LIBRARY COMMITTEE

Wednesday 12th October 2016

UK Research Data Concordat

Description of paper

1. The Concordat on Open Research Data has been developed by a UK multi-stakeholder group including HEFCE, Universities UK and RCUK. This concordat will help to ensure that research data gathered and generated by members of the UK research community is made openly available for use by others wherever possible in a manner consistent with relevant legal, ethical, disciplinary and regulatory frameworks and norms, and with due regard to the costs involved.

2. This paper will highlight and itemise the University of Edinburgh Research Data Service support facilities (existing and planned) which attend to and address the 10 principles of the discipline-agnostic Concordat (detailed below).

Action requested

3. The Committee is asked to note the paper.

Recommendation

4. It is recommended that the paper is circulated to all relevant stakeholders.

Background and context

Principle 1.

Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice.

5. The University of Edinburgh is a world-leading centre of academic excellence with a mission for the creation, dissemination and curation of knowledge. Information Services, a support group within the University, contributes to the University’s mission by striving to provide a Knowledge Management and Information Service appropriate for supporting and enabling learners, researchers and teachers in a world-class University.

6. The University of Edinburgh recognises that research data management is one of the essential areas of responsible conduct of research and provides a framework that supports researchers and their data throughout the course of their research and beyond.

7. It was the first university in the UK to implement a Research Data Management (RDM) policy1 passed by the University Senate in May 2011 requiring that data are managed to the highest standards as part of the University’s commitment to research excellence.

8. In order to implement the RDM policy Information Services (IS) secured significant investment to develop a suite of tools to support researchers and fulfil obligations within changing national and international settings. This was framed by an evolving cross-divisional RDM Roadmap which

---

1 http://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy
detailed a planned programme of effort across four strategic areas: data management planning, active data infrastructure, data stewardship and data management support.

9. The RDM Roadmap\(^2\) and programme were underpinned by a governance model that ensured cooperation amongst Information Services (IS) managers and oversight by an academic-led steering group.

**Principle 2.**

*There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable.*

10. The University of Edinburgh recognises that although many types of data can be shared, it is often not feasible for all data to be made openly available.

11. Open access to data may be restricted if required by legislative, regulatory, contractual, and ethical obligations. This may include:

- **Intellectual Property Rights:** Components of a research dataset covered by third party IPR cannot be made available, unless permission for data sharing has been given by all rights holders.
- **Ethics and Data Protection:** Human subject confidentiality should be maintained throughout the research process, in accordance with the UK Data Protection Act 1998.
- **Sensitivity and Confidentiality:** Data may be restricted if there is a need to protect commercial confidentiality or protect those being studied from harm.
- **Commercial exploitation:** Many research funders allow research projects to restrict or delay the sharing of data if there is a desire to patent aspects of the research.
- **Of national interest:** Data may be restricted if there is a need to protect the national interest or security.

12. Aforementioned legislative, regulatory, contractual, and ethical obligations can be articulated in openly published dataset metadata records held in PURE. For datasets that can ultimately be made openly available an embargo period (of up to five years) can be imposed in Edinburgh DataShare\(^3\) on items to delay making them available until e.g. commercial advantage has been realised, scholarly communication has been published. The forthcoming Data Vault service will securely store data that may need to be restricted until such times that it is possible for it to be shared.

**Principle 3.**

*Open access to research data carries a significant cost, which should be respected by all parties.*

13. The University of Edinburgh appreciates that costing for RDM is a complex issue with many variables to be considered. In general, researchers need to consider the following points when costing RDM:

- **What their research funder requires as a condition of their funding:** Some funders specify a data repository to which the research data must be offered, researchers need to be fully aware of what the funder demands and any costs that will arise from meeting those obligations.

\(^2\) [http://www.ed.ac.uk/files/atoms/files/uoe-rdm-roadmap_-_v2_0_0.pdf](http://www.ed.ac.uk/files/atoms/files/uoe-rdm-roadmap_-_v2_0_0.pdf)
\(^3\) [http://datashare.is.ed.ac.uk/](http://datashare.is.ed.ac.uk/)
demands. They should where possible speak with repository staff when writing their grant application to find out what they need to provide in terms of data quality, metadata, and documentation and how much they charge for the long-term preservation of the data – they also need to find out if the charging model is “pay once store forever” or “annual fees”, if the latter they will need to find out what the funder position is on paying for ongoing costs beyond the life of the grant.

- The amount and complexity of data they are going to collect / create: This is a crucial factor and one that only the researchers will know the answer to, the larger and more complex the dataset being created / collected the greater the effort required and the greater the potential costs.
- Where they will store and back-up the data during the project: The simplest way to address this is to use DataStore. If they estimate that their data volume will exceed the appropriate individual or group allocation then they can buy in additional fully backed-up and resilient storage for £200 per TB per year and this cost should be added into the grant application. If a researcher does decide not to use DataStore then they will need to organise their own back-up systems and calculated any additional costs arising from these and add them to the grant.
- How and when they will create / collect the metadata and documentation required to support the research data: Creating / collecting metadata and documentation requires staff time and effort which may need to be accounted for.
- Where they will archive the data for long term preservation and sharing: Researchers should first offer their data to the repository or archived specified by their funder if one exists, it is up to them to contact the repository and find out what costs, if any, will arise from depositing in that repository. If the funder does not specify a repository then it is the responsibility of the researcher to select the most appropriate long-term home for their data and ascertain costs and other requirements. In this situation Edinburgh DataShare could be used which is free at point of use for all Edinburgh researchers.

**Principle 4.**

*The right of the creators of research data to reasonable first use is recognised.*

14. Edinburgh Datashare Submission Policy states that ‘Items can be deposited at any time, but will not be made publicly visible until any publishers’ or funders’ embargo period has expired.’ An embargo period (of up to 5 years) permits data creators to pursue academic recognition through publication etc. prior to making the data openly available.

15. The visibility of metadata records of datasets generated by University of Edinburgh researchers can be restricted in the Pure Data Catalogue until such times as they can be made publicly through Edinburgh Research Explorer.

16. The University has committed 0.5TB (500GB) of high quality storage (DataStore⁴) with guaranteed backup and resilience to every active researcher. Data currently used as part of the research process are stored via DataStore behind university authorisation and authentication firewalls and are as such not open available. Researchers can choose to share this data via DataSync⁵ with chosen collaborators.

---

⁴ [https://www.wiki.ed.ac.uk/display/ecdfwiki/DataStore+service](https://www.wiki.ed.ac.uk/display/ecdfwiki/DataStore+service)
⁵ [http://www.ed.ac.uk/information-services/computing/desktop-personal/datasync](http://www.ed.ac.uk/information-services/computing/desktop-personal/datasync)
Principle 5.

*Use of others’ data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement.*

17. The University of Edinburgh adheres to the UK Research Integrity Office “Code of Practice for Research.” Researchers depositing or using data need to comply with this code. In addition, university research centres, schools and colleges have discipline-specific research regulations and ethics committees. These committees specify ethical clearance forms and procedures for research undertaken at departmental, school or college level. Academic research staff must comply with “The Data Protection Act 1998” and ensure that appropriate permissions and rights are cleared prior to data collection, secondary data use and/or public dissemination of research outputs. Researchers can seek advice from Records Management in relation to data protection issues including anonymization of personal data, confidentiality agreements, and transferring data to other organisations.

18. Data Library & Consultancy\(^6\) team can researchers to discover and use datasets for analysis, learning and teaching. The Data Library can also help negotiate data licenses and assist with special access requirements to potentially disclosable data.

19. Edinburgh DataShare specifies legal requirements in relation to violations of copyright and data protection in the “Depositor Agreement” which the depositor agrees to when they submit their data. A copy of the Depositor Agreement is sent to each depositor via email after deposit. Through this agreement, the depositor grants permission to the repository to immediately withdraw the data if proof of any legal rights violation is received. In case of withdrawal, the repository reserves the right to retain the metadata record and to state that the data has been removed.

20. Edinburgh DataShare recommends and offers by default the use a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence. Alternatively depositors may manually choose their preferred licence or enter a copyright statement. Edinburgh DataShare policies and procedures comply with the University’s Information Security Policy.

Principle 6.

*Good data management is fundamental to all stages of the research process and should be established at the outset.*

21. The University of Edinburgh has a formal Research Data Management Policy which requires that data are managed to the highest standards as part of the University’s commitment to research excellence. The Policy articulates individual and joint responsibilities of the university and researchers with respect to data management requirements and expectations.

22. Information Services through the Research Data Service provides a suite of tools and support to assist University researchers with data management before, during and after a research project. These include data management planning, working with data, sharing and archiving data. Such activities are complimented by training and awareness raising activities (F-2-F, formal and informal RDM training sessions and courses), RDM promotional material, RDM website\(^7\), and Edinburgh Research Data Blog\(^8\).

---

\(^6\) [http://www.ed.ac.uk/information-services/research-support/data-library](http://www.ed.ac.uk/information-services/research-support/data-library)

\(^7\) [http://www.ed.ac.uk/information-services/research-support/data-management](http://www.ed.ac.uk/information-services/research-support/data-management)

\(^8\) [http://www.ed.ac.uk/information-services/research-support/data-library/edinburgh-research-data-blog](http://www.ed.ac.uk/information-services/research-support/data-library/edinburgh-research-data-blog)
Principle 7.

*Data curation is vital to make data useful for others and for the long-term preservation of data.*

23. The Research Data Service provides a suite of tools and support that map onto the data lifecycle to assist University of Edinburgh researchers and add to the data management value proposition.

24. Edinburgh DataShare is an open multi-disciplinary digital repository of research datasets produced by researchers at the University of Edinburgh. DataShare is designed to be scalable and highly resilient allowing researchers to publish, share, describe, embargo, and licence their data assets. Research data are discoverable through search engines to maximise visibility and impact. The repository includes a standards-compliant metadata schema compatible with repository harvesting protocols, a user interface for deposit and administration, search and browse facilities, item-level and file-level usage statistics, time-stamped submissions and persistent identifiers. DataShare can provide depositors with usage statistics so they know how many times and when their data have been downloaded. DataShare meets with funding body requirements, complies with the University RDM Policy, and assumes responsibility for ensuring researchers’ data are preserved for future use.

25. DataShare was awarded Data Seal of Approval\(^9\) trusted digital repository status in October 2015 in recognition of its commitment to long-term preservation of data.

26. The Data Vault\(^10\) is a closed-access archive storage service offered by Information Services where researchers can safely and securely store their data for the long term. Data is kept safe from accidental deletion and, when combined with a record of the dataset in PURE, can fulfill the expectations of research funders who require long term storage of research data. Data can be copied into the Data Vault from DataStore, and can be placed back in DataStore when a retrieval request is made.

Principle 8.

*Data supporting publications should be accessible by the publication date and should be in a citeable form.*

27. Edinburgh DataShare allows University of Edinburgh researchers to upload, share, and license their data resources for online discovery and re-use by others. The system creates a permanent record, a DataCite persistent identifier (DOI), and a suggested citation, so that work can be formally attributed when used or re-analysed by others. Publishers that require a link to datasets that underpin a publication can link to the data in DataShare via the DOI.

28. University of Edinburgh uses PURE to record descriptive data (metadata) about research data in order to meet institutional and funder policy requirements. Datasets that have been described in PURE are automatically shown as part of a staff member’s online profile in Edinburgh Research Explorer and can be linked to scholarly publications such as journal articles and conference papers. If data are stored in DataShare or an external repository, the DOI can be added to the record in PURE to link them together.

---


29. DataShare records are indexed and harvested by Thomson-Reuters Data Citation Index.

**Principle 9.**

*Support for the development of appropriate data skills is recognised as a responsibility for all stakeholders.*

30. The University of Edinburgh Research Data Service provide a range of data management skills training and support facilities, including:

- **MANTRA**\(^{11}\) (Research Data Management Training) is a free, online non-assessed course with guidelines to help you understand and reflect on how to manage the digital data collected throughout the research process.
- Delivered by the University of Edinburgh in collaboration with the University of North Carolina, the Research Data Management and Sharing MOOC\(^ {12}\) (Massive Open Online Course) uses the Coursera on-demand format to provide short, video-based lessons and assessments across a 5-week period, but learners can proceed at their own pace. Certificates of Accomplishment will be available to any learner who completes a course for a small fee.
- Short courses and interactive workshops open to research staff and postgraduate research students who want to manage their research data effectively and efficiently, and create data management plans as part of their grant applications. The courses and workshops focus on good practice in research data management, working with personal and sensitive data, writing data management plans, and handling data with SPSS. Courses on any aspect of RDM can be tailored for schools, institutes or research groups on demand.

**Principle 10.**

*Regular reviews of progress towards open research data should be undertaken.*

31. The University of Edinburgh RDM Programme set out a four year plan across four strategic areas: data management planning, active data infrastructure, data stewardship and data management support. This was underpinned by a Research Data Management Roadmap (v. 2) which aimed provide clear information to all University staff on the progress that has been made in delivering the RDM programme for the University and on the areas where work remains to be done or where new tools and services are being evaluated or developed for the benefit of researchers across the University. The RDM Roadmap was administered by a Steering Committee chaired by eminent academic with membership representing each of the three colleges, IS and Edinburgh Research and Innovation (ERI). Their role was to provide oversight to the activity of an RDM action group, ensuring the services meet the needs of University researchers. This group contains representatives from across IS each of whom is tasked with delivering separate elements of Roadmap V2.0 whilst ensuring that the entire programme remains integrated.

32. This work was completed in July 2016. A new governance mechanism and strategy are currently being discussed.

**Resource implications**

\(^{11}\) [http://datalib.edina.ac.uk/mantra/](http://datalib.edina.ac.uk/mantra/)

\(^{12}\) [https://www.coursera.org/learn/data-management](https://www.coursera.org/learn/data-management)
There are no resource implications for Library Committee.

**Risk Management**
33. As detailed in the paper, there are significant financial and reputational risks to non-compliance Open Data requirements.

**Equality & Diversity**
34. Equality and diversity have been considered and we do not think there are particular E&D implications for these policies.

**Next steps/implications**
35. Ultimate responsibility for REF and RCUK Open Access policy compliance lies with Schools and Colleges. This paper has been circulated to Research Policy Group.

**Further information**
36. **Author**                      **Presenter**
    Stuart MacDonald               Dominic Tate
    Data Library                   Scholarly Communications
    6th October 2016

**Freedom of Information**
37. This paper can be included in open business.