Activities, solutions and experiences within UK universities to meet the EPSRC research data policy

Monica Duke, DCC, Jonathan Rans, DCC
Verena Weigert, Jisc
University of St Andrews

Interview with Anna Clements, assistant director digital research

12 January 2015

Interview led and recorded by Jonathan Rans, DCC and Verena Weigert, Jisc
Introduction to the University of St Andrews

St Andrews is a public research university. Founded 600 years ago, it is the oldest of Scotland’s four ancient universities. It is divided into four faculties, containing 18 schools. It employs an academic staff of around 1,150 and the student body includes around 1,600 research and taught postgraduates. Anna Clements originally worked in the IT department but moved to the library at the end of 2013. She works closely with the vice-principal for research to drive St Andrews RDM support programme.

Policy, strategy, governance and sustainability

EPSRC roadmap development

The progress that the University of St Andrews has made over the past two years in the development and implementation of its RDM policy and strategy has been strongly catalysed by the EPSRC's expectations.

St Andrews' high-level RDM advisory group initially convened as the roadmap development group. It has a cross-departmental membership chaired by the VP research and includes the chief information officer (CIO), dean of science, representatives from the research policy office and members of the relevant IT teams. This group has undergone a number of transformations in response to changing needs and conditions but has always retained strong representation from senior academics. Because of ongoing reorganisation within the library and IT services the group has not met regularly but is expected to play a central role in the future implementation of RDM strategy.

The roadmap document itself, completed for the May 2012 deadline, is not publicly available but has been strongly influenced by Monash University’s RDM strategy and takes from that a clear emphasis on the roles and responsibilities of both the institution and its academics.

When planning the practical implementation of the roadmap, securing the active engagement of all academic staff was considered of paramount importance, and not just of those that were in receipt of EPSRC funding. To this end, an academic-led review of the potential roadmap activities was led by Professor Simon Dobson from the department of computer science and included representatives from science, medicine, social science and arts and humanities. This consultation exercise aimed to identify the services that researchers need to support their research, irrespective of the compliance issues. These requirements were then incorporated into the framework of a roadmap that responded to the EPSRC’s requirements.

This approach enabled the institution to explore opportunities at the academic level and tie the new RDM activities into other strands of work, particularly those that addressed how St Andrews could respond to the challenges of data-intensive research. In September 2014 part of this response was realised with the establishment of a new institute for data intensive research. It will bring together researchers from all four faculties with an interest in big and/or complex data research and will be expected to contribute expert advice to the ongoing development of data management services. In parallel, there are moves to establish a division of digital research within the library, which would also be expected to play a pivotal role in the ongoing programme of RDM support implementation at St Andrews.

1 St Andrews institute for data-intensive research: idir.st-andrews.ac.uk/about/
An organisational upshot of the academic review was the decision to assign the lead on RDM implementation to the library although the team there will be working closely with colleagues from IT, the research policy office and senior academic staff across the university.

**Policy development**

It is only relatively recently that St Andrews has published an RDM policy² to sit alongside its strategy document. This was drafted by Professor Dobson, Anna Clements and the VP research and was circulated to the research forum³ to gather feedback from academic stakeholders. The forum includes all research directors and others who hold responsibility for research strategy development. After incorporating the amendments suggested by the forum, the policy was ratified by the academic council⁴, the formal policy body at St Andrews, and then published in June 2014. In terms of content and structure it was very useful to have existing policy documents from other institutions to use as examples and the final version is strongly influenced by the University of Edinburgh’s policy⁵, with modifications to reflect the local conditions.

There is also a broad review of other policies relating to the governance of data access and reuse in progress, particularly in the area of intellectual property rights (IPR).

EPSRC expectation I requires that researchers are made aware of all policies with a bearing on the management and access of their data. The primary method for connecting academics to this information is via St Andrews’ RDM web portal, which contains links to all data-related policies including the RDM policy and those dealing with IPR, ethics and legal issues. This will soon be supplemented by a programme of face-to-face advocacy and awareness-raising events.

Anna has also developed a locally customised version of the DCC’s DMPonline, which academics are encouraged to use when developing data management plans (DMPs) for grant submission. The St Andrews templates include tailored guidance, links to relevant policies and contact details for colleagues who can provide specialist assistance. For EPSRC researchers there is no mandate to submit a DMP or technical appendix with their grant applications, although there is an expectation that such a document will be produced by the researcher. St Andrews researchers bidding for EPSRC grants are encouraged to use DMPonline, which contains a specially designed EPSRC template but, without the explicit drive from the research council, the tool has not gained as much traction with them as it has for other funders.

---

² St Andrews RDM policy: [st-andrews.ac.uk/staff/policy/research/researchdata/](http://st-andrews.ac.uk/staff/policy/research/researchdata/)
³ St Andrews Research and teaching staff forum: [st-andrews.ac.uk/staff/research/policies/researchstaffforum/](http://st-andrews.ac.uk/staff/research/policies/researchstaffforum/)
⁴ St Andrews Academic council: [st-andrews.ac.uk/about/senate/academiccouncil/](http://st-andrews.ac.uk/about/senate/academiccouncil/)
⁵ Edinburgh RDM policy: [ed.ac.uk/schools-departments/informationservices/about/policies-and-regulations/research-data-policy](http://ed.ac.uk/schools-departments/informationservices/about/policies-and-regulations/research-data-policy)
Resources

St Andrews understands that it needs to invest in its IT infrastructure both as a response to funder requirements and to protect its research portfolio, particularly in support of the storage of active data. However, it has proved difficult to make the business case without being able to define categorically the volume of data that needs to be accommodated. In the current climate there is little appetite for making a large-scale infrastructure investment without very concrete evidence.

Some information was collected by the research computing team in 2013 and additionally by the academic-led review but the utility of the results for building a convincing business case for storage investment was affected by an uneven response across the faculties. Even with the responses that were received, a fairly large margin for error needs to be factored in as these are figures which researchers typically find hard to estimate accurately, particularly when asked how much data they would expect to need to archive long-term. In terms of human infrastructure, there is an ongoing reorganisation in place to consolidate RDM support into the library within the new digital research division.

Training and guidance

This reorganisation is part of a larger process of organisational change that has had the effect of slowing down the amount of face-to-face advocacy and awareness-raising that could be delivered. However, whilst waiting for the staff who will deliver the training to be put in place, Anna has been able to identify needs and develop materials. The academic review of RDM strategy has been very useful for providing the researchers’ perspective - as expected, there is a preoccupation with the management of active data over the questions of long-term preservation.

The library has a series of training events planned for delivery in March 2015 and these will cover all areas of research support. There are modules on bibliometrics/alt-metrics and OA as well as RDM. These workshops are designed to be accessible to research staff and will be relatively short (around an hour long) in an effort to make them more attractive to those with little time to devote to training.

Centrally run courses for researchers with an interest in RDM are essential, of course, but driving culture change requires that researchers are engaged with the issues across the board. A good place to start is with early career researchers, who already have a formalised programme of training in fundamental aspects of research. At St Andrews this has been exploited by including a module on RDM in the graduate skills programme; it directs students towards the University of Edinburgh’s online MANTRA training materials, available for reuse under the Creative Commons licence. The next step towards cementing RDM as part of early career researcher training will be to incorporate MANTRA materials into the postdoctoral training programme.

Good though these courses are, there is still a need to supplement them with face-to-face advocacy and guidance through a programme of school-level engagement. Given the current restrictions on resource for proactive engagement with researchers, the initial method for meeting the EPSRC’s requirement for awareness

---

6 Plietzsch, Birgit (2013), How much (more) research data do we have, and where do we store it? 18 March 2013, St Andrews Research Computing Blog, retrieved 1 February 2015 from http://research-computing.wp.st-andrews.ac.uk/2013/03/18/how-much-research-data-do-we-have-and-where-do-we-store-it/
7 Mantra RDM training resources: http://datalib.edina.ac.uk/mantra/
raising (expectation I) has been via the institutional RDM web pages⁸, supplemented by regular updates to the research forum. The pages contain links to a variety of policies that have a bearing on RDM and have contact details to put researchers in touch with relevant expertise around the institution. The pages have recently been updated to include more comprehensive guidance and also disciplinary case studies that aim to show the benefits to research of good RDM practice. The university recognises some researchers remain to be convinced that this is something they need to engage with over and above the minimum that is required for funder compliance.

**Infrastructure**

St. Andrews RDM infrastructure visualisation – implementation in progress

Image credited to Prof Simon Dobson and Anna Clements

---

⁸ St Andrews RDM web pages: [st-andrews.ac.uk/staff/research/data/]
Preservation and storage

At the moment schools provide locally maintained data storage but this is in need of refreshment, which offers an excellent opportunity to move towards central storage in a managed way. Doing so will address some aspects of EPSRC expectation VIII, which requires that the institution provides support for data across the full RDM lifecycle.

There is clearly a need for investment in this area, even if the scale is hard to quantify. Researchers have demonstrated some interest in infrastructure that supports the management of active data, and they are looking for storage and services that are easily accessible and flexible. They have less interest in the long-term preservation of data but, by bundling that in with other services, there is a better chance of it gaining traction.

The VP research has recently approved a policy to provide each PI with 0.5TB centrally managed, resilient and secure research data storage, with a costed model for storage over and above this quota.

Responding to the EPSRC’s requirement for long-term data preservation (expectation VII) and the need for mechanisms to make appropriate datasets open access, St Andrews is establishing an institutional data repository. This will be based on the PURE platform, a decision strongly influenced by the fact that the software has already been selected as the basis of institutional CRIS and data catalogue systems. The current publications repository is an instance of PURE linked to DSpace, so a similar set-up was considered for the data repository. However, after further discussion, there did not appear to be any considerable advantage to using a linked version of DSpace and PURE over a standalone instance of PURE to handle data. Both platforms are known to have a number of issues when asked to deal with big or complex datasets and it was felt that DSpace was not able to enhance this provision sufficiently to justify an extra layer of infrastructure. So big data still remains a concern with the institutional repository but this is considered to be an acceptable risk given that data-intensive subjects are more likely to have their own subject repository or datacentre. St Andrews’ approach is that PURE will be a repository for those researchers without access to more appropriate disciplinary datacentres. To support this strategy, there is a need for guidance to lead researchers to the most appropriate repository and to ensure that they retain and deposit a metadata record in the institutional data catalogue for datasets deposited elsewhere.

The catalogue itself, which will allow St Andrews to fulfil EPSRC expectation V, is also based on the PURE platform. Both the repository and catalogue are in the test phase at the moment and we expect them to be launched in the early part of 2015. The back end of the repository is currently being supported by local storage but in future may integrate with the Arkivum system. This is all currently under review.

Deposit workflows

At this early stage, workflows and infrastructure are not in place to drive the automatic deposit of datasets when they reach the appropriate stage of the lifecycle. This is a key area for researchers and Anna is working with colleagues in computer science and chemistry to investigate the automatic ingest of data and metadata from existing facilities.

PURE includes metadata on all awards, including end dates and can be used to provide reports to identify candidates for data deposit. To make this process more manageable, the focus is on data that underpins published outputs which also allows existing systems governing open access to be exploited as a flag to indicate when there should be data due for deposit.
In terms of the metadata deposit, PURE has always accepted metadata on datasets and there are currently 18 publicly available metadata records at the Research@StAndrews portal⁹. There is already guidance in place for researchers on how to deposit research data metadata into PURE and link to journal articles¹⁰. The new version of the software moves datasets into its own top-level entity and includes enhanced metadata, workflow and storage functionality as well as support for creating and storing Datacite DOIs. Datasets, as with all entities in PURE, can be linked to any other entity within PURE, including projects, articles, activities and impact.

St Andrews is in the process of enhancing its existing Research@StAndrews public-facing portal to allow selected elements of the metadata to be exposed and this will certainly include those links to other information and outputs – this is something that other universities are already doing with their catalogues¹¹.

The question of tracking access requests and download instances in response to EPSRC expectation III is still somewhat unresolved as the catalogue is likely to be pointing to datasets held under a wide variety of conditions. For data that is fully open and held within the institutional repository it should be a relatively achievable task. Should the repository be publicly exposed via a web portal then it is a relatively simple task to use google analytics to gather this data. If the repository links through to Arkivum in the future then it may be possible to exploit the functionality of their system to gather download statistics. However, this is not the whole story and in cases where there are access restrictions in place some other method of gathering this information will be required. At present it is hard to say how varied this monitoring will have to be.

**St Andrews’ main RDM challenges**

Naturally, when implementing support for as broad and complex an issue as research data management, there will be whole raft of problems to solve. At St Andrews, these were the most notable to date:

- It is difficult to establish the amount of storage required for RDM, particularly to the level of certainty required to build a compelling evidence-based business case. This is an area where a national strategy on developing the appropriate shared infrastructure could be cost-effective for the sector
- Policy development has been fairly straightforward, as has getting high-level support and ratification. The difficult part has been developing and implementing the underlying action plan, particularly in terms of securing resources
- Agreeing on roles and responsibilities during the organisational restructuring has been time-consuming
- The focus on RDM has drifted during the most intense periods of REF activity. The timing of the EPSRC’s survey of researchers RDM awareness was very helpful as it refocused attention on support provision in the aftermath of REF submission
- Incentivising EPSRC researchers to produce data management plans is challenging in the absence of a funder mandate to submit a plan with the grant application

---

⁹ St Andrews Research portal: [https://risweb.st-andrews.ac.uk/portal](https://risweb.st-andrews.ac.uk/portal)
¹⁰ St Andrews guidance on compliance with RCUK OA policy: [st-andrews.ac.uk/library/services/researchsupport/openaccess/funderpolicies/rcuk/](http://st-andrews.ac.uk/library/services/researchsupport/openaccess/funderpolicies/rcuk/)
¹¹ Edinburgh research Explorer: [research.ed.ac.uk/portal/](http://research.ed.ac.uk/portal/)
Tips for other institutions developing RDM support services

- Get senior academics on board from the start

- Similarly, researchers should be at the centre of the development of policies and procedures. Support services should reflect their motivations and take account of how they do research, and should also take into account the infrastructure that is already in place. Many academics will be engaging with RDM to some extent and it is important for central services to complement this activity whilst encouraging researchers to retain responsibility for the integrity of their data. This reflects lessons that are being learned from the overall development of publications repositories which are probably only now, really engaging researchers as a ‘business as usual’ part of their research process; and arguably this has been driven by the HEFCE REF2020 OA policy

- Invest in infrastructure incrementally - it is hard to argue for a significant investment due to the difficulty of forecasting data storage requirements, so aim for scalable solutions that can accommodate future expansion