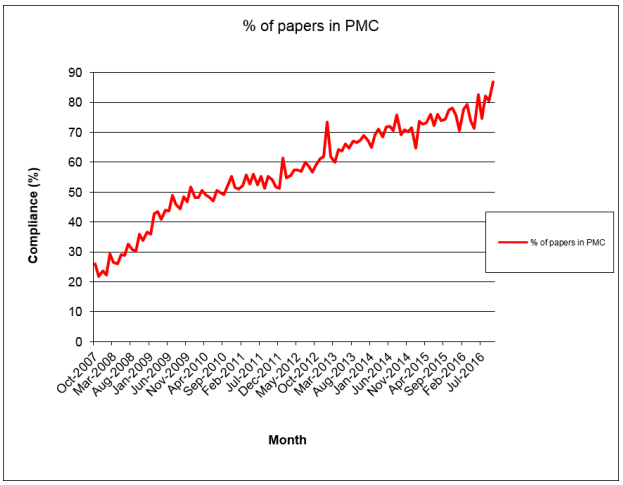
Data management. Learning from the innovators: 2 leading funders compared

1. Data policy from the European Commission and Wellcome

The last decade has seen a significant growth in research data policy. There was a surge in the release of data policies from UK research funders[1] around 2010-2011, and similar cycles are evident in the USA following the OSTP memo,[2] and in Europe. An analysis commissioned by SPARC Europe and updated biannually by DCC shows a continual growth in data policy in recent years.[3] Not all policies are equal though – some have changed the terms of reference and provoked shifts in practice. The EPSRC policy was a game changer in terms of institutional responses to RDM services in the UK (as noted in a parallel SPARC policy briefing[4]). Others are shaping our understanding of openness, FAIR data and DMPs.

Wellcome has been a pioneer in terms of open scholarship. Its open access policy, support of PubMedCentral and annual statistics on uptake have shifted practice and progressively increased the access to published research papers year-on-year. Currently around 80% of Wellcome-attributed articles are made open access in line with its policy. It has also digitised many of its own library collections and actively engages on social media, increasing access and use of the collections. More recently, Wellcome shifted thinking in terms of data management and sharing, calling for Output Management Plans that consider the full breadth of research outputs needed to enable research to be fully understood and reused.



Similarly the European Commission has inspired many member states to initiate open science programmes and RDM infrastructure projects in response to its Open Research Data Pilot, initiated in 2014. The scope, requirements and guidelines relating to the pilot have shifted each year in response to community feedback. In particular, the requirement for openness has been clarified with the introduction of the mantra "as open as possible, as closed as necessary" and the FAIR guidelines seek to reinforce the value of data management, regardless of whether data can be shared openly or not. This responsiveness to community feedback is laudable and augurs well for an appropriate and implementable open data policy in FP9.

[1] Jones, S. (2012) Developments in Research Funder Data Policy in *International Journal of Digital Curation* vol 7 issue 1 doi:10.2218/ijdc.v7i1.219

[2] Stebbins, Michael (2013) *Expanding Public Access to the Results of Federally Funded Research*, https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally- funded-research

[3] SPARC Europe and DCC (2017) *An Analysis of Open Data and Open Science Policies in Europe,* https://sparceurope.org/download/2285/

[4] SPARC Europe and DCC (2017) *Open Data Policy in the UK*, https://sparceurope.org/download/2035/

This case study will review the Wellcome policy on data, software and materials management and sharing[5], as well as the European Commission's Open Research Data pilot,[6] drawing common threads and indicating future policy directions.

2. Data plus: broadening out the research data definition

The term 'data' is often used as shorthand to imply a much broader set of related materials that enable research to be understood and reused. The European Commission's guidelines state that data and associated metadata need to be deposited in repositories for reuse. Moreover, they note that information about the tools needed to validate the results (e.g. specialised software or software code, algorithms and analysis protocols) should also be shared, if not the instruments themselves.

The 2017 Wellcome policy goes further by explicitly calling out a wide range of materials that should be made available to support reproducibility and underpin further research. This includes datasets generated by the research project, original software created in the course of the research, as well as materials like antibodies, cell lines, and reagents. Applicants should consider all their research data and materials together with IP and know-how in an Outputs Management Plan, describing their approach for maximising the value and reuse of these.

3. Active Data Management Plans (DMP)

The Wellcome's new policy is shaping thought on the coverage of DMPs. Funders such as the UK's EPSRC have asked for Software Management Plans, but this is the first case where as funder has been explicit that all types of research output should be considered together in one 'Outputs Management Plan'. Wellcome also calls for the plan to be dynamic and updated throughout the research lifecycle.

The EC pilot is similarly driving change in terms of DMPs. The Open Data Pilot considers DMPs to be a deliverable, first due by month six of the project and then updated throughout the course of the project as important changes occur to the data or intentions regarding management and sharing. The EC has adapted its template for DMPs, changing from a plan that was considered per dataset to one that addresses FAIR data management across the whole project. Disciplinary guidelines are planned and further changes may be made in light of community responses to a survey on the Horizon 2020 DMP template.[7]

4. Open data

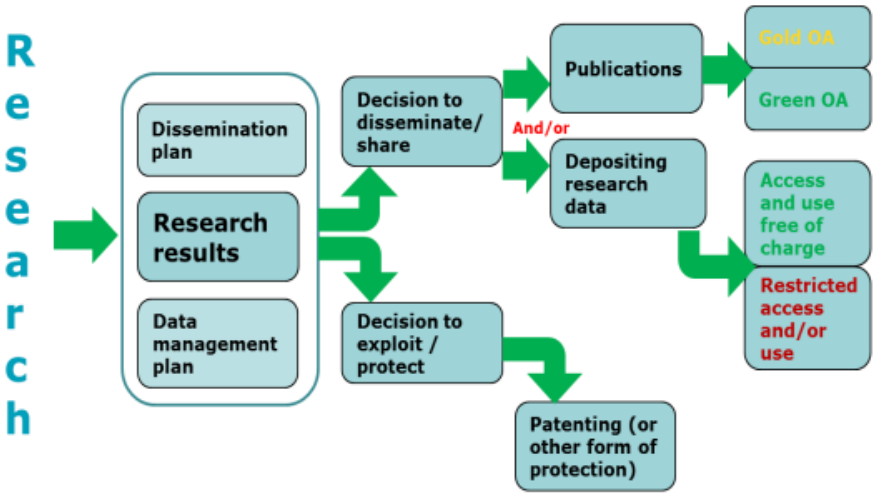
There is often a perceived tension between open data policy and protecting research participants or commercial interests, particularly in the area of health research. The European Commission has sought to address this with its mantra "as open as possible, as closed as necessary" highlighting that the extent of openness should be defined by the project given the nature of the data and specific research context.

[5] Wellcome (2017) *Policy on data, software and materials management and sharing,* https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing

[6] European Commission Directorate General for Research and Innovation (2018) *Open Research Data (ORD) - the uptake in Horizon 2020,* collection of uptake statistics and guidelines relating to the pilot https://data.europa.eu/euodp/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of -horizon-2020

[7] Grootveld, M et al (2018). *OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans* http://doi.org/10.5281/zenodo.1120245

Wellcome expects researchers to manage outputs in a way that will achieve "the greatest health benefit" whether this be by making content openly available or using intellectual property as a tool to help protect and commercialise an original idea, product or technology. This parallels the EC's diagram showing that both commercialisation and open data sharing are legitimate paths to managing outputs.



Wellcome also highlights the timeliness of data sharing, particularly in public health emergencies. It requires that researchers share quality-assured interim and final data as rapidly and widely as possible, and in advance of journal publication. This is a dimension that is increasingly emerging in funder policies.

5. FAIR

The European Commission introduced new guidelines in July 2016 which promote the concept of FAIR data; that is data which is Findable, Accessible, Interoperable and Reusable.[8] FAIR Data is a core part of the Commission's policy. It has convened a FAIR Data Expert group[9] to advise on the steps needed to turn FAIR data into reality. The Action Plan and recommendations produced by this group will inform the EOSC governance structure and future policy.

FAIR has gained incredible momentum since its inception in 2014. It has not only been adopted by the EC, but other researcher funders such as the Irish Research Council and Health Research Board Ireland have referenced FAIR in recent funding calls, and the German and Dutch governments proposed financial support for the GO-FAIR initiative in 2017.[10] Wellcome's policy also alludes to elements of FAIR asking researchers to ensure outputs are discoverable, deposited in recognised community repositories, and are assigned persistent identifiers.

[8] FORCE11 (2014) *The FAIR Data Principles* https://www.force11.org/group/fairgroup/fairprinciples
[9] European Commission Expert Group on 'Turning FAIR Data into Reality' http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3464
[10] Scholz W & Tochtermann K (2017) *GO-FAIR Ready for Take-off* https://www.zbw-mediatalk.eu/en/2017/06/go-fair-ready-for-take-off

6. Enabling environments

The Wellcome policy notes that it is creating an environment that enables and incentivises researchers to maximise the value of their research outputs. It invests in key data resources via its dedicated biomedical resource and technology development scheme, and has provided long-term support for databases and tools such as Ensembl, PDB-Europe and the Open Microscopy Environment.[11] Wellcome also lists example repositories for different types of data.[12]

The European Commission announced its intention to develop a European Open Science Cloud (EOSC) in the Communication on the European Cloud Initiative (COM(2016) 178 final), as a seamless environment for all European researchers to access and work with data. The implementation of the EOSC is already supported through Horizon 2020 funding, and in particular through a dedicated EOSC call in the 2018-2020 work programme on research infrastructures, including e-infrastructures.[13] This includes supporting a robust infrastructure environment beginning by integrating and consolidating e-infrastructures and connecting them to research infrastructures, through the project EOSC Hub. Other supporting projects explore the EOSC governance, national initiatives, FAIR data uptake and innovative services, among other things. In the meanwhile, the European Commission is exploring governance and funding mechanisms of the EOSC with Member States. In March 2018 the EC released a Staff Working Document providing an Implementation Roadmap for the European Open Science Cloud.[14]

It is critical that data policy is complemented by investments in infrastructure and support like those made by the Wellcome Trust and European Commission to ensure they are feasible to implement.

7. Monitoring and rewards

Lessons learned from the implementation of its Open Access policy are informing Wellcome's thinking on monitoring data and other outputs. Compliance is much more difficult to measure here as it involves a degree of subjective judgement. Nevertheless, Outputs Management Plans are reviewed when making a funding decision, and a check is performed at end-of-grant to give a clear message to researchers that they care about data access and will follow up. Sanctions are felt to be premature (they were first implemented in 2012 for open access to publications) but are likely to be brought in at some point.

With respect to incentives, Wellcome has been partnering with other funders around the development of a funder 'blueprint' setting out common steps to recognise data sharing and open research in the grant application process. This includes encouraging applicants to include other outputs alongside publications, and a question on the track record of open practice. Welcome already encourages its committee members to recognise the wide diversity of valuable outputs that result from research. These actions echo recommendations from the Open Science Policy Platform report on rewards and incentives, which concluded that "for the practice of Open Science to become

[11] See: Ensembl at: https://www.ensembl.org PDB-Europe at: http://www.ebi.ac.uk/pdbe and OME at: https://www.openmicroscopy.org

[12] Wellcome Data repositories and database resources https://wellcome.ac.uk/funding/managing-grant/data-repositories-and-database-resources

[13] See https://ec.europa.eu/programmes/horizon2020/en/european-research-infrastructures-including-e-infrastructures-%E2%80%93-horizon-2020-work-programme-2018

[14] See: https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none

mainstream, it must be embedded in the evaluation of researchers at all stages of their career."[15] The report proposes actions for universities, senior researchers and funders to recognise and reward open practices in the vein being pursued by Wellcome.

The European Commission is also reviewing Data Management Plans submitted by active research projects. This is either done by project officers or external experts. An evaluation framework based on the Horizon 2020 template has been developed by the Research Executive Agency to guide assessments, and the FOSTER Plus project and the Digital Curation Centre have provided several training courses to train EC staff on reviewing DMPs. Since the Open Research Data requirements are still a pilot, not formal policy, sanctions are not yet applied.

8. Conclusions

At a Science Europe workshop on harmonising Research data policy and guidelines across Europe,[16] Robert-Jan Smits, the EC's Director-General of Research and Innovation, asserted that the transition to open science is in the hands of the members of Science Europe. They collectively hold over 90% of the research budget so have the power to drive change. A Science Europe study presented at the event found that over 40% of its members do not require DMPs and many only have partial requirements.[17] The attendance and interest in the workshop suggests this will change with more data policies and DMP requirements being released in the run up to 2020. These new policies will hopefully emulate the forward-thinking approach of the Wellcome Trust and the European Commission, driving further change and innovation. In particular, new policies should apply the definition of data broadly and ensure they address infrastructure issues and rewards to promote actual implementation and change in practices rather than lip service to requirements.

There is also an opportunity for funders to practice what they preach with research data policies. These policies are typically made available as PDFs or via non-stable weblinks, causing Daniel Mietchen to call for them to be assigned PIDs at the workshop. Assigning persistent identifiers to versioned policy documents would enable stable referencing and remove ambiguity about what set of requirements pertain to which datasets. Moreover, if machine-readable versions of the policies were provided, tools could automatically act on the requirements they describe, assisting researchers to follow the policies and enabling automated monitoring of the data management and sharing practices that ensue. Research funder data policies play a valuable role in driving FAIR and open data practices. By adhering to the same principles they propose, these policies could become a more active component of the research data ecosystem they are inspiring.

SPARC Europe would like to thank DCC for carrying out this work
March 2018

[15] European Commission (2017) Evaluation of Research Careers fully acknowledging Open Science Practices: Rewards, incentives and/or recognition for researchers practicing Open Science. Report written by the Open Science Policy Platform (OSPP) Working Group on Rewards under Open Science http://doi.org/10.2777/75255
[16] Event page and slides: http://www.scienceeurope.org/policy/policy-areas/research-data/rdm-initiative
[17] See data in presentation at: http://www.scienceeurope.org/wp-content/uploads/2018/02/7_1_SE-RDM-WS-Jan-2018_Survey_Results_Krupavicius.pdf