

learn

LEARN Toolkit of Best Practice for Research Data Management



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

This report is distributed under the terms of the Creative Commons Attribution License (CC-BY) 4.0 <https://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: <https://doi.org/10.14324/000.learn.00>

Contents

Introduction	4
---------------------	---

Part 1 – The Case Studies

Section 1 – Policy and Leadership

Case Study 1	Developing and Implementing the Wellcome Trust’s Data Management and Sharing Policy	8
Case Study 2	Development of a Model Policy for Research Data Management (RDM) at Austrian Research Institutions	12
Case Study 3	Brexit – and its potential impact for Open Access in the UK	17
Case Study 4	Research Data Management supporting Research Integrity and Open Science	23

Section 2 – Advocacy

Case Study 5	Research Data Management Advocacy – what works well	29
Case Study 6	Raising awareness on RDM and engaging stakeholders in Latin America and the Caribbean	33
Case Study 7	UWI St Augustine Campus Libraries and RDM efforts at the UWI, St Augustine Campus	37
Case Study 8	4TU.Centre for Research Data / TU Delft	40

Section 3 – Subject Approaches

Case Study 9	Challenges and Opportunities for Research Data Management in the Arts, Humanities and Social Sciences: a practitioner’s viewpoint	45
Case Study 10	RDM in the Performing Arts: Living Symphonies	57

Section 4 – Open Data

Case Study 11	Why Open Data?	65
Case Study 12	Open Educational Resources: Service setup and Data Management	70
Case Study 13	The handling of research data in the social sciences at University of the Andes – Data Centre (CEDE) – Colombia	73

Section 5 – Research Data Infrastructure

Case Study 14	The Research Data Storage Service at UCL	76
Case Study 15	Scientific Data Management on a Dataverse Network	80
Case Study 16	Delivering the European Open Science Cloud (EOSC): Principle and Practice in delivering Open Science	85

Section 6 – Costs

Case Study 17	Research Data Management at the University of Edinburgh: How is it done, what does it cost?	89
----------------------	---	----

Section 7 – Roles, Responsibilities & Skills

Case Study 18	Training early career researchers	94
Case Study 19	Training subject librarians in Research Data Management	100
Case Study 20	The Emerging Role of the Data Scientist and the experience of Data Science education at the University of Amsterdam	103

Section 8 – Tool Development

Case Study 21	Legal Requirements, RDM and Open Data	115
Case Study 22	Developing a Data Management Plan: a case study from Argentina	119
Case Study 23	Surveying your level of preparation for Research Data Management	123

Conclusions		127
--------------------	--	-----

Part 2 – The Model RDM Policy

• Model Policy for Research Data Management (RDM) at Research Institutions/Institutes	131
• Guidance for Developing a Research Data Management (RDM) Policy	135
• Evaluation Grid of RDM Policies in Europe	139

Part 3 – LEARN Executive Briefing

• English version	168
• Spanish translation	170
• German translation	173
• Portuguese translation	176
• French translation	179
• Italian translation	182

Introduction

In April 2016, Jisc issued an information leaflet on *Managing research data in your institution*.¹ They concluded that ‘data needs to be selected, curated, retained and stored, using appropriate metadata’. The call was timely and aimed at research performing institutions. Similarly, the Research Data Alliance (RDA) also makes an important offering in the research data space.² The RDA is an international organisation focused on the development of infrastructure and community activities aimed to reduce barriers to data sharing and exchange, and to promote the acceleration of data-driven innovation worldwide. With over 4,500 members globally, the RDA comprises individuals, organisations and policy makers representing multiple industries and disciplines, who are committed to building the social, organisational and technical infrastructure needed to reduce barriers to data sharing and exchange, and to accelerating data-driven innovation worldwide.

From 11-17 September 2016, more than 850 data professionals and researchers from all disciplines around the globe convened in Denver, Colorado, for the first edition of International Data Week (IDW). This landmark event, organised by CODATA, the Committee on Data of the International Council of Science (ICSU), the ICSU World Data System (WDS) and the Research Data Alliance (RDA), brought together data scientists, researchers, industry leaders, entrepreneurs, policymakers and data stewards from all disciplines to explore how best to exploit the data revolution in order to improve science and society through data-driven discovery and innovation.³

In the UK, the Digital Curation Centre (DCC) provides access to a range of resources including How-to Guides, case studies and online services. Their training programmes aim to equip researchers and data custodians with the skills they need to manage and share data effectively. The DCC also provides consultancy and support for issues such as policy development and data management planning.⁴

Clearly, research data management is a topic of wide interest. The *Research Data Curation Bibliography* by Charles W. Bailey lists over 620 selected English-language articles, books, and technical reports that are useful in understanding the curation of digital research data in academic and other research institutions.⁵

What issues are current for those involved in RDM? For decision makers, the primary issue is probably that of the associated costs. The 4C project offers an overview of relevant cost models.⁶ One of these is the LIFE model – Lifecycle Information For E-literature – for which one of the LEARN project partners (UCL) was a joint lead.⁷ The LIFE costing model is:

$$L = C + Aq_T + I_T + M_T + BP_T + CP_T + Ac_T$$

where

L = Complete lifecycle cost over time 0 to T.

C = Creation

1. Jisc: <https://www.jisc.ac.uk/guides/research-data-management>; last accessed 8 February 2017.
2. RDA: <https://www.rd-alliance.org/>; last accessed 8 February 2017.
3. CODATA: <http://www.codata.org/>; last accessed 8 February 2017.
4. DCC: <http://www.dcc.ac.uk/about-us>; last accessed 8 February 2017.
5. Bailey, C.W.: *Research Data Curation Bibliography*, version 7: 01/24/2017: <http://digital-scholarship.org/rdcb/rdcb.htm>; last accessed 8 February 2017.
6. 4C: <http://www.4cproject.eu/summary-of-cost-models/>; last accessed 9 February 2017.
7. LIFE3: <http://www.4cproject.eu/summary-of-cost-models/16-community-resources/outputs-and-deliverables/105-life3-costing-model-life3/>; last accessed 9 February 2017.

Aq = Acquisition
I = Ingest
M = Metadata Creation
BP = Bit-stream Preservation
CP = Content Preservation
Ac = Access
T = Period of time over which identified activity lasts

However, there is an elephant in the room with regard to RDM costing. As 4C says, 'There is a sizeable canon of research into cost modelling for digital curation but the research is in many ways preliminary and there has been little uptake of the tools and methods that have been developed. For example, tools to manage and estimate costs have not been integrated into other digital curation processes or tools.'⁸ This has made it extremely difficult for research performing institutions to take RDM forward locally when total costs are unclear, for decision makers do not write blank cheques. Even when costs are known, many institutions are unable or unwilling to reveal their costing activities.

Another issue which is setting the agenda for RDM in Europe is the recent publication of the Commission's High Level Expert Group Report on the European Open Science Cloud.⁹ This Report has at its kernel the benefits which Open Data can bring to research communities. It bemoans the current fragmentation in the European research data landscape and states starkly, 'There is no dedicated and mandated effort or instrument to coordinate EOSC-type activities across Member States'.¹⁰

At institutional level, a baseline was drawn by the LERU Roadmap for Research Data,¹¹ which was published in December 2013. This was the first document to look in detail at the opportunities and challenges which face European research performing organisations in the RDM space. LERU is the League of European Research Universities, comprising 23 members in 12 European countries. Two members of the LEARN project, UCL and the University of Barcelona, are also members of LERU.¹² The LERU Roadmap classified the issues facing their members under seven headings:

- Policy and Leadership
- Advocacy
- Selection, Collection, Curation, Description, Citation, Legal Issues
- Research Data Infrastructure
- Costs
- Roles, Responsibilities, Skills
- Recommendations

Whilst the Roadmap was written on behalf of LERU members, the issues it analysed were in fact generic and can be said to apply to any research performing organisation anywhere in the world – they are by no means exclusive to research-intensive universities. The Roadmap 'presents a series of blueprints which LERU members, indeed any European university, could use to begin to tackle the challenges which research data poses. It also has a series of messages for researchers, research institutions, support services and policy makers'.¹³

8. 4C: <http://www.4cproject.eu/about-us/>; last accessed 9 February 2017.

9. European Commission: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; last accessed 9 February 2017.

10. *Ibid.*, p. 6.

11. LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf; last accessed 9 February 2017.

12. LERU: <http://www.leru.org/index.php/public/about-leru/members/>; last accessed 9 February 2017.

13. LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf, p. 3; last accessed 9 February 2017.

The EU-funded LEARN project took up where the LERU Roadmap finished. The five project partners – UCL, LIBER, Barcelona, Vienna and the UN Economic Commission for Latin America and the Caribbean¹⁴ – identified what was needed to embed the principles of the LERU Roadmap into research performing institutions. Accordingly, the LEARN project set out to provide:

1. A model RDM policy for research performing institutions
2. A Toolkit of Best Practice Case Studies, illustrating the challenges and opportunities identified in the LERU Roadmap
3. 5 Workshops to examine the issues and to produce material for the Case Studies – in London, Vienna, Helsinki, Santiago in Chile, and Barcelona¹⁵
4. An Executive Briefing on RDM challenges and opportunities in six languages
5. A self-assessment survey to allow institutions to assess their level of preparation for RDM, with an analysis of the findings
6. Key Performance Indicators (KPI) to assess whether all elements of the LEARN template for a RDM policy are included in institutional policy work; and a set of KPIs to measure implementation of the policy
7. Lists of Further Reading and a Glossary of technical RDM terms¹⁶

This Toolkit is a major deliverable of the LEARN project. It takes the issues identified above in the LERU Roadmap and provides templates and Case Studies on how success in implementation can be achieved. Many of the formal LEARN outputs and Deliverables are therefore embedded in the Toolkit.

Part 1 consists of 23 Case Studies. These are drawn from issues in the original LERU Roadmap, enlarged by discussions and interactions in the five LEARN Workshops. Briefings in six languages, summarising the main findings in the Roadmap, were prepared for the Workshop participants.¹⁷ Workshops were attended by a range of stakeholders: researchers, research funders, librarians, IT specialists, publishers, senior institutional decision makers, and RDM specialists. The sessions consisted of formal presentations, with discussion and breakout sessions, allowing for wider discourse. In Santiago, the breakout sessions were replaced with panel sessions. Latin America and the Caribbean supplemented their formal Workshop with three mini-Workshops to gain more feedback. LEARN also held a café session at the Amsterdam Open Science Conference in April 2016. Feedback from all these sessions was analysed, resulting in the Case Studies contained in Part 1 of the LEARN Toolkit.

Part 2 of the Toolkit contains the Model RDM Policy produced by the University of Vienna, accompanied by Guidance and an Evaluation Grid of 20 European RDM policies which helped the Vienna Team to formulate the Model LEARN Policy. Part 3 consists of an Executive Briefing in six languages aimed at senior institutional decision makers. It takes the main points identified in the LEARN project and explains how senior decision makers can interact with these issues.

The LEARN Toolkit provides an armoury of best practice for all research performing organisations who wish to develop a persuasive RDM offering. We live in an era of data deluge and institutions who remain unprepared to tackle these challenges/seize these opportunities do so at their peril.

¹⁴ LEARN: <http://learn-rdm.eu/en/partners/>; last accessed 9 February 2017.

¹⁵ LEARN: <http://learn-rdm.eu/en/events/>; last accessed 9 February 2017.

¹⁶ LEARN: <http://learn-rdm.eu/en/category/further-reading/>; last accessed 9 February 2017.

¹⁷ LEARN: <http://learn-rdm.eu/en/outputs/project-materials/>; last accessed 9 February 2017.

Part 1

The Case Studies

Section 1

Policy and Leadership

Developing and Implementing the Wellcome Trust's Data Management and Sharing Policy

Author: David Carr (Programme Manager – Open Research, Wellcome Trust)

Email: d.carr@wellcome.ac.uk

1.1. MAXIMISING THE VALUE OF RESEARCH DATA: A KEY PRIORITY FOR RESEARCH FUNDERS

There is a strong and growing consensus among research funders over the need to ensure that data outputs resulting from the research we support are managed and shared in ways that will deliver the greatest benefit to society. Over recent years, funders around the world have introduced policies requiring that their funded researchers make data available to others in a timely and responsible manner, and plan their approach for managing data as an integral part of planning their research.

The Wellcome Trust¹ is a global research foundation dedicated to improving health for everyone through enabling great ideas to thrive. This case study summarises our experience in implementing our data management and sharing policy over the last decade, drawing out some of the key lessons and remaining challenges.

1.2. WHY DOES DATA SHARING MATTER TO RESEARCH FUNDERS?

In common with other research funders, Wellcome's work to encourage data sharing is driven in large part by a recognition that much of the data currently generated by research represents a vast untapped resource. Enabling researchers and other users to access, combine and use data could open up new avenues for discovery and innovation that might never have been anticipated by the original data generators. In addition, access to the data underlying research findings is critical to ensure that these claims can be scrutinised and reproduced. Data sharing also holds the potential to help reduce avoidable duplication and waste – helping to enable research funds to be allocated effectively and enhancing the efficiency of the research enterprise.

1.3. DEVELOPMENT OF WELLCOME'S DATA MANAGEMENT AND SHARING POLICY

Wellcome's policy on data management and sharing² was published in January 2007, and was updated following a review in 2010. It followed two years after the introduction of our policy on open access to research publications³, and built on Wellcome's work over many years to develop key data resources for the benefit of the research community – notably in the genomics field where we took a lead role in ensuring the data generated in the Human Genome Project was made immediately available, with no restrictions on its use.

Unlike our open access policy, where we were able to set out a clear mandate that all original research papers we fund be made open access within six months of publication, our data management and sharing policy allows for a case-by-case approach. Whilst a research article is a single type of output for which

¹ <http://wellcome.ac.uk>; last accessed 5 February 2017.

² <http://wellcome.ac.uk/funding/managing-grant/policy-data-management-and-sharing>; last accessed 5 February 2017.

³ <http://wellcome.ac.uk/funding/managing-grant/open-access-policy>; last accessed 5 February 2017.

a consistent rule could be applied, the optimal approach for sharing data outputs will vary dramatically depending on the nature of the data and the research context. Furthermore, there are some types of data which cannot be shared openly and where limits on sharing are required – particularly data relating to human research participants.

1.4. KEY POLICY PROVISIONS

Our data management and sharing policy is very similar to those of other major funders – including the UK Research Councils and US National Institutes of Health. The key elements are as follows:

- We expect all of our funded researchers to maximise the availability of research data with as few restrictions as possible;
- We require applicants for funding to submit a data management and sharing plan in cases where their proposed research is likely to generate data outputs that will hold significant value as a resource for the wider research community;
- We commit to review data management and sharing plans, and any associated costs to deliver them, as an integral part of the funding decision;
- We expect all users of research data to acknowledge the sources of their data and to abide by the terms and conditions under which they accessed the original data.

1.5. OUR EXPERIENCE IN IMPLEMENTING THE POLICY

Over the ten years the policy has been in place, we have undertaken periodic reviews to take stock of the data management and sharing plans being submitted and to gauge the perspectives of researchers and reviewers. Following our review of the policy in 2010, we introduced more detailed guidance for applicants⁴ on developing data management and sharing plans – structured around seven key questions that plans should address (Figure 1.1). The overall quality of plans has improved over time, but plans still vary significantly in terms of their levels of detail and specificity. Particularly in areas of research where data sharing resources and practices are less well-established, many researchers and reviewers still do not feel there is sufficient clarity on our expectations of researchers or, in many cases, how best to put data sharing requirements into practice.

1. What data outputs will your research generate?
2. When do you intend to share your data?
3. Where will the data be made available?
4. How will your data be accessible to others?
5. Are limits on data sharing required (including any intended to protect research participants or safeguard intellectual property)?
6. How will key datasets be preserved?
7. What resources are required to deliver the plan?

Figure 1.1 Wellcome guidance on developing a data management and sharing plan: key questions

⁴ <http://wellcome.ac.uk/funding/managing-grant/developing-data-management-and-sharing-plan>; last accessed 5 February 2017.

At present, it is also not clear that the costs of implementing data plans are always being fully factored into funded applications. In 2016, we updated our guidance to give greater specificity on the costs that could be requested in terms of people and skills, data storage and computation, data access, data preservation and deposition.

Finally, our ability to monitor the extent to which researchers put their plans into practice is currently limited. While data sharing does form a key criterion in decisions over renewals for major resources and databases we support, we do not have a process, nor the in-house resources, systematically to track the delivery of plans across the bulk of research we support.

1.6. WIDER CHALLENGES IN DATA SHARING

Different research disciplines are at very different stages in developing the resources and practices required to support data sharing. Several major barriers persist which must be overcome if funder policies to maximise the value of research data are to be successful – key amongst these are:

- **Infrastructure and tools** – building and sustaining the technical resources and tools needed to store, access and analyse vast and complex research datasets;
- **Culture and incentives** – fostering a cultural shift to ensure data sharing is valued and rewarded appropriately;
- **Capacity and skills** – developing the skills necessary in the research community to manage and analyse data effectively;
- **Ethics and governance** – establishing the policy frameworks to ensure data sharing is ethical and fair, and maintains public trust.

1.7. WORKING IN PARTNERSHIP

To take on these challenges, Wellcome's approach over recent years has been to focus on strategically important research areas where we believe there is the potential to work with the community to advance data sharing, and to forge partnerships with other funders to drive change. For example, we have:

- established the Expert Advisory Group on Data Access⁵ in partnership with MRC, ESRC and Cancer Research UK to provide strategic advice to funders on data access for cohort and longitudinal studies across genetics, epidemiology and social sciences;
- joined with a consortium of pharmaceutical companies to support the ClinicalStudyDataRequest.com⁶ platform to enable research access to clinical trials data;
- taken a lead role in working with funders and journals to drive the rapid sharing of research data related to public health emergencies⁷.

⁵ <http://wellcome.ac.uk/what-we-do/our-work/expert-advisory-group-data-access>; last accessed 5 February 2017.

⁶ <http://www.clinicalstudydatarequest.com>; last accessed 5 February 2017.

⁷ <http://wellcome.ac.uk/what-we-do/our-work/data-sharing-public-health-emergencies>; last accessed 5 February 2017.

1.8. EMERGING PRIORITIES

Wellcome is actively exploring how we can build on the work we have done to champion data sharing. In terms of our policy, we are likely to move towards a more holistic approach for the management of outputs. Rather than request a data management and sharing plan in isolation, we would ask researchers to outline a plan for managing and sharing any outputs of value (including software and research materials, as well as data) and to describe their approach where relevant to managing any associated intellectual property. In parallel, we are also actively exploring how to strengthen implementation of our data management and sharing policy – including defining more clearly the roles of reviewers and staff in assessing plans and developing a clearer template for plans. We are particularly keen to explore the scope to work with other funders to implement machine-actionable data management plans that dynamically update as data outputs are generated.

Wellcome is also actively developing new opportunities to advance the broader goals of open research and pilot innovative models to push the boundaries of openness – building on our work to establish the Wellcome Open Research⁸ publishing platform and Open Science Prize⁹.

1.9. CONCLUSIONS

Over the last five to ten years, there has been a growing international recognition of the crucial importance of maximising access to research data. There is a strong policy alignment between major research funders, but significant challenges remain in implementing these policies in practice. Based on Wellcome's experience, key issues for funders to consider in developing and implementing data sharing policies include the need to:

- clarify expectations for researchers as far as possible; and to develop guidance tailored to specific research fields and data types in terms of current best practice for data management and sharing, and the resources available;
- establish a clear process for reviewing and assessing data management plans and the associated costs, and a proportionate mechanism to track plans post-award;
- develop new mechanisms for funders to recognise and reward the contributions of researchers who generate and share high quality datasets and initiate a broader cultural shift;
- consider how best to work in partnership at national and international level to:
 - build and sustain repositories, standards and tools to support data sharing;
 - develop the skills and capacity needed to manage, share and analyse data;
 - harmonise policies and practices wherever possible;
 - advocate and champion the ongoing transition to open science approaches.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁸ <http://wellcomeopenresearch.org>; last accessed 4 March 2017.

⁹ <http://www.openscienceprize.org>; last accessed 4 March 2017.

Development of a Model Policy for Research Data Management (RDM) at Austrian Research Institutions

Authors: Paolo Budroni, (Project Director, e-Infrastructures Austria), Barbara Sánchez Solís (Project coordinator, e-Infrastructures Austria) & Imola Dora Traub (LEARN Project Coordinator) – University of Vienna

Email: paolo.budroni@univie.ac.at / barbara.sanchez.solis@univie.ac.at

2.1. BACKGROUND

The establishment of a model research data management (RDM) policy for Austria is underway. On the one hand, this enterprise has been prompted by rising expectations in the research community, particularly in reaction to the Open Research Data Pilot from Horizon 2020, which has been running since 2014. On the other hand, the results of a comprehensive, quantitative survey regarding RDM in Austria were completed between January and March 2015, as part of the project e-Infrastructures Austria.¹ Over 3000 researchers from 20 out of 21 public universities in Austria, as well as three non-university research institutions, took part in the survey² – this response reflected a 9% response rate in Austria. When asked about desired measures for RDM, more than half of the survey participants expressed an explicit desire for guidelines and policies.

It is worth noting that, at the time of the survey, none of the participating Austrian institutions and none of the large national research grant foundations made use of an RDM policy. Only the Austrian Science Fund (FWF), the grant foundation, included a paragraph in its Open Access Policy dedicated to research data, which stated that “whenever legally and ethically possible, all research data and similar materials which are collected and/or analysed using FWF funds have to be made openly accessible”.³

In early 2016, in order to formulate requisite and explicitly-cited guidelines for competent RDM, the Project Management of e-Infrastructures Austria created a “task force dedicated to finding strategies for the management of research data in Austria”. During the lifespan of this expert group, the FWF called for an “Open Research Data Pilot”.

The Expert Group comprised 22 members⁴ from the stakeholder groups including e-Infrastructures Austria, government ministries, Universities Austria (UNIKO), Vice-Rectors of Research, national research-funding organizations, scientists, scientific libraries, IT-services and research services, and was organised by the Library and Archive Services of the University of Vienna. The Expert Group also tasked a nine-member working sub-group to develop a model for RDM policies in Austrian research institutions. The resulting model policy provides exemplary templates in both German and English, which can be adapted to suit the philosophy and needs of any research institution. This model for RDM policy is the result of six months of collaborative work, and was completed during a meeting of the Expert Group on 2 June 2016.

The Library and Archive Services of the University of Vienna worked concurrently with its partners on the

¹ HRSM project, 2014-2016; project sponsor: BMWF; project management: Library of the University of Vienna; 26 project partners. Website: <https://www.e-infrastructures.at/en/home>; last accessed 5 February 2017.

² *Researchers and their Data: Results of an Austria-wide Survey – Report 2015*. Version 1.2. DOI: <http://dx.doi.org/10.5281/zenodo.32043> - Download e-book: https://phaidra.univie.ac.at/detail_object/o:409473; last accessed 5 February 2017.

³ See also: FWF (Austrian Science Fund): <https://www.fwf.ac.at/en/research-funding/open-access-policy/>; last accessed 5 February 2017.

⁴ For information about the members, see the Appendix below.

implementation of the Horizon 2020 Project LEARN.⁵ It proved advantageous that the leadership of the Work Package 3 (Policy Development and Alignment) of the LEARN Project and the leadership of e-Infrastructures Austria were active at the same time, and that both tasks were managed within the same organisation, i.e. the Library of the University of Vienna. For this reason, findings continually flowed back and forth between the expert groups of the two projects. Furthermore, during this same period, the first three (of five) LEARN workshops were held in London, Vienna and Helsinki, and focused on RDM and policy development.

2.2. EVALUATION OF RDM-POLICIES IN THE SCOPE OF PROJECT LEARN

Between July 2015 and June 2016, the Library of the University of Vienna collected and analysed over 40 European RDM policies. In the course of this preparation phase, it became obvious that in many countries (especially in continental Europe) there have hardly been any published guiding principles regarding RDM. After a further selection process, 20 policies were examined more closely based on (identified) format and content-related criteria.⁶ Using an analysis grid, 11 RDM policies from the United Kingdom, four from Germany, one from the Netherlands and four from Finland were evaluated and checked for possible significant changes during this period at regular intervals.

The most striking results from this analysis related to format and content: it was apparent that research institutions often draw on one another, and sometimes sources were even explicitly referenced. Authorship and the date of publication were not always explicitly stated, and standard formatting did not exist. More than half of the policies analysed made no mention of review periods or revision editions. It was universally clear which topics the policies addressed, and largely, to whom they applied. The concrete objectives of the policies were not directly declared in each case. Roles and responsibilities in research institutions were always mentioned, and in some cases were clearly assigned to specific stakeholders. Only very few institutions explicitly name students as stakeholders worthy of consideration. A position on research funders was taken by most institutions, although, with a few exceptions, costs were only indirectly mentioned.

The term “research data” was defined by most institutions, but the terms “research” and “researcher” only rarely; definitions of other key terms (such as “data management plan”) were also rarely supplied. “Open data” as an issue was a universal concern (although to a varying extent), “restricted data” or “closed data” were mentioned in connection with ethical and legal concerns, if at all. In turn ethical and legal aspects were almost always mentioned, but with widely differing interpretations; in many cases, additional guidelines were referenced. Ownership of data was clearly formulated in about one quarter of the selected policies; it is worth noting, however, that although authorship is mentioned, very few delineations between copyright and rights to use were made.

On the topic of “storage and access” it was notable that data security and open access to research data were strongly emphasised, while long-term archiving was only sporadically mentioned. A specific location for the storage of data was, with a few exceptions, not named; although, some research institutions provide or recommend such services. Externally (with respect to a research project) generated and stored data should also be registered internally. The archival storage period for research data was addressed in about half of the examples analysed; the exact lengths of time, if declared, varied, but 10 years was the length of time most commonly cited. The explicit deletion of data was mentioned in only a very few examples, although

⁵ This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654139; website: <http://learn-rdm.eu/>; last accessed 5 February 2017.

⁶ See also: *Evaluation Grid for RDM Policies in Europe. Survey results, August 2016* in this Toolkit, pp. 139-66.

this issue is best addressed by data management plans (DMP). A DMP was described in all examples (in some more thoroughly than others) or even considered as a mandatory requirement; in several policies there is evidence that a template was used, or a DMP-guidance tool (such as that of Digital Curation Centre).

The topic of “support and training” was universally treated as a necessary component of RDM and was mentioned in all policies. In contrast, the relevance of topics such as “educational data” and “cultural heritage” has not yet entered the consciousness of the research community.

2.3. WORK OF THE EXPERT GROUP E-INFRASTRUCTURES AUSTRIA

The Expert Group task force made use of previous data on the subject of RDM, including the results of the report entitled “Researchers and their Data: Results of an Austrian Survey”, the results of the first LEARN workshop, held in London in January 2016, as well as the results of an online conference with universities in South America.

The Expert Group also formed a nine-person work sub-group, which met regularly every two weeks, and was charged with drafting a policy paper. At first, work was begun in English, as many of the existing policy examples were written in English. Over time, however, drafts were broken down to meet Austrian needs, in both language and meaning. The model policy became more and more concrete with each meeting. The project management of e-Infrastructures Austria ensured a continual flow of information between the work sub-group and the Project LEARN, particularly as the breakout sessions during the second and third workshops became more and more focused on policy development in varying European institutions. E-Infrastructures Austria also set a high standard with the organisation of the four-day “Training Seminar for Research Data Stewardship and e-Infrastructures”,⁷ which looked at operational measures in the field of RDM.

The following duties of the Expert Group are of particular importance:

- Regularly exchanging information regarding the development of a model RDM policy with LEARN project partners, particularly with representatives from South America, in order to compare and standardise terminology;
- Utilising the results of the breakout sessions of the LEARN Workshops;
- Keeping the goals and mission outlined in the LERU Roadmap in consideration;
- Upholding the “FAIR guiding principles for scientific data management and stewardship”;⁸
- Gathering feedback from the Austrian research landscape, particularly with regard to rights and organisational guidelines and terminologies;
- Involving institutional computer centres (ICT);
- Cooperating with legal experts;
- Continually exchanging information with representatives from Austrian research funders and sponsors;
- Comparing the results of the work sub-group with the conclusions drawn after the examination of RDM policies across Europe (see also: *Evaluation Grid for RDM Policies in Europe. Survey results, August 2016* in this Toolkit, pp. 139-66).⁹

⁷ See also University of Vienna: <http://e-seminar.univie.ac.at/en/>; last accessed 5 February 2017.

⁸ See also Wilkinson, Mark and others: ‘The FAIR Guiding Principles for scientific data management and stewardship’ in *Scientific Data* 3:160018 doi: 10.1038/sdata2016.18 (2016); last accessed 5 February 2017.

⁹ Also available as LEARN: <http://phaidra.univie.ac.at/o:459219>.

2.4. CONCLUSION

After the creation of a model policy, and in particular its customisation at the local level, many recommendations can be made to help establish efficient RDM at individual institutions. The establishment of RDM support services has proven indispensable. Therefore, the Expert Group also provided recommendations on an organisational and structural scale. In June 2016, the Expert Group decided to publish the model policy¹⁰ and to enter it in the Universities Austria (UNIKO) “Forum Research” for further comments. From the autumn of 2016 the recommendations will be addressed and local adaptations could begin.

Further documents related to this case study are: 1) *Model policy for research data management (RDM) at Austrian research institutions*; 2) *LEARN Evaluation Grid for RDM Policies in Europe. Survey results, August 2016*.¹¹

2.5. APPENDIX

22 members of the task force dedicated to finding strategies for the management of research data in Austria
- Project e-Infrastructures Austria:

Mag. Maria Seissl, Library of the University of Vienna

Head of the Library and Archive Services of the University of Vienna, Coordination task force

Seyavash Amini, University of Hannover

Legal advisor, e-Infrastructures Austria

Mag. Bruno Bauer, Library of the University of Vienna, Medical University of Vienna

Chair of the General Assembly, e-Infrastructures Austria

Mag. Dr. Andrea Braidt, Academy of Fine Arts Vienna

Vice-Rector for Research

Univ. Prof. Dr. Gerhard Budin, University of Vienna

Coordinator, Think Tank, e-Infrastructures Austria

Dr. Paolo Budroni, Library of the University of Vienna

Project Director, e-Infrastructures Austria; Coordinator, work sub-group; Secretary

Dipl.-Ing. Dr. Michaela Fritz, Medical University of Vienna

Vice-Rector for Research and Innovation

Dipl.-Ing. Raman Ganguly, Central Information Services, University of Vienna

Technical Director, e-Infrastructures Austria

Dipl.-Ing. Florin Guma, IT-Services, University of Salzburg

Representative from university IT Services, e-Infrastructures Austria

¹⁰ e-Infrastructures Austria: <http://phaidra.univie.ac.at/o:459162>; last accessed 5 February 2017.

¹¹ See n. 9.

Dipl.-Ing. (FH) Manfred Halver, *FFG, European and International Programmes*
Research funding

Dr. Peter Kraker, *Know-Center*¹²
Representative from the scientific community, Representative from OANA

Mag. Wolfgang Nedobity, *UNIKO*
General Secretary of the Austrian Universities Conference (UNIKO)

Mag. iur. Sabine Ofner, *Federal Ministry of Science, Research and Economics*

Mag. Eva Ramminger, *University and State Library of Tyrol, University of Innsbruck*
Deputy Chair of the General Assembly, e-Infrastructures Austria

Ao. Univ. Prof. Dr. Andreas Rauber, *Vienna University of Technology*
Representative from the scientific community

Dr. Falk Reckling, *FWF*
Research funding

Mag. Barbara Sánchez Solís, *Library of the University of Vienna*
Project coordinator, e-Infrastructures Austria

Dipl.-Ing. Dr.techn. Maximilian Sbardellati, *University of Music and Performing Arts Vienna*
Representative from university IT Services, e-Infrastructures Austria

MinRat Peter Seitz, *Federal Ministry of Science, Research and Economics*

Mag. Sandra Vidoni, *University of Klagenfurt*
Representative from university Research Services

Mag. Michela Vignoli, *Austrian Institute of Technology (AIT)*
Representative from the scientific community

MinRat Daniel Weselka, *Federal Ministry of Science, Research and Economics*



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

¹² The Know-Center is funded within the Austrian COMET program – Competence Centers for Excellent Technologies – under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Science, Research and Economy, and the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

Brexit – and its potential impact for Open Access in the UK

Author: Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries)

Email: p.ayris@ucl.ac.uk

3.1. ESTABLISHING THE CASE

In the Referendum of 2016¹, the UK's decision to leave the EU has caused both delight and consternation. A fundamental driver for that result was the perception that the UK needed to achieve greater autonomy. In some quarters, this has led to loud calls for individual autonomy. London Mayor Sadiq Khan wants London to be given more autonomy from central government following the UK's vote ... to leave the European Union, saying that the city needs to "take back control."² Autonomy is a powerful and emotive word. It is important to note that autonomy is not the same as independence. As the Mayor has also said: 'I want to send a particular message to the almost one million Europeans living in London, who make a huge contribution to our city – working hard, paying taxes and contributing to our civic and cultural life. You are welcome here. We value the enormous contribution you make to our city and that will not change as a result of this referendum.'³ Nonetheless, the Mayor seeks to establish a new agenda for London in a Brexit world: 'It's not simply a state of mind or an attitude – it's what we are: open for talent, for business, for investment.'⁴

If London is open, what does that mean for universities and their activities? First, it would be helpful to tie down what the role of the university in the early 21st century is. Professor Ronald Barnett at the UCL Institute of Education has said, 'We are now coming to have a sense that what it is to be a university in the 21st century necessarily includes a positive orientation to the world, in all of its aspects. The university – as an idea – is not only networked across the world, not only active in many countries, but takes up a positive stance *towards* the world. Indeed, it has a care for the world, wanting to play its part in helping to improve the world.'⁵ That is a very helpful discussion and offers much in terms of understanding the possible consequences of Brexit.

Many commentators have reacted with fear and alarm to the Brexit vote. Immigration is seen by some as the major issue and as a driver for the 'No' vote in the Referendum. Others note the impact of Brexit on exchange rates, and the perceived damage were the UK to leave the Single Market.⁶ For universities, there are enormous concerns over the possible loss of EU funding in Horizon 2020, the ability of UK universities to recruit overseas students and to retain its EU workforce.⁷ Universities UK has highlighted a key concern as: 'In terms of recruiting EU staff in the longer term, any changes will depend on the kind of relationship

¹ Electoral Commission: <http://www.electoralcommission.org.uk/find-information-by-subject/elections-and-referendums/past-elections-and-referendums/eu-referendum/electorate-and-count-information>; last accessed 3/1/17.

² Business Insider UK: <http://uk.businessinsider.com/sadiq-khan-speech-on-london-independence-after-brexit-and-the-eu-referendum-2016-6>; last accessed 3/1/17.

³ Independent: <http://www.independent.co.uk/news/uk/sadiq-khans-brexit-eu-referendum-response-in-full-there-is-no-need-to-panic-a7100071.html>; last accessed 3/1/17.

⁴ Financial Times: <https://www.ft.com/content/d32b1a42-7a5b-11e6-ae24-f193b105145e>; last accessed 3/1/17.

⁵ Barnett, R, 24 June 2016, EU referendum: will UK HE become less global, more parochial? THE blog: <https://www.timeshighereducation.com/blog/eu-referendum-will-uk-he-become-less-global-more-parochial>; last accessed 3/1/17.

⁶ BBC: <http://www.bbc.co.uk/news/uk-politics-32810887> gives an overview of current issues at the end of 2016; last accessed 3/1/17.

⁷ UUK: <http://www.universitiesuk.ac.uk/policy-and-analysis/brexit/Pages/brexit-faqs.aspx>; last accessed 3/1/17.

the UK negotiates with the EU. However, UUK is committed to highlighting the value of all EU staff, including researchers, scientists and academics, and is urging the UK government to guarantee that those currently working at UK universities can continue to do so after the UK exits the EU.’

Clearly, the current situation poses threats. However, the purpose of this article is to suggest that Brexit is not simply a threat, but also an opportunity. A recent article in *Insights* suggested that Brexit presented opportunities for commercial publishing,⁸ ‘... where some publishers see adversity, others see possibility. While there has been much hand-wringing about economic fallout, nearly half of all publishers see Brexit as an opportunity to make money on exports ...’. The words of Sadiq Khan on the future of London are important here – ‘it’s what we are: open for talent, for business, for investment.’ The emphasis is on the word ‘open’, and it is the argument of this article that Brexit presents not only challenges but also real opportunities for the UK and Open Access, not in terms of autonomy but of freedom – the freedom to innovate and to devise new models for the dissemination of scholarly outputs. These are core values of the Open Access movement and 2017 presents the opportunity to invest time and effort to deliver on them.

3.2. DELIVERING THE GOODS

How has the UK contributed to this vision for an Open Access future? Is it an independent view or one shaped in collaboration with others? What challenges lie ahead for the UK in developing its Open Access position and presence? A study of four themes can help tease out answers to these questions: Open Access policies and mandates, EU copyright reform, new Open Access publishing models and Open Science.

Policies and mandates

Brexit means that the UK will leave membership of the European Union, not that it will be leaving Europe. ‘Brexit means Brexit’, but the nature of the future relationship remains to be worked out. However, Open Access is a European – and indeed a global – agenda, not solely a matter for the EU. Europe is awash with Open Access infrastructure. As of 3 January 2017, OpenDOAR listed 3,285 Open Access repositories. 45.2% of these are based in Europe. Looking at the breakdown of repositories by country worldwide, the top 9 countries with a repository presence are as follows:

Country	%	No.
United States	15	493
United Kingdom	7.6	250
Japan	6.4	211
Germany	5.9	193
Spain	3.8	124
France	3.6	119
Italy	3.3	110
Brazil	2.8	91
Poland	2.7	90
Other	48.8	1604

⁸ Wilcock J, and Miller A, The truth and consequences of Brexit: could a catastrophe for academia be an opportunity for publishers?, *Insights*, 29 (3) (Nov. 2016), 216-23; DOI: <http://dx.doi.org/10.1629/uksg.328/>; last accessed 3/1/17.

The UK does well in terms of its place in this particular league table, coming second overall and ahead of any other European nation.

Arguably, the drive towards Open Access in the UK has been driven by UK funder mandates, by the Finch Review and by the recent HEFCE Open Access requirement for REF2020. Research-intensive universities are on the ball in supporting their researchers in meeting the requirements of Open Access funder policies. UCL (University College London), for example, lists 39 funder policies on its website,⁹ only 4 of which are linked directly to the European Union. It should be noted, however, that these European funders are significant funders of UK collaborative research – the European Research Council, the EU's FP7 programme, Horizon 2020 and Marie Curie. In February 2016, UCL noted, 'UCL has retained and strengthened its position as the top performing university in Europe in the major EU funding scheme Horizon 2020, securing more than €103 million so far. In another significant funding success, UCL researchers have recently been awarded nine highly prestigious European Research Council (ERC) Consolidator Grants, totalling around €15 million and placing the university as the second-placed higher education institution in Europe for the number of grant awards under this scheme. UCL has also been awarded 27 Marie Curie International Fellowships, worth around €6 million.'¹⁰ Clearly, loss of EU research funding will have a major impact on the ability of research-intensive universities to undertake research and so to disseminate the results of that research activity as Open Access outputs. As Universities UK has stated: 'UUK will make the case to government of the importance and impact of our strong research collaboration with European partners, highlighting how EU programmes play a central role in supporting this.'¹¹

Funding is a serious issue, but in other areas the UK has made a significant contribution to the global OA debate. The Finch Report,¹² accepted by Government in July 2012, was key in determining a public policy position in the UK on Open Access. On 16 July, Research Councils UK announced that they were also introducing Open Access requirements.¹³ As it has been implemented, RCUK offers funding to research-intensive universities to disseminate their funded research outputs as Gold OA outputs.¹⁴ In the first 3 years of activity, UCL (University College London) exceeded the targets which RCUK had set. The vast majority of papers made Open Access were Gold, supported by RCUK funding.

Year	RCUK target for OA papers	UCL result for OA papers	% compliance
Year 1	693	797	115%
Year 2	815	963	118%
Year 3	924	991	107%
Year 4 (Apr 16-Mar 17)	1090	798	65% (to Oct 17 2016)

In Europe, the Dutch have also taken a similar strong line on Gold Open Access. 'This gold standard open access is the route the Netherlands has been pursuing aggressively at home, and which it has pledged

⁹ UCL: <https://www.ucl.ac.uk/library/open-access/research-funders>; last accessed 3/1/17.

¹⁰ UCL: <https://www.ucl.ac.uk/news/news-articles/0216/17022016-ucl-excels-EU-research-funding>; last accessed 3/1/17.

¹¹ UUK: <http://www.universitiesuk.ac.uk/policy-and-analysis/brexit/Pages/brexit-faqs.aspx#funding>; last accessed 3/1/17.

¹² Association of Commonwealth Universities: <https://www.acu.ac.uk/research-information-network/finch-report>; last accessed 3/1/17.

¹³ Association of Commonwealth Universities: <https://www.acu.ac.uk/research-information-network/finch-report>; last accessed 3/1/17.

¹⁴ Research Councils UK: <http://www.rcuk.ac.uk/research/openaccess/> for the latest iteration of the RCUK Open Access policy; last accessed 3/1/17.

to steer the whole of the EU towards during its ... presidency.’¹⁵ In fact, a 2016 study¹⁶ found that 5 EU countries want to abandon the traditional subscription model and move to Gold Open Access dissemination: the Netherlands, Hungary, Romania, Sweden and the UK. Clearly, the UK has contributed to this debate, a contribution not solely shaped by the EU.

In the UK, the recent HEFCE mandate for Open Access to support the Research Excellence Framework (REF) 2020 is already being very influential in shaping attitudes to OA dissemination in universities.¹⁷ The REF has enormous influence since the results determine the selective annual allocation of quality-related (QR) grant distribution from the Higher Education Funding Councils. There is every chance that REF OA compliance, rather than the Finch review or even the RCUK OA mandate, will be a game changer for the development of OA in the UK going forward.

European Copyright reform

The European Union is currently engaged in what we believe to be the final stages of copyright reform proposals. In Europe, a number of organisations are taking a leading role in supporting demands for academic-friendly copyright reform, bodies such as LIBER (Association of European Research Libraries) and LERU (League of European Research Universities).¹⁸ For these organisations, the crux of the matter is the need to modernise copyright legislation for the digital age. Their case is focussed on the need for an Exception for Text and Data Mining (TDM) to be enshrined in the new legislation.¹⁹ Text and data mining is the process of deriving information from machine-read material. It works by copying large quantities of material, extracting the data, and recombining it to identify patterns. Copyright legislation is involved in the discussion because of the act of copying. For a digital future, let alone an Open Access future, TDM is an essential tool. Researchers will want to mine content which is both Open Access and material which is available from commercial suppliers, where copyright has typically be assigned to the publisher. LIBER and LERU assert that ‘the right to read is the right to mine’; and that all content, to which researchers have legal access, should be open for TDM. There are also legal barriers which restrict researchers’ abilities to mine the open web. This legal uncertainty hampers research and discoveries, which would act as a foundation for innovation and income generation, creating new jobs for the European economy. It is vital that the draft copyright reform proposal²⁰ currently offered by the Commission embraces all these requirements.

When the UK leaves the EU, where will it stand in relation to the new Directive? There are two issues to consider. There are already copyright-friendly regimes in operation around the world: the USA, Asia, Canada and the UK, for example. In the UK, the Hargreaves review of UK copyright frameworks allows an Exception for TDM, but for non-commercial purposes only.²¹ In this form, this mirrors the current proposal from the EU Commission. For the UK, however, a major issue would be how it should react if the final version of the EU reform package is vastly different from the UK’s current offering. If the EU adopts such advanced and improved proposals before Brexit, it is possible/likely that the EU stipulations would be carried over into UK law, unless they are rejected by Parliament or the courts. If, however, the EU reform package is delayed further and not adopted until after Brexit, how will the UK react? Given research collaborations between

¹⁵ Science Business: <http://sciencebusiness.net/news/77453/few-countries-ready-to-adopt-gold-standard-open-access-to-scientific-journals>; last accessed 3/1/17.

¹⁶ European Commission: http://ec.europa.eu/research/openscience/pdf/openaccess/npr_report.pdf; last accessed 3/1/17.

¹⁷ HEFCE: <http://www.hefce.ac.uk/rsrch/oa/Policy/>; last accessed 3/1/17.

¹⁸ LIBER: www.libereurope.eu and LERU: www.leru.org; last accessed 3/1/17.

¹⁹ LIBER: <http://libereurope.eu/blog/2013/04/25/text-and-data-mining-its-importance-and-the-need-for-change-in-europe/> for the LIBER TDM Factsheet; last accessed 3/1/17.

²⁰ European Commission: <https://ec.europa.eu/digital-single-market/en/news/proposal-directive-european-parliament-and-council-copyright-digital-single-market-for-the-current-version-of-the-proposed-EU-Copyright-Directive>; last accessed 3/1/17.

²¹ CILIP: <http://www.cilip.org.uk/blog/boldly-go-librarians-role-text-data-mining-for-a-library-view-of-hargreaves>; last accessed 3/1/17.

European universities, it would be unacceptable for the UK to have less generous arrangements for TDM than other European partners. This represents a challenge for the UK going forward.

New publishing models

Open Access allows new approaches to scholarly publishing. In the UK, there is a growing amount of interest in the creation of Open Access publishing platforms, often linked to institutional university libraries. One good example of this is UCL Press, the UK's first fully Open Access University Press.²² Grounded in the Open Science/Open Scholarship agenda, UCL Press will seek to make its published outputs available to a global audience, irrespective of the ability to pay, because UCL believes that this is the best way to tackle global 'Grand Challenges'²³ such as poverty, disease, hunger.

The Press focuses its publishing activity on scholarly monographs, scholarly editions, textbooks, edited collections and journals. After 18 months of activity, the Press can report considerable success. It has now surpassed 200,000 downloads for its published outputs. On the website, 43 titles have been published or are in press at the time of writing (3/1/17) – 35 monographs and 8 journal titles. The business model is Open Access, with the university meeting the publishing costs for UCL authors once the submissions have been peer reviewed. For external authors, a Book Publication Charge is levied, which is £5000 for books up to 100,000 words.²⁴ There is a waiver scheme for a number of selected non-UCL authors. The waiver scheme demonstrates UCL's commitment to Open Access publishing and its awareness of the challenges faced by non-funded authors. The UCL Press model is by no means unique to UCL in Europe. However, it is fair to say that institutional Open Access publishing in the UK is fast-growing and self-seeding, and not yet largely driven by other European developments. In this context, Brexit will neither damage nor encourage this home-grown plant to flower.

Open Science

One area in which the European Union has taken a clear leadership role is Open Science. This role was developed under the innovative Dutch presidency of the Union in 2016. The Open Science Conference in Amsterdam in May of that year, and the Council Open Science Conclusions, point to real leadership which the EU has offered.²⁵ The Conclusions have strong ambitions for Open Access.

The Council

'AGREES to further promote the mainstreaming of open access to scientific publications by continuing to support a transition to immediate open access as the default by 2020, using the various models possible and in a cost-effective way, without embargoes or with as short as possible embargoes, and without financial and legal barriers, taking into account the diversity in research systems and disciplines, and that open access to scientific publications should be achieved in full observance of the principle that no researcher should be prevented from publishing; INVITES the Commission, Member States and relevant stakeholders, including research funding organisations, to catalyse this transition; and STRESSES the importance of clarity in scientific publishing agreements.'

²² UCL: www.ucl.ac.uk/ucl-press/about, and UCL Press: <http://www.ucl.ac.uk/ucl-press/publish>, last accessed 3/1/17.

²³ UCL: <https://www.ucl.ac.uk/grand-challenges>; last accessed 31/1/17.

²⁴ UCL Press: <http://www.ucl.ac.uk/ucl-press/publish>; last accessed 3/1/17.

²⁵ Netherlands EU Presidency 2016: <https://english.eu2016.nl/latest/events/2016/04/04/open-science-conference> and Council of the European Union: <http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>; last accessed 3/1/17.

Full Open Access by 2020 is a very ambitious vision. As a member of the EU, the UK is committed to support this objective. After Brexit, depending on the nature of the future relationship between the EU and the UK, the United Kingdom probably will not be mandatorily subject to this requirement going forward. In the UK itself, there is no current equivalent mandate for 100% OA compliance by 2020. The nearest directive is probably the HEFCE requirement for the Research Excellence Framework, also 2020. However, not all research produced in the UK is submitted to the REF. The EU ambition for OA, therefore, is more expansive than the public position in the UK. It has to be said, however, that the UK position on 2020 may be more realistic in terms of the ability to attain the stated objective.

One of the major early deliverables from the Open Science agenda is a bold vision for a European Open Science Cloud (EOSC) of research objects. The Commission has appointed a High Level Expert Group (HLEG) to advise on progress in the Cloud, which is a metaphor for an Internet of data, and the HLEG has recently released its Report.²⁶ I was honoured to be a member of the Group that compiled this document. One of the major observations it contains is that the majority of challenges to reach a functional EOSC are 'social rather than technical'. Another major finding is that there is an 'alarming shortage of data experts both globally and in the European Union'. The Report also determines that the technical components needed to create a first generation EOSC are largely in existence already, but that they are 'lost in fragmentation and spread over 28 member states and across different communities'. There is a real challenge facing the UK, and indeed Europe, if the UK is not a member of the EOSC going forward. Research is global; it does not stop at national boundaries. The UK will suffer if its research data is not visible as part of this European collaboration. Europe, and indeed research communities across the globe, will also be the poorer if they cannot seamlessly access UK research outputs alongside other European findings.

3.3. CONCLUSION

The argument of this paper is that, no matter what sort of relationship the UK develops with the European Union post Brexit, Brexit itself poses not only challenges but also presents opportunities. The Mayor of London has written about a new agenda. The UK has already achieved much in the field of Open Access policy and infrastructure, much without direct dependence on European parallels. Indeed, new models of scholarly publishing, developing quickly in the UK, have the power to redefine how the outputs of research are shared and made available.

Nonetheless, there remain challenges. Loss of funding from bodies such as the European Research Council and programmes like Horizon 2020 would have a detrimental effect on the amount of research which the UK can undertake. And while Brexit may give the UK freedom from European jurisdiction, that must not lead to isolation. The European Union has taken a major leadership role in propounding Open Science approaches. It would be a disaster for the UK, were leadership in this important global agenda to be lost in a country that has cut itself off from wider partnerships and collaborations.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

²⁶ European Commission: <https://ec.europa.eu/research/openscience/index.cfm>, released on 11 October 2016; last accessed 3/1/17.

Research Data Management supporting Research Integrity and Open Science

Author: Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries)

Email: p.ayris@ucl.ac.uk

4.1. OVERVIEW

The purpose of this Case Study is to explore and anchor the concept of RDM in the landscape of research integrity. Once positioned in this space, it is then possible to develop ideas around RDM to support emerging agendas. One of the most important agenda items facing 21st century researchers is Open Science. This Case Study then looks at how RDM can contribute to the Open Science debate and to the benefits to Society that Open Science is said to bring.

4.2. RESEARCH INTEGRITY

All well-managed research performing organisations should have codes of conduct for research integrity, which are developed at institutional level and/or at national level.¹ These codes provide frameworks for best practice in research practice and conduct, establishing principles, guidelines or norms for the ethical, effective and legal conduct of research enquiry. By way of example, this Case Study looks at the framework for research integrity in place in UCL (University College London).²

UCL has a Statement on Research Integrity³ and an accompanying Code of Conduct for Research.⁴ The Statement on Research Integrity makes clear: 'It is the view of UCL that everyone involved with research has a joint responsibility for ensuring high standards of integrity throughout the research process, from the creation of methodology and data collection through to publication and authorship.'

The UCL Statement is itself grounded in UCL 2034⁵, the UCL institutional strategy. Principal Theme 1 of this strategy is 'Academic leadership grounded in intellectual excellence'. In 2012, Universities UK published the *Concordat to support research integrity* and the five commitments set out the UK's determination to maintain high standards of rigour and integrity in its research.⁶

'This concordat⁷ seeks to provide a comprehensive national framework for good research conduct and its governance. As signatories to and supporters of the concordat to support research integrity, we are committed to:

¹ There is a European Code, developed by ALLEA and the European Science Foundation in 2011, a new version of which is to appear in spring 2017, and which serves as a reference document for EU-funded Horizon 2020 projects. See ALLEA: http://www.allea.org/wp-content/uploads/2015/07/Code_Conduct_ResearchIntegrity.pdf; last accessed 7 February 2017.

² Key documents and statements are laid out at UCL: <http://www.ucl.ac.uk/research/integrity/integrity-at-ucl>; last accessed 8/1/17.

³ UCL: <http://www.ucl.ac.uk/research/integrity/pdfs/UCL-Statement-On-Research-Integrity.pdf>; last accessed 8/1/17.

⁴ UCL: <http://www.ucl.ac.uk/srs/governance-and-committees/resgov>; last accessed 8/1/17.

⁵ UCL: <http://www.ucl.ac.uk/2034/>; last accessed 8/1/17.

⁶ Universities UK (UUK): <http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/research-concordat.aspx>; last accessed 8/1/17.

⁷ Universities UK: <http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2012/the-concordat-to-support-research-integrity.pdf>; last accessed 8/1/17.

- maintaining the highest standards of rigour and integrity in all aspects of research
- ensuring that research is conducted according to appropriate ethical, legal and professional frameworks, obligations and standards
- supporting a research environment that is underpinned by a culture of integrity and based on good governance, best practice and support for the development of researchers
- using transparent, robust and fair processes to deal with allegations of research misconduct should they arise
- working together to strengthen the integrity of research and to reviewing progress regularly and openly'

UCL welcomed the 2012 *Concordat to Support Research Integrity* and agrees with the five commitments contained within. The four elements of integrity within the concordat, which UCL sees as Principles of Integrity, reflect UCL's existing Code of Conduct for Research. It is expected that all staff (including honorary staff), students, visitors and collaborators are aware of and adhere to both the Code of Conduct for Research and the Principles of Integrity as set out below (taken in their entirety from the concordat).⁸

- **Honesty** in all aspects of research, including in the presentation of research goals, intentions and findings; in reporting on research methods and procedures; in gathering data; in using and acknowledging the work of other researchers; and in conveying valid interpretations and making justifiable claims based on research findings.
- **Rigour**, in line with prevailing disciplinary norms and standards: in performing research and using appropriate methods; in adhering to an agreed protocol where appropriate; in drawing interpretations and conclusions from the research; and in communicating the results.
- **Transparency and open communication** in declaring conflicts of interest; in the reporting of research data collection methods; in the analysis and interpretation of data; in making research findings widely available, which includes sharing negative results as appropriate; and in presenting the work to other researchers and to the general public.
- **Care and respect** for all participants in and subjects of research, including humans, animals, the environment and cultural objects. Those engaged with research must also show care and respect for the stewardship of research and scholarship for future generations.

These statements and principles set the framework for the performance of research at UCL. The third of the principles from the UUK *Concordat* is important for the topics of Open Science and Research Data Management⁹. How can this principle be delivered? It is to this subject that this Case Study is devoted.

4.3. RESEARCH DATA MANAGEMENT AS A FEATURE OF OPEN SCIENCE

In February 2015, the European Commission published a Report entitled *Validation of the results of the public consultation on Science 2.0: Science in Transition*.¹⁰ Inter alia, the Report looked at the barriers that researchers encounter in moving to Open Science approaches. The top two concerns,¹¹ which acted as barriers, were identified as:

⁸ UCL: <http://www.ucl.ac.uk/research/integrity/pdfs/UCL-Statement-On-Research-Integrity.pdf>; last accessed 8/1/17.

⁹ The new European Code (see footnote 1) is also expected to include transparency and open communication.

¹⁰ European Commission: http://ec.europa.eu/research/consultations/science-2.0/science_2_0_final_report.pdf; last accessed 8/1/17.

¹¹ *Ibid.*, p. 10.

- Concerns about quality assurance – 53% fully agreed that this was a barrier; 35% partially agreed
- Lack of credit-giving for Science 2.0 [Open Science] – 50% fully agreed, 38% partially agreed

The Report then looked at how these barriers could be removed, and the types of intervention that would be needed to do this. The answers to the questions of interest to this Case Study are given in Figure 4.1 below.¹² Comparison of the figures is interesting. There was not much interest amongst researchers in intervention in the metrics space. Concerns about the lack of Open Access to research publications and research data scored highly – in fact, this was the most significant total in the validation exercise.

Question/Issue	Need to Intervene
	Yes
Foster Open Science – raise awareness	52%
Traditional Metrics do not capture Open Science	22%
Develop research infrastructures	56%
OA to publications and data	63%

Figure 4.1: Agreement for Policy Actions [abbreviated] from 2015 EC Report on Science 2.0

The EC's validation exercise points to the realisation amongst researchers that there is a need to raise awareness of these issues, particularly around the issues of Open Access to publications and Open Research Data. How far has the research community travelled in attaining these goals? Again, the UK and UCL's work can act as a helpful example.

The UK has a well-established framework for mandating Open Access to publications. UCL Policy states that all publications should be deposited by the author in UCL Discovery¹³ upon being accepted for publication, copyright permissions allowing. All papers intended for inclusion in REF 2020, the UK's national Research Evaluation Framework,¹⁴ must be deposited within 3 months of acceptance. This is supplemented by funder mandates such as those from Research Councils UK¹⁵ and the Wellcome Trust.¹⁶

For Research Data Management, the picture is much less clear. In July 2016, a group of research stakeholders issued a *Concordat on Open Research Data* – HEFCE, RCUK, Wellcome Trust and UUK. The purpose of this document is 'to ensure that the 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.'¹⁷ The *Concordat* sets out ten principles for Open Research Data, and these are highlighted below. The *Concordat* is important because it amplifies UCL's commitment in its Research Integrity frameworks to openness in collecting, analysing and reporting research data.

¹² For the full set of results, see *ibid.*, p. 32.

¹³ UCL: <http://discovery.ucl.ac.uk>; last accessed 8/1/17.

¹⁴ Higher Education Funding Council for England (HEFCE): <http://www.hefce.ac.uk/rsrch/oa/Policy/>; last accessed 8.1.17; see also UCL: <http://www.ucl.ac.uk/library/open-access>; last accessed 8/1/17.

¹⁵ Research Councils UK (RCUK): <http://www.rcuk.ac.uk/research/openaccess/>; last accessed 8/1/17.

¹⁶ Wellcome Trust: <https://wellcome.ac.uk/funding/managing-grant/open-access-policy>; last accessed 8/1/17.

¹⁷ HEFCE, RCUK, Wellcome Trust, UUK: <http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf/>, [p. 1]; last accessed 8/1/17.

Number	Principle
1	Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice
2	There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable
3	Open access to research data carries a significant cost, which should be respected by all parties
4	The right of the creators of research data to reasonable first use is recognised
5	Use of others' data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement
6	Good data management is fundamental to all stages of the research process and should be established at the outset
7	Data curation is vital to make data useful for others and for long-term preservation of data
8	Data supporting publications should be accessible by the publication date and should be in a citeable form
9	Support for the development of appropriate data skills is recognised as a responsibility for all stakeholders
10	Regular reviews of progress towards open research data should be undertaken

Figure 4.2: The 10 Principles of the UK's Concordat on Open Research Data

4.4. UCL ACTIVITY TO DELIVER THE OPEN DATA AGENDA

UCL has a well-established pattern of activity in supporting Open Access to publications. It has established UCL Press as the UK's first fully Open Access University Press.¹⁸ One of its objectives is 'To embed and explore Open Access approaches as the principal means of dissemination for academic work in a digital world.'¹⁹ The challenge for research-intensive universities like UCL is to expand this activity into all relevant fields of Open Science, including Research Data Management. UCL has a Research Data Management policy, which stresses:

'The purpose of this Policy is to provide a framework to define the responsibilities of all UCL members and to guide researchers and students in how to manage the data, enabling research data to be maintained and preserved as a first class research object and made available to the widest possible audience for the highest possible impact.'²⁰

The position taken by the UCL policy on Open Data is that research data should be as open as possible, as closed as necessary.

Supported by this policy, UCL is taking practical steps to deliver pan-university RDM systems and services. Top-level activities are illustrated in Figure 4.3 below.

¹⁸ UCL Press: <https://www.ucl.ac.uk/ucl-press>; last accessed 8.1.17.

¹⁹ UCL Press: <https://www.ucl.ac.uk/ucl-press/about>; last accessed 8/1/17.

²⁰ UCL: <http://www.ucl.ac.uk/library/research-support/research-data/policies>; last accessed 8/1/17.

Action	Description
Policy	
1	Established an Open Science Policy Platform, chaired by the Pro-Vice-Provost (UCL Library Services), to oversee the embedding of Open Science approaches into UCL
2	UCL has re-organised its committee structures for technical IT developments to support research – with an emphasis of RDM storage, archiving, publication and discovery systems
3	Pro-Vice-Provost (UCL Library Services) working on Open Science incentives for UCL's review of reward and promotion systems, including research data
4	Pro-Vice-Provost co-chairs UCL's Bibliometrics Working Group, which is looking at new metrics systems for research evaluation
5	Pro-Vice-Provost (UCL Library Services) works with Vice-Provost (Research)'s Office to ensure that Open approaches are represented in all appropriate UCL Committees
Practice	
6	UCL Library Services has appointed 1.5 FTE staff members specifically to work on research data advocacy across the whole of UCL; Subject Liaison Librarians trained in RDM in order to support academic colleagues
7	Training in Open Science, including RDM, now offered to UCL Postgraduate Researchers via the Doctoral School Development Programme
8	Baseline survey undertaken across UCL to identify starting point for advocacy, training and raising awareness about RDM and Open Data
9	UCL investing in development of platforms for storage, archiving and publication of research data
10	UCL Library Services has identified <u>funder requirements</u> for RDM and actively advocates these to researchers across UCL as part of a <u>comprehensive support offering</u> for RDM ²¹

Figure 4.3: Outline of current RDM activity to embrace Open Science approaches in UCL

Figure 4.3 outlines ten activities which UCL has prioritised to embed RDM in sound practice for research integrity, leading to Open Science and Open Data, where that is possible. It is important to note that this pan-UCL activity represents a collaborative approach across UCL Divisions. The *LERU Roadmap for Research Data* emphasised in 2013 that work in pursuing RDM institutionally was a collaborative effort. Recommendation 23 captures the spirit of this when it says, 'Involve a broad range of stakeholders in training development and delivery, such as heads of graduate schools with a responsibility for training programmes, the HR department, research librarians, IT directors, accrediting bodies and policy makers'.²²

4.5. CONCLUSION

Research performing organisations should all have sound research integrity frameworks which support the research-intensive nature of their work. Sound RDM activity is key to this framework as research data becomes increasingly recognised as a component of Open Science. The UK's *Concordat on Open Research Data* is a Best Practice example of what is required in order to move on the agenda for Open Research Data at an institutional level. Figure 4.3 in this Case Study shows how roles and responsibilities for this are allocated to a number of stakeholders and different parts of the university. RDM is a key part of the research agenda in the 21st century, and research performing organisations have to be proactive and flexible to embrace the challenges that agenda poses.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

²¹ The UCL RDM website is available as UCL: <http://www.ucl.ac.uk/library/research-support/research-data>; last accessed 8/1/17.

²² LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf, p. 32; last accessed 8/1/17.

Section 2

Advocacy

Research Data Management advocacy – what works well

Author: Myriam Fellous-Sigrist (Research Data Support Officer, Library Services, University College London)

Email: m.fellous-sigrist@ucl.ac.uk

5.1. INTRODUCTION: A RESEARCH-INTENSIVE UNIVERSITY

University College London (UCL) ranks among the top twenty universities in the world and is one of the most successful British research institutions at attracting funding. It is also one of the leading institutions in promoting Open Access to publications. Most academic disciplines are represented in its 380 research departments, units, institutes and centres¹. UCL is home to 12,000 research staff and research students².

In the context of *Science as an Open Enterprise*, a report produced by the Royal Society in June 2012, a Research Data Executive Services Group was created in the university to write the UCL Research Data Policy (published in August 2013) and to oversee Research Data services. Chaired by the Director of Library Services, this Group met regularly between 2011 and 2016 and reported to another cross-university board, the Research Information and IT Services Group.

Two Research Data Support Officers (RDSOs) work as part of Liaison and Support Services within UCL Library Services and in close collaboration with the UCL Information Services Division (IT), as well as several other central services. The RDSOs coordinate Research Data Management (RDM) advocacy and support. As is the case in many other UK universities, these positions were created relatively recently; the first RDSO having started in May 2015, and the second in November 2016. One of the main drivers to create these posts was a change in research funders' requirements³ which prompted UK research-intensive universities to provide greater support for researchers in terms of interpreting and complying with funders' policies.

5.2. RESEARCH DATA MANAGEMENT ADVOCACY AND SUPPORT

The Research Data Management Team's activities can be divided into three interwoven missions:

i Advocacy

In terms of advocacy, the team promotes best practice in data management and sharing, and communicates about services available within the university to support researchers throughout their research projects. For instance, one of the early tasks completed by the first Research Data Support Officer was to create a website dedicated to information about Research Data Management⁴.

Built with the help of a Working Group of library colleagues (see below), the website provides information on research funders' and UCL's policies, Data Management Plans (DMPs), as well as on best practices.

¹ As listed in the UCL Departments A to Z at UCL: www.ucl.ac.uk/departments/a-z/, (accessed 4 August 2016).

² Figure from UCL Human Resources as of 1st October 2015 and UCL Student and Registry Services as of 1st December 2015.

³ In particular new research data policies were established by the Engineering and Physical Sciences Research Council in 2015 and by the European Commission's Horizon 2020 programme in 2016.

⁴ The website can be found at UCL: www.ucl.ac.uk/research-data-management (accessed 3 November 2016).

The website and all research data-related services are mainly promoted via short presentations given upon invitation in faculty and research department meetings. Other advocacy activities have included participation in university-wide induction events for new staff members; presentations for staff members in other professional services; delivering 1- to 3-hour bespoke workshops on data management in research departments; and conducting a university-wide Research Data Management survey.

ii Support

One-to-one and research group support are other key areas of activity. Responding to email and phone enquiries is part of the day-to-day support offered to researchers and research students. Meetings are proposed when the user needs advice on several topics, or when a discussion is required with the whole research group. In addition, the Research Data Management Team reviews Data Management Plans (DMP) written as part of research grant applications or as project deliverables. The review consists of feedback on the content and layout of the plan, but also advice on where to find funders' requirements and guidance to write the plan, and information on relevant university or external resources to improve the plan. Users are offered the opportunity to submit a final draft version of their DMP for a last review.

iii Training

A structured training programme enables the dissemination of information about resources and best practices to both researchers and to non-research staff in central services. A separate training programme to introduce Subject Librarians to Research Data Management has run since 2015. Four sessions with thirty participants in each have been convened to date. Since December 2016 a training programme for PhD students has been co-organised by the Research Data Management Team, the Research Integrity Team (in the Research Office), and the Doctoral School. Embedding RDM advocacy in communications about research support in general enables the Research Data Management Team to reach a wide range of researchers who are at early stages of their projects and careers. In addition to the structured programmes, tailored training sessions are delivered upon demand.

5.3. WHAT WORKED: COLLABORATION, PRESENTATIONS, IMMEDIATE HELP AND INFORMATION GATEWAY

A Collaboration across central services

Several of the activities described above were put together thanks to collaboration with other university central services. Joint work with the Information Services Division, the Research Office, the Ethics Committees and Legal Services has resulted in more efficient promotion of RDM services. Moreover, daily liaison both with the Research IT Services and within Library Services has established a growing network of research support and research data experts. Given that the RDM advocacy programme across the university is still relatively novel, and that the Research Data Management Team is still quite small, this network is proving most valuable. In terms of communications, this enables all of the teams to make the most of each invitation to give presentations at faculty or department staff meetings, and thereby to multiply the opportunities to speak about data management planning and the help available within the university.

The Library Working Group dedicated to RDM is another network on which the team can rely. The Group provides discipline-specific knowledge and essential support for short-term projects such as the building of the website, and designing and promoting the cross-university survey. This Group is formed of thirteen volunteers (Librarians, Records Manager, Digital Curation Manager) who work on a specific project each summer; not all Working Group members are required to participate in all projects. It is planned to offer more training to these members and other librarians so that in the future they can answer basic RDM enquiries and review Data Management Plans.

B Presentations at staff meetings and review of Data Management Plans

It has been possible to draw a direct link between verbal presentations given in departmental and faculty meetings and a subsequent increase of email enquiries received. The average allocated time for such presentations is only 10 minutes, but this enables us to point to a range of university services which are often previously unknown to researchers, and to answer several questions. Regarding feedback on our service, it is the reviewing of DMPs which triggers the highest volume of positive feedback. This includes comments such as:

"I wanted to say a huge THANK YOU [sic] for your time with this feedback and also meeting with me in person. It was incredibly helpful."

"Your comments are very detailed and helpful. Thank you especially for looking at it so quickly. I can't seem to find your internal number on the website. I just wanted to give you a ring to say thank you!"

"Thanks again for your thoughtful comments on our draft. It has helped a lot to revise it!"

Although reviewing one DMP can take up to two hours, this assistance has an extensive and immediate impact on the user's grant application and future project. It also enables an opportunity to point to services, and to explain several aspects of data management to a researcher. These documents are moreover valuable source material to analyse the types of data being produced in the university.

C Gateway to other university services

Being seen as a first point of contact is one of the Research Data Management Team's objectives. Because of the size of the user community and the variety of its enquiries and needs, the Team aims to develop an excellent and up-to-date knowledge of who does what in the institution rather than attempting to become knowledgeable in each and every aspect covered by data management. This strategy has so far been successful and in three cases it has enabled putting researchers in touch with the relevant experts. It has proved the right one whenever several experts had to be brought to one table to help a research group. In a similar perspective, the RDM website has been planned as a gateway to find research support resources across the university.

5.4. THE CHALLENGES OF OPERATING IN A VERY DIVERSE RESEARCH INSTITUTION

While many of our approaches have worked well, working in such a large research-intensive institution does pose challenges. Because only generic knowledge can be developed by a small Research Data Management Team, advocacy has so far failed to extend to discipline-specific needs. This can be frustrating

for both service providers and users alike. To help to address this, the RDM Team now recommends that faculty and department data experts should act as the primary contact for subject-specific questions. The idea is that Research Data Management central services are available to complement help offered by local research support staff (such as permanent data managers, research managers and IT officers) who are able to maintain a subject disciplinary expertise. We also aim to foster and support a network of subject-specific data managers across all faculties.

The diversity of research contexts within the university also forces the RDM Team to prioritise its advocacy effort. The strategy so far has been to help users who self-identify as needing it. This can be because their funder has explicit requirements on research data; or because they generate a large amount of digital data. As a result the Team has been slower to assist and reach out to potential users who do not appear to have data management issues. For example, internally-funded and student projects are not forced to use DMPs; or there can be a misunderstanding of what “research data” encompasses. Solutions found so far to help to overcome these limitations include targeting students via the new training programme, giving presentations in all faculties and stressing in our communications that help is available for all researchers whatever their discipline or type of data produced.

5.5. CONCLUSIONS: MEASURING SUCCESSES

Measuring the success of a service focussing on advocacy and awareness-raising is extremely difficult in terms of metrics. Having only existed for less than two years at the time of writing, relatively little hard data is available. Furthermore, being an entirely new area of engagement for the institution, there is no baseline data from which to measure successes. Instead, we have focussed on qualitative data to promote and inform service development. An institution-wide survey has provided a baseline in terms of key areas, including awareness and understanding of RDM, and we would hope that this can be repeated periodically in an attempt to measure the impact that the RDM Team is having and to further inform service development.⁵



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁵ This survey is described more fully in the Toolkit below, pp. 45-56.

Raising awareness on RDM and engaging stakeholders in Latin America and the Caribbean

Authors: Gabriela Andaur & Wouter Schallier (Consultant and Chief of the Hernán Santa Cruz Library, UN Economic Commission for Latin America and the Caribbean [ECLAC])

Email: wouter.schallier@cepal.org / gabriela.andaur@cepal.org

6.1. THE OVERALL CHALLENGE

Raising awareness among relevant stakeholders is critical for the success of any Research Data Management (RDM) initiative, as their participation and collaboration will be needed for the development and implementation of related policies and programmes. The UN Economic Commission for Latin America and the Caribbean (ECLAC), in its role as a partner institution of the LEARN Project, had as one of its missions to raise awareness and engage RDM stakeholders within Latin America and the Caribbean (LAC).

However, the task constituted a significant challenge due to the geographical dimensions of the region and the socio-cultural diversity within it. For that reason, ECLAC had to develop a strategy that involved several actions, including gathering information about the current state of LAC in regards to RDM; identifying relevant stakeholders; liaising with them to understand their needs and expectations, and planning targeted activities taking into account the particularities of people and institutions within the region.

6.2. RDM IN LAC: STATE OF THE ART

The first step was gathering information about past and current developments in RDM in LAC. This would lead to the identification of institutions, people and projects related to research data, in terms of data creation, management, preservation, access, and policy development.

Due to the complexities in collecting information from such a large variety of countries – each one being a whole universe of people and organisations – six countries were selected as the starting point and main focus of research: Argentina, Brazil, Chile, Colombia, Mexico and Peru. Information was gathered using freely-available publications in several formats, mainly institutional websites, and complemented by interviews with stakeholders when necessary.

This initial approach allowed ECLAC to get a first overview of the RDM landscape in LAC. It could be established that – although isolated or relatively unknown – there are several initiatives from scientific communities and organisations related to the management of research data.

One of the trends identified in the region is the promotion of the management of research data through national legal initiatives in the domain of access to scientific information. The most prominent case is Argentina, where the enactment of the law n° 26899¹ in 2013 set new requirements for individuals and organisations whose research is publicly funded and led to the creation of the National System of Repositories

¹ See <http://repositorios.mincyt.gob.ar/recursos.php>; last accessed 3 March 2017.

(SNRD, by its name in Spanish). Other examples are Peru (law n° 30035, 2015²) and Mexico (*Reform to Science and Technology* law in 2014³), where national repositories are expected to gather research publications and data and make them available to the public.

Brazil is another country making progress in the field of RDM, where efforts are being made by organisations related to scientific development. Among them is FAPESP, the funding agency of the Sao Paulo State, currently in the process of developing agency-wide Research Data Management Plans, and the Brazilian Institute of Science and Technology (IBTC), whose Rede Cariniana – a network of digital preservation services available to Brazilian universities – will make available research data generated by researchers in all fields of knowledge.

These are a few examples of initiatives found in the research phase, which also included discipline-specific repositories in a variety of fields such as social sciences, economics and biodiversity. Their identification served as a basis for the definition of the activities that were undertaken by ECLAC in the following months.

6.3. STAKEHOLDERS IDENTIFICATION

The identification of stakeholders was undertaken along with the gathering of general information about RDM in LAC. During the process, ECLAC sought to identify people and organisations, taking into consideration two criteria: the representation from, at least, the six selected countries, and the presence of the most relevant professional sectors or roles normally involved in the management of data (such as researchers, librarians, IT professionals, policy makers and research funders). The size of the initial list of stakeholders grew throughout the research (reaching over 400 people), and its quality improved mainly thanks to the collaboration of the same stakeholders, who provided useful references to people and projects within particular fields, sectors and/or countries, thus helping ECLAC to build a credible network of contacts within LAC.

6.4. UNDERSTANDING STAKEHOLDERS NEEDS AND EXPECTATIONS

After the main group of stakeholders was defined, over 30 meetings were planned and held, either in person or virtually, with three main objectives in mind: to present the LEARN Project and its goals, to better understand the current state of development of RDM in each country and institution, and to identify the strengths and needs perceived by each stakeholder in this respect.

The meetings proved useful in fulfilling these objectives, although some challenging aspects of working with a diverse group of stakeholders over a large geographic area started to emerge. For example, it became apparent that there was not a single use and understanding of terms related to RDM and the scope and purpose of RDM itself. Moreover, one of the first findings was that Research Data Management was not a commonly used term in LAC, meaning that the difference between RDM and other related terms (such as Open Science, Open Data or Open Access) was not necessarily clear. This was identified as a potential barrier to effective communication with stakeholders.

² CONCYTEC: http://portal.concytec.gob.pe/images/stories/images2013/portal/areas-institucion/dsic/reglamento_repositorio_nacional_alicia.pdf; last accessed 8/2/17.

³ Mexican Government: http://www.diputados.gob.mx/LeyesBiblio/pdf/242_081215.pdf; last accessed 8/2/17.

ECLAC was able to identify different levels of understanding about the implications of RDM and to perceive that stakeholders had different interests and expectations in terms of their collaboration with LEARN. However, they had something in common: they wanted to learn more about RDM and they were also interested in knowing other people and organisations with experience in this area, in particular within the Latin American and Caribbean spectrum. This prompted ECLAC to plan new activities to that end.

6.5. TARGETED ACTIVITIES

Having in mind the differences, needs and expectations of stakeholder groups, ECLAC organised a series of online mini-workshops, designed to serve two main purposes: first, to allow stakeholders to meet and know about each others' experience in RDM and, second, to present and discuss issues about the management of research data, which could also help in setting a common understanding of RDM concepts, as a theoretical ground to build upon in future activities.

The first mini-workshop, titled "Research Data Management (RDM): An overview", was held on 20 April, 2016. A second event, more specific in terms of content, was held on 30 June, 2016, and consisted of a discussion of the current state of development of RDM in one Latin American country, Peru. Both events were held using a virtual platform, and lasted one hour.

A third mini-workshop was held in Port of Spain, Trinidad and Tobago, on 24 November, 2016. This event was different from the first two mini-workshops, as it was an on-site full-day event, focused on the developments in and particular characteristics of the Caribbean context.



Figure 6.1 LEARN Workshop at UN ECLAC, Santiago, Chile

6.6. CREATING A FORUM FOR REGIONAL COLLABORATION

The participation of ECLAC in the LEARN Project considered, from the beginning, the organisation of one regional event, which was held on 27 October 2016 at the UN ECLAC premises in Santiago, Chile. The event was titled “Implementation of policies and strategies in Latin America and the Caribbean”. The programme and activities were strongly tied to the findings of the team in previous activities, and resulted in the gathering of around 90 people representing the regional and professional diversity of stakeholders from Latin American and the Caribbean.

This event, and the three previous mini-workshops, allowed ECLAC to advance in a significant manner its mission of raising awareness on RDM-related issues and engaging stakeholders in Latin America and the Caribbean. They also provided a forum for stakeholders in which they met, learned about other people and organisations’ work, shared their experiences and started a discussion about strategic areas of development. It is hoped that these experiences will also contribute to the creation of alliances and joint projects to foster the development of RDM both across LAC and beyond.

6.7. LESSONS LEARNED

The experience of ECLAC in raising awareness on RDM throughout Latin America and the Caribbean provided several lessons. First, it proved how important it is to identify stakeholders and to understand their situation and needs prior to the planning of specific actions.

Expectations on each side must be known, as any action will have to take into account what each party can provide and what it expects to receive. In this respect, actions should be taken to make sure that appropriate communication channels in all directions are in place, and to identify potential barriers, such as language, preconceptions on a given topic, or different organisational cultures and procedures, among others.

A diverse pool of stakeholders requires a close examination of each one of them before looking at the big picture. This will help any organisation to deliver a clear message and to plan and execute targeted activities relevant and useful to all RDM stakeholders which will, in turn, encourage their engagement in the management of research data.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

UWI St Augustine Campus Libraries and RDM efforts at the UWI, St Augustine Campus

Authors: Marsha Winter and Shamin Renwick, Alma Jordan Library, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago

Email: Marsha.Winter@sta.uwi.edu / Shamin.Renwick@sta.uwi.edu

7.1. WHAT ARE THE ISSUES?

The University of the West Indies (UWI) is unique, in that it is a multi-campus institution located in different countries in the English-speaking Caribbean. After 60 years of existence, it presently has over 45,000 students. Researchers—academic staff and postgraduate students—are actively engaged in many research initiatives at the various faculties, centres and units. However, the notion of research data management (RDM) is still in its infancy. Moreover, universities in the Caribbean have been outpaced by their counterparts in the developed nations with regards to RDM. The key issues that face the UWI, at this time, are:

- Lack of awareness
- Coordination of efforts
- Training
- Costs of implementation of RDM across the campuses of the region.

The LERU Roadmap for Research Data¹ describes six high level sets of issues in introducing research data management at an institutional level:

- Policy and Leadership
- Advocacy
- Selection and Collection, Curation, Description, Citation, Legal Issues
- Research Data Infrastructure
- Costs
- Roles, Responsibilities and Skills.

In terms of placing the UWI in this matrix, the institution is at the earliest stage i.e. policy and leadership.

7.2. AWARENESS

At the UWI St. Augustine (STA) Campus located in Trinidad and Tobago, In September 2015, the UWI STA Campus Libraries participated in a two-day Annual Research Expo which highlighted the research conducted on the campus and the assistance provided for this. One of the objectives of the STA Campus Libraries on this occasion was to show the ways in which the Libraries provide valuable support throughout the entire research cycle from the formulation of the idea, preparation of the literature review and the actual study, gathering data, documentation, publications and archiving data.

¹ LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf; last accessed 7 February 2017.

The Campus Libraries were becoming increasingly concerned about how researchers managed data after it was collected and analysed, as well as its availability for further study. As a result, a survey was designed and administered to researchers—both faculty and postgraduate students—to determine their awareness of data management practices; the size of the data they generally managed; how they stored and archived their data; and how they perceived the Libraries having a role in assisting them managing their data during and at the end of their research cycle, if at all.

Based on the survey results, it was clear that researchers were generally not fully aware of what RDM involved or the four key components of RDM:

- create data and plan for its use
- organise, structure and describe data
- store and preserve data
- search for and share it.

Some of those interviewed felt that emailing their data to their personal email account was a form of archiving their research. Furthermore, when they were asked how they saw the Library helping them with managing data, they were unable to say. These responses were indicative of the fact that the notion and elements of RDM were unfamiliar to the group of respondents at the UWI STA Campus. Also, of interest, was that most people handled small amounts of data < 50 MB and not the large data sets the Libraries had expected.

7.3. COORDINATION OF RDM EFFORTS

Another major issue is that there are multiple departments and initiatives at UWI STA that provide support to researchers, but there is little communication at the moment among them.

One project, the Research Information Management System (RIMS), is an online tool used to identify researchers at UWI with specific knowledge and skills. RIMS allocates each researcher a profile in the database where they can update personal information; learn about current research activities on the campus; access internal funding sources; and locate information on and apply for internal and external grants. Through RIMS, researchers can access training and assistance with the development of research proposals (UWI. ORDKT, 2016²).

Another venture is the Trinidad and Tobago Research and Development Impact (RDI) Fund. This Fund, provided by the Trinidad and Tobago Government but managed by the Office of the Principal of the St. Augustine Campus, offers a maximum of US\$ 300,000 to researchers to develop projects in priority areas such as agriculture, crime, violence and citizen security, public health, climate change and related environmental issues, finance and entrepreneurship, technology and society, and economic diversification and sector competitiveness. Since the establishment of the Fund in 2012, eighty-five (85) concept notes have been received and thirty-one (31) grants totalling over US\$ 2,000,000 have been approved and awarded. Despite these successes, RDM has not been an integral requirement for researchers accessing these funds.

² University of the West Indies (UWI). Office of Research, Development and Knowledge Transfer (ORDKT) (2016) Research and Information Management System. Available at: <https://sta.uwi.edu/ordkt/rims.asp> (accessed 15 May 2016).

Due to the structure of UWI, coordination of RDM would be a challenge to implement. The UWI comprises three physical campuses located at St. Augustine in Trinidad and Tobago; Cave Hill in Barbados; and Mona in Jamaica and a fourth campus, the Open Campus which is both a virtual campus and also consists of seventeen (17) centres located on various islands throughout the English-speaking Caribbean. Each territory has its own governmental policies to which it adheres as well as a distinct cultural landscape. The geographical and administrative issues provide major challenges to setting policies and implementing RDM across the UWI campuses. Furthermore, the question of which department will take the lead in developing the necessary infrastructure across the four campuses is a prime concern. Currently, the institutional repository (IR) called UWISpace is managed at the UWI STA Campus Libraries. Recently, the staff at the Alma Jordan Library based at the St Augustine Campus visited Harvard University in order to gain insight into how *Dataverse*—software used for data management—functions with the view to testing and deploying it across the UWI Campuses.

7.4. TRAINING

For RDM to be successfully implemented, staff must be properly trained to provide support to researchers to assist them during the RDM cycle. Academic libraries have carved out a niche in this area in North American countries. At present, the expertise among the UWI librarians is not at the level to provide the necessary RDM support. Although the technical information technology (IT) expertise may exist, testing and implementing software is just one aspect of implementing RDM.

7.5. COST

At this time, all the UWI campuses are undergoing severe budget cuts due to economic setbacks experienced by the contributing UWI territories. In Trinidad and Tobago, the UWI STA Campus overall budget was cut by more than 14% over the last year and departments on the Campus have had to make sometimes drastic adjustments to cope with the diminished allocations. Implementing RDM across the campuses would involve considerable costs associated with the necessary storage and infrastructure and equipping staff with the appropriate skills.

7.6. CONCLUSION

For the UWI, the key issues are identifying those ready and willing to take charge and to drive RDM at the various Campuses. The STA Campus Libraries have shown initiative by deploying an RDM awareness pilot survey among researchers on the campus and sending staff to acquire knowledge of *Dataverse* software. Libraries would have a critical role to play in the RDM implementation process. Nevertheless at UWI, RDM cannot be realised without the collaboration of all relevant departments.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

4TU.Centre for Research Data / TU Delft

Author: Alastair Dunning - Research Data Services, TU Delft Library

Email: A.C.Dunning@tudelft.nl

8.1. BACKGROUND

The 4TU.Centre for Research Data began as a collaboration between three of the Dutch Technical Universities (Delft, Eindhoven and Twente) that form the wider 4TU Federation (which also includes Wageningen University). Since its inception in 2008, the Centre has used a variety of methods to convince and incentivise researchers to deposit their data. So far, over 7,000 datasets have been deposited, published with Digital Object Identifiers (DOIs) and are openly available for reuse.

The screenshot shows the homepage of the 4TU.Centre for Research Data. The header includes the title '4TU.Centre for Research Data' and navigation links for RSS, Contact, Search, and a flag icon. A red seal indicates '2014 DSA 2017'. The main content is organized into several sections:

- Search in datasets:** A search bar with options to search in data sets or the website.
- Twitter: 4TU.ResearchData:** A feed of tweets, including one from 16 February about FAIR Data Principles and another from 09 February about the Dutch National Plan Open Science website.
- Can we help?:** A section featuring a 3D orange cube icon with symbols like '@', '!', and 'Q'.
- 4TU. Centre for Research Data...:** A section describing the center's mission and services, including a long-term archive, permanent access, and data management support.
- ...for you:** A list of benefits for researchers, such as increasing visibility, being citable with DOIs, and linking publications to underlying data.
- Popular datasets:** A list of featured datasets like 'Real-life event logs', 'BPI Challenge 2014', 'CoSeLoG project', 'CFD in drinking water treatment', 'IDRA', and 'Femur samples'.
- Deposit your data:** A section with a large orange arrow icon pointing upwards.

The footer contains logos for TU Delft, TU/e, UNIVERSITEIT TWENTE, and WAGENINGEN University & Research.

Figure 8.1 - 4TU.Centre For Research Data Homepage [<http://researchdata.4tu.nl/en/home/>]

8.2. METHODOLOGY

Various methods have been used. Some have been implemented by all three of the founding universities working together, while others have been taken up by individual universities.

Case Studies on Depositing with 4TU.Research Data are available at <http://researchdata.4tu.nl/en/researchers-about-4turesearchdata/>.¹

Published case studies have demonstrated different reasons researchers have for depositing data, drawn from scientists across the 3 universities. They have been short essays (illustrated with photographs), designed for publication on the Centre's website, but also re-used in presentations and other publicity material. One case study featured the dataset deposited by Bas Hensen, who wished to demonstrate the validity of the results from his team's ground-breaking experiment, that countered claims made by Einstein. Meanwhile Herman Russchenberg wanted his weather conditions data (which recorded rainfall and other meteorological events) to be used by other researchers on a global basis - this reason provided the basis of the case study.

i Roadshows

A series of lunchtime lectures has been organised for researchers within TU Delft. They have been organised and hosted at a departmental level so that staff from 4TU.ResearchData can tailor their presentations according to different disciplinary requirements. The roadshows have been developed in combination with other staff from the library. This has allowed the roadshows to present information on a variety of issues (Open Access, Current research information system implementation etc.). This has meant that more researchers have attended, so they can find out about whichever issue is pertinent to them.

ii Financial Incentives

4TU.ResearchData are currently planning to release two sets of funding to incentivise researchers from the three technical universities to deposit data. The first is a 'Data Rescue' fund. This will provide funds to researchers to allow them to prepare data so that it is suitable for depositing in the 4TU archive. Data preparation can mean giving the research team time to anonymise data, add documentation, or convert the data into formats suitable for publication.

The second is a 'Data Publication' fund. This will give researchers the time and money to write data reviews of their data for a suitable Data Journal (e.g. the *Geoscience Data Journal*) and provide them with suitable Article Processing Charges, if required. The data will then also be published in the 4TU archive.

iii Working with ICT and Projects

Within TU Delft, working with other partners in the university also helps spread our message. Colleagues in the ICT department often provide advice to researchers on data storage and data processing during projects. We therefore organise regular meet ups with the relevant ICT staff to inform each other about our work. This helps ensure that staff from the ICT department are also capable of passing information to researchers on why they should deposit data.

¹ 4TU Centre for Research Data: <http://researchdata.4tu.nl/en/researchers-about-4turesearchdata/>; last accessed 7 February 2017.

Similarly, we also work with the research funding department of TU Delft. Given the requirements for good data management from the EU in Horizon 2020 and also from the Dutch funding agency NWO, the research funding team began to understand the importance of good data management in successful project proposals. This, in turn, has increased the likelihood of more data deposits when such projects begin to produce data.



Figure 8.2 TU Delft Library, CC-BY-SA, M8Scho

8.3. BUILDING INSTITUTION-WIDE DATA STEWARDSHIP

While the methods referred to above have been useful, it is still a rather piecemeal approach. Given the importance of good data management to the entire scientific research lifecycle, a holistic approach was required.

Therefore the next step has been to get the entire institution to consider good data management. In 2015, TU Delft began its Data Stewardship project, with the goal of introducing policies and best practices for data management in each of the university's eight faculties. This is being achieved with the support of the senior management of the university, who have introduced a broader Open Science programme, with the goal of promoting all types of openness in scholarly communication (e.g. open education, open access).

Being able to work with key stakeholders and persuade them why research data is important is essential. Faculty secretaries, who are the senior administrators within each of the faculties, have an important role to play here, at least in the context of TU Delft. They can help shape policy at a faculty level, but can also gauge and weigh the different responses to data management amongst a faculty's staff. This extra knowledge proves valuable in creating data management policies that can work in tandem with the researcher.

Continuing to find allies in the other support services is also important. For example, connecting with the Graduation School, which provides generic training for students, has allowed the Data Stewardship project to see how it can embed training on effective data management. This is a much longer scale piece of work to implement. It requires numerous stages, and will take a few years to complete. The four identified stages are:

a) initial fact finding within faculties

This has involved interviewing researchers and senior administrators on their attitudes and current behaviour in terms of managing their research data. Particular attention has been paid to the varying practices and methodologies within different disciplines, and the impact these have on data management.

b) development of a draft policy

Based on the above, a draft policy was written stating potential roles and responsibilities for stakeholders within the university. It also offered faculties specific options on how they would deal with the following three areas: training for PhD students, data management plans and training for researchers.

c) ongoing conversations about implementing such a policy within separate faculties

The draft policy is then used to continue discussion on the implementation of good data management, with a focus on the staff and infrastructure required at each stage of the research lifecycle.

d) implementation of processes and policies

It is envisioned that there will be funds made available to allow the faculties to put into practice the demands made in the university-wide policy, for example with regard to PhD training or the creation of individual Data Management Plans for each project. Most importantly, it is hoped that Data Stewards can be embedded in each faculty to provide tailor made help for the different disciplines within the university.

At time of writing the second phase of activity has just been completed.

8.4. CONCLUSION

The 4TU.Centre for Research Data has been in existence for nearly ten years. It has therefore had ample opportunity to explore various methods for incentivising researchers to share their data.

While they have limited reach, the early steps taken (roadshows, published case studies) are essential to get initial contact with stakeholders in the university. The most likely way one has of convincing researchers to deposit their data is if their disciplinary peers are already doing this. Therefore the local case studies are important, offering personal testimony. Roadshows are also important as the face to face contact helps build trust, and gets Library staff out of the Library and into the faculties and departments.

However, to advance all this to the next level, a wider institutional approach is required. TU Delft's Data Stewardship project identifies and works with key influencers throughout the university. Engaging the necessary stakeholders and implementing policies - as opposed to engaging individual researchers on a one to one basis - is essential for any university wishing to see Data Stewardship work at scale.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Section 3
Subject Approaches

Challenges and Opportunities for Research Data Management in the Arts, Humanities and Social Sciences: a practitioner's viewpoint

Author: Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries)

Email: p.ayris@ucl.ac.uk

9.1. CONTEXT

This Case Study is based on conclusions drawn from a study of the challenges and opportunities for Research Data Management in the Arts, Humanities and Social Sciences in a research-intensive university – UCL (University College London).¹ The SCONUL strategic dataset (with data taken from HEAS – Higher Education Statistics Agency) for UCL records 6,286 FTE staff and 31,793 FTE students; the mean for research universities (defined as universities who are members of Research Libraries UK) was 2,661 staff FTE and 19,712 FTE students.

The survey was open to all UCL research staff and research students² and available online over 5 weeks in January and February 2016. The 67 questions dealt with respondents' awareness of policies and UCL services; with their practices of data management planning, data creation, storage and sharing. Finally, they were asked about their needs in terms of support and training. All questions addressed the respondents' most recent research project.

306 fully completed surveys were received (out of 619 unique surveys sent in). 130 research departments, institutes, centres and units were represented among the responses (out of a total of 380³) and were drawn from all UCL faculties.

The majority of responses came from research staff members, who are collaborating with other researchers on their project (either based within their department or external to UCL) and who have received external funding for it. The detailed responses from 3 of UCL's Schools and Faculties are given in the Appendix. These are the Faculties of Arts and Humanities, Laws, and Social and Historical Sciences.

9.2. ANALYSIS

9.2.1 Pan-UCL findings

Overall, the findings across UCL were as follows:

A very positive 70% of respondents are aware of UCL's and of their funder's policy on research data,

¹ UCL Discovery: <http://discovery.ucl.ac.uk/1540140/>; last accessed 9 February 2017.

² In this report, "research staff" encompass two categories of staff used by UCL Human Resources: "Academics" and "Researchers" (both full-time and part-time employees). It does not include "Teachers". "Research students" refer to full-time Graduate Research students.

³ As listed in the UCL Departments A to Z (<http://www.ucl.ac.uk/departments/a-z/>, accessed 4 August 2016).

and 60% of respondents know about the UCL services related to Open Access. However, the level of awareness is problematic when it comes to internal research data-specific services: both the Research Data Management website (online since September 2015) and the Research Data Storage facility (available since 2012) are unknown to 60% of the participants.

The most common types of digital data created by respondents are spreadsheets, texts, databases and images. Remarkably perhaps, the answers also show that 30% of respondents produced non-digital data as part of their most recent projects and another 30% collected personal or sensitive data. Half of the respondents produced less than 100 GB of data over the lifetime of their project.

Data storage and archiving practices are also shown to be problematic. The most common method for storing research data was by using a personally-owned computer (45% of responses); the other favourite choices were a UCL computer, an external hard drive/USB stick or a cloud service. At the end of their project, half of respondents left their data on existing storage and, worryingly, 20% either could not recall exactly where they had archived their data, or had no plans for long-term preservation.

Among those who archived their data, 50% did it for their own re-use; for 20% of research staff it was because of funders' requirements. Half of the respondents have already shared their data with other researchers. Among them, only 25% did not have any concern when sharing data. When concerns were expressed, they were linked to legal questions, misinterpretation and time spent to collect the data.

A very large proportion of respondents (71%) said they thought about data management very early on in their projects, and a third indicated having someone in their team or department responsible for RDM. Yet, when asked what challenges they faced when managing their research data, a long list of problems enumerated by 217 participants is striking.

What is also surprising is that respondents mainly described challenges that are linked to handling data during their projects (storage, dealing with large volumes of data, good record keeping and backing-up procedures). This could indicate that they are not aware of where to find central information on these issues; or that the help available (whether at the central, faculty or department level) is not sufficiently adapted to assist with these essential measures.

Among the options proposed to them, respondents indicated that they would like help primarily in the following areas: storage and preservation of data; writing Data Management Plans; costing data management; data sharing and Open Access to publications. They would prefer to receive such assistance through online resources, training sessions in their department and regular drop-in sessions.

9.2.2 Faculty-specific findings

In terms of levels of awareness of policies and services, Arts, Humanities and Social Science researchers showed low levels of awareness of internal UCL RDM facilities and services. 66% of Arts and Humanities researchers did not know of the UCL RDM website. For Laws, half did not know. With regard to usage, only 10% of Arts and Humanities researchers and 8% of Social and Historical Sciences researchers had actually utilised it. This compares with, say, the Faculty of Engineering, where only 10% of researchers knew about the RDM website and 66% did not (Engineering had 67 respondents from 72 surveys).

As to the creation and analysis of data, an interesting picture emerges. For Arts and Humanities researchers, the most important type of data created was textual, with spreadsheets coming a close second. For Laws

it was databases, with text coming a close second. In Social and Historical Sciences, the most popular forms were spreadsheets and photographs/digitised images, followed closely by databases and text. In Engineering, again the most popular formats were text, followed by spreadsheets and other images. Clearly, there is not the difference between the disciplines that might have been imagined.

Where there *is* a difference is in the size of the datasets created. For Arts and Humanities, only 3 respondents created datasets of 1-10 GB. The figures in Laws are too small to use to draw comparisons. In Social and Historical Sciences, the single biggest category for size of dataset creation was 1-10 GB. This contrasts with the Faculty of Engineering, where 11 of the 58 respondents were creating datasets of 1-10 GB, 12 datasets of 10-100 GB, and 11 datasets of 100 GB-1 TB.

When it came to storing and archiving research data, researchers in the Arts and Humanities commonly selected the hard drive of a personal PC or laptop, or an external hard drive/USB stick. For Laws, the most popular storage medium was an external hard drive/USB stick. In Social and Historical Sciences, the preferred media were the same as for Arts and Humanities.

For long term archiving, researchers in Arts and Humanities preferred subject repositories, or repositories external to UCL. In Laws and Social and Historical Sciences, researchers preferred existing storage. In the latter, many respondents admitted to having no archive plans. Engineers showed a pattern similar to Social and Historical Sciences. They preferred to use existing platforms for long-term storage.

Researchers were asked if they had any concerns about sharing their research data with others. Arts and Humanities on the whole had no concerns, and the same is true for Laws. Social and Historical Sciences did have concerns, however, and the most cited reasons were confidentiality issues/IPR or Data Protection.

Finally, researchers were asked what kinds of support they needed. In Arts and Humanities, researchers cited three main areas: storage and preservation of data, Open Access to publications and Data Management Plans. The same three preferences were cited in Social and Historical Sciences. In Laws, however, the most requested area for help was Open Access to publications, followed by a group who felt they needed no help. By way of illustration, in Engineering the most mentioned common areas for additional support were storage and preservation of data and Data Management Plans.

9.3. CONCLUSIONS AND RECOMMENDATIONS

A number of conclusions and recommendations can be drawn from the UCL survey, which illustrate the challenges and opportunities for research data management in the Arts, Humanities and Social Sciences:

Recommendation n°1

In all disciplines, research funders expect grant applicants and grant holders to explain how they will manage their data and to comply with their Data Management Plans. Being aware of these policies and services is a key element to writing successful funding applications. The earlier researchers receive assistance, the lesser the eventual risks for their projects.

- Faculties and research departments are encouraged to promote data storage solutions which comply with institutional and funder policies.
- Establish a central information service to support research data management

activities, such as <http://www.ucl.ac.uk/library/research-support/research-data>, and ensure this is continually promoted across the institution.

- Where possible, Heads of Departments should periodically invite the institutional Research Data Management team to give brief presentations to staff and research students on what assistance is available to them, including on 1-to-1 support and review of Data Management Plans.
- PhD students should be urged to attend courses on research support as part of an institutional Doctoral Skills Development Programme.⁴

Recommendation n°2

Training and support opportunities for both research staff and research students should not overlook the aspects around personal/sensitive data and databases as a large proportion of researchers use these as part of their projects.

Using personal computers and commercial cloud services to store research data represents a clear security risk for any data and a potential breach of security regulations if these are personal/sensitive data. An increasing number of funders currently expect that research data should be preserved for at least 10 years. Whether using an institutional facility or a discipline-specific repository, researchers should ensure that they know how to find reliable archiving facilities.

Recommendation n°3

The lack of clarity on where to find solutions to all of the challenges cited by research staff is a worrying observation. Academic faculties should be strongly encouraged to consider appointing/designating permanent staff members to assist researchers with data management in their subject disciplines.

Following these recommendations will help to avoid rushed and potentially costly short-term decisions; a lack of support when problems arise; and the outdating of skills and standards.

9.4. APPENDIX

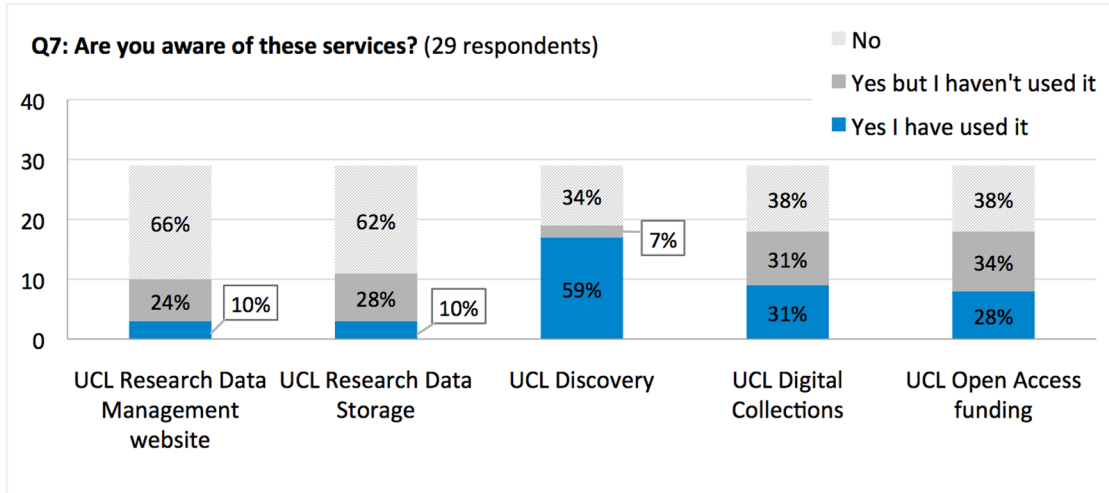
Selected detailed responses to the UCL Questionnaire

9.4.1 Faculty of Arts and Humanities

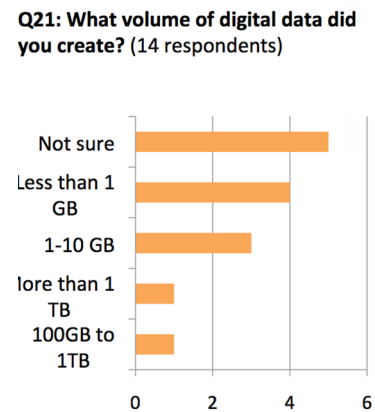
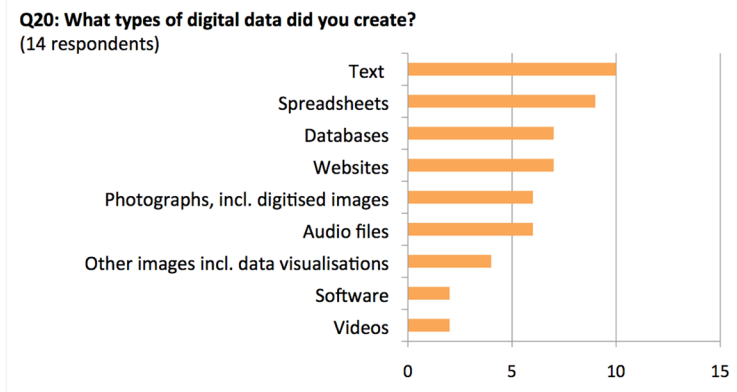
16 completed surveys were sent (out of 37 unique surveys transmitted) from researchers and research students in the Faculty. The overview below highlights responses to some of the key questions. It should be read after the Executive Summary and in conjunction with the whole report. The Research Data Management team is available to discuss the results.

⁴ See, for example, <http://courses.grad.ucl.ac.uk/>; last accessed 29 January 2017.

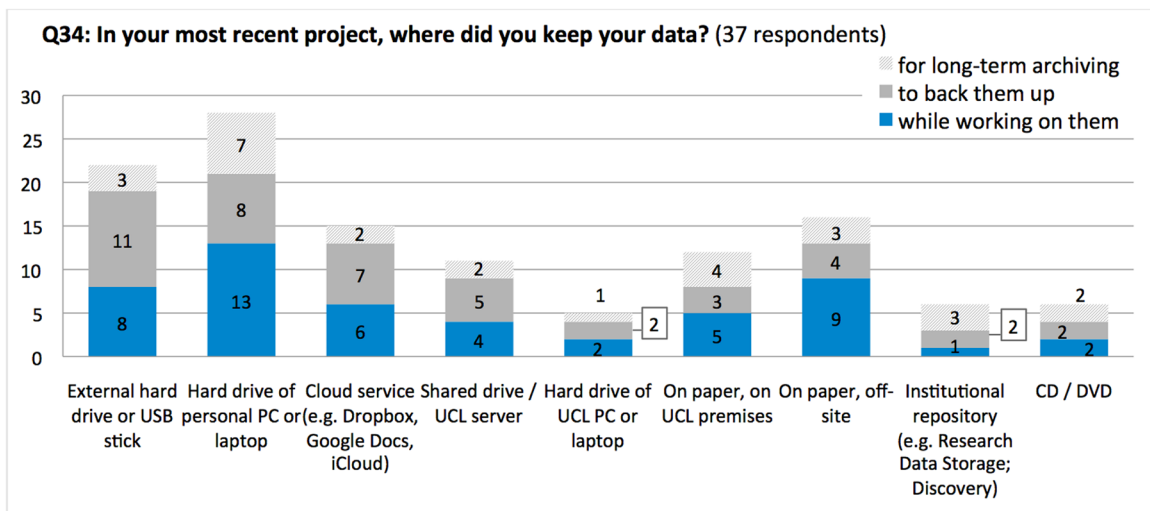
Awareness

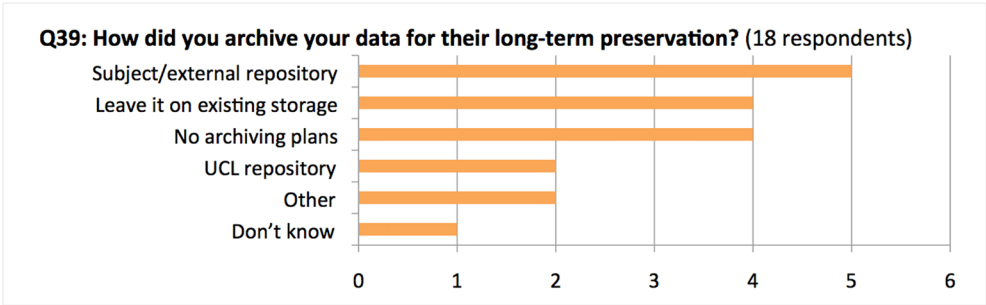


Creating & analysing data

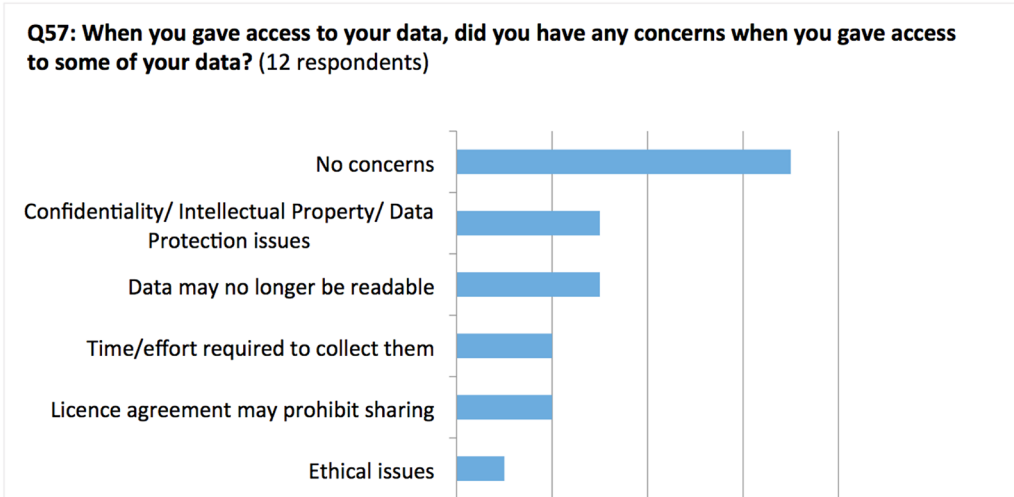
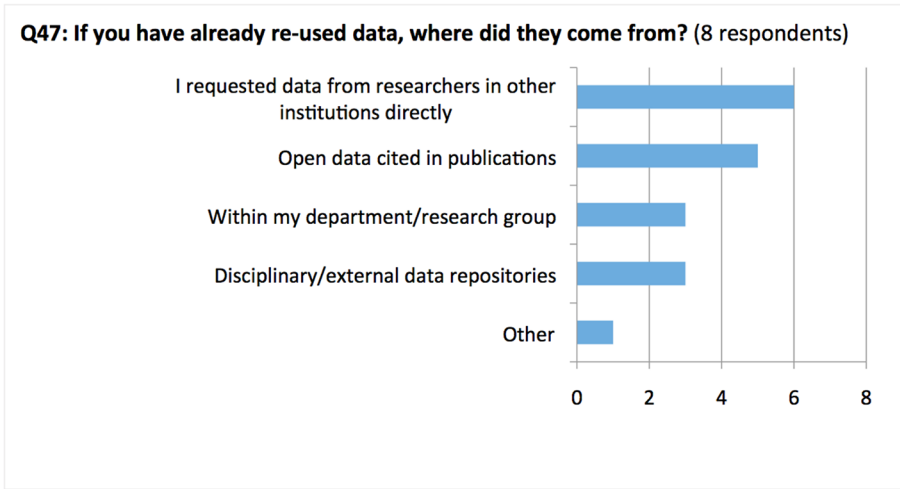


Storing & archiving data

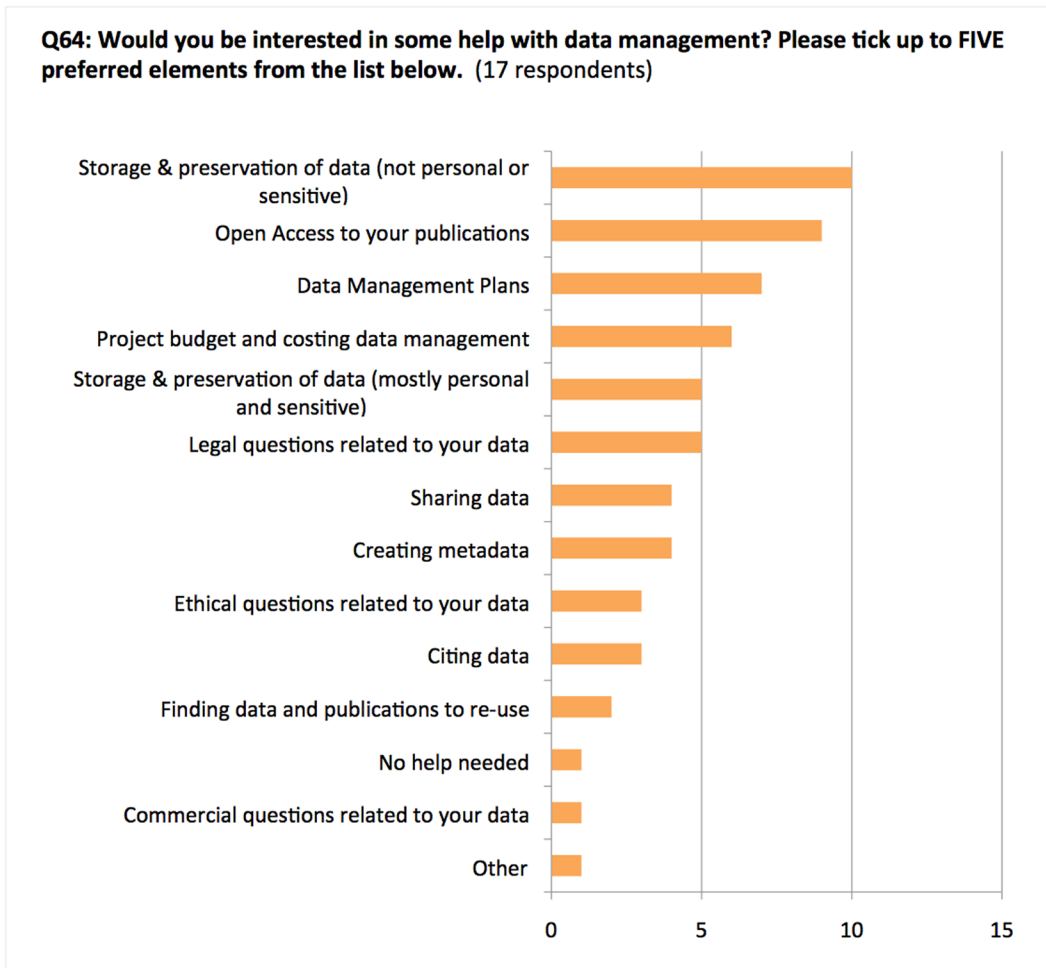




Re-using & sharing data



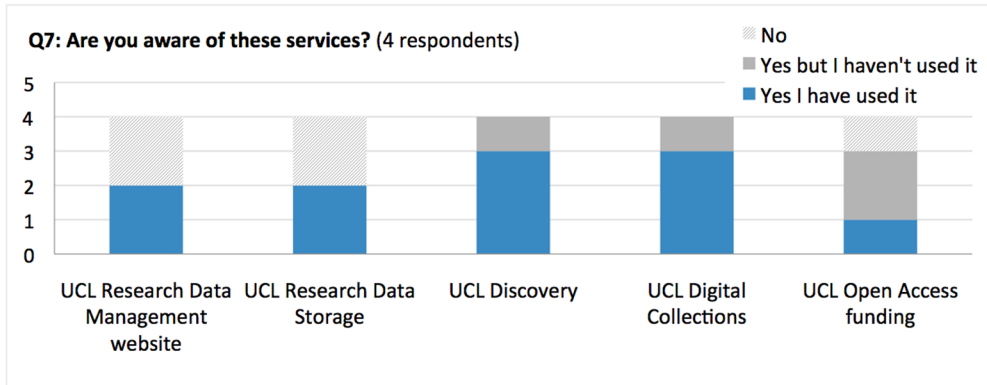
Support needed



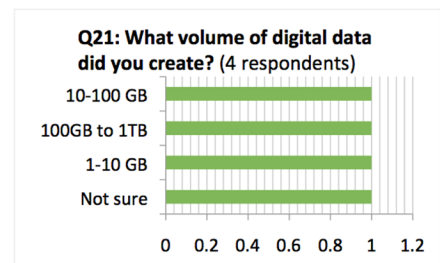
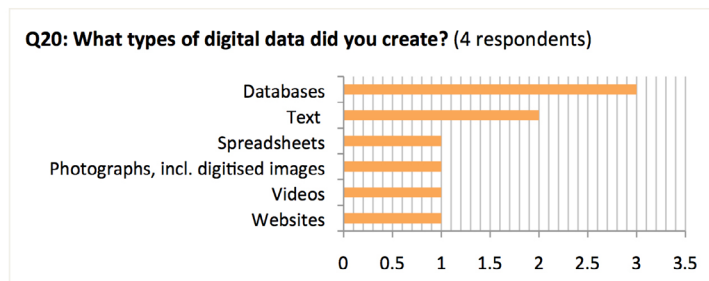
9.4.2 FACULTY OF LAWS

4 completed surveys were sent (out of 5 unique surveys transmitted) from researchers and research students in the Faculty. The overview below highlights responses to some of the key questions. It should be read after the Executive Summary and in conjunction with the whole report. The Research Data Management team is available to discuss the results.

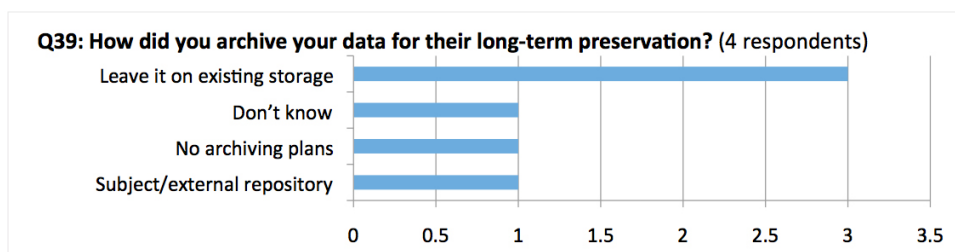
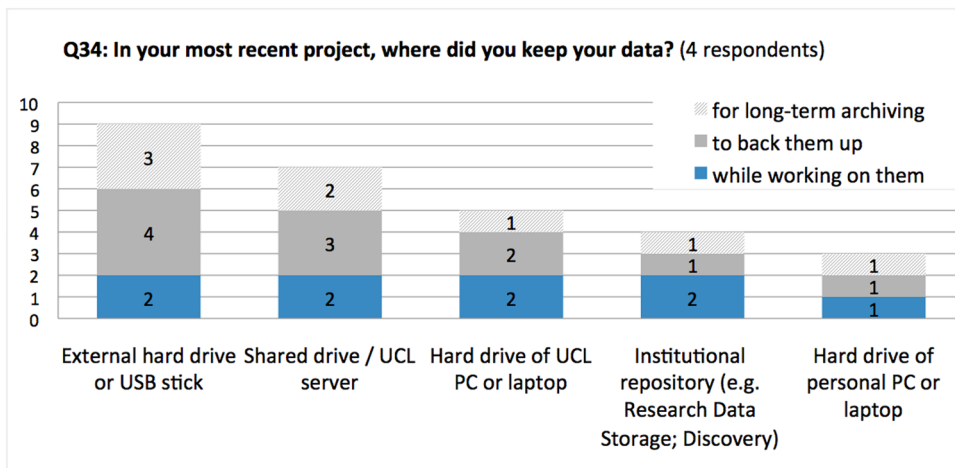
Awareness: policies, UCL services & Data Management Plans



Creating & analysing data

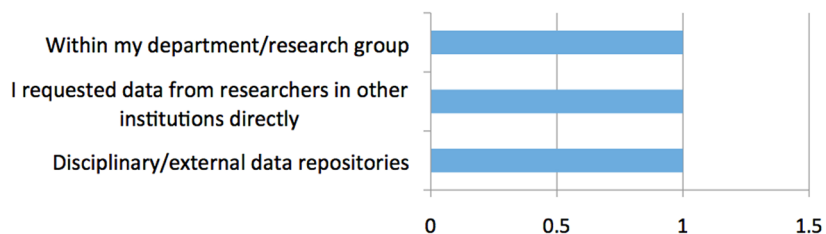


Storing & archiving data

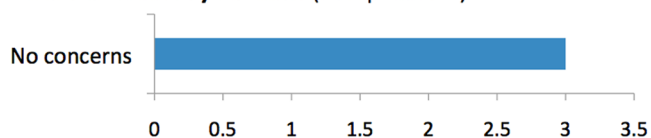


Re-using & sharing data

Q47: If you have already re-used data, where did they come from? (1 respondent)

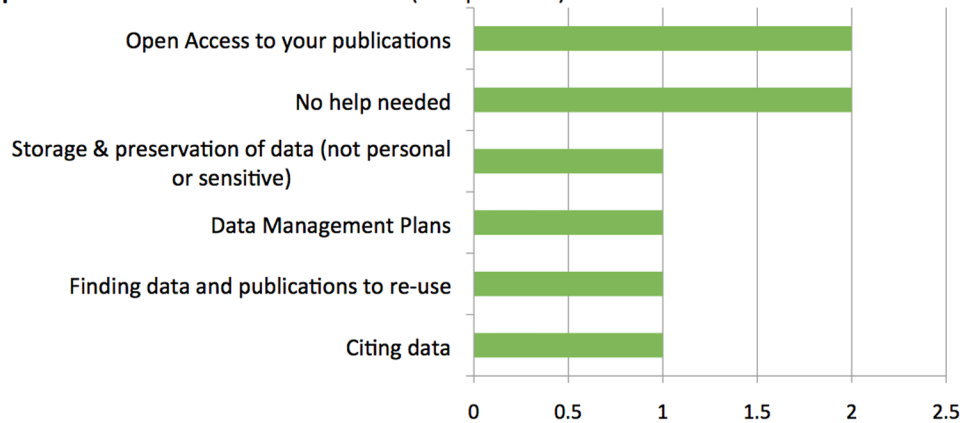


Q57: When you gave access to your data, did you have any concerns when you gave access to some of your data? (3 respondents)



Support needed

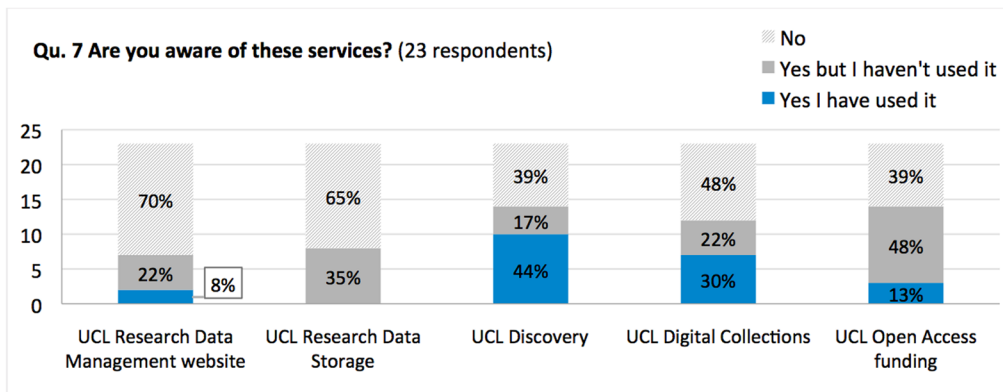
Q64: Would you be interested in some help with data management? Please tick up to FIVE preferred elements from the list below. (4 respondents)



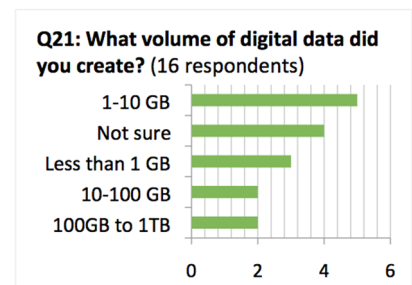
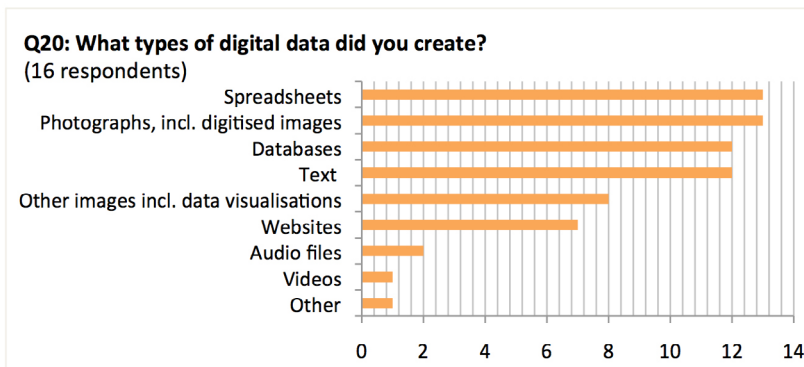
9.4.3 Faculty of Social & Historical Sciences

17 completed surveys were sent (out of 28 unique surveys transmitted) from researchers and research students in the Faculty. The overview below highlights responses to some of the key questions. It should be read after the Executive Summary and in conjunction with the whole report. The Research Data Management team is available to discuss the results.

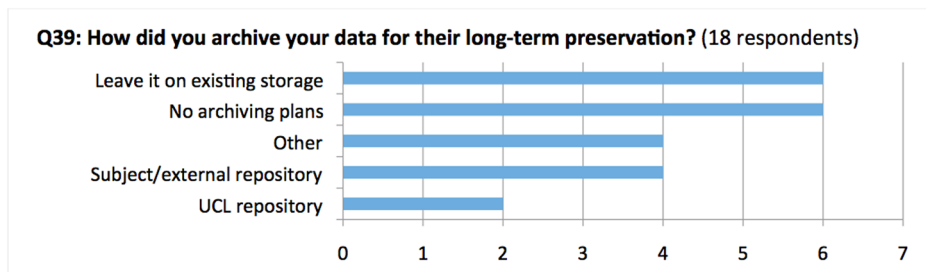
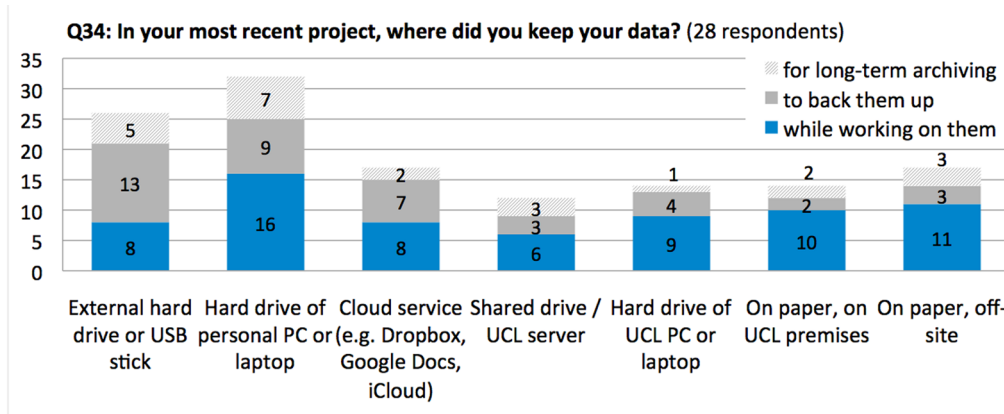
Awareness: policies, UCL services & Data Management Plans



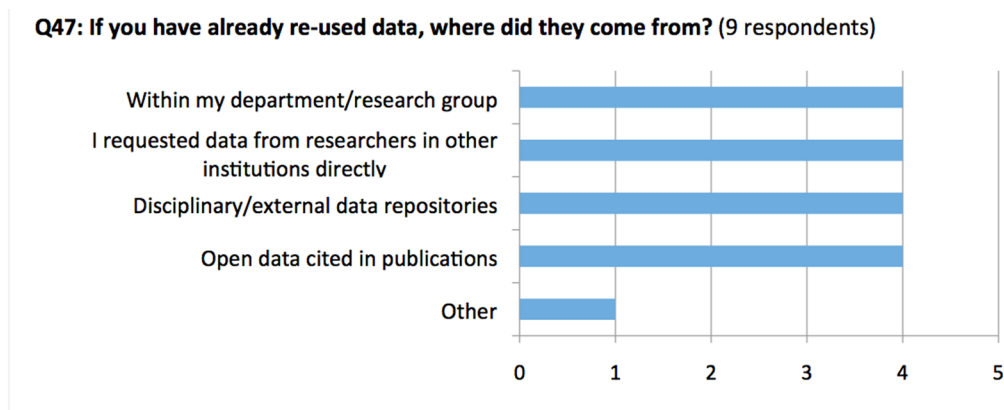
Creating & analysing data



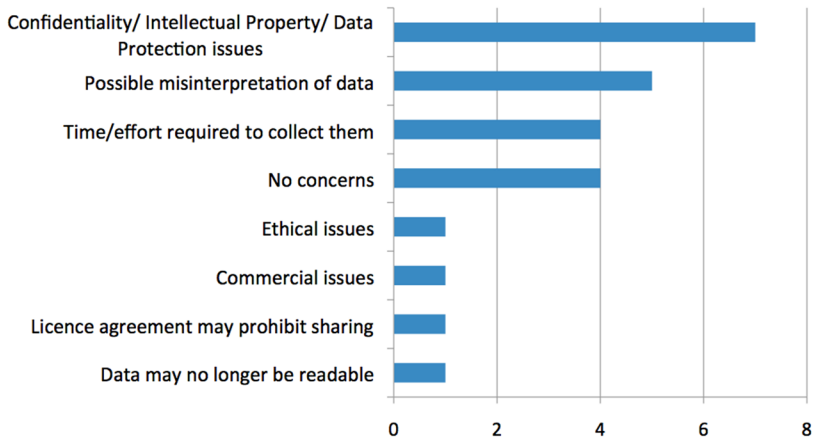
Storing & archiving data



Re-using & sharing data



Q57: When you gave access to your data, did you have any concerns when you gave access to some of your data? (12 respondents)



Support needed

Q64: Would you be interested in some help with data management? Please tick up to FIVE preferred elements from the list below. (17 respondents)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Case Study

10

RDM in the Performing Arts: Living Symphonies by Daniel Jones & James Bulley (Unit for Sound Practice Research, Goldsmiths, University of London)

Authors: James Bulley (Unit for Sound Practice Research, Goldsmiths, University of London) & Andrew Gray (Library Services, Goldsmiths, University of London)

Email: a.gray@gold.ac.uk

10.1. INTRODUCTION

*Living Symphonies*¹ is a landscape sound installation by James Bulley and Daniel Jones², which toured across four different forests³ in the UK in the summer of 2014. The work portrays the thriving activity of the forest's wildlife, plants and atmospheric conditions, creating an ever-changing sound symphony heard from a network of 24 speakers hidden throughout the forest itself. Working with ecologists and wildlife experts across the UK, Jones/Bulley developed highly detailed maps of the flora and fauna that inhabited each forest site where the installation was to take place.

Each species in the surveyed area was depicted by a unique set of musical motifs that portrayed their changing behaviour over day and night, coming to life as the species awakened; moving, developing and interacting just as the organism would. Dozens of these motifs were heard at any moment when the piece was live, spatialised across the space of the forest and heard back through a three-dimensional speaker system. In total there were some 15,000 fragments of sound within the sound score, making up musical movements for over a hundred different organisms.

10.2. FUNDER REQUIREMENTS

The piece was commissioned and funded as a collaborative work by Sound and Music, the Arts Council England and the Forestry Commission England. All copyright in the work, including that of the datasets, remained with the artists and there was no requirement to make any such data publicly available. A required outcome was a toolkit for touring public artworks, produced and published by the Forestry Commission England. This toolkit is openly accessible and available here⁴.

10.3. SURVEY DATA

In order to undertake the piece, the artists collected a large array of datasets over a year-long period of in-depth research and development. This data was used both to create and contextualize the artwork. A table of datasets captured during the project is shown in Figure 10.1.

¹ Living symphonies: <http://www.livingsymphonies.com>; last accessed 5 March 2017.

² James Bulley (b. 1984) and Daniel Jones (b. 1983) are an artist duo whose collaborative practice explores the boundaries of sound art, music, and process-based composition: <http://jones-bulley.com/biography/>; last accessed 5 March 2017.

³ The forest sites for the 2014 tour were as follows; Thetford Forest (24–30 May 2014), Fineshade Woods (20–26 June 2014), Cannock Chase (26 July–1 August 2014), and Bedgebury Pinetum (26 August–7 September 2014).

⁴ Sound and Music: <http://soundandmusic.org/create/planningandproducingartworksinthenaturalenvironmenttoolkit>; last accessed 5 March 2017.

Type of dataset	Format	Size	Capture Tools/ Software	Backup/Storage	Raw	Prepared	Shared with	Archived	Accessible
Detailed ecological surveys of four forest sites captured by teams of volunteers.	.xls / .csv	20 Mb	Google Docs / iPad	Google Drive / download onto Dropbox < > hard drives	Yes	No	Internal	Yes - onto hard drives	Not currently
Photographic surveys of the four forest sites (each metre square of the 30x20m forest sites), used for survey reference and as documentation	.tiff / .jpeg	30 Gb	Canon 6D DSLR camera / Photoshop	Dropbox < > hard drives	Yes	Yes as collage images of each site	Internal mainly (but some on website / used for press)	Yes - onto hard drives	Not currently
Illustrations of selected organisms by the artist Katie Scott. This material was used in visual explanations for the audience and in various marketing materials.	.tiff / .eps	400Mb	Adobe Illustrator / Hand drawn	Dropbox < > hard drives	Yes	Yes print versions etc available	Internal mainly (but some on website / used for press)	Yes - onto hard drives	Not currently
Reference field recordings of the four forest sites. These recordings were used to identify bird song and animal calls, and as an aid in balancing the sound mix of the piece.	.wav	30 Gb	Zoom H6, Tascam P2, DPA4060 Microphones / Pro Tools	Dropbox < > hard drives	Yes	Yes (edited selection available for use)	Internal only	Yes - onto hard drives	Not currently
Site and development documentation, across the four sites. This material was used as a documentation of the process of the creation of the piece for future papers and presentations about the piece.	HD film in .mov format and HR photography in .tiff format	10 Gb	Canon 6D DSLR camera / Photoshop	Dropbox < > hard drives	Yes	Yes (edited selection available for use)	Internal mainly (but some on website / used for documentation film)	Yes - onto hard drives	Not currently
Recordings of the scored musical fragments that make up the piece, derived from sessions with over 50 musicians in studios across the UK.	.wav / Pro Tools sessions (.pts) / Ableton Live sessions (.als)	150 Gb	Numerous microphones. Recorded into Pro Tools (.pts) as .wav files and then worked in Ableton Live (.als)	Dropbox < > hard drives	Yes	Yes (but only for use within the master score for the piece)	Internal only	Yes - onto hard drives	Not currently
3D software simulations of the forest sites created from the survey data	software package - python, C++, MaxForLive	500Mb	Custom software	Dropbox < > hard drives	Yes	/	Internal only	Yes - onto hard drives	Not currently

Figure 10.1 Table of datasets

10.4. BACKUP AND STORAGE

Working in remote forests across England was a challenge for capturing and storing data, as Internet/network access was extremely limited. As a result, the data was regularly backed up and duplicated onto hard drive storage, before then being synchronized to cloud storage at a later point. For immediate 'transfer' purposes all data gathered was placed into Dropbox (for sharing with partners including press organisations, Sound and Music and Forestry Commission England) and then transferred to external hard drive storage (copies were synced and held both at the Jones/Bulley studio and in personal artist studios offsite). Dropbox was used for its ease of use, stability and simple sharing interface.

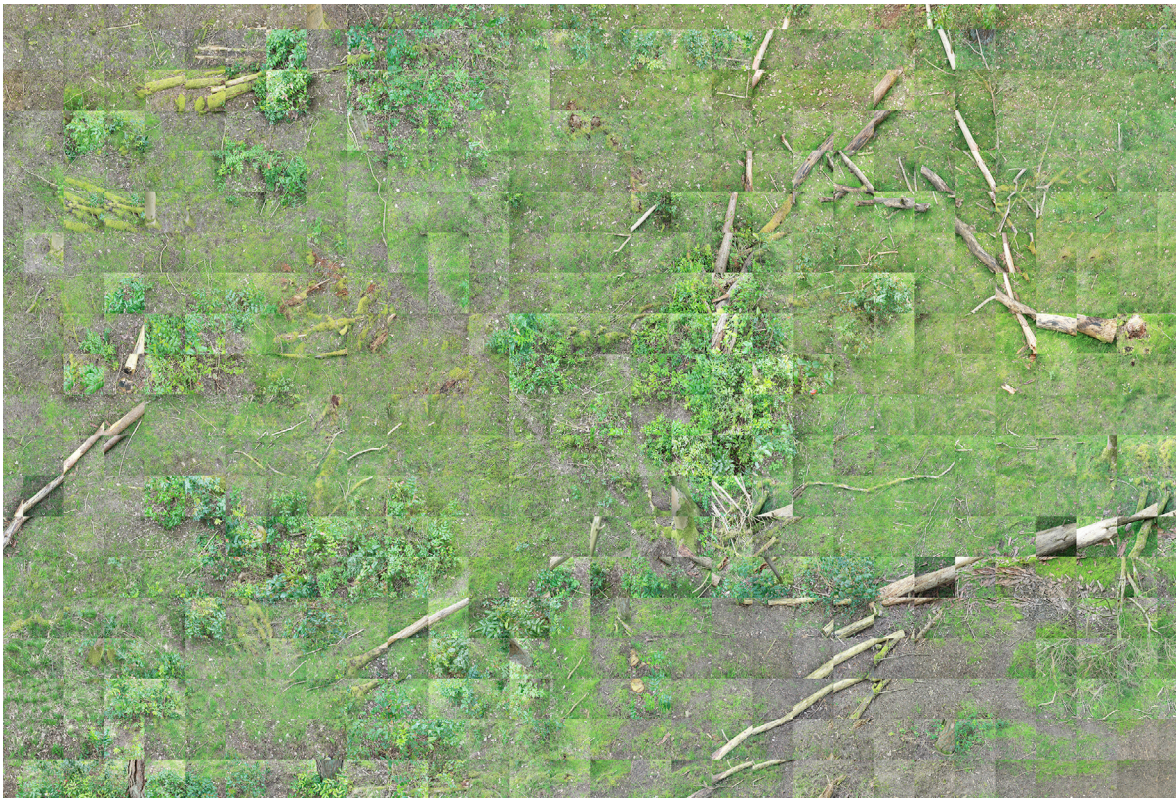


Figure 10.2 Thetford Forest Photographic Survey, 2014 (Photograph: James Bulley)

10.5. ANCILLARY DATA

During the live period as the installation toured, there were a number of additional datasets that were captured by the artists and the production team as part of the project.

A table of datasets captured during the project is included in Figure 10.3

Type of dataset	Format	Size	Capture Tools/ Software	Backup/ Storage	Raw	Prepared	Shared with	Archived	Accessible
Written testimonials (blogposts, handwritten feedback forms regarding audience experience)	.doc / paper	5Mb	Journal articles / written testimonies on paper	Dropbox < > Hard disks / physical backup in studio boxes	Yes	Yes	Some public, some internal only	Yes - onto hard drives	Not currently
Press articles and coverage (BBC news, Nature Journal video feature, Guardian feature etc)	.pdf captures	3Gb	Paparazzi .pdf screen capture software / print to .pdf function on Google Chrome. Videos as downloads (or sent in links from producers)	Dropbox < > Hard disks	Yes	Yes	Internal only	Yes - onto hard drives	Not currently
Video documentation of the sites (both with and without audience presence)	.mov HR files	150Gb	Canon 6D DSLR camera	Dropbox < > Hard disks	Yes	No	Internal only (possible future use)	Yes - onto hard drives	Not currently
Audio documentation of the piece live at each site	.wav files	50Gb	recorded on ZoomH6 with DPA4060 microphones (and various others)	Dropbox < > Hard disks	Yes	Yes (edited highlights selected and used on video documentation)	Internal only (possible future use)	Yes - onto hard drives	Not currently
Photographic documentation of the piece and the forest sites	.tiff files / .jpeg files	10Gb	Canon 6D DSLR camera	Dropbox < > Hard disks	Yes	Yes (edited highlights package created for press use and website use)	Internal only (possible future use)	Yes - onto hard drives	Not currently
Captures of the weather data	.csv files / .xls files	20Mb	Weather station through custom software	Dropbox < > Hard disks	Yes	No	Internal only	Yes - onto hard drives	Not currently

Figure 10.3 Table of ancillary datasets

Class	Behaviour Group	Family	Code	bbreviation	Species	Latin name	Scientific Family	Scientific Genus	Dominance	Radius	Length [1]	Wingspan	Speed [2]	Activity Pattern	Weather	Social Behaviour [3]	Food Sources	Flowers
Mammal	individual	Deer	M.01		Roe Deer	Capreolus capreolus	Cervidae	Capreolus	5	8	120		0.42 [4]	crepuscular, nocturnal		social (4)	grass, leaves, berries, ivy, hestiar	
					Fallow Deer	Dama dama	Cervidae	Dama										
					Reeves' Muntjac	Muntiacus reevesi	Cervidae	Muntiacus										
					Red Deer	Cervus elaphus	Cervidae	Cervus										
Mammal	individual	Fox	M.02		Red Fox	Vulpes vulpes	Canidae	Vulpes	4	8	70		0.5	nocturnal, crepuscular		territorial	mammals, birds, berries, nuts, worm	
Mammal	individual	Rabbit	M.03		European Rabbit	Oryctolagus cuniculus	Leporidae	Oryctolagus	3	8	38		3 [5]	nocturnal, crepuscular		solitary	grass	
Mammal	individual	Hare	M.04		European Brown Hare	Lepus europaeus	Leporidae	Lepus	3	8	60		4	nocturnal, crepuscular [6]		solitary	grass, berries, moss	
Mammal	individual	Badger	M.05		European Badger	Meles meles	Mustelidae	Meles	4	8	75		0.25	nocturnal, crepuscular		social (6)	worm, mammals, insects, reptiles, birds, berries	
Mammal	individual	Weasel	M.06		Least Weasel	Mustela nivalis	Mustelidae	Mustela	3	8	20		1.7 [7]	continuous		territorial	vole, mouse, frog	
					Stoat	Mustela erminea	Mustelidae	Mustela										
Mammal	individual	Mouse	M.07		Woodmouse	Apodemus sylvaticus	Muridae	Apodemus	2	8	9		0.9 [8]	nocturnal [9]	less active when cold/wet	solitary	seeds, berries, insects, worms, snails, fungus	
Mammal	individual	Shrew	M.08		Common Shrew	Sorex araneus	Soricidae	Sorex	2	8	7		0.9 [10]	continuous [11]		territorial [12]	snail, spider, worm, frog, mouse, vole	
Mammal	individual	Vole	M.09		Field Vole	Microtus agrestis	Cricetidae	Microtus	2	8	11		0.9	crepuscular, nocturnal [13]		solitary	grass	
					Bank Vole	Myodes glareolus	Cricetidae	Myodes										
Mammal	individual	Mole	M.10		European Mole	Talpa europaea	Talpidae	Talpa	3	8	14		0.9	continuous		territorial	worm, nuts	
Mammal	individual	Squirrel	M.11		Grey Squirrel	Sciurus carolinensis	Sciuridae	Sciurus	3	8	21		2.5 [14]	diurnal [15]		solitary	seeds, nuts, berries, fungus, bark	
Mammal	individual	Hedgehog	M.12		European Hedgehog	Eriaceus europaeus	Eriaceidae	Eriaceus	3	8	20		0.05 [16]	nocturnal		solitary	beetle, worm, snail	
Mammal	individual	Bat	M.13		Brown Long-eared Bat	Plecotus auritus	Vespertilionidae	Plecotus	4	8	40		5 [17]	nocturnal [18]		social (4)	fly, moth	
					Common Pipistrelle	Pipistrellus pipistrellus	Vespertilionidae	Pipistrellus										
					Soprano Pipistrelle	Pipistrellus pygmaeus	Vespertilionidae	Pipistrellus										
					Daubenton's Bat	Myotis daubentonii	Vespertilionidae	Myotis										
Reptile	individual	Lizard	R.01		Common Lizard	Lacerta vivipera	Lacertidae	Lacerta	2	8	10		0.2 (move in bursts of 0.5-1.5s)	diurnal	bask in sun; speed and pause duration change logarithmically below 25C (0.01@5C - 0.2@25C; 3s@5C - 0.1s@25C) [19]	solitary	insects	
					Grass Snake	Natrix natrix	Colubridae	Natrix										
Reptile	individual	Snake	R.02		Common European Adder	Vipera berus	Viperidae	Vipera	3	8	80		0.5	diurnal	bask in sun; speed 0.25 m/s @ 15C; 0.45m/s @ 30C	solitary	reptiles, mouse, vole, shrew	
Reptile	individual	Slow Worm	R.03		Slow Worm	Anguis fragilis	Anguidae	Anguis	2	8	35		slow moving unless startled	diurnal	bask in sun	solitary	worm, spider, insects, snail	
Reptile	individual	Frog	R.04		Smooth Newt	Amphiba	Salamandridae	Lissokliron	2	8	8			nocturnal [20]	diurnal during wet weather	solitary	snail, spider, insects	
					European Common Frog	Rana temporaria	Ranidae	Rana										
Bird	individual	Wren	B.01		Wren				2	8	10	15	6.5 [21]	diurnal		solitary	insects	
					Treecreeper													
					Nuthatch													
Bird	individual	Robin	B.02		Robin	Muscicapidae			2	8	14	21	11	diurnal		solitary	insects, worm [22]	
Bird	individual	Finch	B.03		Bullfinch				2	8	15	26	11	diurnal		territorial	insects, seeds	
					Chaffinch													
					Greenfinch													
					Goldfinch													
					Hawfinch													
Bird	individual	Tit	B.04		Crossbill				2	8	12	22	8	diurnal		social	insects, spiders	
					Great Tit													
					Blue Tit													
Bird	individual	Goldcrest	B.05		Long-tailed Tit				2	8	9	15	6.5	diurnal			insects, spiders	
					Coal Tit													
Bird	individual	Thrush	B.06		Goldcrest				2	8	9	15	6.5	diurnal			insects, spiders	
					Firecrest													
Bird	individual	Thrush	B.06		Songthrush				2	8	23	36	11 [23]	diurnal			insects, worms, spiders, seeds	
					Mistlethrush													
Bird	individual	Blackbird			Blackbird													
Bird	individual	Nighthjar	B.07		European Nighthjar	Caprimulgidae			3	8	26	60		crepuscular, nocturnal [24]	prefer warm, dry, still nights		moth, fly, dragonfly	
Bird	individual	Warbler	B.08		Blackcap				2	8	11	19		diurnal			insects, berries	
					White Throat Warbler													
					Garden Warbler													
					Willow Warbler													
					Grasshopper Warbler													
Bird	individual	Dove	B.09		Common Chiffchaff													
Bird	individual	Woodpecker	B.10		Stockdove				2	8	40	70		diurnal		gregarious	berries, nuts, seeds, insects	
					Wood Pigeon													
Bird	individual	Woodpecker	B.10		Green Woodpecker				3	8	24	40		diurnal [25]		solitary	ants, nuts, seeds, berries, leaves	
					Great Spotted Woodpecker													
					Lesser Spotted Woodpecker													
Bird	individual	Magpie	B.11		Jay				2	8	40	60		diurnal			insects, mouse, vole, shrew, berries, nuts	
Bird	individual	Magpie	B.11		Magpie				2	8	40	60		diurnal			insects, mouse, vole, shrew, berries, nuts	
					Garrulus - Pica Pica													
Bird	individual	Cuckoo	B.12		Cuckoo	Cuculidae			3	8	32	58		diurnal			spider, beetle, moth, butterfly, insects	
Bird	individual	Rook	B.13		Rook	Corvus frugilegus			3	8	45	80		diurnal			worms, insects, seeds	
Bird	individual	Jackdaw	B.14		Jackdaw	Corvus monedula			3	8	33	70		diurnal			insects, worm, mouse, berries	
Bird	individual	Crow	B.15		Common Crow	Corvus corone			3	8	46	100		diurnal			insects, worms, mouse, vole, berries	
Bird	individual	Pheasant	B.16		Golden Pheasant				3	8	30	85		diurnal			seeds, berries, leaves, insects	
					Common Pheasant													

10.4 Excerpt from Living Symphonies full organism survey, 2014

Flowers	Berries	Nuts	Thetford	Fineshade	Cannock	Bedgebury	Composition Notes	Instrumentation	Composition description	Players	Tone Row	Hz	Root BPM	BPM range	Metre
			3	1				Trombone, Timpani	French Horn prepared techniques, dynamic, interlocking patterns with timpani - opening of Javanese Sinfonietta?	Hywel Jones, James Bulley	[Bb], C, D, Eb, F, G, A	87.3-1000	105	95-115	3/4
			3		3	2		Euphonium	clarinet melodies - rhythmic, complex not obvious, jump structure, trumpet swoops of prepared technique timbre material in harmony, trombone low soft rhythmic notes underlay	Hywel Jones	[F], G, A, Bb, C, D, Eb, E	140-1900	110	100-120	4/4
			3		3	1		Euphonium (extended techniques)	Euphonium - detailed interlocking patterns of slide sounds on 1st, 3rd then 4th... melody from breath sounds on long drawn out high held notes (top end of range)	Hywel Jones	[D], E, F, G, A, Bb, C	140-1300	115	105-125	3/4
			3					Female Voice, Alto	light, counter rhythms, leaping parts, gissandi	Havva Basto, Laurel Sita	[A], B, C, D, E, F, F#, G, G#	164-880	70	60-80	4/4
			3		2	3		5-part male voice choir with off-beat percussive rhythms and tambourine	7/4ish rhythm, punctuations by clap, lunch winding eastern sounding grouped voice melody, fast counterpoints, SF above variations etc, tambourine to add to clap over time, shake to end to percussion over time also	Milo Fitzpatrick (D Bass), Tenor Voices, Rosie Bergenzi, JJBDDJJ	[C], Bb; A, C, D, F	50-1200	90	80-100	5/4
			3		1	3		Alto Clarinet	Glissando, leaping clarinet	Charly Richardson	[F], G, Ab, Bb, C, D, E	800-5,000	110	100-120	4/4
			3		1	3		Extended percussive improvised fragments using the keys of the clarinet	filtered to be fairly high frequency... rhythmic	Charly Richardson	Non-tonal	1200+	95	85-105	4/4
			3		3	3		Piano prepared	Short, punctuated fragments of high pitched sound, glimmering, sharp	Keir studio - Prepared piano - Keir Vine	?	1200+	117	117	3/4
			3		2	2		Prepared Chimes, Prepared Piano, metallophone	Short, punctuated fragments of high pitched sound, glimmering, sharp	Keir Vine	Non-tonal	1200+	110	100-120	4/4
			3		3	1		Hong and Steel Drum chords and rhythms	Chance based rhythms, with underlay of prepared sustained textures and pad (from contact mic scrape material)	Keir Vine	Bb, Eb, G, F, D, B, E	whole	110	100-120	4/4
			3	2	3	3		Tambourine and Castanet phase rhythms	scraping, rattling, bright, fast, long varying rhythm motifs, complex - textural backing by prepared techniques, heavily generative, quite rapid patterns	Rosie Bergenzi, JJBDDJJ, Keir Vine	Non-tonal	1000-10,000	110	100-120	7/4
			2	2	3	1		Soprano Sax & Cello quartet	short rhythmic breath like patterns, layered, tonal, simple quiet melody on clarinet above - mid range pitch-wise	Charly Richardson, Peter Gregson	[C], C, D, A, B, E, F#	1000-7,000	115	105-125	4/4
			3		3	3		Glass Harmonica duet	based on vespers, textural, synthetic	James Bulley	Ab, B, Eb; C, E, Gb	2,500-10,000	120	110-130	4/4
			3		2	2		Double-stopped violin and extended techniques including glissando and harmonics	double stopped, dragged out light motifs, scatty, extended techniques, harmonics, ghost notes, scrapes	Simon	[D], F#, A, E, G, B, C#	198-4,400	95	85-105	4/4
			3	3	3	1		Double-stopped cello and extended techniques including glissando and harmonics	Slides, gliss.	Peter Gregson	[C], A; E, G, B	100-4,000	90	80-100	4/4
						1		Cello	double bass harmonics, dragged gritty notes - sustained, bowed / cello extended techniques, bow etc sounds etc & harmonics - low pitch range for tonal sounds	Peter Gregson	[C], E, G, A, B	100-4,000	90	90	4/4
			3		3	2		Peter Gregson	extended techniques, reed flutters, melodic ideas, gutural low end of the oboe	Cello	[Bb], Db, F, C, Eb, Gb, A	100,4,000	100	90-110	3/4
			3		3	3		Concert flute duet, short descending melodies and appoggiato sequences	short melodic composition - short melodic on flute, processed live, reedy, addition of breathy textures, overtones, timbral - spectral	Katie English	[D], G, A; E, C	>	100	90-110	4/4
			3		3	3		Short clarinet motifs, rapid trills and short melodies	melodies on flute, processed live, reedy, addition of breathy textures, swoops, timbral, spectral	Charly Richardson	[Eb], G, Bb; Ab, C, D, F	>	100	90-110	4/4
			3	2	3	3		Arabic influenced legato flute melodies		Katie English	[A], B, C, E; F, G, D	>	100	90-110	4/4
			3		3	3		Cello trills, extended techniques, rapid melodic motifs		Peter Gregson	[Bb], Db, F; Eb, Gb, A, C	>	110	100-120	4/4
			3	3	3	3		Soprano Saxophone melodies and rhythms		Charly Richardson	[F#], B, C#, F; G#, A#, D#	>	100	90-110	4/4
			3	3	3	3		Ascending and descending violin melodies with counterpoint		Simon Hewitt Jones	[G], D, F#, A, B, C, E, F,	>	100	90-110	4/4
			3					Tuba, Euphonium	spectrally composed, textural - extended techniques on both instruments, overblows, breath timbres, sung notes etc...	David Aird, Hywel Jones	[Bb], C, D, Eb, F, G, A	>	110	100-120	4/4
			3	3	3	3		Bass Flute melodies and counterpoints	spectrally composed, textural - extended techniques on both instruments, overblows, breath timbres, sung notes etc...	Katie English	[Bb], D; C, Eb, F, G, A	>	95	85-105	4/4
			3	3	3	3		Clarinet quartet with tuba and euphonium	reedy tone, swoops or raves - tonal harmonics, but dampened (sostenuto/practice pedal), prepared slightly? Very light dipping short melodic phrases above in mid-range	Hywel Jones, David Aird, Charly Richardson	[Bb], F, A, Eb, G, C	500-4,000	100	80-100	4/4
			3		3	3	Lesser Spotted is very rare	Rapid marimba rhythms, with cabasa counterpoint on the off-beat	Not mapped off woodpecker rhythms, but a much grayer textural version built from 4-5 layers of interlocking rhythms	Keir Vine	D#, G#, A#, C,	500-4,000	110	100-120	4/4
			3		3	3		Extended techniques on the harmonium, mechanical percussive sounds and pedal wheezing	Extended techniques	Keir Vine	Non-tonal	100-5,000	110	100-120	4/4
			1		3	3		Flute melodies with flutter-tongued trills and rhythmic extended techniques on flute keys	Sad, plaintive, similar to messian's oration. A play on the cuckoo's call... slow ponderous... possible duet with cello chorist?	Katie English	[C#], G#, E; D#, F#, A, B, C	800-1400	95	85-105	4/4
			3		3	3		Accordion	Wheezing, grabbed chords, half melodies, croak - extended techniques	James Bulley	?	100-8,000	110	100-120	4/4
			3		3	3		Harmonica	motif led	Theo Lampert-Crook	? A, C, D, F	100-8,000	100	90-110	4/4
			3		3	3		Melodica	Minor	Theo Lampert-Crook	[C], D, Eb, F, G, Ab, Bb	100-8,000	110	100-120	3/4
			3					Trombone	Sung & played notes (extended techniques)	Hywel Jones	Bb, F, D, A	100-8,000	105	95-115	4/4

10.4 Excerpt from Living Symphonies full organism survey, 2014

[This taxonomy details every living organism (in genus groups) and its related music across all four sites of the 2014 tour of *Living Symphonies*]

10.6. SHARING OF DATA

The sharing of the data that underpins *Living Symphonies* has been a complex and near impossible task. Whilst the partner organisations did create a toolkit that explored the touring of the piece (which was a prerequisite of the Arts Council funding that the piece obtained), it has not been possible to make available the vast majority of the above data in any coherent way. It is clear that most of this data would be very useful to many other researchers and artists (as proven by the interest of numerous academics, musicians and ecologists). However, in order to achieve this there would need to be funding allocated to provide the time for the adequate preparation of the datasets with related material to explain and contextualise them. Some of the photography and video has been used to make short reference films and to provide visual context to document the occurrence of the work, but it has not been possible for the artists to make the following datasets available due to a lack of funding, time constraints surrounding its curation and contextualization, i.e. ranges of data and editing of documentation material, and issues in hosting such large quantities of material. Bracketed after these datasets are the avenues that the artists would hope and plan to make the material available through if possible:

- *forest survey data* (Goldsmiths Data Repository – data.gold.ac.uk, livingsymphonies.com)
- *field recordings* (Goldsmiths Data Repository – data.gold.ac.uk, freesound.org)
- *weather datasets* (Goldsmiths Data Repository – data.gold.ac.uk, livingsymphonies.com)
- *photography* (Goldsmiths Data Repository – data.gold.ac.uk, flickr.com)
- *film* (Goldsmiths Data Repository – data.gold.ac.uk, livingsymphonies.com)
- *custom unique software* (Goldsmiths Data Repository – data.gold.ac.uk, [github](https://github.com))
- *sound score materials* (Goldsmiths Data Repository – data.gold.ac.uk, freesound.org)

10.7. CONCLUSION

Whilst much discussion has occurred in recent years surrounding research data management in the context of science-centred and text-based research outputs, very little of this has involved confronting the problems facing artist-researchers working outside these areas. As a result of fundamental differences in the commissioning and funding structures for art projects, there is insufficient funding and understanding on the part of the artists and institutions involved as to how or even why it is worth making this data available. *Living Symphonies* provides a case study that highlights a large and wide-ranging array of datasets that would undoubtedly be useful for researchers across numerous disciplines. In this instance the artists/researchers are comfortable with the vast majority of the data being made available under one of the more openly accessible of Creative Commons licenses – in this instance this would not affect any further income for the artists as the pieces in themselves are unrepeatable due to their site-specific nature. The artists believe this would be the right thing to do, given the publicly funded nature of the project. This data will remain unavailable unless there is adequate funding and planning from the outset for projects such as these.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Section 4

Open Data

Case Study

11

Why Open Data?

Author: Professor Geoffrey Boulton (University of Edinburgh and President of the International Council for Science's Committee on Data for Science and Technology)

Email: G.Boulton@ed.ac.uk

11.1. THE DIGITAL REVOLUTION

At about the turn of the millennium, the global volume of data and information that was stored digitally overtook that stored in analogue systems on paper, tape and disc. The result has been a digital revolution, with the global data acquisition rate now 40 times greater (35×1000^7 bytes) than 10 years ago, still accelerating and driven in part by the massive reduction in the cost of digital storage. In 2003, the human genome was sequenced for the first time. It had taken 10 years and cost \$4 billion. It now takes 3 days and costs \$1,000.

The unprecedented rate that we are able to acquire, store, manipulate and instantaneously communicate vast amounts of digital data and information has profound implications for all fields of science and scholarly research as well as for economies and societies. It is crucial that these implications are explored to the maximum effect by the research and scholarly communities in all parts of the world. Part of the opportunity lies in exploiting "Big Data", where enormous fluxes of data stream into computational and storage devices, often from a great diversity of sensors and sources; in "Linked Data", where semantic linking between different datasets opens opportunities for eliciting much deeper meanings (of great potential relevance for many global challenges such as infectious disease, disaster risk reduction and migration); in the myriad opportunities that arise from blending the physical and digital realms through the "Internet of Things"; and in the powerful but problematic potential of machine learning. The fundamental benefits derived from these approaches are in elucidating patterns and relationships that have previously been beyond our capacity to resolve and both to characterize and to simulate the dynamics of complex systems.

11.2. SCIENCE¹ AS AN INHERENTLY OPEN ENTERPRISE

Openness has been the bedrock on which modern science has been built. The rules of the game were established in the late seventeenth century, when scientific ideas began to be published in open journals rather than hidden in the private correspondence of gentlemen. A further crucial step was the requirement by journal editors that truth claims must be accompanied by the evidence (the data) on which they were based. This permitted others to attempt replication of the observational or experimental evidence and to scrutinise the logic of the proposed relationship between evidence and concept. Failure on either count indicated error. It is a process termed "self correction" by historians of science, tellingly characterised by Arthur Koestler in writing: "The progress of science is strewn, like an ancient desert trail, with the bleached skeletons of discarded theories that once seemed to possess eternal life". If there is a scientific method, this is it, the power of the negative. Albert Einstein characterised it as: "No amount of experimentation can prove me right. A single experiment can prove me wrong."

¹ The word science is used here to mean the systematic organisation of knowledge that can be rationally explained and reliably applied. It is used, as in most languages other than English, to include all domains, including humanities and social sciences as well as the STEM (science, technology, engineering, medicine) disciplines.

11.3. THE BRIGHT SIDE

Like all revolutions that have not yet run their course, it is often difficult to distinguish reality and potential from hype. But powerful, real discoveries have now emerged in the elucidation of previously unsuspected patterns and relationships. In genomics, rapid sequencing and advanced computing power permit systematic testing of relationships between genetic variations and specific traits and diseases, rather than using trial and error, with profound implications for medicine, agriculture, the production of biofuels and the process of drug discovery. The advent of the modern computer has long permitted simulation of the dynamics of highly coupled complex systems, their sensitivity to small variations in initial conditions and their capacity to produce “emergent behaviours” that were not evident from their individual components. We can now add to this by the use of big, linked data to characterise complexity, and by iterating between characterisations and simulation, to follow and forecast the evolution of complex systems, as is now done in modern high-resolution weather forecasting. Only however if data is routinely made “intelligently open” (accessible, intelligible, assessable and re-usable),² can the full benefit of such approaches be realised.

11.4. THE DARK SIDE

However, the vast and complex data volumes that many scientists are now able to access also challenge the open approach required for self-correction. This arises from the difficulty of making such data sets open to scrutiny, together with the metadata, the computer code used in analysis, and the logic of any “learning machine” used in the process. It is hardly surprising that many of us fail this standard, or have succumbed to the temptation to keep our data under wraps so that it can be milked again for further publications. A current debate in the *New England Journal of Medicine*³ about the rights and wrongs of openness in medical research epitomises this conflict; between the public interest in openness and the interests of scientists’ careers in maintaining data ownership. Moreover, the recent attempts to replicate the results of highly regarded papers, in areas as diverse as pre-clinical oncology, social psychology and economics, with replication rates never exceeding 25%, illustrate the consequences of not rigorously presenting all the data and metadata. Without this, self-correction cannot work. If we are to maintain the credibility of the scientific process, we need to regard absence or inadequate presentation of data and metadata as scientific malpractice and to re-establish standards of reproducibility for a data-rich age. Without this we run the risk of the digital explosion overwhelming the processes that ultimately maintain scientific rigour.

11.5. ADAPTING TO CHANGE

Information and knowledge have always been essential drivers of human material and social progress, and the technologies by which knowledge is stored and communicated have been determinants of the efficiency of these processes. The digital revolution is a world historical event as significant as Gutenberg’s invention of moveable type, and certainly more pervasive. A crucial question for the research and scholarly community is the extent to which our current habits of storing and communicating data, information and the knowledge derived from them are fundamental to creative knowledge production and its communication for use in society, irrespective of the supporting technologies, or whether many are merely adaptations to an

² The Royal Society (2012), *Science as an open enterprise*. Royal Society: <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>; accessed 5 February 2017.

³ STAT: <https://www.statnews.com/2016/08/10/data-sharing-science-nejm/>; accessed 5 February 2017.

increasingly outmoded paper/print technology. Do we any longer need expensive commercial publishers as intermediaries in the communication process? Do conventional means of recognising and rewarding research achievements militate against creative collaboration? Has pre-publication peer review ceased to have a useful function? These are non-trivial questions that need non-trivial responses.

Both individuals and institutions need to adapt. The recently published Accord on Open Data⁴ sets out principles and responsibilities. It advocates a normative principle at the level of individuals:

“Publicly funded scientists have a responsibility to contribute to the public good through the creation and communication of new knowledge, of which associated data are intrinsic parts. They should make such data openly available to others as soon as possible after their production in ways that permit them to be re-used and re-purposed.”

and an operational principle that:

“The data that provide evidence for published scientific claims should be made concurrently and publicly available in an intelligently open form. This should permit the logic of the link between data and claim to be rigorously scrutinised and the validity of the data to be tested by replication of experiments or observations.”

A positive reaction to the Accord from the International Union of Crystallography⁵ included an even stronger clarion call to action:

“We urge the worldwide community of scientists, whether publicly or privately funded, always to have the starting goal to divulge fully all data collected or generated in experiments.”

Such statements from the global research community about the open ethos of scientific inquiry, and its relevance to the need of humanity to use ideas freely, should be echoed by universities as part of their traditional role in preserving, re-assessing and creating knowledge and communicating it, in questioning received wisdom rather than blandly regurgitating it. They are also important in combating a countervailing trend towards the privatisation of knowledge, of which some universities are part, by succumbing to injunctions to see themselves largely as instruments of national wealth creation, where intellectual output is marketable property rather than public good. In contrast, the technologies at our fingertips have a key enabling potential for “open science”, in which publicly funded science is done openly, its data are open to scrutiny, its results are available freely or at minimal cost, and results and their implications communicated more effectively to a wide range of stakeholders. Moreover scientific knowledge ‘producers’ should cease to think of knowledge ‘users’ as passive information receivers, or at best as contributors of data to analyses framed by scientists, but potentially as respected allies in the co-framing of issues and the co-production of actionable knowledge⁶.

⁴ Science International 2015: Open Data in a Big Data World; available at www.science-international.org; accessed 5 February 2017.

⁵ ICUr: *Open Data in a Big Data World: A position paper for crystallography*; available at <http://www.iucr.org/iucr/open-data>, accessed 5 February 2017.

⁶ Hackmann, Heide and Boulton, Geoffrey: *Science for a sustainable and just world: a new framework for global Science policy?* UNESCO World Science Report 2015, pp. 12-14; available at UNESCO: <http://unesdoc.unesco.org/images/0023/002354/235406e.pdf>; accessed 5 February 2017.

11.6. INFRASTRUCTURES FOR OPEN DATA

Whilst universities must respond to these ethical challenges in their own ways, they must also respond to the need to manage their data in ways that they believe to best reflect their mission. Several years ago, rigorous data management was seen by many universities merely as a cost, as an “unfunded mandate”. Increasing numbers of universities now see open data as a necessary part of their future and plan to position themselves to exploit the opportunities that it offers. Some of the essential principles of good research data management have now been established as a result of hard won experience^{7 8}, many of which are shared in this volume. The “hard” infrastructure of high performance computing or cloud technologies and the software tools needed to acquire and manipulate data in these settings are only part of the problem. Much more problematic is the “soft” infrastructure of national policies, institutional relationships and practices, and incentives and capacities of individuals. For although science is an international enterprise, it is done within national systems of priorities, institutional roles and cultural practices, such that university policies and practices need to accommodate to their national environment. The iceberg figure reflects this (figure 11.1). The easy part is the visible part comprising the hardware and software tools required by a national open data system and any consents required for data use. Below the surface lie issues of process and organization. What is the ecology of the national research system? Do funders recognize and respond to the open data imperative? And is there adequate support for data management, data science advice and training? Then there are the people. Do they have the skills required to exploit the potential of the digital revolution? Are there incentives for researchers to make their data intelligently open? And does the mindset of a researcher accept the ethos of the first principle in the Accord?

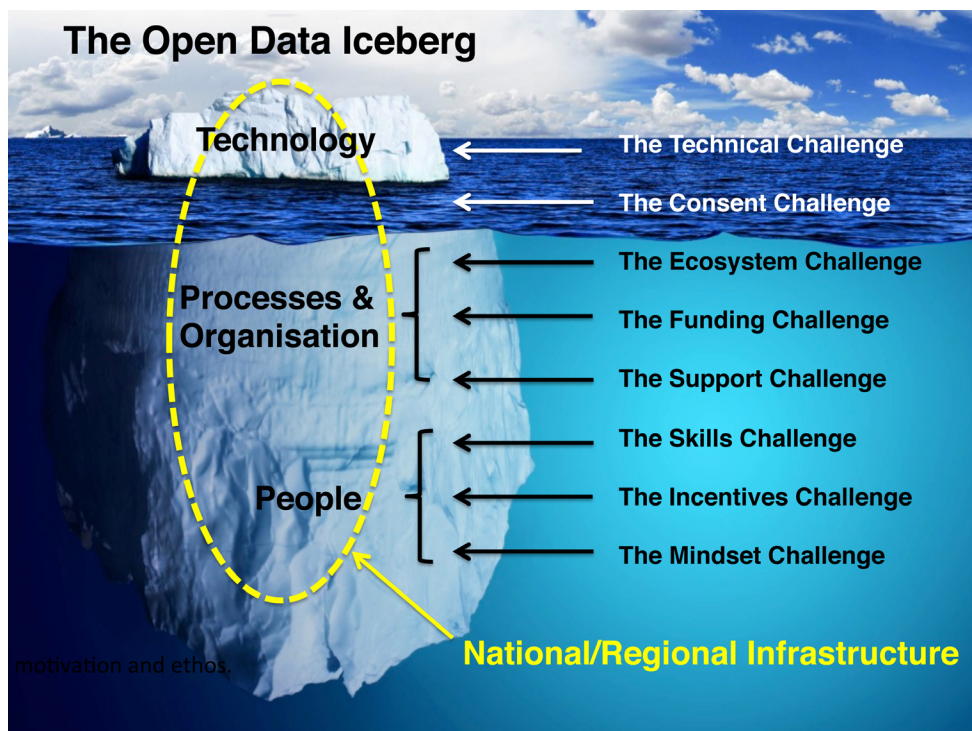


Figure 11.1

⁷ CODATA, 2015: *Current Best Practice for Research Data Management Policies*; available at <http://dx.doi.org/10.5281/zenodo.27872>; accessed 5 February 2017.

⁸ LERU Roadmap for Research Data: <https://www.fosteropenscience.eu/content/leru-roadmap-research-data>; accessed 5 February 2017.

There are, however, important developments in support of open data beyond the confines of the university, with which universities can engage to their considerable benefit, if only to relieve themselves of the burden of being data management islands. “Open data platforms” are currently being developed where the needs of users are matched with hardware/software provision and data managerial skills, and created within individual disciplines (e.g. US National Institute of Health⁹, Elixir-europe programme¹⁰) or multi-national geographic regions (e.g. Open Science Platform for Africa; Latin America and the Caribbean Platform; European Open Science Cloud).

11.7. A DECADAL VISION

Nearly two decades ago, Tim Berners-Lee proposed that datasets that relate to the same or related phenomena could be semantically linked in ways that integrate different perspectives,¹¹ and thereby offer much deeper understanding than merely using the web as a means of retrieving documents. Such a semantic web for science has the potential to integrate data from many sources to gain insight into complex relationships. It could, for example, be a means of integrating data from the natural and social sciences that are highly relevant to many complex global challenges; or of integrating data from the “internet of things”, where almost any device with its own power source is able to acquire non-trivial information about its environment. Such a development is impeded by two barriers: a failure of many disciplines to define their own vocabularies and ontologies, which impedes the efficiency with which they are able both to locate and use data relevant to their own discipline; and a failure to adhere to standards that enable interoperability between disciplines. A strategic initiative is currently being launched by the International Council for Science’s Committee on Data for Science and Technology, together with international science unions and associations, in the form of a Commission on Data Standards for Science to tackle these two major issues. It has great potential not only to enhance scientific understanding, but also the way that science is able to engage with the wider public in a more truly open science. This will require a major, decadal effort from across the science community, and could prove to be a profound step that will fundamentally change the way that science is done in the 21st century, through an unprecedented capacity to integrate data from disparate disciplines in ways that profoundly increase the potential of science to address major global challenges.



This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654139.

⁹ NIH: <https://www.ott.nih.gov/nih-ott-open-data-initiative>; accessed 5 February 2017.

¹⁰ ELIXIR: www.Elixir-europe.org; accessed 5 February 2017.

¹¹ Berners-Lee, Tim; Hendler, James; Lassila, Ora: ‘The Semantic Web’. *Scientific American Magazine*, 17 May 2001; available at https://www.sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf; accessed 5 february 2017.

Open Educational Resources: Service setup and Data management

Authors: Davor Orlic (Knowledge 4 All Foundation Ltd.) & Clive Young (UCL ISD Learning Technology & Media Services)

Email: davor.orlic@ijs.si

12.1. WHAT ARE THE ISSUES?

UCL (University College London) has demonstrated institutional commitment to open education in its Education Strategy 2016–21¹ with a priority to introduce an open education resources service to “provide a showcase for UCL education and for student-generated content”. The proposed service is explicitly aligned to UCL’s educational initiatives, especially the development of research-based education, the UCL Connected Curriculum.

To further the priority an Open Education Special Interest Group (SIG) was established and in 2016 this commissioned a survey and report to determine current practices and possible futures. The aim of this internal report, *Open Education Initial Scoping Study*, was an “investigation scoping the potential, practicalities and possible future actions to support open education initiatives across UCL in response to UCL’s education strategy”. The approach was deliberately wide-ranging and comprised a review of current open education activities and actors across UCL. It included open educational resources and course activities including massive open online courses (MOOCs), open data in teaching and learning, and open textbooks, again emphasising connections to the Connected Curriculum.

The main finding was that there were many perspectives, experiences and small-scale initiatives in this area across the institution. There was a need to enable and invigorate a common clear ethos of “open” across faculties and campuses. The report indicated positive reactions to the philosophy of “open” with strong support in many (but not all) academic areas. There was an appreciation among stakeholders that there is not a single correct way of doing it. It was felt that open education at UCL can best be introduced by focusing on openness via a set of specific dimensions, such as content, technology, and pedagogies.

Through the SIG UCL is assessing how initiatives can be brought together and also how existing policies and projects are used to support a more coordinated and purposeful approach. UCL is already recognised as a European leader in its commitment to open access to research. *UCL Discovery*, UCL’s open access repository for UCL research publications, is well established as a mainstream service. UCL Library Services and UCL Digital Education are in discussions about piloting an Open Education Resources (OER) repository using the same platform as UCL Discovery. Another development is that UCL has recently launched MediaCentral, a media repository that showcases and provides access to media-based teaching and other items.²

Additionally, UCL Press³ is the UK’s first fully Open Access university press, launched in 2015. Since then 43 titles have either been published or are in press, including three textbooks. This is seen as a major opportunity in UCL to change the current commercial business model for textbook publishing. Open Access textbooks present the institution with an opportunity to make an offering in the Open space which

¹ UCL: https://www.ucl.ac.uk/teaching-learning/sites/teaching-learning/files/migrated-files/UCL_Education_Strategy_Final_Web.pdf; last accessed 5 February 2017.

² UCL: <http://mediacentral-stream.ucl.ac.uk/Home/>; last accessed 5 February 2017.

³ UCL: <http://www.ucl.ac.uk/ucl-press/>; last accessed 5 February 2017.

will promote access to, and use of, textbooks by the end user. Students support OER, especially open textbooks, because OER in digital format are accessed at no cost and print copies are also available at relatively low cost.

The notion of linking some UCL Press titles to MOOCs is also being explored. UCL has already run three centrally-funded MOOCs on the *Futurelearn*⁴ platform with three others in production for launch in 2017. UCL academics are also involved in several other MOOCs as well as open access courses on *UCL eXtend*,⁵ UCL's externally-facing virtual learning environment.

12.2. COSTS

In addition to building on and connecting currently-funded open initiatives, UCL Library Services and UCL Digital Education are jointly seeking additional funding to begin piloting the OER service, with a projected implementation in 2017. In terms of OER content production, a major benefit will be cost-effectiveness because of the ability to share and re-use resources. However, while OER bring down total expenditures, they are not cost-free. New OER can be assembled or simply re-used/re-purposed from existing open resources, and RDM storage and hosting facilities can be re-used for OER. This is a primary strength of OER and, as such, can produce major cost savings. OER need not be created from scratch. On the other hand, it should be recognised that there are some costs in the assembly and adaptation process.

12.3. SERVICE PROVISION

The pilot OER service will be run by staff from across the UCL Library Services and UCL Digital Education teams. The teams will run the activity of planning and organising people, faculty, infrastructure, communication and the material components of an OER service in order to guarantee its quality and the interaction between the UCL Press as an RDM service provider and UCL faculty and students as users. The pilot will include a requirements analysis for the staff and end-user functionality of the repository (e.g. subject descriptor taxonomies, workflows to support quality assurance, branding). UCL will also identify and develop exemplar OERs to test the technical system and develop/document support processes. As indicated in the opening section a major requirement is to raise the profile of OER and input to the pilot through a programme of training and advocacy, workflows and case studies. From an educational impact perspective, the team will evaluate and make recommendations for turning the demonstrator into an established and sustainable service.

12.4. INDICATIVE COSTS

For indicative capital setup costs, the OER service will be connected to the UCL Research Data Service which required in the region of £1 million to be established, with presumably little additional cost for the introduction of OER RDM. The pilot will clarify actual staff support requirements, but potential staff costs can be estimated as a two year half-time grade 8 post – less than £60,000 including full economic costing. The existing Library, Digital Education and academic development teams will co-ordinate activity to encourage a culture of 'open' amongst existing professional and academic staff and students, including top-down strategic direction, energy and encouragement.

⁴ Future Learn: <https://www.futurelearn.com/>; last accessed 5 February 2017.

⁵ UCL: <https://extend.ucl.ac.uk/>; last accessed 5 February 2017.

12.5. SCOPE OF THE SERVICE

The Digital Education Team will support academic colleagues, colleges, units and individual researchers to publish educational media (podcasts, video, files, etc.) under an appropriate Creative Commons Licence. The team will also promote and support the use of OER in teaching at UCL via its user communities, websites, news publications, workshops and events. In partnership with Library Services, it will offer staff and students training in copyright licensing and how to prepare materials for open publication and platforms for dissemination. Furthermore UCL encourages members of the university to use, create and publish OERs to enhance the quality of the student experience, provided that resources used have undergone quality assurance, are fit-for purpose and relevant.

12.6. WHERE DO UCL OER SERVICES SIT?

The OER service will be a partnership between the UCL Digital Education Team, UCL Library Services and UCL Press. UCL Digital Education is based within the Information Services Division (ISD) and provides support, advice and training for all aspects of Learning Technology, e-learning and open and distance learning across the whole of UCL. The Division of UCL Library Services runs UCL Discovery,⁶ UCL's open access research repository, and UCL Press is a department of UCL Library Services.

12.7. POLICY DEVELOPMENT

Currently there is no existing national OER policy in the UK to support institutions in OER adoption and no current OER institutional repository in UCL. There is a UCL Research Data policy⁷ and a LERU Working Group produced the LERU Roadmap for Research Data⁸ which contains a range of guidance and information on Open Data and which could be extended or adopted for OER. UCL's current policy and advocacy for RDM are fully in line with the LERU Roadmap. In terms of a potential future UCL policy in OER, it should articulate and expand upon UCL's position on OERs and provide guidelines for practice in learning and teaching and RDM procedures. The University could encourage staff and students to use, create, and publish OERs via institutional and other repositories and help them track usage and impact.

12.8. CONCLUSIONS

OER as a service with its RDM setup is not so difficult to setup if other data management systems are already in place in a University setting. We contend that is why an OER service should be embedded into UCL's Library services and main Content Management System tools.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁶ UCL: <http://discovery.ucl.ac.uk>; last accessed 5 February 2017.

⁷ UCL <http://www.ucl.ac.uk/isd/services/research-it/documents/uclresearchdatapolicy.pdf>; last accessed 5 February 2017.

⁸ LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf, last accessed 5 February 2017.

The handling of research data in the social sciences at University of the Andes – Data Centre (CEDE) – Colombia

Authors: Nicolás Fuertes (Data Centre Manager) & María Alejandra Galeano (Research Assistant), CEDE University of the Andes

Email: nd.fuertes1359@uniandes.edu.co / ma.galeano258@uniades.edu.co

13.1. HABEAS DATA¹ VS OPEN DATA

In Colombia, there is a regulatory framework that protects the publication of personal data such as names, municipality, identification documents, phone numbers and addresses. In addition any information regarding children's development and victims of conflict is also protected by this regulation. For this reason, Centro de Datos (CEDE) at University of the Andes has created a range methods in order to publish research data and make it as open as possible, but without revealing the elements protected by Habeas Data. These methods include: confidential agreements, users with different levels of access, and a special data processing room located in the School of Economics at the University of the Andes.

13.2. TYPES OF USERS AND ACCESS POLICY

As a dissemination platform for its information CEDE uses its web page <https://datoscede.uniandes.edu.co/>. Nonetheless, due to the sensitivity and the costs of the data only users with the appropriate level of granted access clearance can access this information. To this end the platform has three different categories of users: professors from the Economics department, other University of the Andes members, and finally external users. The first have free access to all of the data sets the web page offers, the main reason for this being that the information has been collected or requested by them. The second group has limited access to the information; an authorisation from a professor of the University, along with a brief summary of the investigation for which the data is needed, are necessary in order to access restricted information. Lastly, the third group can only access the public data available in the web page, and no further information is allowed due to the contracts by which the University obtains restricted information.

13.3. ANONYMISATION PROTOCOLS

In order to publish restricted data, our CEDE created an anonymisation protocol. The algorithm used consists of:

- Generating a random number in order to identify the household. This number is generated based on the interviewer identification number, order of the municipalities visited and order of the interviews conducted in the house;
- Generating a consecutive number based on the interview order and on the order of the municipalities visited.

¹ Habeas Data: https://en.wikipedia.org/wiki/Habeas_data; last accessed 7 February 2017.

13.4. RESTRICTED DATA AND CONFIDENTIAL AGREEMENTS

In line with the Open Data initiative, CEDE has classified its datasets into five different groups: open data, public use, licensed data, external repository and data that is not available. The data that is considered licensed is mainly information from public institutions, which hold material that is relevant for economic research, but where that information is not generally freely accessible. To access this type of data, the University subscribes to confidentiality agreements or contracts. However, due to the sensitive nature of this data, its availability is subjected to an authorization process by which professors attest to the use of the data by a student, subject to a prior discussion as to its necessity within the specific research process. The data that contains information protected by Habeas Data also has an additional step of authorization, in which the researcher has to sign a confidentiality clause where they agree not to reveal information about particular individuals, not to use the data for other purposes, and/or not to circulate the information among other people that lack the necessary permission to use it.

13.5. DATA PROCESSING ROOM

As a contingency measure to protect the information subject to Habeas Data and to guarantee access to the information, CEDE jointly with the National Department of Statistics (DANE) have made available a data processing room. In this space researchers can access information with no anonymity process in order to estimate more complete economic models. To access this room, researchers have first to request the information from DANE by way of an email in which they give details of the specific data set and give a brief introduction to the research project they are working on. After this, DANE grants access by creating a designated user file into which the information requested is placed. Finally, researchers sign a confidentiality agreement and schedule an appointment to work on the computers available in the data processing room. These computers provide no Internet connection or USB ports, in order to control the entry and extraction of data. When researchers want to enter or take information, they must first send an email to DANE concerning the material they want to enter or file, as well as the output format, to obtain the results from the calculation made in the data processing room. Under no circumstances is the raw data made available out of this room.

13.6. CONCLUSIONS

For CEDE it is important to make the data available and as open as possible. However, in this process, it is necessary to respect the law related to personal data protection in Colombia. CEDE has identified and operates some methods that are useful to open the data without revealing the protected elements within it. For this reason, the University believes that it is important to find a way under the law to publish this data and make it as open as possible.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Section 5

Research Data Infrastructure

The Research Data Storage Service at UCL – A LEARN Case Study

Author: James A J Wilson (Head of Research Data Services UCL)

Email: j.a.j.wilson@ucl.ac.uk

14.1. DATA STORAGE DURING THE 'ACTIVE' PHASE OF RESEARCH

The active phase of a research project comprises the generation or collection of data, its processing, and its analysis. If data is well managed during this phase then it can considerably simplify the job of preparing data for longer-term preservation and access after the end of a project, but it is not easy for institutions constructively to intervene in many elements of research, which are often highly specific to the requirements of a particular project. Data storage is the one element that almost all researchers depend upon, and where institutions can offer a generic central service. However, even in this realm there is a wealth of options available to most researchers, from laptop hard drives and memory sticks, to commercial cloud services such as Dropbox.

By providing researchers with a storage service that is both easy to use and includes helpful collaboration mechanisms, an institution can however gain some measure of control over how their data assets are managed, and facilitate the smooth path of data and associated metadata through the research data lifecycle.

14.2. THE RESEARCH DATA STORAGE SERVICE AT UCL

The development of the Research Data Storage (RDS) Service at UCL was motivated from the outset by the necessity of assisting researchers to comply with the requirements of research funders. UCL sought to develop a data storage service that had the 'resilience and disaster recovery to assure the safety of research data'; 'multiple and intuitive user interfaces to meet a broad set of user experiences', a 'service wrap to make the Service useful to more users', and the 'capacity to increase the user base across UCL'. A tender for physical storage to enable the objectives of a data storage service was issued in 2012, and the service opened to researchers in June 2013.

Use of the service has grown exponentially since that time. As of December 2016, the service hosts approximately 760 TB of research data before replication and redundancy, 1.791 PB in total. All faculties at UCL have at least one project that is using the service.

The service is offered to research projects, rather than individual researchers. This helps with the assignment of useful metadata, as projects can be cross-referenced with administrative information held in grants databases and other UCL information sources. In practice, the service does not prohibit the creation of unofficial projects, as that would effectively proscribe the use of the storage by 'unfunded' research, a mode of working which is common in the humanities and social sciences.

When signing up for an allocation of project storage space, the authorisation of a Principal Investigator is required. The PI must vouch that no personal data (as opposed to research data) is held in the system, and that they recognise their legal obligations under the UK Data Protection Act 1998 and otherwise.

Research Data Services
James Wilson

Register a new project...

Principal Investigator

Project administrator (optional)

Start date

End date

Volume of data (TB)

Title

Description

UCL users

I accept that this allocation will be used for Project, not Personal data.

I confirm that we will not store data on this service for which we have obligations under the UK Data Protection Act 1998 or any other pre-existing third party arrangement.

Figure 14.1 The RDS New Project Registration Page

To be assigned a new project, the PI must also provide a start/end date for their project and some basic descriptive metadata. Projects can request between 1 and 5 TB storage, or contact the service directly if they need more. The minimum allocation of 1TB reflects the fact that the service was originally developed with large-scale data users in mind, as this community was least well served by alternative solutions, although the Storage Service is available to all UCL researchers however much data they anticipate generating.

14.3. UNDERLYING INFRASTRUCTURE

There are two different storage technologies under the bonnet of the RDS Storage Service: *General Parallel File System* (GPFS) block storage; and *Web Object Storage* (WOS). This was seen as a good combination, as the fast GPFS component can cater for users who require data to be staged to UCL's high-performance computing facilities, whilst the highly scalable object storage provides a cost-effective way of managing the bulk of UCL research data. The *Integrated Rule-Oriented Data System* (iRODS) is used as the management layer for data in the object store.

14.4. SUPPORT REQUIREMENTS

Besides the need to keep the infrastructure up to date and ensure that the service is running smoothly from a technical perspective, the RDS team works with UCL Library Services to assist researchers with interesting use-cases to make the most of the service by ensuring their workflows are rationalised. At present, some common administration processes, such as changing permissions in project groups, are also still a semi-manual process, although web interfaces are being developed to allow users to do more of this themselves.

14.5. COSTS AND PRICING

At the time of writing the Research Data Services team consists of 4 full-time employees (4 FTE), although not all of this staffing resource is dedicated to keeping the storage service ticking over. Monitoring, patching, bug-fixing, service communications, support and consultancy, and service management take about 2.5 FTE at present, with the rest of the time going towards future service development (including a UCL institutional repository), a re-architecting of the present service, and technology monitoring and assessment. An unusually high proportion of staff time over the last year has been spent dealing with issues affecting the object storage. Once the service is more mature, and more of its administrative processes automated, we would expect it to require less staff time to maintain. In addition to the core team, the service requires a small amount of resource from the UCL helpdesk team and the Data Centres team.

Hardware and support costs for a storage service will vary according to the specific deal arranged with the supplier(s). The current RDS capacity was achieved via two purchases: an initial purchase of just under a petabyte of GPFS storage and 240 TB of WOS storage, plus servers, support, and other small items of equipment, for around £740,000 in 2012; and an expansion of 2.88 PB of WOS for a little under £600,000 in February 2014. 1.2 PB of this was later converted to GPFS.

The service itself is currently offered free of charge to UCL researchers, although those with particularly large requirements (>10TB) are asked to contribute to costs if they are able. As the service scales up, this model is unlikely to remain viable, so a new pricing model is currently under development to ensure long-term sustainability.

The new pricing model will almost certainly allow a storage allocation up to a certain point free of charge, with charges applying for quantities beyond this as yet unset level. This should enable small and unfunded projects to continue using central storage, with all its benefits both to researchers and institution in terms of being able to manage data over the long term. More data-intensive projects, on the other hand, will be expected to include their required data storage capacity in their grant applications – passing their exceptional costs on to the research funder.

Although demand for the service is anticipated to continue to grow exponentially, the costs are expected to be offset in part by the falling price of storage. We are seeking to move to a purchasing strategy of buying storage according to more of a just-in-time model in future, as it makes little sense in owning constantly depreciating capacity standing idle. It is possible that some sort of cloud capacity will be used as well, but it is recognised that the costs of cloud storage add an unpredictable and potentially expensive component to the service model.

14.6. FUTURE REQUIREMENTS

At present, the RDS Service is a push-in / pull-out service. However, many of our users want to be able to use their allocated storage space as though it were available as a mounted drive. This prospect is challenging given the large file sizes the service needs to cater for, but various technologies are being assessed for suitability.

Other improvements and functionality that users have requested include:

- a. File versioning
- b. Dropbox-like sync and share functionality
- c. The ability to add non-UCL collaborators to projects (which is currently possible, but only by adding the collaborator as an honorary member of UCL, which is a bureaucratic process)

As of December 2016, the RDS is engaged in a major project to expand capacity and better address user requirements.

14.7. LESSONS LEARNED

Some things to consider when setting up a storage service for active research data:

- Ensure that your choice of underlying storage technology is mature and reliable – this is a situation where being an early adopter is not necessarily a good strategy;
- Have a clear policy as to what the service can and cannot offer;
- Ensure a daily back-up is in place;
- Run induction sessions to understand new users and their requirements;
- Communicate clearly the benefits of institutional storage over personal storage;
- Recommend a single graphical interface for less technical users, plus programmatic access for the more technically adept;
- Invest time in developing a clear reporting system that is independent from the underlying infrastructure;
- Understand how your institution's identity and group management systems work;
- Have a plan B for if something goes catastrophically wrong!



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Scientific Data Management on a Dataverse Network at IBICT

Authors: Miguel Ángel Márdero Arellano and Alexandre Faria De Oliveria, (Coordinator & Technological Solutions Coordinator, Brazilian Network of Digital Preservation Services, CARINIANA, IBICT¹)

Email: miguel@ibict.br / alexandreoliveira@ibict.br

15.1. INTRODUCTION

In the 20th century, at the time when the State was the main driving force for the management of information and documentation resources and services, the creation of intermediary information systems in Brazil became a matter of strategic importance. As a result of this, it was necessary to construct a scientific technical infrastructure, as well as to train qualified personnel in the management of the production of, access to and preservation of information in science and technology. Brazil's government conferred upon the Brazilian Institute of Information in Science and Technology (IBICT) the responsibility to become the standard-bearer for core competencies in the process of treatment of, access to, and dissemination of information.

The Cariniana is a distributed preservation network, funded by IBICT, committed to national and international cooperation, promoting the management and dissemination of digital preservation practices and developing a sustainable digital preservation programme to support Brazilian universities and research centers' needs and requirements. In 2012 IBICT recognized the need to address digital preservation issues, and it adopted the LOCKSS (Lots of Copies Keeps Stuff Safe²) approach as suitable for the needs of the Cariniana network. Its main focus concerns open access publications in Brazil. The network preserves journals, and doctoral theses, and it is just starting to cover scientific data to be deposited in a research data repository. The experimental phase, using LOCKSS open source, covered a year in 2013, and was supervised by LOCKSS staff from the University of Stanford.

In 2015, the implementation of Cariniana's Dataverse³ repository added significant new services to the digital preservation network which will help specialized libraries' staff to deal with the demand from researchers for a trusted space for their datasets. IBICT is making available a repository for research data that is responsible for long term preservation and good archival practices, while researchers can share, keep control of, and receive recognition for their data. In addition, the repository supports the sharing of research data with persistent data citation and enables reproducible research.

15.2. WHAT MOTIVATED THE REPOSITORY OF SCIENTIFIC DATA AT THE CARINIANA NETWORK?

The Cariniana Network resulted from the need to create a digital preservation service of Brazilian electronic documents to ensure continuous access to these documents throughout time. The creation of the project for the preservation of research data was based on the idea that the more copies of a document that are stored in different places, the safer they will be. First, a centralized storage structure is used; then the content goes through distributed computer resources, with the participation of institutions that support electronic documents.

¹ IBICT: <http://www.ibict.br/>; accessed 5 February 2017.

² LOCKSS: <https://lockss.org>; accessed 5 February 2017.

³ IBICT Dataverse Network: <http://repositoriopesquisas.ibict.br>; accessed 5 February 2017.

Initially, the activities were carried out jointly with the University of Brasilia. In the first phase, the Dataverse network is being used for the addition and storage of research documents from individuals, institutional projects, and electronic journals. After that, the possibility of integration with the LOCKSS platform of the partner institutions will be used as the preservation repository of the stored material. Offering digital preservation services includes integrating the scientific data content of the connected institutions into a unified pattern; these mechanisms must facilitate the automation of processes of identification, storage, validation, and conversion of the content into new digital formats.

IBICT started a pilot project in 2015 and one of its objectives is to be a valuable contributor to the development of research data repositories in Brazil. The Cariniana Dataverse network is developing information products to promote the practice of digital curation at institutions with important collections in digital format. Coordinated by IBICT, the Dataverse repository is used by the Cariniana team to help network partners become proficient at using methods of insertion and storage of electronic documents in research data repositories.

ibict Instituto Brasileiro de Informação em Ciência e Tecnologia

POWERED BY THE **Dataverse Network** PROJECT v. 3.8.2

IBICT Dataverse Network [Search](#) [Advanced Search](#) [Tips](#) [Create Account](#) [Log In](#)

Search this Dataverse Network

O IBICT está disponibilizando um repositório de dados de pesquisa que cuida da preservação a longo prazo e de boas práticas de arquivamento, para que os pesquisadores possam compartilhar, manter o controle e obter o reconhecimento de seus dados. O repositório suporta o compartilhamento de dados de pesquisa com citação de dados persistentes, permitindo que eles sejam reproduzidos. O Dataverse é um grande repositório aberto de dados de todas as disciplinas, criado pelo Institute for Quantitative Ciências Sociais da Universidade de Harvard. O repositório Dataverse do IBICT fornece um meio disponível gratuitamente para depositar e descobrir conjuntos de dados específicos arquivados por colaboradores das instituições participantes da rede Cariniana.

Dataverses

30 Dataverses

A Dataverse is a container for research data studies, customized and managed by its owner.

RECENTLY RELEASED DATERVERSES

Marcos Aurélio Gomes	07/12/2016
Datasets de Conferências	10/11/2016
Datasets de Teses	10/11/2016
Datasets de Dissertações	10/11/2016
Guilherme Ataíde	08/11/2016

[View More >](#)

Faculdade de Tecnologia Dataverses

Dataverses: 1

Repositório de pesquisa da Faculdade de Tecnologia da UnB.

Studies

131 Studies, 399 Files, 83 Downloads

A study is a container for a research data set. It includes cataloging information, data files and complementary files.

RECENTLY RELEASED STUDIES

Gestão do conhecimento	19/01/2017
Viabilidade técnico-econômica de plantas solares	19/01/2017
Resoluções ANEEL	19/01/2017
Modelagem e simulação de sistemas CPV-T	19/01/2017
Revista Brasileira de Energia Solar	19/01/2017

[View More >](#)

MOST DOWNLOADED STUDIES

Normas do Tamanho da Categoria para 905 palavras da Língua portuguesa by Janczura, Gerson Américo; Castilho, Goiara Mendonça; Oliveira, Nelson Rocha	22
Gestão dos Dados Científicos dos Pesquisadores dos Programas de Pós-Graduação em Ecologia do Nordeste Brasileiro by Anjos, Renata; Dias, Guilherme	11
Resumo de Atividades Rede Oriade 2015 by Mardero Arellano, Miguel Angel	9
O documento na Ciência da Informação: em busca da construção de um objeto a partir da concepções da Documentação e da Diplomática by Rabello, Rodrigo	8
Normas de associação livre para 1004 palavras da língua portuguesa by Janczura, Gerson Américo; Castilho, Goiara Mendonça; Keller, Victor Nahuel; Oliveira, Nelson Rocha	7

[View More >](#)

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
http://doi.org/10.17616/R3K36C
IBICT Dataverse Network

Figure 15.1: Homepage of IBICT Dataverse Network Portal

Dataverse is a large repository open to data from all disciplines and hosted by the Institute for Quantitative Social Science at Harvard University. The Dataverse repository at IBICT provides free-of-charge an available means to deposit, find, and access specific datasets that are being archived by researchers from the participating organizations. It will act as a steward of digital content, is open for data deposits from our institutions' affiliated partners, and it shares content with all their researchers and librarians.

The Dataverse repository includes a relatively simple self-service ingest workflow for researchers; it also has the ability to share with trusted groups of researchers prior to publication, and it helps them fulfil Data Management Plan requirements. The Cariniana team was interested in Dataverse because it can be easily installed and maintained, and it can be brought online with a relatively small staff.

Nonetheless, the main reason Cariniana chose to make a Dataverse repository available to its partner institutions was the ability to integrate it with other systems; that is, LOCKSS and Archivematica⁴ for distributed and local long-term preservation, OJS⁵ for data publication, and DSpace⁶ for interoperability.

15.3. WHAT SCIENTIFIC DATA IS ARCHIVED?

Thanks to the technical cooperation agreements established for OJS journals, the Dataverse repository had allowed initial collaboration and support on the implementation of a scientific data preservation service. The target of the service is content from institutional and individual projects. IBICT Dataverse's member service provides an individual space for deposit of archives or datasets for researchers, or a community of researchers and institutions. All the data that is being archived is automatically identified, linked, and supported with access mechanisms.

The datasets in Dataverse are considered a structural archive with standardized metadata to maximize its compatibility and retrieval. This helps researchers meet the requirements of the funding institutions for verification of research project data. All the metadata records are made available for research, and the service allows uploading of datasets identified by the author or institutional owner. Information from the dataverses can be used by local libraries in helping users with better-informed answers to their queries. Furthermore, Cariniana is collaborating with its partners' institutions' libraries in their management decisions regarding researchers' data that needs to be processed by digital curation and archiving processes.

⁴ Archivematica: <http://archivematica.org>; accessed 5 February 2017.

⁵ PKP: <https://pkp.sfu.ca/ojs/>; accessed 5 February 2017.

⁶ DSPACE: <http://dspace.org>; accessed 5 February 2017.

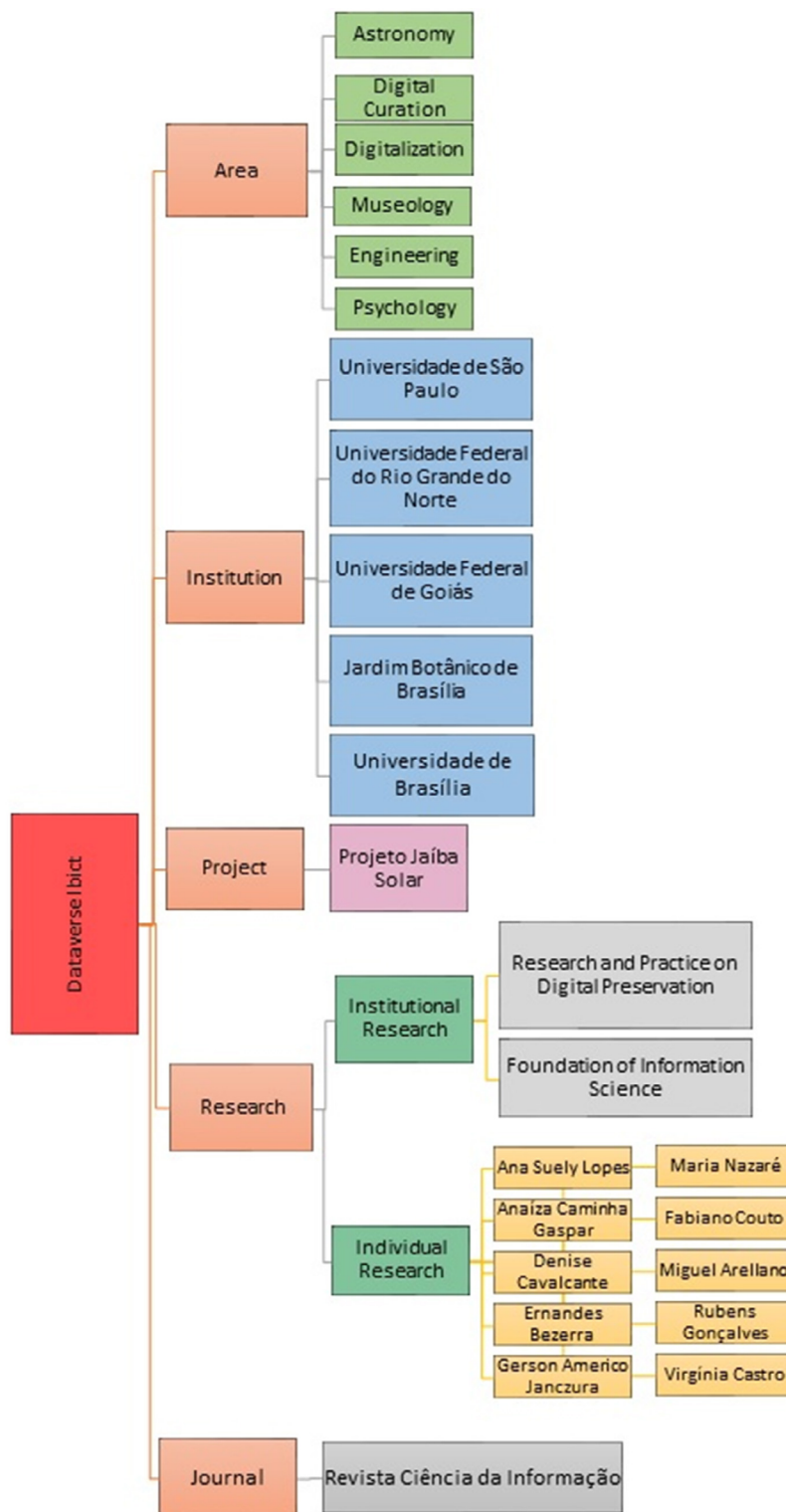


Figure 15.2: Organogram of IBICT Dataverse Network

Currently, the repository is being used by twenty-nine Cariniana collaborators, from six institutions, who have deposited data as members of the network. Researchers may store any type of digital data on any subject specialism. The Dataverse repository hosts 89 studies uploaded on 23 dataverses from institutional and individual research projects.

15.4. WHAT COMES NEXT?

The repository requirements and policies on access, privacy, and reuse need to be well defined. Cariniana staff are establishing a curator service that will help organize data for preservation. A partner institution is planning financially to support the translation of the system's user interface into Portuguese, and the Cariniana team has produced a user manual. The migration of Dataverse to version 4.0 will be accomplished in the current year. This procedure will be combined with an institutional strategy at IBICT for secure backups and the assignment of a persistent URL.

IBICT's International Technical and Scientific Support Committee is establishing guidelines and recommendations for the repository. The Cariniana network is now discussing and planning further collaborative action and an integration of efforts. Its members consider Dataverse a trusted repository that allows for the long-term persistent access to scientific data, with mechanisms that incorporate full identification and the prevention of digital obsolescence.

IBICT considers it fundamental to establish efficient management of data for the development of scientific research, with quality in all aspects related to organization, documentation, archiving, and sharing of scientific information. As the project evolves, new challenges arise, revealing new approaches to the scientific information digital workflow. In Brazil there are still relatively few research data repositories, but the impact of the contributions from the field of Library and Information Science is growing.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Delivering the European Open Science Cloud (EOSC): Principle and Practice in delivering Open Science

Author: Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries)

Email: p.ayris@ucl.ac.uk

16.1. THE EUROPEAN OPEN SCIENCE CLOUD

In 2016, the European Commission's High Level Expert Group (HLEG) on the European Open Science Cloud published their Report.¹ I was privileged to be a member of this HLEG. The Report is designed to establish a vision for the future of Research Data, particularly Open Data, in Europe. The main findings were:²

- The majority of the challenges to reach a functional EOSC are social rather than technical.
- The major technical challenge is the complexity of the data and analytics procedures across disciplines rather than the size of the data per se.
- There is an alarming shortage of data experts both globally and in the European Union.
- This is partly based on an archaic reward and funding system for science and innovation, sustaining the article culture and preventing effective data publishing and re-use.
- The lack of core intermediary expertise has created a chasm between e-infrastructure providers and scientific domain specialists.
- Despite the success of the European Strategy Forum on Research Infrastructures (ESFRI), fragmentation across domains still produces repetitive and isolated solutions.
- The short and dispersed funding cycles of core research and e-infrastructures are not fit for the purpose of regulating and making effective use of global scientific data.
- Ever larger distributed data sets are increasingly immobile (e.g. for sheer size and privacy reasons) and centralised HPC alone is insufficient to support critically federated and distributed meta-analysis and learning.
- Notwithstanding the challenges, the components needed to create a first generation EOSC are largely there but they are lost in fragmentation and spread over 28 Member States and across different communities.
- There is no dedicated and mandated effort or instrument to coordinate EOSC-type activities across Member States.

The Report proposed 4 Recommendations for Policy development, 4 for Governance and 7 for the next phase of Implementation.³ Policy Recommendation 4 captures the essence of the HLEG vision for the EOSC: 'Frame the EOSC as the EU contribution to an Internet of FAIR Data and Services underpinned with open protocols.' The Recommendations on Implementation spell out the need for a Roadmap for future development of the EOSC, with rules of engagement and a light-touch framework for governance. How easy will it be to deliver this bold vision? Let us consider two challenges: Funding and FAIR research data.

¹ European Commission: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; last accessed 8/1/17.

² Ibid., p. 6.

³ Ibid., p. 7.

16.2. CHALLENGES: FUNDING AND FAIR RESEARCH DATA

Funding is an obvious challenge. The Report emphasises that the need is not so much one to build new infrastructures, but rather to make what Europe already has less silo-based and more interoperable:

Based on the consensus that most foundational building blocks of the Internet of FAIR data and Services are operational somewhere, but that they operate in silos per domain, geographical region and funding scheme, we recommend that early and strong action is taken to federate these gems. Optimal engagement is required of the e-infrastructure communities, the ESFRI communities and other disciplinary groups and institutes. Several of these cross-ESFRI building blocks begin to operate in individual Member States. Simultaneously, the wealth of small and large industrial players in Europe should be engaged. All partners and stakeholders that adhere to standards and sign off on the Rules of Engagement (RoE) should be eligible.⁴

Even if this is true, the costs for delivering the EOSC are for significant sums of money. Overall, it is estimated that €2 billion is needed from the Commission's Horizon 2020 funding pot, as well as additional public and private investment of €4.7 billion to develop further the European data infrastructure. Of this €4.7 billion, €0.2 billion is needed for widening the user base to the public and private sectors, €1 billion for the EU-wide Quantum technologies flagship and €3.5 billion for data infrastructure.⁵

Stakeholder Group / Need	Funding requirement to deliver EOSC
<i>European Commission</i>	
Horizon 2020	€2 billion
<i>Public / Private funds</i>	
Widening User Base to public and private sectors	€0.2 billion
EU-wide Quantum technologies flagship	€1 billion
Data infrastructure	€3.5 billion
TOTAL	€6.7 billion

Figure 16.1: Perceived funding requirement to deliver the EOSC

The financial requirement is significant indeed.

What can be said about the Report's central insistence on FAIR data? The vision of the Report is for the EOSC to be technically conceived as an Internet of FAIR data and Services. It points to a parallel with the early development of the Internet:

The creation of NSFNET, choice of the TCP/IP standard and the authorised development of Domain Names enabled the boom of the Internet in the 1990's, where the development of the HTTP and HTML drove its major application domain, the largely textual WWW. This combination of authorisation, key support by a major leading agency (NSF) and a dedicated community (W3C) setting and enforcing minimal standards allowed virtually everyone to start building standard-compliant tools and services in the ecosystem.⁶

⁴ Ibid., p. 14.

⁵ European Commission: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; last accessed 8/1/17. For Quantum technologies, see European Commission: <https://ec.europa.eu/digital-single-market/en/news/european-commission-will-launch-eu1-billion-quantum-technologies-flagship>; last accessed 8/1/17.

⁶ European Commission: http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf; last accessed 8/1/17.

FAIR research data is **F**indable, **A**ccessible, **I**nteroperable and **R**eusable.⁷ A move to FAIR data as the default principle for research data management is a significant step. The 2016 UCL study of research data analysed the reasons why UCL researchers would not share their data.⁸

Question	Result/Response
Nobody asked me to	40%
Confidentiality / Intellectual Property / Data Protection	25%
Ethical Issues	10%
Other	7%
Time / effort required to collect them	6%
Possible misinterpretation of data	6%
Commercial Issues	3%
Licence agreement prohibits sharing	2%
Data no longer readable	1%

Figure 16.2: Question 55. Why didn't you give access to your data? (118 respondents)

Whilst FAIR research data is essential to underpin the EOSC as an Internet commons of data available for sharing and re-use, it is clear that the academic community has some way to go to see Open Data as the default position for research data they are creating/using, as the UCL survey shows.

16.3. CONCLUSION

The publication of the EOSC Report represents a watershed in the vision for the creation of a European, and in the long term a global, commons of FAIR research data. Johannes Gutenberg's invention of moveable type printing in the West at the end of the fifteenth century revolutionised the way ideas were recorded and disseminated. The Protestant Reformation, and the Counter Reformation, would not have been possible without the aid of the printing press. FAIR, Open Data and developments such as the European Open Science Cloud have the potential to have a similar impact in the 21st century. Studies such as the one undertaken by UCL, however, underline the challenges that need to be overcome to deliver the EOSC vision. Researchers are engaged on a journey, and it is the mission of the LEARN Toolkit of Good Practice to help them arrive at their chosen destination.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁷ FORCE11: <https://www.force11.org/group/fairgroup/fairprinciples>; last accessed 8/1/17. For an early discussion, see M.D. Wilkinson, M. Dumontier, B. Mons et al., 'The FAIR Guiding Principles for scientific data management and stewardship' at Nature: <http://www.nature.com/articles/sdata201618>; last accessed 8/1/17. Published in *Scientific Data*, 3:160018, DOI: 10.1038/sdata.2016.18.

⁸ For a full description of this survey, see chapter 9 of this Toolkit, pp. 45-56.

Section 6

Costs

Research Data Management at the University of Edinburgh: How is it done, what does it cost?

Authors: Robin Rice (Data Librarian & Head, Research Data Support) and David Fergusson (Head of Research Services at the University of Edinburgh Information Services Group)

Email: R.Rice@ed.ac.uk

RESEARCH DATA MANAGEMENT PROGRAMME (2012-2016)

17.1. Beginnings: Data Asset Framework project

Research Data Management (RDM) began to be addressed at the University of Edinburgh in 2009, when an RDM website was published within Information Services giving advice on various RDM topics, with pointers to contacts in the University who could help. This came about because of the results of an investigation into data management practices in three university departments as part of the Digital Curation Centre's Data Asset Framework (DAF, then Data Audit Framework) in 2008.¹ Specifically, the project steering group recommended four further actions:

1. University policy in research data management;
2. Training for staff and postgraduates;
3. Web page guidance on research data management;
4. Gap analysis of existing support services.

17.2. Policy development

Another committee, led by the Director of Library & University Collections, developed the university's Research Data Management Policy, passed by the University Court in May, 2011.² This outlined the roles and responsibilities of researchers and the university itself for RDM good practice, noting that some services would need to be developed to fulfil the institution's obligations. The Data Library team began piloting training for PhD students, leading to a Jisc-funded project, Research Data Management Training (MANTRA, 2010), which created an open educational resource with 9 modules on data management and handling, still maintained and in use today.³

17.3. RDM Roadmap

The gap analysis activity, led by a new, academic-led RDM Steering Group, resulted in an RDM Roadmap, designed for high level planning that would become a living document as goals were met and new ones added. Eventually the Roadmap covered the time period January 2012 through July 2016,⁴ and covered four categories of service:

- RDM planning: support and services for planning activities typically performed before research data is collected / created;

¹ DCC: <http://www.data-audit.eu/>; last accessed 12 February 2017.

² University of Edinburgh: <http://www.ed.ac.uk/is/research-data-policy>; last accessed 12 February 2017.

³ EDINA: <http://datalib.edina.ac.uk/mantra/>, last accessed 12 February 2017.

⁴ University of Edinburgh: <http://www.ed.ac.uk/is/rdm-roadmap>; last accessed 12 February 2017.

- Active data infrastructure: facilities to store data actively used in current research activities and to provide access to that storage, and tools to assist in working with the data;
- Data stewardship: tools and services to aid in the description, deposit, and continuity of access to completed research data outputs;
- Data management support: awareness raising and advocacy, data management guidance and training.

17.4. Business case to fund the Roadmap

Capital and recurrent funds were secured from the university to cover the human and physical infrastructure needed to support the services. As stated in the RDM Roadmap, the business case submitted to the University IT Committee in June 2012 estimated a cost of £1M one-off, and £250K recurrent to implement the RDM Policy. In some cases, services already existed and just needed to be brought under the governance of the RDM steering group, such as Edinburgh DataShare, first set up as a demonstrator for managing research data in institutional repositories (DISC-UK DataShare project, 2007-09⁵), but becoming through completion of Roadmap goals the University's institutional research data repository.⁶ The recurrent funds enabled some new RDM-specific posts to be created, as an adjunct to existing support roles across Information Services.

SUPPORTING INFRASTRUCTURE FOR RESEARCH DATA MANAGEMENT (RDM) AT THE UNIVERSITY OF EDINBURGH

17.5. Efficiencies

The University of Edinburgh provides a consolidated underlying infrastructure for a large number of services that require data storage and data management. This has a number of advantages in both providing increased usability for the end user, all their data is accessible in 'one place' (although through numerous services), thus more closely corresponding to their mental model; and providing efficiencies of scale for the operation of the services through avoiding fragmentation and duplication of infrastructures.

17.6. Scale

The infrastructure at Edinburgh is composed of a primary base layer of a fast parallel storage file system (using GPFS), approximately 9 Petabytes in total. This allows large numbers of concurrent accesses without degrading the performance. This is then presented to a range of desktop services through a presentation layer of servers that export the correct protocols for Windows, Mac, Linux, etc. In addition, the infrastructure also serves the University compute cluster for large scale analysis tasks.

17.7. Infrastructure staffing

In order to provide this underlying converged infrastructure a small operations staff run the infrastructure, and financial support for these posts is shared across specific services (and also specific research activities and projects). There are two specific posts which support the RDM services directly on DataStore, 1 Senior Systems Engineer and 1 Junior Systems Engineer. These provide office hours support for the service and an approximately 99% availability service.

⁵ DISC-UK: <http://www.disc-uk.org/datashare.html>; last accessed 12 February 2017,

⁶ University of Edinburgh: <http://datashare.is.ed.ac.uk/>; last accessed 12 February 2017.

17.8. DataStore

The large scale shared high performance storage infrastructure at the University of Edinburgh was initiated in 2005 under the umbrella of the ECDF (Edinburgh Compute and Data Facility). The DataStore service has grown from that starting point to provide the main storage, back-up and disaster recovery infrastructure for research data, group data and personal data. Storage on DataStore is currently charged internally to the University at £175/TeraByte/year.

FROM RDM PROGRAMME TO RESEARCH DATA SERVICE

17.9. Transition and new website

The transition from RDM Programme to Research Data Service has been completed and the final Roadmap has been signed off by the steering group, with acknowledgement of some minor missed targets that will be rolled into ongoing service improvements. The new service website⁷ reflects all of the service components including DataStore and DataShare and is organised according to a vision which takes into account the full user experience of using the service in the context of doing their research, and in becoming a one-stop shop for any research data-related needs:

- User-friendly navigation and headings (instead of brand names, for example 'Active Data Storage' under the general category, 'Working with Data' instead of 'DataStore');
- Tools and support categorised according to a simplified data lifecycle corresponding to before, during, and after a research project;
- Generic and customised training and support available on demand.

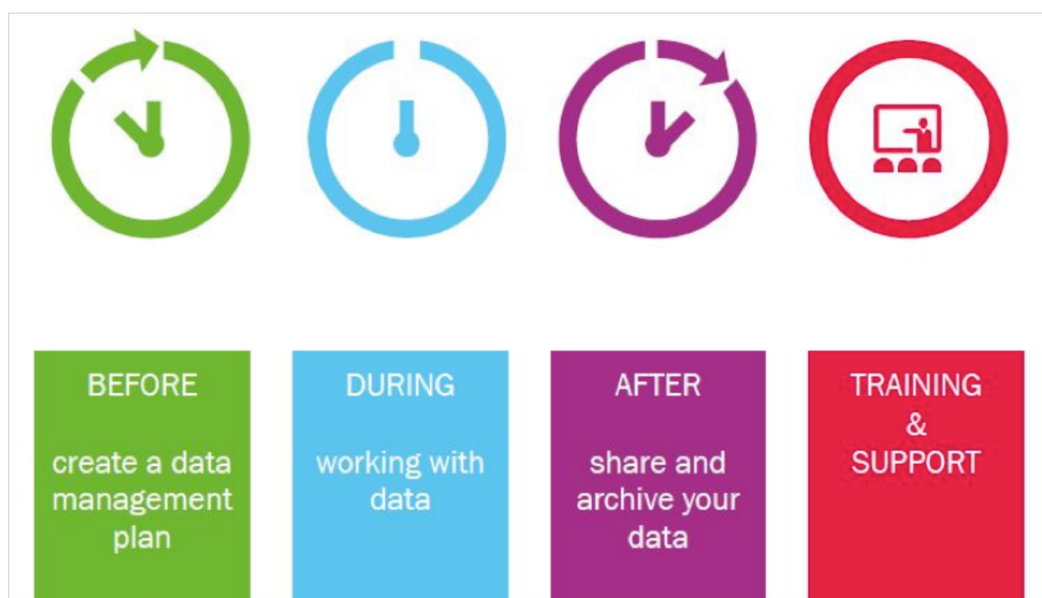


Figure 17.1: Data lifecycle and training needs

⁵ DISC-UK: <http://www.disc-uk.org/datashare.html>; last accessed 12 February 2017,

⁶ University of Edinburgh: <http://datashare.is.ed.ac.uk/>; last accessed 12 February 2017.

⁷ University of Edinburgh: www.ed.ac.uk/is/research-data-service; last accessed 12 February 2017.

17.10. Service team

The line is difficult to draw on the exact RDM team because of the necessary and pre-existing contributions from staff across Information Services, namely: Library & University Collections, IT Infrastructure, EDINA and Data Library, Digital Curation Centre, and User Services Division. Using service management framework language, the Research Data Service requires the following roles to be filled: Business Owner (representing the customer, currently filled by the Chair of the Steering Group), Service Owner, Service Operations Manager, and Virtual Team; the latter three are all staff members of Information Services.

17.11. Funded RDM posts

The Virtual Team is large, and includes IS staff who contribute to the service in any substantial way, and who were already providing data services of some sort before the RDM Programme began. However, the posts specifically funded by the service itself are as follows:

- 1 RDM Service Coordinator (0.7 FTE) (Library & University Collections)
- 1 Information Systems Developer (Library & University Collections)
- 1 Senior Systems Engineer (IT Infrastructure)
- 1 Junior Systems Engineer (IT Infrastructure)
- 3 Research Data Service Assistants (2.0 FTE) (EDINA and Data Library)
- 1 Software Engineer (0.5 FTE) (EDINA and Data Library)

The RDM Service Coordinator and the Research Data Service Assistants make up the front-facing staff, along with other existing posts such as data librarians.

17.12. Staff Budget

Median staff costs are 42,000 GBP for senior staff and 34,000 GBP for junior staff. The RDM budget including a small amount for operational costs (events, printing, minimal travel expenses) is 350,000 GBP for 2016-17, although due to normal staff changes and turnover this may vary some from projected expenditure. The funding model builds on the original recurrent university funds and employs cost recovery where practicable, especially via line items in research grant proposals. Hardware costs are considered capital spend.

17.13. Ongoing work

Current project activity will lead to additional service components being incorporated: Data Vault and Data Safe Haven. (Data Vault is currently being offered by appointment only but the aim is to move to a self-service workflow as soon as possible.) Data Safe Haven, an active data infrastructure for sensitive data, is due to be rolled out in August, 2017. Models to sustain their operations are being developed as part of the project activity.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Section 7

Roles, Responsibilities & Skills

Training early career researchers

Authors: Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries) & Ignasi Labastida (Head of the Research Unit at the Learning and Research Resources Centre [CRAI] of the University of Barcelona)

Email: p.ayris@ucl.ac.uk / ilabastida@ub.edu

18.1. RESEARCH AND RESEARCH-BASED EDUCATION

Research – blue sky and applied – is fundamental to the mission of research-intensive universities. As such, it is enunciated in the Mission Statements of such institutions. The University of Barcelona for example, a research-intensive university in the Catalan region, states that ‘The University of Barcelona is a public institution committed to the environment, whose mission is to provide a quality public service of higher education primarily through the study, teaching, research and effective management of knowledge transfer.’¹ This is its Mission. Research also features prominently in its Vision, ‘Barcelona University must be a university that offers comprehensive training, ongoing and critical evaluation at the highest level, and research which is both advanced and efficient.’²

The importance of research in a university is also captured in the Mission Statements of university organisations. LERU, the League of European Research Universities, advocates:³

- education through an awareness of the frontiers of human understanding;
- the creation of new knowledge through basic research, which is the ultimate source of innovation in society;
- and the promotion of research across a broad front in partnership with industry and society at large.

Learning through research and enquiry is a fundamental feature of study in a research-intensive university. Universities, with a strong tradition of producing world-class research, wish to demonstrate that excellence not only in their research outputs but also in the learning experience of their students, both undergraduate and postgraduate.

University College London (UCL), a research-intensive university in the UK, has developed a model for research-based education via its Connected Curriculum initiative.⁴ This is made up of six inter-connected strands of activity:

1. Students connect with researchers and with the institution’s research;
2. A through line of research activity is built into each programme;

¹ ‘La Universitat de Barcelona és una institució de dret públic compromesa amb l’entorn, la missió de la qual és prestar el servei públic (de qualitat) de l’ensenyament superior principalment per mitjà de l’estudi, la docència, la recerca i una gestió eficaç de la transferència del coneixement.’ University of Barcelona: <http://www.ub.edu/pladirector/en/missio.html>; last accessed 29/1/17.

² ‘La Universitat de Barcelona ha de ser una universitat que inclogui una formació integral, continuada i crítica del més alt nivell, i una recerca avançada i eficient.’ University of Barcelona: <http://www.ub.edu/pladirector/en/missio.html>; last accessed 29/1/17. University of Barcelona: <http://www.uab.cat/doc/PlaDirector13-15EN>; last accessed 29/1/17.

³ LERU: <http://www.leru.org/index.php/public/about-leru/mission/>; last accessed 7/1/17.

⁴ UCL: <https://www.ucl.ac.uk/teaching-learning/education-initiatives/connected-curriculum>; last accessed 7/1/17.

3. Students make connections across subjects and out to the world;
4. Students connect academic learning with workplace learning;
5. Students learn to produce outputs – assessments directed at an audience;
6. Students connect with each other, across phases and with alumni.

Research data can also be seen as a learning object. In a digital environment, research outputs cannot be restricted/limited to traditional written works such as journal articles or monographs. Nowadays, research outputs consist of a mixture of objects, amongst which can be found written works and data. One of the building blocks for these publications is research data. Via digital networks, it is possible to share both publications and the underlying data to anyone who can access them. The emergence of research data as a major source of information is now becoming apparent. To take advantage of this revolution researchers, especially early career researchers, need to be trained in best practice in research data management. This Case Study offer one example of how this can be done.

18.2. EARLY CAREER RESEARCHERS

In 2001, a US study sponsored by the Pew Charitable Trusts found that:⁵

Students in 11 arts and sciences disciplines from 27 institutions and 1 cross-institutional program [...] were surveyed. Responses were received from 4,114 students, a response rate of 42.3%. Results suggest that the training doctoral students receive is not what they want, nor does it prepare them for the jobs they take. Many students do not understand what doctoral study entails, how the process works, and how to navigate it effectively. There is a mismatch among the purpose of doctoral education, the aspirations of the students, and the realities of their careers within and outside academia.

In 2017, the situation is better. The UCL Doctoral School, in its Code of Practice, stresses:⁶

UCL offers a programme for the development of generic research and personal transferable skills to help you develop the skills necessary not only for successful completion of your degree but also to equip you for later life and for the workplace ... The specific menu of courses and other training opportunities should be discussed between you and your Supervisors using the skills self assessment section of UCL's Research Student Log. The self-assessment process is based on a national framework, the Researcher Development Framework.

It follows that the need for skills development has been identified and courses/materials put in place. One of those training needs concerns research data management.

⁵ Golde, Chris M., Dore, Timothy M., *At Cross Purposes: What the Experiences of Today's Doctoral Students Reveal about Doctoral Education*, Wisconsin University, Madison, 2001; available at <http://files.eric.ed.gov/fulltext/ED450628.pdf>; last accessed 7/1/17.

⁶ UCL: <http://www.grad.ucl.ac.uk/codes/Graduate-Research-Degrees-Code-of-Practice-1617.pdf>; last accessed 7/1/17.

18.3. THE LERU DOCTORAL SUMMER SCHOOL AS A MODEL OF BEST PRACTICE

LERU itself has produced a LERU Roadmap for Research Data.⁷ Chapter 6 looks at Roles, Responsibilities and Skills and identifies a need for training for early career researchers, for academics and for support staff. The training needs and routes for skills development are clearly identified in Figure 18.1 below. The separate categories are not mutually exclusive. All stakeholders – student/PhD + Senior Researcher + Librarian + Data Scientist need to work together to share knowledge and Best Practice. Nonetheless, the categorisation in Figure 18.1 does attempt to codify the learning needs of each stakeholder group and how these needs can realistically be met. It accepts that there is a graduated series of learning needs, starting with postgraduate/PhD students, and which increase in complexity as early career researchers become Senior Researchers. In this model, Librarians have a new role to play in the research space. They need to acquire new skills and to impart that knowledge to the groups that they train. This partnership is crucial in embedding RDM skills into the research landscape. Finally, there is the emerging new career of Data Scientist, and this is discussed more fully in the section below on the European Open Science Cloud.

WHO	Postgrad/PhD	Senior Researcher	Librarian	Data Scientist
WHEN	Early stages of postgraduate study	As needed, or at beginning of research project/proposal state	CPD for subject librarians/during library education	Discipline-specific academic courses (doctoral)/CPD
WHAT	Basics of data management practice, FAIR ⁸ principles, data citation, data evaluation. Competence in legal and ethical issues	Training on discipline-specific data management practices; an understanding of the FAIR principles; how to write a data management plan (tailored as necessary to funder requirements), data reuse skills. Competence in legal and ethical issues	Data curation. An understanding of the FAIR principles. Some disciplinary-specific e-research methods (TDM)/data collection skills, IT skills. Competence in legal and ethical issues	Discipline-specific skills for data management/exploitation/interoperability. An understanding of the FAIR principles. Competence in legal and ethical issues
HOW	Credited models	Practical training	Accredited CPD/ Professional courses	Professional (academic) courses and accredited CPD

Figure 18.1: Training needs and routes for skills development

Having identified the training needs, how can those needs be met? The LERU Roadmap suggests that, for most categories of user, what is required are credited models and/or professional courses. LERU universities have taken this to the next stage by devising a format for a formal Summer School to train PhD students new to research and to RDM. The first meeting was held in Leiden in Summer 2016.⁹ This is a taster for future activity, which is currently being discussed in the LERU network.

The Programme for the Summer School¹⁰ had as its ambition the creation of the ‘new generation of data scientists’. Each of the 21 LERU member universities¹¹ was invited to send one or more members of their doctoral programme to attend the week, the intention being that having received training in Leiden they

⁷ LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf; last accessed 7/1/17.

⁸ Force11: <https://www.force11.org/group/fairgroup/fairprinciples>; last accessed 29/1/17.

⁹ LERU Doctoral Summer School (2016): <http://www.dtls.nl/fair-data/fair-data-training/leru-summer-school/>; last accessed 7/1/17; and for general background LERU Doctoral Summer School (2016): <http://www.dtls.nl/first-kind-leru-doctoral-summer-school-data-stewardship/>; last accessed 7/1/17.

¹⁰ LERU Doctoral Summer School (2016): <http://www.dtls.nl/fair-data/fair-data-training/leru-summer-school/programme/>; last accessed 29/1/17.

¹¹ LERU: <http://www.leru.org>; last accessed 29/1/17.

could then return to their organisations and cascade that knowledge around local doctoral candidates. 38 doctoral students attended the event from 21 universities and associated hospitals.¹²

The format of the Programme was to aim for a mixture of keynote speakers on specific topics, speakers to lead in particular thematic areas and student presentations/discussions. The Summer School highlighted a number of issues, which are likely to form the core of RDM training activity going forward. Some of the more prominent are listed here:

- The importance of research data being FAIR (Findable, Accessible, Interoperable and Reusable)¹³
- The importance of data management plans in providing a framework for the creation, storage, and sharing of research data¹⁴
- Licensing issues and an explanation of the meaning of the Creative Commons suite of licences and its use in research data¹⁵
- Big Science is Open Science¹⁶
- The future infrastructure for Open Science¹⁷

18.4. TOP-LEVEL ISSUES CONCERNING RESEARCH DATA FOR THE LERU SUMMER SCHOOL

FAIR data is one of the building blocks of the new information age. If research data is findable, accessible, interoperable and reusable, it increases in value as a tool for supporting innovation and new discoveries. Effective licensing of research data, when needed, increases their usefulness and makes it clear what the terms of re-use are. One of the drawbacks of the early development of Open Access is that many of the published research outputs tagged as Open Access outputs have no accompanying licence. This makes it difficult to understand exactly what the rules for reuse are in every case. Moreover, the lack of a licence has to be interpreted as all rights reserved in accordance with copyright law. Can an Open Access publication, with no accompanying licence, be re-used for commercial advantage?

Not all research data is big data. Many collections of data form part of a long tail of data creation, where research data has been created/collected to support the publication of a particular article, or a lecture to taught-course students. The term 'big data' is sometimes overused and brings with it legal issues such as privacy into the discussion. Nonetheless, the best future for research data, whether big or small, is that it is open where that is legally possible. Finally, to deliver and perform Open Science, infrastructure is needed – not simply technical platforms but also training and skills development programmes to create the 'new generation of data scientists'. All this and more was discussed in a focussed and intensive week in the LERU Doctoral Summer School.

A particularly important part of the model for the Summer School was the balance between formal presentations and the opportunity for students themselves to present case studies using their own research data, and to interact with speakers.¹⁸ 'With data sharing, new scientific discoveries can be made' was one

¹² LERU Doctoral Summer School (2016): <http://www.dtls.nl/fair-data/fair-data-training/leru-summer-school/students/>; last accessed 29/1/17.

¹³ FAIR: <https://www.force11.org/group/fairgroup/fairprinciples>; last accessed 7/1/17

¹⁴ Digital Curation Centre: <http://www.dcc.ac.uk/resources/data-management-plans>; last accessed 7/1/17.

¹⁵ Creative Commons: <https://creativecommons.org/licenses/>; last accessed 7/1/17.

¹⁶ For a description of Big Data in the science ecosystem, see Royal Society, *Science as an Open Enterprise* (London: Royal Society, 2012), available at <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>; last accessed 7/1/17.

¹⁷ European Commission: <http://ec.europa.eu/research/openscience/index.cfm>; last accessed 7/1/17.

¹⁸ Some of this is captured on the summary video at LERU: <https://www.youtube.com/watch?v=sSQPY5Mc5Rs>; last accessed 7/1/17; and in the tweets recorded at LERU Summer School (2016): <https://twitter.com/hashtag/lerusummerschool2016?src=hash>; last accessed 7/1/17.

of the recorded tweets. Other participants felt that the Summer School was a valuable mirror to reflect how science is done in the twenty-first century. The participants expressed real enjoyment at being able to participate in the event. In fact, they wished they had had more time to discuss the new information that they were learning each day. With feedback like this, the objective of the Summer School to provide solid training in data stewardship for the next generation of future leaders does not seem to have been unrealistic.

18.5. THE FUTURE PATTERN OF SKILLS DEVELOPMENT: THE EUROPEAN OPEN SCIENCE CLOUD?

In July 2016, the European Commission published the Report of its High Level Expert Group on the European Open Science Cloud:¹⁹

... The European Open Science Cloud (EOSC) aims to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market. It should enable trusted access to services, systems and the re-use of shared scientific data across disciplinary, social and geographical borders. The term cloud is understood by the High level Expert Group (HLEG) as a metaphor to help convey both seamlessness and the idea of a commons based on scientific data. This report approaches the EOSC as a federated environment for scientific data sharing and re-use, based on existing and emerging elements in the Member States, with lightweight international guidance and governance and a large degree of freedom regarding practical implementation. The EOSC is indeed a European infrastructure, but it should be globally interoperable and accessible. It includes the required human expertise, resources, standards, best practices as well as the underpinning technical infrastructures. An important aspect of the EOSC is systematic and professional data management and long-term stewardship of scientific data assets and services in Europe and globally. However, data stewardship is not a goal in itself and the final realm of the EOSC is the frontier of science and innovation in Europe.

Important in this summary of activity, for present purposes, is the recognition of the importance of skills development. The LERU Roadmap itself identified the category of Data Scientist as the summation of skills development in terms of research stakeholders. This identification is further developed in the EOSC Report by emphasising the absolute importance of developing the role of data stewards to deliver the vision of a global commons of scientific data. The Report suggests:

A first cohort of core data experts should be trained immediately to translate the needs for data driven science into technical specifications to be discussed with hard-core data scientists and engineers. This new class of core data experts will also help translate back to the hardcore scientists the technical opportunities and limitations.²⁰

Elsewhere, the Report puts figures to the training requirement:

The number of people with these skills needed to effectively operate the EOSC is, we estimate, likely exceeding half a million within a decade. As we further argue below, we believe that the implementation of the EOSC needs to include instruments to help train, retain and recognise

¹⁹ European Commission: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>, p. 6; last accessed 7/1/17.

²⁰ Ibid., p. 16.

this expertise, in order to support the 1.7 million scientists and over 70 million people working in innovation. The success of the EOSC depends upon it.²¹ These are significant numbers. It will take a significant investment in European teaching infrastructures to develop the curricula, agree success criteria for measuring successful delivery and finance this huge training undertaking. Commissioner Moedas (Research, Science & Innovation), however, has highlighted the need for skills development and has said, ‘Such recommendations deserve detailed consideration by the scientific community and other stakeholders.’²²

Research performing organisations need to start somewhere, as the LERU Roadmap makes clear. In this context the model of the LERU Doctoral Summer School seems a measured, successful and immediate response that such bodies need to make to manage the training needs implicit in the data deluge.

18.6. CONCLUSION

The purpose of this Case Study has been to look at the role of research performing organisations in skills development for early career researchers, and to set those needs in the context of the growing importance of research data and the emerging role of the data steward. In 2001, a North American survey decried the value and use of much doctoral training activity. In 2017, examples from Western Europe show that levels of provision, and the understanding of the need, have improved. The LERU Roadmap for Research Data (2013)²³ identified clear needs for research data training and the LERU Doctoral Summer School in Leiden (2016) provides a Best Example Case Study of how that training can be delivered in practice. The scale of the need, however, is illustrated by the Report of the High Level Expert Group on the European Open Science Cloud – half a million skilled data stewards needed in the next 10 years. If true, there is no room for complacency and every need for action. The LERU Summer School provides an excellent model for the training seminars needed to deliver generic skills and subject-specific insights into the emerging activity of data stewardship.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

²¹ Ibid., p. 12.

²² Ibid., p. 4.

²³ LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf; last accessed 7/1/17.

Training subject librarians in Research Data Management

Authors: June Hedges (Head of Liaison and Support Services, Library Services, University College London) and Myriam Fellous-Sigrist (Research Data Support Officer, Library Services, University College London)

Email: m.fellous-sigrist@ucl.ac.uk

19.1. INTRODUCTION

University College London (UCL) ranks among the top twenty universities in the world and is one of the most successful British research institutions at attracting funding. Almost all academic disciplines are represented in its 380 research departments, units, institutes and centres¹. UCL is home to 12,000 research staff and research students².

UCL Library Services run eighteen libraries which support UCL's teaching and research activities, including one in the award-winning School of Slavonic and East European Studies building and several that provide services to both UCL and the National Health Service. The combined staff in UCL Library Services totals 263 FTE (full-time equivalents). Amongst this number are around 30 subject liaison and site librarians who have responsibility for supporting the research and teaching of the institution. These librarians are the primary points of contact for academics, researchers, UCL staff and students. They provide subject-specific support and advice on resources and collections, offer training to staff and students, and promote and provide training on the various teaching and research support services that the Library offers, including open access services.

Two Research Data Support Officers work as part of the same team as well as in close collaboration with the UCL Information System Division (IT Services) and several other central services. These officers coordinate Research Data Management (RDM) advocacy and support across the institution. To ensure the long-term sustainability and scalability of the RDM support service, as well as sufficient subject discipline support, the RDM team aims to foster several support networks of subject-specific experts across the university. The subject liaison and site librarians form one of these networks.

19.2. INTRODUCING LIBRARIANS TO RDM THROUGH WORKSHOPS

The first UCL Research Data Policy (launched in August 2013) was accompanied by introductory presentations on RDM and related service developments for Library Services' staff. A programme of three day-long workshops was subsequently planned to inform and train library staff about current issues in Research Data Management, from key definitions up to the review of Data Management Plans. These workshops took place in 2015; they gathered between 30 and 35 participants each. The sessions were designed and delivered by Data Management experts from the Information School of the University of Sheffield. The outline for each workshop was as follows:

Session 1: This workshop provided an introduction to research data and its management in the context of UCL and the Library's role. Topics covered included the nature of data and research data services; data

¹ As listed in the UCL Departments A to Z (at www.ucl.ac.uk/departments/a-z/; accessed 4 August 2016).

² Figure from UCL Human Resources as of 1st October 2015 and Registry Services as of 1st December 2015.

management planning; information security; and issues around Text and Data Mining. At the end of the session, librarians were tasked with talking to a researcher about RDM in advance of the next workshop.

Session 2: This workshop started with presentations on what participants had learnt from their conversations with researchers. This was followed by a discussion about the survey approach to gathering further information about data management in the university. Librarians looked at identifying key choices in planning training and at the issues around selecting, describing and citing data. At the end of the day, participants were tasked with group exercises to prepare the last workshop.

Session 3: In this workshop, participants heard from each group what their ideas and plans were for addressing the various aspects of RDM support identified in the first two sessions. Librarians presented group reports on practical RDM, requirements gathering (and in particular the Data Asset Framework method³), data sharing, data sources and RDM websites.

The first two sessions comprised presentations by the workshop leaders introducing a new area, approach or issue coupled with group activities which led to the final project of creating a plan on how to respond to key aspects of RDM. This was done in small groups working together outside the workshops to prepare their response and also a brief presentation to be delivered at the final workshop.

Feedback from the workshops was positive and librarians welcomed the thoroughness of the programme, which did cover all the key aspects of RDM. They also welcomed the opportunity to work collaboratively with colleagues from across Library Services. The only negative feedback received was that the workshops had been spread over a 6-month period with 2-3 month intervals between each.

In 2016 a fourth workshop focused on central RDM services run by UCL Library Services and the UCL Information Systems Division jointly. The 30 participants discussed the roles and interaction of these different services, and how these are explained to researchers and research students across all disciplines. This event was also planned as a networking opportunity which gathered together for the first time librarians, Research IT staff and departmental data managers.

Feedback received from this event showed that it helped to meet colleagues across the university and to understand how the different services join up. Theoretical presentations were considered to be less useful, and several participants suggested that having small group activities to put into practice what was said in the presentations would enhance their learning; this would include creating flow-charts to explain data storage processes within the university and drafting discipline-specific guidance for researchers and students.

19.3. INVOLVING LIBRARIANS IN RDM THROUGH PARTICIPATION IN A WORKING GROUP

Subsequent to formal RDM training, librarians were given the opportunity to apply their knowledge by actively contributing to a new Working Group. The Library RDM Working Group was created in 2015. It supports the two Research Data Support Officers with discipline-specific knowledge and essential staff resource for short-term projects. This Group is made up of thirteen volunteers (Librarians, Records Manager, Digital Curation Manager, Research Data Support Officers) who work on a specific project each summer; not all Working Group members are required to participate in all projects.

³ This method is explained by the Digital Curation Centre at www.dcc.ac.uk/resources/tools/data-asset-framework/; accessed 20 December 2016.

In 2015, the Group first concentrated on building the new RDM website, and second on designing and promoting a cross-university RDM survey. For the first activity, all Group members were trained to improve their knowledge of the Library Content Management System and to write for the web. They worked in pairs to draft, edit and publish online the webpages that they had chosen to work on. The website⁴ was completed in two months and launched at the start of the academic year 2015/2016. It featured nine how-to guides, a section about the university's and research funders' policies on research data, key definitions about RDM, a searchable list of Frequently Asked Questions, and a selection of resources and tools to learn more about RDM. The website is regularly updated since its launch and new resources have since been added. The survey was designed, tested and promoted by five members of the RDM Working Group between the summer 2015 and winter 2015-2016. The exercise was primarily aimed at finding information about awareness, practices and needs related to RDM across all faculties. Analysis of the results helped assess what support was needed by researchers with regard to RDM, and how this should be prioritised by UCL Library Services and other research support services across the university.

In summer 2016, the Group worked on creating discipline-specific resources to help researchers throughout their research projects; such resources include RDM guidance, metadata standards, data repositories and ethics guidelines. A second completed project was the design of a course template to introduce research students to RDM. The template consists of a series of presentation slides, a lesson plan and guidance to deliver the course. The course was tested in autumn and winter 2016 by three members of the Group with cohorts of Masters and PhD students. It serves as an essential basis to develop future courses on RDM at a more advanced level and aimed at further communities across the university.

RDM Working Group members cite their primary reason for volunteering to take part as being the opportunity to extend their knowledge of RDM both to fulfil personal interest, but also to provide extended research support to the departments with which they work. In the case of subject liaison librarians, a greater knowledge of RDM has been a means to establish new points of contact within the academic communities that they support.

19.4. CONCLUSIONS

RDM training will continue within UCL Library Services to ensure that subject liaison and site librarians' knowledge stays up to date. Currently, future plans include a 'train the trainer' session to help them deliver introductory courses on RDM in research departments. A session on reviewing Data Management Plans (DMPs) is also being designed as several librarians have expressed the need to be able to follow-up with enquiries on DMPs once they have delivered the introductory course.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁴ UCL: www.ucl.ac.uk/research-data-management; accessed 3 November 2016.

The Emerging Role of the Data Scientist and the experience of Data Science education at the University of Amsterdam

Author: Yuri Demchenko (Senior Researcher, System and Network Engineering Research Group. University of Amsterdam)

Email: y.demchenko@uva.nl

20.1. MOTIVATION AND BACKGROUND

Modern research requires new types of specialists that are capable of supporting all stages of the research data lifecycle – from data production and input to data processing, storage, and the publishing and dissemination of scientific results, which can be jointly defined as key components of the emerging profession of Data Science (DSP).

To address this demand from research and industry, the Horizon 2020 Programme is funding the EDISON Project (Grant 675419, INFRASUPP-4-2015: CSA),¹ the goal of which is to build the Data Science profession for European research and industry. This includes the definition of Data Science and data handling-related professional profiles (or occupations), corresponding core competences and skills, the Data Science Body of Knowledge and a Model Curriculum that together comprise the EDISON Data Science Framework. This work is done with the involvement of the main stakeholders from the research community, industry, data preservation and handling community, universities and professional training organisations.

The University of Amsterdam is coordinator and a base organisation for the EDISON Project; other partners include the University of Stavanger (Norway), the University of Southampton (UK), Engineering Italy, EGI.eu, FTK (Germany), and Inmark Europe (Spain). The project benefits from multiple Data Science-related initiatives and academic activity and effective cooperation between Computer Science and multi-disciplinary departments, University Library and IT departments. It is also supported by such external initiatives as the Amsterdam Data Science Centre and Amsterdam School of Data Science (ASDS). On the other hand, all project recommendations find their practical pilot implementation at the University of Amsterdam and in cooperating organisations. This includes four Data Science and Big Data programmes, Research Data Management (RDM) training (together with the University Library), training for researchers, programmes and course catalogue services for universities and students, and advice for companies.

20.2. STAKEHOLDERS AND THEIR ROLE IN DATA SCIENCE EDUCATION

To create a foundation for the sustainable education and training of future Data Science professionals and Core Data Experts to support present and future data-driven research, the EDISON Project involves and is cooperating with multiple stakeholders, relevant bodies and communities. This includes but is not limited to the following:

- **Academic and research departments** are key for developing and teaching educational courses on Data Science and Research Data Management: four different Data Science programmes have started in the 2016-17 academic year, targeting different demand sectors in research and industry

¹ EDISON: <http://edison-project.eu/>; accessed 5 February 2017.

(see below for a description of the programmes). Course development, teaching and support is provided primarily by departmental staff with some facility services maintained by ICT departments.

- **The University Library** is involved in two main activities: (i) it provides basic training for researchers and contributes to the more general academic education for students in RDM; (ii) it cooperates with the ICT department in developing and implementing university-wide RDM services, infrastructure and policy.
- **The ICT department** supports Data Science education by providing and maintaining HPC facilities and services. The ICT department cooperates with the University Library in implementing RDM infrastructure and policy university-wide.

20.3. EDISON DATA SCIENCE FRAMEWORK

The EDISON Data Science Framework (EDSF),² is a core product of the EDISON Project that provides a basis for the definition of the whole ecosystem for education, training and professional development in core Data Science and Data Management-related competences and skills. An important component of EDSF is the Data Science professional family that provides a basis for defining customisable educational and training programmes for different target professional groups. Figure 20.1 below illustrates the main EDSF components:

- CF-DS – Data Science Competence Framework
- DS-BoK – Data Science Body of Knowledge
- MC-DS – Data Science Model Curriculum
- DSP - Data Science Professional profiles and occupations taxonomy
- Data Science Taxonomy and Scientific Disciplines Classification (including Vocabulary)

The proposed framework provides a basis for other components of the Data Science professional ecosystem:

- EDISON Online Education Environment (EOEE)
- Education and Training Marketplace and Directory
- Data Science Community Portal (CP) that also includes tools for individual competences benchmarking and personalized educational path building
- Certification Framework for core Data Science competences and professional profiles

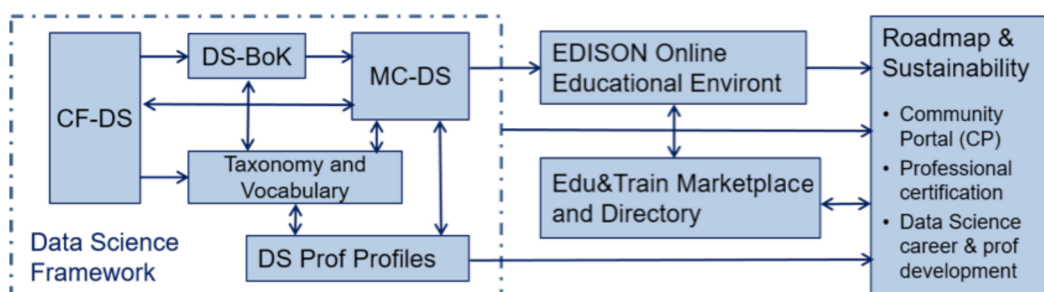


Figure 20.1 EDISON Data Science Framework components.

² EDISON: <http://edison-project.eu/edison/edison-data-science-framework-edsf>; accessed 5 February 2017.

The CF-DS includes common competences required for the successful work of Data Scientists in different work environments in industry and in research and throughout the whole career path. Future CF-DS development will include coverage of domain-specific competences and skills and will involve domain and subject matter experts.

The DS-BoK defines the Knowledge Areas and Knowledge Units for building Data Science curricula that are required to support specified Data Science competences. DS-BoK is organised by Knowledge Area Groups (KAG) that correspond to the CF-DS competence groups. DS-BoK incorporates best practices in Computer Science and domain-specific BoK's and includes KAs based on the Computing Classification System (CCS2012), components taken from other BoKs and proposed new KAs to incorporate new technologies used in Data Science and their recent developments.

The MC-DS is built based on CF-DS and DS-BoK where Learning Outcomes are defined based on CF-DS competences, and Learning Units (LU) are mapped to Knowledge Units in DS-BoK. Three mastery (or proficiency) levels are defined for each Learning Outcome to allow for flexible curricula development and profiling for different Data Science professional profiles.

The DSP profiles and Data Science occupations taxonomy are defined based on, and as an extension to, the European Skills, Competences, Qualifications and Occupations (ESCO) framework. The DSP profiles definition provides an instrument to create effective organisational structures and corresponding roles to support the whole data management lifecycle. For example, in the area of professional data handling/management, the following taxonomy is proposed: Professional (data handling/management): Data Stewards, Digital Data Curator, Digital Librarians, Data Archivists. DSP can also be used for building individual career paths and corresponding competences and skills transferability between organisations and sectors of the economy.

20.3.1. Data Science Competence Framework (CF-DS)

The Data Science Competence Framework (CF-DS)³ has been built based on an extensive study of the demand and supply side of the Data Science job market, organisational structures and roles as well as existing practices and standards in the area of competences and skills management. The figure below [20.2] presents the following competences:

Three competence groups identified in the NIST document and confirmed by the analysis of collected data:

- Data Analytics including statistical methods, Machine Learning and Business Analytics
- Engineering: software and infrastructure
- Subject/Scientific Domain competences and knowledge

Two newly identified competence groups that are in high demanded and are specific to Data Science

- Data Management, Curation, Preservation (new)
- Scientific or Research Methods (new)

³ EDISON: <http://edison-project.eu/data-science-competence-framework-cf-ds>; last accessed 9 February 2017.

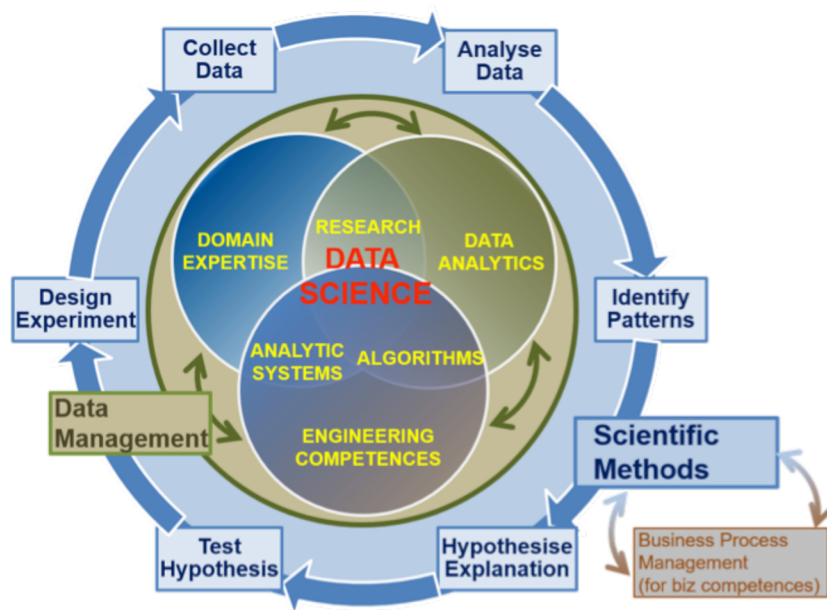


Figure 20.2. Data Science competence groups

Knowledge of scientific research methods and techniques makes the Data Scientist profession different from all previous professions. For business-related professions, a similar role belongs to business process management in areas that need to be adapted to a new data-driven agile business model, in particular, to adopt continuous data-driven business processes improvement.

Data management, curation and preservation are already included in existing (research) data-related professions such as data steward, data archivist, data manager, digital librarian, data curator, and others. Research data management is an important component of European Research Area policy. Companies also recognise the need for data management skills when they start using data-driven technologies.

The identified demand for general competences and knowledge of Data Management and Research Methods needs to be addressed in future Data Science education and training programmes, as well as being included in re-skilling training programmes. It is important to mention that knowledge of Research Methods does not mean that all Data Scientists must be talented scientists; however, they need to understand general research methods such as formulating an hypothesis, applying research methods, producing artefacts, and evaluating an hypothesis (so called 4 steps model). Research Methods training is already included into Masters programmes and for graduate students.

The identified competence areas provide a basis for defining education and training programmes for Data Science-related jobs, re-skilling and professional certification.

Other skills commonly recognised are referred to as “soft skills” or “social/professional intelligence”: interpersonal skills or team work, the ability to cooperate. In many cases, an organisation expects the Data Scientist to provide a kind of literacy advice and guidance on related data analysis and management technologies.

20.3.2. Data Science Body of Knowledge (DS-BoK)

The DS-BoK should contain the following Knowledge Area groups (KAG) that are defined after CF-DS competence groups:

- KAG1-DSDA: Data Analytics group including Machine Learning, statistical methods, and Business Analytics
- KAG2-DSENG: Data Science Engineering group including Software and infrastructure engineering
- KAG3-DSDM: Data Management group including data curation, preservation and data infrastructure
- KAG4-DSRM: Scientific or Research Methods group
- KAG5-DSBPM: Business process management group
- KAG6-DSDKX: Data Science Domain Knowledge group, which includes domain-specific knowledge

Universities can use DS-BoK as a reference to define knowledge areas that they need to cover in their programmes depending on the primary demand groups in research or industry. Domain-specific knowledge can be acquired as a part of academic education or as postgraduate professional training at the graduate's work place. It is also commonly recognized that KAG6-DSDKX is essential for the practical work of a Data Scientist, which means that Data Scientists need to have sufficient understanding of specific subject domain-related concepts, models, organisation and corresponding data analysis methods to effectively communicate with domain-related specialists for data collection, insight and the presentation of results.

20.3.3. Data Science Model Curriculum (MC-DS)

The initial Data Science Model curriculum provides two basic components for building customisable Data Science curricula: (1) the definition of a learning outcomes (LO) based on the CF-DS competences, including their differentiation for different proficiency levels, e.g. using Bloom's Taxonomy, (2) definition of the Learning Units (LU) that map to the LOs for target professional groups, which need to be defined in accordance with existing academic discipline classifications such as the 2012 ACM Computing Classification System (CCS2012).⁴

20.3.4. Data Science Professional Profiles Definition (DSPP)

The proposed Data Science Professional profiles (DSPP)⁵ definition is based on the analysis of the demand in research and industry in data-related professions as well as in current company practices in defining new data-related organisational roles. The identified professional profiles are classified using ESCO taxonomy⁶, and necessary extensions are proposed to support the following hierarchy of data handling-related occupations (see Figures 20.3 & 20.4):

- Managers: Chief Data Officer (CDO), Data Science (group/department) manager, Data Science infrastructure manager, Research Infrastructure manager
- Professionals: Data Scientist, Data Science Researcher, Data Science Architect, Data Science (applications) programmer/engineer, Data Analyst, Business Analyst, etc.
- Professionals (database): Large scale (cloud) database designers and

⁴ ACM: <http://www.acm.org/about/class/class/2012>; last accessed 9 February 2017.

⁵ EDISON: <http://edison-project.eu/data-science-professional-profiles-definition-dsp>; accessed 5 February 2017.

⁶ European Commission: <https://ec.europa.eu/esco/portal/home>; last accessed 9 February 2017.

- administrators, scientific database designers and administrators
- Professional (data handling/management): Data Stewards, Digital Data Curator, Digital Librarians, Data Archivists
- Technicians and associate professionals: Big Data facilities operators, scientific database/infrastructure operators
- Support and clerical workers: Support and data entry workers.

The competences and skills required for different professions are defined in the DSP Profiles document in accordance with the Data Science Competence Framework (CF-DS). An example of mapping CF-DS competences to identified data handling-related occupations is provided.

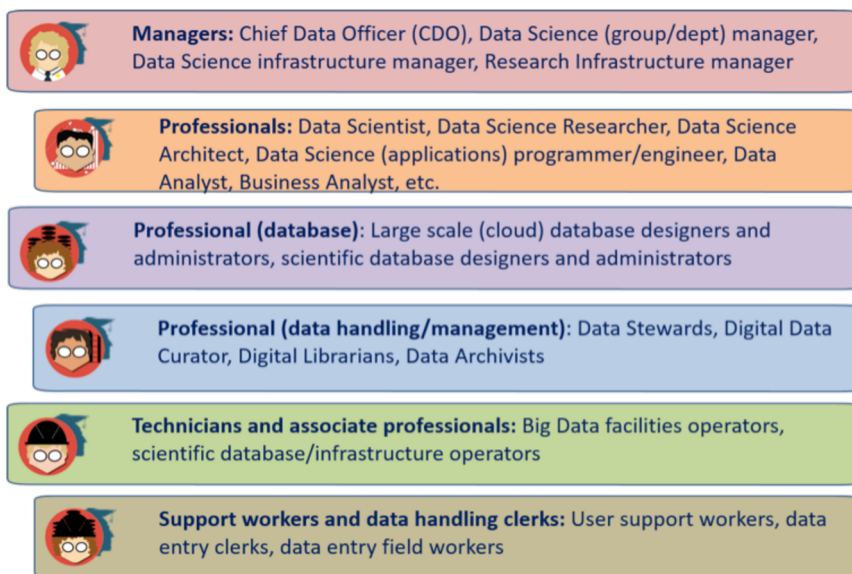


Figure 20.3 Data Science Professions family groups

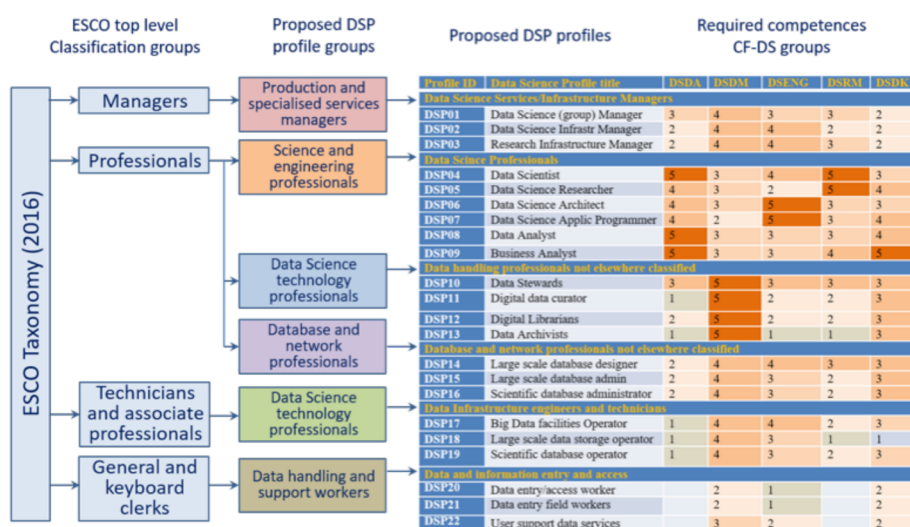


Figure 20.4. Data Science Professions family groups with hierarchy

20.4. DATA SCIENCE PROGRAMMES IMPLEMENTATION AT THE UNIVERSITY OF AMSTERDAM (UVA)

The University of Amsterdam is starting 4 new Data Science programmes and tracks that are based on/originate from different departments, and which are aimed at different industries and target groups from Computer Science, Business Administration, and multidisciplinary studies. They are primarily intended to answer the needs of the Dutch economy (i.e. industry, research and public services) which is to a large extent international. The programmes and tracks are developed by the departments independently, but all of them use general EDISON recommendations.

i. Artificial Intelligence and Data Science (specialisation)

(<http://gss.uva.nl/future-msc-students/information-sciences/content26study-programme/profile-data-science.html>)

Track - Master

At the core of Data Science are methods for the analysis of large volumes of data. Recently much more data has become available in electronic form, and methods for the analysis and modelling of these data for prediction, classification and optimisation have become much more effective. Recent technical innovations, such as Deep Learning, provide increasingly powerful tools that make it possible to find complex patterns in very large datasets.

Much of the Master's Artificial Intelligence (AI) degree is about Data Science. The obligatory courses on Machine Learning address key technology and theory for modelling large amounts of data. The courses on Machine Learning, Natural Language Processing, Information Retrieval and Computational Intelligence all have a strong focus on data-driven methods. For the "AI courses" in the curriculum, students can choose advanced courses on these topics: Machine Learning 2, Computer Vision 2, Natural Language Processing 2, Information Retrieval 2, Deep Learning, Data Mining Techniques, Information Visualisation and Probabilistic Robotics. All these courses are about modelling data. These can be complemented by courses outside AI, for example on distributed computer systems, privacy and ethical questions, or on statistics.

Within programme: Artificial Intelligence

Organisation: UvA

Language: English

Duration: 5 months

ii. Big Data Engineering

(<http://gss.uva.nl/future-msc-students/information-sciences/content28/computer-science.html>)

Track - Master

In the Internet era, data is at the centre of the stage. We all continuously communicate via social networks, we expect all information to be accessible online continuously, and the world's economies thrive on data processing services where revenue is created by generating insights from raw data. These developments are enabled by a global data processing infrastructure, connecting everyone from small company computer clusters to data centres run by world-leading IT giants. In the Big Data Engineering track, you study the technology from which these infrastructures are built, allowing you to design and operate solutions for processing, analysing and managing large quantities of data. This track is part of the joint Masters in

Computer Science, in which renowned researchers from both the Vrije Universiteit Amsterdam (VU) and the University of Amsterdam (UvA) contribute their varied expertise in one of the strongest Computer Science programmes available in Europe.

Within programme: Computer Science

Organisation: UvA + VU

Language: English

Duration: 2 years

iii. MBA Business Analytics & Data Science

(<http://abs.uva.nl/programmes/mba/content2/mba-big-data.html>)

Track – Master MBA

This MBA in Big Data and Business Analytics is intended for hands-on Big Data specialists, for people in leadership roles working with Big Data and for Entrepreneurs. The curriculum of this MBA is highly multidisciplinary, with courses from A (analytics), B (business) and C (computer science), and with projects to practise and implement the integration of these three aspects.

Furthermore, the curriculum is a mix of state-of-the art theory taught by renowned academic professors, and it includes practical applications of this knowledge taught by people with extensive industry experience. In the curriculum, much time will be devoted to the '21st century skills' - the skills required to become successful in this age: entrepreneurship / entrepreneurial attitude, flexibility, teamwork, communication skills and ethics.

Key features:

- Two-year part-time programme (2 evenings per week);
- Balanced curriculum consisting of Business courses (e.g. strategy, finance, marketing, HRM), Analytics courses (e.g. statistics, econometrics, system optimization) and Computer Science courses (e.g. machine learning, data visualisation);
- All lecturers combine theory with practical applications;
- Silicon Valley study trip and Big Data Thesis Project will be part of the programme;
- Degree: Master of Business Administration (MBA) granted by the University of Amsterdam;

Organisation: Amsterdam Business School, UvA

Language: English

Duration: 2 years

iv. Data Science

(<http://gss.uva.nl/future-msc-students/information-sciences/content/data-science.html>)

Track - Master

In the one-year Data Science Master's track, you will acquire knowledge of the theories and tools used in data science. We will teach you how to use these tools for working with data in different domains, such as Healthcare, Media and Communication, Smart City, Life Sciences and Digital Humanities. Graduates have an integrated view on the possibilities and development of data science in society. Students will benefit from the strong collaboration with Amsterdam Data Science (ADS), bringing together leading researchers across

the entire life cycle of data science, from expertise in machine learning and information retrieval to human computer interaction and large-scale data management.

Within programme: Information Studies

Organisation: UvA + VU

Language: English

Duration: 1 year

20.5. RESEARCH DATA MANAGEMENT EDUCATION AND TRAINING

Research Data Management training is recognised as essential for practising researchers of all scientific domains and important for academic Data Science education. It is typically covered by training programmes for postgraduates, PhD students and researchers; however it is rarely covered by existing or planned academic programmes and courses. It has been identified that to cover the wide needs of the research and academic community, the RDM curriculum and training materials must allow easy customisation and localisation to adjust to the trainees' background and local infrastructure resources, as well as to cater for the needs of specific scientific domains.

The EDISON Project has addressed RDM training and education as a priority issue in order to contribute to raising standards in general competences and skills related to working with research data and with the variety of modern data including social (network) data, environmental data and business data. The EDSF provides a basis for defining a general RDM training program that covers the major practical aspects of RDM; this can be also considered as an important component of more general data literacy training.

The proposed customisable RDM training program

The following RDM training program has been constructed based on an extensive study of existing RDM training programmes and resources, in particular collected at the Data Management Clearinghouse⁷ and by the RDA US directory of RDM resources⁸. It covers most topics available in currently-available RDM training programmes and curricula, has a modular structure and provides the possibility of expanding into more specific data management topics that may be required by specific groups of practitioners.

A Research Data Management training or education programme should contain the following essential modules (allowing extension and adoption to particular target communities):

A. Use cases for data management and stewardship

- Preserving the Scientific Record

⁷ DM-Clearinghouse, 2016. Data Management Training Clearinghouse. Available at <https://www.sciencebase.gov/catalog/item/56d88012e4b015c306f6cfc>; accessed 5 February 2017.

⁸ RDA-US-RDM, 2016, *RDA US directory of RDM resources*, 2016. Available at https://docs.google.com/spreadsheets/d/10RTW-nZk0x_mPQw2VAlttcc656MV9EeCaDe2IM4umb4/edit#gid=0; accessed 5 February 2017.

B. Data Management elements (organisational and individual)

- Goals and motivation for managing your data
- Data formats
- Creating documentation and metadata, metadata for discovery
- Using data portals and metadata registries
- Tracking data usage
- Backing up your data
- Data security and integrity
- Data Management Plan (DMP) (also a part of hands on session(s))

C. Responsible Data Use Section (Citation, Copyright, Data Restrictions)

- Handling sensitive data
- Ethical issues, obtaining consents

D. Open Science, Open Access and Open Data (Definition, Standards, Open Data use and re-use, open government data)

- Research data and open access
- Repository and self-archiving services
- PID identifier for data and ORCID identifier for researchers
- Stakeholders and roles: engineer, librarian, researcher
- Open Data services: ORCID.org, Altmetric Donut, Zenodo

E. Hands on and labs:

- a) DMP design
- b) Metadata and tools
- c) Selection of licenses for open data and contents (e.g. Creative Common and Open Database)

The proposed RDM training program has been taught at the Data Science workshop since May 2016 at Amsterdam Business School, University of Amsterdam⁹ organized by the EU Erasmus+ Eduworks¹⁰ Project. The program contained two major parts: general RDM topics, and Data Management Plan (DMP) design that was presented as a hands-on exercise. The training materials were developed jointly by the EDISON Project UvA in cooperation with the University Library and are available under a CC BY licence. Further development is expected in the framework of the proposed RDA Working Group on RDM literacy.

20.6. REQUIRED RESOURCES

A successful Data Science education programme depends on the availability of 3 key components: (1) teaching staff, (2) computing and lab facilities, and (3) a pool of experts/advisers and related topics for course and thesis projects. All three components create challenges and require advanced planning. The following offerings are made available by the relevant departments:

⁹ Amsterdam Business School: <http://abs.uva.nl/>; accessed 5 February 2017.

¹⁰ Eduworks: <http://www.eduworks-network.eu/>; accessed 5 February 2017.

1. Teaching staff: Core teaching staff are provided by departments hosting the programme or track; associate teaching staff from industry provide specialised courses; local industry experts are invited to give selected lectures; leading domain researchers and experts are invited to give lectures, seminars and colloquia.
2. Computing and lab facilities: Computer classes are operated by departments and supported by ICT departments; high performance computing facilities are provided by the SURFsara Dutch research HPC facility; departments are actively using research and educational grants from the major cloud and Big Data providers such as Amazon Web Services, Microsoft Azure, IBM Watson and BlueMix to give students the opportunity to learn about leading industry platforms and applications.
3. A pool of experts and project development topics: departments maintain a network of external experts and collaborating research and technology organisations that provide advice on students' projects and host students' thesis projects.

A common problem and gap in developing consistent Data Science programmes is setting up a professional Data Management course that would cover both Research Data Management and industry data management and governance topics. The EDISON Project is cooperating with departments to develop core Data Management courses including Research Data Management courses and training for students and researchers.

20.7. COORDINATION OF RELEVANT ACTIVITIES INSIDE UVA

For coordination purposes and for the exchange of experience, UvA has created the Data Science Interest Group and a corresponding mailing list that has become an important forum for coordinating activities between departments, projects and collaborating organisations. This important role belongs to the Amsterdam Data Science Centre (ADS)¹¹ which is a joint initiative of 10+ companies and institutions in the Amsterdam area; the recently-established Amsterdam School of Data Science (ASDS) also has an important role to play.¹²

20.8. CONCLUSIONS

- The EDISON Data Science Framework (EDSF), the product of the EDISON Project, provides a strong background for building customisable Data Science programmes, including Research Data Management education and training programs;
- The Data Science Professional profiles (DSPP) play an important role since they define the whole spectrum of the data-related organisational roles currently present and required by research organisations and industry;
- The University of Amsterdam has created an effective cooperative and creative environment for coordinating efforts from multiple departments in establishing Data Science-related programmes. The EDISON Project provides necessary recommendations and materials for building consistent Data Science programmes;
- The University Library cooperates with the ICT department and research and teaching departments on RDM training and in setting up academic courses on Data Management;
- The EDISON Project maintains an extensive network of Data Science experts and Champion Universities that run or implement Data Science programmes in Europe;
- The current success of the EDSF makes it critically important to ensure the

¹¹ Amsterdam Data Science: <http://amsterdamdatascience.nl/>; accessed 5 February 2017.

¹² Amsterdam School of Data Science: <https://www.schoolofdatascience.amsterdam/>; accessed 5 February 2017.



Section 8

Tool Development

Legal requirements, RDM and Open Data

Author: Ignasi Labastida (Head of the Research Unit at the Learning and Research Resources Centre [CRAI] of the University of Barcelona)

Email: ilabastida@ub.edu

21.1. INTRODUCTION

Quite often when we talk about legal issues related to research data, we fall into discussions about privacy and personal data. This issue is fundamental when data are gathered from personal surveys or clinical trials, for instance. In these cases, researchers should follow the standard procedures established by their institutions through dedicated committees, for example an ethics or bioethics commission. In many of these cases, data cannot be shared openly. Only some aggregated data or anonymised data can be shared following a strict procedure¹.

In this chapter, I would like to focus on the legal tools we have to make data open once we have overcome all the possible barriers to providing data gathered or created during research activities. For the purpose of this case study, I will use the term open as defined by the open definition: “*Open data and content can be freely used, modified, and shared by anyone for any purpose*”.²

First, I will look at how copyright deals with data and afterwards I will review the different options we have to share data openly. It is important to know how researchers can share data because reusability is one of the FAIR principles that research data must fulfil.³ As stated in the principles, data and metadata must be released with a clear and accessible data usage licence.

21.2. DATA AND COPYRIGHT

To analyse what the different options to license data are, we must first review which rights are involved. Data is a complex term in relation to copyright because there are many formats of research output that can be considered as data depending on the discipline. For instance, data can be numbers, texts, or images. This variety of formats involves a different treatment when applied to copyright. It is clear that facts or dates cannot be copyrighted by anyone and therefore they fall outside any protection. In those cases there is no need to use a licence, and the best practice is to state that all this kind of data is under public domain.

However when data are texts or images, copyright has to be taken into account. Generally when there is a degree of originality exploitation rights appear, and there is a need to use a licence to authorise wide reuse: otherwise data should be considered with all rights reserved.⁴ Even in cases where images have a lack of originality, some legislation grants the performers some exploitation rights, shorter than the ones granted when images are considered works.⁵

¹ An example of challenges and compromises in anonymising data can be read in: Benjamin Saunders, Jenny Kitzinger, and Celia Kitzinger, “Anonymising interview data: challenges and compromise in practice”, *Qualitative Research* 2015 Oct; 15(5): 616–632, doi: 10.1177/1468794114550439 (last accessed 29/01/2017).

² Open Definition: <http://opendefinition.org/> (last accessed 29/01/2017).

³ FORCE11: <https://www.force11.org/group/fairgroup/fairprinciples> (last accessed 29/01/2017).

⁴ Current copyright laws do not require any procedure to get exploitation rights. Therefore in the absence of a copyright notice, the “all rights reserved” regime should be applied.

⁵ For instance in Spain, the agent of a “mere photograph” has such a right for 25 years. More on the situation of non-original photographs: Thomas Margoni, “The digitisation of cultural heritage: originality, derived works and (non) original photographs”, <http://www.ivir.nl/publicaties/download/1507.pdf> (last accessed 29/01/2017)

Moreover, data are not usually released individually, but rather as part of a compilation or a database. This way of presenting data could be protected by copyright in two different ways. Again, if the compilation or database has a degree of originality, it can be protected as any other creative work, as I have mentioned before in relation to data. The originality has to be found in the selection or arrangement of the data. This protection is granted even if the compiled data by themselves are not copyrightable. Furthermore, in the European Union and a few other countries, databases with a lack of originality in the selection or in the arrangement of data may have another layer of protection by means of the so called *sui generis* (i.e. of its own kind) right.

This right recognises the substantial investment in compiling a database and grants the creator a period of protection of fifteen years. During this time, nobody can extract the whole content or a substantial part of the database and reuse it without consent. Again this protection is granted to any database whereas its content could be protected, or not, by copyright. Therefore we must take into account these different layers of protection in order to share data openly. In the next section I will review some of the licences that we can use.

21.3. LICENCES FOR DATA AND DATABASES

Probably when we deal with licences for open content, the first set of legal texts that come to us is the one provided by Creative Commons (CC).⁶ However there are other options that fulfil the requirements to deal with all the possible layers of protection in a database.

21.3.1 Use of Creative Commons Licences for data and databases

With almost 15 years of experience, the suite of licences developed by Creative Commons (CC) provides a good solution to share any content that falls under the scope of copyright protection. Therefore, if we want to share data that could have some protection due to its originality or its format, we can consider using them, as we can if we want to share a database with originality in the selection or the arrangement of its elements.

Currently CC offers a standard set of six licences that provide for different degrees of reusability. Any of the six licences grants the right to reproduce, distribute and communicate in public the licensed material for non-commercial purposes. Depending on the licence, it is even possible to grant those exploitation rights for commercial purposes. Four of the six licences also grant the transformation right that permits the creation and dissemination of derived works. When the transformation right is granted, the licensor can require that the possible derived works be disseminated using the same licence as the original work or an equivalent one. This requirement is inspired by the copyleft⁷ clauses that originally carried free software licences.

It is important to note that CC also has a public domain mark⁸ that can be used to identify public domain works. This tool has been used in some governmental material and in cultural and heritage institutions.

Until the current version 4.0, CC licences approached the *sui generis* database right in different ways. Initially, and due to its US copyright inspiration, there was no mention of this right because it is not recognised in US copyright law. In version 3.0, some of the ported versions developed by European CC affiliates introduced

⁶ For a detailed explanation of the types of licences, go to: <https://creativecommons.org/share-your-work/licensing-types-examples/> (last accessed 29/01/2017).

⁷ An arrangement whereby software or artistic work may be used, modified, and distributed freely on condition that anything derived from it is bound by the same conditions.

⁸ For a detailed explanation of the Public Domain Mark, go to <https://creativecommons.org/share-your-work/public-domain/pdm/> (last accessed 29/01/2017).

the issue into their local texts and they mainly proposed to waive the *sui generis* right when licences were attached to databases.

In version 4.0, where in principle there will be no porting process other than translations, the *sui generis* database right has been included in a dedicated section of the legal code. The current version treats this right as any other exploitation right. This means that if a licence prohibits the work to be reused for a commercial purpose, it implies that the extraction and reuse of all or a substantial part of the elements in a database cannot be for commercial exploitation.

Therefore the requirements of the four elements of the CC licences have the following implications when applied to the *sui generis* database right:

- Attribution: Any extraction and reuse of all or a substantial part of elements from the licensed database requires a proper acknowledgment of its creator and any others designated to receive appropriate credit;
- Non Commercial: Any extraction and reuse of all or a substantial part of elements from the licensed database cannot be for a commercial purpose;
- Non Derived Works: It is not allowed to build a new database with all the elements, or a substantial part of them, extracted from the original licensed database;
- Share Alike: It is allowed to build a new database with all the elements, or a substantial part of them, extracted from the original licensed database, but this new database has to be licensed under the same licence or an equivalent one.

21.3.2. Licences created for data and databases

Before having the abovementioned *sui generis* right included in the six standard licences, CC created a legal tool aimed at scientific databases. This tool is called CC0 and it is both a waiver and a licence at the same time.⁹

Sometimes CC0 is seen as a pure public domain dedication and it raises some concerns in those countries where the copyright law does not allow the placing of a work into the public domain before the protection term expires or the waiving of all copyright rights, especially moral rights. In fact CC0 is not a full waiver of rights. CC0 works on two levels: first, the rights holder waives all rights over the work or content to the fullest permitted by law;¹⁰ second, all the unwaivable rights are then granted to the fullest permitted by law to the user, acting as a licence without any requirements. If there are still some rights that cannot be waived or licensed by the applicable law they remain with the corresponding rights holder.

Before the release of CC0, the Open Knowledge Foundation created the Open Data Commons project to provide legal solutions for open data. This initiative launched three licences addressed to share data and databases openly: the Open Database Licence, the Attribution Licence, and the Public Domain Dedication and Licence.¹¹ The first is a pure copyleft licence allowing for wide reuse, with the requirement to use the same licence when creating a derived licence. The second licence only requires a proper attribution in a

⁹ For a detailed explanation of the Public Domain Mark, go to Creative Commons: <https://creativecommons.org/share-your-work/public-domain/cc0/> (last accessed 29/01/2017).

¹⁰ As is explained in the CC0 dedicated FAQ, no legal instrument can ever eliminate all copyright interests in a work in every jurisdiction. Creative Commons: https://wiki.creativecommons.org/wiki/CC0_FAQ#Does_CC0_really_eliminate_all_copyright_and_related_rights.2C_everywhere.3F (last accessed 29/01/2017).

¹¹ For a detailed explanation of the three licences, go to Open Data Commons: <https://opendatacommons.org/licenses/> (last accessed 29/01/2017).

similar way to the Attribution Licence from Creative Commons. Finally, the Public Domain Dedication and Licence works in the same way as the CC0 tool.

Finally we can mention a couple of licences created to allow the reuse of public sector information: the Open Government Licence from the United Kingdom¹² and the French Open Licence/*Licence Ouverte*.¹³ Both these licences grant a full reuse of the information attached to them, acknowledging the corresponding sources.

21.4. CONCLUSION

Before starting to think about the most suitable licence to be applied for reusability, it is important to check that data can be legally released and that there are, for instance, no implications for privacy, security or confidentiality. It is important to use a licence that takes into account all the possible layers of protection applicable to data: authors' rights, neighbouring or related rights, and especially the *sui generis* database right. If we pursue wide reusability, we must avoid licences that restrict some uses, for instance commercial purposes or the creation of derived materials. Licences that only require an acknowledgement of the source and the creators of data and/or databases fulfil the goal of providing complete reusability.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

¹² Full text of the licence available at The National Archives: <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/> (last accessed 29/01/2017).

¹³ Full text of the licence available at Etalab: <https://www.etalab.gouv.fr/licence-ouverte-open-licence> (last accessed 29/01/2017).

Developing a Data Management Plan: a case study from Argentina.

Author: Fernando-Ariel López, Chief of Institutional Communications and Training Sectors (CAICYT – CONICET)

Email: flopez@conicet.gov.ar

22.1. INTRODUCTION

This case study will describe the experience of the *Centro Argentino de Información Científica y Tecnológica del Consejo Nacional de Investigaciones Científicas y Técnicas* (CAICYT-CONICET)¹ in the research, development and implementation of a Research Data Management Plan for the *Observatorio Nacional de la Degradación de Tierras y Desertificación* (ONDTyD)² and for CONICET.

22.2. A RESEARCH DATA MANAGEMENT PLAN BY CAICYT-CONICET

Several international organisations related to the field of science and technology (National Research Agencies, Funders, University Consortia, etc.) have started to require that research project funding applications be accompanied by a Research Data Management Plan (DMP) elaborated by the lead researcher and/or the group of researchers applying for funds.

The DMP allows for, on the one hand, the organisation of research data for researchers and, on the other, the ability to diagnose, characterise and predict, based on the information contained in the DMP, thus making it a valuable instrument for institutions managing Science and Technology. Furthermore, the DMP becomes a fundamental tool to assess and evaluate the potential impact (social, economic, cultural, etc.) implied in the development of research projects.

In Argentina there exists legislation and regulations that provide a framework and formalise the requirement for Data Management Plans (DMP):

- Data Management Plans are required by the law 26899 “[Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos](#)”, enacted in November 2013 and revised in November 2016;⁴
- Resolution CONICET 2705/15 and Institutional Repository Policies “[CONICET Digital](#)” require open access to publications and data funded by CONICET to researchers and institutes affiliated to CONICET;⁵
- The CONICET Data Policy [in development] will be aligned with the law 26899 and the Resolution CONICET 2705/15, requiring and regulating Data Management Plans and other aspects of data sharing.

¹ English translation: Argentinean Centre of Science and Technology Information of the National Council of Science and Technology Research, <http://www.caicyt-conicet.gov.ar/>, last accessed 02/02/2017.

² English translation: National Observatory of Soil Degradation and Desertification, <http://www.desertificacion.gob.ar/>, last accessed 02/02/2017.

³ English translation: “Creation of Institutional Open Access Repositories, Own or Shared”.

⁴ SNRD: <http://repositorios.mincyt.gob.ar/recursos.php>; last accessed 02/02/17.

⁵ CONICET: <http://ri.conicet.gov.ar/themes/Mirage/RD%2020150710-2705.pdf>, last accessed 02/02/2017.

22.3. WHAT IS A DATA MANAGEMENT PLAN (DMP)?

A research data management plan (DMP) is a document elaborated by a researcher or a group of researchers, where the following is defined:

- What data will be created and how;
- How data will be described, organised, stored and managed;
- Who will be responsible for each of these activities;
- How data will be shared, explaining any use restriction that could apply.

The data management plan (DMP) is a live document, which evolves until the end of the research and its subsequent publication. Usually, a DMP is required at the following points in time: (1) at the time of requesting funding, accompanying the research project proposal; (2) once the project has started; (3) half way through the project; (4) at the end of the research project.

22.4. PROBLEMS WITH RESEARCH DATA

The National Observatory of Soil Degradation and Desertification (ONDTyD) is a national system for the evaluation and monitoring of soil across different scales (national, regional and pilot sites), based on an integral, interdisciplinary and participatory approach. It is sustained by a network of science and technology, and political organisations that provide data and knowledge and, at the same time, are also users of that information. Interactive maps, publications and an online geospatial data repository are being developed for their visualisation. The goal of ONDTyD is to identify the causes of desertification, to anticipate environmental risks and to collaborate in the restoration of affected ecosystems.

In the methodology developed, ONDTyD uses indicators of biophysical and socioeconomic vectors. However, the researchers were not aware of the lifecycle of their data, data management practices, documentation of their use, re-use, licenses or long-term preservation. The result was multiple versions of data from various sources and a lack of standardisation.

ONDTyD invited CAICYT-CONICET to collaborate in the improvement of these areas of their ongoing research project, whose indicators have varying levels of progress in terms of data collection.

22.5. DEVELOPMENT OF THE RESEARCH DATA MANAGEMENT PLAN

The first task was to discover the level of awareness of the field of data management amongst the researchers and to identify the research practices, documentation generated and group workflows at ONDTyD. We established regular meetings with the group coordinators, with specific researchers, as well as other meetings of a more general nature with the whole group. These meetings allowed us to understand, determine and reach consensus among participants about research data lifecycles and workflows.

We continued with the identification, analysis and comparison of research data management plans required by the Digital Curation Centre (DCC, UK), Horizon 2020 (European Union), the National Science Foundation (NSF, USA) and the Australian Research Council (ARC, Australia), as specified in the Information Laboratory of CAICYT – CONICET’s working paper “Analysis of Data Management Plans”.

The following action was to develop a Research Data Management Plan for ONDTyD, incorporating a *data dictionary* which was also developed (the dictionary specifies what information is required and incorporates definitions and alternative answers to the questions of the DMP). Furthermore, a section on *Best Practices* was included, referring to: (a) Data formats, (b) Folders and files structure, (c) Version control, and (d) Metadata schemas.

The ONDTyD-DMP includes the sections: (a) Administrative data; (b) Data collection; (c) Documentation and metadata; (d) Storage and security copies; (e) Selection and preservation; and (f) Data re-use.

22.6. PLATFORM FOR DMP MANAGEMENT, TRAINING AND SUPPORT

The next phase was to develop and to implement a digital tool to enable the research group (located across different provinces and cities in Argentina) to load, edit, and store and publish remotely a Data Management Plan (ONDTyD-DMP).

We identified and compared different online platforms for the management of a DMP. For diverse reasons, the tool selected was *DMPonline*⁶ developed by the Digital Curation Centre (DCC, UK). Following acquisition, we then undertook the customisation and translation of the platform for use by the ONDTyD.

To ensure the implementation and correct use by all members of the Observatory, the next step was to deal with training and support:

- Development of a workshop entitled “Scientific Data: quality, normalisation and visualisation”
- Development of a virtual course about the ONDTyD-DMP, which incorporated information on the required sections and best practices (mentioned above).
- Establishment of a support helpline, to answer questions emerging in the process of filling out the ONDTyD-DMP

22.7. IMPACT

After meeting and exchanging information with ONDTyD, the combined workgroup deemed it necessary to reconsider some methodological decisions, resulting in the enhancement of data, their documentation and the management of research data created and to be generated in the future. In this way the group of researchers of ONDTyD improved their understanding and skills in the management of research data.

The Fundación Williams⁷ - the project funder - make clear its interest in incorporating the DMP as an integral element in the process of receiving future research funding applications.

Based on the previous experience and the work carried out with ONDTyD, at the request of the Gerencia de Desarrollo Científico⁸ of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina), we:

- Developed a Research Data Management Plan for CONICET. The DMP is of a generic nature and has 3 levels of information detail for the presentation of a project:

⁶ For further information about DMPonline, see DCC, UK: https://dmponline.dcc.ac.uk/about_us, last accessed 02/02/2017.

⁷ English translation: Williams Foundation: <http://www.fundacionwilliams.org.ar/>, last accessed 02/02/2017.

⁸ English translation: Scientific Development Division of the National Council of Science and Technology Research.

1. Global: Necessary general aspects that provide information about who is responsible for data, their basic characteristics and related legal aspects;
 2. Management: Consideration of concrete aspects of management and decision-making for data documentation and re-use;
 3. Data Set: Reference to specific aspects of scientific data generated in research projects funded by CONICET.
- Launched a CONICET DMP Pilot Survey, as part of the call for Strategic Projects of CONICET, with the following objectives: (a) to know the treatment of data generated by researchers, and (b) to draw attention to the interests and needs of researchers, research agencies and funders.
 - Accepted an invitation to participate in the Consultant group on Scientific Data Management of CONICET, for the establishment of: (a) a Data Policy for CONICET, and (b) a Roadmap for the Management of Scientific Data at CONICET.

22.8. CONCLUSIONS

It is fundamental to acquire an appreciation of the discipline and to know research practices and workflows of specialised research groups in the thematic area. It is also important to allow for constant feedback from research groups and/or researchers in each thematic area to reach consensus in regard to data lifecycle, data management plans, metadata, etc.

The DMP enables researchers to plan the creation and collection, as well as the organisation, of data. A good DMP will multiply the possibilities for data use, re-use, and the impact of research in the scientific community and in society at large.

The requirement of a DMP by institutions that manage and fund research in Science and Technology constitutes an important input for diagnosis and prediction, necessary for the development of infrastructure and for the evaluation and measurement of potential and/or real impact (social, economic cultural, etc.) that a piece of research and its funding imply.

ONDTyD's digital platform to manage their DMPs was developed and implemented. The platform should be flexible, modular and interoperable with repositories of data, publications, etc.

Training and support of the researchers at ONDTyD have proved vital elements to success with the implementation and development of DMPs, the implementation of which will facilitate future use and re-use of data.

The Research Data Management Plan of the Observatorio Nacional de la Degradación de Tierras y Desertificación (PGD – ONDTyD) and the Digital Platform for DMP Management were developed by the Centro Argentino de Información Científica y Tecnológica (CAICYT) of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) with the support and funding of Fundación Williams.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Surveying your level of preparation for Research Data Management

Authors: Ignasi Labastida (Head of the Research Unit at the Learning and Research Resources Centre [CRAI] of the University of Barcelona) & Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries))

Email: ilabastida@ub.edu / p.ayris@ucl.ac.uk



SURVEY: Is your institution ready for managing research data?

The LEARN project has compiled the following survey as a self-assessment tool to assist institutions discover how ready they are for managing research data. The survey is based on the issues posed to institutions by the LERU Roadmap for Research Data published at the end of 2013, and available at: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf.

The survey has thirteen questions addressing the main elements to be taken into account in developing an institutional strategy for research data management. Each question has three possible answers represented by green, yellow or red answers. The more 'green light' responses recorded, the readier an institution probably is for managing its research data. We encourage you to complete the questionnaire online which is available at: <http://learn-rdm.eu/en/rdm-readiness-survey/>, with a link straight through to the questionnaire at <http://goo.gl/forms/m6PGJ34tGr>. The survey is available in both English and Spanish.

The Survey is iterative, in that it can (once taken) be re-taken at regular intervals. Changes in the scores will themselves illustrate the level of progress made in the intervening period.

1. POLICY

- My institution has a policy on research data
- My institution is working on a policy on research data
- My institution has no policy regarding research data

2. LEADERSHIP

- My institution has a steering committee on research data
- My institution is working on setting up a working group to develop services and policies on research data
- There is no dedicated group on research data at my institution

3. ROLES

- My institution has established new roles to steward the management of research data
- Some staff are shifting part of their work to involve the management of research data
- There is no one dedicated to research data

4. INFORMATION (SERVICES)

- My institution has an information point/helpdesk/webpages on research data management
- There is someone at/in the university library/research office who can give advice on research data management to researchers
- No service at my institution provides clear information on research data management

5. DISSEMINATION (AWARENESS)

- My institution has created some materials on the management of research data
- There are some links with information on research data on the library/research office website
- Researchers need to look outside my institution for information on the management of research data

6. INFRASTRUCTURE

- My institution provides an infrastructure to manage research data through the complete research cycle
- My institution provides some services for managing data but not through the complete research cycle
- Researchers need to use external facilities to manage their data

7. COST MODEL

- My institution has established a list of free and paid for services based upon an analysis of costs
- My institution offers some services for free and some need to be paid but there is not a public list of paid services
- There has not been any analysis regarding the cost of managing research data at my institution

8. LEGAL

- There is a protocol to define who is the owner of research data produced
- My institution has a policy on intellectual property rights (IPR) but there is no mention of research data
- My institution does not have a policy on IPR

9. SELECTION OF DATA

- There are protocols, laid down by bodies such as the university or the research funder, to define which data has to be kept, shared, archived, etc
- My institution gives some advice about the preservation of research data
- My institution has not established any guidance about which research data should be kept

10. PUBLICATION AND SHARING

- There are protocols, laid down by bodies such as the university or the research funder, defining which data has to be published, where and under which terms of use
- My institution allows researchers to publish research data in our institutional repository or in a disciplinary repository (outside the institution)
- My institution does not have a protocol or a place to publish research data

11. TRAINING

- My institution has scheduled regular training sessions on research data management addressed to researchers, students and staff
- My institution offers training sessions in research data management upon demand
- There are no training sessions about how to manage research data

12. REVISION AND UPDATES

- My institution has established a roadmap to review and, if needed, update its policy and services on research data
- My institution is developing services for managing research data, but there is no scheduled calendar for reviews
- My institution has yet to start the conversation to create a working group on research data

13. OPEN DATA

- My institution publishes research data openly by default and it has established a set of exceptions to waive this policy
- My institution allows researchers to share data openly but there is no formal policy established
- My institution does not publish any data openly



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Conclusions

Conclusions

The 23 Case Studies in the LEARN Toolkit span, but expand on, all seven themes of the original LERU Roadmap.¹ Overall, they underline the challenges and opportunities identified in the 2013 document, but now offer solutions to address a range of issues. They are grouped as follows:

- Policy and Leadership
- Advocacy
- Subject approaches
- Open Data
- Research Data Infrastructure
- Costs
- Roles, Responsibilities and Skills
- Tool development

POLICY AND LEADERSHIP

The LERU Roadmap advocated that ‘Every LERU member should develop and promulgate an institutional data policy’.² The LEARN Toolkit provides the tools to do this, with a model RDM policy and guidance in Part 2 of the compilation. Additionally, the Case Studies support the call for policy leadership and alignment. Case Study 1 from the Wellcome Trust argues that there is broad agreement on policy amongst research funders on the importance of RDM, whilst identifying key challenges which remain. The Executive Briefings in six languages in Part 3 are designed for senior decision makers, to support them in delivering sound solutions. Case Study 2 describes the process of developing a model RDM policy for Austria, based on the LEARN template, which acts as a framework and which can be customised at a local level. Case Study 3 looks at Brexit and its potential impact on Open Science, concluding that perhaps the greatest threat currently lies in a possible lack of engagement in the UK with the European Open Science Cloud. Case Study 4 looks at linking the practice of RDM with research integrity frameworks.

ADVOCACY

Many of the Case Studies are devoted to the theme of advocacy. The LERU Roadmap stressed that LERU members, researchers and research funders should ‘Promote best practice in data management, citation and interoperability to increase the visibility of data’.³ This is true of the Case Studies from both Latin America/the Caribbean and Europe. Some interesting themes around advocacy are identified. Case Study 5 makes the point that RDM advocacy to researchers is in its infancy. Accordingly, qualitative rather than quantitative measures and approaches currently predominate. However, the institution in this Case Study has undertaken a wide-ranging internal survey which will provide a baseline for future activity. Case Study 6 emphasises that what is needed is to identify RDM stakeholders, ensure good communication, and develop implementation plans. Case Study 7 links leadership and advocacy by asking the question ‘Who has leadership for RDM at an institutional level in the University of the West Indies?’. Case Study 8 underlines the challenges involved in RDM advocacy. In this institution, after years of activity, the difficulties in changing institutional culture with regard to RDM remain.

¹ LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf; last accessed 9 February 2017.

² Ibid., paragraph 22.

³ Ibid., paragraph 43.

SUBJECT APPROACHES

The Case Studies in the LEARN Toolkit look particularly at RDM issues in the Arts, Humanities and Social Sciences. Case Study 9 looks in detail at the challenges and opportunities at UCL (University College London). It identifies that many researchers in these disciplines do not use, or are unaware of, UCL-supported RDM solutions and that there is a need for advocacy to these communities. Case Study 10 is from the Performing Arts. Discussion on RDM has centred on the sciences (in the English sense, excluding the Arts, Humanities and Social Sciences). Because of how Arts projects are funded and structured, there are special problems and challenges relating to RDM – which this Case Study identifies.

OPEN DATA

The issue of Open Data is treated in several Case Studies. Case Study 11 argues that research is an inherently Open enterprise and that universities increasingly see Open Data as part of their future. Case Study 12 looks at Open Data in educational resources, showing that there is a place for research data in taught provision, where the university supports research-based or research-led education. Case Study 13, on the other hand, shows how closed data is managed in a practical way in the University of the Andes in Colombia.

RESEARCH DATA INFRASTRUCTURE

A number of the Case Studies look at the provision of infrastructure. Case Study 14 from UCL (University College London) draws lessons from its Research Data Storage service. Case Study 15 analyses a collaborative RDM service in Brazil, using the Dataverse network. Case Study 16 looks at the recent Report from the European Commission's High Level Expert Group on the European Open Science Cloud and discusses the vision for that development, based on research data which is FAIR (Findable, Accessible, Interoperable and Re-usable).

COSTS

LEARN itself is not an economic study, but several of the Case Studies look at RDM costs. Case Study 14, for example, gives some detailed costings for the Research Data Storage service at UCL (University College London). Case Study 16 looks at the projected costs for delivering the European Open Science Cloud. Case Study 17 from the University of Edinburgh gives detailed costings and analysis for the formation of its research data services.

ROLES, RESPONSIBILITIES AND SKILLS

A number of the Case Studies emphasise the need for training and skills development for all stakeholders in the RDM landscape. Case Study 18 looks at training for early career researchers by analysing the 2016 LERU Doctoral Summer School. Case Study 19 looks at training subject liaison librarians in research data management. Chapter 20 looks at the EDISON project, which is creating a data science profession for Europe.

TOOL DEVELOPMENT

A number of chapters look at tool development to support RDM. Case Study 21 looks at legal requirements and shows how the use of licences can establish frameworks for sharing, re-use and compliance. Case Study 22 looks to Argentina and the development of Data Management planning, concluding that good Data Management Plans will deliver good research. Finally, chapter 23 looks at the LEARN Readiness survey. The survey allows research performing institutions to assess their level of preparation for RDM by answering 13 questions. Using a traffic light marking scheme of red, amber or green, the survey will be marked and enable those taking the test to see how prepared they are. The test can be taken iteratively, so that over a period an institution can measure its progress in RDM activity.

CONCLUSION

Research data is the new currency of the digital age. From sonnets to statistics, and genes to geodata, the amount of material being created and stored is growing exponentially. The LERU Roadmap identifies a serious gap in the level of preparation amongst research performing organisations. This gulf is prominent in areas such as policy development, awareness of current issues, skills development, training, costs, community building, governance, disciplinary/legal/terminological and geographical differences. The LEARN Toolkit is designed to identify sound solutions and proposals for these challenges and opportunities. By adopting recommended LEARN practices, templates and guidance, all those involved as stakeholders in RDM can introduce best practice into their institutions.

Part 2

The Model RDM Policy



Model Policy for Research Data Management (RDM) at Research Institutions/Institutes

1. PREAMBLE

The *[name of research institution]* recognizes the fundamental importance of research data¹ and the management of related administrative records in maintaining quality research and scientific integrity, and is committed to pursuing the highest standards. The *[name of research institution]* acknowledges that correct and easily retrievable research data are the foundation of and integral to every research project. They are necessary for the verification and defence of research processes and results. RDM policies are highly valuable to current and future researchers. Research data have a long-term value for research and academia, with the potential for widespread use in society.

2. JURISDICTION

This policy for the management of research data applies to all researchers active at the *[name of research institution]*. The policy was approved by the *[dean/commission/authority]* on *[date]*. In cases when research is funded by a third party, any agreements made with that party concerning intellectual property rights, access rights and the storage of research data take precedence over this policy.

3. INTELLECTUAL PROPERTY RIGHTS

Intellectual property rights (IPR) are defined in the work contract between a researcher and his or her employer. IPRs might also be defined through further agreements (e.g. grant or consortial agreements). In cases where the IPR belong to the institution that employs the researcher, the institution has the right to choose how to publish and share the data.

4. HANDLING RESEARCH DATA

Research data should be stored and made available for use in a suitable repository or archiving system, such as *[name of institutional repository/archiving system, if applicable]*. Data should be provided with persistent identifiers.

It is important to preserve the integrity of research data. Research data must be stored in a correct, complete, unadulterated and reliable manner. Furthermore, they must be identifiable, accessible, traceable, interoperable, and whenever possible, available for subsequent use.

¹ See definitions of "research", "researchers" and "research data" in the Annex.

In compliance with intellectual property rights, and if no third-party rights, legal requirements or property laws prohibit it, research data should be assigned a licence for open use.²

Adherence to citation norms and requirements regarding publication and future research should be assured, sources of subsequently-used data explicitly traceable, and original sources can be acknowledged.

Research data and records are to be stored and made available according to intellectual property laws or the requirements of third-party funders, within the parameters of applicable legal or contractual requirements, e.g. EU restrictions on where identifiable personal data may be stored. Research data of future historical interest and the administrative records accompanying research projects should also be archived.

The minimum archive duration for research data and records is 10 years after either the assignment of a persistent identifier or publication of a related work following project completion, whichever is later.

In the event that research data and records are to be deleted or destroyed, either after expiration of the required archive duration or for legal or ethical reasons, such action will be carried out only after considering all legal and ethical perspectives. The interests and contractual stipulations of third-party funders and other stakeholders, employees and partner participants in particular, as well as the aspects of confidentiality and security, must be taken into consideration when decisions about retention and destruction are made. Any action taken must be documented and be accessible for possible future audit.

5. RESPONSIBILITIES, RIGHTS, DUTIES

The responsibility for research data management during and after a research project lies with [*name of research institution*] and its researchers and should be compliant with codes for the responsible conduct of research.

5.1 RESEARCHERS ARE RESPONSIBLE FOR:

- a. Management of research data and data sets in adherence with principles and requirements expressed in this policy;
- b. Collection, documentation, archiving, access to and storage or proper destruction of research data and research-related records. This also includes the definition of protocols and responsibilities within a joint research project. Such information should be included in a Data Management Plan (DMP), or in protocols that explicitly define the collection, administration, integrity, confidentiality, storage, use and publication of data that will be employed. Researchers will produce a DMP for every research project.³
- c. Compliance with the general requirements of the funders and the research institution; special requirements in specific projects should be described in the DMP;
- d. Planning to enable, wherever possible, the continued use of data even after project completion. This includes defining post-project usage rights, with the assignation of

² Concrete recommendations for licensing should be listed and be available to the researchers.

³ A Data Management Plan (DMP) is a structured guideline (document or online tool) which depicts the entire lifeline of data and can be updated if needed. Data management plans must assure that research data are traceable, available, authentic, citable, properly stored and that they adhere to clearly defined legal parameters and appropriate safety measures governing subsequent use. Ideally, DMPs should be delivered in a machine actionable format.

- appropriate licences, as well as the clarification of data storage and archiving in the case of discontinued involvement at the *[name of university/research institution]*;
- e. Backup and compliance with all organisational, regulatory, institutional and other contractual and legal requirements, both with regard to research data, as well as the administration of research records (for example contextual or provenance information);
 - f. To ensure appropriate institutional support, it is required that new research projects are registered at the proposal stage at *[name of research institution/central body]*.

5.2 THE *[NAME OF RESEARCH INSTITUTION]* IS RESPONSIBLE FOR:

- a. Empowerment of organisational units, providing appropriate means and resources for research support operations, the upkeep of services, organizational units, infrastructures, and employee education;
- b. Support of established scientific practices from the beginning. This is possible through the drafting and provision of DMPs, monitoring, training, education and support, while in compliance with regulations, third-party contracts for research grants, university/institutional statutes, codes of conduct, and other relevant guidelines;
- c. Developing and providing mechanisms and services for the storage, safekeeping, registration and deposition of research data in support of current and future access to research data during and after the completion of research projects;
- d. Providing access to services and infrastructures for the storage, safekeeping and archiving of research data and records, enabling researchers to exercise their responsibilities (as outlined above) and to comply with obligations to third-party funders or other legal entities.

6. VALIDITY

This policy will be reviewed and updated as required by the head of/the director of the *[name the research institution]* every *[two years]*.

Annex: Definitions of Research, Researchers and of Research Data

1. Research is any creative and systematically performed work with the goal of furthering knowledge, including discoveries regarding people, culture and society, in addition to the use of such knowledge for new applications.

2. Researchers refers to all research-active members of an institution including employees and doctoral candidates. Persons not directly affiliated with an institution, but who, for purposes of research, make use of or are physically present at the institution, are also included in the term. Visiting researchers or collaborators may also be expected to comply with the policy.

3. Research data refers to all information (independent of form or presentation) needed to support or validate the development, results, observations or findings of a research project, including contextual information. Research data include all materials which are created in the course of academic work, including digitisation, records, source research, experiments, measurements, surveys and interviews. This includes software and code. Research data can take on several forms: during the lifespan of a research project, data

can exist as gradations of raw data, processed data (including negative and inconclusive results), shared data, published data and Open Access published data, and with varying levels of access, including open data, restricted data and closed data.

Three further approaches, each dealing with different aspects of research data, may help to find the proper definition for individual research institutions:

a. According to the **LERU Roadmap for Research Data**⁴ (LERU Research Data Working Group, Advice Paper No. 14 – December 2014):

“Research data, from the point of view of the institution with a responsibility for managing the data, includes: All data which is created by researchers in the course of their work, and for which the institution has a curatorial responsibility for at least as long as the code and relevant archives/record keeping acts require, and third-party data which have originated within the institution or come from elsewhere.”

b. The Australian **Griffith University**⁵ presents the following definition⁶:

“Research data are factual records, which may take the form of numbers, symbols, text, images or sounds, which are used as primary sources for research, which are commonly accepted in the research community as necessary to validate research findings.”

b. The **University of Minnesota**⁷ definition of research data⁸:

“Research data are data in any format or medium that relate to or support research, scholarship, or artistic activity. They can be classified as:

- *Raw or primary data: information recorded as notes, images, video footage, paper surveys, computer files, etc.*
- *Processed data: analyses, descriptions, and conclusions prepared as reports or papers*
- *Published data: information distributed to people beyond those involved in data acquisition and administration.”*



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁴ LERU: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf; last accessed 14/2/17.

⁵ Griffith University: https://intranet.secure.griffith.edu.au/__data/assets/pdf_file/0003/716106/ARI_DataManagement_Pt1_Apr2015.pdf; last accessed 14/2/17.

⁶ See also: Ingrid Dillo – Data Archiving and Networked Services (DANS), *Certification as a means of providing trust*, Florence, Fondazione Rinascimento Digitale, 2012 and Data Management at UTSA.

⁷ University of Minnesota: <https://www.lib.umn.edu/datamanagement/whatdata> ; last accessed 14/2/17

⁸ See also: Ingrid Dillo – Data Archiving and Networked Services (DANS), *Certification as a means of providing trust*, Florence, Fondazione Rinascimento Digitale, 2012.



Guidance for Developing a Research Data Management (RDM) Policy

This document provides the essential elements of a Research Data Management (RDM) Policy and is part of the LEARN Toolkit containing the **Model Policy for Research Data Management (RDM) at Research Institutions/Institutes**.

The elements below may be used to define research data, explain RDM, illustrate workflows, point out benefits and give information about funding agency requirements. Please note that in order to facilitate the measurability of the policies and their impact, they should be created in a machine actionable format. Furthermore, indicators may be used for automated validation processes.

Elements	Description
Header info	Document title Institutional logo
Title of policy	Description of the pursued issue
Subtitle	If necessary: extension of the title
General remarks before getting started	<ul style="list-style-type: none"> > Research data is one part of the knowledge capital of research institutions. In data-driven science, good data management promotes discovery, efficiency, and increases reliability by ensuring consistent quality with a high level of comparability. The policy may be strongly connected to strategic alignments and strategic management. It could help in building the bridge from technical requirements to skills and competencies. > Research data management is considered as a whole in the policy (including research records, methods, software, code etc.). > These principles will determine the organisation's behaviour. > These principles also apply to the behaviour of individuals within the institution. > The policy (with annexed documents) should contain definitions, indicating answers to these questions: <ul style="list-style-type: none"> • What is "research data"? • What is "research"? • Who is a "researcher"? > The following should be clear: <ul style="list-style-type: none"> • Authorship of the policy. It should be clear who defines the policy ("the speaking entity") and why this entity (author of the policy) defines the policy. What is the role of "the speaking entity" (authorship)? • Aim of the policy. Why does a research institution/institute have a policy? What is the goal of the policy? What does the institution want to achieve? • Subject. According to the statutes of the institution and its published guidelines: What is the subject of the policy?

<p>Preamble</p> <p>Refers to Point 1 of the Model Policy</p>	<p>The preamble describes the context:</p> <ul style="list-style-type: none"> > It is an introductory statement or a description of an initial situation. > It defines why there should be a policy and how to contextualize it within the institution. This part has to be localised by each institution and aligned with the prevailing philosophy and mission of the institution. > Scientific disciplines and organizations produce and manage different types of materials which might have different guiding principles. It is essential that consistency is brought to the field in the form of research institution/institute-level policies. > The fundamental truths or propositions that serve as the foundation for the chain of reasoning of the policy should be described.
<p>Jurisdiction</p> <p>Refers to Point 2 of the Model Policy</p>	<ul style="list-style-type: none"> > The scope of the policy must be defined according to space and time. > The relationship between the policy and research institution/institute and non-research institution/institute guidelines and statutes must be clarified in the policy. > Compliance with legal and contractual provisions must be maintained.
<p>Intellectual Property Rights</p> <p>Refers to Point 3 of the Model Policy</p>	<p>According to the FAIR principles, the fundamental purpose of rights definition is to encourage re-use and collaboration.</p> <ul style="list-style-type: none"> > In this section, rights must be defined according to the questions: <ul style="list-style-type: none"> • Who owns research data? • And who holds rights in such data? <p>This is a fundamental question. With regard to research data protected by law, this question can be answered by legal advisers.</p> <ul style="list-style-type: none"> > The following aspects must be considered: <ul style="list-style-type: none"> • terms of use • questions of licensing and subsequent use of data • data protection aspects, including relevant legal requirements • privacy rights, usage rights, exploitation rights and copyrights > In cases where no law fittingly applies to a specific piece of research data, the policy will apply to intellectual property rights, etc. > The policy must take into account all contracts made with funders, as well as contracts between researchers and their institutions, which have precedence. <p>You might include the following sentence:</p> <p>The research institution will make research data available under an open licence, unless legal obligations, third party rights, intellectual property rights and privacy rights preclude this. The licence is selected according to the type of data and in order to label the data and facilitate its utilization. An example for a Source Code Licence would be the General Public Licence (GPL). For all other kinds of data, CCO or CCBY licences can be used. Data which are not subject to any copyright restrictions should be clearly marked as such with for instance the Creative Common Public Domain Mark. In some cases copyright belongs to the institution that employs the researcher, so there may be a question regarding who has the right to choose a licence.</p>

<p>Handling research data</p> <p>Refers to Point 4 of the Model Policy</p>	<ul style="list-style-type: none"> > This section refers to all processes for dealing with one's own and other people's data throughout and after the scientific discovery process. > The policy refers to any research data generated within the institution, for instance in education, cultural heritage and institutional management. > It is important to define how research data are to be changed, documented, used, secured, archived, publicized and the conditions under which data may subsequently be used. Thus, this section reflects the FAIR data principles, meaning that data are Findable, Accessible, Interoperable and Re-usable. > It should be clear which exceptions exist in the policy and to what extent they apply. This may also concern the "right to be forgotten" (deletion of data). > Concerning deletion (deleting): This defines which data can or must be deleted and who decides to carry this out. > Concerning retention of data: The minimum recommended period for retention of research data is 10 years. However, in some particular cases it should be considered that: <ul style="list-style-type: none"> • for short-term research projects that are for assessment purposes only, such as research projects completed by students, retaining research data for 12 months after the completion of the project may be sufficient • for some research projects retaining research data for 15 years or more may be necessary (e.g. clinical trials) • for other areas (e.g. gene therapy, seismological data), research data must be retained permanently • if the work has community or heritage value, research data should be kept permanently, preferably within a national collection > The policy should contain a statement showing which policy takes precedence when research is funded by external funders, and showing the expectations placed by the institution on external research partners. > Concerning storage and access: The policy should address where data will be stored and how it will be accessed. If possible, there should be a recommendation for the use of institutional research infrastructures. > If needed or foreseen, regulations for <ul style="list-style-type: none"> • open data • restricted data • and/or closed data should be specified
---	--

<p>Responsibilities, Rights, Duties</p> <p>Refers to Point 5 of the Model Policy</p>	<ul style="list-style-type: none"> > This section defines the coverage of the policy: <ul style="list-style-type: none"> • institutional • faculty-wide (or other organizational units) • discipline-wide • group(s) of people covered: such as research staff, research support staff, IT services, students > The scope and coverage of the policy should be checked: <ul style="list-style-type: none"> • Does the policy include all research data? • Does the policy include/exclude a selection of the non-digital results of research processes? > Regulations concerning the responsibilities, rights and duties of the following persons and institutions should be formulated with regard to research data: <ul style="list-style-type: none"> • researchers and research data producers (e.g. PhD students) • funders and funders' regulations (the policy should acknowledge that funders have rights and regulations, and show that these will be given precedence where appropriate) • institutions • research supporting entities (for example, libraries, IT services, research support centres, etc.) > If necessary, there should be a recommendation for institutional research infrastructure. > Questions around the costs of RDM (including stewardship of data) as stated in a data management plan (DMP), as well as who bears those costs, should be well defined. This could also include costs that occur after a project has ended. > It is important to define roles, responsibilities and competencies in order to assign objectives and define time frames. Relevant questions: <ul style="list-style-type: none"> • Who is in charge of ensuring legal compliance? • Who will provide legal advice? • Who is in charge of the quality of the content? • Who is in charge of defining acceptable formats? • Who is in charge of maintaining the currency of formats over time? • Who will provide technical support? • Who will promote services? • Who will provide training?
<p>Approval of the policy, periodic review, validity and timeline</p> <p>Refers to Point 6 of the Model Policy</p>	<ul style="list-style-type: none"> > This pertains to the date of release of the policy and how long the current policy will be valid. This can be done on a regular basis, which may be externally defined, or based upon needs. The key dates must be included. > The policy should be subjected to periodic review. The changes in each revision must be listed. > The relevant questions here are: <ul style="list-style-type: none"> • How long are the terms of the policy valid? • Who/which body is responsible for reviewing and updating the policy? • What should be done after the end of the defined timeline or period?
<p>Footer info</p>	<ul style="list-style-type: none"> • Page number • Version number • Status • etc.
<p>Annexes</p> <p>Refers to Annex of Model Policy</p>	<ul style="list-style-type: none"> • Definition of key terms • Excerpts from / links to relevant funder policies or expectations • List of related institutional policies (with links)

See also the LEARN Project Glossary: <http://learn-rdm.eu/en/dissemination/glossary/>; last accessed 12/2/17



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.



Evaluation Grid of RDM Policies in Europe

Between July 2015 and June 2016, the Library of the University of Vienna (as the leader of Work Package 3 – Policy Development and Alignment of the LEARN Project) collected and analysed over 40 European RDM policies. In the course of this preparation phase it became obvious that in many countries (especially in continental Europe) there have been published hardly any guiding principles regarding RDM. After a further selection process, 20 policies were examined more closely based on (identified) format and content-related criteria. Using the following analysis grid, 11 RDM policies from the United Kingdom, four from Germany, one from the Netherlands and four from Finland (see list at the end of this document) were evaluated and checked for possible significant changes during this period at regular intervals. This compact overview is also supplemented by a detailed evaluation of the selected policies with extensive comments (see below).

Criteria	Status - Overview		
	It was NOT taken into consideration	It was PARTLY taken into account	It has been CONSIDERED
Number of institutions: 20			
Authorship		### ##	
Validity		### ##	##
Review	### ##		##
Subject			### ## ##
Scope and coverage		##	## ##
Preliminaries and definitions		##	## ##
Institutional awareness, support and services		##	## ##
Objectives (“what and how”)		### ##	##
Roles and responsibilities		##	## ##
DMP			### ## ##
Costs	##	## ##	
External		##	##
Ownership	##	## ##	##
Retention	## ##		##
Deletion	### ## ##		
Legal aspects		##	## ##
Ethics		##	##
Open data / restricted data / closed data		## ##	##
Storage and access		##	## ##
Metadata curation	##		##
Exceptions	##	##	##
Research infrastructure	## ##	##	##
Long tail of data / head of project data	## ## ##		
Educational data	## ## ## ##		
Cultural heritage	## ## ## ##		

Criteria	Status - Detail			
Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
1. Authorship. It should be clear who defines the policy	University of Bath Universität Bielefeld Universität Heidelberg	STFC *University of Birmingham University of Bristol University of Cambridge University of Edinburgh University of Glasgow University of Leeds University of the Arts London Aalto University University of Helsinki Radboud University Humboldt-Universität zu Berlin Universität Göttingen	UCL University of Oxford Tampere University of Technology University of Turku	STFC: Drawn up by an internal technical working group (information only on the website, not in the document) UCL: Author mentioned by name University of Bristol: Approved by Senate Well-defined authorship only in the previous draft version, not in the updated document University of Cambridge: Approved by Research Policy Committee University of Edinburgh: Approved by University Court RDM Roadmap: authors mentioned by name University of Glasgow: Information only on the website, not in the document Approved by Research Strategy and Planning Committee Tampere University of Technology: Detailed description of working process Working group chaired by Vice President for Research University of Turku: Decision of the Rector Radboud University: Poor information on the authorship (Executive Board) Humboldt-Universität zu Berlin: Approved by Academic Senat

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>2. Validity. The date of the release of the policy should be clear. It should also be clear how long the terms of the policy are valid</p>	<p>STFC University of the Arts London</p>	<p>University of Birmingham University of Bristol University of Cambridge University of Glasgow University of Leeds University of Oxford Tampere University of Technology University of Helsinki Humboldt-Universität zu Berlin Universität Bielefeld Universität Göttingen Universität Heidelberg</p>	<p>UCL University of Bath University of Edinburgh Aalto University University of Turku Radboud University</p>	<p>UCL: Approving policy Ensuring resources Implementation</p> <p>University of Bath: Date of last modification is indicated</p> <p>University of Cambridge: Date of last modification is indicated “The University acknowledges that a full implementation of this policy framework will be a long-term process.”</p> <p>University of Edinburgh: Aspirational policy: implementation will take some years RDM Roadmap: Timeframe August 2012 – July 2016</p> <p>University of Glasgow: Information about release only available on the website, not in the document</p> <p>University of Leeds: Institutional RDM Policy Evolution on the website</p> <p>University of Turku: Realisation is followed with indicators Policy and implementation are developed</p> <p>Universität Bielefeld: Information about release only available on accompanied webpage “Resolution on RDM”</p> <p>Universität Göttingen: Information about release only available on the website, not in the document</p> <p>Universität Heidelberg: Date of last modification is indicated</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>3.Review. The policy should be a subject to periodic review</p>	<p>STFC</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Edinburgh</p> <p>University of the Arts London</p> <p>Aalto University</p> <p>Tampere University of Technology</p> <p>University of Turku</p> <p>Radboud University</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p>	<p>University of Glasgow</p> <p>University of Helsinki</p> <p>Universität Heidelberg</p>	<p>UCL</p> <p>University of Bath</p> <p>University of Cambridge</p> <p>University of Leeds</p> <p>University of Oxford</p>	<p>UCL: Reviewed at least every 3 years by RIIISG and UCL Research Data and Network Services Executive</p> <p>University of Bristol: Revision History only in the previous draft version, not in the updated document</p> <p>University of Cambridge: Regularly reviewed by the Open Access Project Board</p> <p>University of Glasgow: Information on website: “This policy replaces the previous Draft”</p> <p>University of Leeds: Research and Innovation Board is responsible for reviewing and updating the policy</p> <p>University of Oxford: Research and Information Sub-Committee is responsible for updating of the policy</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>4.Subject. It should be clear what the subject of the policy is</p>		<p>University of Birmingham</p> <p>University of Edinburgh</p> <p>Universität Bielefeld</p> <p>Universität Heidelberg</p>	<p>STFC</p> <p>UCL</p> <p>University of Bath</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Aalto University</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Radboud University</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Göttingen</p>	

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>5. Scope and coverage. The scope and the coverage of the policy should be defined</p>		<p>University of Birmingham</p> <p>University of Cambridge</p> <p>University of Edinburgh</p> <p>University of Leeds</p> <p>Aalto University</p> <p>Radboud University</p> <p>Universität Heidelberg</p>	<p>STFC</p> <p>UCL</p> <p>University of Bath</p> <p>University of Bristol</p> <p>University of Glasgow</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p>	<p>University of Bristol: Applies to all research conducted by University staff and postgraduate research students (PGRs) regardless of whether or not the research is externally funded but not to taught postgraduate students or undergraduates</p> <p>University of Glasgow: “For all staff, including technical and other support staff and persons with honorary positions and students carrying out or supporting research at, or on behalf of, the University.”</p> <p>University of Oxford: “Researchers, departments/faculties, divisions, central administrative units and service providers and, where appropriate, research sponsors and external collaborators, need to work in partnership to implement good practice (...)”</p> <p>University of the Arts London: The policy applies to all staff involved in externally funded research at the University, especially where the funding body requires a DMP</p> <p>Aalto University: RDM Policy “to make data management easier for individual researcher”</p> <p>Humboldt-Universität zu Berlin: Policy is addressed to all researchers</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>6. Preliminaries and definitions.</p> <p>A policy should contain key RDM terms, indicating answers to these questions:</p> <p>a. What is “research data”?</p> <p>b. What is “research”?</p> <p>c. Who is a “researcher”?</p>	<p>University of Edinburgh *</p> <p>University of Helsinki</p>	<p>University of Birmingham</p> <p>University of Cambridge</p> <p>Aalto University</p> <p>Radboud University</p> <p>Universität Heidelberg</p>	<p>STFC</p> <p>UCL</p> <p>University of Bath</p> <p>University of Bristol</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Tampere University of Technology</p> <p>University of Turku</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p>	<p>STFC: Implicit definition of data their policy applies to Detailed data definition</p> <p>UCL: Research Detailed data definition</p> <p>University of Bath: Research Research data</p> <p>University of Birmingham: Data Management</p> <p>University of Bristol: Research data Data Steward</p> <p>University of Cambridge: Research data</p> <p>University of Glasgow: Data (research data) Metadata DMPs Persistent object identifier</p> <p>University of Leeds: Research data Research data lifecycle</p> <p>University of Oxford: Research data Research Researcher</p> <p>University of the Arts London: Research data</p> <p>Aalto University: Policy applies to digital research materials produced, used and revised in research projects, i.e. research data. (Physical materials shall be excluded). “The concept of research data is not specified further in this policy (...).”</p> <p>Tampere University of Technology: Research material DMP</p> <p>University of Turku: Glossary of Open Science and Research project (http://avointiede.fi/keskeinen-sanasto)</p> <p>Radboud University: Metadata DMP</p> <p>Humboldt-Universität zu Berlin: Research data</p> <p>Universität Bielefeld: Research data</p> <p>Universität Göttingen: Research data RDM</p> <p>Universität Heidelberg: Lifecycle</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
7. Institutional awareness, support and services.		STFC University of Bath University of Leeds University of the Arts London Radboud University Humboldt-Universität zu Berlin Universität Bielefeld	UCL University of Birmingham University of Bristol University of Cambridge University of Edinburgh University of Glasgow University of Oxford Aalto University Tampere University of Technology University of Helsinki University of Turku Universität Göttingen Universität Heidelberg	<p>UCL: Identify and implement training or skills development (by the Heads of Department)</p> <p>University of Birmingham: University provides training, support, advice, guidelines and templates for RDM and DMPs</p> <p>University of Bristol: Training and guidance for researchers Support with DMP and in depositing research data in the University's Research Data Repository</p> <p>University of Cambridge: Dedicated website providing guidance in good data management practice</p> <p>University of Edinburgh: Training, support and advice for RDM and DMPs</p> <p>University of Glasgow: Own webpage for support with RDM at Glasgow, Funder Requirements, Storage and Costs, Creating – Organising – Accessing Data Discipline-specific data management training, support and advice, particularly on aspects such as data ownership and ethics Local guidance and support to assist researchers in developing and implementing DMPs</p> <p>University of Leeds: Training, support and advice on RDM</p> <p>University of Oxford: University should provide necessary resources for services and training</p> <p>Tampere University of Technology: Training and orientation for university community including students Support for identifying and solving legal issues</p> <p>University of Helsinki: Training as part of studies and staff training</p> <p>University of Turku: Support for researchers for identifying and solving legal and ethical issues related to research data Training as part of studies and staff training University community is informed about data management and media visibility of data is followed</p> <p>Radboud University: Services on the website</p> <p>Universität Bielefeld: Commitment of the university to support implementation and quality-assurance is available on accompanied webpage "Resolution on RDM"</p> <p>Universität Heidelberg: "Kompetenzzentrum Forschungsdaten" (consulting and support)</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>8. Objectives ("what and how").</p> <p>It should be clear what should be done and how it should be done</p>		<p>STFC</p> <p>University of Edinburgh</p> <p>University of Leeds</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Radboud University</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p> <p>Universität Heidelberg</p>	<p>UCL</p> <p>University of Bath</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Glasgow</p> <p>Aalto University</p>	<p>University of Bath: Registration of data within 12 months</p> <p>Aalto University: 5 Principles for open access publishing of research data (recommended to be acknowledged in data management in general)</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>9. Roles and responsibilities.</p> <p>Definition of the responsibilities, tasks and instruments, of:</p> <p>a. the researchers / P.I. (data producing entity)</p> <p>b. research supporting entities (e.g. research services, libraries, IT services)</p> <p>c. the institution</p>		<p>STFC</p> <p>University of Edinburgh</p> <p>University of Leeds</p> <p>Aalto University</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Heidelberg</p>	<p>UCL</p> <p>University of Bath</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Glasgow</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Radboud University</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p>	<p>UCL: Detailed description of responsibilities Students as Data Creators UCL Research Data Service</p> <p>University of Bath: e.g. Data Steward Data loss Contact for queries</p> <p>University of Birmingham: PI Researchers Students University All those undertaking research within the University (including students) have a responsibility to manage their data effectively</p> <p>University of Bristol: Researchers PI Data Steward Postgraduate Research Students and Supervisor University</p> <p>University of Cambridge: University University staff and students</p> <p>University of Edinburgh: PI University</p> <p>University of Glasgow: Researchers School and College Level Support University Services</p> <p>University of Leeds: Responsible owners PI University Researchers Research and Innovation Board</p> <p>University of Oxford: Researchers University</p> <p>University of the Arts London: University PI Director of Research Management and Administration (RMA) Research assistants</p> <p>Radboud University: Researcher Project leader Director of research institute Director of education University</p> <p>University of Turku: Each university community member</p> <p>Humboldt-Universität zu Berlin: Obligation of researchers includes instructing students and doctoral candidates about handling of research data properly "Researchers should take responsibility for deciding at what time and on what legal terms research data may be accessed."</p> <p>Universität Bielefeld: Researcher PI Rektorat (on accompanied webpage "Resolution on RDM")</p> <p>Universität Göttingen: PI Researcher University</p> <p>Universität Heidelberg: PI University</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>10. DMP. The policy should specify a requirement to complete a DMP (either institutional or funder)</p>	<p>Humboldt-Universität zu Berlin</p>	<p>University of Oxford University of the Arts London Aalto University</p>	<p>STFC UCL University of Bath University of Birmingham University of Bristol University of Cambridge University of Edinburgh University of Glasgow University of Leeds Tampere University of Technology University of Helsinki University of Turku Radboud University Universität Bielefeld Universität Göttingen Universität Heidelberg</p>	<p>STFC: Consistent with DMPs of other facilities National and international recommendations for best practice (DCC guidance)</p> <p>University of Birmingham: From 2015, all new research proposals must include DMPs or protocols</p> <p>University of Bristol: DMP should be written before research commences DMP guidance for specific funders on website DMP template for PGR students – DMP online tool by DCC</p> <p>University of Cambridge: Guidance by University of Cambridge and DCC If funders require a DMP, such plan needs to be prepared according to funders' requirements Researchers should update their DMPs regularly, ensure that at the end of the project all their research outputs, together with their location, are indicated in their DMPs and deposit their final DMPs into an appropriate repository</p> <p>University of Edinburgh: All new research proposals must include DMPs</p> <p>University of Glasgow: Researchers have to produce DMP for every research project that will generate a dataset</p> <p>University of Leeds: DMP must be created for each proposed research project or funding application to allow costing and infrastructure planning. Once project is approved DMP should be updated</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>11. Costs.</p> <p>Questions around the costs of RDM should be well defined</p>	<p>University of Edinburgh</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Radboud University</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Heidelberg</p>	<p>STFC</p> <p>UCL</p> <p>University of Birmingham</p> <p>University of Cambridge</p> <p>University of Glasgow</p> <p>Aalto University</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Universität Göttingen</p>	<p>University of Bath</p> <p>University of Bristol</p> <p>University of Leeds</p>	<p>STFC: Efficient and cost-effective research Appropriate to use public funds</p> <p>University of Birmingham: Researchers should seek to cover direct costs of RDM from research funder DMP will include costing RDM</p> <p>University of Bristol: Time and any likely cost for storage and management should be explicitly written into research applications, including instances where data will need to be made publicly available or curated for many years beyond the project lifetime. Funders: costs relating to storage and management of research data are legitimate costs and can be included within a research proposal. These costs can generally only cover the lifetime of the grant so any work needed to make the data available for sharing at the end of the project should be built into the proposal. Research Data Service's Anticipating the costs of RDM document Potential costs for larger deposit</p> <p>University of Glasgow: Costs are not mentioned in the policy, but on the website: "Cost of storing data (...): Research Data is £1800 per-terabyte (excluding VAT). This is a one-off charge and guarantees secure data storage for ten years. "</p> <p>University of Leeds: Guide for costing and infrastructure planning is available on the website. Researchers should seek to recover the direct costs of managing research data generated by projects from the research funder</p> <p>Aalto University: Opening access to research data shall be implemented in a cost-effective manner</p> <p>Radboud University: "Previous research suggests that a centralised service for data management at Radboud University would be more cost effective than management at an institutional level."</p> <p>Universität Göttingen: "Specific requirements have to be aligned among all stakeholders and may involve additional funding."</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>12. External.</p> <p>The policy should contain a statement on the primacy of external funding requirements and about external research partners</p>	<p>Tampere University of Technology *</p> <p>University of Turku *</p> <p>Humboldt-Universität zu Berlin *</p> <p>Universität Göttingen</p>	<p>University of Edinburgh *</p> <p>University of Leeds *</p> <p>University of the Arts London *</p> <p>Aalto University *</p> <p>University of Helsinki *</p> <p>Radboud University *</p> <p>Universität Bielefeld *</p> <p>Universität Heidelberg</p>	<p>STFC *</p> <p>UCL *</p> <p>University of Bath *</p> <p>University of Birmingham *</p> <p>University of Bristol *</p> <p>University of Cambridge *</p> <p>University of Glasgow *</p> <p>University of Oxford</p>	<p>UCL: Collaborative research</p> <p>University of Birmingham: Data retained elsewhere should also be recorded with University Funder-compliant storage Third-party Intellectual Property rights in “Code of Practice for Research” (3 Research Data)</p> <p>University of Bristol: Where research is carried out under a grant or contract: terms of agreement will determine ownership and rights to exploit the data External research partners Third party funded research data of PGRs should be passed on to supervisor before the student leaves the University</p> <p>University of Cambridge: “The University is committed to achieving compliance with the data policies of its external research sponsors, publishers and governmental agencies, and requires its staff and students to abide by terms and conditions agreed with third parties. The University also recognises that such third parties’ policies are evolving and that they may require higher levels of data accessibility and dissemination in the future.”</p> <p>University of Edinburgh: Data retained elsewhere should be registered with the University</p> <p>University of Glasgow: Data retained elsewhere should also be recorded with University data registry Researchers have to “familiarise themselves with relevant funder data policies and expectations and endeavour to comply with these policies.”</p> <p>University of Leeds: Research Funder data requirements available on the website Data held outside the University should be recorded in the University data registry</p> <p>University of Oxford: Overview of major research funders’ data policies (DCC)</p> <p>Universität Heidelberg: Data retained elsewhere should also be recorded with the University</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>13.Ownership.</p> <p>The questions concerning the ownership of research data should be taken into consideration</p>	<p>University of Edinburgh **</p> <p>University of the Arts London *</p> <p>Radboud University *</p> <p>Universität Bielefeld *</p> <p>Universität Göttingen</p>	<p>STFC *</p> <p>UCL *</p> <p>University of Birmingham *</p> <p>University of Cambridge *</p> <p>University of Leeds *</p> <p>University of Oxford *</p> <p>Tampere University of Technology *</p> <p>University of Turku *</p> <p>Humboldt-Universität zu Berlin *</p> <p>Universität Heidelberg</p>	<p>University of Bath *</p> <p>University of Bristol *</p> <p>University of Glasgow *</p> <p>Aalto University *</p> <p>University of Helsinki</p>	<p>UCL: Owner is responsible for preserving research data</p> <p>University of Bath: Student's data</p> <p>University of Birmingham: in "Code of Practice for Research" (3 Research Data)</p> <p>University of Bristol:</p> <ul style="list-style-type: none"> • Where research is carried out under a grant or contract: terms of agreement will determine ownership • Where no external contract exists: University normally has ownership of primary data generated in the course of research undertaken by researchers in its employment • University does not automatically own student Intellectual Property (IP) <p>Suitable agreements for ownership should be established and agreed in writing by parties concerned before a project starts</p> <p>University of Glasgow: Researchers have to:</p> <ul style="list-style-type: none"> • "Clearly state who owns the data that are being generated through the research activity. Where this is not clear, researchers will work with IPR specialists in Research Strategy and Innovation, the Library and College support teams to verify data ownership as early as possible in the research data lifecycle." • "Ensure that, when leaving the University (for retirement or a position elsewhere), data of long-term value which were generated using University resources are deposited in the Institutional Data Repository for long-term storage and preservation." <p>University of Leeds: Responsibilities of the responsible owners</p> <p>Aalto University: Ownership of copyright protected research data is transferred to the University if the data is created in externally funded research project of the University</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>14. Retention.</p> <p>The length of time and criteria on what research data is required to be kept, should be defined.</p>	<p>University of Glasgow</p> <p>University of Leeds</p> <p>University of the Arts London</p> <p>Aalto University</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p> <p>Universität Heidelberg</p>	<p>University of Birmingham</p> <p>University of Cambridge</p> <p>University of Edinburgh</p> <p>Humboldt-Universität zu Berlin</p>	<p>STFC</p> <p>UCL</p> <p>University of Bath</p> <p>University of Bristol</p> <p>University of Oxford</p> <p>Radboud University</p>	<p>STFC: Original data retained for the longest possible period 10 years after end of project reasonable minimum Not re-measurable data: retain in perpetuity</p> <p>UCL: Min. 10 years after publication Plan for custodial responsibilities</p> <p>University of Bath: Data must be retained for 10 years. "Researchers should avoid retaining data using methods that might not persist for 10 years, such as use of project websites or personal computing equipment."</p> <p>University of Birmingham: in "Code of Practice for Research" (3 Research Data): 10 years clinical, major social, environmental or heritage importance: 20 years</p> <p>University of Bristol: "In order to meet funder requirements around the storage, preservation and accessibility of research data, unless otherwise agreed the University is expected to keep a copy of any significant research data for a specified period after the end of the research (generally 10 years)."</p> <p>University of Cambridge: As long as data seems to be valuable to data creator or to others, or required by funder/other regulatory requirements</p> <p>University of Edinburgh: Research data of future historical interest (and records of University) will be offered and assessed for deposit and retention</p> <p>University of Oxford: Min. 3 years after publication As long as they are of continuing value</p> <p>Radboud University: "The retention period for research data is a minimum of ten years." The minimum retention period for Radboud University is longer than the code of academic practice suggests. A longer minimum period can be applied by each discipline. A maximum period cannot be defined, because it is dependent on the discipline."</p> <p>Humboldt-Universität zu Berlin: Researchers are committed to secure their research data for the long term</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>15. Deletion.</p> <p>It should be clear how the deletion of data should be carried out and who decides about it</p>	<p>STFC</p> <p>University of Bath</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Edinburgh</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>University of the Arts London</p> <p>Aalto University</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Radboud University</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p> <p>Universität Heidelberg</p> <p>*</p>	<p>UCL</p> <p>Tampere University of Technology</p>	<p>University of Oxford</p>	<p>UCL: Suggests recommendation in DMP for destruction of research data</p> <p>University of Oxford: Reason: agreed period of retention has expired or legal or ethical reasons Should be done in accordance with legal, ethical, research funder and collaborator requirements (confidentiality and security)</p> <p>Tampere University of Technology: Intentional destruction of data in DMP</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
16. Legal aspects.		UCL University of Edinburgh University of Oxford University of the Arts London Radboud University Humboldt-Universität zu Berlin Universität Bielefeld Universität Göttingen	STFC University of Bath University of Birmingham University of Bristol University of Cambridge University of Glasgow University of Leeds Aalto University Tampere University of Technology University of Helsinki University of Turku Universität Heidelberg	<p>STFC: Compromising data integrity (modification of data or incorrect metadata) considered as serious breach of policy Users acknowledge source of data</p> <p>University of Bath: Guidance on selecting licence for research data</p> <p>University of Bristol: "Researchers must ensure that they abide by licences or terms of use when using or sharing third party data." "(...) Exclusive rights to research data must not be assigned, licenced or otherwise transferred to external parties."</p> <p>University of Cambridge: Intellectual Property Rights</p> <p>University of Glasgow: Exclusive rights to reuse or publish should not be handed over to commercial publishers or agents without retaining rights to make data openly available for re-use (unless this is condition of funding)</p> <p>University of Leeds: Guidance on sharing and publication of research data Relevant legislative frameworks</p> <p>Aalto University: Guidelines with "Rules of handling of information materials" Recommended license for research data: CC BY 4.0, metadata: CCO 1.0, software: MIT Licence User rights of third parties – University may charge a fee for the use of research data</p> <p>Tampere University of Technology: Security and data protection Authors appropriately acknowledged by reuse Fee and restrictions on data sets processed for industry or society</p> <p>University of Helsinki: Good practice for attribution of authorship University of Helsinki must always be indicated as the source of data Fee for data sets processed for business and society</p> <p>University of Turku: Attribution of authorship University of Turku must always be indicated as the source of data University has at least rights of use Fee for data sets processed for business and society Creator's right to primary use of research data Commercial utilisation and related protection of rights</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>17. Ethics.</p> <p>The ethical use/reuse of data, particularly how it affects potential reuse should be considered</p>	<p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p>	<p>STFC</p> <p>UCL</p> <p>University of Edinburgh</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>Radboud University</p> <p>Universität Göttingen</p>	<p>University of Bath</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>Aalto University</p> <p>University of Turku</p> <p>Universität Heidelberg</p>	<p>University of Bristol: University has developed methods to provide controlled access to sensitive data Ethics of Research Policy and Procedure</p> <p>University of Cambridge: Research Ethics Policy</p> <p>University of Glasgow: Researchers have to ensure that sensitive data is properly managed (Data Protection Policy, Confidential Data Policy)</p> <p>University of Leeds: Guidance on good practice in ethics and ethical review</p> <p>Aalto University: Guidelines for ethical principles, responsible conduct of research and processing of personal data</p> <p>University of Helsinki: Protection of confidential information Data security and protection</p> <p>University of Turku: Processing and preservation of personal data and sensitive material in DMP</p>
<p>18. Regulations for:</p> <p>a) open data</p> <p>b) restricted data</p> <p>c) closed data</p> <p>should be made a subject of discussion</p>	<p>Radboud University</p> <p>Universität Bielefeld</p>	<p>UCL</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Edinburgh</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Göttingen</p> <p>Universität Heidelberg</p>	<p>STFC</p> <p>University of Bath</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>Aalto University</p> <p>University of Turku</p>	<p>STFC: Length of proprietary period specified in DMP Data publicly available</p> <p>University of Birmingham: in “Code of Practice for Research” (3 Research Data)</p> <p>University of Bristol: Open and restricted data mentioned in relation to storage</p> <p>University of Cambridge: “There is a balance between openness and duties under professional codes and legal obligations” Make research data as widely and openly available as possible</p> <p>University of Glasgow: Publicly funded research data openly available with as few restrictions as possible</p> <p>Aalto University: Research data is not opened if the opening would violate privacy, safety, security, terms of project agreements or legitimate concerns of private partners</p> <p>Tampere University of Technology: All research materials open by default</p> <p>University of Turku: Leading theme in data policy is openness Openness can be limited for justified reason</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>19. Storage and access.</p> <p>It should be addressed by the policy where data will be stored and how it will be accessed.</p>	<p>Universität Bielefeld</p>	<p>University of Edinburgh</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Heidelberg</p>	<p>STFC</p> <p>UCL</p> <p>University of Bath</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>Aalto University</p> <p>Radboud University</p> <p>Universität Göttingen</p>	<p>STFC: Published data to publication available within 6 months Use of different repositories</p> <p>UCL: Research data: attributable, citable, identifiable, retrievable, available, secure (...) Long-term preservation</p> <p>University of Bath: Security measures</p> <p>University of Birmingham: Security of research data</p> <p>University of Bristol: University's Research Data Repository – limited amount of free storage Long-term retention Statement on how to access supporting data of published outputs should be ensured by researchers Information security policies</p> <p>University of Cambridge: Publicly accessible discipline-based or institutional repository When depositing research data into external data repositories, researchers should choose repositories which support Open Researcher and Contributor ID (ORCID)</p> <p>University of Edinburgh: National or international data service or domain repository or a University repository</p> <p>University of Glasgow: Researchers have to:</p> <ul style="list-style-type: none"> • “Work with IT Services and College IT teams to identify storage requirements that may exceed those currently offered by the institution. • “Store their data during the course of their research in accordance with guidance from IT Services and funder requirements.” • “Deposit data in a reputable repository for long term preservation and sharing.” <p> University Services have to “provide a dedicated institutional research data repository with appropriate security and backup.”</p> <p>University of Leeds: All relevant research data should be offered and assessed for deposit and preservation in an appropriate University, national or international data service or domain repository: Guidance</p> <p>University of Oxford: Planning for the ongoing custodianship (at the University or using third party services) of data after the completion of research or, in event of departure or retirement from the University Agreement with the head of department/faculty as to where data will be located and how this will be stored</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
				<p>Aalto University: Research data and necessary software to access data shall be easily accessible Embargo period can be agreed upon Data chosen for long-term preservation shall be safely stored and curated Necessary software stored together with research data</p> <p>Tampere University of Technology: Long-term preservation and reuse All materials must be retrievable and citable</p> <p>University of Helsinki: Discoverability and citability</p> <p>University of Turku: Discoverability and citability</p> <p>Humboldt-Universität zu Berlin: Long-term preservation Open Access Declaration</p> <p>Universität Göttingen: “Storage and archiving of digital research data is carried out within the technological and informational infrastructure of the University or in acknowledged external or internal subject repositories.”</p> <p>Universität Heidelberg: Long-term preservation Open-Access-Policy</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
20. Metadata curation.	University of Bristol University of Edinburgh University of Leeds University of Oxford Humboldt-Universität zu Berlin Universität Bielefeld Universität Göttingen Universität Heidelberg	UCL University of Birmingham University of the Arts London Aalto University	STFC University of Bath University of Cambridge University of Glasgow Tampere University of Technology University of Helsinki University of Turku Radboud University	<p>STFC: Sufficient metadata to enable re-use</p> <p>University of Birmingham: Sufficient metadata description to aid discovery and re-use</p> <p>University of Cambridge: Metadata Guidance</p> <p>University of Glasgow: Definition of metadata Support by the University Services</p> <p>University of the Arts London: To enable discoverable, accessible and effective re-use</p> <p>Tampere University of Technology: Metadata describes structure of data and how it was created Must specify owner and legal restrictions</p> <p>University of Helsinki: Metadata must contain owner and legal restriction</p> <p>University of Turku: Metadata must contain owner and legal restriction</p>
21. Exceptions. It should be clear what exceptions there are in the policy and what their extent is	UCL University of Cambridge University of Edinburgh University of Leeds Tampere University of Technology Humboldt-Universität zu Berlin Universität Bielefeld Universität Heidelberg	University of Birmingham University of the Arts London Aalto University Radboud University Universität Göttingen	STFC University of Bath University of Bristol University of Glasgow University of Oxford University of Helsinki University of Turku	<p>University of Bristol: "The policy does not currently apply to taught postgraduate students or undergraduates (apart from in exceptional circumstances)."</p> <p>University of Oxford: "(...) Where research is supported by a contract with or a grant to the University that includes specific provisions regarding ownership, retention of and access to data, the provisions of that agreement will take precedence."</p> <p>University of Helsinki: "This policy does not cover the physical resources on which research data are based (e.g., paper materials) or the use of biological research material."</p> <p>University of Turku: "The data policy does not apply to physical and biological materials and the University's practices related to them are presented in the research infrastructure policy of the University of Turku."</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>22. There should be a recommendation for institutional research infrastructure</p>	<p>STFC UCL University of Birmingham University of Bristol University of Edinburgh University of Oxford University of the Arts London Humboldt-Universität zu Berlin Universität Bielefeld Universität Heidelberg</p>	<p>University of Bath University of Cambridge University of Leeds Aalto University Universität Göttingen</p>	<p>University of Glasgow Tampere University of Technology University of Helsinki University of Turku Radboud University</p>	<p>University of Cambridge: Infrastructure and training to promote best practice in data management amongst academics</p> <p>University of Glasgow: Technical infrastructure and services</p> <p>University of Leeds: Costing and infrastructure planning</p> <p>Tampere University of Technology: Tools and services</p> <p>University of Helsinki: Tools and services</p> <p>University of Turku: Tools and services “(...) Data infrastructure is built and developed together with national and international parties, taking into account the services and infrastructures that they offer.”</p> <p>Universität Göttingen: Services for research data infrastructure</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>23. Researchers should know how to deal with:</p> <p>a. the long tail of data</p> <p>b. the head of project data</p>	<p>University of Bath</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Edinburgh</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Aalto University</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Radboud University</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p> <p>Universität Heidelberg*</p>	<p>STFC</p> <p>UCL</p>		<p>STFC: Very large data sets</p> <p>UCL: Curate smaller collections of digital research data</p>

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>24. Educational data should be mentioned in the policy</p>	<p>*</p> <p>STFC</p> <p>UCL</p> <p>University of Bath</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Edinburgh</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Aalto University</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Radboud University</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p> <p>Universität Heidelberg</p> <p>*</p>			

Number of institutions: 20	It was NOT taken into consideration *	It was PARTLY taken into account *	It has been CONSIDERED *	Comments
<p>25. Cultural heritage should be an issue</p>	<p>*</p> <p>STFC</p> <p>UCL</p> <p>University of Bath</p> <p>University of Birmingham</p> <p>University of Bristol</p> <p>University of Cambridge</p> <p>University of Edinburgh</p> <p>University of Glasgow</p> <p>University of Leeds</p> <p>University of Oxford</p> <p>University of the Arts London</p> <p>Aalto University</p> <p>Tampere University of Technology</p> <p>University of Helsinki</p> <p>University of Turku</p> <p>Radboud University</p> <p>Humboldt-Universität zu Berlin</p> <p>Universität Bielefeld</p> <p>Universität Göttingen</p> <p>Universität Heidelberg</p> <p>*</p>			

Selected European Policies

STFC (UK): *STFC scientific data policy*– (April 2016)

Very detailed policy addressing most of the identified main topics. Contains also advice: “Any deliberate attempt to compromise [data] integrity, e.g. by the modification of data or the provision of incorrect metadata, will be considered as a serious breach of this policy.”

[<http://www.stfc.ac.uk/stfc/cache/file/D0D76309-252B-4EEF-A7BFAF6271B8EC11.pdf>; last accessed 07/03/2017]

UCL - University College London (UK): *UCL Research Data Policy* (2 August 2013)

Clearly arranged policy addressing most of the identified main topics with particular focus on roles and responsibilities (data creators, students, supervisors and researchers) | UCL Research Data and Network Services Executive | Director of UCL Library Services and UCL Records Manager | RIISG | Vice Provost (Research) | Provost).

[<http://www.ucl.ac.uk/isd/services/research-it/documents/uclresearchdatapolicy.pdf>; last accessed 07/03/2017]

University of Bath (UK): *Research Data Policy* (9 April 2014)

The policy is only available on the website of the university (not as pdf document) and the policy text is complemented by: Research Data Policy guidance (27 March 2015). Very detailed policy addressing most of the identified main topics. Contains also limitation: “Researchers should avoid retaining data using methods that might not persist for 10 years, such as use of project websites or personal computing equipment.”

[<http://www.bath.ac.uk/research/data/policy/research-data-policy.html> and <http://www.bath.ac.uk/research/data/policy/research-data-policy-guidance.html>; last accessed 07/03/2017]

University of Birmingham (UK): *University RDM policy* (May 2014)

The policy is only available on the website of the university (not as pdf document) under the overarching topic: “Principles of Research Data Management”. The policy is a single-page ten-point paper addressing most of the identified main topics.

[<https://intranet.birmingham.ac.uk/as/libraryservices/library/research/rdm/rdm-principles.aspx>; last accessed 07/03/2017]

University of Bristol (UK): *Research Data Management and Open Data Policy* (19 October 2015)

The University of Bristol provided a draft policy in June 2014 (Research Data Management Principles). The content of the draft version was substantially expanded but the clearly arranged document history disappeared. Very detailed policy addressing most of the identified main topics. Guidance with additional information is also provided on the website (not as pdf document): *Research Data Management and Open Data Policy Guidance*. The first issue the policy addresses is “Ownership of Data”. Document for guidance about costs: *Anticipating the costs of research data management* (October 2015). Contains also commitment: “(...) Funders require that research data is preserved after the end of a project (typically for at least 10 years). There is a cost to the technical curation of data which cannot be built into project funding, therefore the University is committing to meeting these costs”.

[http://www.bristol.ac.uk/media-library/sites/university/documents/governance/UOB_RDM_Policy.pdf, <https://data.bris.ac.uk/rdm-policy-guidance/> and <https://drive.google.com/drive/folders/0B-sxe4ro-QTTSzIIrDBUeHIGY0U>; last accessed 07/03/2017]

University of Cambridge (UK): *Research Data Management Policy Framework* (23 April 2015)

The policy is only available on the website of the university (not as pdf document). Detailed policy addressing most of the identified main topics. The focus is on the responsibilities of the University, staff and students; e.g.: “The University is responsible for managing a dedicated website providing guidance for the University’s academics in good data management practice.” Contains also a collection of RDM policies of major research funders in the UK.

[<http://www.data.cam.ac.uk/university-policy>; last accessed 07/03/2017]

The University of Edinburgh (UK): *Research Data Management Policy* (16 May 2011)

The policy is only available on the website of the university (not as pdf document) and is complemented

by the *Research Data Management (RDM) Roadmap. August 2012-July 2016* (September 2015) from the Information Services RDM Policy Implementation Committee. It is a 1 page document with 10 points partly addressing a large part of the identified main topics.

[<http://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy> and http://www.ed.ac.uk/files/atoms/files/uoerdm-roadmap_-_v2_0_0.pdf; last accessed 07/03/2017]

University of Glasgow (UK): *Good Management of Research Data Policy* (19 November 2015)

The draft version was updated (significantly expanded) at the end of 2015. Very detailed policy. No information about retention or deletion. Contains also advice: "It should be noted by all research staff that many major funders now mandate certain research data management actions and failure to meet funder expectations can lead to sanctions as detailed in funder data policies. In addition to this, failure to implement good research data management can potentially lead to situations which expose researchers to research misconduct allegations."

[http://www.gla.ac.uk/media/media_435489_en.pdf; last accessed 07/03/2017]

University of Leeds (UK): *University of Leeds Research Data Management Policy* (June 2015)

The first final version of the policy: *Research Data Management Policy* (July 2012) as part of a policy timeline is available as pdf document, the updated version is only available on the website. The older version had comments on sufficient metadata, the new version has comments on costing and infrastructure planning instead. Clearly arranged 1 page policy with 10 points, many embedded links and additional guidance for most of the identified main topics. Contains a list of benefits (9) by implementing the policy.

[<https://library.leeds.ac.uk/research-data-policies>] (Last accessed 07/03/2017)

University of Oxford (UK): *Policy on the Management of Research Data and Records* - (9 July 2012)

Clearly arranged, precisely formulated policy with 12 points addressing most of the identified main topics. Contains well-defined information about deletion and reference to other university policies.

[http://blogs.bodleian.ox.ac.uk/wp-content/uploads/sites/126/2014/01/Policy_on_the_Management_of_Research_Data_and_Records.pdf ; last accessed 07/03/2017]

University of the Arts London (UK): *UAL Research Data Management Policy* (2014)

Clearly arranged policy addressing most of the identified main topics. The text is divided into the following topics: Background | Aims | Principles | Scope (What does it cover? Who does it apply to?) | Roles and responsibilities | Workflow.

[<http://www.arts.ac.uk/media/arts/research/documents/UAL-Research-Data-Management-Policy-2014.pdf>; last accessed 07/03/2017]

Aalto University (FIN): *Aalto University Research Data Management Policy* (10 February 2016)

The precisely formulated policy seems to be rather a strategy focusing on the promotion of open access publishing (+ 5 Principles for open access publishing of research data). Roles and responsibilities are hardly addressed.

[<https://tinyurl.com/hol4fyt>; last accessed 07/03/2017]

Tampere University of Technology (FIN): *TUT Research data policy* (21 January 2016)

The policy (divided into 6 points with the main focus on DMPs) is only available on the website of the university (not as pdf document). Contains a detailed description of the work of the Research data policy working group.

[<http://scienceport.tut.fi/researchdataservices>; last accessed 07/03/2017]

University of Helsinki (FIN): *Research Data Policy* (11 February 2015)

The policy (divided into 8 points) is only available on the website of the university (not as pdf document). Contains also licence: "© 2015 University of Helsinki, licenced under the Creative Commons Attribution 4.0 International licence". Addresses most of the identified main topics but "Definition" and "Retention".

[<http://www.helsinki.fi/kirjasto/en/get-help/management-research-data/research-data-policy/>; last accessed 07/03/2017]

University of Turku (FIN): *Open science and research data policy of the University of Turku* (9 February 2016)
Clearly arranged, well-designed policy with 12 points which indicates other “utilised data policies”: University of Helsinki, Concordat on Open Research Data coordinated by the Research Councils UK, and training sessions of the Open Science and Research project of the Ministry of Education and Culture. Main topics are: Starting points | Responsibility (of each university community member) | Legal and ethical issues | Data Management Infrastructure | Training, Orientation and Instructions | Communication | Realisation of data policy.
[<https://www.utu.fi/en/news/Documents/datapolitiikka-en-2.pdf>; last accessed 07/03/2017]

Radboud University (NL): *University policy for storage and management of research data* (25 November 2013)
A brief summary with 4 main elements of the policy is available on the university website. The policy focuses mainly on the storage of selected data (including dissertations, Bachelor’s and Master’s theses) with a well-defined retention period (minimum 10 years) and on the responsibilities within the university. The university RDM policy will be supplemented by each research institute: 9 thematic focal points are listed which should be included in these policies (Responsibilities | Selection on data | Metadata | Storage | Safety of data | Retention | Accessibility and reuse | Privacy of sensitive data | Support and training). Contains also limitation: “The principles of validation and reproducibility imply that storage on a PC/laptop or a mobile device is not an option.”
[<https://tinyurl.com/hgt9xog>; last accessed 07/03/2017]

Humboldt-Universität zu Berlin (D): *Humboldt-Universität zu Berlin Research Data Management Policy* (8 July 2014)
Clearly arranged policy with main focus on the responsibilities of researchers (individual topics have not been disclosed in detail). Contains also unusual obligation: “Researchers should take responsibility for deciding at what time and on what legal terms research data may be accessed.” The policy is complemented by: *Guidelines - A supplement to the Humboldt-Universität zu Berlin Research Data Management Policy* (25 August 2014).
[<https://www.cms.hu-berlin.de/de/dl/dataman/hu-rdm-policy/view>; last accessed 07/03/2017]

Universität Bielefeld (D): *Principles and guidelines on handling research data at Bielefeld University* (19 July 2011)
The policy is only available on the website of the university (not as pdf document) and is complemented by: *Resolution on Research Data Management* (12 November 2013). Brief policy with general statements mainly about responsibilities.
[<https://data.uni-bielefeld.de/en/policy> and <https://data.uni-bielefeld.de/en/resolution>; last accessed 07/03/2017]

Universität Göttingen (D): *Research data policy of the Georg-August University Goettingen (incl. UMG)* (28 August 2014)
A 1 page document with 10 points partly addressing most of the identified main topics.
[<http://www.uni-goettingen.de/en/488918.html>; last accessed 07/03/2017]

Universität Heidelberg (D): *Research Data Policy. Richtlinien für das Management von Forschungsdaten* (18 July 2014)
The policy is only available in German on the website, not as pdf document. It deals mainly with legal and ethical issues
[<https://www.uni-heidelberg.de/universitaet/profil/researchdata/>; last accessed 07/03/2017]



This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654139.

Part 3

LEARN
Executive
Briefing



Executive Briefing

MANAGING KNOWLEDGE ASSETS FOR RESEARCH AND INNOVATION IN THE 21ST CENTURY

The problem

Research data is the new currency of the digital age. From sonnets to statistics, and genes to geodata, the amount of material being created and stored is growing exponentially. However, the [LERU Roadmap for Research Data](#) identifies a serious gap in the level of preparation amongst research performing organisations. This gulf is prominent in areas such as policy development, awareness of current issues, skills development, training, costs, community building, governance, disciplinary/legal/terminological and geographical differences.

The solution

This **LEARN Executive Briefing** will help decision and policy makers identify sound solutions. In addition, stakeholders can follow the [LEARN Toolkit of Best Practice Case Studies](#), all of which will help organisations to grapple with the data deluge. LEARN also provides a self-assessment survey.¹

Research Data Policy

Every research performing organisation should have a research data policy, which lays down a framework for how research data is curated and managed. Research funders should also have a research data policy, stipulating the obligations that a researcher is expected to meet as a condition of the funding received. LEARN has created a model Research Data Management policy for research performing organisations, along with guidance for the implementation of this policy.² The model LEARN policy can be both adapted and adopted by individual research performing organisations, by regional, national and/or international consortia.

FAIR Data

Best practice indicates that research data should be FAIR³:

Findable – **A**ccessible – **I**nteroperable – **R**eusable

To be findable, the data should be adequately described, using standard taxonomies and ontologies where possible. To be accessible, research data should ideally be open data, available for sharing and reuse. Not all research data can be open, but best practice indicates that such data should be “as open as possible, as closed as necessary”⁴. Research data should also be interoperable, capable of being processed by machines using vocabularies which follow FAIR principles. To be reusable, metadata describing the data should meet domain-relevant community standards.

¹ All available at <http://learn-rdm.eu>; last accessed 16/12/16.

² As in n. 1 above.

³ See <https://www.force11.org/group/fairgroup/fairprinciples>; last accessed 12/12/16.

⁴ European Commission - Guidelines on FAIR Data Management in Horizon 2020 p.4 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf); last accessed 12/12/16.

Research Data Management Stewardship

It is important that researchers plan the collection, curation, description and dissemination of their research data at the start of their research. This information is best captured in a Research Data Management plan, which provides a framework for research data stewardship.⁵

Infrastructure

To curate their research data, researchers and research performing organisations need access to the requisite digital eco-systems. These may be maintained locally; or they may be commercial services, subject domain offerings or regional/national/international platforms. Different subject communities and individual countries will want to provide such facilities in different ways. Commonly, the platform(s) will need to offer the following services:

- Storage, for researchers who are actively collecting data;
- A publication platform, where research data and related software can be made available for sharing and re-use;
- Archive facilities, to allow research data to be curated for the long term, often in response to the requirements of research funders;
- A discovery service, which will allow researchers and citizens to search for research data deposits both locally and across the Internet.

The European Commission is promoting the European Open Science Cloud.⁶ The EOSC is a metaphor to help convey both seamlessness and the idea of a commons based on scientific data. The EOSC will be a federated environment for the sharing and re-use of scientific data, based on existing and emerging elements in the Member States, with lightweight international guidance and governance and a large degree of freedom regarding practical implementation.

Training

The prevalence of research data requires all researchers, new and established, to equip themselves with the skills and tools to be confident in a data-driven environment. The lead needs to be taken by research performing organisations and, in many cases, by their institutional libraries.

Funding

Research data management comes with costs. There is no one method for assessing these costs, but a number of costing models exist to help, for example the 4C Project.⁷

Conclusion

Research data can drive innovation and stimulate new discoveries, to the great benefit of Society. All stakeholders in the research workflow have a role to play. This Executive Briefing highlights what researchers and research performing organisations need to do to rise to this exciting challenge.

⁵ For further information, see <http://www.dcc.ac.uk/resources/data-management-plans>; last accessed 12/12/16.

⁶ See <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; last accessed 12/12/16.

⁷ For further information, see <http://www.4cproject.eu/summary-of-cost-models>; last accessed 12/12/16.





Resumen ejecutivo – Spanish translation

GESTIONAR LOS ACTIVOS DEL CONOCIMIENTO PARA LA INVESTIGACIÓN Y LA INNOVACIÓN EN EL SIGLO XXI

El problema

Los datos de investigación son la nueva divisa de la era digital. Desde sonetos a datos estadísticos, y desde genes a datos geoespaciales, la cantidad de material que se crea y se almacena crece de manera exponencial. Sin embargo, la [Hoja de Ruta para los Datos de Investigación de la LERU](#) identifica una brecha significativa en relación al nivel de preparación entre las instituciones donde se realizan actividades de investigación. La brecha es considerable en áreas como el desarrollo de políticas, el conocimiento de los temas de actualidad, el desarrollo de habilidades, la formación, los costes, la creación de comunidades, la gobernanza, las diferencias disciplinarias/legales/terminológicas y geográficas.

La solución

Este **Resumen Ejecutivo de LEARN** ayudará a quien debe tomar decisiones y elaborar políticas a identificar buenas soluciones. Además, todas las partes interesadas pueden seguir el Documento de [LEARN](#) de Instrumentos de Buenas Prácticas creado a partir de casos prácticos, que ayudarán a las organizaciones a capear el diluvio de datos. LEARN también ofrece una herramienta de autoevaluación.¹

Política sobre datos de investigación

Cada institución que realiza actividades de investigación debería tener una política sobre datos de investigación, que establezca un marco para gestionar y conservar los datos de investigación. Las organizaciones que financian la investigación también deberían tener una política sobre datos, estipulando las obligaciones que deben cumplir los investigadores al recibir financiación. LEARN ha creado un modelo de Política de Gestión de Datos de Investigación para las organizaciones que realizan actividades de investigación, junto con una guía para su implementación.² El modelo de política de LEARN puede ser adaptado y adoptado de manera individual por instituciones que realizan investigación y por consorcios regionales, nacionales y/o internacionales.

Datos según los principios FAIR

Las buenas prácticas indican que los datos de investigación deberían seguir los principios FAIR³:

Findable (Encontrables) – **A**ccessible (Accesibles) – **I**nteroperable (Interoperables) – **R**eusable (Reutilizables)

¹ Disponibles en <http://learn-rdm.eu>; último acceso 07/01/17.

² Como en 1 anterior.

³ Consultar <https://www.force11.org/group/fairgroup/fairprinciples>; último acceso 07/01/17.

Para poder ser encontrados, los datos deberían ser descritos de una manera adecuada, utilizando taxonomías y ontologías estándar cuando fuera posible. Para ser accesibles, los datos de investigación, idealmente deberían ser datos abiertos, es decir estar disponibles para ser compartidos y reutilizados. No todos los datos de investigación pueden ser abiertos, pero las buenas prácticas muestran que los datos deberían ser “tan abiertos como sea posible, y tan cerrados como sea necesario”⁴. Los datos de investigación deberían ser también interoperables, pudiendo ser procesados por máquinas utilizando vocabularios que sigan los principios FAIR. Para ser reutilizables, los metadatos que describen los datos deberían seguir los estándares de cada comunidad relevante en el dominio.

Administración de los datos de investigación

Es importante que los investigadores planifiquen la recopilación, conservación, descripción y difusión de sus datos al inicio de su actividad investigadora. La mejor manera de registrar esta información es mediante un Plan de Gestión de Datos, que ofrece una buena estructura para la administración de los datos de investigación.⁵

Infraestructuras

Para conservar los datos de investigación, los investigadores y las instituciones que desarrollan actividades de investigación, necesitan acceder a ecosistemas digitales adecuados. Pueden ser mantenidos localmente, o bien pueden ser servicios comerciales, estar dirigidos a dominios científicos específicos o ser plataformas regionales/nacionales/internacionales. Cada comunidad temática o cada territorio puede optar por ofrecer estas instalaciones de forma diferente. En principio, la(s) plataforma(s) deben ofrecer los siguientes servicios comunes:

- Almacenamiento, para los investigadores que están recopilando datos de manera activa;
- Una plataforma de publicación, que ofrezca al público los datos de investigación y el software relacionado para ser compartidos y reutilizados;
- Servicios de archivo, para poder conservar los datos a largo plazo, a menudo como respuesta a los requerimientos de los financiadores de la investigación;
- Un servicio de localización, que permita a los investigadores y a la ciudadanía buscar los repositorios de datos de investigación a nivel local y a través de Internet.

La Comisión Europea promueve la Nube Europea de Ciencia Abierta (EOSC).⁶ La EOSC es una metáfora para ayudar a transmitir tanto la fluidez, como la idea de un bien común creado a partir de los datos científicos. La EOSC será un entorno federado para compartir y reutilizar datos científicos, a partir de elementos ya existentes e incipientes en los estados miembro, que contará con pautas internacionales y gobernanza ligeras, así como un amplio grado de libertad en relación a su implementación práctica

Formación

El predominio de los datos de investigación requiere que todos los investigadores, los nuevos y los ya establecidos, se doten de las habilidades y las herramientas para sentirse seguros en este entorno orientado a los datos. Las instituciones donde se desarrolla la investigación deben encargarse de esta formación, y en muchos casos, desde sus bibliotecas institucionales.

⁴ European Commission - Guidelines on FAIR Data Management in Horizon 2020 p.4 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf); último acceso 07/01/17.

⁵ Para más información, consultar <http://www.dcc.ac.uk/resources/data-management-plans>; último acceso 07/01/17.

⁶ Consultar <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; último acceso 07/01/17.

Financiación

Existen costes asociados a la gestión de datos de investigación. No hay un único método para analizar estos costes pero existen varios modelos que pueden ayudarnos, por ejemplo el que propone el Proyecto 4C.⁷

Conclusión

Los datos de investigación pueden impulsar la innovación y estimular nuevos descubrimientos en beneficio de la sociedad. Todos los participantes en el proceso de la investigación tienen un papel que desempeñar. Este resumen ejecutivo resalta lo que los investigadores y las instituciones que desarrollan actividades de investigación necesitan para abordar este apasionante reto.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁷ Para más información, consultar <http://www.4cproject.eu/summary-of-cost-models>; último acceso 07/01/17.



Leitfaden für Führungskräfte – German Translation

MANAGEMENT VON WISSENSBESTÄNDEN – EIN ZENTRALES ANLIEGEN FÜR FORSCHUNG UND INNOVATION IM 21. JAHRHUNDERT

Die Problemstellung

Forschungsdaten sind der Rohstoff für die Wissenschaft im digitalen Zeitalter. Von Sonetten zu Statistiken, von Genfaktoren zu Geodaten – die Menge an Materialien, die generiert und gespeichert werden, wächst exponentiell. Dennoch wird durch die [LERU Roadmap for Research Data](#) evident, dass Forschungseinrichtungen in sehr unterschiedlichem Ausmaß für digitales Datenmanagement gerüstet sind. Diese Kluft tritt vor allem im Zusammenhang mit Policy-Fragen, dem Bewusstsein für aktuelle Themen und Problemstellungen, Kompetenzentwicklung, Ausbildung, Kosten, Netzwerkbildung und Governance sowie disziplinären/rechtlichen/terminologischen und geographischen Unterschieden besonders deutlich zutage.

Die Lösung

Der vorliegende Leitfaden **LEARN Executive Briefing** soll Führungskräften und EntscheidungsträgerInnen dabei helfen, tragfähige Lösungen zu ermitteln. Darüber hinaus können sich alle Beteiligten an den „[LEARN Toolkit of Best Practice](#)“ Case Studies orientieren, allesamt Studien, die dazu geeignet sind, Forschungsinstitutionen bei der Bewältigung der Datenflut zu unterstützen. LEARN bietet auch eine Umfrage zur Selbstbewertung im Hinblick auf Forschungsdatenmanagement an.¹

Policy für Forschungsdatenmanagement

Jede Forschungseinrichtung sollte über eine Policy verfügen, in der Regelungen für die Kuratierung und das Management von Forschungsdaten definiert sind. Forschungsförderer sollten ebenfalls über eine Forschungsdatenmanagement-Policy verfügen, in der Verpflichtungen festgehalten sind, die von den ForscherInnen als Leistung für die erhaltene Finanzierung erwartet werden. LEARN hat eine Vorlage für eine Forschungsdatenmanagement-Policy und einen Leitfaden zur Umsetzung erstellt.² Diese LEARN Modell-Policy lässt sich problemlos adaptieren und ist daher für unterschiedlichste Institutionen und Konsortien geeignet, sei es auf regionaler, nationaler und/oder internationaler Ebene.

Daten nach dem “FAIR”-Prinzip

Best Practice im Umgang mit Forschungsdaten bedeutet, dass diese FAIR³ sein sollten:

Findable (auffindbar) – **A**ccessible (zugänglich) – **I**nteroperable (interoperabel bzw. kompatibel) – **R**eusable (nachnutzbar)

¹ Verfügbar unter <http://learn-rdm.eu>; letzter Zugriff 16.12.2016.

² Wie in 1 oben.

³ Siehe <https://www.force11.org/group/fairgroup/fairprinciples>; letzter Zugriff 12.12.2016.

“Findable” (also „auffindbar“), bedeutet, dass die Daten auf adäquate Weise beschrieben sein sollten, und zwar – wo dies möglich ist - unter Verwendung von Standardtaxonomien und -ontologien. Um „accessible“ (also „zugänglich“) zu sein, sollten Forschungsdaten idealerweise offene Daten sein, also für Sharing und Nachnutzung zur Verfügung stehen. Es können zwar nicht alle Forschungsdaten offen sein, aber Best Practice bedeutet in diesem Zusammenhang, dass der Zugang zu solchen Daten so offen wie möglich und so geschlossen wie nötig sein sollte. Forschungsdaten sollten auch „interoperabel“ bzw. kompatibel sein, in einem maschinenlesbaren Format zur Verfügung stehen und mit einem Vokabular versehen sein, das auf den „FAIR“-Grundregeln basiert. Durch die Beschreibung der Datensätze mit Metadaten, die den jeweiligen disziplinspezifischen Standards entsprechen, unterstützt man das Prinzip der Nachnutzung („reusable“).

Research Data Stewardship

Es ist wichtig, dass ForscherInnen bereits zu Beginn ihrer Forschungsarbeiten die Erfassung, Kuratierung, Beschreibung, Nachnutzung und Dissemination ihrer Forschungsdaten planen. Diese Informationen sind idealerweise in einem Datenmanagementplan erfasst, der somit das Gerüst für einen verantwortungsvollen Umgang mit Forschungsdaten darstellt.⁴

Infrastruktur

Für den adäquaten Umgang mit ihren Forschungsdaten benötigen ForscherInnen sowie Forschungseinrichtungen Zugang zu den erforderlichen digitalen Ökosystemen. Diese können auf lokaler Ebene betrieben werden, es kann sich dabei jedoch auch um kommerzielle Servicedienste, themen- oder disziplinbezogene Angebote, oder um regionale, nationale und internationale Plattformen handeln. Unterschiedliche Wissenschaftscommunitys und unterschiedliche Länder werden derartige Einrichtungen möglicherweise auf unterschiedliche Art und Weise nutzen und bereitstellen wollen. Üblicherweise wird/werden die Plattform(en) folgende Dienstleistungen anbieten:

Storage für ForscherInnen, die aktiv Daten sammeln;

- Eine Publikationsplattform, über die Forschungsdaten und dazugehörige Software geteilt und nachnutzbar gemacht werden;
- Archivfunktionen, die eine langfristige Kuratierung der Forschungsdaten ermöglichen, oftmals als Reaktion auf die von Forschungsförderern gestellten Anforderungen;
- Ein Suchdienst, der es ForscherInnen wie auch der Öffentlichkeit ermöglicht, Datenbestände sowohl auf lokaler Ebene als auch über das Internet zu durchforsten.

Die Europäische Kommission fördert die European Open Science Cloud.⁵ Diese EOSC ist keine Cloud-Lösung im eigentlichen Sinn, sondern eine Metapher für den reibungslosen Austausch von Daten und für den Gedanken von wissenschaftlichen Daten als Gemeinschaftsgut. Die EOSC wird ein gemeinsames Umfeld für die sichere Bereitstellung und Nachnutzung von Forschungsdaten sein, das sich aus bereits bestehenden und noch in Entwicklung befindlichen Infrastrukturen in den Mitgliedsstaaten zusammensetzt, mit einem Minimum an internationaler Führung und Steuerung und einem Maximum an Freiheit hinsichtlich der praktischen Umsetzung.

Training

⁴ Für weitere Informationen siehe <http://www.dcc.ac.uk/resources/data-management-plans>; letzter Zugriff 12.12.16.

⁵ Siehe <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; letzter Zugriff 12.12.2016.

Durch die neuen technischen Möglichkeiten und Arbeitsweisen ist es notwendig, dass sich alle ForscherInnen, etablierte genauso wie JungforscherInnen, all jene Fähigkeiten und Instrumente aneignen, die nötig sind, um sich in einer datengesteuerten Umwelt sicher zu bewegen. Die Führungsrolle sollte dabei von Forschungseinrichtungen und – in vielen Fällen – von deren institutseigenen Bibliotheken übernommen werden.

Finanzierung

Forschungsdatenmanagement ist immer mit Kosten verbunden. Es gibt keine bestimmte, vorgegebene Methode zur Berechnung dieser Kosten, jedoch existiert eine ganze Reihe von Kostenmodellen, die bei dieser Berechnung behilflich sein können, darunter das Projekt 4C.⁶

Schlussfolgerung

Forschungsdaten sind der Motor für Innovationen, sie sind der Ausgangspunkt für neue Entdeckungen und von großem Nutzen für die gesamte Gesellschaft. Alle jene, die in die Arbeitsabläufe von Forschungsprojekten eingebunden sind, haben ihre ganz bestimmten Rollen zu erfüllen. Das vorliegende „Executive Briefing“ zeigt auf, was ForscherInnen und Forschungseinrichtungen leisten müssen, um sich dieser spannenden Herausforderung zu stellen.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁶ Für weitere Informationen siehe <http://www.4cproject.eu/summary-of-cost-models>; letzter Zugriff 12.12.2016.



Informe Executivo – Portuguese translation

GESTÃO DE ATIVOS DE CONHECIMENTO PARA PESQUISA E INOVAÇÃO NO SÉCULO XXI

O problema

Os dados de pesquisas são a nova moeda da era digital. De sonetos a estatísticas, e de genes a geodados, o volume de material que está sendo criado e armazenado cresce exponencialmente. Contudo, o [Roteiro LERU para dados de pesquisas](#) identifica uma grave lacuna no nível de preparação entre organizações de pesquisa. Essa lacuna é proeminente em áreas como formulação de políticas, conscientização acerca das questões atuais, desenvolvimento de habilidades, capacitação, custos, construção de comunidades, governança, diferenças disciplinares, jurídicas, terminológicas e geográficas.

A solução

Este **Informe executivo LEARN** ajudará os tomadores de decisões e formuladores de políticas a identificar boas soluções. Além disso, as partes interessadas podem seguir o kit [LEARN](#) de estudos de caso de boas práticas, que ajudará as organizações a lidar com o enorme volume de dados. LEARN também proporciona uma pesquisa de auto-avaliação.¹

Política de dados de pesquisas

Toda organização de pesquisa deve ter uma política de dados de pesquisas, que inclua um esquema de gestão dos dados de pesquisas. Os financiadores das pesquisas também devem ter uma política de dados de pesquisas, estipulando as obrigações que um pesquisador deve cumprir como condição do financiamento. LEARN criou uma política modelo de gestão dos dados de pesquisas para organizações, bem como orientações para a implementação da política.² A política modelo de LEARN pode ser adaptada e adotada por organizações de pesquisa ou por consórcios regionais, nacionais e internacionais.

Dados FAIR

A boa prática indica que os dados de pesquisas devem ser FAIR³:

Facilmente encontráveis – **A**cessíveis – **I**nteroperáveis – **R**eusáveis

Para serem facilmente encontráveis, os dados devem ser descritos adequadamente, usando taxonomias e ontologias padrão sempre que possível. Para serem acessíveis, os dados de pesquisas idealmente devem ser abertos, disponíveis para compartilhamento e reutilização. Nem todos os dados de pesquisas podem ser abertos, mas a boa prática indica que esses dados devem ser “tão abertos quanto possível, e tão protegidos quanto necessário”⁴. Os dados de pesquisas também devem ser interoperáveis, capazes de serem processados por máquinas usando vocabulários que seguem os princípios FAIR. Para serem

¹ Disponíveis em <http://learn-rdm.eu>; acessado pela última vez em 16/12/16.

² Como em 1 acima.

³ Veja <https://www.force11.org/group/fairgroup/fairprinciples>; acessado pela última vez em 12/12/16.

reusáveis, os metadados que descrevem os dados devem cumprir padrões das comunidades relevantes ao domínio.

Gestão dos dados de pesquisas

É importante que os pesquisadores planejem a coleta, curadoria, descrição e disseminação dos dados no início da pesquisa. A melhor maneira de captar essa informação é mediante um plano de gestão dos dados de pesquisas, que proporciona um quadro para a administração dos dados de pesquisas.⁵

Infraestrutura

Para fazer a curadoria dos dados de pesquisas, os pesquisadores e organizações de pesquisa precisam ter acesso a ecossistemas digitais. Esses ecossistemas podem ser mantidos localmente ou podem ser serviços comerciais, ofertas de domínios de temas ou plataformas regionais, nacionais e internacionais. Diferentes comunidades de temas e países proporcionarão esses meios de maneira diferente. Em geral, as plataformas devem oferecer os seguintes serviços:

- armazenamento, para pesquisadores que coletam dados;
- plataforma de publicação, onde os dados de pesquisas e respectivos softwares podem ser disponibilizados para compartilhamento e reutilização;
- serviços de arquivo, para permitir a curadoria dos dados de pesquisas no longo prazo, em geral como resposta às exigências dos financiadores da pesquisa;
- um serviço de descoberta, que permite que os pesquisadores e cidadãos busquem depósitos de dados de pesquisas localmente e em toda a Internet.

A Comissão Europeia está promovendo a Nuvem Europeia de Ciência Aberta (EOSC).⁶ A EOSC é uma metáfora para ajudar a transmitir uniformidade e a ideia de um ambiente de compartilhamento voltado a dados científicos. A EOSC será um ambiente federado para compartilhamento e reutilização de dados científicos, baseado em elementos atuais e emergentes dos Estados membros, com leve orientação e governança internacional e alto grau de liberdade no tocante à implementação prática.

Capacitação

A prevalência de dados de pesquisas requer que todos os pesquisadores, novos e estabelecidos, se equipem com habilidades e ferramentas para serem confiantes num ambiente baseado em dados. A liderança deve ser assumida por organizações de pesquisa e, em muitos casos, por suas bibliotecas institucionais.

⁴ European Commission - Guidelines on FAIR Data Management in Horizon 2020 p.4 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

⁵ Para mais informações, veja <http://www.dcc.ac.uk/resources/data-management-plans>; acessado pela última vez em 12/12/16.

⁶ Veja <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; acessado pela última vez em 12/12/16.

Financiamento

A gestão dos dados de pesquisas tem múltiplos custos. Não há um método único para avaliar esses custos, mas existem vários modelos de cálculo de custos que podem ajudar, por exemplo, o projeto 4C.⁷

Conclusão

Os dados de pesquisas podem orientar a inovação e estimular novas descobertas, trazendo grandes benefícios para a sociedade. Todas as partes no fluxo de trabalho de pesquisa têm um papel a desempenhar. Este Informe Executivo destaca o que os pesquisadores e organizações de pesquisa devem fazer para enfrentar esse desafio.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁷ Para mais informações, veja <http://www.4cproject.eu/summary-of-cost-models>; acessado pela última vez em 12/12/16.



Synthèse – French Translation

LA GESTION DES DONNÉES DU SAVOIR POUR LA RECHERCHE ET L'INNOVATION AU 21^{ÈME} SIÈCLE

Le problème

Les données de la recherche représentent la nouvelle monnaie d'échange à l'ère numérique. Des sonnets aux statistiques, en passant par les gènes et les données géographiques, la quantité de matériel créé et stocké croît de façon exponentielle. Cependant la [Feuille de Route de LERU pour les Données de la Recherche](#) a identifié d'importantes inégalités entre les organismes de recherche en termes de préparation pour faire face à ces questions. L'écart est particulièrement sévère dans les domaines suivants : l'élaboration de politiques de gestion des données, la sensibilisation aux difficultés actuelles, le développement des compétences, la formation, les coûts, la création de communautés de pratique, la gouvernance, ainsi que les différences entre les disciplines, systèmes législatifs, terminologies et espaces géographiques.

La solution

La présente **Synthèse du projet LEARN** aidera les décideurs à prendre des décisions solides. En outre, le Kit de Bonnes Pratiques de [LEARN](#) est à disposition de tous les acteurs de la recherche; organisé autour de cas d'étude, ce kit permettra aux organismes de recherche de faire face au déluge de données. Le projet LEARN a également développé un outil d'auto-évaluation.¹

Politique des données de la recherche

Tout organisme de recherche devrait adopter une politique des données de la recherche énonçant les responsabilités dont sont investis les chercheurs lorsqu'ils reçoivent un financement. Le projet LEARN a créé un modèle de politique pour la gestion des données dans les organismes de recherche; ce modèle est accompagné de conseils pour mettre en place une telle politique.² Le modèle de politique proposé peut être adapté et adopté par chaque organisme mais aussi par des consortiums régionaux, nationaux et/ou internationaux.

Les données « FAIR »

Selon les bonnes pratiques, les données de la recherche doivent être « FAIR »³:

Faciles à trouver (Findable) – **A**ccessibles (Accessible) – **I**nteropérables (Interoperable) – **R**éutilisables (Reusable)

Pour être faciles à trouver, les données doivent être correctement décrites, si possible à l'aide de taxinomies et d'ontologies. Pour être accessibles, elles doivent, dans l'idéal, être ouvertes (open data) et disponibles

¹ Tous ces outils sont disponibles à l'adresse suivante: <http://learn-rdm.eu> (consulté le 12/01/17).

² Voir 1 ci-dessus.

³ Voir www.force11.org/group/fairgroup/fairprinciples (consulté le 12/01/17).

pour le partage et la réutilisation. Les données de la recherche ne peuvent pas toutes être ouvertes mais les bonnes pratiques indiquent que ces données doivent être «aussi ouvertes que possibles et aussi fermées que nécessaires » («as open as possible, as closed as necessary »)⁴. Ces données doivent aussi être interopérables et lisibles par des machines utilisant un vocabulaire conforme aux principes « FAIR ». Pour être réutilisables, les métadonnées qui décrivent les données doivent respecter les standards établis par le champ de recherche dont il est question.

Plans de gestion des données de la recherche

Il est recommandé aux chercheurs de planifier la collecte, le traitement, la description et la diffusion de leurs données dès le début de leur recherche. Ecrire un Plan de gestion des données de la recherche (Data Management Plan) permet de rassembler ces éléments et d'établir un programme de gestion durant la recherche.⁵

Infrastructures

Afin de traiter leurs données les chercheurs et organismes de recherche doivent pouvoir accéder à un écosystème numérique adéquat. Ces infrastructures peuvent être gérées par chaque institution; elles peuvent aussi être fournies par des services commerciaux ou par des entrepôts de données spécialisés dans un champ de recherche particulier; ou bien elles peuvent prendre la forme de plateformes régionales, nationales ou internationales. Chaque champ de recherche et chaque pays devra trouver le système qui lui convient. De manière générale, ces infrastructures doivent proposer les services suivants :

- Le stockage de données, pour les chercheurs qui en collectent activement;
- Une plateforme de publication, où les données de la recherche et les logiciels associés peuvent être partagés et réutilisés;
- Un système d'archivage, permettant que les données soient traitées et préservées sur le long terme, en conformité avec les exigences des financeurs;
- Un système permettant de chercher parmi les données entreposées, afin que chercheurs et citoyens puissent découvrir ces données (qu'elles soient accessibles in situ ou en ligne).

La Commission Européenne encourage l'usage de l'European Open Science Cloud (EOSC).⁶ L'EOSC est une métaphore qui entend exprimer un processus cohérent et sans obstacle, ainsi que l'idée que les données de la recherche constituent un bien commun (commons). L'EOSC sera un environnement fédérateur permettant de partager et réutiliser ces données. Cet environnement sera fondé sur un ensemble d'infrastructures déjà en place ou émergentes dans les Etats Membres et aura une supervision internationale allégée; une large part de liberté sera accordée aux questions pratiques d'utilisation.

Formation

La prévalence des données de la recherche oblige tous les chercheurs, qu'ils soient débutants ou expérimentés, à s'armer de compétences et d'outils leur permettant de travailler en confiance dans cet environnement où les données abondent. Les organismes de recherche doivent prendre l'initiative en termes d'offre de formation; souvent, la tâche en revient aux bibliothèques universitaires ou de recherche.

⁴ European Commission - Guidelines on FAIR Data Management in Horizon 2020 p.4 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf). (Consulté le 12/01/17).

⁵ Pour plus d'informations, voir www.inist.fr/donnees/co/module_Donnees_recherche_26.html ou www.dcc.ac.uk/resources/data-management-plans (consulté le 12/01/17).

⁶ Voir <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud> (consulté le 12/01/17).

Financement

La gestion des données a un coût. Il n'y a pas de méthode unique pour estimer ce coût mais plusieurs modèles de calcul, tels que le 4C Project, sont disponibles.⁷

Conclusion

Les données de la recherche peuvent favoriser l'innovation et stimuler de nouvelles découvertes pour le plus grand bénéfice de la Société. Tous les acteurs du cycle de la recherche ont un rôle à jouer. Cette Synthèse met en valeur ce que les chercheurs et organismes de recherche doivent faire pour être à la hauteur de ces défis passionnants.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁷ Pour plus d'informations, voir www.4cproject.eu/summary-of-cost-models (consulté le 12/01/17).



Briefing esecutivo– Italian Translation

COME VALORIZZARE LA GESTIONE DELLA CONOSCENZA PER MIGLIORARE LA RICERCA E L'INNOVAZIONE NEL 21° SECOLO

La problematica

Nell'era digitale i dati della ricerca valgono oro. Dai sonetti alle statistiche, dai dati genetici ai geodati: la quantità di materiali creati e memorizzati in formato digitale cresce in maniera esponenziale. La [LERU Roadmap for Research Data](#) (tabella di marcia per i dati della ricerca) identifica tuttavia gravi lacune per quanto riguarda il livello di preparazione delle organizzazioni attive nel campo della ricerca. Questo divario emerge in modo evidente per quanto riguarda lo sviluppo di politiche di gestione dei dati, la consapevolezza di problemi ancora aperti, lo sviluppo di competenze, la formazione, i costi, la creazione di comunità, la *governance* e le differenze disciplinari/legali/terminologiche e geografiche.

La soluzione

Il presente vademecum **LEARN Executive Briefing** vuole essere una guida a disposizione degli attori coinvolti nei processi decisionali e politici al fine di individuare soluzioni sostenibili. Inoltre, i soggetti interessati possono beneficiare del [LEARN Toolkit of Best Practice Case Studies](#), e dei suoi strumenti risultanti da studi casistici di buone pratiche. Ogni singolo strumento aiuterà le istituzioni che operano nell'ambito della ricerca ad affrontare meglio le questioni riguardanti il cosiddetto "diluvio di dati" (data deluge). LEARN offre anche uno strumento quale il questionario di autovalutazione.¹

Research Data Policy (Policy in materia di dati della ricerca)

Ogni organizzazione che svolge attività di ricerca dovrebbe dotarsi di una propria politica dei dati, per poter definire un quadro di riferimento su come curare e gestire i dati della ricerca. Anche i finanziatori della ricerca dovrebbero dotarsi di una tale *policy*, per stabilire gli obblighi che i ricercatori sono tenuti ad adempiere come premessa per la concessione dei finanziamenti. LEARN ha creato un modello di "policy" per la politica di gestione dei dati della ricerca (Research Data Management) per le organizzazioni che svolgono attività di ricerca, oltre a delle linee guida per l'attuazione di questa policy.² Il modello fornito da LEARN può essere adattato e adottato sia da singole organizzazioni di ricerca, che da associazioni attive nella ricerca a livello regionale, nazionale e/o internazionale.

¹ Il tutto è disponibile su <http://learn-rdm.eu>; ultimo accesso: 16/12/16.

² Come sopra1.

I principi FAIR

Secondo le buone pratiche, i dati della ricerca devono essere FAIR³:

Findable – **A**ccessible – **I**nteroperable – **R**eusable (reperibili – accessibili – interoperabili – riutilizzabili)

Per essere “findable”, i dati devono essere adeguatamente descritti, utilizzando, ove possibile, ontologie e tassonomie standardizzate. Per essere “accessible”, i dati della ricerca dovrebbero essere idealmente “open data” (ad accesso aperto), disponibili per la condivisione e il riutilizzo. Non tutti i dati della ricerca possono essere “open data”, ma secondo le buone pratiche l’accesso dovrebbe essere “il più aperto possibile e chiuso solo quanto necessario”. Inoltre, i dati della ricerca dovrebbero essere “interoperable”, ossia in grado di essere elaborati da sistemi operativi con linguaggi conformi ai principi FAIR. Infine, per essere “reusable”, i metadati descrittivi devono conformarsi alle norme comunitarie rilevanti in materia.

Research Data Management Stewardship (Gestione responsabile dei dati della ricerca)

È importante che i ricercatori provvedano a pianificare la raccolta, conservazione, descrizione e divulgazione dei propri dati della ricerca sin dall’inizio dell’attività. Queste informazioni dovrebbero essere integrate in un Research Data Management Plan che fornisca un quadro di riferimento per una gestione responsabile dei dati della ricerca.⁴

Infrastruttura

Per curare i dati della propria ricerca, i ricercatori e le organizzazioni attive nella ricerca devono poter accedere agli appositi ecosistemi digitali; questi sistemi possono essere mantenuti a livello locale; in alternativa è possibile ricorrere a servizi commerciali, a offerte specifiche a seconda del settore, oppure a piattaforme regionali/nazionali/internazionali. Le diverse comunità scientifiche e/o disciplinari e i singoli Paesi adotteranno soluzioni differenti per rendere fruibili tali strutture. Di norma, la piattaforma, o le piattaforme dovranno essere in grado di offrire i seguenti servizi:

- archiviazione dei dati prodotti e raccolti dai ricercatori;
- una piattaforma di pubblicazione che consenta la condivisione e riuso dei dati della ricerca e relativi software;
- funzioni di archiviazione, affinché i dati della ricerca possano essere conservati nel lungo termine, spesso in risposta ai requisiti imposti dai finanziatori della ricerca;
- un *discovery service* (servizio di ricerca) che consenta ai ricercatori e alla comunità, di esplorare gli archivi dei dati della ricerca sia localmente che attraverso Internet.

La Commissione Europea ha deciso di promuovere la European Open Science Cloud.⁵ L’EOSC è una metafora ideata per trasmettere sia il concetto di un’integrazione di sistemi in modo continuo, che l’idea di un bene comune basato su dati scientifici. L’EOSC sarà un ambiente federato per la condivisione e il riutilizzo di dati scientifici, basato su elementi esistenti ed emergenti negli Stati membri, con un minimo di dirigenza e di forme di *governance* internazionali e un massimo di libertà in termini di attuazione pratica.

³ Cfr. <https://www.force11.org/group/fairgroup/fairprinciples>; ultimo accesso: 12/12/16.

⁴ Per ulteriori informazioni: <http://www.dcc.ac.uk/resources/data-management-plans>; ultimo accesso 12/12/16.

⁵ Cfr. <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; ultimo accesso 12/12/16.

Formazione

Per un'adeguata diffusione dei dati della ricerca è imprescindibile che tutti i ricercatori, dalle nuove leve a quelli già affermati, si dotino delle competenze e degli strumenti necessari per muoversi con sicurezza in un ambiente caratterizzato da dati. Un ruolo di capofila in questo senso dovrebbe essere svolto dalle organizzazioni attive nella ricerca e, in molti casi, dalle biblioteche istituzionali ivi presenti.

Finanziamento

La gestione dei dati della ricerca comporta costi. Non esiste un unico metodo specifico per valutarli, bensì una serie di preziosi modelli di calcolo, ad esempio il progetto 4C.⁶

Conclusioni

I dati della ricerca possono dare una spinta verso l'innovazione e offrire lo spunto verso nuove scoperte, a grande beneficio della società. Tutti coloro che sono coinvolti nei processi scientifici sono chiamati a svolgere la propria parte. Il presente Executive Briefing mette in evidenza ciò che i ricercatori e le istituzioni attive nella ricerca sono invitati a fare per raccogliere questa sfida coinvolgente e stimolante.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁶ Per ulteriori informazioni: <http://www.4cproject.eu/summary-of-cost-models>; ultimo accesso 12/12/16.

