Open access to research data: opportunities, risks and implications
28th March 2013, 2-4.15pm

Report

Synopsis:
The event aimed to provide an opportunity to consider the implications of a move towards open access to research data and to offer a forum for discussion for researchers in the University.

The event was opened by Professor Peter Bath, Deputy Chair of the University Research Ethics Committee (UREC). Presentations were given by Mr Martin Lewis, Dr Tom Webb, Professor Peter Fleming, Dr Farida Vis and Dr Bridgette Wessels.

Martin Lewis is Director of Library Services & University Librarian. He is interested in open access to research data and was a member of the University’s Research Data Management Steering Group.

Dr Tom Webb is a Research Fellow in the Department of Animal & Plant Sciences and has experience of sharing and accessing ecological data sets.

Professor Peter Fleming is a Professor of Industrial Systems and Control in the Department of Automatic Control & Systems Engineering. He frequently collaborates with industrial partners and was a member of the University’s Research Data Management Steering Group.

Dr Farida Vis is a Research Fellow in the Information School. She is interested in open data issues and is a founding member of Open Data Manchester, an organisation for people who are interested in realising the potential of open data to benefit citizens, business and public bodies in Greater Manchester and beyond.

Dr Bridgette Wessels is a Senior Lecturer in the Department of Sociological Studies. She is currently undertaking a European Union funded project on ‘Policy RECommendations for Open Access to Research Data in Europe (RECODE)’.

Professor Peter Bath is the Deputy Chair and Representative for the Faculty of Social Sciences on the University Research Ethics Committee (UREC). He is Professor of Health Informatics in the Information School.
**Session 1:**

The event opened with a presentation by Martin Lewis entitled ‘Open access to research data’. The presentation can be viewed and downloaded from: [www.shef.ac.uk/ris/other/gov-ethics/grippolicy/education/dataevent](http://www.shef.ac.uk/ris/other/gov-ethics/grippolicy/education/dataevent)

Martin’s key messages were:

- Open access to research data sits at the intersection of research data management, open access to publications and open data.
- It is still early days for research data management – researchers are the experts in managing research data.
- There are lots of benefits of research data management such as maximising the impact of data-intensive research, improving the likelihood of success in future grant applications for such research, assurance of research integrity and enhanced data security.
- Research data management facilitates data sharing and collaboration, maximises the opportunities for new research based on reuse and recombination of data from various sources and supports the principles of open access to publicly-funded research outputs recognised by RCUK and OECD.
- However, there are also many challenges such as the volume of data, underdeveloped policies and standards, lack of ownership, heterogeneity of formats, IP, skill/infrastructure/resource and scales of provision.
- A number of external data curation/data storage organisations exist for various types of research, but there are significant gaps; there are therefore challenges relating to the long-term curation of, and access to, data in many disciplines.
- *Science as an open enterprise* (published by the Royal Society in 2012) recommends that ‘scientists should communicate the data they collect and the models they create, to allow free and open access, and in ways that are intelligible, assessable and useable for other specialists in the same or linked fields wherever they are in the world. Where data justify it, scientists should make them available in an appropriate data repository...’.
- *Science as an open enterprise* also recommends that ‘Universities and research institutes should play a major role in supporting an open data culture’.
- The Digital Curation Centre provides information on the curation policies of the main UK research funders.
- Research data management should be considered at the project’s outset and is very hard to retrofit.
- The University undertook a scoping project on research data management in 2011/12 which raised a number of areas for action. A University RDM Coordinator has been appointed for 2 years to take forward these actions. Professional services will be working together to support this area.
- There is a University research data management policy which is part of the University’s Good Research & Innovation Practices Policy.
- There is further support on research data management on the Library's web pages.
- Before we can open research data for sharing, we have to manage and curate it effectively.
- Not all data needs to be shared – this should be a researcher-led decision.

**Session 2:**

Dr Tom Webb presented on his viewpoint and experience of open access to research data in a presentation entitled ‘Open research data = win-win’. The presentation can be viewed and downloaded from: [www.shef.ac.uk/ris/other/gov-ethics/grippolicy/education/dataevent](http://www.shef.ac.uk/ris/other/gov-ethics/grippolicy/education/dataevent)

Tom’s key messages were:
In his view, open research data is win-win. He is an ecologist and his research is a long way from commercialisation. His research looks at the factors that influence the distribution of species; in ecology, most research is conducted at a small-scale in the short-term (more than 60%) whilst most questions currently of relevance to wider society are large-scale, long-term (there is currently considerable pressure towards policy-relevant ecology). Macro-ecologists work at larger scales but these are normally regional rather than global and in these studies the grain-size (i.e. size of site/population) is often larger. This means that there is a gap to be filled in ecology which can be helped by data sharing.

An example of such research is a project on marine ecosystems – this was a large bid (£4m) with lots of partners. However, it would have been impossible to do the research with the required detail on a large scale within the budget so the only option was to collate and use existing data. This involved lots of quality control (including a clear data policy reflecting sensitivities of the data and those involved) and use of effective metadata.

Another example of such research was on a project using the MacroBen database which involved the integration of 44 datasets. The willingness of the data owners was key to the success of this project.

People can be wary about sharing data but Tom believes that other researchers will often use it to answer a totally different research question to the researcher’s own.

There is a moral case for data sharing as well: others should be able to verify what you have done – the data is the actual scholarship behind the journal article.

In his view, open access to research data is more important than open access to publications, and can lead to collaborations, impacts, outputs and credit to the researcher, their department and their institution.

Research is undertaken to advance science and society. Researchers should consider the following questions:

- What were you funded to do?
- Will open data help or hinder progress towards that goal?

Session 3:

Professor Peter Fleming presented on his viewpoint of open access to research data. Peter’s key messages were:

- He undertakes engineering research within the Rolls Royce University Technology Centre so his perspective concerned undertaking research with industry.
- His research involves working with algorithms and methods/models (e.g. to monitor the health of engines), so his interest in data sharing relates to reproducible scholarship; i.e. making the methods available to other researchers to be tested on different datasets and hopefully expose their strengths and weaknesses.
- Data/methods need to be captured in a shareable format to facilitate this. We need an institutional way to address research data management – but it shouldn’t involve too much administration for researchers and it should be devolved to enable discipline-specific approaches.
- Research data management, even at a local level, means that individuals within teams need to work in an organised way.
- The real world studies that he undertakes for Rolls Royce aim to improve the performance of specific machine component. The work is based upon data provided by the company and the results of the research would never be shared due to commercial confidentiality.
Session 4:

Dr Farida Vis presented on her viewpoint and experience of open access to research data. The presentation can be viewed and downloaded from: www.shef.ac.uk/ris/other/gov-ethics/grippolicy/education/dataevent

Farida’s key messages were:
• It is very difficult to share large social media datasets. Twitter, and other social media providers, will not allow it to take place.
• One of her research projects investigated allotments – this was an unfunded project with multiple experimental methods. From the start there was strong engagement from outside academia with various communities.
• The government has recently released an open data portal for government data which includes information about allotments – it is possible to search by datasets but the data tends to be quite basic – is this actually making it easier for people to find out the things they want to know?
• The researchers wanted to find out more about allotment demand and provision so enriched the data on data.gov.uk by using GIS to show real-world problems. They put in Freedom of Information (FoI) Act requests to all UK councils to ask for information on allotment agreements and waiting lists which provided valuable data on local situations.
• The story was picked up by the Guardian which covered the open data angle of the research. It was then subsequently covered by the Daily Mail which covered the fact that people have to wait a long time to get an allotment.
• At this point, a lot more users started getting involved and interested in the project. This highlights that the general public appreciate having access to data on topics that genuinely concern and interest them, rather than open data per se.
• The project has now got EPSRC funding and they have just launched a second round of FoI requests but do have restrictions this time.
• The researchers want to release the data from the project but need to consider how and where this would be done. They do not want a commercial development company to capitalise on the research that they undertook with input from members of the community.

Session 5:

Dr Bridgette Wessels presented on ‘Policy research on open access: an introduction to the RECODE project. The presentation can be viewed and downloaded from: www.shef.ac.uk/ris/other/gov-ethics/grippolicy/education/dataevent

Bridgette’s key messages were:
• RECODE is an EU-funded project entitled ‘Policy RECommendations for Open access to research Data in Europe’ – the University of Sheffield is one of eight partners in the project. The partners are based across Europe and include research councils, publishers and libraries.
• The aims of the project are:
  • To leverage existing networks, communities and projects to address challenges within the open access and data dissemination and preservation sector.
  • To provide a space for European stakeholders to work together to provide common solutions to issues relevant to open access and data dissemination and preservation.
  • To provide over-arching recommendations for a policy framework to support open access to European research data.
The project is looking at five case studies from a range of disciplines, each with their own challenges, including the following:

- Particle physics: huge amounts of numerical data collected by hundreds of academic partners
- Clinical data: includes personal data with issues of quality control, data protection and ethical considerations
- Computational modelling of human physiology: may not be reproducible and can be hard to validate
- Geological data: has issues of interoperability
- Archaeological data: has issues with preservation

The RECODE package has three grand challenges:

- Infrastructure and technology
- Ethical and legal issues
- Institutional issues

Further information is available on the [RECODE website](http://www.recodewebsite.com).

**Session 6:**
Following the presentations, the speakers formed a panel for discussion with the participants. The following questions were discussed:

**I work in Psychology in cognitive neuroscience and I have data that I want to share. How do I go about it? Where do I put it?**
The panel said that the data would need to have metadata to explain the data and ensure it could be used by others. There are a number of data centres, although these are normally discipline specific. It was suggested that the researcher contacts ESRC to ask about depositing it in the UK Data Archive. The ERC, Data Tryad and Figshare are all other options for depositing data. It was acknowledged that there are lots of gaps in disciplinary coverage.

**What institutional support can the University provide to researchers?**
This is an area that the University is currently considering – the aim is to minimise difficulties for researchers in research data management and in depositing their data.

**I have oral history data from communities in Africa but I don’t feel that I have the right to make this data publicly available as I don’t have consent from the participants to do this. I feel it would change the relationship between myself (as the researcher) and the participants fundamentally if I said I would make it open to anyone.**
The panel agreed that this was a difficult situation and highlighted some of the difficulties raised by qualitative data. If the data were anonymised and made available it would risk losing some of its context and meaning. This issue would need to be considered from the very start of the research process – e.g., a copyright contract with the participant might be possible, but this may not be appropriate in the research context. It also raises a question of ‘whose data is it anyway?’ – does it belong to the researcher or the participant? Twitter provides another example of difficulties with ownership – the users are the creators but the company wants to retain ownership and will not allow the data to be deposited.

**At what point in time should data be shared – pre or post-publication?**
There were a number of different views on this question. Some favoured early release, pre-publication, as the researcher will have already done enough work not to be scooped. Others favoured released after publication once the findings have been officially released. It was also
noted there is recognition and credibility in properly managed datasets that further supports the research.

The point was also raised that it is assumed that other users of the data will be other researchers but the end users could be financial, pharmaceutical or public service providers (all of whom currently use open data from the government). This shows emerging power relations around the reuse of data and highlights the need to share data responsibly.

It seems that there is an asymmetry between publicly-funded and privately-funded research. Open access to research data seems a noble idea in academia but others may take your data and use it to make money. This does highlight how a public good can become a private profit. It asks a fundamental question: where does science sit in society? It was noted that creative commons licences could be used to protect the data from being used for commercial purposes. The panel was also in agreement that there are sometimes very good reasons for withholding and not releasing data. The national data centres have well developed policies which can help to inform such decisions.

Each participant was also provided with the section on research data management from the University’s Good Research & Innovation Practices Policy, which can be viewed here: [www.shef.ac.uk/ris/other/gov-ethics/grippolicy/practices/all/rdm](http://www.shef.ac.uk/ris/other/gov-ethics/grippolicy/practices/all/rdm)