

DEVELOPING OPEN DATA POLICIES

Open Data and Open Access

“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” – European Commission 2012 (http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf).

Research data are the data, files, and other records, produced in research or that evidence and validate research results. Providing Open Access to research data is seen as a key activity complementary to making publications Open Access.

Open Data policies are still relatively few in number and have not been analysed as have Open Access policies. There are however some policies in existence, and models and guides to creating and implementing them.

What an Open Data policy covers (Digital Curation Centre)

Research by the Digital Curation Centre (DCC) suggests the following elements to include in an Open Data policy (<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- **Data centre:** provision of a data centre to curate unpublished electronic resources or data;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.

The model Open Data policy: the RECODE Project

The EU-funded RECODE Project recommends the following content for a model policy (http://recodeproject.eu/wp-content/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf):

- **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.
- **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.

What makes an Open Data policy effective

Open Data policies have had too short a life to support the type of analysis undertaken by PASTEUR4OA on Open Access policies (<http://www.pasteur4oa.eu/resources>).

However, by analogy with the PASTEUR4OA analysis, it can be inferred that factors driving effectiveness include:

- Making policies mandatory;
- Insisting on timing of deposit to be at the latest when outputs are published;
- Linking to the data from publications;
- Technical specifications to enable re-use;
- Clear licensing statements.

Policies of funders

The United Kingdom has strong policies for open access to both research publications and research data.

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 (below) information on what the policies of these funders include (and also links direct to the texts of these policies) (<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-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).

These comprehensive policies may also be used as models by other funders.

David Ball

April 2016