of these issues can consult a selective bibliography dealing with publications about this space.[13] The foci for this activity are quite diverse: researcher behaviour, incentives and rewards, changes in the ecology of scholarly communication, technical issues, and the challenges of building and operating data infrastructure.

Role of data infrastructure

This paper will focus specifically on data infrastructure, interpreted broadly: hardware (storage and associated computer hardware), software, content and format standards, and human actors. As the bulk of the data needed by and generated by researchers is increasingly managed electronically, the role of this data infrastructure is becoming critical. One of the pressing challenges in data infrastructure is how best to enable data interoperability across boundaries. These boundaries include those between countries, between disciplines, and between producers of research data and the consumers of those data. A new organization, the Research Data Alliance (RDA), has been brought into existence specifically to address those boundaries from an infrastructure perspective.

Genesis of the Research Data Alliance

Early activity

In 2010 the European Commission initiated the creation of the *Riding the Wave* report.[4] This included a set of recommendations,

which led to a EU–US workshop in Lyon in October 2011. By 2012 two separate activities were underway to address the challenge of how best to co-ordinate infrastructure support for data interoperability. In Europe, the effort was called the Data Access Interoperability Task Force (DAITF). In the US, the National Science Foundation (NSF) together with the National Institute of Standards and Technologies (NIST) was developing a proposal for a DataWeb Forum. These two proposals were then discussed at the International Conference on Research Infrastructure in Copenhagen in March 2012, with a presentation on the DataWeb Forum proposal being made at the DAITF workshop held at that event. Part of the inspiration behind both proposals was the success of the grass-roots Internet Engineering Task Force (IETF) organization[14] and its processes.

individuals found themselves talking about shared concerns

Coming together

Individuals associated with both activities attended an EU–Australian workshop on research data infrastructure in Brussels in June 2012 and found themselves talking about shared concerns over morning tea. At additional conferences and meetings in Copenhagen and Barcelona in October 2012, supported by the NSF and the EU Commission, both the funders and those active at these meetings discussed how to combine the two strands of activity and also how to involve Australia (in large part because of its commitment to research data infrastructure expressed through the funding of the Australian National Data Service21 from 2009 onwards). A series of regular and intense videoconferences in the last quarter of 2012 and the first quarter of 2013 resulted in the creation of the RDA. This involved the development of governance arrangements, an initial website, establishing a regular schedule of meetings during the startup phase, and the arrangements for the First Plenary meeting of the new organization.

Funder commitment

The NSF provided primary funding for RDA/US (RDA members from the US) to support the development and operations of the RDA organization, build the RDA commu-

nity within the US, and provide participant support for key US members to attend RDA events. The NSF also provided funding to a number of additional data projects and infrastructure efforts to support the sharing of data across disciplines and geographic boundaries. NSF and NIST supported the hosting of the Second Plenary.

The EC funded a project called iCORDI (later renamed RDA/EU), commencing in September 2012, to provide similar functions in a European context, and to support the First Plenary.

The Australian Commonwealth government provided additional funding to the Australian National Data Service (ANDS[15]) to take part in RDA activities, and to support the Third Plenary.

At the time of writing (mid-2014) these commitments are all current, but are also finite in length. The RDA is seeking other research funding agencies who might also wish to join the RDA Colloquium (the group of those agencies who are currently contributing to the RDA).

Current state of the Research Data Alliance

Vision, mission and principles

The ultimate RDA vision is a world where researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society. In support of this vision, RDA sees as its mission to work to build the social and technical bridges that enable this open sharing of data.

Membership of the RDA is open to anyone who agrees to support its guiding principles. These are:[16]

- *Openness* – Membership is open to all interested individuals who subscribe to the RDA's Guiding Principles. RDA community meetings and processes are open, and the deliverables of RDA working Groups (see below) will be publicly disseminated.
- *Consensus* – The RDA moves forward by achieving consensus among its membership. RDA processes and procedures include appropriate mechanisms to resolve conflicts.
- *Balance* – The RDA seeks to promote bal-

anced representation of its membership and stakeholder communities.

- *Harmonization* – The RDA works to achieve harmonization across data standards, policies, technologies, infrastructure, and communities.
- *Community-driven* – The RDA is a public, community-driven body constituted of volunteer members and organizations, supported by the RDA Secretariat.
- *Non-profit* – The RDA does not promote, endorse, or sell commercial products, technologies, or services.

The RDA accomplishes its mission through two important mechanisms: Working Groups and Interest Groups.

Working Groups

Working Groups are 'comprised of experts from the community that are engaged in *creating deliverables that will directly enable data sharing, exchange, or interoperability*' [emphasis added]. RDA endorsement is dependent upon the Working Group committing to produce deliverables within an 18-month time frame that will be implemented and adopted by one or more specific communities. Working Group deliverables include, but are not limited to, technical specifications and implementation practices, conceptual models or frameworks, implemented policies, and other documents and practices that improve data exchange. Working groups and their deliverables undergo a 'community review process'.[17] Note that the expectation is that a Working Group will form quickly, tackle an achievable defined piece of work, deliver a solution, get the solution adopted, and then disband. A wide range of Working Groups are currently active.[18]

Interest groups

Interest Groups are 'comprised of experts from the community that are committed to directly or indirectly enabling data sharing, exchange, or interoperability. RDA endorsement is dependent upon the Interest Group *serving as a platform for communication and coordination among individuals, outside and within RDA, with shared interests*' [emphasis added]. They produce important deliverables such as surveys, recommendations, reports, and Working Group case statements. Interest groups must

have international participation and a demonstrated community. They should not be for promoting specific projects or technologies. Interest Groups remain in operation as long as they remain active, subject to periodic evaluation of their activity and its relevance to RDA aims. An interest group that has been inactive for six months will be asked to disband.

With respect to function and outcomes, Interest Groups may do one or more of the following:

1. serve as a platform that leads to the formation of one or more Working Groups;
2. support communication and co-ordination among a cluster of related Working Groups/ Interest Groups that may be grouped by theme (research domain, data publishing, data lifecycle component, etc.);
3. enable better communication and co-ordination across different Working Groups/ Interest Groups (all domain-specific groups, all education groups, between technically oriented and domain-specific groups, etc.);
4. serve to communicate and co-ordinate with a specific community outside the RDA, fostering synergies, bringing new groups/members to the RDA and conversely bringing Working Group activities to the attention of external parties.[17]

Unlike a Working Group, there is no requirement for an Interest Group to disband within a defined time period. A wide range of Interest Groups are also currently active.[19]

Plenaries

Twice a year, in March and September, the RDA community comes together in plenary sessions to meet and work together. Plenaries are not like conventional conferences – the focus is on providing time for Working and Interest Groups to get together. A typical day might involve everyone in a large theatre for the first 90-minute session, followed by a large number of parallel working sessions for the rest of the day. The most recent plenary had between 10 and 12 tracks in each of three blocks of 90 minutes each day. In other words, each day had around 30 individual sessions used by Interest and Working Groups to get together and engage in the sort of high-bandwidth communication that is only possible in person.

Working Groups commit to produce deliverables with 18 months

attendance at the plenaries has been steadily growing

Attendance at the plenaries has been steadily growing. Plenary One in Gothenburg in March 2013 attracted 230 registrations (although attendance was constrained by the size of the venue). Plenary Two in Washington DC in September 2013 had 369 registrations. Plenary Three in Dublin in March 2014 had 503 registrations. Plenary Four in Amsterdam in September 2014 is planning for over 700 registrations.

Governance

RDA activity began in earnest in August 2012 with the establishment of an international Steering Group by funding agencies from the US, EU and Australia. During the startup phase, the Steering Group was charged with developing the RDA by defining its charter and organizational structures, and with promoting its aims and mustering support for its activities. Now that the organization is up and running, the governance arrangements have largely transitioned to a number of bodies with specific responsibilities.

attendees at Plenary Three came from 35 countries

The RDA Council is responsible for the oversight, sustainability, and overall success of the RDA. The Council's responsibilities include approval of candidate interest and Working Groups to ensure alignment with RDA goals. Administration for the RDA is carried out by the Secretariat, led by a Secretary General. The Technical Advisory Board provides technical expertise and advice to the Council. It also assists in developing and reviewing RDA Interest and Working Groups to promote their impact and effectiveness. The Organizational Advisory Board (OAB) provides organizational advice to the RDA Council on the directions, processes, and mechanisms of the RDA.[20]

Lessons learned

So, a little over one year into the evolution of the RDA, what have we learned?

The right thing at the right time

While the progress of the RDA since its official birth in March 2013 is no doubt due in part to those involved in its creation, its success is also due to its timing and to changes underway in the research system. The found-

ing group, responding to this, shared a sense of the importance of research data, the need to better co-ordinate solutions to the barriers to its wider use, and an urgent sense that the time to act was now.

Facilitating conversations

A critical step in reducing the barriers to data interoperability is meeting with others who have the same challenges in data reuse or who are developing solutions that might be applicable. This requires the ability to have effective conversations with others facing the same challenges. A real part of the value of RDA is in its provision of physical (plenaries) and online (the RDA website and its organic group spaces) locations to have these conversations.

Momentum begets momentum

The initial momentum enjoyed by the RDA at its first two plenaries (and all of the associated Interest Group and Working Group activity) has become (in the short term at least) self-perpetuating. People come to RDA plenaries because other people come to RDA plenaries. Presumably, this will eventually be limited by the number of people who see the RDA as useful to them (or by the availability of venues with enough space for 10–20 parallel working sessions).

Providing a nexus

Another benefit of the success of RDA is that it is rapidly becoming the place to meet with colleagues who also work with research data. At the last plenary a number of attendees commented that if all the RDA did was to bring data practitioners together twice a year this would still be a useful function. The value of creating this concentration of people is apparent from the list of collocated events planned around Plenary Four in September 2014.[21] At the time of writing, this numbered ten events, and is likely to rise.

Challenges of globality

Of course, any new organization comes with teething pains, and the RDA is no exception. One of the more significant challenges is turning its global ambitions into reality. While the attendees at Plenary Three came

from 35 countries, the bulk came from the US and Western Europe. The membership of the various governance groups is also over-represented from those regions. And the ability of those who cannot travel to plenary events to contribute effectively to the face-to-face sessions is highly constrained. The RDA leadership recognizes these challenges, and is actively working to address them. But there is still quite a way to go.

Conclusion

Of course, none of this activity (while exciting) will ultimately matter if it does not make a difference. The real test for the RDA will be in whether it actually reduces barriers to data interoperability. This will depend on the outputs of RDA Working Groups being taken up. And this, in turn, depends on RDA Working Groups successfully completing. A number are on track to do so at Plenary Four in September 2014, and this event will feature a showcase of their outputs. The next task is then to promote wider uptake and reduce actual barriers to data sharing.

As indicated in the introduction, the IETF provided inspiration to the founders of the RDA. The unofficial motto of the IETF is often quoted as 'rough consensus and running code'.[14] The equivalent for the RDA might be 'rough consensus and exchanged data'. If it is going to make a difference, this will need to be true. All those involved in the organization at the moment are working hard to bring about this result.

References

1. RDA, Plenary Meetings. Available at https://rd-alliance.org/plenary-meetings.html (accessed 17 June 2014).
2. Digital Curation Centre, International Digital Curation Conference (IDCC). Available at http://www.dcc.ac.uk/events/international-digital-curation-conference-idcc (accessed 17 June 2014).
3. EUDAT Project, EUDAT Conferences. Available at http://www.eudat.eu/conferences (accessed 17 June 2014).
4. Wood, J., Andersson, T., Bachem, A., *et al.* 2011. *Riding the Wave: How Europe Can Gain from the Rising Tide of Scientific Data. Final report of the High Level Expert Group on Scientific Data*, submission to the European Commission (2011). Available at http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf (accessed 17 June 2014).
5. Knowledge Exchange 2011. *A Surfboard for Riding the Wave: Towards a Four Country Action Programme on Research Data*. Available at http://www.knowledge-exchange.info/Default.aspx?ID=469 (accessed 17 June 2014).
6. National Science Foundation, NSF Data Management Plan Requirements. Available at http://www.nsf.gov/eng/general/dmp.jsp (accessed 17 June 2014).
7. National Institutes of Health, NIH Data Sharing Policy. Available at http://grants.nih.gov/grants/policy/data_sharing/ (accessed 17 June 2014).
8. Wellcome Trust, Policy on data management and sharing. Available at http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm (accessed 17 June 2014).
9. European Commission, Guidelines on Data Management in Horizon 2020. Available at http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 17 June 2014).
10. Australian Research Commission, Combined FAQs for Discovery Program Schemes. Available at http://www.arc.gov.au/pdf/DP15/Combined%20FAQs%20for%20Discovery%20Program%20Schemes%202014-15_Version%204_28%20Feb%202014.pdf, sections 3.12 and 3.13 (accessed 17 June 2014).
11. *Nature*, Special issue on data sharing, 9 September 2009. Available online at http://www.nature.com/news/specials/datasharing/index.html (accessed 17 June 2014).
12. *Science Magazine*, Special online collection: Dealing with Data, 11 February 2011. Available online at http://www.sciencemag.org/site/special/data/ (accessed 17 June 2014).
13. Bailey, C., *Research Data Curation Bibliography*, version 4. Available at http://digital-scholarship.org/rdcb/rdcb.htm (accessed 17 June 2014).
14. IETF, *The Tao of IETF: A Novice's Guide to the Internet Engineering Task Force*. Available at http://www.ietf.org/tao.html (accessed 27 June 2014).
15. Australian National Data Service (ANDS). Available at http://ands.org.au/ (accessed 29 June 2014).
16. RDA, Guiding Principles. Available at https://www.rd-alliance.org/get-involved.html (accessed 27 June 2014).
17. RDA, Interest Group and Working Group definition. Available at https://www.rd-alliance.org/group-process-procedures.html (accessed 27 June 2014).
18. RDA, List of Working Groups and links to their organic group areas. Available at https://www.rd-alliance.org/group-type/working-group (accessed 27 June 2014).
19. RDA, List of Working Groups and links to their organic group areas. Available at https://www.rd-alliance.org/interestgroup-list.html (accessed 27 June 2014).
20. RDA, Organisation. Available at https://www.rd-alliance.org/organisation.html (accessed 27 June 2014).
21. RDA, Plenary Four Collocated Events. Available at https://rd-alliance.org/plenary-meetings/fourth-plenary/plenary4-colocated-events.html (accessed 27 June 2014).

Andrew Treloar
Email: andrew.treloar@ands.org.au
Web: andrew.treloar.net

Andrew Treloar is Co-chair of the Research Data Alliance Technical Advisory Board and Director of Technology for the Australian National Data Service, Monash University, Australia.