Davidson, J., Jones, S., and Molloy, L. (2014) Big data: the potential role of research data management and research data registries. In: *IFLA World Library and Information Congress 80th IFLA General Conference and Assembly* , Lyon, France, 16-22 August 2014.

http://eprints.gla.ac.uk/102160/

Deposited on: 3 February 2015

Big data: the potential role of research data management and research data registries

Joy Davidson
Digital Curation Centre, University of Glasgow, Glasgow, United Kingdom
joy.davidson@glasgow.ac.uk

Sarah Jones
Digital Curation Centre, University of Glasgow, Glasgow, United Kingdom
sarah.jones@glasgow.ac.uk

Laura Molloy
Digital Curation Centre, University of Glasgow, Glasgow, United Kingdom
laura.molloy@glasgow.ac.uk

Abstract:

Universities generate and hold increasingly vast quantities of research data – both in the form of large, well-structured datasets but more often in the form of a long tail of small, distributed datasets which collectively amount to 'Big Data' and offer significant potential for reuse. However, unlike big data, these collections of small data are often less well curated and are usually very difficult to find thereby reducing their potential reuse value. The Digital Curation Centre (DCC) works to support UK universities to better manage and expose their research data so that its full value may be realised. With a focus on tapping into this long tail of small data, this presentation will cover two main DCC, services: DMPonline which helps researchers to identify potentially valuable research data and to plan for its longer-term retention and reuse; and the UK pilot research data registry and discovery service (RDRDS) which will help to ensure that research data produced in UK HEIs can be found, understood, and reused.

Initially we will introduce participants to the role of data management planning to open up dialogue between researchers and library services to ensure potentially valuable research data are managed appropriately and made available for reuse where feasible. DMPs provide institutions with valuable insights into the scale of their data holdings, highlight any ethical and legal requirements that need to be met, and enable planning for dissemination and reuse. We will also introduce the DCC's DMPonline, a tool to help researchers write DMPs, which can be customised by institutions and integrated with other systems to simplify and enhance the management and reuse of data.

In the second part of the presentation we will focus on making selected research data more visible for reuse and explore the potential value of local and national research data registries. In particular we will highlight the Jisc-funded RDRDS pilot to establish a UK national service that aggregates metadata relating to data collections held in research institutions and subject data centres. The session will conclude by exploring some of the opportunities we may collaboratively explore in facilitating the management, aggregation and reuse of research data.

1. Universities and research data management

The Jisc-funded Digital Curation Centre (DCC) supports UK universities to better manage and exploit their research data. Over the past few years we have engaged with over 90% of the research-active universities in the UK through our roadshow series. In parallel we have run a programme of institutional engagements (1) in which we worked on a one-to-one basis with UK HEIs to help them improve their research data management (RDM) practices. As a result, the DCC has gained valuable insights into the requirements and maturity of the sector as a whole. The Institutional Engagement programme allowed us to identify RDM needs and to assist institutions in developing policies and support services in response.

Many of the research intensive universities we support are involved in large, international collaborations, very often producing what might be called 'Big Data'. These well resourced collaborations tend to pose less of a concern for universities, as the projects are of such a scale that they have dedicated facilities, staffing and support systems in place. The more pressing challenge for universities relates to their responsibility for all that remains - a large collection of small, heterogeneous datasets with a divergent set of RDM requirements. Researchers are not always aware of good practice in RDM or the support provided by their institutions and may generate their data in a relatively undisciplined manner. This can make data difficult to locate - even in the short term - and poses a real risk that published research findings cannot be validated if questions should arise. Moreover, user requirements for data deposit and dissemination are diverse and complex, as shown in the six use cases for Edinburgh DataShare (2). Handling such a disparate range of outputs and needs can be challenging. The scale of data being produced means that it is often difficult to get a handle on what is held, where it is held, and by whom. Enabling data reuse is also a major challenge as there is often no central catalogue of data holdings with clear definitions of locations, access rights and restrictions. The remainder of this paper will focus on two relevant tools to support universities with these challenges - DMPonline and the Research Data Registry and Discovery Service (RDRDS).

DMPonline (3) is the DCC's web-based tool to help researchers write data management plans (DMP). It includes a number of templates that meet specific funder requirements, and can be customised by institutions to provide tailored guidance and details of local support. Many UK universities encourage researchers to write DMPs for all research activity – not just that which is externally funded - and are attempting to build data management planning into existing workflows (e.g. the grant application process). To simplify the workflow and avoid duplication of effort, several UK HEIs are aiming to integrate DMPonline with other university systems, such as Current Research Information Systems (CRIS), so that relevant data are shared between them and do not have to be re-entered.

While DMPs help institutions to get a sense of the data being created on current research projects, tools to aid the discovery of data post-project are also needed. Many UK universities

are developing catalogues or inventories of data holdings to facilitate reuse. In parallel, the DCC is running a pilot project to harvest metadata from these catalogues and collate it into a national registry - the Jisc Research Data Registry and Discovery Service (RDRDS) (4). The Australian National Data Service (ANDS) software is being used for the pilot. Research Data Australia is 'an Internet-based discovery service designed to provide rich connections between data, projects, researchers and institutions, and promote visibility of Australian research data collections in search engines' (5). Use of DMPonline and the RDRDS may encourage universities to collate intelligence on their research data holdings and open up their rich collections for reuse.

2. Data Management Planning

Data management planning is a topic of increasing importance to researchers, their host institutions, publishers and research funders alike. Indeed, many research funders now require a data management plan as part of grant proposals, and several UK universities have also mandated them for all research activity undertaken by staff. The DCC has led the development of guidance and resources to support data management planning since 2010. We initially compared funders' data policies, examining requirements for DMPs and using this analysis authored the Checklist for a Data Management Plan (6). We also produced the first online tool to assist in the data management planning process - DMPonline.

Data management planning in the context of big science has been addressed admirably by the MaRDI-Gross project (7). While it may seem flippant to suggest that reasonable advice is simply to 'read and implement the OAIS specification', the nature of big science is sufficiently different to make data management planning a question of defining and formalising existing practice, since much of the infrastructure, financial support and technical expertise is already in place. Data management planning is a far greater challenge – and subsequently provides most benefit – in the context of small science. For lone researchers or less well-resourced teams, there is far greater likelihood of gaps in RDM expertise or responsibility falling to an overworked research assistant or PhD student. Ad hoc and ill-informed practices can result, exacerbating the challenges universities face when they are required to preserve and maintain access to this content.

The process of writing DMPs is beneficial in a number of ways. It helps researchers to make informed decisions, to pre-empt challenges, to mitigate risks such as data loss, and to identify additional costs that should be considered. By considering issues such as where to store data, who will perform back-up, and how data security and access rights will be managed, researchers can identify gaps in support provision and skillsets. Planning at the outset encourages an early dialogue about requirements with university support services and means that appropriate provision can be costed in to grant applications and provided.

Some universities are proactively identifying common issues and setting triggers to alert library and IT services to researchers' needs for support. At the University of Leicester, for example, RDM questions have been built into their grants costing system. Researchers are asked whether they will be creating sensitive data or large volumes of data. Affirmative answers are flagged and trigger an automated email to the research liaison team. The discussions between the researcher and the research liaison team that follow help to ensure appropriate plans are in place to manage data through the active phase of research to post-project dissemination and potential reuse, as noted in the diagram below.
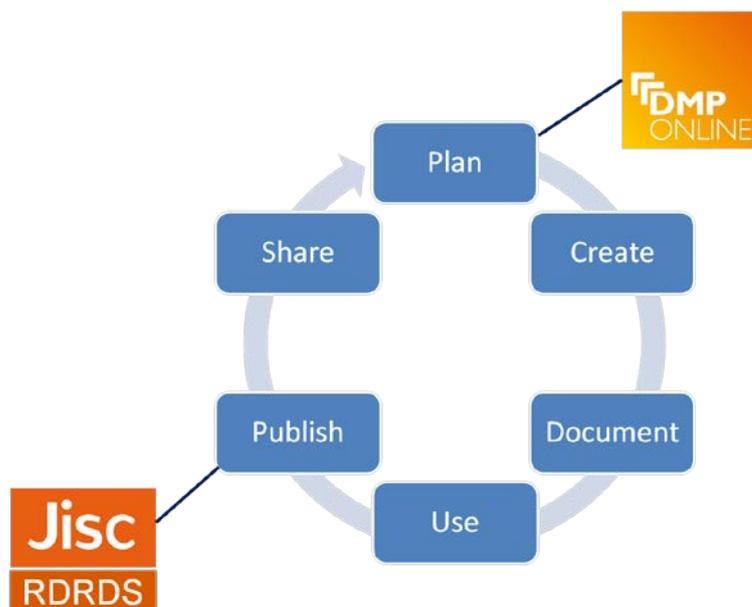
Figure 1: a typical research lifecycle, showing where the DCC's DMPonline and Research Data Registry and Discovery Service tools provide support.

Data management planning helps to identify potential barriers to data sharing, such as confidentiality agreements and opportunities to commercialise the research. Risks can be highlighted early on so they can be mitigated appropriately, for example by applying embargo periods or restricting access to the data to an approved group of users. Universities may also seek to mine and exploit the data being collected via their DMPs to help them plan for improvements to their own RDM services, as has been done by a few universities in the United States (8, 9).
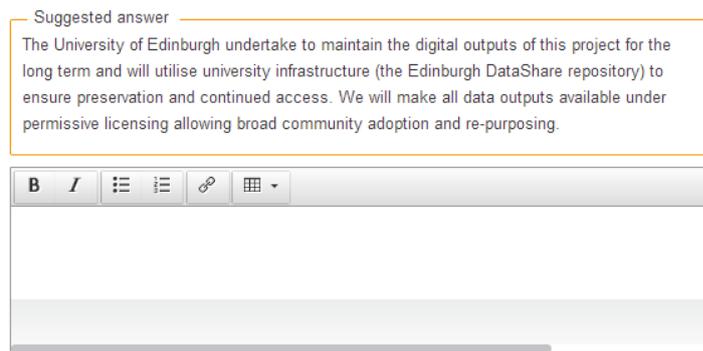
3. DMPonline

DMPonline is a web-based tool to help researchers and research support staff to produce data management and sharing plans that meet funders' mandates. It was first demonstrated at the Jisc conference in London in 2010 and has since undergone regular updates and some major redevelopment. The tool has received international recognition, and was shortlisted along with the work of US colleagues on DMPTool for the DPC's Digital Preservation Awards at the end of 2012. A major new version of DMPonline was released in autumn 2013. This allows far greater flexibility in terms of adding templates, custom guidance, examples and suggested answers. Demand for customisations and foreign language versions is growing rapidly, particularly across Europe in light of the Horizon 2020 data management planning requirements for their Open Data Pilot (10).

There are a number of ways in which institutions can customise DMPonline to make it more useful in their local context. They can create their own templates to provide questions and guidance that researchers should respond to, add tailored guidance to help researchers answer funders' questions, and provide examples and suggested answers. We are also adding new features to allow institutions to brand the tool, see statistics on use, and potentially a DMP review function.

Figure 1: a screengrab of an institutional customisation in DMPonline, showing tailored guidance and example answers.

As use of DMPonline grows, we are gathering valuable feedback to help us to define a clear plan of action and prioritise which new features are most needed. One of the priority areas of work is to add features that integrate DMPonline with other systems and underpin use of the DMP as a living document to aid communication throughout the project. Some feature requests being considered to do this are flagging the status of plans, allows comments inside the system and pushing data out to other services such as registries. We use GitHub so others can download and contribute to the code, and are currently working with institutions to integrate DMPonline with their local systems. Based on the number of comments and suggestions we receive through GitHub, there is clear evidence that there is a community of users who want to help develop the tool further. We intend to make the most of user input and plan to establish a DMPonline community group to allow external parties to help steer the future direction we take with the ongoing development of the tool.

DMPonline – and the act of data management planning in general – helps to ensure that even the smallest amount of research data can be identified at the earliest stages of the research process as having potential value. This allows institutions to ensure adequate resources are provided to curate the data over time to maximise its chances of discovery and reuse.

4. The Jisc Research Data Registry and Discovery Service (RDRDS)

In order to realise the full value that the exploitation of well-managed research data can deliver to researchers and to funders, it is crucial to plan for its retention and curation.

> 'Ensuring that data is curated and archived correctly, with appropriate descriptions and metadata about the source and method of collection, will allow data to be reused most effectively. It will also ensure that data can be reused with confidence in the analysis and that conclusions can be drawn from data which someone else has collected.' (11)

In this respect, UK higher education and research have undertaken much activity over the last few years to improve RDM support provision and practice. Much of this activity was funded through the Jisc Managing Research Data programmes (12) and supported by the DCC

Institutional Engagements programme. While RDM practice is improving, without adequate discovery mechanisms even well-managed data may go unused. Discovery alone is not enough. End users will need to be able to make decisions about the data they find and whether or not it is suitable for their research. This is highlighted in the recent G8 Open Data Charter. This states under 'Principle 2: Quality and Quantity' that it will:

> 'make sure that data are fully described, so that consumers have sufficient information to understand their strengths, weaknesses, analytical limitations, and security requirements, as well as how to process the data' (13)

Many funding bodies support subject-specific data centres to ensure that research data can be collected and made visible for reuse. A core function of subject-specific data centres is to focus on standardised and high quality description of each dataset to help reusers to make informed judgements about its relevance and potential value. However, the number of sustainably-funded, subject-specific data centres is relatively small and there are many disciplines without access to an appropriate subject-specific data centre. Increasingly, the task of describing and making research data visible is falling to university libraries and repository staff. A number of UK universities are now developing institutional data repositories which can provide preservation and exposure for this research data. Whilst this approach supports an overall improvement in the sustainable curation of research data, it may also increase the number and size of separate silos of research data which is detrimental to the overall notion of discoverability. In addition, the quantity and quality of the metadata captured by institutions varies and there is a role for international bodies to develop standards and to provide support and guidance on good practice in this area.

To exploit the full potential of this emerging mixed data repository landscape, Jisc has invested in the development of a pilot UK Research Data Registry and Discovery Service (RDRDS). The RDRDS aims to increase the discoverability of research datasets arising from UK research by exposing it at a national level. The RDRDS will not hold data, judge the quality of datasets or dictate what is exposed: rather, it works in partnership with a network of UK subject-specific data centres and university-based institutional data repositories to harvest dataset metadata records, and so promote the discoverability of research data held by all partner institutions. Partner institutions will remain responsible for the selection and stewardship of the datasets. At the time of writing there are nine partner universities and six subject-specific data centres. The data centre members are currently all RCUK-funded, but membership will expand in the second phase of RDRDS pilot activity (due to start in 2014) and it is hoped that other subject-specific data centres and databases will become active participants, as well as a larger set of UK universities.

Technical development of the first pilot instance of the service followed the lead of the Research Data Australia portal described earlier. A new implementation of the open software platform was used for the first phase of pilot activity, and comparative work with alternative software will be explored in the upcoming second pilot phase. Research activity is frequently international in nature: the current pilot work cannot accommodate details of all research data that may potentially be of interest to UK researchers, desirable though that may be. Rather, it aims to develop a means by which the metadata describing research datasets held by UK research institutions is harvested and exposed both by a dedicated RDRDS portal and also by promotion to generic search engines. Ideally, the UK RDRDS and Research Data Australia and will be joined by similar mechanisms from other national contexts, to constitute a global

network of discoverable research data. This global network will expose and promote the reuse of the massive datasets resulting from 'big science' research and hosted at long-established, well-funded national data centres. However, alongside them will be the metadata of the 'long tail' of smaller datasets resulting from individual researchers, small research teams and low- or no-budget research work at universities and research institutions.

5. Short-term opportunities for ongoing research and development

While there has been substantial progress in developing procedures, tools, and support for research data management and sharing there is still a long way to go. Below, we outline a few of the areas where we feel there is the greatest potential for collaboration in the near future to realise benefits.

1) **Harvesting metadata from CRIS to institutional data catalogues**
   EuroCRIS (14) has been working to refine the CERIF data model to better handle research data information within current research information systems. We recommend that further efforts are put into refining and agreeing standards for integration between CRIS and institutional data catalogues. Not only will this help to reduce duplication of effort, but it would also help to improve consistency and quality of the metadata records in both systems. CASRAI (15) - the Consortia Advancing Standards in Research Administration Information - may also have a role to play in developing profiles for sharing research information between institutional and external systems.

2) **Include equipment data into data management plans and institutional catalogue records**
   We believe that the equipment used to capture and process research data forms an integral part of its provenance. In this respect, we are keen to integrate aspects of equipment data into DMPonline templates and to further explore potential synergies with the Equipment.data group (16) being led by the University of Southampton. Information about data capture and processing equipment may also be extremely valuable to expose through institutional data catalogues and possibly, as a result, through RDRDS. With this approach, it might someday be possible to search for research data outputs by data capture or processing type as well as by topic.

3) **Deposit DMPs into IRs and ensure that a link is included in the metadata**
   Currently, DMPs tend to be submitted as part of grant applications and there is little evidence that they are updated over the life of the project or shared beyond the project partners and funders. We feel that this misses opportunities. DMPs provide evidence of how the data was handled during the active phase of the research and could be crucial for helping secondary users to assess the quality and trustworthiness of the data they are considering using. We recommend that researchers consider making their updated DMPS available at the end of the project through their institutional repository or – perhaps more attractively – by publishing as a data paper. The Ubiquity Press metajournals (17) offer an ever-growing range of discipline-specific data journals where data papers (essentially public versions of DMPs) are published as scholarly outputs in their own right. The data papers are peer reviewed to 'highlight openly archived data with high reuse potential, and provide recognition for the producers of the data'. DCC and Ubiquity Press have had initial discussions about enabling the export of plans within DMPonline to the data paper templates used by

Ubiquity Press. This could be a valuable incentive for researchers to update and share their data management plans.

4) **Use of RDRDS by institutions to gather usage statistics**
There is potential for the Research Data and Discovery Service (RDRDS) to evolve to supplement the information available to institutions in the periodic gathering of statistics on data usage. Use of the RDRDS in this way may help institutions when assessing research impact during future research excellence framework (REF) exercises in the UK and could be useful in the re-appraisal of research data once retention periods have expired. Indeed, it is likely that the re-appraisal of research data held in institutional repositories may be a labour-intensive process and is an activity that is often omitted from current RDM support services. There is a need for agreed standards and metrics to help assess the ongoing value of research data. Statistics made available by national level discovery services such as RDRDS could conceivably supplement other measures taken by data repositories in supporting automated approaches to assessment and re-appraisal thereby making it more efficient.

5) **Use of RDRDS by funders**
There is potential value in the Research Data and Discovery Service (RDRDS) being used by research funders to help ensure that data outputs have been shared as outlined in their data management and sharing policies. To this end, ongoing communication and collaboration between national services like RCUK's Gateway to Research (18), ResearchFish (19) and RDRDS is valuable. We anticipate that this will be a key aspect for the second phase of RDRDS pilot work. The Gateway to Research aims to enable the public and particularly SMEs to search for and analyse information on publicly funded research activity. ResearchFish is the system that RCUK uses to gather, monitor and track information on publicly-funded research outputs. In addition to the potential value to funders in terms of the practicalities of monitoring compliance, services such as RDRDS could help them to review the impact their funding has had and to help identify trends and gaps where future funding may be needed.

6) **Use of RDRDS by researchers**
Just as HEIs and funders may use national data registry and discovery services to supplement their efforts in tracing research impact at a funder or institutional level, researchers may also wish to make use of such resources to help track use of their outputs. In this respect, embedding researcher identifiers such as ORCID (20) into DMPs, institutional data catalogues and research information systems will be helpful.

6. Longer-term challenges

Significant effort has gone into identifying data that may have longer-term value and to make that data more visible for reuse. This in itself is a huge step forward to realising the potential of big data – whether it is large, structured datasets or the long tail of smaller, disparate research datasets. However, to fully tap into this potential, there are still some major challenges that need to be addressed by the community.

1) **Providing tools to analyse and manipulate the data**

Much of the data held within institutional repositories will require specialist software to render, analyse and manipulate the data. To ensure that data reuse is a seamless as possible, we would benefit from the choice of being able to either move data to the tools, or to move analysis tools to the data – this latter, most likely in cloud based services. Some subject-specific data centres provide access to web-based tools and services but more development in this area would be welcome; for instance, the Natural Environment Research Council's British Atmospheric Data Centre provides access to the CEDA Web Processing Service (21). It is likely that the provision of specialist tools and services will not be feasible for each institutional data repository service due to the costs and complexities around software licensing. In the short term, there may be scope for regional or disciplinary cooperation between HEIs to share services and costs. Over the longer term, it will be necessary for the curation community (researchers, HEIs, funders, archives, libraries, publishers and vendors) to work towards reaching agreement on developing models for sustainable licenses for archived data reuse.

2) **Common agreement on attribution stacking**
There have been ongoing efforts to develop guidance for data citation which is helping to standardise approaches. For example, the DCC released a 'How-to' guide on data citation (22) in 2012 and Force11 released their 2014 Joint Declaration on Data Citation Principals (23). However, as researchers mash up datasets from multiple sources, additional guidance and support will be needed to help researchers to accurately cite these derived resources. For the data reuse culture to grow, the attribution stacking problem must be addressed; it may be helpful to develop mechanisms to automate the process. DMPs could usefully include information on rights and citation for the benefit of secondary users. Building in automated attribution services into data repositories, which can be reflected in national research data registries and discovery services, may be an area worth exploring as data discovery and reuse services become more widely used.

7. Conclusions

There is great potential to collaborate in the short term to make the long tail of research data easier to manage, to find, to assess and to reuse so it is as valuable as larger, more structured 'Big Data'. Data management planning can help institutions to get a sense of the research data being created on current research projects and to highlight data with potential value. Local and national tools to aid the discovery of research data post-project are also helping to make sure that data – of any size – is more visible and can be found and assessed for reuse. Use of DMPonline and the UK RDRDS allows universities to collate intelligence on their research data holdings and open up their rich collections for reuse. The current activities and potential future opportunities described in this paper complement each other in making the description and discoverability of research datasets of all disciplines and sizes more efficient and reliable. However, as noted some approaches must still be experimental, and there is still much to be done.

of DMPonline, DCC Institutional Engagements, and the UK pilot research data registry and discovery service, as well as Jisc for funding these initiatives.

Author Biographies
Joy Davidson is Associate Director of the Jisc funded Digital Curation Centre (DCC). Joy was also currently PI for the Jisc funded Data Management Skills Support Initiative – Assessment, Benchmarking and Classification (DaMSSI-ABC) project and is also a core member of the EC-funded Collaboration to Clarify the Costs of Curation (4C) and the Facilitate Open Science Training for European Research (FOSTER) projects. She is involved in several international working groups including the Knowledge Exchange Primary Research Data Working Group, Research Data Canada (RDC) Education & Training Subcommittee, and the RIN Informal digital literacies group.

Sarah Jones is a Senior Information Support Officer with the Digital Curation Centre (DCC), a UK national service providing support to the higher education sector in all aspects of research data management. Since 2011 her principal focus has been on the DCC's institutional engagement programme, in which she has been leading the provision of support to a range of universities. This involves advising on RDM strategy and assisting them to deliver a range of support services. Sarah coordinates work on DMPonline and undertakes research on data policy and data management planning.

Laura Molloy is a researcher at the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow, working in the areas of digital curation, digital preservation and research data management. A member of DCC staff, she has experience in the design and delivery of training and skills development for researchers and support professionals, including work on the Jisc funded Incremental project and the Data Management Skills Support Initiative (DaMSSI). She was one of three Evidence Gatherers across the Jisc Managing Research Data programme. EU-funded work includes the delivery of outreach and training events for the Planets project, online outreach materials for Digital Preservation Europe (DPE) and development of a curriculum framework for digital curation in the cultural heritage sector for the DigCurV project. She currently coordinates the pilot UK RDRDS initiative.

References
1) DCC Institutional Engagements: http://www.dcc.ac.uk/tailored-support/institutional-engagements
2) Rice, R. (2013). *Six Use Cases for Edinburgh DataShare.* Presentation given at the Repository Fringe Conference, 31 July – 2 August 2013, Edinburgh. Retrieved from http://www.slideshare.net/rcrice/rice-repofringe2013
3) DMPonline: https://dmponline.dcc.ac.uk
4) Jisc Research Data Registry and Discovery Service: first pilot phase is described at http://www.dcc.ac.uk/projects/research-data-registry-pilot

5) ANDS Research Data Australia:https://researchdata.ands.org.au/home/about

6) DCC. (2013). *Checklist for a Data Management Plan*. v.4.0. Edinburgh: Digital Curation Centre.

7) Gray, N., Carozzi, T. D. & Woan, G. (2012). *Managing research data in big science*. LIGO Project Report P1000188. University of Glasgow. Retrieved from http://arxiv.org/abs/1207.3923

8) Nicholls, N. (2014). *Establishing Services and Resources to support Faculty Writing Data Management Plans: Lessons Learned from an Engineering Pilot*. Presentation given at the International Digital Curation Conference, 24-27 February 2014, San Francisco. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/IDCC14/Parallels/GreenNicholls_ParallelA1.pdf

9) Wells Parham, S., & Doty, C. (2012). 'NSF DMP Content Analysis: What Are Researchers Saying' in *Bulletin of the American Society for Information Science and Technology.* Volume 39, Number 1. Retrieved from https://smartech.gatech.edu/bitstream/handle/1853/48707/OctNov12_Parham_Doty.pdf?sequence=1

10) European Commission (2013). 'Commission launches pilot to open up publicly-funded research data' (webpage). Retrieved from http://europa.eu/rapid/press-release_IP-13-1257_en.htm

11) Department for Business, Innovation and Skills (2013). Seizing the data opportunity: a strategy for UK data capability. Retrieved from https://www.gov.uk/government/publications/uk-data-capability-strategy

12) Managing Research Data (JiscMRD) programme 2009-11: http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx and Jisc Managing Research Data Programme 2011-13 (MRD02): http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

13) Cabinet Office (2013) *G8 Open Data Charter and Technical Annex*. Retrieved from https://www.gov.uk/government/publications/open-data-charter

14) EuroCRIS: http://www.eurocris.org/Index.php?page=homepage&t=1

15) CASRAI: http://casrai.org/

16) Equipment.data: http://equipment.data.ac.uk

17) Ubiquity Press Metajournals: http://www.metajnl.com

18) RCUK Gateway to Research: http://gtr.rcuk.ac.uk

19) ResearchFish: https://www.researchfish.com

20) ORCID: http://orcid.org/

21) CEDA Web Processing Service (WPS): http://ceda-wps2.badc.rl.ac.uk/submit/choose

22) Ball, A. (2012). 'How to License Research Data'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/how-guides/license-research-data

23) Force 11 Joint Declaration of Data Citation Principles https://www.force11.org/datacitation