Open Data Policy Case Study

Open Data Policy in the UK, August 2017

1.   Introduction

Over the last ten years, the UK data policy landscape has been shaped by the overarching drive to open-up data and maximise the impact and reach of research. This is part of a wider movement called open science, which declares that data is a public good and as such should be openly available where possible.

The development of data policies alongside other initiatives, for example the recent FAIR data principles[1], continue to bring much needed clarity around the concepts of data management and sharing, and help to maximise the value from research investments.
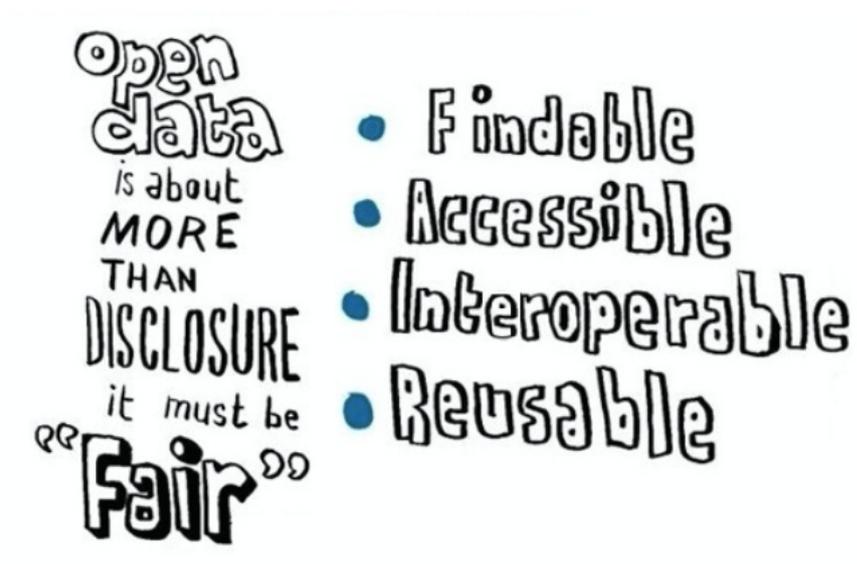

Image: FAIR data principles. Source: illustration CC-BY Scientific Data, Nature Publishing Group

2.   The UK Concordat on Open Research Data

The role of data policy is to advance data availability, data archiving and data management while emphasizing the importance of data as a core output of public research[2]. The launch of the Concordat on Open Research Data[3] in July 2016 by four of the UK's leading research organisations (RCUK, HEFCE, Universities UK and Wellcome Trust) under the patronage of the UK Open Research Data Forum, is a key development in the harmonisation of UK research data policy.

---

[1] FORCE11. (2016). The FAIR DATA Principles, FORCE11, https://www.force11.org/group/fairgroup/fairprinciples
[2] Neylon, C. (2011, November 10). Reflections on research data management [blog], http://www.dcc.ac.uk/news/reflections-research-data-management
[3] RCUK (2016). *Concordat on Open Research Data*, Research Councils UK, http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf/

The concordat proposes a set of principles for working with research data including the different roles and responsibilities needed to support the research process. Rather than being a rulebook of specific behaviours and activities, it offers a set of expectations of good practice developed by the research community itself. These principles highlight the underlying need to establish open research data as the desired outcome for publicly-funded research:

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

The Concordat in its current form outlines the responsibilities of researchers, their employers and funders, and complements existing guidelines and frameworks, for example the 2011 RCUK Common Principles. The public consultation on the draft concordat, which included universities, research groups, individuals and various organisations across the sector, highlighted some well-known issues imperative to those tasked with implementing data policy:

- Avoiding bias in policy terminology towards STEM subjects.
- Defining digital versus non-digital data, data, code and research articles.
- Recognising the trade-offs between the risks and benefits of sharing data.
- Rewarding the effort put into managing data.
- Addressing the cost of providing services and infrastructure.

3.   Policy Development in the UK

These requirements undoubtedly lead us a step closer toward harmonization across data policies in the UK and toward a more coordinated investment in data infrastructure at a national level. This is a process that has been gathering pace for more than a decade now, and given how interdisciplinary and collaborative research is, has been closely aligned to and influenced by international policy moves (see History of major developments in data policy in Figure 1).
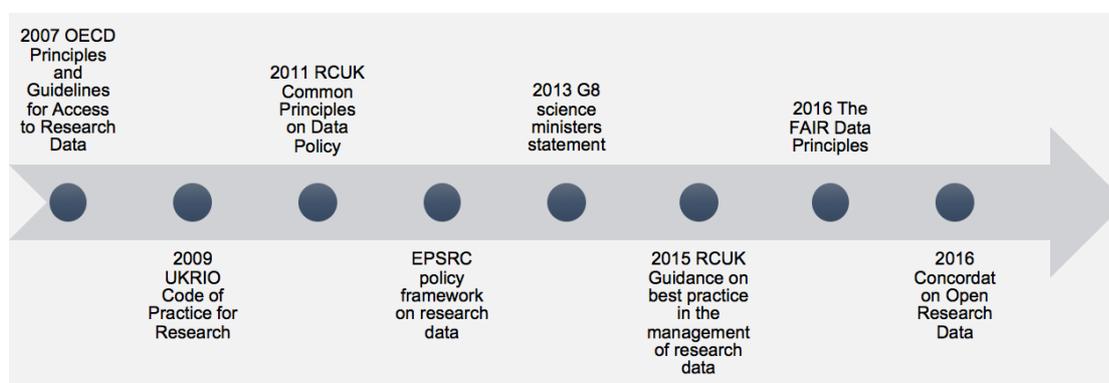

Figure 1: History of major developments in data policy

In 2007, the 30 member countries from the Organisation for Economic Co-operation and Development (OECD) issued Principles and Guidelines for Access to Research Data from Public Funding[4]. Paramount to the 13 principles listed in the document was the notion that publicly funded research data was a public good produced in the public interest and therefore openly available to all. In response the UK coordinating body, the Research Councils UK (RCUK), have issued similar codes and principles in the form of Expectations for Societal and Economic Impact[5] in 2008 and the Policy and Code of Conduct on the Governance of Good Research Conduct[6] the following year. (The latter was replaced by the RCUK Policy and Guidelines on Governance of Good Research Conduct). Similar requirements were also proposed in the Code of Practice for Research issued by the UK Research Integrity Office (UKRIO)[7] in 2009. In July 2015, RCUK released further Guidance on best practice in the management of research data[8] adding an explanatory text for each of the seven 'common principles'. This guidance informed the RCUK consultation on the aforementioned draft Concordat on Open Research Data[9].

Another key development for data policy in the UK was the release by the EPSRC (the Engineering and Physical Sciences Research Council) its policy framework on research data[10]. The policy set out EPSRC's principles and expectations concerning how the institutions would ensure that research data was appropriately managed and shared, placing the duty, unlike earlier data policies, on the research organisations to provide appropriate

---

[4] OECD (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Organisation for Economic Co-operation and Development, https://www.oecd.org/sti/sci-tech/38500813.pdf

[5] RCUK (2008). *Expectations for Societal and Economic Impact*, Research Councils UK, http://www.rcuk.ac.uk/Publications/archive/StatementofExpectationon/

[6] RCUK (2009). *Policy and Code of Conduct on the Governance of Good Research Conduct*, Research Councils UK, http://www.rcuk.ac.uk/publications/researchers/grc/

[7] UKRIO (2009). *Code of Practice for Research: promoting good practice and preventing misconduct*, UK Research Integrity Office, http://ukrio.org/publications/code-of-practice-for-research/

[8] RCUK (2015). *Guidance on best practice in the management of research data*, Research Councils UK, http://www.rcuk.ac.uk/documents/documents/rcukcommonprinciplesondatapolicy-pdf/

[9] RCUK (2015). *Concordat on Open Research Data*, Research Councils UK, http://www.rcuk.ac.uk/media/news/160728/

[10] EPSRC (2011). *EPSRC policy framework on research data*, Engineering and Physical Sciences Research Council, https://www.epsrc.ac.uk/about/standards/researchdata/

policies and infrastructure, and raise awareness of researchers' responsibilities regarding data management and sharing. The eight expectations contained in the framework arose from seven core principles, which aligned with the core RCUK principles on data sharing and generated considerable attention and debate in part due to two deadlines imposed on institutions. The first requirement was to have a clear roadmap in place to align policies and processes with the EPSRC's expectations by May 2012; this then had to be implemented to ensure full compliance by May 2015. The failure to comply, it was declared, could lead to the denial/withdrawal of funding. The EPSRC mandate has had a major impact on the policy landscape and is often cited a key driver for institutional data policy, which, in turn, has led to changing practices towards research data across the universities.

Despite the various RCUK policies and guidelines released in this period, a more coordinated approach was still needed. Each of the seven Research Councils had released their own policies too, and the EPSRC approach to place the onus on institutions, led to a much more compliance focused culture, with institutions checking they met the criteria of all key funders. While the basic principles of these policies were comparable, the specific requirements varied. AHRC, for example, ask for content to be kept accessible for 3 years, several funders state a period of 10 years after the completion of a research project, while the EPSRC expects data to be securely preserved for a minimum of 10 years from the date that any privileged access period expires, or from the last date on which access to the data was requested by a third party[11]. The lack of coherence and consistency was placing an additional strain on universities. Mapping across the different positions in each Research Council policy to identify a baseline to work from led to several calls for more harmonisation.

The significant development came in April 2011 with the introduction of the RCUK's Common Principles on Data Policy[12], which aimed to provide an overarching framework for individual Research Council policies on data policy, and to a certain extent succeeded in bringing in more harmonisation across the research councils. These reiterated the OECD declaration about the importance of data as a public good and highlighted other important aspects such as discoverability of data; legal, ethical and commercial constraints on data release; data citation; the right of first use for data creators; and finally, the cost-effective methods for data management and sharing[13]. This has now been taken further with the Concordat.

The wider international policy landscape, articulated by bodies such as EU, OECD or the G8 Science Ministers, will continue to influence the UK data policy in the short term. A degree of uncertainty however is to be expected following the June 2016 Brexit referendum. On a positive note the 'direction of travel' towards greater openness, continued strengthening of research data policy towards open data and support for open innovation is unlikely to change[14]. Research data policy took also a prominent place in recent G8 Science Ministers' statements[15]:

- To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, and acknowledging the legitimate concerns of private partners.

---

[11] Jones, S. (2008, rev. 2011) *Summary of UK Research Funders' Expectations for the Content of Data Management and Sharing Plans*, http://www.dcc.ac.uk/sites/default/files/documents/resource/policy/FundersDataPlanReqs_v4%204.pdf

[12] RCUK (2011). *RCUK Common Principles on Data Policy*, Research Councils UK, http://www.rcuk.ac.uk/research/datapolicy/

[13] Jones, S. (2012). Research data policies: principles, requirements and trends in Pryor, G., ed., *Managing Research Data*. London: Facet Publishing.

[14] DCC (2016). *Common Directions in Research Data Policy*, Digital Curation Centre, http://www.dcc.ac.uk/resources/briefing-papers/common-directions-research-data-policy/common-directions-research-data-pol

[15] G8 science ministers statement: London, 12 June 2013, https://www.gov.uk/government/publications/g8-science-ministers-statement-london-12-june-2013

- Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.
- To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including recognition of researchers fulfilling these principles, and appropriate digital infrastructure.

This suggests not only a commitment to an 'open by default' position on data, but highlights a need for more investments into infrastructure, both technological and organisational.

Elsewhere, a similar nation-wide approach can be observed in the Netherlands where the importance of open data to advance and accelerate research is addressed in the Nationaal Plan Open Science[16] policy. The document lists the key ambitions and challenges as being:

- To provide full open access to publications in 2020.
- To make research data optimally suited for reuse.
- To incorporate recognition and rewards into the broader research landscape.
- To establish a 'clearing house' for all information regarding available research support.

The policy, led by the Dutch government's Ministry of Education, Culture and Science, is considered an example of a 'softer'[17] approach to policy making; encouraging and inspiring in tone, rather than mandating compliance.

In parallel, disciplinary concordats have been also emerging. In public health research, the World Health Organization (WHO) and the Wellcome Trust have in 2008 initiated discussions about the common Code of Conduct on the sharing of data of public health importance, which later led to the joint statement of purpose[18] issued in 2011 by the Wellcome Trust and other UK funders of public health research, which stated:

We, as funders of health research, intend to work together to increase the availability to the scientific community of the research data we fund that is collected from populations for the purpose of health research, and to promote the efficient use of those data to accelerate improvements in public health.

The statement expressed both immediate and longer-term aspirations, and called for access to and use of data, on condition that it should be equitable, ethical and efficient.

---

[16] TUDelft (2017). *Nationaal Plan Open Science*, TU Delft Library, doi:10.4233/uuid:9e9fa82e-06c1-4d0d-9e20-5620259a6c65

[17] SPARCEurope (2017). *An Analysis of Open Data and Open Science Policies in Europe*, SPARC Europe, http://sparceurope.org/new-sparc-europe-report-analyses-open-data-open-science-policies-europe/

[18] Wellcome Trust (2011). *Sharing research data to improve public health: full joint statement by funders of health research*, Wellcome Trust, https://wellcome.ac.uk/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health

4.   Conclusions

Diversity is a key feature of this complex landscape of institutions, funding regimes, and disciplinary practices. Unsurprisingly there have been voices calling for an introduction of a common template to tackle variations in policy and 'to promote coordination and collaboration between key stakeholder groups, fostering the alignment of policies, and clarity on the allocation of roles, responsibilities, resources and services'[19]. According to one of the reports the UK could do more to:

> *locate policies and support for open research data more firmly in relation to wider policies for research; in establishing mechanisms for co-ordination between the different agencies with a stake in policy development for open research data; and in establishing programmes to promote and support the development of capabilities, capacity and infrastructure in a co-ordinated and coherent fashion'[20].*

At the other end of the scale, we must also acknowledge the intricate relationship between policy and practice. In the UK, in some areas (in genomics, for instance) there is a long standing and established collaboration between researchers and policy makers. In other areas, this is still missing. The policies are not being rigorously implemented and the attitudes and behaviours of researchers are sometimes seen as a barrier to policy development[21]. More could be done to understand the different disciplinary cultures and the existing social norms that guide the researchers' attitude to data sharing[22]. However, the data context and disciplinary practices vary greatly across domains so it is impossible to get specific policy requirements that apply to all – some degree of tailoring is inevitable.

Data policies can only be effective if they are developed with researchers and the wider research community in mind, and not delivered 'top down'. Researchers' needs, policy goals and the emergent data landscape must be well integrated to deliver improved outcomes for all. If a policy is to take effect it must be closely followed by services offering real value. Naturally, resource implications push for join up efforts here, and a more harmonised policy landscape could help to foster more coherent and interoperable research data infrastructure. By agreeing common principles that we are all working towards, we can have a clearer policy position to communicate to researchers and a more directive set of requirements to guide implementation through the delivery of data services.

SPARC Europe would like to thank DCC for carrying out this work

D|C|C

www.dcc.ac.uk

---

[19] Open Research Data Taskforce (2017). *Research Data Infrastructures in the UK*, Open Research Data Taskforce, http://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf
[20] Open Research Data Taskforce (2017). *Research Data Infrastructures in the UK*, Open Research Data Taskforce, http://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf
[21] Open Research Data Taskforce (2017). *Research Data Infrastructures in the UK*, Open Research Data Taskforce, http://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf
[22] Neylon C (2017). *Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers*. Research Ideas and Outcomes 3: e14673. https://doi.org/10.3897/rio.3.e14673