Against the Grain

Manuscript 8407

Future Proofing IRs with Data and Software

Natasha Simons

Chris Erdmann

Follow this and additional works at: https://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

Future Proofing IRs with Data and Software

by **Natasha Simons** (Associate Director, Australian Research Data Commons) <natasha.simons@ardc.edu.au>

and **Chris Erdmann** (Library Carpentry Community & Development Director, California Digital Library)
<Christopher.Erdmann@ucop.edu>

Disclaimer: The views here represented the authors' personal views and not those of their employers.

Librarians around the world have contributed a huge amount of energy, expertise, time and enthusiasm into running institutional repositories since they were first initiated over 15 years ago (Smith, 2019).[1] A central overarching purpose of these institutional repositories (IRs) has been to capture, describe, make accessible and preserve the research outputs produced at the institution to which the repository belongs. Services have been developed that support the repository to achieve this purpose and the overarching goal of delivering value to the institution and its research community.

Initially, outputs held in the repository were publications, specifically the open access versions of published articles. In more recent years, and in line with changes to research policy and practice, the scope of IRs has expanded to include other research objects such as grey literature (conference papers, unpublished reports etc.), research data and software. However, IRs operate in a competitive landscape that includes professional network and profile systems with deposit features such as ResearchGate as well as discipline and generalist repositories. Unfortunately, IRs are underutilised by the research community compared with these other options. They are often seen as a last resort option, even by the institutions they support. This puts IRs at risk despite the sustainability that is inherent in linkage to the institution. In this article, we raise opportunities for IRs to expand further into data and software in order to stand out in a competitive landscape. We consider the advantages and challenges and make suggestions for future proofing to increase the competitiveness of IRs and ensure sustainability.

Meeting the FAIR Data and Software Challenge

Research data management is a major challenge for universities and other research institutions. The need for effective management and availability of data has been highlighted by research funding agencies, scholarly publishers, discipline communities and government. Published articles are still the prime currency for academic tenure and promotion; however, an increasing number of policies from scholarly publishers require a data availability statement that includes the location and access conditions for the data used to support the article findings. The push for data — and more recently, software — availability from publishers supports credibility, verification and reproducibility of research findings in the published article.

While a large proportion of research data is made available through the supplementary section of journals, there is a push from sections of the research community away from this option and toward deposit in data and software repositories in order to improve curation and support long term preservation and access. For example, the Commitment Statement in the Earth, Space and Environmental Sciences[2] which came out of the AGU Enabling FAIR Data Project (Findability, Accessibility, Interoperability, and Reusability) signed by a cross section of stakeholders including key scholarly publishers explicitly states:

> "Publishers will strive to...Direct all core research outputs (data, software, appropriate samples and sample descriptions) to FAIR-aligned repositories, following the FAIR Principles (Wilkinson et al, 2016)[3] (e.g., using CoreTrustSeal[4] certification). This means that supplements will no longer be used as the primary archive for data."

A variety of repositories are available to support long-term access to research data including discipline, generalist and institutional. These repositories differ in a variety of ways including governance, cost, scope, community engagement and so on. The increasing demand for data to be made available in repositories that are FAIR-aligned presents a real opportunity for IRs to grow and increase in value to the research community. Before looking at ways to do this, it's worthwhile to outline the pros and cons of IRs for data and software deposit.

Advantages and Challenges

When **PLOS** first announced their data deposit policy in 2013, a group of librarians wrote a letter to **PLOS**[5] to raise the fact that they had not mentioned institutional repositories as an option. The group made a strong case for including IRs as an option, citing their strengths including trust within the research community, expertise of library staff, persistence of the repository (and hence the data deposited), assistance and guidance for researchers, and overall sustainability of the IR as a system and a service. They wrote:

> "In summary, academic libraries already play an integral role in your efforts to make data more widely accessible and therefore are valuable additions to the established repositories recommended for archiving data related to publications in your journals. Our request is that you include institutional repositories when providing guidance to researchers on how they might comply with your policy."

This is likely to have been an oversight and **PLOS** responded positively, promptly updating their author guidance to include institutional repositories. However, it is worth noting that IRs were not on the immediate radar and this reflects a deeper issue that many researchers are simply unaware that their institution has an IR. A recent study of Spanish Universities (Borrego, 2017)[6] compared article deposit between IRs and ResearchGate (RG). The authors found that researchers preferred to use RG for their articles primarily because they were unaware of the existence of the IR. At the same time, they were aware of the advantages offered by RG. This is not surprising, given the RG marketing strategy as a universal platform for research communities and as a promotional tool. However, it indicates there is a lot more work to do in raising awareness of the IR in general, whether for article or data deposit.
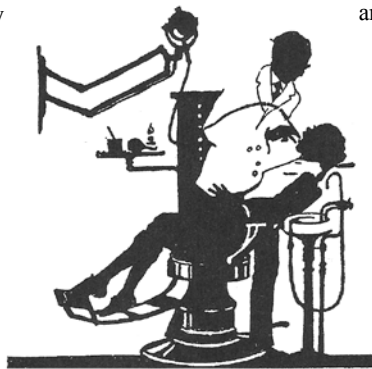
One of the major hurdles in making researchers aware of the existence of the IR is that the IR tends to sit outside of research workflows. This makes it challenging as the IR is effectively hidden from view and relies on library or institutional campaigns to raise awareness of its existence, scope and purpose. Even with awareness, researchers are often unwilling to take the extra time to use an IR because it involves taking additional steps to deposit data outside of existing publication workflows. Contrast this to the ease of submitting data to a discipline or generalist repository with publisher integration, where data is deposited as part of the article submission and review process.

Many institutions lack a policy that specifically states researchers should (or must) deposit their data in the IR. If they do mention the IR, it's often listed as the repository of last resort, following discipline options. While researchers often follow discipline lines and institutional policy is challenging to enforce, mention of the IR in such a policy indicates some level of support for the repository from the institution. Without it, awareness of the existence of the IR may escape senior managers and senior researchers at the institution as well as others who could be making use of the repository as a service.

A 2012 study of the Australasian community (Simons & Richardson, 2012)[7] showed that staffing is a major concern for IRs. Many reposi-

tories are operated on a part-time basis and for the majority, there is no long-term repository sustainability plan (such as would cover budget, technical upgrades and staffing). Staffing and planning are major factors in providing consistency, quality and sustainability of IR services.

The same study showed that technical staffing was also an issue. Often technical upgrades, bug fixes and features are the responsibility of a university IT department. The challenges here include the repository being low on the IT priority list and the specialist knowledge required for the particular IR system. This impacts the ability of the IR to compete on a technical level with other options such as ResearchGate, discipline or generalist repositories.

Implications

Given this (non-exhaustive) list, what would be some good options for IRs to consider to improve uptake and better compete in the research landscape? The following are a few suggestions:

1. Champion FAIR data and software

As mentioned in the letter to **PLOS**, IRs have in their favour that they are trusted. The expertise of the librarians who run them is valuable and a part of this trustworthiness. However, publishers and sections of the research community are looking for a stronger definition of trustworthiness and a way of easily identifying which repositories enable FAIR data and software (e.g., Zenodo Principles).[8] The Core Trust Seal certification — while currently still at low levels of adoption among data repositories — is gaining traction as a way of demonstrating to a repository's users and funders that they have been evaluated by an independent authority and endorsed for their trustworthiness. Publishers are also looking to such certification as a way of guiding authors for data and software deposit. A cohort of data repositories was recently assembled to apply for CoreTrustSeal (CTS) certification as part of the American Geophysical Union Coalition Enabling FAIR Data Project, and other discipline communities are considering following suit. Could IRs consider coming together as a cohort to advance CTS and put themselves clearly on the map as a FAIR data and software deposit option?

2. Participate and leverage Scholix and Make Data Count initiatives

The Scholix initiative[9] comes from the World Data System and the Research Data Alliance, offering a global framework for linking the data and software held in repositories with the articles held by publishers. It presents an opportunity to showcase data held in the IR alongside the article made available through a publishers' portal. Are librarians aware of this initiative, and can further steps be taken to expose data- and software-article links to Scholix so that the data and software in IRs is showcased?

The Make Data Count project[10] is a global initiative to have a standard and fair way to compare metrics for data sharing. Metrics are key to measure research impact and therefore can help demonstrate the value of data deposit. Repositories can contribute to this initiative by submitting COUNTER compliant usage reports of views and downloads to DataCite so that you can see use over time in DataCite search. How can more IRs contribute to MDC?

3. Consider how to spend time

Staff time is one of the most valuable assets to an IR. It may be tempting for staff to spend time convincing individual researchers to deposit their data and software into the repository, but is this time well spent? We suggest it is better to spend time advocating that the repository be included in institutional policy and procedures, raising awareness of the IRs existence and value among senior policy makers and staff, working with IT to develop technical features that make the repository more attractive to researchers, collecting feedback and improvement suggestions from researchers and considering how the repository can be included in institutional system workflows, e.g., mention of the IR at the time of writing a Data Management Plan. It is also worth considering how IR staff can be upskilled in new areas such as FAIR data and software in order to apply these concepts to the IR setting.

4. Engage in long term planning

While a positive feature of IRs is their sustainability, this is achieved because of their link to the institution. Many IRs lack a long-term plan, and spending time developing one would enable the library to conduct a review and analysis of the IR within the broader environment and plan out budget, staffing, advocacy, policy and a features roadmap. This kind of planning is needed to take the IR forward from a last to a first option for the community it supports.

Conclusion

Institutional Repositories operate in a highly competitive environment. They rely on institutional support for funds and staffing which is potentially at risk given low uptake levels by researchers. We have discussed some of the challenges and opportunities for IRs to expand further in data and software which we hope will contribute to a rich discussion that will help IRs future proof themselves and thrive in a competitive environment.

References

Smith, Ina (2015). Open access infrastructure. UNESCO Publishing. p. 7. ISBN 978-92-3-100075-1. *https://unesdoc.unesco.org/ark:/48223/pf0000232204*

Coalition for Publishing Data in Earth and Space Sciences (COPDESS). Enabling FAIR Data Project: Author Guidelines. Retrieved from *http://www.copdess.org/enabling-fair-data-project/author-guidelines/.*

Wilkinson, M. D., **Dumontier, M.**, **Aalbersberg, I. J.**, **Appleton, G.**, **Axton, M.**, **Baak, A.**, ... and **Bouwman, J.** (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3.

CoreTrustSeal. Retrieved from *https://www.coretrustseal.org/.*

(2015, November 2). Open Letter to PLoS – Libraries Role in Data Curation. Retrieved from *https://datacurepublic.wordpress.com/open-letter-to-plos-libraries-role-in-data-curation/.*

Borrego, Á. (2017). Institutional repositories versus ResearchGate: The depositing habits of Spanish researchers. Learned publishing, 30(3), 185-192.

Simons, N., and **Richardson, J.** (2012). New Roles, New Responsibilities: Examining Training Needs of Repository Staff. Journal of Librarianship and Scholarly Communication, 1(2), eP1051. DOI: *http://doi.org/10.7710/2162-3309.1051.*

Zenodo. About: Principles. Retrieved from *http://about.zenodo.org/principles/.*

SCHOLIX. Retrieved from *http://www.scholix.org/.*

Make Data Count. Retrieved from *https://makedatacount.org/*

Endnotes

1. *https://unesdoc.unesco.org/ark:/48223/pf0000232204*
2. *http://www.copdess.org/enabling-fair-data-project/author-guidelines/*
3. *https://www.nature.com/articles/sdata201618*
4. *https://www.coretrustseal.org/*
5. *https://datacurepublic.wordpress.com/open-letter-to-plos-libraries-role-in-data-curation/*
6. *https://doi.org/10.1002/leap.1099*
7. *https://jlsc-pub.org/articles/abstract/10.7710/2162-3309.1051/*
8. *http://about.zenodo.org/principles/*
9. *http://www.scholix.org/*
10. *https://makedatacount.org/*