



13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June 2016,
Scotland, UK

Using a CRIS to support communication of research: mapping the publication cycle to deposit workflows for data and publications

Federica Fina^{a*}, Jackie Proven^a

^aUniversity Library, University of St Andrews, St Andrews, KY16 9TR

Abstract

This paper describes a case study to explore how we continue to develop our CRIS and support the University's research needs and how it has become an embedded tool for researchers to manage their research outputs and to enable Open Access and Open Data.

The paper will show how we used researchers' feedback and comments to develop a simple and easy to remember workflow mapped against existing and familiar research lifecycles. We examine some of the technical, practical and cultural issues we have encountered in implementing these workflows, and show how the CRIS as a single portal has streamlined tasks and reduced duplication of effort.

© 2016 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the Organizing Committee of CRIS2016.

Keywords: CRIS, Pure, Open Access, Research Data Management, workflows

1. Introduction

1.1. Policies scenario

Organisations and funders have a growing expectation of openness by default. Both Open Access to publications and more recently Open Data have an important role in this gradual move to Open Research. Ambitious policies have been developed to build on the opportunities this 'open by default' can deliver: to drive innovation, support sharing and derive new knowledge built on outputs. Funders such as the Wellcome Trust,^{1,2} ERC³ and RCUK^{4,5} stress the need to derive maximum value from funded research, and individual researchers need to consider compliance with funder policies as well as maximising their visibility and impact. To achieve this 'open by default' scenario a cultural change has to be introduced among researchers. Funders and institutions have introduced new expectations and requirements. Organisations must have in place services and resources to allow compliance, inform and guide researchers as mentioned in the first of the nine EPSRC expectations:

*"Research organisations will promote internal awareness of these principles and expectations and ensure that their researchers and research students have a general awareness of the regulatory environment and of the available exemptions which may be used, should the need arise, to justify the withholding of research data"*⁶

Institutions are also expected to manage the research outputs and maintain a publicly available catalogue:

* Corresponding author. Tel.: +44 (0)1334 462343.

E-mail address: ff23@st-andrews.ac.uk

“Research organisations will ensure that appropriately structured metadata describing the research data they hold is published (normally within 12 months of the data being generated) and made freely accessible on the internet;”⁶

An institutional CRIS offers the functionalities that both researchers and support staff need to meet the ‘open-by-default’ agenda.

1.2. CRIS at St Andrews background

The University of St Andrews has been using a research information infrastructure since 2006 starting with an in-house current research information system (CRIS) called Research Expertise, which was then replaced by Pure (Elsevier) in 2010.^{7,8} The CRIS allows managers, support staff and researchers to access institutional data related to research at St Andrews and held in other corporate systems. These data include staff information, student records, and research grants.

The initial focus of the CRIS was that of managing research assessment, but policy and cultural shifts have meant a steady increase in the need to use the CRIS to support the open agenda. This saw the Pure CRIS being integrated with the Open Access (OA) institutional repository (DSpace⁹) in 2010, following a pilot service in 2008 linking Dspace with Research Expertise.^{10,11}

Subsequently, with the introduction of Open Data policies, in 2015 the CRIS became a research data catalogue and a research data repository. During 2014 we explored several options for deposit and storage of data, including extending the Pure/Dspace integration or using Pure as a catalogue and Dspace separately as our data repository. The primary driver was an easy interface to encourage researcher engagement, and the familiarity of Pure combined with its developing data functionality and easier technical implementation led to this choice.

The CRIS at St Andrews is now a single source of information but also a single interface for both administrators and users to meet the open agenda requirements.

1.3. Case study

With the use of a case study the paper explores how we continue to develop the CRIS and support the University’s research needs and how it has become an embedded tool for researchers to manage their research outputs and to enable Open Access and Open Data.

The paper describes the workflows and procedures the Open Access (OA) and Research Data Management (RDM) support teams have put in place in St Andrews to use the CRIS as a single tool for research office, administrators, library and individual researchers to manage information, data sources and storage as efficiently as possible. We examine some of the technical, practical and cultural issues we have encountered in implementing these workflows, and show how the CRIS as a single portal has streamlined tasks and reduced duplication of effort.

2. Questions faced

In recent years the introduction of new and stricter Open Access requirements has brought changes in researchers’ publication practices, for example having to think ahead at grant application stage about the potential cost of Article Processing Charges (APCs) and later on considering the most appropriate journals, correct licenses, additional funder requirements of co-authors; and individual journals’ policies.

The addition of funders’ research data policies meant that researchers had to modify their entire research practices, from estimating data volumes at an early stage of the grant application all the way to the publication of raw data.

These changes, inevitably led researchers to have lot of questions for both OA and RDM teams.

2.1. Open Access

Until very recently OA policies have been primarily associated with funder requirements, with only certain researchers developing awareness and making their work open by default. The introduction of the new OA policy¹² for the next REF – requiring that *“to be eligible for submission to the post-2014 REF, authors’ outputs must have been deposited in an institutional or subject repository”* - has changed this significantly. The focus is now on deposit of accepted manuscripts with the aim of making Open Access a standard part of scholarly communication. We began preapring for the REF/OA policy in 2014, becoming a partner on a Jisc OA Good Practice project (Lessons in Open Access Compliance for Higher Education (LOCH¹³) led by University of Edinburgh). The University implemented the policy in January 2015 with the key message to ‘deposit in Pure’.

While a simple message is easier to communicate, it inevitably leads to questions. Many of these are nuanced and varied across disciplines, and the regular outputs of the OA Good Practice projects combined with our own experience proved invaluable in enabling us to respond and support researchers appropriately. The most common questions are often deceptively simple:

- Why do I have to do this?
- When and how do I upload my article?
- When and how will my manuscript be made public?

The ‘why’ is often practical, for example when a researcher has been depositing preprints in arXiv for years and sees deposit in Pure as an additional process; or it may be a more philosophical question based on a need for evidence of the benefits of OA. Interestingly, as the deadline of 1 April 2016 approached the ‘why’ largely disappeared and the headline became ‘Open access is here: make sure you are ready’. The important questions for our researchers are therefore when and how to go about the deposit process, with additional queries and concerns about how OA will happen, if they need to ‘revisit’ their deposit or complete any further actions.

Our aim is to make this a cohesive process, integrating with their established publishing activities where possible.

2.2. *Research Data Management*

The most common researchers’ questions related to the recently introduced research data policies are:

- Why do we have to do this?
- How and when should I deposit the data?
- Where and when can I obtain a DOI for my data?
- How do I link, in the CRIS, my data to my publication?

After an initial stage of questioning the reasons behind the policies, they have now accepted these new requirements and are trying to understand how and when to deposit data in the CRIS. The step that created and still creates some confusion to researchers is obtaining a unique persistent identifier (DOI = digital object identifier) for the datasets in the CRIS. Some expected the DOI to be automatically issued by the CRIS as soon as a record was created, while others, unaware of all the implications of creating a unique identifier, often expected the RDM team to provide them with one in a matter of minutes. At this point we realised that clear instructions and transparency on the processes behind the scenes were essential.

3. **Workflow**

With the aim of answering the above questions and at the same time producing user guides that clarify the deposit workflows for both publications and research data, we started a review of how we used to communicate these processes. Furthermore, we were asked by a head of department to produce a simple graphical representation of both workflows and to highlight where the two meet. This prompted us at mapping the OA and RDM workflows against the publication process with which researchers are very familiar.

Initially we considered merging the OA and RDM workflows into one but given that the audiences do not always overlap, we deemed it more appropriate to keep the workflows as two separate entities but referenced against the same process for clarity (Figure 1). These allows us to use one of the two or both as required by the occasion.

3.1. *Open Access*

Figure 1 shows the high level mapping of the deposit process against the familiar publishing lifecycle. The OA workflow has two main options for ‘deposit’:

- **Create from template:**
Once a paper has received the final acceptance decision, the author’s accepted manuscript (AAM) is deposited in Pure. A basic record is created in the appropriate template with the date of acceptance and minimal metadata. The AAM is uploaded and given 2 properties: ‘version’ and ‘public access to file’. As an author is logged in to Pure, their name and correct affiliation is populated automatically by the CRIS. They are also able to select from an authority list of journal titles. The CRIS interface provides help tips and prompts, and an indication of mandatory fields.
We have produced a very simple checklist, and a more detailed user guide, to step researchers through this process. Linking of other content such as underlying data and projects is included in the process.
- **Import from external source:**
Instead of manually entering data, the CRIS allows for bibliographic details to be imported from a range of sources. The need to deposit at acceptance reduces the available sources – though we are actively lobbying for metadata to be available from publishers (eg via Crossref) earlier. One main option used primarily in Maths and Physics is to import from arXiv. This has the benefit of building on practice already established, and being very easy to do via the CRIS interface. Researchers are even able to set up notifications so they are prompted to import papers added to arXiv by co-authors. The API allows the pdf to be imported along with the metadata. The functionality of the CRIS allows for matching of all internal authors, recognition of duplicates, and the ability to edit the metadata and file properties before saving to the authors’ profile.

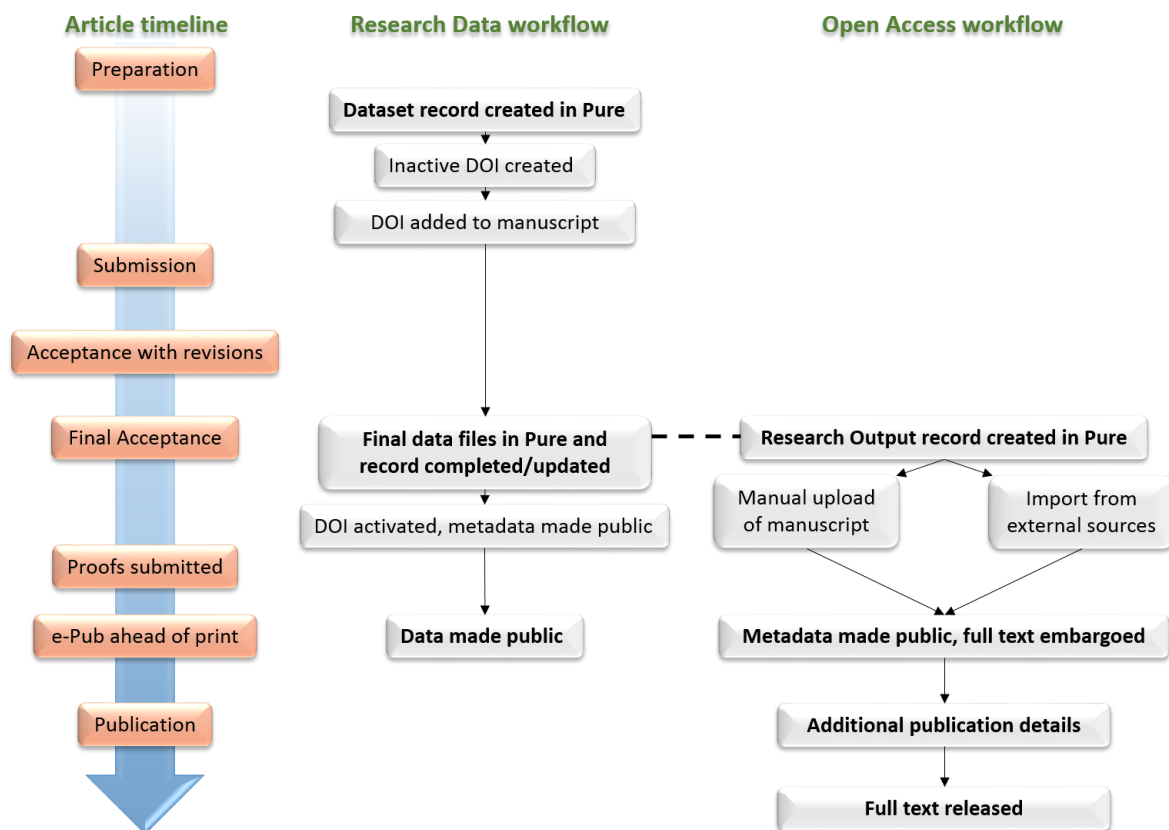


Fig. 1 RDM and OA workflows mapped against the publication process, well-known to researchers. Dotted line: related content link. Article timeline not to scale.

Once the deposit is complete, the user has the ability to set additional properties in order to manage access. Metadata is made public through our Pure Portal by default, but can be temporarily set as ‘backend’ until policies are checked or additional bibliographic details are added. Files can be set as Open, Closed or with an embargo date so that full text is released at a later date. The full record and AAM are only sent to the repository once all elements are set as Open.

3.2. Research Data Management

Differently from publications, for datasets the CRIS only offers manual entry; therefore records must be created from template. The workflow illustrated in Figure 1 was developed mainly considering the timescale and the steps to create a DOI resolving to a URL (minting a DOI), but also external factors that have an effect on the when of each step. In brief, the most important aspect we considered are: funders require authors to include in the publication the persistent identifier pointing to the dataset; most of the journals check that the DOI resolves to a landing page before publishing the article; the timescale between submission of the proofs and publication can be as short as a few hours. We therefore advise authors to create a basic Pure dataset record a few days before submitting the manuscript. At this stage it is enough for the authors to only fill in the mandatory fields. We are then notified by the CRIS that a new record has been created and will issue, by knowing the final syntax, an inactive DOI (coining stage) which the authors can add to the manuscript and proceed with submission to the journal. Once the article is accepted (this can take from a few months to more than one year) and the data are in their final version, authors can upload the files in the CRIS and complete the dataset record. The CRIS does not notify administrators of files being uploaded, so we ask authors to email the RDM team. At this point we can proceed with the validation of the record, which results in the publication of the metadata on the portal, and the activation of the DOI (minting). As shown in Figure 1, we advise authors to make sure they upload the files in the CRIS and the DOI is activated before submitting the proofs to the journal. The reason for this is that some journals put the publication process on hold until the DOI is active and resolves to a landing page.

Until December 2015, minting a DOI was a manual process involving the creation of a metadata xml file and used to take on average 15 min per record. Only recently, with the CRIS upgrade to our current version (Pure 5.4.3), the DOI minting process has become extremely efficient thanks to the integration with DataCite.¹⁴ We can now mint a DOI in a few seconds at the click of a button. Even though the activation of a DOI is now a fast process, authors still need to prepare the data, upload them, complete the record and email us; so we still advise them to do all of this before submitting the final proofs.

Once the article is published online, the data files are made publicly available on the St Andrews research portal.¹⁵

As shown in Figure 1, at final acceptance stage both publication and underpinning data should be in the CRIS, it is at this stage

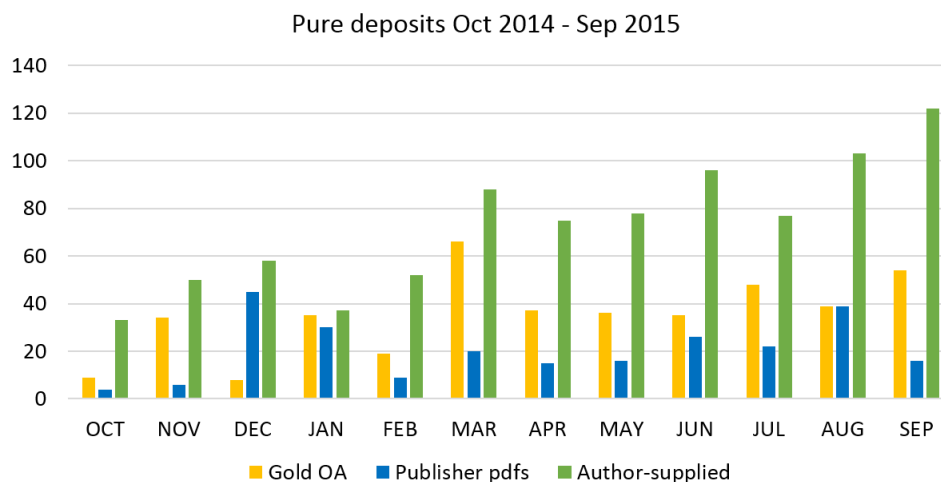


Fig. 2 Number of articles deposited in Pure, with the author-supplied manuscripts in green showing the greatest rise

that the two can be linked. Both dataset and article templates in the CRIS offer the option of creating links between other content. The links will then show on the public facing portal.

4. Roles

4.1. Open Access

The CRIS enables researchers to control and maintain their own research profile, and the expectation is that deposit will normally be an independent process. This encourages researchers to engage fully with Pure so that they become familiar with the system, are responsible for checking all information about their related research activities and ensure their publication lists are complete. However some of the manual entry required is perceived to be onerous for prolific authors, or difficult to remember for those in disciplines where only one or two articles are published a year. While our guides go some way to support these researchers, there is a significant demand for mediated deposit.

Through the work of the LOCH project we developed a model of ‘distributed mediation hubs’. This entails School research administrators being trained by Library staff in the deposit process, and researchers then simply forward their AAMs by email to the relevant School contact for deposit on their behalf. As well as saving researchers’ time, we have seen increased efficiency where local staff have knowledge of specific disciplinary practice, the dynamics of research groups, and the details of grants acknowledged in papers. The linking of related content, including projects with the relevant funder and grant ID, can be supported well through distributed mediation. Initial dialogue such as confirmation of date of acceptance can also take place at a local level, and the number of queries passed on to the central OA team are gradually reduced. The model was piloted successfully in our School of Chemistry, leading to a deposit rate of 90% by early 2016. Similar models of mediation are being rolled out to other departments as demand arises and where local support is available. We do not have sufficient central resource to implement a full mediation service, and it is also recognized that the skills of the OA team in the Library are best used for the validation process – checking copyright policies, verifying bibliographic data, enhancing metadata and applying correct embargoes and licences.

Our flexible approaches across the University resulted in a significant rise in author-driven deposits, measured during the pilot phase in 2015 (Figure 2).

4.2. Research Data Management

As for publications, we encourage researchers to actively engage with the CRIS by creating records and uploading files. In addition to the reasons listed above and differently from articles, the authors’ input is essential in providing meaningful, discipline-specific description of the dataset. Furthermore, when dealing with datasets many variables are involved such as data ownership, commercial agreements, legal and ethical issues, which often only the authors know the details of. Support staff would need to obtain all these information from the authors before proceeding with the deposit, adding extra communication exchange and possibly confusion. For all the above reasons we have not yet implemented ‘mediation hubs’.

While we do not explicitly offer mediated deposit, often researchers seek the RDM team assistance asking to create records and upload files in the CRIS on their behalf. At this initial stage, while researchers are still understanding the policies and adapting their everyday data management practices, we are happy to assist them in depositing their data in Pure. The long term aspiration, however, is to see the majority of researchers to independently deposit their datasets in the CRIS and provide all the discipline-specific information to aid discoverability and re-use.

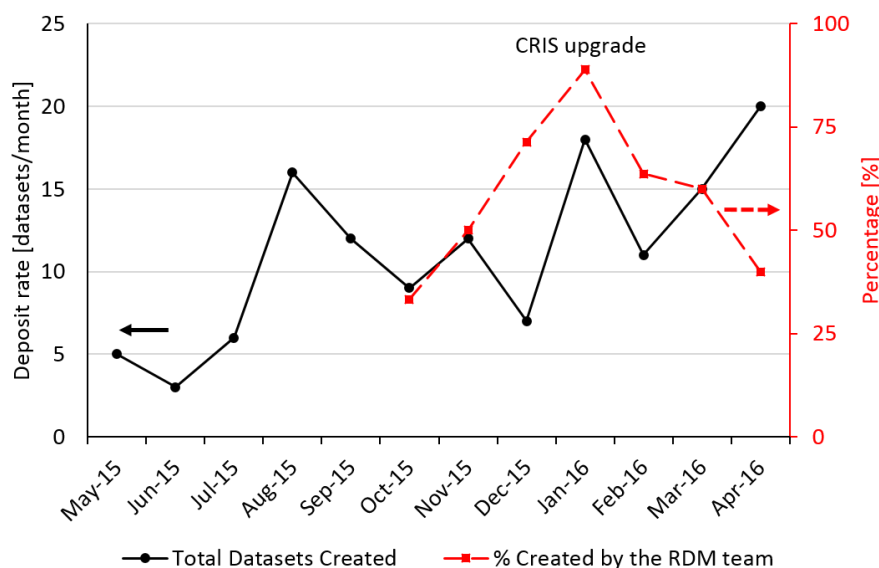


Fig. 3 Number of datasets created over time (black line) and percentage created by the RDM team on behalf of the authors (red line).

We have, therefore, started monitoring the total number of datasets created each month and the percentage of datasets created by the RDM team. Figure 3 shows that the total deposit rate is steadily increasing since May 2015 with 20 datasets created in the CRIS in April 2016. The peak of deposit in August 2015 was a direct consequence of the compliance check we performed on published articles acknowledging EPSRC. In October 2015 we started keeping a record of how many of the datasets were created by the RDM team each month. The percentage saw an initial increase, reaching its maximum of 89% in January 2016. After a series of tutorials on how to deposit datasets in Pure and the dissemination of an updated user guide, the value has been constantly decreasing. In April 2016 only 40% of the datasets were created by the team on behalf of the authors. These values indicate that the workflow mapped against the publication process and the advocacy strategy are indeed effective.

5. Joint advocacy approach

Even though we developed two independent workflows on how to deposit in the CRIS, we have modified our face-to-face advocacy approach. We have started delivering joint OA and RDM Pure tutorials when possible. The aim of joint events is to communicate a complete overview of the deposit process and to highlight overlaps and similarities between the two workflows. Furthermore, having representatives of both teams present at an event offers the advantage of being able to address a wide variety of questions.

In addition, we hold joint drop-in sessions in schools and departments. These have so far taken two different forms, coffee breaks where members of both teams join academics during their breaks; or traditional format where the teams book a room in the school and anyone with questions can just drop-in.

We have also developed close collaboration between support teams which has a positive effect on our enquiry workflow and in raising awareness. Often authors send a single email with queries about Open Access and Open Data and responses are copied to both teams so that answers are provided in one thread and the communication kept simple and effective. Taking inspiration from authors' practices of a single email we are testing a new awareness raising strategy. The dataset deposit process starts earlier than that of publications, therefore authors are likely to be already in communication with the RDM team. At manuscript acceptance stage, when authors ask for the activation of the DOI (Figure 1), it is normal practice for the RDM team to link the dataset template to other content such as projects and publications in the CRIS; at this point, if no record exists for the relative publications, a reminder is then sent to the author and the OA team is copied in the communication. In a similar fashion, the OA team send a gentle reminder to the authors, copying the RDM team, if during their validation activity they find no underpinning datasets.

6. Conclusions

Having mapped both RDM and OA workflows to existing publication practices makes the process clearer to researchers and this can be noticed by the increased deposit rates for both publications and datasets.

Joint advocacy events allow us to better communicate messages and clarify processes. But the benefits are not all for researchers, our joint approaches to support have also resulted in increased knowledge of the processes and requirements between teams, developing our skills and experience.

Exploiting the CRIS functionality for linking content types, we are able to support a cohesive workflow, and enable end-users to discover related content such as all publications linked to a grant, activities related by author or journal, or all datasets relating

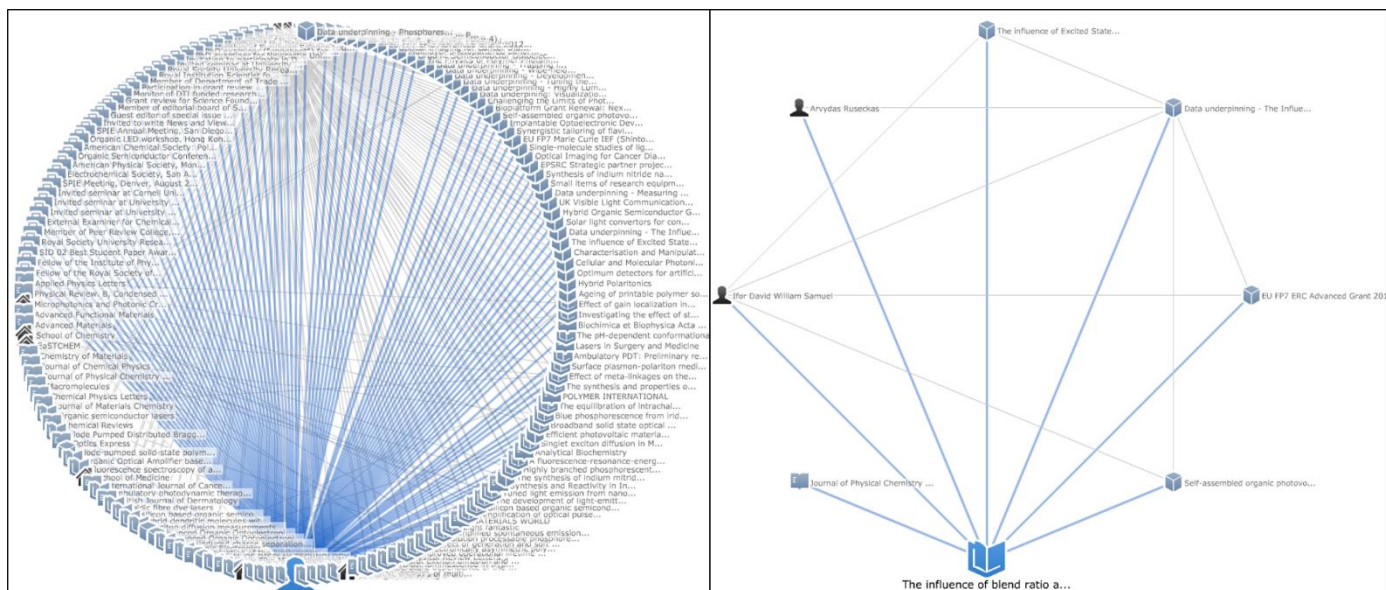


Fig. 4 Pure portal graph of relation for a) an author and b) a publication.

to a research centre. Figure 4a shows an example of all research activities linked to a person and Figure 4b shows a sample set of relations for a specific publication.

Acknowledgements

We would like to thank the OA and RDM teams for the help in developing workflows and we also thank Jisc for funding the OAGP Pathfinder LOCH project and our project partners University of Edinburgh and Heriot Watt University.

References

1. Wellcome Trust Open Access Policy - <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm>
2. Wellcome Trust Policy on Data Management and Sharing - <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>
3. EC Open Access - <http://ec.europa.eu/research/openscience/index.cfm?pg=openaccess>
4. RCUK Policy on Open Access - <http://www.rcuk.ac.uk/documents/documents/rcukopenaccesspolicy-pdf/>
5. RCUK Common Principles on Data Policy - RCUK Common Principles on Data Policy
6. EPSRC, Clarifications of EPSRC expectations on research data management - <https://www.epsrc.ac.uk/about/standards/researchdata/expectations/>
7. Clements A, McCutcheon V. Research data meets research information management: Two case studies using (a) Pure CERIF-CRIS and (b) EPrints repository platform with CERIF extensions. *Procedia Computer Science* 2014;**33**:199-206
8. Pure - <https://www.elsevier.com/solutions/pure>
9. DSpace Repository software. <http://www.dspace.org/>
10. Proven, J. and Aucock, J. 2011. Increasing uptake at St Andrews: Strategies for developing the research repository. *ALISS Quarterly* 6(3): pp6-9 <http://hdl.handle.net/10023/1824>
11. Poster proposal for OPEN REPOSITORIES 2008 from St Andrews University LIS: Project to integrate research publications deposit through a centralised deposit workflow: optimising the relationship and functionality of the St Andrews Research Expertise Database and the St Andrews Digital Research Repository. <http://hdl.handle.net/10023/467>
12. Open access and research assessment policy - <http://www.hefce.ac.uk/rrsch/oa/>
13. LOCH project - <http://libraryblogs.is.ed.ac.uk/loch/>
14. DataCite - <https://www.datacite.org/>
15. St Andrews research portal - <https://risweb.st-andrews.ac.uk/portal/>