Developing Open Data policies

Open Data and Open Access

The 2012 European Commission recommendation on access to and preservation of scientific information states that: "*open access to scientific research data enhances data quality, reduces the need for duplication of research, speeds up scientific progress and helps to combat scientific fraud*".[1]

Research data are defined as the data, files, and other records, produced in research or that evidence and validate research results, or more specifically as units of information: facts or numbers used as the basis for reasoning and calculation.[2]

While Open Access to research outputs has a long history and development, Open Data have only come into scope recently. However Open Access to research data is fast becoming recognised by stakeholders, including researchers, research funders, data managers, research institutions and publishers, as a key activity complementary to Open Access to scientific publications. Guy and Ploeger report that the current situation is that Europe, North America and Oceania lead the way with regard to request or encouragement policies and statements related to sharing research data. The emerging norm is for funders to push for data release in parallel with activity happening at individual institutions.[3]

Given that Open Data policies are still relatively few in number and have had only a brief life, it is not yet possible to determine the factors that make them effective, as PASTEUR4OA has been able to do with Open Access policies.[4]

There are however some policies in existence, and models and guides to creating and implementing them, on which this document is based.

In the longer term, a recently announced project, led by the International Development Research Centre, entitled *Exploring the opportunities and challenges of implementing open research strategies within development institutions* (http://www.idrc.ca/EN/Programs/NE/Pages/ProjectDetails.aspx?ProjectNumber=108131) will carry out case studies to "assist with the refinement of the guidelines for implementation of open research data policies". These case studies may provide pointers to what makes a policy effective.

What an Open Data policy covers

Following an extensive survey of policies and reports, Guy and Ploeger note that, as might be expected at this stage in their development, the content of policies they have examined varies regarding mandates. However they conclude that most will cover the following elements:

- **Timing**: when publication should take place;
- **Data plan**: requirements for a technical management plan;
- **Access and sharing**: what exactly will need to be available for public use;
- **Long term curation**: data creation and sustainability;

---

[1] European Commission 2012 recommendation on access to and preservation of scientific information: http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdfrecommendation-access-and-preservation-scientific-information_en.pdf

[2] Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

[3] For a full exposition see: Guy, M. and Ploeger, L. 2015. *PASTEUR4OA Briefing Paper: Open Access to Research Data*. Available at: http://pasteur4oa.eu/sites/pasteur4oa/files/resource/PASTEUR4OA%20Briefing%20Paper_FINAL_0.pdf.

[4] See for instance: Swan, A. 2015. *Open Access policy effectiveness: A briefing paper for research funders*. Available at: http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/Policy%20effectiveness%20-%20funders%20final.pdf.

- **Monitoring**: any monitoring that will be carried out by the funding body and guidance available;
- **Storage**: details of the appropriate repository(ies) or data centre(s) to be used;
  Costs: where costs can be claimed from and when.

This analysis echoes the structure of the Digital Curation Centre's (DCC) database of UK funders' policies, which identifies the following common elements (http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies):

- **Data**: a datasets policy or statement on access to and maintenance of electronic resources;
- **Time limits**: set timeframes for making content accessible or preserving research outputs;
- **Data plan**: requirement to consider data creation, management or sharing in the grant application;
- **Access/sharing**: promotion of OA journals, deposit in repositories, data sharing or reuse;
- **Long-term curation**: stipulations on long-term maintenance and preservation of research outputs;
- **Monitoring**: whether compliance is monitored or action taken such as withholding funds;
- **Guidance**: provision of FAQs, best practice guides, toolkits, and support staff;
- **Repository**: provision of a repository to make published research outputs accessible;
- **Data centre**: provision of a data centre to curate unpublished electronic resources or data;

- **Costs**: a willingness to meet publication fees and data management / sharing costs.

CODATA publishes a useful collection of statements outlining **the policies of a number of organisations, mostly in the area of the environmental sciences (**http://www.codata.info/resources/databases/data_access/policies.html**). The organisations are mainly intergovernmental (such as the** World Meteorological Organization and the Intergovernmental Oceanographic Commission of UNESCO) and international (**such as the** International Council for Science and the International Social Science Council).

There is a wide range of requirement and permissiveness, not only between but also within these statements. T**he** World Meteorological Organization's policy for instance states: "Members **shall** provide on a free and unrestricted basis essential data and products which are necessary for the provision of services in support of the protection of life and property and the well-being of all nations" (emphasis added). Then goes on to state: "Members **should** also provide the additional data and products which are required to sustain WMO Programs" (emphasis added).

The Intergovernmental Oceanographic Commission of UNESCO's policy states: "Preservation of data needed for long-term global ocean programs is **required**", and "International data archives **must** include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining the data" (emphasis added).

The model Open Data policy

The RECODE Project (http://recodeproject.eu/) - an FP-7 project funded by the European Union – suggests that a model Open Data policy should include the following elements (for the complete descriptive text see *Policy recommendations for open access to research data* [5]):

- **Open access as default.** The policy should set open access for research data as the default and mandatory requirement and provide appropriate support and funding.

---

5 RECODE. 2014. *Policy recommendations for open access to research data*. Available at: http://recodeproject.eu/wp-content/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf.

- **Responsibilities.** The policy should assign responsibilities and set out the expectations for the main stakeholders.
- **Target content.** The policy should be explicit on which data should be open. Open access should be required for research data used to validate scientific claims in publications.
- **Data Management Plan (DMP).** The policy should require grant applicants who will generate data to provide a DMP as the main tool through which to address comprehensively data management.
- **Time of deposit.** The policy should require data supporting publications to be made open ideally at the latest at the same time with the publications and link to it.
- **Locus of deposit.** The policy should require deposit in certified and trusted repositories and/or data centres.
- **Technical specifications to allow reuse.** To enable research data reuse and citation funders should require information on metadata, DOI, interoperability of systems, machine readability and mineability and software in the policy.
- **Licensing research data.** The policy should require that research data is accompanied by licensing describing the terms of use.
- **Provisions for long-term availability.** Policies should include provisions for the long-term availability of data, since re-use and availability are primary reasons for open access to research data.
- **Compliance with policy.** The policy should make statements regarding compliance to it by the researchers and clarify measures for non-compliance.

Developing an Open Data policy

RECODE also provides a good practical checklist specifically for funders developing policies:

- *Have you mapped relevant international policies for open access to research data?*
- *Have you involved stakeholders and the research community in developing the policy?*
- *Have you assessed the available infrastructures that are necessary for the implementation of your policy?*
- *Have you estimated the costs for data management and preservation?*
- *Does your policy include statements on:*
  - Open access as the default and mandatory position and possibility for closed access is offered when necessary
  - Distribution of responsibilities to involved parties
  - Target data for open access
  - Time of deposit
  - Locus of deposit
  - Technical specifications
  - Licensing
  - Requirement of Data Management Plan
  - Compliance and monitoring statement
- *Do you require grant applicants to offer information regarding data management at the application stage?*
- *Do you include open access to research data as a clause in your grant agreements?*
- *Do you offer guidance to researchers in your website and otherwise to enable them to comply with your policy?*
- *Have you made provisions to provide incentives to researchers for making their research data open?*
- *Have you established a monitoring and compliance mechanism?*
- *Have you decided how and when to evaluate the efficacy of your policy?*

What makes an Open Data policy effective

As noted at the start of this document, it is not yet possible to carry out the sort of rigorous statistical analysis achieved by PASTEUR4OA on Open Access policies: Open Data policies have had too short a life. However one might, by analogy with the PASTEUR4OA results, surmise that mandatory policies are more effective than non-mandatory ones.[6]

This has not yet been realised by many funders. According to the SHERPA/JULIET database, the situation as of early 2016 (http://www.sherpa.ac.uk/juliet/index.php?la=en&mode=advanced) was that only 42 of the funders with a data archiving policy **require** archiving, 112 do not.

One can also surmise that other factors likely to drive usage are: insisting on timing of deposit to be at the latest when outputs are published; linking to data from publications; technical specifications to enable re-use; and clear licensing statements.

Policies of funders

RECODE's *Policy recommendations for open access to research data* holds that "UK research funders, the Research Councils UK and the Wellcome Trust, are global pace-setters in policy development for research data and in comprehensively developing relevant supporting services". The DCC provides in useful tabular form information on what the policies of these funders include (http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies), and also links to the texts of these policies:

Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

(● denotes full coverage, ◐ partial coverage and ○ no coverage; for explanations of the terms used see the section *What an Open Data policy covers* above).

It is interesting to note that:

- All policies cover data;
- Nearly all policies cover all the policy stipulations, but two policies do not cover monitoring;
- Levels of support are variable.

Other organisations are also starting to require data to be open. An editorial in the *BMJ* notes that the International Committee of Medical Journal Editors is proposing "that authors of manuscripts

---

6 Swan A, Gargouri Y, Hunt M and Harnad S (2015) *Open Access policy: numbers, analysis, effectiveness* http://pasteur4oa.eu/sites/pasteur4oa/files/deliverables/PASTEUR4OA%20Work%20Package%203%20Report%20final %2010%20March%202015.pdf.

considered for publication in its members' journals must agree to share de-identified individual patient data no later than six months after publication".[7]

David Ball
February 2016

---

7 Murray, M.J. 2016. "Thanks for sharing: the bumpy road towards truly open data". *BMJ* 2016;352:i849. Available at: http://www.bmj.com/content/352/bmj.i849.