

EIFL Digital Research Literacy Training Programme

Writing a data management plan
26 May 2021

David Ball, David Ball Consulting
davidball1611@gmail.com

This Morning

- Open Data and its complexities
- DMP Checklist
- The resources provided in the EIFL Training Programme

Research Data Definitions

- Research data can be defined simply as whatever is either produced in research or evidences research outputs.
- The European Commission's definition is: "information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation" (*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).
- Examples include: statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings, images.
- Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge.

Open Data Exemptions

- *“As open as possible, as closed as necessary”* – European Commission
- Exemptions:
 - Incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited
 - Incompatible with the need for confidentiality in connection with security issues
 - Incompatible with rules on protecting personal data
 - Embargoes?

Concordat On Open Research Data (UKRI et al.)

- To ensure that the research data gathered and generated by members of the UK research community is made openly available for use by others wherever possible in a manner consistent with relevant legal, ethical and regulatory frameworks and norms.
- To establish a set of expectations of good practice with the intention of establishing open research data as the desired position for publicly-funded research over the long-term.

<https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/>

Concordat On Open Research Data / 2

- Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form.
- They may include, for example, statistics, collections of digital images, sound recordings, transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript.

FAIR Data Principles (FORCE11)

- **Findable:** easy to find the data and the metadata for both humans and computers - persistent identifiers (PIDs);
- **Accessible:** data should be retrievable by their identifier using a standardised and open communications protocol;
- **Interoperable:** data should be able to be combined with and used with other data or tools. The format of the data should therefore be open and interpretable for various tools;
- **Re-usable:** metadata and data should be well described so that they can be replicated and/or combined in different settings
(<https://www.force11.org/group/fairgroup/fairprinciples>)

Research Data Policies

Emerging policies vary, but 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
- **Monitoring:** any monitoring that will be carried out by the funding body and guidance available
- **Storage:** details of the appropriate repository or data centre to be used
- **Costs:** where costs can be claimed from and when

Data Management Plans

- DMPonline is a web-based tool that supports researchers to develop data management and sharing plans
<https://www.dcc.ac.uk/dmponline> (Data Curation Centre)

DMP Checklist

Good model from Data Curation Centre

<https://www.dcc.ac.uk/DMPs/checklist>; includes:

- Administrative Data
 - pertinent project ID; project name and description; funder; principal investigator's name and persistent ID; relevant policies of funder and institution.
- Data Collection
 - what data will be collected or created, and how; type, format and volume; sharing and long-term access; standards and methodologies; quality assurance.

DMP Checklist / 2

- Documentation and Metadata
 - to accompany the data to ensure they can be read and interpreted in the future; metadata standards; methodologies and definitions.
- Ethics and Legal Compliance
 - Consent for preservation and sharing; protecting identity; secure storage; institutional ethics committees.
 - Copyright and intellectual property rights; ownership of data; licences; embargoes e.g. for commercial exploitation; relevant policies of funder/institution.

DMP Checklist / 3

- Storage and Back-up
 - Costs and responsibilities; recovery; security and access control; collaborators.
- Selection and Preservation
 - What must be retained (legal, contractual, ethical), and for how long; what must be destroyed; possible re-use.
 - preservation plan; repository; costs; after funding has expired.

DMP Checklist / 4

- Data Sharing
 - Discovery; conditions for sharing; timing of deposit/availability; embargoes.
- Responsibilities and Resources
 - Responsibilities: for compliance with policies etc.; for implementing and reviewing the DMP; for data capture, metadata production, data quality, storage and backup, data archiving & data sharing.
 - Resources: staff, training, technology.

The Training Programme

Content

- A wide range of formats (videos, courses, guides, practical exercises ...) which can be customised
- For use with different audiences: trainers, library staff, researchers, managers...
- A lot of content within the documents (videos, webinars, articles, presentations...): delve down to find individual relevant elements, such as individual graphs, tables...
- Creative Commons licences

<https://eifl.net/resources/eifl-digital-research-literacy-training-programme-outline-librarians>

Embedding Open Science

- Use of the documents assembled is key to winning hearts and minds within institutions, but Open Science needs to be embedded:
 - In academic and research policies, with the committed backing of senior management.
 - In decisions on researchers' appraisal and progression – only papers deposited in the institutional repository count for progression.
 - In the research and academic committees of the institution
 - In the involvement of champions internal and external.
 - In workflows of researchers, supported by library staff.
- Open Science must no longer be something new and external.