IJDC | *General Article*

The Administrative Load of Sharing Sensitive Data: Challenges and Solutions?

Kirsty Merrett
University of Bristol

Zosia Beckles
University of Bristol

Stephen Gray
University of Bristol

Debra Hiom
University of Bristol

Kellie Snow
University of Bristol

Damian Steer
University of Bristol

Abstract

Sharing data openly has become a straightforward process at the University of Bristol. The University's top funders mandate or recommend data sharing as a condition of funding, and many publishers require access to research data to enable results of published articles to be verified. The University has provided a dedicated data repository to support this since 2015, and demand for open publication has risen steadily since its inception. However, an increasing number of requests for sharing data relate to data that has ethical, legal or commercial sensitivities and so cannot be published openly.

Rather than discuss the wide-ranging ethical implications of data sharing, this practice paper will focus on the secure sharing of sensitive data that has ethical approval and, where required, has the necessary consent in place, from the perspective of an institution that has already decided to undertake the work inherent in sharing sensitive data. The specific purpose is to detail the workflow and administrative tasks integral in this and to highlight the types of challenges encountered.

Received 21 January 2018 ~ *Accepted* 21 January 2018

Correspondence should be addressed to Dr Kirsty Merrett, Research Data Service, Library Services, University of Bristol, Augustine's Courtyard, Orchard Lane, Bristol BS1 5DS. Email: j.k.merrett@bristol.ac.uk

An earlier version of this paper was presented at the 13th 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/

Introduction

The Research Data Service (RDS) has introduced a facility for researchers to share data through two controlled means, as either restricted data (where a registration is required) or as controlled data (where a committee assess the request). There is a defined appeals process for rejected applications that directs applicants to the Information Rights Officer as Freedom of Information requests.

Whilst it is rewarding to share data which otherwise may have remained in the hands of one research group, and potentially, or possibly, shared without any legally binding agreement, the administrative checks and audit trail which support data release through controlled routes have proved complex and unwieldy. The processes and policies are being fine-tuned as experience with this grows, but processes are also being challenged as new combinations of data type, requestor status and institutional affiliations come to light.

Making the correct decision about restrictions on data access and data release is imperative, as the right decision makes it more efficient for the applicant and the University. Both 'open' and 'controlled' statuses bear implications. As the recent PLoS/PACE case (PLoS ONE Editors, 2017) illustrates, higher authorities can overturn decisions designed to limit risk factors to participants.

This paper will detail how processes and workflows have evolved over time, and will provide a model for developing good practice within the institution. This undertaking has not been done lightly, and builds on the foundations laid by the team in the past five years. Ensuring researchers know what constitutes sensitive data is key to this endeavour, and the team has worked diligently to crystallize this. The term 'sensitive data' is clearly defined as referring 'to data relating to people, animal or plant species, data generated or used under a restrictive commercial research funding agreement, and any data likely to have significant negative public impact if released.'[1] There is a wealth of support available to researchers throughout the research lifecycle to encourage them to consider carefully any sensitivity inherent in their data and the implications of sharing. The Research Data Service provides detailed guidance for writing Data Management Plans (DMPs) for 14 funders. This encourages researchers to think up front about funder requirements for data sharing and any sensitivity the research may have. There is a dedicated webpage for sensitive data[2] which includes an online 'Sensitive Data Bootcamp tutorial'[3] with guidance developed by the Australian National Data Service,[4] a short video on 'Sharing data from Research Participants'[5] explaining the access levels of the

1 University of Bristol – Dealing with sensitive data: http://www.bristol.ac.uk/staff/researchers/data/dealing-with-sensitive-data/
2 See Footnote 1
3 Sensitive Data Bootcamp tutorial: https://data.blogs.bristol.ac.uk/bootcampsd/
4 ANDS Sensitive Data Guide: https://www.ands.org.au/__data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2018.pdf
5 Sharing data from research participants [VIDEO]: https://www.youtube.com/watch?time_continue=1&v=gFNznhqrpIs

repository and a guidance document 'Sharing research data concerning human participants.'[6]

Having provided researchers with a solid understanding of what constitutes sensitive data, the University Data Repository then provides researchers three different levels of data access at the publications end of the process – open, restricted and controlled.[7]

- **Open data:** The most permissive data access level, suitable for data where there are no particular sensitivities. Where research participants are involved, they have given consent to share anonymized data as 'Open data'; the risk of re-identification is considered as extremely low.
- **Restricted data:** There is some degree of sensitivity involved, e.g. research participants have not given explicit consent to share as 'Open data'. However, the risk of re-identification of participants is considered low. Data is made available to approved bona fide researchers, after they have signed a data access agreement.
- **Controlled data:** There is a large degree of sensitivity involved, e.g. explicit consent from research participants is not in place to share as 'Open data' and the risk of re-identification of participants is medium to high. Requests from bona fide researchers are referred to an appropriate Data Access Committee (DAC) for approval before data can be shared under a data access agreement.

Though many researchers deposit open datasets in the repository as a matter of course, awareness regarding different access levels is on the increase. This growth is partly due to the nature of research undertaken in specific research groups and the academic culture of working across teams on different research projects, but also as a result of discussing the publication of legacy research. As repository capacity and researcher willingness to share has evolved over time, it is unfortunately a lack of forethought in historic consent and patient information sheets that inadvertently mean the data cannot be shared, or at least shared openly. This issue is the same for many repository staff. As Corti (2011) sums up, the researcher may not have intended to close down data sharing, it is often simply that they have failed to 'use the language so they are not going to prohibit [it].'

Aim

The RDS aimed to address the issue of access to sensitive research data through procedures and supporting workflows that enabled researchers to discharge the administrative tasks and decision making of sharing sensitive data to either the RDS (restricted data), or DAC (controlled data). These routes provide a long-term access solution that negates the risk associated with the 'contact the author' route of sharing sensitive data. Added benefits to the University and researcher are the provision of an

6 Sharing research data concerning human participants: https://www.youtube.com/watch?time_continue=1&v=gFNznhqrpIs

7 Data Access levels of he University of Bristol Research Data Repository and FAQs: https://drive.google.com/drive/folders/0B-sxe4ro-QTTN19yNnBwZXFGSlE

independent decision-making body with consistent and transparent processes and a verifiable audit trail. The controlled data routes also supply a clearer picture of who holds sensitive data, provides details of whom it has been shared with, and under what conditions.

Methods

Having identified the need for some form of controlled data access, the RDS set about collaborating with senior University staff to draw up the processes and workflows to support it. Three Task and Finish meetings were convened in the second half of 2015, with members comprised from Faculty Research Ethics Committees, Research Enterprise and Development (Research Governance; Contracts), IT Services (Advanced Computing Research Centre; IT Governance and Risk) and Library Services.

The Task and Finish Group made recommendations to the University's Ethics of Research Committee, following guidance published by the Expert Advisory Group for Data Access[8] (EAGDA). These included the creation of a University of Bristol DAC to oversee access requests for non-open data and to shape and review policies for data sharing, including proportionate governance procedures. A standard Data Access Agreement was introduced to replace the many data transfer agreements in use.

At a glance decision trees (Figure 1) and detailed team workflows were drawn up and agreed by the Task and Finish Group; these have subsequently been refined and revised by the RDS and DAC to accommodate the more complex data access scenarios that have arisen since sensitive data release has commenced. Data transfer procedures have also been expanded to accommodate alternative methods.[9]

Workflows

The majority of requests made to the Research Data Service have been for controlled data; therefore the rest of this paper will concentrate on this level of access. There are three reasons for this focus. Firstly, as illustrated by Figure 1, the application process, background checks and administrative tasks for 'restricted data' are the same, but 'controlled data' requires additional steps. Secondly, the RDS has received 11 requests for access to controlled data but only two requests for restricted data; as different combinations and scenarios have emerged and inevitably challenged the yes/no workflow, more contradictions have arisen with controlled data. Finally, by definition, this level of access carries more risks to participants, so it is important to demonstrate to the community that Data Stewards can hand over the responsibility of sharing the most sensitive of data to Research Data teams as access procedures are 'firm but fair' to both the requester and owner, remain flexible enough to allow for differences between institutions, but retain the spirit of the consent agreed to by research participants.

8 Expert Advisory Group for Data Access: https://wellcome.ac.uk/what-we-do/our-work/expert-advisory-group-data-access

9 If the RDS stores the data, our Senior Technical Officer liaises with the requestor's institution to arrange transfer to a ssh/scp/sftp host where possible, or on an encrypted physical drive where not. If the Data Steward holds the data, they prepare it for this type of transfer.

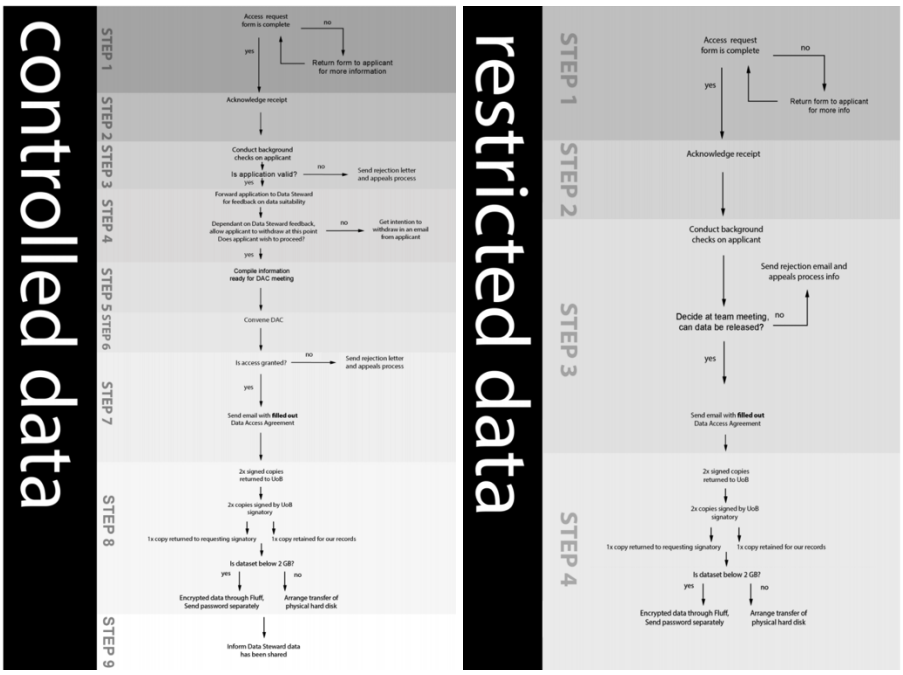

Figure 1. At a glance decision trees produced for the Task and Finish Group.

It is worth emphasising that the bulk of the administrative work is undertaken prior to the DAC being convened; it is the RDS's task to ensure that all the details are present, that any clarification required has already been sought and the answer is satisfactory. The DAC give their time freely, but they are senior academic and professional services staff with full schedules; therefore when the DAC are called upon to make a decision, it must be a productive and efficient use of their time. As demand grows, there are scheduling difficulties, and so for the more clear cut requests, or requests for datasets we have previously released and know well, a 'virtual' DAC is often more effective; it can be called quickly, there are no scheduling problems, and as long as all of the documentation is in order, the turn around from request to decision is quicker. To facilitate sharing information between the DAC and all team members, the RDS uses an institutional Google Drive account (i.e. not a personal Google account) for the storage of the information pertaining to the request, though this is due for review (see the section Next Steps).

The following workflow details the administrative tasks for a clear-cut and tidy application for access to research data. However, it has rarely been as straightforward as this would suggest for a variety of reasons, which will be discussed in due course.

1. **Request for access to controlled data is received**

    a) Request form is completed via a Google Form, with a receipt sent on completion;

    b) A notification email is sent to team members;

c) A folder is added to the team's Google Drive to deposit information from the applicant and Data Steward;

d) Details are added to a Google Sheet, which details the progress of all Data Access requests.

2. Acknowledge receipt

a) An acknowledgement email is sent to the applicant;

b) Include the date by which the applicant will be notified (we endeavour for 20 working days from receipt, in line with FOI requests);

c) Add a date to the RDS calendar to prompt response (changed if new information is received, see Step 3).

3. Access request form checks

a) We have (or have access to) the requested data;

b) Institutional affiliation is given and evidenced (i.e. via a screenshot of applicant's institutional contact directory with onscreen date if possible);

c) Institutional email address is given and evidenced (i.e. via a screenshot of applicant's institutional contact directory with onscreen date if possible);

d) Appropriate institutional signatory is nominated (i.e. via a screenshot of institutional signatory's entry in the institutional contact directory, or name on contracts/governance page with onscreen date if possible);

e) Institutional data protection/data security/information security policies are checked;

f) Applicant has ethical approval in place and this has been provided (if required);

g) If funded, commissioned or sponsored, evidence has been provided;

h) Check that the planned research does not contravene the University's Data Access Agreement;

i) If further clarification is required, contact the applicant, using the clarification email template. The 20 working day period resets (to day 1) when any new information is received (the date is then changed in the RDS Calendar). If clarification is not received within the 20 working day period, the application is rejected on the grounds of having incomplete information;

j) If checks are completed successfully, proceed to the next step.

4. **Data steward's assessment**

   a) The current Data Steward should offer an opinion as to the suitability of the data to the proposed research (Note: the Data Steward cannot 'reject' the application but Data Steward feedback may lead to the applicant's withdrawal if the data is deemed not suitable for planned research);

   b) This opinion should be shared with the applicant and may lead to:

      1. A revision of the planned research in which case the 20 working day period resets (to day 1) when any new information is received, or;

      2. A withdrawal of the application, in which case, request that an intention to withdraw is sent in written form to end the process, or;

      3. Continuation of the application process, in which case proceed to the next step.

5. **Compile information for the DAC, including**

   from the applicant:

   a) Completed data access request form from applicant;

   b) Ethical approval letter from applicant (if applicable);

   c) Funding, commissioning or sponsorship documents (if applicable);

   d) Evidence from background check (see Step 3);

   e) Any other information provided (research protocol, existing correspondence with the Data Steward);

   from depositor:

   f) Ethical approval letter/plan from Data Steward (if applicable). Note: if no ethical approval letter can be located or letter/plan does not cover data, Research Governance to provide a statement on appropriateness for sharing before the DAC meeting, if possible;

   g) Patient information sheet from Data Steward (if applicable);

   h) Blank consent form from Data Steward (if applicable);

   i) Information about any third party agreements from Data Steward (if applicable);

   j) Funding letter, collaboration agreement etc. from Data Steward (if applicable). Check for any documents in the University's Online Management of Research

Contracts and Applications drive if required; Note: if no contracts information exists or contracts information does not cover data, ask the Research Contracts office to provide a statement on appropriateness for sharing before the DAC meeting, if possible.

6. **Convene Data Access Committee**

    a) Decide on a suitable membership for the DAC (i.e. is the Data Steward available? Are Research Contracts/Governance representatives required?);

    b) If no DAC meeting is already arranged within the 20 working day response period, search for a suitable time/venue and create meeting;

    c) If no real world meeting is possible, send information via email and proceed using the virtual DAC route;

    d) Share the Google Drive folder specific to the request with the DAC;

    e) Email each DAC member with information pack, to include information listed in Step 5;

    f) Prepare printed information pack for each DAC member before real world meetings;

    g) For virtual meetings, log all responses in a Google sheet on receipt;

    h) For physical meetings, minute a summary of the discussion to the meeting notes.

7. **Notification of decision**

    a) If application rejected send rejection template and appeals information to applicant;

    b) If application accepted send an acceptance email to the institutional signatory and attach a completed Data Access Agreement (send a copy to the applicant). Ask for two signed paper copies to be returned.

8. **Signed Data Access Agreement is received**

    a) Send two signed copies to The University of Bristol signatory (Deputy CIO) to sign on behalf of the University;

    b) Once signed on behalf of both parties, store one copy in locked pedestal in the RDS team office;

    c) Return one copy to the requesting signatory for their records.

9. **Inform Data Steward**

    a) Inform Data Steward data can now be shared;

b) Data Steward compiles data.

10. **Arrange data delivery**

If <2GB:

a) Encrypt using 7-zip, assign strong password and send to via secure file transfer (currently only our Senior Technical Officer can do this);

b) Email password to applicant under separate cover.

If >2GB:

c) Arrange delivery with applicant on encrypted physical drive;

d) Email password to applicant under separate cover.

Results

The Research Data Service have received over a dozen requests for sensitive data release since the Task and Finish group first met, and has subsequently released seven datasets to institutions worldwide through use of the DAC, with a full audit trail for each request. The remaining requests have either been withdrawn, rejected, or are still in process.

Whilst the RDS started with what seemed to be a fairly straightforward process, the workflow and procedures have needed to become more agile and responsive, as researchers with different circumstances from both within the University and other institutions have come to light.

Administrative Flexibility

The sequence of events is never quite what was anticipated in Figure 1. For example, it may be that the Data Steward who wants to share data with an external researcher initiates contact with the RDS, rather than the process starting when a researcher submits a Data Access request form; or, the RDS may only become aware of the intent to share data at the point we are approached for a Data Access Agreement, which entails starting from the beginning despite the Data Steward having agreed that data will be shared. Arranging access to sensitive data has to tactfully navigate the tensions between the level of administration required for an audit trail, a transparent and clear cut decision making process, and researcher sensibilities, without making it so onerous a process that researchers are tempted to resort to more casual arrangements. Therefore, instead of the simple step-by-step box ticking exercise first envisioned, the workflow is more often now a dialogue between the applicant, Data Steward and RDS staff, with the RDS liaising with professional services staff in Research Governance, Contracts, the Secretary's Office and IT Services about what needs to be clarified and checked at each stage to progress the request.

Supporting Information

Regardless of whether it is the Data Steward or the applicant who has initiated the request process, both parties need to provide the Research Data Service with the relevant information required to make the decision – from the Data Steward, the consent forms, contacts and ethical approval letters, for applicants, research proposals and Research Ethics Committee approvals, and where available, research protocols and funding agreements. This can sometimes prove to be a stumbling block, but it is a fundamental part of the decision making process. Providing evidence that no contractual conflicts exist is difficult, as contracts can reside in any number of administrative offices (e.g. procurement, finance or contracts offices) and the RDS is reliant on the Data Steward understanding the implications of contracts they have entered into.

Compiling Documentation for the DAC

There are three principle challenges to compiling the information for the DAC; checking institutional affiliations, confirming an appropriate signatory, and cross-referencing details of the data access request with the documentation received.

Affiliation

Our initial workflow required the following:

- Institutional affiliation is given and evidenced (i.e. via a screenshot of applicant's institutional contact directory with onscreen date if possible).
- Institutional email address is given and evidenced (i.e. via a screenshot of applicant's institutional contact directory with onscreen date if possible).

However, different institutions have different approaches to staff contact directories – many only have contact details behind an institutional sign-on screen, some institutions allow researchers to 'opt out' of adding some contact details in the contact directory and they may be missing a phone number or email, and on occasion, the researcher is completely absent from the contact directory. Researchers use different email addresses on their application to those provided in the directory, for example departmental or private email addresses. In one instance, an applicant changed names after a change in marital status, and the contact directory had not been updated. These niggles need unpicking and there are often additional conversations that take place with the applicant or other departments to confirm they are a bona fide researcher.

Crucially, the researcher must be under a contract with their institution, as the contract provides the supporting governance processes and risk mitigation; postgraduate researchers and undergraduates do not qualify. Applications are welcome from all organisations with established research governance processes.

Additionally, researchers with 'honorary status' have applied for access, some of which only existed in Universities' RIS pages, or occasionally, on departmental web pages, rather than in contact directories. Ultimately, the decision was made that if the RDS remain meticulous in confirming a named institutional signatory (see below) and if the named person is prepared to sign on behalf of the institution for the researcher, then

we have checked and provided evidence with all due diligence and have resolved the issue.

Institutional signatory

The most common problem is the lack of an appropriate institutional signatory. This needs to be a person from the contracts, legal or research office (rather than supervisors, Head of Department or Research Directors) and it must be a named person (i.e. no blanket Research Office email addresses). As this person is taking on institutional responsibility for the data by signing the Data Access Agreement, the RDS need to evidence to the DAC that they fulfil the criteria and are named on Research Office/Secretary's Office websites, but at times, this has been difficult to do. Furthermore, one institution rejected our request to use a person from the contracts office, and instead, discharged the responsibility to the Head of School, a change in process that required agreement from the Secretary's Office.

Application and supporting documentation cross referencing

It is imperative that the ethical approval, research title, researcher name, duration of research, funders and so on all correspond to the original application. Early Career Researchers or administrative staff are sometimes tasked with the duty of applying for the data, and it is only when discussions are had about honorary status, ethics documents or research titles that it transpires that they will not be handling the data, or sometimes, even involved with the research. It is imperative that the researcher who will be handling the data is the person who applies, as their credentials and experience in handling sensitive data are a part of the DAC's decision making process. Additionally, details regarding the proposed study and the applicant are part of the Data Access Agreement, which is a legal document, so this needs to be accurate.

Increase in Workload

The number of queries about use of the repository for access to sensitive research data has risen almost 80% in a year, from 19 discrete queries in 2016 to 34 in 2017. There are a small but growing number of restricted and controlled datasets (287 open, 19 restricted and seven controlled)[10] but there is a disproportional increase in RDS time dedicated to supporting sensitive data access, one that belies the simple statistics of the number of controlled datasets.

The dialogue that takes place at each stage is time consuming, and requires a number of concurrent conversations with various parties. When there are a number of simultaneous requests, it can become difficult to maintain momentum on each request, and there is always a risk that a small administrative error may be made, a box may not be ticked, or an affiliation may not be as rigorously researched as it could have been. However, at the present rate of enquiries, the RDS is managing the administration of these requests successfully. The administrative process has been shared internally with units who wish to share their own data whilst using the RDS's existing policies and auditing tools.

10 Statistics taken from: https://data.bris.ac.uk/data/dataset?level=top&_Access_limit=0

Conclusions

Sensitive data can be released securely and safely, with robust verification checks and within reasonable time scales. Navigating sensitive data release has proved time consuming, but rewarding; in addition to providing access to publicly funded data and supporting the global reach of researcher endeavours, the project has strengthened the profile of the Research Data Service within the institution and has provided opportunities to work alongside colleagues from other departments and divisions.

Over time, the route will build a clearer map of who has sensitive research data, minimise the risks involved by encouraging researchers to share through more formal routes, and will provide practice based knowledge on how to facilitate data sharing. In addition, the Data Access Committee is available for research groups and professional services to consult when clarification or advice is required.

Next Steps

The Research Data Service will continue to fine tune processes as our experience of handling sensitive data release grows. Each iteration of data release provides us with new circumstances to discuss and resolve with the Data Access Committee. Repeated requests for identical or near identical datasets may provide opportunities to streamline some processes and shorten the time between request and release. Moving to a 'virtual committee' for straightforward requests may alleviate pressure on the Data Access Committee and administrative difficulties of scheduling physical meetings.

Internal applications from both research and honorary contract University of Bristol staff have been directed to the RDS, and we are starting to work through 'light-touch' processes for these applications.

The RDS is attending demonstrations by a number of suppliers of Case Management Systems, alongside Research Enterprise and Development and the Strategic Programmes and Projects office. This will eventually replace use of Google Drive.

References

Corti, L. (2011). *Re-use, sharing, and archiving sensitive research data: A practical overview* [Presentation file]. Retrieved from http://www.dspace.cam.ac.uk/handle/1810/236354

The PLoS ONE Editors. (2017). Expression of concern: Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: A cost-effectiveness analysis. *PLoS ONE, 12*(5): e0177037. doi:10.1371/journal.pone.0177037