

Service Integration to Enhance Research Data Management: RSpace Electronic Laboratory Notebook Case Study

Stuart Macdonald
University of Edinburgh

Rory Macneil
Research Space

Abstract

Research Data Management (RDM) provides a framework that supports researchers and their data throughout the course of their research and is increasingly regarded as one of the essential areas of responsible conduct of research. New tools and infrastructures make possible the generation of large volumes of digital research data in a myriad of formats. This facilitates new ways to analyse, share and reuse these outputs, with libraries, IT services and other service units within academic institutions working together with the research community to develop RDM infrastructures to curate and preserve this type of research output and make them re-usable for future generations. Working on the principle that a rationalised and continuous flow of data between systems and across institutional boundaries is one of the core goals of information management, this paper will highlight service integration via Electronic Laboratory Notebooks (ELN), which streamline research data workflows, result in efficiency gains for researchers, research administrators and other stakeholders, and ultimately enhance the RDM process.

Received 16 January 2015 | *Accepted* 10 February 2015

Correspondence should be addressed to Stuart Macdonald, Research and Learning Services, Edinburgh University Main Library, George Square, Edinburgh EH8 9LJ. Email: stuart.macdonald@ed.ac.uk

An earlier version of this paper was presented at the 10th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



Introduction

‘Sound research rests on the ability to evidence, verify and reproduce results – managing your data enables all three’ (Hodson, 2013).

The principle that data generated from publicly-funded research should be openly shared whenever possible has been asserted by OECD (2007), Research Councils UK (n.d.), and more recently in the Royal Society Report ‘Science as an open enterprise’ (2012). This has afforded a pivotal shift towards greater scrutiny of data generated from the research process through coordinated action by funders, with significant responsibilities falling to universities and their researchers, as seen in the EPSRC Policy Framework on Research Data (2011) which obliges research institutions to have support infrastructure for research data management and storage in place by May 2015.

Not all of the drivers for improved management and sharing of research data are ‘top-down’. Jisc’s Managing Research Data programme¹ (2011-2013) piloted RDM services in universities in conjunction with a series of institutional engagement projects undertaken by the Digital Curation Centre to provide tailored support to increase RDM capability. Both initiatives were influential in establishing research data management frameworks in UK universities to support policy makers and university administrators as they adapt to and are accountable for to an ever broader range of digital research outputs and artefacts.

Effecting durable and coherent change across the many distinct units within an institution is indeed a major challenge bearing in mind financial constraints, investment in existing technologies, research intensiveness, domain practice and expertise, with territorial legal and cultural practices notwithstanding. As a result, RDM programmes and subsequent service delivery across institutions are at varying levels of maturity and complexity. Institutions are now having to perform cost-benefit analysis on the internal development of active data infrastructure, storage, data repositories and catalogues versus off-the-shelf solutions whose interoperation depends, to a greater or lesser degree, on the open source or proprietary nature of the product with attendant staffing overheads and commitment.

We are arguably entering a new phase of RDM service development with the need to both consolidate and integrate services to maximise investment. Uptake of those services remains core to RDM support and programme development as they strive to meet researcher requirements without harming institutional research competitiveness. For this to be fully realised institutions may be required to:

- Commit to cross-system resourcing, collaboration, and information flow (both within and between organisations) with stronger ties being forged ‘between libraries, information and computing services’, which will assist innovation and help to make RDM infrastructures sustainable and embedded within academic practice (Macdonald and Martinez, 2010);
- Develop scalable, resilient, adaptable and (where appropriate) interoperable solutions ‘using open standards, interfaces, and definitions to allow new tools and systems to be accommodated when resources allow’ (Menziez et al., 2011);

¹ Managing Research Data:

http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

- Gain buy-in from the primary stakeholders, the local research community themselves.

In order to further maximise aforementioned investments, whilst taking into account current budgetary constraints, new phases of institutional RDM service development may need to extend beyond the ‘traditional’ areas of libraries, information and computing services and harness those service capabilities and opportunities offered by publishers, developers, research instrument providers, and other stakeholders engaged in the research data lifecycle. Service consolidation, integration and tool development is also likely to be further enhanced by the recent Jisc Research Data Spring project,² part of the research at risk co-design challenge area, which aims to find new technical tools, software and service solutions that will improve researchers’ workflows and the use and management of their data.

RDM Service Integration

Since the University of Edinburgh Senate passed the Research Data Management (RDM) Policy³ in May, 2011, Information Services (IS) has been working with colleagues across the University to determine how best to implement the policy. Following the establishment of a Steering Committee and Implementation Group to oversee the rollout of a suite of services effective across all research areas, funding was secured to establish infrastructure for secure storage, management, sharing and preservation of research data in the University. This was articulated by way of an RDM Roadmap⁴ to communicate strategy and milestones for effective planning and implementation of services developed to ‘support researchers and fulfil obligations within a changing national and international setting’ as discussed in detail by Rice et al. (2013).

At time of writing the Research Data Management (RDM) Programme planning and pilot activity, and initial roll-out of primary services have been completed with the following services on offer to the research community:

- **DMPonline**⁵ – an online tool by the Digital Curation Centre that helps researchers to produce a data management plan (DMP) to cater for the whole lifecycle of a project;
- Research Data **MANTRA**⁶ – an online course designed for researchers planning to manage digital data as part of the research process;
- Edinburgh **DataShare**⁷ – the online digital repository of multi-disciplinary research datasets produced at the University of Edinburgh;

2 Jisc Research Data Spring: <http://www.jisc.ac.uk/rd/projects/research-data-spring>

3 University of Edinburgh Research Data Management Policy: <http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy>

4 University of Edinburgh Research Data Management Roadmap: <http://www.ed.ac.uk/schools-departments/information-services/about/strategy-planning/rdm-roadmap>

5 DMPonline: <http://dmponline.dcc.ac.uk>

6 MANTRA: <http://datalib.edina.ac.uk/mantra>

7 Edinburgh DataShare: <http://datashare.is.ed.ac.uk/>

- **DataStore**⁸ – a central facility to store data actively used in current research activities, free at point of use allocation (currently 0.5TB). Researchers can assign up to 50% of their free individual allocation to shared project spaces. Additional capacity can be purchased, with support for very large data (>1PB) hosting available. DataSync, a secure Dropbox-type utility is due to launch early in the 2015 to perform cross-platform synchronisation of data files held on DataStore using the open source file-hosting tool ownCloud⁹.

Work is ongoing with a second phase of work which attends to maturation of those existing services, and their embedding in the consciousness of the research communities through extensive awareness raising and training activities. Upcoming Roadmap milestones will also subsequently tackle requisite interoperation between existing and planned RDM services and external providers as discussed in the following section.

Requirements gathering and consultation exercises have been conducted to inform the design of a Data Asset Registry (DAR), a catalogue of data assets produced by University researchers to aid discovery, access and reuse, the outcome of which recommended that the PURE Current Research Information System dataset type be utilised as the means to record dataset metadata (with data upload functionality disabled). Further investigations will take place to streamline workflow and metadata exchange between the data catalogue and Edinburgh DataShare. A similar requirements gathering activity was conducted to inform the development of a Data Vault, a secure, private and long-term archive or ‘vault’ of data that is only accessible by the creator or their representative. Again, at time of writing the focus of attention is on developing a front-end web application that finesses the functional requirements relating to versioning, authorisation and authentication, deletion and retention, file transfer and system integration, rather than on the back-end storage facility with associated file and storage security functions and integrity checking, which could be developed and managed internally.

According to evidence in a recent report (Hettrick, 2014) by the Software Sustainability Institute (SSI) surveying software generation as part of the research process in Russell Group institutions, 92% of academics use research software, 69% say that their research would not be practical without it, whilst 56% develop their own software (with 21% of those have no training in software development!). The report substantiates anecdotal evidence that software development is prevalent across the research spectrum at the University of Edinburgh and was instrumental in initiating discussions with colleagues from across IS around software preservation. As a result it was agreed that further understanding of the number of local research projects that are creating software is required in addition to tools to measure software uptake/usage in local research, along with the availability of high-level guidance around software development.

Hosted by the Data Library, Edinburgh DataShare was built as an output of the Data Information Specialists Committee-UK (DISC-UK) DataShare project¹⁰, which explored pathways for academics to share their research data over the Internet at the Universities of Edinburgh, Oxford and Southampton (2007-2009). The project was funded as part of the Jisc Repositories and Preservation Programme.

⁸ University of Edinburgh Data Storage and Backup Services: <http://www.ed.ac.uk/schools-departments/information-services/research-support/data-management/data-storage>

⁹ OwnCloud: <http://owncloud.org>

¹⁰ DISC-UK DataShare Project: <http://www.disc-uk.org/datashare.html>

As such, in terms of maturity it has been part of the IS Service portfolio for four years, over which time much work has been done to engage with researchers in terms of encouraging data deposit. Development of its DSpace platform has given rise to interoperability and integration with a range of service providers, tools and registries, e.g.:

- OAI-PMH-compliant metadata records are harvested by the Thomson Reuters Data Citation Index and discoverable through conventional search engines;
- DataCite Digital Object Identifiers (DOIs) are assigned to research datasets in order to provide unique and perpetual identifiers for data, to allow easy citation and discoverability;
- The SWORD protocol (a specialism of the Atom Publishing Protocol) has been harnessed for batch data deposit of large files (size and volume) from remote computers via an API.

The screenshot shows the Edinburgh DataShare website. At the top, there is a navigation bar with the University of Edinburgh logo and links for 'University Homepage', 'IS Homepage', and 'Research Support'. Below this is a 'Contact Us' and 'Login' button. The main content area is titled 'Information Services' and includes a search bar, a 'Deposit Your Data' button, and a list of research communities. A 'Spotlight' section highlights 'NMJ-morph' with confocal microscope images. The 'Latest Items' section lists recent uploads such as 'ARCHER CSE Performance Data and Service Report: Q1 2014'.

Figure 1. Edinburgh DataShare.

The RDM Programme has also been partnering with Research Space¹¹ to integrate its next generation Electronic Laboratory Notebook (ELN) RSpace with the university's emerging RDM infrastructure. This integration is discussed in the following section.

¹¹ RSpace ELN: <http://lab-ally.com/products/rspace-eln/>

RSpace Electronic Laboratory Notebook Case Study

Like Edinburgh's RDM policy, RSpace has been designed to facilitate the implementation of the policies from funding agencies and councils discussed above that encourage open sharing and effective preservation of research data. Developed initially over a two year period (2012 – 2013) in close consultation with a range of representatives from the University of Wisconsin, including researchers, PIs, IT managers, and professionals from research data management and the commercialisation office, RSpace is the first ELN specifically designed to meet the variant needs of a large research university.

Early discussions with Wisconsin confirmed evidence from interactions with other universities that in order to meet the needs of large research universities an ELN would need to appeal to three different constituencies within the institution: researchers, labs/PIs, and administrators – including IT managers, data librarians, commercialisation officers, and research data administrators. The early development of RSpace reflected the information conveyed by colleagues in Wisconsin about the needs of all three communities, but focussed primarily on the needs of researchers and PIs/labs. Useful discussions were had about 'institutional' needs, including making data from RSpace available for a data repository and a data archive, and the high level requirements needed to support these needs, but no development work in these areas was carried out at this stage.

In late 2013, starting with a series of chance discussions rather than from a premeditated plan, it emerged that Edinburgh's RDM plans intersected in promising ways with plans for the further development of RSpace. In particular, an opportunity developed to build out the 'institutional' back end of RSpace in a way that would allow it to integrate with the various strands of RDM at Edinburgh (discussed above), which at that time were becoming clearer and/or beginning to take concrete shape.

Over the course of 2014 the institutional back end of RSpace was duly developed in a way that made it possible to (a) integrate RSpace with the three central pillars of the new RDM infrastructure at Edinburgh (DataStore, DataShare and Data Vault), and (b) build a platform into RSpace on which similar integrations could be carried out at other large research institutions with relatively little customisation. Among other things, a configurable export-to-XML capability was developed, enabling the export of documents, folders, and combinations thereof at the individual researcher and lab level. This provided the platform for subsequent integration with DataShare (described below) and planned integration with Data Vault.

A preparatory step was to integrate RSpace with EASE, Edinburgh's authentication and authorisation service. The scene was then set for RSpace to be integrated with DataShare, which was undertaken as a joint project by Research Space working with Edinburgh University Data Library.

After completion of the integration, once logged in to EASE, RSpace users at Edinburgh are able to deposit data directly into DataShare. They do this via an easy-to-use wizard that takes the researcher through a series of steps whereby they select the data for deposit, fill in metadata in a form (shown in Figure 2) required for deposit into DataShare and submit the form. This results in a deposit to DataShare in exactly the same format as other DataShare deposits. This process at the front end is enabled by backend integration using DataShare's SWORD API, METS description headers, and data from RSpace bundled into XML zip files.

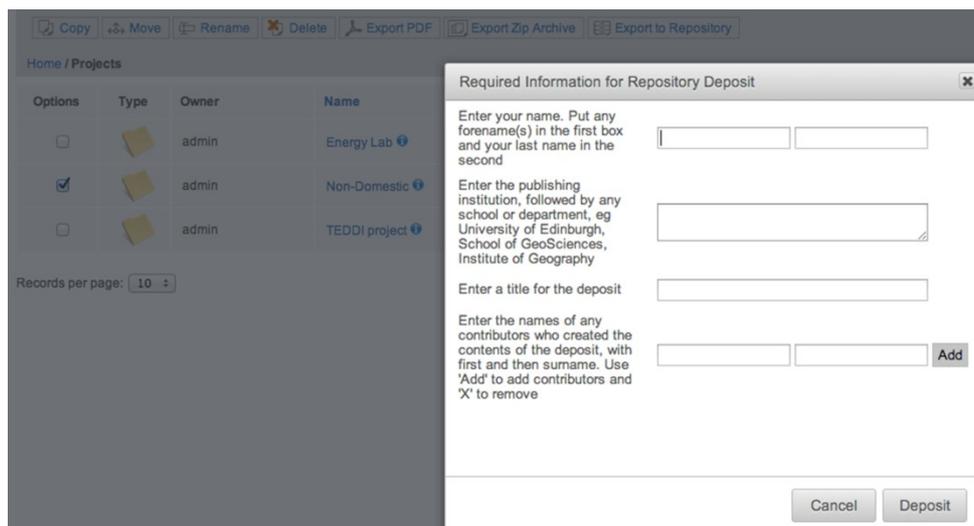


Figure 2. Depositing data from RSpace into DataShare.

As noted above, requirements are still being gathered for Data Vault. The plan is to develop an integration between RSpace and Data Vault that allows deposits in a similar fashion as are currently made into DataShare. It is anticipated that this will involve deposits of data that are in the form of XML zip archives.

Perhaps the most interesting result of the integration of RSpace into Edinburgh's RDM to emerge so far is reflected in researchers' reactions to the virtually simultaneous (albeit coincidental) availability of RSpace and DataStore. After the DataShare integration was completed, RSpace worked with the IT Infrastructure team in Edinburgh University's Information Services on an integration between RSpace and DataStore. This enables researchers to access files in DataStore via a simple two-step procedure. First, they are given the option of locating and selecting for use one or more file stores on DataStore to which they have access.

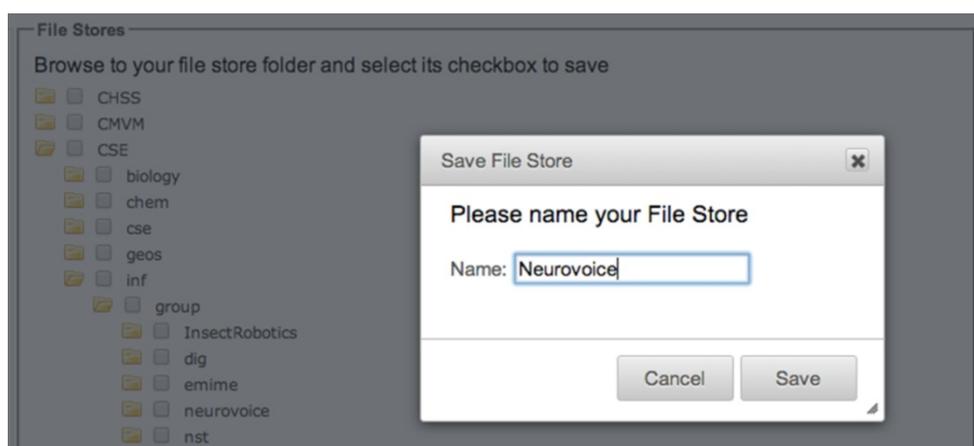


Figure 3. Setting up a file store in RSpace.

These file stores are then exposed in a convenient tree structure in RSpace. Whenever the researcher wants to reference a file in a document in RSpace, they can locate the file in the tree, and by clicking on it a link to the file is created in the document.

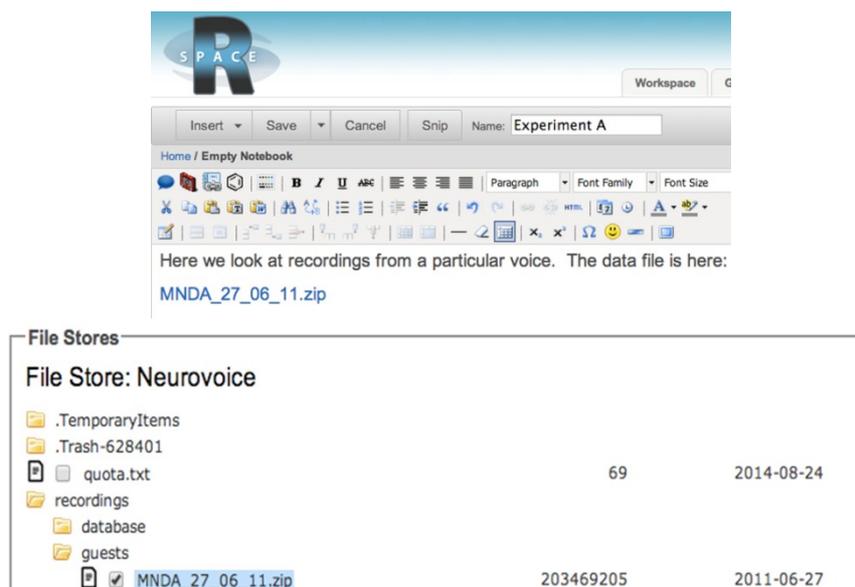


Figure 4. Linking to DataStore files in RSpace.

When an initial, limited trial of RSpace was rolled out to ten labs in November, 2014, researchers from no less than nine of the labs reported that it was the ability to use RSpace in conjunction with the DataShare repository that was of most benefit. One researcher summarised this as follows:

‘My plan for workflow would be generally to deposit my data in DataStore either from the wet lab instruments (gel photos, elisa data, etc, and also possibly directly from an iPad) or from in silico data analysis I’ve been doing, and then link to it from within RSpace.’

The ability to record data in RSpace and conveniently link to files thus appears to be a viable and attractive data management solution for researchers that fits into and enhances their natural workflow.

The integrations with DataStore, DataShare, and in future Data Vault will enable RSpace to serve as a fourth pillar in the RDM infrastructure at Edinburgh, a kind of infrastructural glue that facilitates gathering and sharing of data at each stage of the research data lifecycle: data entry by individual researchers; sharing and analysing of data by groups of researchers (labs), and finally export of data into a repository (DataShare) for publication and public access, and an archive (Data Vault) for long term preservation of the data. This is captured in the Figure 5.

As noted, a small scale initial trial of RSpace was started in November, 2014. The trial will be publicised and rolled out to a larger subset of the Edinburgh research community in early 2015. In conjunction with that, and in anticipation of facilitating wider and streamlined adoption of RSpace, Research Space and Edinburgh plan to develop training resources, in particular a MANTRA training module specifically covering the RSpace ELN.

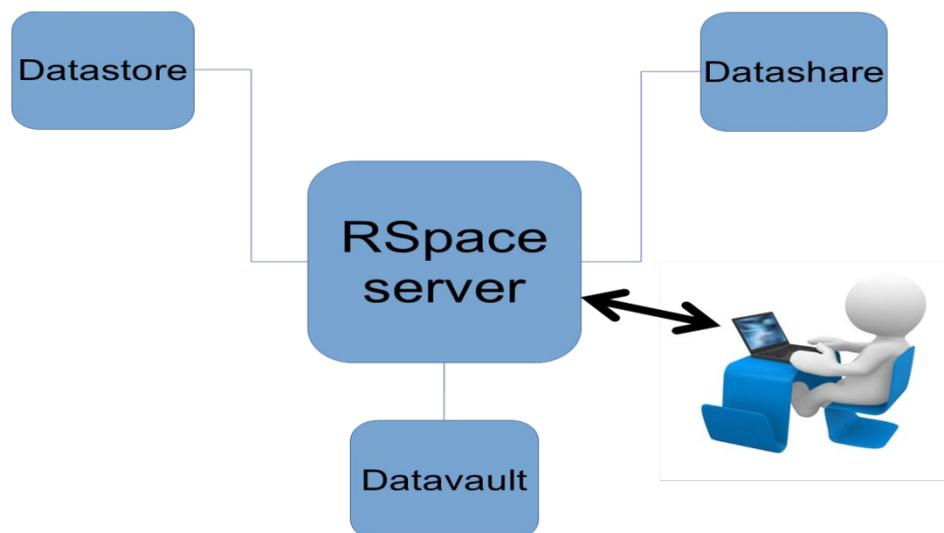


Figure 5. RDM infrastructure.

Conclusions

The University of Edinburgh RDM Programme has invested considerable technological and human effort in establishing and subsequently embedding RDM services into the local research setting. To consolidate the relative successes so far, the RDM Programme is now entering a new developmental phase exploring innovative opportunities to further enhance and streamline the management of research data. This will include investigation of component integration as well as data capture from research instrumentation and software (spectrometer, scanner, ELN and streaming of research output directly into RDM service solutions).

ELN integration with RDM service infrastructure at the University of Edinburgh enables RSpace to serve as an important pillar in RDM at Edinburgh. As infrastructural glue it supports and facilitates the gathering and unified sharing of data at each stage of the research process. This paves the way for new and innovative ways to capture and unify raw data flow from research instruments in laboratories and other data intensive research environments, directly to the scalable service solutions. Integration of this nature and on this scale has the potential to relieve the burden on the researcher in terms of time and effort through seamless transfer of data products. It attends to funder requirements and reinforces both institutional responsibility and commitment in relation to the management of those data assets generated by researchers funded by the public purse. These are indeed exciting developments as we embark on the vision of a cohesive 'cradle to cradle' RDM service ecosystem.

References

EPSRC. (2011). *EPSRC policy framework on research data*. Retrieved from <http://www.epsrc.ac.uk/about/standards/researchdata/>

- Hettrick, S. (2014). It's impossible to conduct research without software, say 7 out of 10 UK researchers [Web log message]. Retrieved from Software Sustainability Institute Blog: <http://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers>
- Hodson, S. (2013). Seven rules of successful research data management in universities [Web log message]. Retrieved from Guardian Blog: <http://www.theguardian.com/higher-education-network/blog/2013/jul/16/research-data-management-top-tips>
- Macdonald, S., & Martinez-Uribe, L. (2010). Collaboration to data curation: Harnessing institutional expertise. *New Review of Academic Librarianship*, 16(S1), 4–16. doi:10.1080/13614533.2010.505823
- Menzies, K., Birrell, D., & Dunsire, G. (2011). An investigation of information systems interoperability in UK universities: Findings and recommendations. *New Review of Information Networking*, 16(2), 92–140. doi:10.1080/13614576.2011.619917
- OECD. (2007). *OECD principles and guidelines for access to research data from public funding*. Retrieved from <http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>
- Research Councils UK. (n.d.). *Research Councils UK common principles on data policy*. Retrieved from <http://www.rcuk.ac.uk/research/datapolicy/>
- Rice, R., Ekmekcioglu, C., Haywood, J., Jones, S., Lewis, S., Macdonald, S., & Weir, T. (2013). Implementing the research data management policy: University of Edinburgh roadmap. *International Journal of Digital Curation*, 8(2), 194–204. doi:10.2218/ijdc.v8i2.283
- Royal Society. (2012). *Science as an open enterprise: Open data for open science*. Retrieved from <https://royalsociety.org/policy/projects/science-public-enterprise/Report>