biodiversa

The network promoting pan-European research on biodiversity, ecosystem services and Nature-based solutions

BELMONT FORUM

Guidance document on data management, open data, and the production of Data Management Plans

To cite this report

Goudeseune L., Le Roux X., Eggermont H., Bishop W., Bléry C., Brosens D., Coupremanne M., Davis R., Hautala H., Heughebaert A., Jacques C., Lee T., Rerig G., Ungvári J. (2019). Guidance document for scientists on data management, open data, and the production of Data Management Plans. BiodivERsA report. 48 pp.

DOI : 10.5281/zenodo.3448251

Main contact for this report

Lise Goudeseune: l.goudeseune@biodiversity.be
(BelSPO/Belgian Biodiversity Platform)

Photography credits

Pixabay cover, p. 2, 11, 16, 18, 20, 26, 40, 44, 48
Edwin Brosens p. 22
Pexels p. 31, 43
Sophie de Grissac p.34
iStockPhoto p. 39
Flickr p. 4, 8, 12
© Cyril FRESILLON/CNRS Photothèque p. 3
© Roland GRAILLE/CNRS Photothèque p. 15

Acknowledgements

The work on DMP has been guided by an ad hoc working group composed of:

- Wade Bishop, University of Tennessee (USA)
- Claire Bléry, BiodivERsA, FRB (France)
- Dimitri Brosens, INBO/BBPf (Belgium)
- Maxime Coupremanne, DEMNA/BBPf (Belgium)
- Rowena Davis, Belmont Forum e-I&DM (USA)
- Hilde Eggermont, BiodivERsA, RBINS/BBPf (Belgium)
- Lise Goudeseune, RBINS/BBPf (Belgium)
- Harri Hautala, AKA (Finland)
- André Heughebaert, BBPf (Belgium)
- Cécile Jacques, BiodivScen, FRB (France)
- Erica Key, Belmont Forum (USA)
- Tina Lee, Belmont Forum e-I&DM (USA)
- Xavier Le Roux, BiodivERsA, FRB/INRA (France)
- Gaby Rerig, DFG (Germany)
- Alexandre Roccatto, FAPESP (Brazil)
- Robert Samors, Belmont Forum e-I&DM (USA)
- Maria Uhle, NSF (USA)
- Judit Ungvári, Belmont Forum e-I&DM, NSF (USA)
- Luciano Verdade, FAPESP (Brazil)
- Jean-Pierre Vilotte, ANR (France)

TO CONTACT BIODIVERSA

BiodivERsA Coordination

Xavier Le Roux, Coordinator and CEO
xavier.leroux@biodiversa.org
Tel: +33 (0)6 31 80 38 20

BiodivERsA Secretariat

Claire Bléry, Secretariat Executive Manager
claire.blery@fondationbiodiversite.fr
Tel: +33 (0)1 80 05 89 37

Fondation pour la Recherche sur la Biodiversité
195 rue St Jacques, 75005 Paris, France
www.biodiversa.org

FOLLOW US:

@BiodiversaPlus

TO CONTACT BELMONT FORUM

Belmont Forum Secretariat

Erica Key, Executive Director
info@belmontforum.org
Tel: +598 2600 2798 ext. 230

Belmont Forum Data Management Contact Point

Judit Ungvári, AAAS S&T Policy Fellow
jungvari@nsf.gov
Tel: +1 703-292-2968

c/o Inter-American Institute for Global Change Research
Avenida Italia 6201, Edificio Los Tilos 102,
Montevideo, Uruguay 11500
www.belmontforum.org

FOLLOW US:

@Belmont_Forum

To contact the BiodivScen project Secretariat:
Patricia Kammerer - patricia.kammerer@fondationbiodiversite.fr

TABLE OF CONTENTS

1. INTRODUCTION

1. INTRODUCTION

1.1. THE IMPORTANCE OF SCIENTIFIC DATA & DATA MANAGEMENT

Scientists are expected to generate new knowledge often based on the generation of new data. For a long time, their major production consisted of scientific publications presenting this new knowledge and the methods used to obtain the data underlying scientists' conclusions… often ignoring the publication of well organised and described datasets (Fig. 1).

However, and as for many other activities, organising, achieving, and making accessible data are increasingly important in science, for improving traceability and fostering data sharing.

A recent survey (CrowdFlower, 2016) indicated that data scientists do not spend most of their time building algorithms, exploring data or doing predictive analyses: they actually spend most of their time cleaning and organising data (Fig. 2)! This is referred to as 'data wrangling' and is sometimes compared to digital janitor work.

Figure 1: How scientists' achievements have traditionally been evaluated (Herrema, 2014)

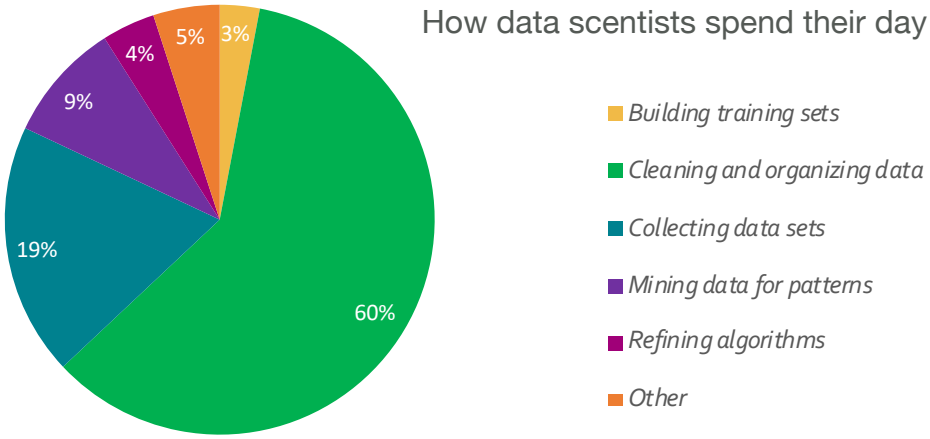


Figure 2: How data scentists spend their day (redrawn after CrowdFlower, 2016)

In this context, data organisation and formatting, and ultimately data sharing and use or re-use, have become of paramount importance to science. This has even become more critical as research is changing rapidly. Digital technologies now make the conduct of science more collaborative, more international and more open to the world, and researchers as well as research programmers and funders have to embrace these changes (European Commission, 2016a). In particular, a more Open Science requires a more systematic open access to scientific data. This in turn requires researchers to take into account and apply adequate approaches and tools for data management, including the production of Data Management Plans (DMPs).

1.2. THE PRESENT GUIDANCE DOCUMENT AND WHAT IS EXPECTED FROM RESEARCH PROJECTS

The purpose of this document is to help the projects funded through joint Calls for transnational research projects to update and develop their DMPs. It has been developed by BiodivERsA and the Belmont Forum part of their joint programme 'BiodivScen', but can also be useful beyond this specific context.

The projects funded are encouraged to make the data produced and used during and after the lifetime of the project "as open as possible".

As part of their research project, each project team should appoint a Data Manager and create and implement a Data Management Plan (DMP) to enable sharing of research data.

In each application form, project applicants were asked to answer a series of questions on their intentions and plans regarding the management of the research data produced or used during their project.

Starting from this information, which was considered a draft Data Management Plan, **projects were expected to provide an updated and improved Data Management Plan** in the first months after the start of the project. The funded consortia will be supported and helped through:

» a devoted 'Data Management Plan' workshop;

» the present guidance document.

So it can also be used beyond the context of the BiodivScen programme.

1.3. EXISTING GUIDELINES AND TEMPLATES

Many guidelines and toolkits have already been published, most notably (but not limited to):

» the Belmont Forum's toolkit and guide (Belmont Forum e-I&DM, 2018) with a section offering resources specifically matched with the requirements of "the Belmont Forum's" data management expectations and requirements;

» the guidelines developed by the European Commission for H2020 projects (European Commission, 2016b);

» the generic guidelines for drafting a DMP based on the BelSPO-funded project SAFRED (Milotic et al, 2018).

There is no mandatory template to use: all projects are different and the DMPs should be adapted to the focus of the research and the type and volume of data that will be used or produced. However, we strongly encourage the use of one of the following tools:

1/ The template proposed in this document (see 4.2 A proposed template to structure your DMP when drafting your proposals) includes a series of questions to help you structure your DMP and make sure all aspects are covered.

2/ The Data & Digital Outcomes Management Plan (DDOMP) offers a condensed structure based on a series of questions covering the whole data management scope. It was produced by the Belmont Forum e-Infrastructures and Data Management initiative for CRAs (Collaborative Research Actions).

The present document provides funded applicants with centralized and summarized information on:

1. open science and open/FAIR data principles;
2. data management concepts and needs in the context of this type of transnational funded projects; and
3. relevant tools and resources.

This guidance is intended to be a living document, to be updated as needed so it can also be used beyond the context of the BiodivScen programme.

A short glossary (see Annex I - Glossary) contains the most important terms used in the data management field.

2. MAIN PRINCIPLES & POLICIES FOR DATA MANAGEMENT

2. MAIN PRINCIPLES & POLICIES FOR DATA MANAGEMENT

2.1. WHAT ARE OPEN DATA & OPEN SCIENCE ?

The holistic concept of Open Science refers to a movement which sets out a broader vision of having all scientific outputs open and endeavours to make science freely and easily accessible to everyone. This movement also particularly supports science in its integration into the digital era.

> "**Open** means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)." (source: The Open definition)

Science is characterised by the collection, analysis, interpretation, publication of data and its integration to existing knowledge. Therefore, Open Science encompasses many aspects, including the concepts of Open Access, Open Data, Open Standards, Open Education, etc. that facilitate the diffusion of scientific knowledge.

Open Data (sometimes referred to as **Open Access to Data**) is the idea that data should be freely available to everyone to use and re-publish as they wish, without restrictions from copyright, patents or other mechanisms of control.

It has to be distinguished from related concepts such as **Open Access** (referring to having papers published in free and open journals), and **Open Source** (referring to programmes or software with publicly accessible code that can be shared and modified) (see Fig. 3).

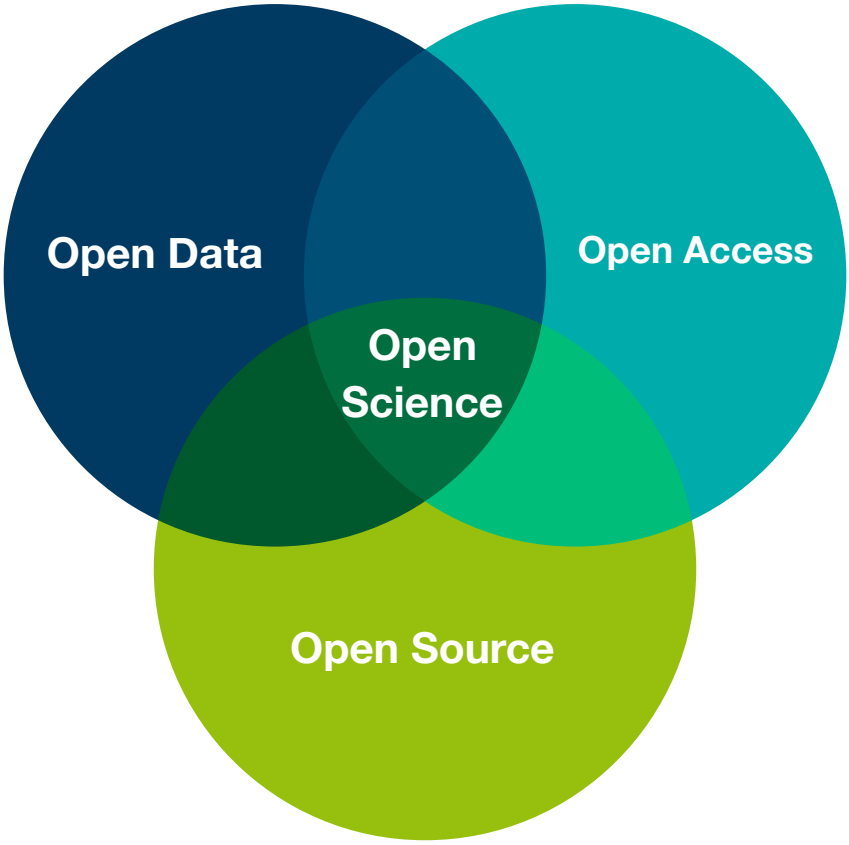


Figure 3: Graphical representation of the different aspects of Open Science (after Jomier, 2017)

2.2. POLICY AND PRINCIPLES FOR THIS GUIDANCE DOCUMENT

One of the objectives of both the Belmont Forum and BiodivERsA is to promote and encourage the funded projects to make their data as open as possible, though with restricted or closed access where appropriate and necessary. Openness of research data indeed promotes its reuse.

For instance, the funding agencies from BiodivERsA and the Belmont Forum participating to the BiodivScen programme have agreed on common Open Data principles and a data policy document was submitted to the research teams during the application process.

The present guidance document intends to help the projects to follow and comply with this data policy, which principles are aligned with the **Data and Digital Outputs Management Plan** (DDOMP) developed by the Belmont Forum and the Open Science principles promoted for developing the European Research Area on Biodiversity (BiodivERsA, 2016; European Commission, 2016a).

2.3. THE FAIR PRINCIPLES

The principles discussed above are based on, and in line with, the FAIR principles (FORCE11, 2014; Wilkinson et al, 2016 ; SNSF, 2017b), a set of guiding principles which define a range of qualities a published dataset (or any digital research object) should have in order to be:

F

FINDABLE

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

A

ACCESSIBLE

Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

I

INTEROPERABLE

Metadata and data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

R

REUSABLE

Data and collections have a clear usage licenses and provide accurate information on provenance.

(according to LIBER, 2017)

These principles provide guidance for scientific data management and stewardship and are relevant to all stakeholders in the current digital ecosystem. They directly address data producers and data publishers to promote maximum use of research data.

THE FAIR GUIDING PRINCIPLES

FINDABLE

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

ACCESSIBLE

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

INTEROPERABLE

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

REUSABLE

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

according to Wilkinson et al, 2016

2.4. OPEN DATA PRINCIPLES AT THE EUROPEAN LEVEL

The European Commission, as a European public funding body, also developed Open Science principles related to research data. Horizon 2020 requires that data of publicly funded research projects must be accessible to anyone, free of charge, and that each beneficiary must ensure Open Access to all peer-reviewed scientific publications relating to its results (European Commission, 2016c).

The EU practice with open research data is currently being piloted in the Open Research Data (ORD) pilot stating that open access to data becomes the default setting for Horizon 2020 projects (projects are 'as open as possible, as closed as necessary').

Part of this pilot includes development of appropriate Data Management Plan (DMP) models. A DMP is required for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. However, projects that opt out are still encouraged to submit a DMP on a voluntary basis.

FAIRsFAIR (Fostering FAIR Data Practices In Europe), a project funded through Horizon 2020, aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle with emphasis on fostering FAIR data culture and the uptake of good practices in making data FAIR.

2.5. THE EU GDPR AND OPEN DATA

In 2018, the GDPR[1] (General Data Protection Regulation) came into force and raised many questions regarding its impact/consequences on data management practices and openness of data in research projects.

According to the EU, the aim of the Regulation is not to refrain data sharing but rather to create a framework that should ease data management and make it more transparent. In that aspect, "*protecting data and opening data*" are not excluding each other, they actually share similar goals (European Data Portal, 2018). The moto related to Open Data is "*as open as possible, as closed as necessary*". Therefore projects should make sure the way their data is used, produced, and shared is GDPR compliant.

Since GDPR deals exclusively with **personal data**, the only situation when it directly affects Open Data is when Open Data includes personal data. Is considered personal data any information that relates to an identified or identifiable living individual e.g. name, surname, home address, email, location data, IP address, etc. This data can only be processed with a **preliminary clear and explicit** consent of the person it relates to. Another option is to render the data **anonymous** in such a way that the individual is not or no longer identifiable (in an irreversible way). Then, it is no longer considered personal data (European Commission, 2018b).

The University of Helsinki, for example, has its own **GDPR Privacy Template** (Raymond, 2019): this form is intended for research participants to be informed about (and give their consent to) the processing of their personal data. The Belgian DMPTool (see also in: 4.1 Tools & templates for developing a DMP) integrates a GDPR functionality. It gives the possibility to create a DMP either with or without GDPR questions (depending on whether the project intends to process personal data or not).

1. The Regulation (EU) 2016/679 or General Data Protection Regulation was approved and adopted by the EU Parliament in April 2016, and came into force on 25th May 2018. (European Commission, 2018b)

2.6. OPEN DATA POLICIES AT AGENCIES' AND NATIONAL LEVELS

In addition, researchers should be aware of and ensure that they follow and comply with the institutional rules and requirements on data management issued by their organisation or funding agency[2], as well as to the policies at national level. In the Annex II, we provide useful links to (information on) national and funders' Open Data/Open Access policies.

There are a number of different policies and strategies at different levels (including funders, national and EU). As there is no common standard format for data management plans, they are difficult to compare and evaluate. The Science Europe Working Group on Research Data has recently developed a framework for discipline-specific data management practices (Science Europe, 2018).

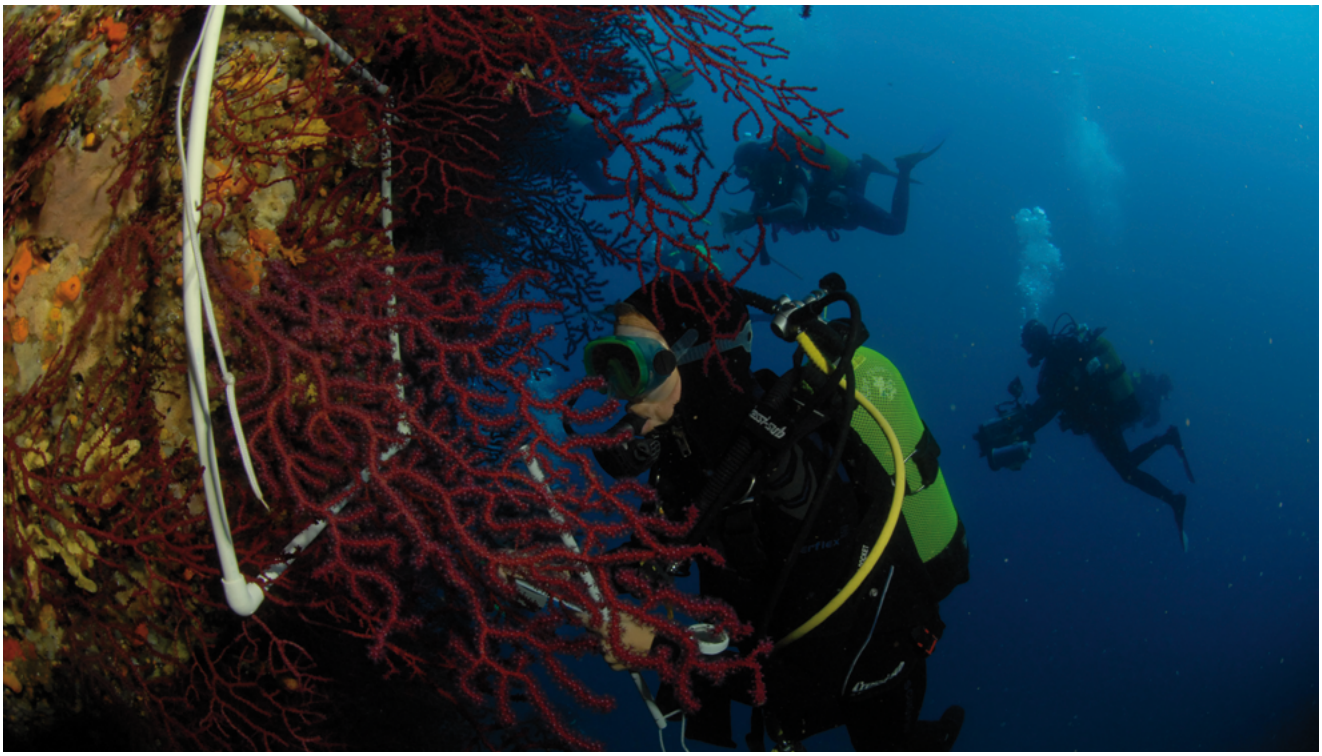
In addition, researchers can use the policy comparison tool developed by The Belmont Forum e-Infrastructures & Data Management project which compares over 20 data management plans from participating Belmont Forum agencies, many of them being also BiodivERsA members.

How researchers can deal with multiple (different and possibly conflicting) Open Data policies/ strategies between different funders, countries, EU,…:

In those instances where there are data policy conflicts among different funders of a given transnational project, researchers should raise the issue to the concerned funders and ask them help to resolve the conflicts.

In the context of joint calls whether with the Belmont Forum or with BiodivERsA, in case data policy conflicts among a project's funders arise, researchers are also recommended to inform the Data Management Working Group or secretariat of the initiative, which may provide help in solving the issue.

2. Based on the results of a survey launched in 2017 by the Belmont Forum and BiodivERsA, it appeared that many BiodivScen agencies have open access policies (for publications), but only a few of them have open data policies (related to data).



3. WRITING A DATA MANAGEMENT PLAN

3. WRITING A DATA MANAGEMENT PLAN

3.1. WHAT IS A DMP AND WHAT ARE THE BENEFITS FOR A RESEARCH PROJECT?

A Data Management Plan (DMP) is a formal document that describes the data management life cycle for the data to be collected, processed and/or generated by a research project (European Commission, 2016c).

It can be distinguished from a Data Curation Profile (DCP): while the DMP is the general plan or blueprint, the DCP is a tool to gather the relevant pieces for data management, to inform the construction behind the plan. In that sense, DCPs inform DMPs.

"*The goal of a DCP is to give curators the information needed to best prepare the data, manage its access and use, and design other value-added services, such as enhanced metadata applications and analytical tools, to facilitate sharing and preservation.*" (Bishop & Grubesic, 2016).

WHAT ARE THE BENEFITS OF DEVELOPING A DMP?

- If prepared early, increases efficiency during the project
- Data collected and stored in a more structured way
- Avoid or minimise risk of data loss
- Useful if collaborators leave
- Enable share/re-use of data and guarantees research reproducibility
- Increase verifiability of research
- Increases longevity of project by helping to make data available even after project ends

(adapted from: Milotic et al, 2018)

A DMP should not be confused with a data paper, which is a peer reviewed scientific publication describing a dataset. Publishing a data paper helps increasing the visibility and credibility of research, and therefore might be a good way to share interesting datasets generated by a project (see Scientific Data, Nature's dedicated website for data papers, or most Pensoft Publishers journals).

3.2. HOW TO WRITE AND IMPROVE YOUR DMP?

DMPs are unique; their content, composition, and structure can vary greatly as they depend on the project and the data generated.

However, to make sure all aspects have been considered and, if relevant, covered, you will find a list of questions (in section 4.2 A proposed template to structure your DMP when drafting your proposals) that should be addressed when writing a DMP. These questions can also be used as a framework (sections) to structure a DMP.

A few general recommendations and best practices that apply to all types of projects and their DMPs:

- Drafting of a DMP should be considered early in the design of a project (i.e. ideally long before the project even starts);
- Analytical and methodological issues related to data should be written into your research plan;
- A DMP is a living document and should be reviewed and updated whenever needed during the project lifetime;
- The content of the DMP (answers to the questions) should be well structured, comprehensive and detailed, but the answers should be simple, compact, and straightforward;
- Abbreviations should be explained (if they are many, a list of abbreviations should be enclosed); references properly cited; all online tools, websites, and repositories listed with working hyperlinks such as DOI's or other persistent identifiers;
- Bullet points and tables should be used to present data types, file formats etc. in a concise way;
- Freely and easily accessible Open Science tools should be used, as much as possible;
- Your data management efforts and information should be listed on a single webpage, e.g. your project webpage.

4. DEVELOPING YOUR DATA MANAGEMENT PLAN: A PROPOSED TEMPLATE

4. DEVELOPING YOUR DATA MANAGEMENT PLAN: A PROPOSED TEMPLATE

4.1. TOOLS & TEMPLATES FOR DEVELOPING A DMP

The European Commission proposes a DMP template **for Horizon 2020 projects**. Applicants from European countries are invited to take this into account. This is particularly important when research projects are co-funded by the European Commission.

The e-Infrastructure & Data Management Toolkit developed by the Belmont Forum offers a wide range of training resources to suit the myriad needs of researchers, but the resources can also be matched to data management questions asked during different stages of the Belmont Forum proposal and award process. Included among training resources are video presentations on the Data Curation Profile process.

The DMPTool is an online tool that helps projects create, review, and share data management plans that meet institutional and funder requirements. It contains templates (according to the research field, institution, etc.) presented as forms that need to be filled in. The DMP Tool provides a structure for the DMP and, once completed, it is possible to download it in various formats.

It has been released under two main versions:

» The US version which contains the US (and Brazilian) templates, and US-based guidance.

» The UK-DCC version which gives access to European templates, to the new unified European templates suggested by Science Europe, as well as to the ERC guidelines. Their templates match the demands and suggestions of the Guidelines on Data Management in Horizon 2020.

In addition, the tool was used as a basis to develop versions and templates adapted to national/organisations' regulations in several countries[3].

The DMPTool and the Digital Curation Centre both published **lists of public DMPs** (see here and here) that have been created with their tool and that can be downloaded for inspiration, as well as a list of example DMPs from various fields.

The Data Stewardship Wizard is a tool to help researchers build a comprehensive and FAIR-based DMP using a logical flow chart and a smart questionnaire.

GitHub is a website and service, the largest repository of open source software, which can be used to collaborate, store, manage, and update DMPs (e.g. the GloBAM Data Management Plan). It can be used in association with Bookdown, a tool to convert content written in R Markdown into a website or pdf.

The Data Curation Profiles Toolkit provides helpful guidance for data management, and it contains a DCP directory with various examples of DMPs from different disciplines.

The **TrIAS project**, which aims to build an open data-driven framework to support policy on invasive species, is an example of a data management plan for a biodiversity project (Groom et al, 2017).

Below, we propose a **generic DMP** structure that can be very useful when you will write the DMP section of your application and then develop/update your DMP.

3. For example, the Belgian version of the tool, including templates from Belgian institutions can be found here: https://dmponline.be.
For the Finnish version, see https://www.dmptuuli.fi.
Institutions and Universities at the state of Sao Paulo, because of FAPESP policies, have established their own basic DMP templates, in Portuguese, which are housed and maintained at https://dmptool.org.
The Canadian bilingual tool is called DMP Assistant, and it is housed by the Portage Network: https://portagenetwork.ca/.

4.2. A PROPOSED TEMPLATE TO STRUCTURE YOUR DMP WHEN DRAFTING YOUR PROPOSALS

Hereunder, we propose a generic structure for a DMP organised in eight sections, with a list of questions that should be asked and, if relevant, addressed and answered. This list is as comprehensive as possible, but should be adapted to each project and each DMP.

I. DATA MANAGER(S)

» Who will be **responsible for managing the data**? Who will ensure that the data management plan is carried out? *e.g. a Data Manager will take care of the DMP and coordinate the work of the data collectors/providers/users from each WP*

» Will a **specialized and experienced** data expert be part of the project team?

It is recommended to have one (or several) appointed Data Manager(s) to produce the DMP and overview the data management practices before, during, and after the project. This should ideally not be the Project Coordinator itself, and it is a great advantage to have people experienced in data management in the team.

It is recommended also to store the DMP as a project document, which can be updated whenever needed (for example: to update the information about upload of new data etc.).

II. DATA IDENTIFICATION & DESCRIPTION

» What's the **purpose** of the research?

» **What datasets** of long-term value do you expect that the project will collect, process, and produce?

» What is the data? How and in what **format** will the data be collected? Is it numerical data, image data, text sequences, or modelling data? *e.g. microscopic images, video recordings of interviews, etc.*

» **How much** data will be generated for this research?

» **How long** will the data be collected and how often will it change?

» Are you using data that **someone else produced**? If so, where is it from?

» Will there be **other types of material** of long-term value produced? If so, what are your plans for ensuring these are also available over the long-term? *e.g. samples of specimens collected on field; physical collections, software, curriculum materials, etc.*

The DMP should consider, not only primary data, but also secondary data; not only peer-reviewed publications but also communications material, documents for stakeholders, etc. In addition, it should include physical objects such as specimens, etc.

Description of the data and datasets should be complete and detailed: type, flows, quantity, format, etc.

This applies both to data collected/generated as to data (e.g. from other research projects) that will be (re)used.

"Long-term" means those datasets that, over time, will or may be of value to others within your research community and/or the wider research and innovation community.

III. DATA ORGANISATION & EXCHANGE (INTERNALLY, DURING THE PROJECT)

» How do you intend to **manage the data** during the life of the project to ensure their long-term value is protected?

» What is your strategy for **organising your data**? How do you organise your folders and name your files? What directory and file naming convention will be used?

» **What data will be shared** among your colleagues/partners, when, and how?

» **Where** will the data be held during the project?

» **Who** has the right to **manage this data**?

» Who will have **access**? How do you **exchange files** (and other information) with your collaborators?

» How do you take care of **consistency and quality of the data**?

Be specific and complete in explaining WHEN exactly the data will be available (e.g. at any moment of the project) and WHERE (e. g. name of repository or website).

IV. DATA STORAGE AND BACK-UP (DURING THE PROJECT)

» What is your **data storage and backup** strategy?

» How much **data storage** do you need for the project and what is the **estimated increase per month**?

» How frequently do you do your **backups**? At how many **independent locations**?

» What are your local storage and backup **procedures**? Will this data require **secure storage**?

V. DATA SHARING, STANDARDS[4] & METADATA (WITH EXTERNALS)

» Are you using a **file format** that is standard to your field? If not, how will you document the alternative you are using?

» Which **methodology** and **standards** will be applied? *e.g. NetCDF (Network Common Data Format) files (for multidimensional spatio-temporal data)*

» What tools or software are required to **read or view the data**?

» What **supporting documentation** will you be creating and make publicly available in order to make the data understandable by other researchers and support the longer-term re-use of the data?

» Are you using **metadata** that is standard to your field? How will the metadata be managed and stored?

It is essential to associate datasets with metadata so that other researchers can understand how the data was collected and under which conditions it can be reused.

VI. DATA RESTRICTIONS

» **How open** will the data and outputs be?

» Will you be dealing with **sensitive/personal/restricted data and why**? *e.g. spatial/temporal information about endangered species; personal information from interviews;…*

» Do you expect there will be any **restrictions** on how the data can be accessed or reused (e.g. due to Intellectual Property Rights (IPRs))?

» Does sharing the data raise **privacy, ethical, legal, or other confidentiality**[5] concerns? (Please note that there may be country-dependent rules when it comes to these issues).

» Do you have a plan to protect or anonymize data, if needed?

Data should be as open as possible, though with restricted or closed access where appropriate and necessary; for example, if there are sensitive data involving human subjects. Depending on the nature of the research, the degrees of data openness may vary, extending from fully open to strictly confidential data.

The reason for restricting the access or use of some data should be explained and justified.

As for personal data, make sure it complies with the EU General Data Protection Regulation (see 2.5 The EU GDPR and Open Data).

4. There are many standards options for (meta)data. A few of them are listed under 4.3 Standards of the guidance document.
5. For more information, please visit: https://libraries.mit.edu/data-management/share/confidentiality/

VII. DATA PUBLISHING & LICENSING

» Where and how will the data be published?

» Under which **licences** will the data be published? Have **possible licensing** issues been considered?

» Will the research be published in a journal that requires the **underlying data** to accompany articles?

» Will there be any **embargoes** on the data? If yes, explain why, until when, and what happens when the embargo is over.

» What are your intended **Open Access publications practices** (Green, Golden, Hybrid[6])?

Data should be made available as soon as the results of the research have been published.

VIII. DATA ARCHIVING (AFTER THE PROJECT ENDS)

» How will the data be managed **after the project ends** to ensure their long-term availability?

» Will the data be published with a **Digital Object Identifier** (DOI) and/or be placed in a **recognized long-term repository** or data centre, and when will this take place?

» How will you be **archiving the data**? Will you be storing it in an archive or repository for long-term access? If not, how will you preserve access to the data?

» How will you prepare data for **preservation or sharing**? Will the data need to be anonymized or converted to more stable file formats?

» Are **software or tools** needed to use the data? Will these be archived?

» **How long** should the data be retained? 3-5 years, 10 years, or forever?

Be specific and complete in explaining WHEN and FOR HOW LONG exactly the data will be available (e.g. after project ends, during 5 years,...) and WHERE (e. g. name of repository or website). Vague statements like "could be/may be/we plan to…" should be avoided and replaced with more accurate terms.

The DMP should include the full list of data storages/ repositories/catalogues/websites (with working hyperlinks).

It is highly recommended to use Digital Object Identifiers (DOIs).

We encourage projects to store the data for as long as it is possible: 5-10 years storage is seen as the minimum period of time.

IX. COSTS

» What are the estimated costs for managing your data and other materials during/after the project?

» How have you accounted for the costs to ensure long-term availability?

e.g. cloud hosting costs at www.example.com will be of 15,000€/year, and the budget is included in the project.

6. See Annex I — Glossary for a description of the different OA publication types. Licensing and publishing references and tools can be found under 4.4 Licensing, citing & publishing of the guidance document.

5. TOOLS & RESOURCES

5. TOOLS & RESOURCES

5.1. REPOSITORIES FOR DATASETS

The funded research projects should store and make available their projects' data via key national and international archives and storage services. Data should be submitted to discipline-specific, community-recognized repositories where possible, or to generalist repositories if no suitable community resource is available. Data management, listing and archiving services are provided by an increasing number of electronic repositories. Some institutions have their own repository (e.g. at the organisation's level).

The CoreTrustSeal organisation provides certification for repositories that want to validate their quality and trustworthiness, by following a series of requirements jointly developed by the Data Seal of Approval (DSA) and the World Data System of the International Science Council (ISC-WDS). The list of certified repositories is available on their website.

As a general rule, repositories ask payment for depositing and hosting research data, while downloading datasets remains free of charge. Projects may include the costs of data storing and sharing in the projects budgets, but this has to be done at the proposal stage and the eligibility of costs has to be checked by each national team with its funder.

The lists hereunder consists of a selection of recommended repositories: they are either general repositories, which are open to all fields of research, or centralised and domain-specific repositories, focussing on biodiversity research and related areas.

5.1.1. GENERAL REPOSITORIES

Data Archiving and Networked Services (DANS) provides the following services:

» DataverseNL, to store and share research data during the project.

» After a project ends, research data are sustainably stored and shared via the online archiving system EASY.

» Provide and add research projects and publications to the science portal NARCIS.

DANS also guides other archives, research institutes and research financiers to questions relating to data management, certification and subjects such as FAIR, open access and software sustainability.

DataHub allows the creation of tools and applications for data; it also hosts thousands of datasets.

Dataverse is an open source research data repository software installed in many institutions worldwide. The Harvard Dataverse is the largest repository (+ 60,000 datasets) and, although it is focussed on social sciences, it is open to researchers from all research fields.

The Dryad Digital Repository is a curated, general-purpose repository for a wide diversity of data types, based on open source software.

EUDAT offers heterogeneous research data management services and storage resources through a geographically distributed network across 15 countries. Services include: data hosting/registration/management/sharing; data management planning; data discovery; data access/interface/movement; identity and authorization.

figshare is an interface designed for academic research data management and research data dissemination. Users can deposit all file types (research papers, FAIR data, and non-traditional research outputs) and make them available in a citable, shareable, and discoverable manner.

Mendeley Data is an open research data repository, where researchers can upload and share their

research data. Datasets can be shared privately amongst individuals, as well as published to be shared with the world.

The OpenAIRE project website provides access links to publications, datasets, software and other research products from EU-funded projects.

The Registry of Research Data Repositories is a global registry of research data repositories that covers research data repositories from different academic disciplines.

Zenodo is a catch-all, open access research data repository for EC funded research, created in 2013 by OpenAIRE and CERN to provide a place for researchers to deposit datasets. It enables the sharing and showcasing of multidisciplinary research results.

5.1.2. REPOSITORIES FOR THE BIODIVERSITY AND ENVIRONMENTAL RESEARCH DOMAIN

» The Arctic Biodiversity Data Service (ABDS), is an online data portal that aims to provide a framework for aggregating and improving access to Arctic biodiversity data. It is an online, interoperable and circumpolar data and metadata management system.
→ Research field(s): arctic biodiversity

» The Dynamic Ecological Information Management System (DEIMS-SDR) is an information management system that allows the user to discover long-term ecosystem research sites around the globe, along with the data gathered at those sites and the people and networks associated with them.
→ Research field(s): ecology, environmental sciences

» The Earth Microbiome Project (EMP), is a global catalogue of microbial taxonomic and functional diversity on the planet.
→ Research field(s): genetic diversity, microbial diversity

» The GEOSS Portal offers a single entry point to store, share and unlock Earth Observation data (data, imagery and analytical software packages) from archives all over the world. It connects users to existing databases and portals and provides summarised and up-to-date information.

» In particular, **GEO BON** promotes the structuration and mobilization of biodiversity data (see GEO BON, 2017). Note that for Europe, the **EU BON** initiative proposes a European biodiversity portal (see http://biodiversity.eubon.eu/tools). The EU BON's Data Publishing and Dissemination Toolbox (DPDT) is a set of standards, guidelines, recommendations, workflows and tools designed to ease scholarly publishing of biodiversity-related data.
→ Research field(s): biodiversity for GEO BON and EU BON; more generally Earth sciences for GEOSS

» The German Federation For Biological Data (GFBio e.V.), a consortium of biodiversity relevant repositories (PANGAEA, EBI and eight repositories of collections and museums), is the national contact point for issues concerning the management and standardisation of biological and environmental research data during the entire data life cycle. It mediates expertise and services between the GFBio data centres and the scientific community.
→ Research field(s): biological and environmental research data

» The Global Biodiversity Information Facility (GBIF) is an international open data infrastructure allowing anyone, anywhere to access and share primary biodiversity data. The GBIF network uses Darwin Core standard, which forms the basis of GBIF's index of more than a billion of species occurrence records. Includes the GBIF Integrated Publishing Toolkit (IPT), which is a free open source software tool used to publish and share biodiversity datasets through the GBIF network.
→ Research field(s): biodiversity, taxonomy, species diversity and distribution

» The Knowledge Network for Biocomplexity (KNB) is an international repository intended to facilitate ecological and environmental research.
→ Research field(s): biodiversity, ecology, environmental sciences

» MoveBank is a free, online database of animal tracking data hosted by the Max Planck Institute for Ornithology. It contains the MoveBank Data Repository, designed for long-term data archiving. The Repository complements MoveBank's flexible tools for sharing, managing, and analysing animal movement data throughout all stages of research by providing a way to formally publish completed research datasets.
→ Research field(s): biodiversity, animal tracking

» The National Center for Biotechnology Information (NCBI), provides access to biomedical and genomic information (databases, software, literature, etc.). It includes the Sequence Read Archive (SRA) that stores raw sequence data from "next-generation" sequencing technologies; GenBank, an annotated collection of all publicly available DNA sequences; etc.
→ Research field(s): biodiversity (genetic diversity), biotechnology, bioinformatics

» The NERC Data Centres of The Natural Environment Research Council (NERC) are subject-based environmental Data Centres to store and distribute data from its own research programmes and data that are of general use to the environmental research community.
→ Research field(s): environmental & space sciences

» PANGAEA, an information system, is operated as an Open Access library aimed at archiving, publishing and distributing geo-referenced data from Earth system research. The system guarantees long-term availability of its content through a commitment of the hosting institutions.
→ Research field(s): Earth & environmental sciences

» The Swedish Species Gateway (Artportalen) is a Swedish species observation system and tool for gathering and providing information about species occurrence (plants, animals, fungi). Anyone (private individuals, scientists, authorities) can report species and search from the over 61 million records.
→ Research field(s): biodiversity

Note that more and more journals take into account the need to make data fully accessible. Further, the Biodiversity data journal (https://bdj.pensoft.net/) uses a narrative (text) and data integrated publishing workflow to stimulate mobilisation, review, publication, storage, dissemination, interoperability and re-use of biodiversity data.

Nature's Scientific Data provides on their website a list of recommended repositories, organised according research field, with a list dedicated to taxonomy and species diversity.

Many other repositories exist for specific fields (climate/weather, health, chemistry, social sciences, etc.) that could be used in the context of biodiversity research.

5.2. STANDARDS FOR BIODIVERSITY DATA

The wide use of standards by the biodiversity research community greatly improved the interoperability of published datasets. The most important biodiversity data standards are listed here, more information can be found on the TDWG website.

» Access to Biological Collections Data (ABCD) Data Exchange Standard is an evolving comprehensive standard for the access to and exchange of data about specimens and observations.

» DarwinCore are standards for sharing biodiversity data. It includes a glossary of terms intended to facilitate the sharing of information about biological diversity by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.

» The Ecological Metadata Language (EML) developed by the Knowledge Network for Biocomplexity (KNB) is a metadata standard developed for the Earth, environmental and ecological sciences.

» Frictionless Data (by Open Knowledge International) shortens the path from data to insight with a collection of specifications and software for the publication, transport, and consumption of data.

» Open Geospatial Consortium (OGC) standards developed open standards for geospatial content and services, sensor web and Internet of Things, GIS data processing and data sharing.

5.3. LICENSING, CITING & PUBLISHING

Open data publishing practices can increase the visibility of your research and can help your data sets become a valuable output. Open science publishing does not mean that a researcher is denied credit for their work. In many cases, visibility is increased with open access licenses such as the Creative Commons Attribution (CCBY), which preserves credit.

The e-Infrastructures and Data Management Project has been working with publishers to develop a **Data Accessibility Statement** (DAS) that links publications to their underlying data sets (Belmont Forum e-I&DM, 2018c).

The **Plan S** is an initiative for open-access science publishing that was launched by Science Europe and a consortium of funding agencies in September 2018. The plan requires scientists and researchers who benefit from state-funded research organisations and institutions to grant open access to their publications as of 2020. (Science Europe, 2019)

Below is a list of resources to better understand and address licensing and other publication considerations.

» Creative Commons (CC) is a non-profit corporation providing different types of copyright-licenses, known as Creative Commons licenses, free of charge to the public.

» DataCite is an organisation that provides persistent identifiers (DOIs) for research data.

» Directory of Open Access Journals (DOAJ), is a community-curated online directory that indexes and provides access to over 9,000 open access Journals.

» Open Data Commons, part of Open Knowledge International, provides three types of licenses: Public Domain Dedication and License (PDDL), Attribution License (ODC-By), and Open Database License (ODC-ODbL).

» ORCID (Open Researcher and Contributor ID) provides persistent digital identifiers for researchers and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between researchers and professional activities ensuring that their work is recognized.

5.4. OTHER TOOLS & RESOURCES

Hereunder is a list of other tools that might be useful either because they include a range of different services (e.g. resources, tools, advice,...) or because they focus on a particular aspect that do not fit in the sections above (e.g. impact of research).

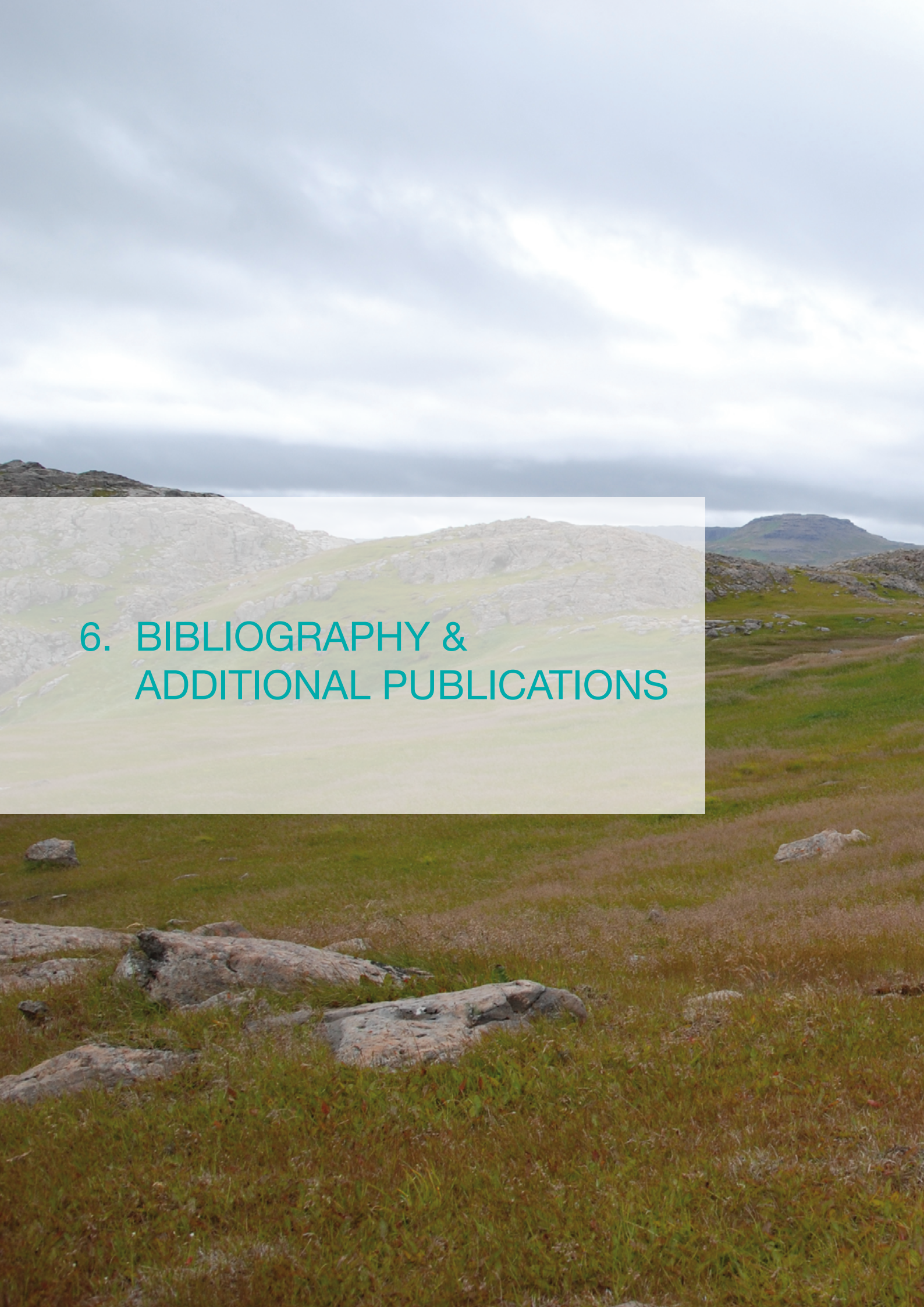
» Data Observation Network for Earth (DataONE) provides access to data across multiple member repositories, supporting search and discovery of Earth and environmental data. DataONE also provides best practices in data management through educational resources and materials.

» The FAIR self-assessment tool, developed by the Australian Research Data Commons, give researchers the possibility to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).

» FAIRsharing is a curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.

» ImpactStory is an open-source website that helps researchers explore and share the online impact of their research.

» Open Science Framework (OSF) is a tool to connect the entire research process, also allowing for 'controlled access' of project data making it easy to collaborate beyond the consortium.



6. BIBLIOGRAPHY & ADDITIONAL PUBLICATIONS

6. BIBLIOGRAPHY & ADDITIONAL PUBLICATIONS

You will find in this section a list of references and sources on data management practices and open data. Subsection 5.1 includes research papers and documents like guidelines and toolkits and policy documents, while sub-section 5.2 lists webpages and information websites. This section also serves as bibliography for the references cited in this guidance document (but all these references are not necessarily cited in the text).

6.1. PUBLICATIONS AND DOCUMENTS

Belmont Forum e-I&DM (2018a). Belmont Forum e-Infrastructures & Data Management Toolkit. Last update 11 December 2018.

https://bfe-inf.github.io/toolkit/index.html

Belmont Forum e-I&DM (2018b). Data and Digital Outputs Management Annex. May 2018.

https://www.belmontforum.org/archives/resources/data-and-digital-outputs-management-annex-full

Belmont Forum e-I&DM (2018c). Belmont Forum Data Accessibility Statement and Policy. October 2018.

https://www.belmontforum.org/archives/resources/data-accessibility-statement-das

Belmont Forum e-I&DM (2019). Data and Digital Outputs Management Plan (DDOMP) Guide: A Step-By-Step User Guide for Building a Successful Data Management Plan. Last update 10 February 2019.

https://www.belmontforum.org/archives/resources/data-and-digital-outputs-management-plan-ddomp

BiodivERsA (2016). BiodivERsA strategic research and innovation agenda (2017-2020) - Biodiversity: a natural heritage to conserve, and a fundamental asset for ecosystem services and Nature-based Solutions tackling pressing societal challenges. 86 pp.

https://www.biodiversa.org/968

Bishop, B. W., & Grubesic, T. H. (2016). Geographic Information: Organization, Access, and Use. New York: Springer. https://www.springer.com/gp/book/9783319227887

CrowdFlower (2016). CrowdFlower Data Science Report 2016. https://www.kdnuggets.com/2016/04/crowdflower-2016-data-science-repost.html

Deutsche Forschungsgemeinschaft - DFG (2015). Guidelines On The Handling Of Research Data In Biodiversity Research. Working Group on Data Management of the DFG Senate Commission on Biodiversity Research. Information for Researchers No. 36. https://www.dfg.de/en/research_funding/announcements_proposals/2015/info_wissenschaft_15_36/index.html

Dunning A., de Smaele M. & Böhmer J. (2017). Are the FAIR Data Principles fair? International Journal of Digital Curation: vol 12 No 2 (2017).

https://doi.org/10.2218/ijdc.v12i2.567

European Commission (2016a). Open Innovation, Open Science, Open to the World – a vision for Europe. DG R&I report, 104 pp.

http://publications.europa.eu/resource/cellar/3213b335-1cbc-11e6-ba9a-01aa75ed71a1.0001.02/DOC_2

European Commission (2016b). Guidelines on FAIR data management in Horizon 2020 (3rd ed.). European Commission - Directorate-General for Research and Innovation. Version 3.0. 26 July 2016.

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission (2016c). H2020 Online Manual – Open access & Data management. Last update 26 January 2016.

http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

European Commission (2016d). Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 3.1. 25 August 2016.

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

European Commission (2016e). Horizon 2020 templates: Data management plan (DMP) v1.0 – 13.10.2016.

http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx

European Commission (2018a). Turning FAIR into reality. Final report and action plan from the European Commission expert group on FAIR data.

https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283

Groom Q., Brosens D., Adriaens T., Vanderhoeven S. (2017). TrIAS Data Management Plan. Version 1.2.

https://osf.io/kg3av/

Hansen K. K., Buss M., Haahr L. S. (2018). A FAIRy tale: A fake story in a trustworthy guide to the FAIR principles for research data. http://doi.org/10.5281/zenodo.2248200

Herrema A. (2014). Data? uhh..overthere. Publication and data. Het Bouwteam.https://www.fosteropenscience.eu/content/cartoonpublication-and-data

Jomier J. (2017). Open science: Towards reproducible research. Information Services & Use, 37(3), 361–367. https://doi.org/10.3233/ISU-170846

Jones S. (2011). How to Develop a Data Management and Sharing Plan. DCC How-to Guides. Edinburgh: Digital Curation Centre.

http://www.dcc.ac.uk/resources/how-guides

LIBER (2017). Implementing FAIR Data Principles : The Role of Libraries. LIBER (Association of European Research Libraries) factsheet.

https://libereurope.eu/article/implementing-fair-data-principles-role-libraries/

Michener W.K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol 11(10): e1004525.

doi:10.1371/journal.pcbi.1004525

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004525

Milotic T., Desmet P., Van Hoey S. (2018). *Good enough practices in data management and how to translate these in data management plans (DMPs).* Technical note, 28 September 2018.

https://doi.org/10.5281/zenodo.1421739

Nature Editorials (2018). Everyone needs a data-management plan. Nature, 555(7696), 286–286.

https://doi.org/10.1038/d41586-018-03065-z

OECD (2015). Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, 25. Paris: OECD Publishing.

https://doi.org/10.1787/5jrs2f963zs1-en

Raymond, C. (2019). Privacy Policy/Notice for scientific research: EU General Data Protection Regulation, Art. 12–14. University of Helsinki, 22nd May, 2019. Available at: https://www2.helsinki.fi/sites/default/files/atoms/files/privacy-notice_national-security-survey-2019.pdf

Schiermeier, Q. (2018). Data management made simple. Nature, 555(7696), 403–405.

https://doi.org/10.1038/d41586-018-03071-1

Schmidt B., Gemeinholzer B., Treloar A. (2016). Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. PLOS ONE 11(1): e0146695. https://doi.org/10.1371/journal.pone.0146695

Science Europe (2018). Presenting a framework for discipline-specific research data management. Science Europe Guidance Document D/2018/13.324/1.

https://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf

Stall S., et al. (2018). Data sharing and citations: New author guidelines promoting open and FAIR data in the Earth, space, and environmental sciences. Science Editor 2018;41(3):83-87.

https://www.csescienceeditor.org/article/data-sharing-and-citations-new-author-guidelines-promoting-open-and-fair-data-in-the-earth-space-and-environmental-sciences

Swiss National Science Foundation - SNSF (2017a). Data Management Plan (DMP) - Guidelines for researchers.

http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/data-management-plan-dmp-guidelines-for-researchers.aspx

Swiss National Science Foundation - SNSF (2017b). SNSF Explanation of the FAIR data principles.

http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

Wieczorek J., Bloom D., Guralnick R., Blum S., Döring M., Giovanni R., et al. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715.

https://doi.org/10.1371/journal.pone.0029715

Wilkinson M. D., Dumontier M., Aalbersberg, I. J., Appleton G., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3, 160018*. http://doi.org/10.1038/sdata.2016.18

6.2. WEB PAGES & WEB RESOURCES

Belmont Forum e-Infrastructures & Data Management (2015). Open Data Policy and Principles. Adopted at the Belmont Forum's 10th Plenary Meeting, Oslo, 2015. Retrieved from: https://www.belmontforum.org/about/open-data-policy-and-principles

Belmont Forum e-Infrastructures & Data Management (2017): www.bfe-inf.org

European Commission (2014). Open Science. Last update 28 October 2014. Retrieved from:

http://ec.europa.eu/research/openscience/index.cfm?pg=home

European Commission (2016). OpenAIRE (2016). How to create a Data Management Plan? Nov 2016. Retrieved from:

https://www.openaire.eu/how-to-create-a-data-management-plan

European Commission (2017). Facts and Figures for open research data. Last update 20 May 2017. Retrieved from:

https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en

European Commission (2018b). Data protection. Retrieved from:

https://ec.europa.eu/info/law/law-topic/data-protection_en

European Data Portal (2016). What is Open Data? Last update 13 May 2016. Retrieved from:

https://www.europeandataportal.eu/elearning/en/module1/#/id/co-01

European Data Portal (2018). Protecting data and opening data. Retrieved from:

https://www.europeandataportal.eu/fr/highlights/protecting-data-and-opening-data

European Union Open Data Portal (2016). Open Research Data (ORD) - the uptake in Horizon 2020. Last update 10 May 2016. Retrieved from:

https://data.europa.eu/euodp/en/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020

French Agency for Research – ANR (2018). La politique science ouverte de l'ANR. Retrieved from:

https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/

FORCE11 (2014). The Fair Data Principles. Retrieved from: https://www.force11.org/group/fairgroup/fairprinciples

Foster (2018). Open Source in Open Science. Last update 23 June 2018. Retrieved from:

https://www.fosteropenscience.eu/foster-taxonomy/open-source-open-science

GEO BON (2017). A roadmap for the mobilization of structured biodiversity data. 2 July 2017. Retrieved from:

https://geobon.org/a-roadmap-for-the-mobilization-of-structured-biodiversity-data

GO FAIR (2017). FAIR Principles. Last update 24 December 2017. Retrieved from:

https://www.go-fair.org/fair-principles/

JAMK (2017). Open Science and Research in JAMK. JAMK University of Applied Sciences. Retrieved from:

http://www.jamk.fi/en/Research-and-Development/open-science-and-research-in-jamkavoin-tiede-ja-tutkimus/open-science-and-research-in-jamk/

MIT Libraries (2014). Write a Data Management Plan. Published on 16 December 2014. Retrieved from:

https://libraries.mit.edu/data-management/plan/write

Open Access Belgium (2015). Glossary. Last update 7 September 2015. Retrieved from:

https://openaccess.be/support/glossary

Open Knowledge International (2019). Open Data Handbook: Guides, case studies and resources for government & civil society on the "what, why & how" of open data. First issued in 2010, last update in February 2019. Retrieved from: http://opendatahandbook.org

OpenAIRE (2016). How to create a Data Management Plan? Nov 2016. Retrieved from:

https://www.openaire.eu/how-to-create-a-data-management-plan

OpenAIRE (2018). What is a Data Management Plan (DMP) and how do I create one? Last update 28 September 2018. Retrieved from:

https://www.openaire.eu/what-isa-data-management-plan-and-how-do-i-create-one

Science Europe (2019). Plan S. Making full and immediate Open Access a reality. Coalition S website. Retrieved from: https://www.coalition-s.org

SpringerNature (2018). Research Data Policies: Data Availability Statements. Retrieved from:

https://www.springernature.com/gp/authors/research-data-policy/data-availability-statements/12330880

University of Cambridge (2015). Crafting your data management plan. Published on 24 March 2015. Retrieved from:

https://www.data.cam.ac.uk/data-management-guide/creating-your-data/data-management-plan

We hope that this guidance document, developed by BiodivERsA and the Belmont Forum, will help you improve the way you handle, store and make accessible on the long term the data you produced during your research projects.

The guide should be viewed as a living document. Thus, would you have any suggestion to complement or amend it, please contact the DMP Working Group (l.goudeseune@biodiversity.be).

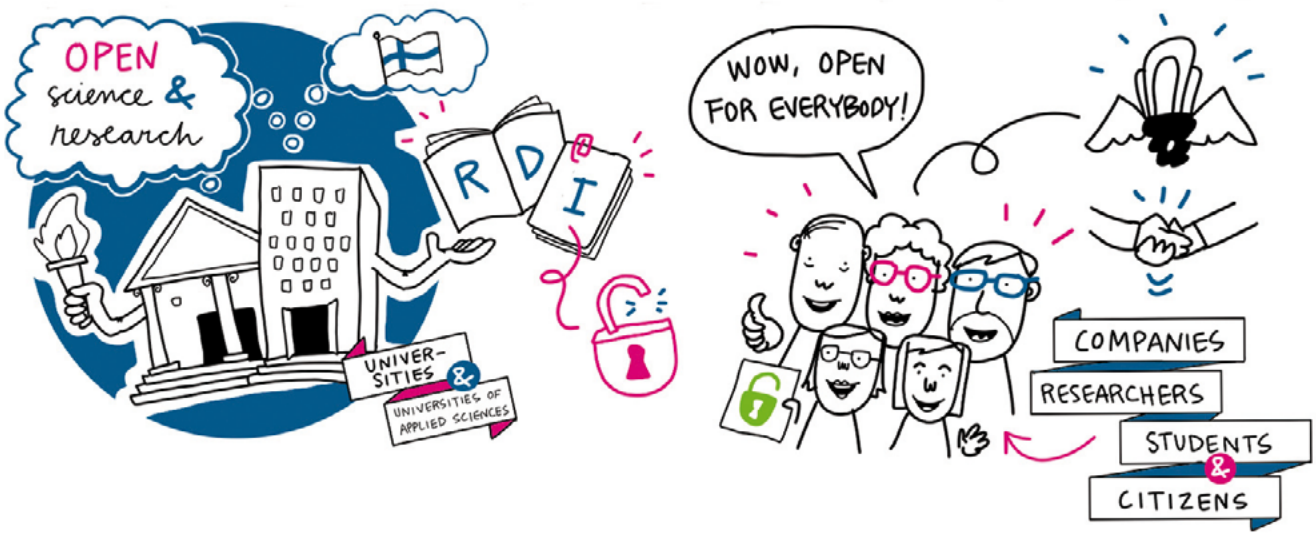


Figure 4: Open data contributes greatly to open science which should have more impact on a broader range of stakeholders (image taken from: JAMK, 2017; artist: Linda Saukko-Rauta)

ANNEX I: GLOSSARY


6
PS

ANNEX I: GLOSSARY

You will find here is a list of recurring terms and names related to data management that are used and referred to in this guidance document. See also the EU Open Data Portal glossary as well as the Glossary from Open Access Belgium.

Access to Biological Collections Data (ABCD) **Data Exchange Standard** is an evolving comprehensive standard for the access to and exchange of data about specimens and observations.

https://www.tdwg.org/standards/abcd

Biodiversity Information Standards (TDWG) is a not-for-profit scientific and educational association that works to develop open standards for the exchange of biodiversity data, facilitating biodiversity informatics.

https://www.tdwg.org/

Creative Commons (CC licence) is a non-profit corporation dedicated to making it easier for people to share and build upon the work of others, consistent with the rules of copyright. Six licenses are possible, combining four basic elements: the attribution, the derivatives, the commercial use, and the 'share-alike' principle. This allows the creator to mark his/her work with the freedom he/she wants it to carry, so others can share, remix, use commercially, or any combination thereof.

https://creativecommons.org

DarwinCore are standards developed for sharing biodiversity data. It includes a glossary of terms intended to facilitate the sharing of information about biological diversity by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.

https://www.tdwg.org/standards/dwc; Wieczorek et al, 2012

Data Availability Statements (DAS) provide a statement about where data supporting the results reported in a published article can be found - including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. The Belmont Forum has a Data Accessibility Statement that was developed in conjunction with the science publishing community - including Springer Nature. (SpringerNature, 2018)

Data and Digital Outputs Management Plan (DDOMP) is the updated version of the "data policy" developed by the Belmont Forum. (https://www.belmontforum.org/archives/resources/data-and-digital-outputs-management-annex-full)

DMP is the commonly used acronym for **Data Management Plan**. It is a formal document that specifies how research data will be handled both during and after a research project. It identifies key actions and strategies to ensure that research data are of a high quality, safe, sustainable and – where possible – accessible and reusable. A DMP should be considered a 'living' document - it is ideally created before or at the start of a research project, but updated when necessary as the project progresses. Planning for data management is therefore not a one-off event, but a process. (OpenAIRE, 2018)

DOI stands for **Digital Object Identifier**, a code used to permanently and stably identify (usually digital) objects. DOIs provide a standard mechanism for retrieval of metadata about the object, and generally a means to access the data object itself (Wilkinson et al, 2016).

Ecological Metadata Language (EML) is a metadata standard developed by the Knowledge Network for Biocomplexity (KNB) for Earth, environmental and ecological sciences. It allows a full description of the dataset, including methods for data collection.

https://eml.ecoinformatics.org

FAIR means the Guiding Principles for scientific data management and stewardship (published in Wilkinson et al, 2016). According to FAIR, data needs to be Findable, Accessible, Interoperable, and Reusable.

Gold Open Access (or "Open Access publishing"): providing immediate and permanent Open Access to the articles published in Open Access journals.

https://openaccess.be/support/glossary

Green Open Access (or "self-archiving"): providing Open Access by publishing in any journal, and then deposit an Open Access version of the publication in a repository.

https://openaccess.be/support/glossary

Hybrid Open Access: some publishers allow immediate Open Access to some of their articles on condition that an 'Article Processing Charge' (APC) is paid. In such journals, two types of articles coexist: those freely accessible and others only accessible through a subscription. Although being mainly responsible for the recent rise in available Open Access publications, 'hybrid Open Access' doesn't take away the need for paying subscriptions.

https://openaccess.be/support/glossary

Metadata: set of structured data describing physical or digital resources. They are an essential element for sharing information about publications and data sets. Set standards for metadata are essential for the interoperability between electronic resources. Metadata are categorized in descriptive, administrative and structural metadata.

https://openaccess.be/support/glossary

Open Access (OA) can be defined as the practice of providing online access to scientific information that is free of charge to the user and that is reusable. Open access to 'scientific information' refers to two main categories:

- → Peer-reviewed scientific publications (primarily research articles published in academic journals)
- → Scientific research data: data underlying publications and/or other data (such as curated but unpublished datasets or raw data)

https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-access_en

---

Open Data is the idea that data should be freely available to everyone to use and re-publish as they wish, without restrictions from copyright, patents or other mechanisms of control. Within the context of scientific publishing, it refers to the principle of not only providing access to open access publications, but also to the underlying datasets, on which the research is based.

https://openaccess.be/support/glossary/

Open Geospatial Consortium (OGC) standards: open standards for geospatial content and services, sensor web and Internet of Things, GIS data processing and data sharing.

http://www.opengeospatial.org/standards

Open Science: to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction (OECD, 2015).

OpenAIRE is a EU-funded project to support the implementation of Open Access in Europe. It has started in 2009 and is now in its third phase. The OpenAIRE website provides access links to publications, datasets, software and other research products from EU-funded projects. OpenAIRE also offers services such as: Open Science training, OpenAIRE Mining Service, Open Science Helpdesk, etc.

https://www.openaire.eu

Plan S is an initiative for Open Access publishing that was launched in September 2018. The plan is supported by cOAlition S, an international consortium of research funders. Plan S requires that, from 2020, scientific publications that result from research funded by public grants must be published in compliant Open Access journals or platforms.

https://www.coalition-s.org/

Research Data Management (RDM) concerns the organization of data, from its entry to the research cycle to the dissemination and archiving of valuable results.

https://www2.le.ac.uk/services/research-data/rdm/what-is-rdm

SPARC is a membership organisation to promote open data, but also open access and open education.

https://sparcopen.org

ANNEX II: LINKS TO (AND INFORMATION ON) NATIONAL AND FUNDERS' OPEN DATA/ OPEN ACCESS POLICIES

ANNEX II: LINKS TO (AND INFORMATION ON) NATIONAL AND FUNDERS' OPEN DATA/OPEN ACCESS POLICIES

You will find here is a list of recurring terms and names related to data management that are used and referred to in this guidance document. See also the EU Open Data Portal glossary as well as the Glossary from Open Access Belgium.

AUSTRIA

FWF - Open Access policy: http://www.fwf.ac.at/en/research-funding/open-access-policy/

FWF - Research Data Management: https://www.fwf.ac.at/en/research-funding/open-access-policy/research-data-management

For a summary of the on open data/access initiatives in Austria, see OpenAIRE's National Open Access Desk for Austria: https://www.openaire.eu/item/austria

BELGIUM

BELSPO Open Science Mandate: https://www.belspo.be/belspo/openscience/mandate_en.stm

FWO - Open Access policy: https://www.fwo.be/en/the-fwo/research-policy/open-access/

F.R.S.-FNRS – Open Access policy: https://www.frs-fnrs.be/docs/Reglement_OPEN_ACCESS_EN.pdf

Open Access and Open Research Data Funder Policies (BelSPO, FWO, FRS-FNRS): https://openaccess.be/open-access-in-belgium/policies-mandates-in-belgium/

Belgian Open Access legislation: https://openaccess.be/belgian-open-access-legislation/

Brussels Declaration on Open Access (2012): https://openaccessbelgium.files.wordpress.com/2012/11/signedbrussels-declaration-on-open-access.pdf

For a summary of the on open data/access initiatives in Belgium, see OpenAIRE's National Open Access Desk for Belgium: https://www.openaire.eu/item/belgium

BRAZIL

FAPESP - Plano de Gestão de Dados: http://www.fapesp.br/gestaodedados/ and https://fapesp.br/12632/portaria-cta-no-012019.

BULGARIA

For a summary of the on open data/access initiatives in Bulgaria, see OpenAIRE's National Open Access Desk for Bulgaria: https://www.openaire.eu/item/bulgaria

CANADA

NSERC - Policies and Guidelines Open Access: http://www.science.gc.ca/eic/site/063.nsf/eng/h_547652FB.html

FRQ-The Fonds de recherche du Québec Policy on Open Access: http://www.frqnt.gouv.qc.ca/en/science-ouverte

CZECH REPUBLIC

For a summary of the open data/access initiatives in the Czech Republic, see OpenAIRE's National Open Access Desk for the Czech Republic: https://www.openaire.eu/os-czech-republic

DENMARK

Denmark's National Strategy for Open Access: https://ufm.dk/en/research-and-innovation/cooperation-between-research-and-innovation/open-access/Publications/denmarks-national-strategy-for-open-access/national-strategy-for-open-access-english.pdf

For a summary of the on open data/access initiatives in Denmark, see OpenAIRE's National Open Access Desk for Denmark: https://www.openaire.eu/os-denmark

ESTONIA

For a summary of the on open data/access initiatives in Estonia, see OpenAIRE's National Open Access Desk for Estonia: https://www.openaire.eu/item/estonia

FINLAND

AKA - Open Science: https://www.aka.fi/en/research-funding/responsible-science/open-science/

Key documents of Open Science and Research Initiative: https://avointiede.fi/en/policies/policies-open-science-and-research-finland/open-science-and-research-initiative

For a summary of the on open data/access initiatives in Finland, see OpenAIRE's National Open Access Desk for Finland: https://www.openaire.eu/item/finland

FRANCE

ANR policy on Open Science: https://anr.fr/en/anrs-role-in-research/values-and-commitments/open-science/

National Plan for Open Science: https://cache.media.enseignementsup-recherche.gouv.fr/file/Recherche/50/1/SO_A4_2018_EN_01_leger_982501.pdf

Minister of Higher Education, Research and Innovation – Open Science: https://www.ouvrirlascience.fr/category/open_science/

For a summary of the on open data/access initiatives in France, see OpenAIRE's National Open Access Desk for France: https://www.openaire.eu/item/france

GERMANY

BMBF - Open Access Strategy: https://www.bmbf.de/de/open-access-das-urheberrecht-muss-der-wissenschaft-dienen-846.html

DFG – Open Access: www.dfg.de/lis/openaccess/en

DFG – Handling of Research Data: www.dfg.de/proposal_process/research_data

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities: https://openaccess.mpg.de/Berliner-Erklaerung

For a summary of the on open data/access initiatives in Germany, see OpenAIRE's National Open Access Desk for Germany: https://www.openaire.eu/item/germany

National Research Data Infrastructure: NFDI4BioDiversity is a consortium dedicated to deliver practical tools for researchers and improve data availability for environmental studies/ecology. It is part of a greater network within the German National Research Data Infrastructure (NFDI).

GREECE

The Greek Law already supports Open Data activities with provisions for public bodies' infrastructures and introduction of Creative Common licenses. Open Access policies are progressively adopted by Higher Education Institution (HEI) of Greece. Public funders have not yet adopted Open Access policies. Nonetheless, Greece is currently developing a national Open Access/Open Science action plan.

For a summary of the on open data/access initiatives in Greece, see Open AIRE's National Open Access Desk for Greece: https://www.openaire.eu/item/greece

IRELAND

Irish EPA - Research Data Archive: http://www.epa.ie/researchandeducation/research/safer/

Ireland's National Principles for Open Access Policy Statement: https://www.dri.ie/sites/default/files/files/National%20Principles%20on%20Open%20Access%20Policy%20Statement%20(FINAL%2023%20Oct%202012%20).pdf

For a summary of the on open data/access initiatives in Ireland, see OpenAIRE's National Open Access Desk for Ireland: https://www.openaire.eu/item/ireland

LITHUANIA

RCL - Guidelines on Open Access to Scientific Publications and Data: https://lmt.lt/data/public/uploads/2016/09/eng_-atvira-prieiga-_-galutinis.pdf

For a summary of the on open data/access initiatives in Lithuania, see OpenAIRE's National Open Access Desk for Lithuania: https://www.openaire.eu/item/lithuania

NORWAY

Norwegian Research Council - Open Data policies: Open access to research data (forskningsradet.no)

National goals and guidelines for open access to research articles: https://www.regjeringen.no/en/dokumenter/national-goals-and-guidelines-for-open-access-to-research-articles/id2567591/#_ftnref2

For a summary of the on open data/access initiatives in Norway, see OpenAIRE's National Open Access Desk for Norway: https://www.openaire.eu/item/norway

POLAND

NCN - Open Access Policy: https://ncn.gov.pl/finansowanie-nauki/otwarta-nauka?language=en

Ministry of Science and Higher Education - Open access to scientific publications: https://www.gov.pl/web/edukacja-i-nauka/otwarty-dostep-do-publikacji-naukowych

For a summary of the on open data/access initiatives in Poland, see OpenAIRE's National Open Access Desk for Poland: https://www.openaire.eu/item/poland

PORTUGAL

FCT - Open Access Policy: https://www.fct.pt/acessoaberto/

For a summary of the on open data/access initiatives in Portugal, see OpenAIRE's National Open Access Desk for Portugal: https://www.openaire.eu/os-portugal

ROMANIA

Open Government Partnership – National Action Plan 2018-2020: https://www.opengovpartnership.org/wp-content/uploads/2018/11/Romania_Action-Plan_2018-2020_EN.pdf

Currently, the national strategic framework for Open Science and Open Access is being developed: the initiative is carried out by experts from the Open Science Knowledge Hub-UEFISCDI, part of a project financed through structural funds, developed in partnership with of the Ministry of Research, Innovation and Digitization.

The Open Science team inside UEFISCDI - Open Science Knowledge Hub Romania (OSKH) - is the OpenAIRE NOAD, RDA Node Romania and involved in the National Initiatives for Open Science (NI4OS) – EC funded project to support the EOSC development. Also, UEFISCDI is part of the EOSC Association.

For a summary of the on open data/access initiatives in Romania, see OpenAIRE's National Open Access Desk for Romania: https://www.openaire.eu/item/romania

SLOVAKIA

For a summary of the on open data/access initiatives in Slovakia, see OpenAIRE's National Open Access Desk for Slovakia: https://www.openaire.eu/item/slovakia

SPAIN

Spanish State Scientific, Technical and Innovation Research Plan (PEICTI)

https://www.ciencia.gob.es/Estrategias-y-Planes/Planes-y-programas.html

For a summary of the on open data/access initiatives in Spain, see OpenAIRE's National Open Access Desk for Spain: https://www.openaire.eu/item/spain

SWEDEN

Formas policy on Open access: Open science - Formas

A Formas webpage that includes information on DMP: Good to know before you apply - Formas

The Swedish Research Council (Vetenskapsrådet) two links on open access and DMP:

https://www.vr.se/english/mandates/open-science/open-access-to-research-data.html

https://www.vr.se/english/applying-for-funding/requirements-terms-and-conditions/producing-a-data-management-plan.html

The Swedish National Library:

Open Access and Bibsam Consortium – Kungliga biblioteket – Sveriges nationalbibliotek – kb.se

For a summary of the on open data/access initiatives in Sweden, see OpenAIRE's National Open Access Desk for Sweden: https://www.openaire.eu/item/sweden

SWITZERLAND

SNSF policy on Open Research Data: http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx

Swiss National Strategy on Open Access: https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/Open_Access/Open_Access_strategy_final_e.pdf

For a summary of the on open data/access initiatives in Switzerland, see OpenAIRE's National Open Access Desk for Switzerland: https://www.openaire.eu/item/switzerland

THE BELMONT FORUM

Data Accessibility Statement Policy: https://zenodo.org/record/1476871#.XMIAH6LwKGR

THE NETHERLANDS

NWO - Open (FAIR) data: https://www.nwo.nl/en/policies/open+science/data+management

National Plan Open Science: https://www.openscience.nl/en

For a summary of the open data/access initiatives in the Netherlands, see OpenAIRE's National Open Access Desk for the Netherlands: https://www.openaire.eu/item/netherlands

TURKEY

For a summary of the on open data/access initiatives in Turkey, see OpenAIRE's National Open Access Desk for Turkey: https://www.openaire.eu/item/turkey

UNITED STATES OF AMERICA

NSF – Data Sharing Policy: https://www.nsf.gov/bfa/dias/policy/dmp.jsp

This guidance document was developed part of the BiodivScen programme jointly implemented by the Belmont Forum and BiodivERsA.

DOI : 10.5281/zenodo.3448251

The BiodivScen project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776617.