

University of Luxembourg

Research Data Management and Data Management Plans



UNIVERSITÉ DU
LUXEMBOURG

MAISON DU SAVOIR

Research Data Management & Data Management Plans

Research Facilitators

Jonathan England

Librarian for Scholarly Communication and Research Process;
ORBi^{lu} coordinator



jonathan.english@uni.lu



[@jonatorvue](https://twitter.com/jonatorvue)



<https://orcid.org/0000-0001-6715-8628>

What is research data?

Definitions

- DATA = “factual information used as a basis for reasoning, discussion or calculation” ([Merriam-Webster](#))
- RESEARCH DATA = “recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings” ([EPSRC](#))
- **RESEARCH DATA = all of the information that you use as an integral part of your research**
- Are excluded: incidental or administrative data generated in the course of personal activities, desktop or mailbox backup, or data produced by non-research activities (teaching, administrative tasks) [[LINK](#)]



Digitalbevaring.dk

Types of research data

- *Observational*: data captured in real time that is usually unique and irreplaceable (e.g. remote sensing data, survey data, field recordings, sample data).
- *Experimental*: data captured from lab equipment that is often reproducible (e.g. gene sequences, magnetic field data, electron microscope images)
- *Models or simulation*: data generated from test models where model and metadata may be more important than output data from the model (e.g. climate models, economic models).
- *Derived or compiled*: resulting from processing or combining 'raw' data (e.g. text and data mining, compiled databases, 3D models).
- *Reference or canonical*: a static or organic conglomeration or collection of datasets, probably published and curated (e.g. gene sequence databanks, collection of letters or archive of historical images).

https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained



https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained

Examples of research data

- Documents (text, MS Word), spreadsheets
- Scanned laboratory notebooks, field notebooks, diaries
- Online questionnaires, transcripts, surveys or codebooks
- Digital audiotapes, videotapes and other digital recording media
- Scanned photographs or films
- Transcribed test responses
- Database contents (video, audio, text, images)
- Digital models, algorithms, scripts
- Contents of an application (input, output, logfiles for analysis software, simulations)
- Documented methodologies and workflows
- Records of standard operating procedures and protocols



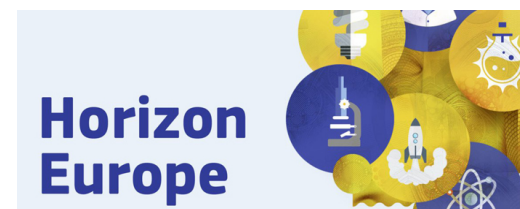
Examples of research data formats

- Text - flat text files, Word, Portable Document Format (PDF), Rich Text Format (RTF), Extensible Markup Language (XML).
- Numerical - Statistical Package for the Social Sciences (SPSS), Stata, Excel.
- Multimedia - jpeg, tiff, dicom, mpeg, quicktime.
- Models - 3D, statistical.
- Software - Java, C.
- Discipline specific - Flexible Image Transport System (FITS) in astronomy, Crystallographic Information File (CIF) in chemistry.
- Instrument specific - Olympus Confocal Microscope Data Format, Carl Zeiss Digital Microscopic Image Format (ZVI).

What is a DMP?

A formal ‘living’ document

- 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.
- More and more research funders require a DMP as part of the grant proposal process, or after funding has been approved.
- It is a ‘living’ document that is ideally created before or at the start of a research project, but updated when necessary as the project progresses. It’s a process...



[OpenAIRE FAQ - https://www.openaire.eu/faqs#ifaqCat-170](https://www.openaire.eu/faqs#ifaqCat-170)



What makes a good DMP?

- No universal DMP template and not all funders provide with a template to use
- There are no absolute right answers
- But be **clear, specific and detailed...**
- And **justify decisions**
- The DMP is to prove to the funder that the researcher has taken time to reflect on what to do, that consideration has been given and the approach seems reasonable
- And that your data is “As open as possible, as closed as necessary” (FAIR principles)

Venkataraman, S. (2018, November). RDM, Open Research and DMP presentations and associated files. Zenodo. <http://doi.org/10.5281/zenodo.1489929>



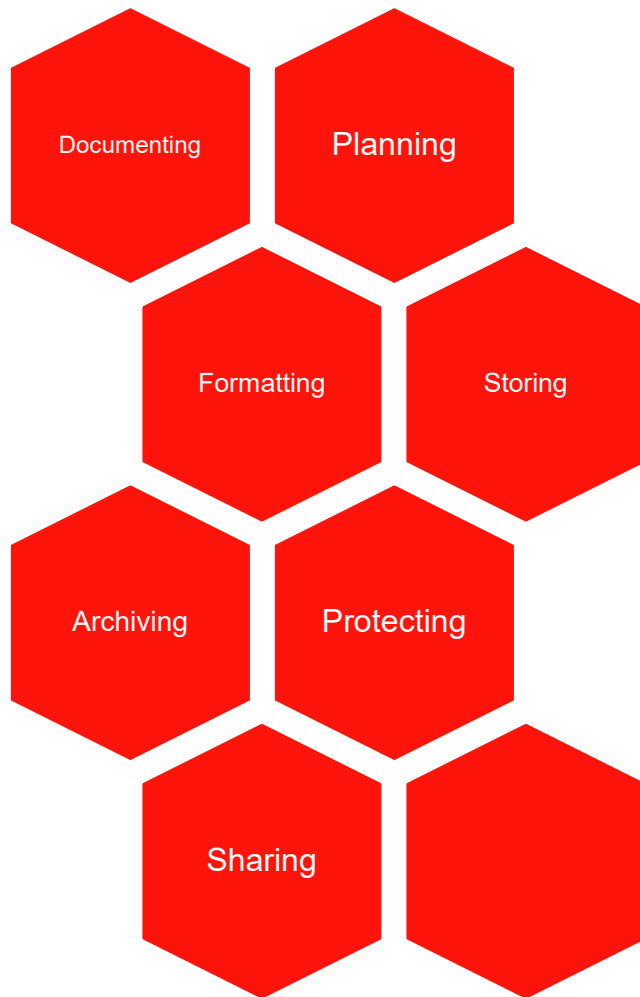


Useful tools

- [DMPonline](#) – provide a number of templates of different funders
- [DCC template](#)
- [H2020 template \(DCC website\)](#)
- Downloadable [H2020 template](#) (EC website)
- Downloadable [ERC template](#) (EC website)
- [H2020 guidelines](#)
- [FNR Data Management Policy](#)
- [OpenAIRE FAQ](#)



Best practices in RDM



- A Data Management Plan outlines what you will be doing with the data
- How you manage, process, format, store, document, etc. the data might be different during and after the project.

Storing

during the
project

- Where is your data stored? *ATLAS, laptop, USB stick?*
- Is a backup created on a regular basis?
- Who has access to the data? Can someone delete/overwrite it by accident?
- What would happen if your file became corrupt? Do you create versions?
- Dealing with personal data? How will you ensure the protection of this data?
- How are you sharing with collaborators? *Dropbox?*
- What if something happened to you? Anyone else has access to the data?
- Think worst case scenario (Zombie attack?), you might not be safe, but your data will be preserved.



Not so good

Best practice

Storing

Laptops, USB sticks are not necessarily secure (not GDPR compliant) and can be lost, then what...

Storing your data on ATLAS is secure, backups are done daily, and you can access it from anywhere through Cisco AnyConnect

Versions

Do you 'Save' or 'Save as' your files? What happens if the file gets corrupted or deleted?

Take the habit to 'Save as' and create different versions of your file (see best practices for file naming). Any mistakes can be easily undone. Version once a day or when you have made big changes to a file

Backup

Having all your versions/"backups" in one folder. But what if the folder gets deleted by mistake?

A backup is different from versioning. Save a copy of a folder (and subfolders) in an actual backup folder that is write- and/or access-restricted

Not so good

Best practice

Sharing with collaborators (outside the University)

Dropbox and Google Docs/Sheets have their servers in the USA, which can be a problem with GDPR. They can still be used, but you need to justify why in the DMP

'Dropit' can share any files with external collaborators (including read-only or read-write options, versions history, password protected, etc.)

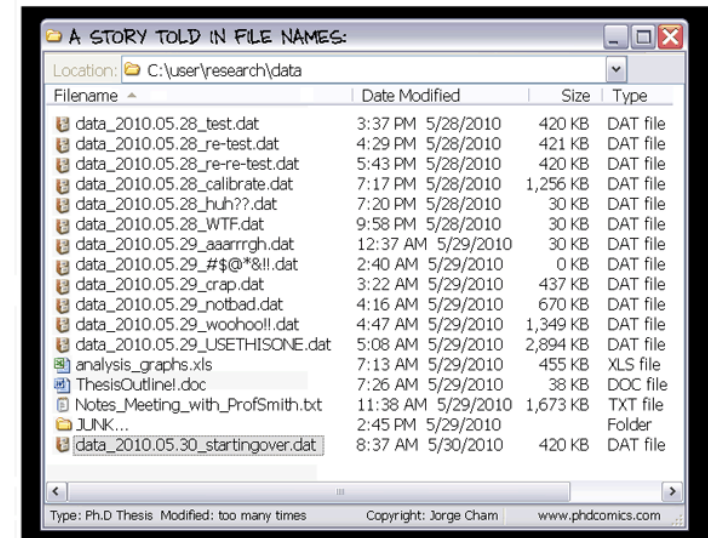
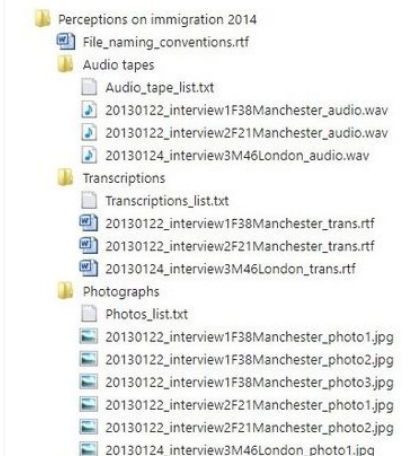
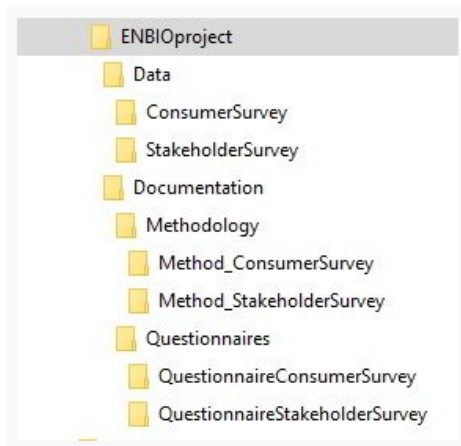
Access

Letting everyone in your team or all collaborators have access to the data can be a problem when dealing with personal data and you run the risk of files being deleted by mistake. But being the only one with the data is not safe either. What if something happens to you?

Decide on who 'needs' access to the data (e.g. one collaborator only), create (sub)folders with limited access, etc.

Folders structure

- Structuring your files in folders is important to make it easier to locate and organise files and versions. Even more important when collaborating with others
- Structure depends on the study. All material relevant to the data should be included (i.e. documentation “README” file, information on data collection and data processing procedures).



[Consortium of European Social Science Data Archives - CESSDA Training - File naming and folder structure - https://www.CESSDA.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure](https://www.CESSDA.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure)



File naming

NO

myabstract.docx

Joe's Filenames Use Spaces and Punctuation.xlsx

figure 1.png

fig 2.png

JW7d^(2sl@deletethisandyourcareerisoverVx2*.txt

YES

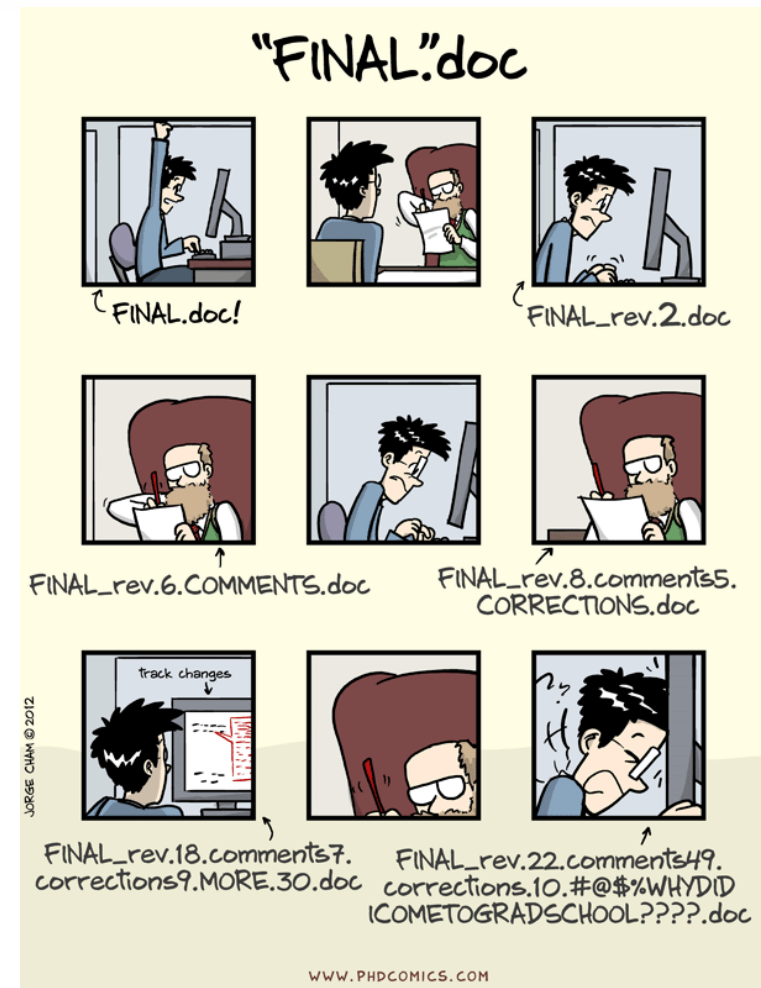
2014-06-08_abstract-for-sla.docx

joes-filenames-are-getting-better.xlsx

fig01_scatterplot-talk-length-vs-interest.png

fig02_histogram-talk-attendance.png

1986-01-28_raw-data-from-challenger-o-rings.txt



[Bryan, J. \(2015\) Naming things -](#)

<https://speakerdeck.com/jennybc/how-to-name-files>



File naming

- Machine readable:

Avoid spaces, punctuation, accented characters, case sensitivity

Use “_” and “-” to recover metadata from filenames

“_” underscore used to delimit units of metadata I want later

“-” hyphen used to delimit words so my eyes don’t bleed

Same using R’s ability to narrow file list by regex:

```
> list.files(pattern = "Plasmid") %>% head
[1] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A01.csv"
[2] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A02.csv"
[3] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A03.csv"
[4] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_B01.csv"
[5] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_B02.csv"
[6] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_B03.csv"
```

[Bryan, J. \(2015\) Naming things -](https://speakerdeck.com/jennybc/how-to-name-files)

<https://speakerdeck.com/jennybc/how-to-name-files>



File naming

■ Human readable:

Name contains info on content

Default ordering by chronological and/or logical order

Put something numeric first (e.g. date)

Use YYYY-MM-DD (ISO 8601) for dates (or YYYYMMDD)

Left pad other numbers with zeros

if you don't left pad, you get this:

```
10_final-figs-for-publication.R  
1_data-cleaning.R  
2_fit-model.R
```

which is just sad

Example:

<date><type><ID1><gender><age><datatype><ID2>

Need batch renaming of automatically generated files?

[LINK](#)

chronological
order

```
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_H01.csv  
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_H02.csv  
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_H03.csv  
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_platefile.csv  
2014-02-26_BRAFWTNEGASSAY_FFPEDNA-CRC-1-41_A01.csv  
2014-02-26_BRAFWTNEGASSAY_FFPEDNA-CRC-1-41_A02.csv  
2014-02-26_BRAFWTNEGASSAY_FFPEDNA-CRC-1-41_A03.csv  
2014-02-26_BRAFWTNEGASSAY_FFPEDNA-CRC-1-41_A04.csv
```

logical
order

```
01_marshall-data.r  
02_pre-dea-filtering.r  
03_dea-with-limma-voom.r  
04_explore-dea-results.r  
90_limma-model-term-name-fiasco.r  
helper01_load-counts.r  
helper02_load-exp-des.r  
helper03_load-focus-statinf.r  
helper04_extract-and-tidy.r
```

Example:

- 20130311_interview2_audio.wav
- 20130311_interview2_trans.rtf
- 20130311_interview2_image.jpg

[Bryan, J. \(2015\) Naming things - https://speakerdeck.com/jennybc/how-to-name-files](#)



File format (1/2)

Open formats, what for?

▶ So that users might read my documents unhindered

Users exchanging reports.



CLOSED FORMAT, IDENTICAL SOFTWARE

Alice uses the software program "Carcera⁽¹⁾". She records her report in a closed format (one that does not permit interoperability), then sends it to Bob, who has the same software program. He can read the document, modify it and send it back to Alice.



PROGRAMS WITH CLOSED FORMATS, DIFFERENT SOFTWARE

The following day, Alice sends her report to Albert. He doesn't have the same software program, which refuses to open the document. Albert has no other choice than to acquire the Carcera software used by Alice, with the hope it is compatible with his computer.

So that your documents might be read more easily by other people, without you having to worry about which software they use, choose open formats.

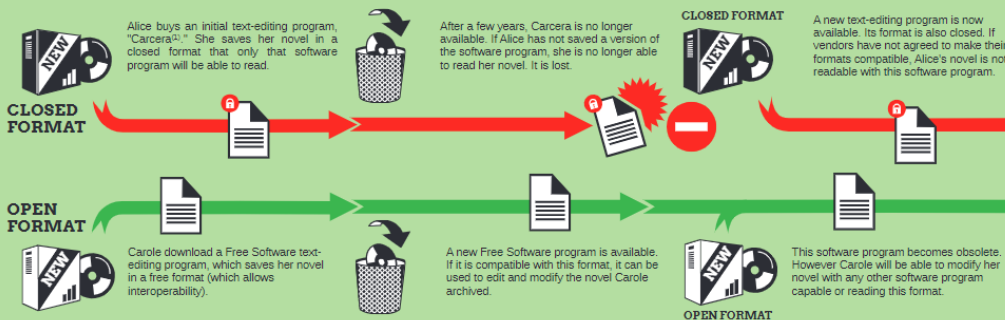


PROGRAMS WITH OPEN FORMATS, DIFFERENT SOFTWARE

Carole, another user, chooses to record her report in open format (allowing for interoperability) and sends it to David. David can read the document, modify and record it, either by using the same open format software or by using another interoperable software.

▶ To ensure the longevity of my documents

Alice and Carole use text editors to write, save, and preserve a novel.



The availability and longevity of your documents, saved in a closed format, depends on the decisions of software vendors.

In your interest, choose software programs that save your documents in open formats. They are not dependent on any particular software program.

Document designed by April with Free Software. Graphics: Antoine Bardelli. Free Art License 1.3 or higher / Creative Commons Paternity, Share alike 2.0 or higher / GFDL 1.3 or higher.

<https://www.april.org/en/open-formats>

File format (2/2)

If you are not aware of any disciplinary standards these are some good file formats for the preservation of the most common data types:

- Textual data: XML, TXT, HTML, PDF/A (Archival PDF)
- Tabular data (including spreadsheets): CSV (instead of Excel's proprietary XLSX)
- Databases: XML, CSV
- Images: TIFF, PNG, JPEG (note: JPEG is a 'lossy' format which lose information when re-saved, so only use it if you are not concerned about image quality)
- Audio: WAV, FLAC, MP3 (lossy format)

[University of Cambridge - Data Management Guide - Choosing Formats - https://www.data.cam.ac.uk/data-management-guide/creating-your-data/choosing-formats](https://www.data.cam.ac.uk/data-management-guide/creating-your-data/choosing-formats)



Planning

- Who is responsible for which part of data management?
- Are new skills required for any activities?
- Do you need extra resources to manage data, such as people, time or hardware?
- Have you accounted for costs associated with depositing data for longer-term preservation and access?



Digitalbevaring.dk

[UK Data Service - Data management checklist -
https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx](https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx)



RDM costs covered by funders

- There are no default costs that can be given, but there are [tools that can help](#)

- Indirect costs
 - Human resources (e.g. data curator)

- Overhead costs
 - Existing university services and infrastructure

- Direct costs
 - Costs related to database or dataset access (e.g. licensing costs)
 - Data storage and backup costs (e.g. cloud storage)
 - Data formatting or data cleaning costs (e.g. if external service)
 - Transcription of qualitative data (e.g. interview transcripts)
 - Data encryption software
 - Digitalisation of analogue or paper-based research data
 - Data storage and preservation costs in a repository (e.g. <https://datadryad.org/>; 120USD/20GB)

Sources: Tom Jakobs (FNR),
OpenAIRE



Protecting

- Do your data contain confidential or sensitive information? If so, have you discussed data sharing with the respondents from whom you collected the data?
- Have you discussed it with the Data Protection Officer? (DPO – Sandrine Munoz). And the Chief Information Security Officer? (CISO – Niek Nigg)
- Are you gaining written consent from respondents to share data beyond your research?
- Do you need to anonymise data, for example, to remove identifying information or personal data, during research or in preparation for sharing?



Consent

Ethics

GDPR

Confidentiality

[UK Data Service - Data management checklist -
https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx](https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx)

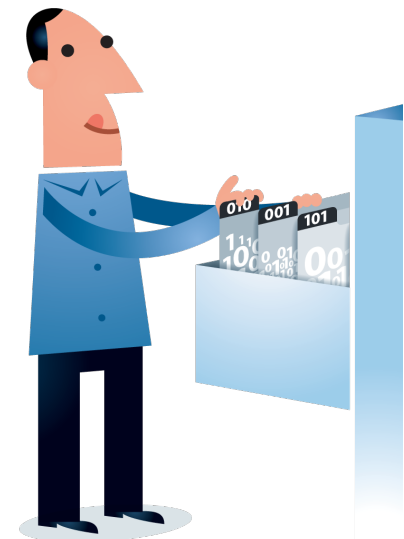


Archiving

- While storing issues concern the data during the research project, archiving focuses on what happens to the data after the end of the project?
- Who has access to which data during and after research? Is there a need for access restrictions? How will these be managed after you have finished your PhD ? or when you are dead (*in a long time I hope for you*)...
- How long will you store your data for and do you need to select which data to keep and which to destroy?
- Answers to those questions will depend on the best practices in your field of research

[UK Data Service - Data management checklist -
https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx](https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx)

Long-term
preservation



Digitalbevaring.dk



Sharing

*Licences
Copyright*

FAIR principles

*after the
project*

- Do you intend to make all your data available for sharing or how will you select which data to preserve and share?
- How and where will you preserve your research data for the longer-term?
- How will you make your data accessible to future users?
- Have you established who owns the copyright in your data? Might there be joint copyright?
- Have you considered which kind of license is appropriate for sharing your data and what, if any, restrictions there might be on re-use?
- If you are purchasing or re-using someone else's data sources have you considered how that data might be shareable, for example negotiating a new licence with the original supplier?
- Can you preserve for the long-term, personal information so that it can be used in the future?

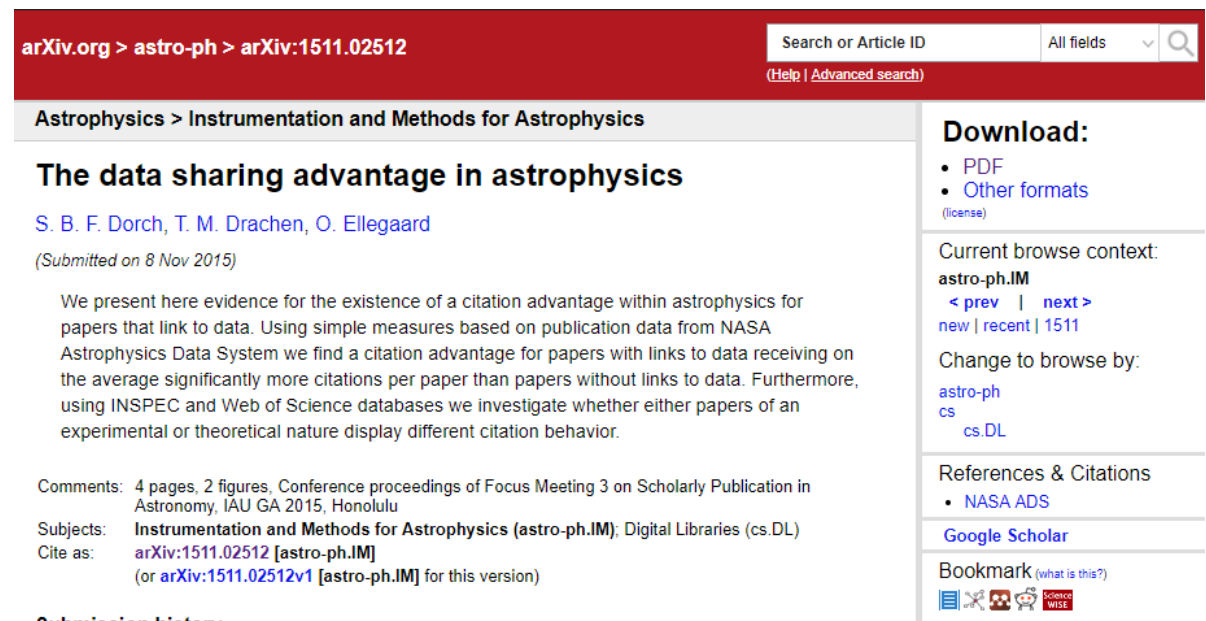
[UK Data Service - Data management checklist -
https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx](https://www.ukdataservice.ac.uk/manage-data/plan/checklist.aspx)



FAIR principles

How useful would it for your own research to have access to others data?

... publish data in your own interest...



arXiv.org > astro-ph > arXiv:1511.02512

Search or Article ID All fields

(Help | Advanced search)

Astrophysics > Instrumentation and Methods for Astrophysics

The data sharing advantage in astrophysics

S. B. F. Dorch, T. M. Drachen, O. Ellegaard

(Submitted on 8 Nov 2015)

We present here evidence for the existence of a citation advantage within astrophysics for papers that link to data. Using simple measures based on publication data from NASA Astrophysics Data System we find a citation advantage for papers with links to data receiving on the average significantly more citations per paper than papers without links to data. Furthermore, using INSPEC and Web of Science databases we investigate whether either papers of an experimental or theoretical nature display different citation behavior.

Comments: 4 pages, 2 figures, Conference proceedings of Focus Meeting 3 on Scholarly Publication in Astronomy, IAU GA 2015, Honolulu

Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM)**; Digital Libraries (cs.DL)

Cite as: arXiv:1511.02512 [astro-ph.IM]
(or arXiv:1511.02512v1 [astro-ph.IM] for this version)

Submission history

From: S. B. F. Dorch [view email]
[v1] Sun, 8 Nov 2015 18:13:05 UTC (491 KB)

Download:

- PDF
- Other formats (license)

Current browse context: astro-ph.IM
< prev | next >
new | recent | 1511


Change to browse by:

- astro-ph
- cs
- cs.DL

References & Citations

- NASA ADS

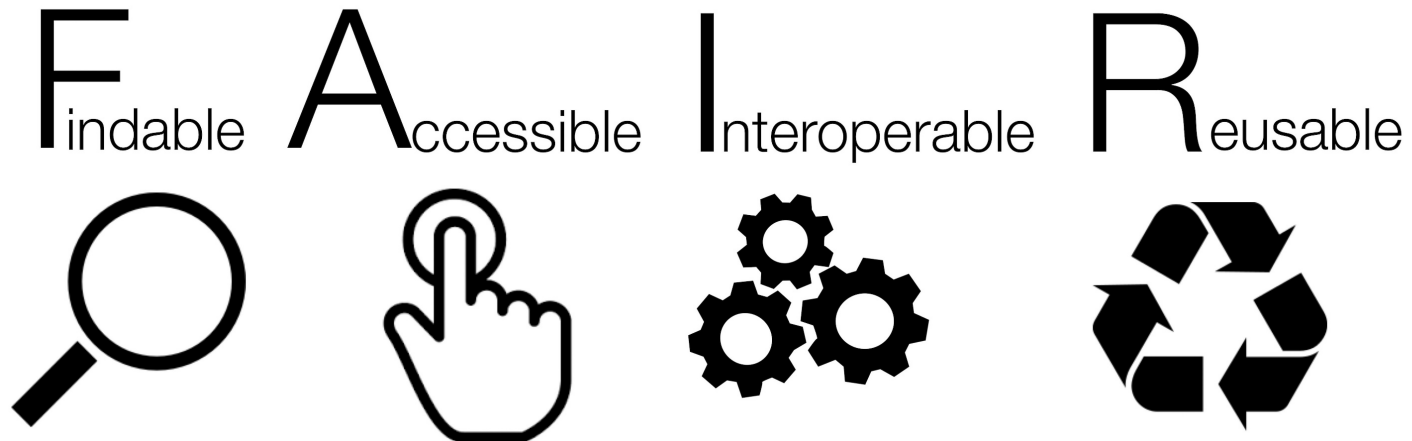
Google Scholar

Bookmark (what is this?)




But wanting to share data is not enough...

... it has to be reusable by others...



A set of principles that describe the attributes data need to have to enable and enhance reuse, by humans and machines

What is FAIR DATA?

**Persistent
Identifiers**



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

FINDABLE



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

ACCESSIBLE

Repository

**Open file
formats**



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

INTEROPERABLE



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

REUSABLE

Licence

LIBER



“As open as possible, as closed as necessary”

- FAIR data does not mean it has to be open. You can make data open, but **if it is not FAIR it is useless**:
 - Findable – the data is online, but no one can find it because it has no metadata attached to it and no PIDs
 - Accessible – the data is open but you need to “contact the author to have access to it”
 - Interoperable – the data is open, findable, accessible and reusable but no documentation is included and the data’s variable names are not explained
 - Reusable – the data is open, findable and accessible but you did not provide a licence, so people do not know if they can reuse the data
- Data **can be FAIR and closed**, e.g. sensitive data can still be accessible on an online platform available to authorised people (a licence would still be needed to provide reuse, but the new data would not be openly shareable either).

Persistent Identifiers

ORCID

Connecting Research
and Researchers



Repository



Data sharing made easier: use Repository Finder to find the right repository for your data

2,000 Data Repositories and Science Europe's Framework for Discipline-specific

Three new DOI Fabrica features to simplify account management
Last month month we launched DOI



Documenting (1/4)

- Imagine your new on a project and trying to decipher other people’s datasets. What kind of information and metadata would you like to have to understand what the data is about, how it was processed, etc.

- Compile this information for your own research data, it is like a “README” file

- Distinguish:
 - Project-level documentation
 - Data-level documentation
 - Metadata and other resources

<https://www.ukdataservice.ac.uk/manage-data/document.aspx>

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Documents/Documentation-and-metadata>



Documenting (2/4)

- Project-level documentation:
 - For what purpose was the data created?
 - What does the dataset contain?
 - How was data collected?
 - Who collected the data and when?
 - How the data processed?
 - What possible manipulations were done to the data?
 - What were the quality assurance procedures?
 - How can the data be accessed?

Example:

<https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6986&type=Data%20catalogue#!#documentation>



Documenting (4/4)

- Metadata = machine-readable data documentation = data about data:
 - Author, creation date, etc.
 - Check out the Dublin Core Metadata Generator, metadata templates and metadata standards

⊖ Metadata templates

Metadata can, at its simplest, be stored in a single text file. However, you can also use a metadata template to help you structure your metadata or to see how your metadata appears in *.html.

Below we provide examples of metadata templates that you can use when compiling documentation. Or just for inspiration to take a look at typical fields which are often required. It is always possible to include additional documentation beyond what is suggested.

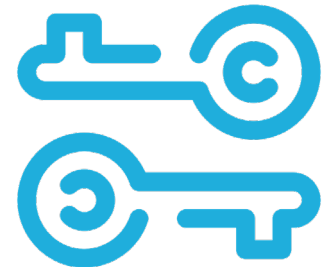
- [Create a codebook](#) about your research to accompany the dataset (DDI Alliance, 2017a).
- Download the York University (n.d.) [Library Metadata Template](#), Dublin Core;
- Have a look at the Georgia Tech Library (n.d.) [Metadata Template](#);
- Use the [Dublin Core Metadata Generator](#) (dublincoregenerator, n.d.);
- Have a look at the Cornell University (n.d.) [guide to writing “readme” style metadata](#) (with downloadable template);
- Use the [ISO 19115-2 Metadata Editor](#) (GRIIDC (2015)) web application;

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata>

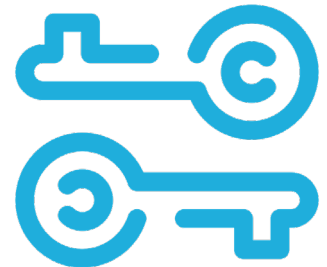


Open License: middle ground between © and Public Domain

- Removes ambiguity over what others can and cannot do with your work
- As rights owner, you modify the conditions of what can and cannot be done with your work (combining different conditions)
- Removes uncertainty about © status for re-user
- You still keep (certain) rights, but you grant certain reuses without them needing to contact you
- Infringements against the conditions of the license are a © infringement
- Different types of licenses are suitable for different types of content

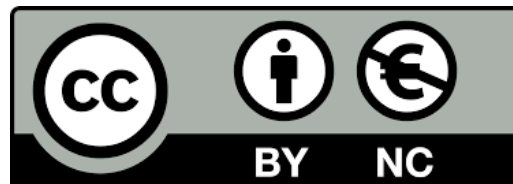
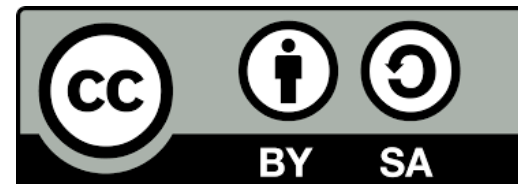


- A Creative Commons license (or the public domain tool) is universally recognisable, juridically sound, easily applicable and leaves the user in no doubt about the intentions of the author.
- CC licenses let you easily change your copyright terms from the default of “all rights reserved” to “some rights reserved.”
- The easier you make it for somebody to figure out their re-use rights, the more likely it is that they will respect them!
- Be clear and unambiguous when creating and displaying your work's metadata (i.e. the information about your work)
- Use license logo and link to the human-readable license text
- Make sure people can reach you or the platform when in doubt



Research Data Management & DMP

webinar – April 2021



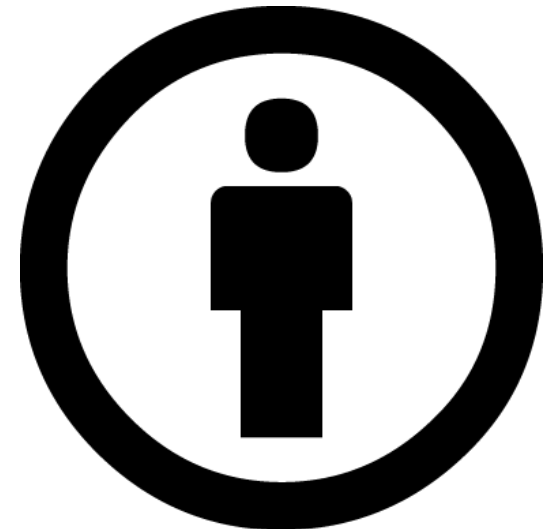
ATTRIBUTION (“BY”)

You are free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material
- for any purpose, even commercially

Under the following terms:

- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



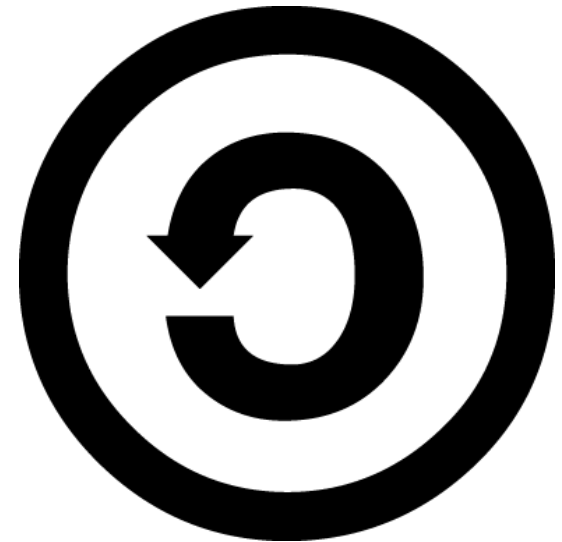
SHARE ALIKE (“SA”)

You are free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material
- for any purpose, even commercially

Under the following terms:

- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made
- ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.



NON-COMMERCIAL (“NC”)

You are free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material

Under the following terms:

- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made
- NonCommercial — You may not use the material for commercial purposes



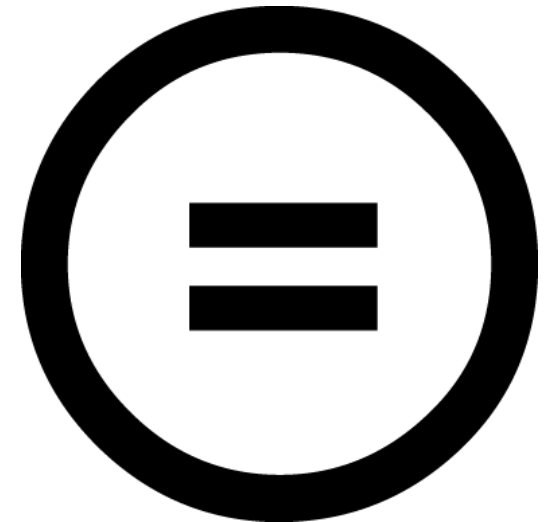
NO DERIVATIVES (“ND”)

You are free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material

Under the following terms:

- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made
- NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.

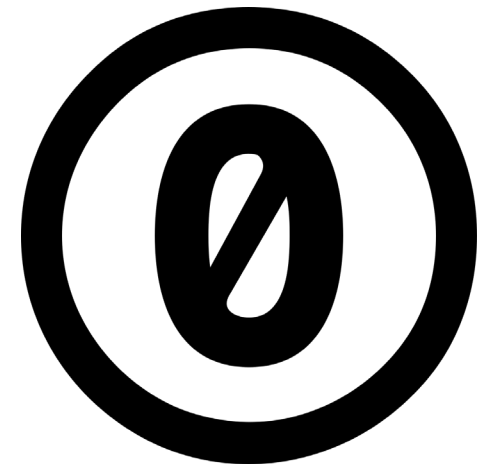


CC 'ZERO'

- It is different from 'Public Domain' because it is an actual licence, but the rights are the same

You are free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material
- for any purpose, even commercially
- No Attribution required — You are free to give appropriate credit (best practice) but it is not required



The CC 0 is the **most appropriate licence for sharing data**, as it avoid attribution stacking, allows commercial reuse such as consulting and allows meta-analyses to be freely carried out.



Research Data Management & DMP

webinar – April 2021

LICENSES

MOST FREE



ATTRIBUTION
CC BY

This license lets you distribute, remix, tweak, and build upon the original work, even commercially, as long as you credit the original creation. This is the most accommodating of licenses offered.



ATTRIBUTION-SHAREALIKE
CC BY-SA

This license lets you remix, tweak, and build upon the original work even for commercial purposes, as long as you credit the original work and license your new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on the work should carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia.



ATTRIBUTION-NODERIVS
CC BY-ND

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to the original work.



ATTRIBUTION-NONCOMMERCIAL
CC BY-NC

This license lets you remix, tweak, and build upon the original work non-commercially. Your new works must be non-commercial and acknowledge the original work, but you don't have to license your derivative works on the same terms.



ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE
CC BY-NC-SA

This license lets you remix, tweak, and build upon the original work non-commercially, as long as you credit the original work and license your new creations under the identical terms.



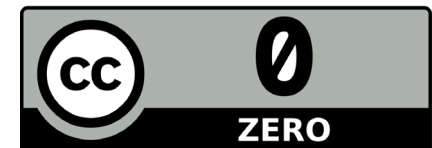
ATTRIBUTION-NONCOMMERCIAL-NODERIVS
CC BY-NC-ND

This license is the most restrictive of the six main licenses, only allowing you to download the original work and share it with others as long as you credit the original work. You can't change the original work in any way or use it commercially.

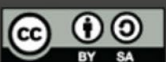
LEAST FREE



for publications
(including
AAM/postprint)



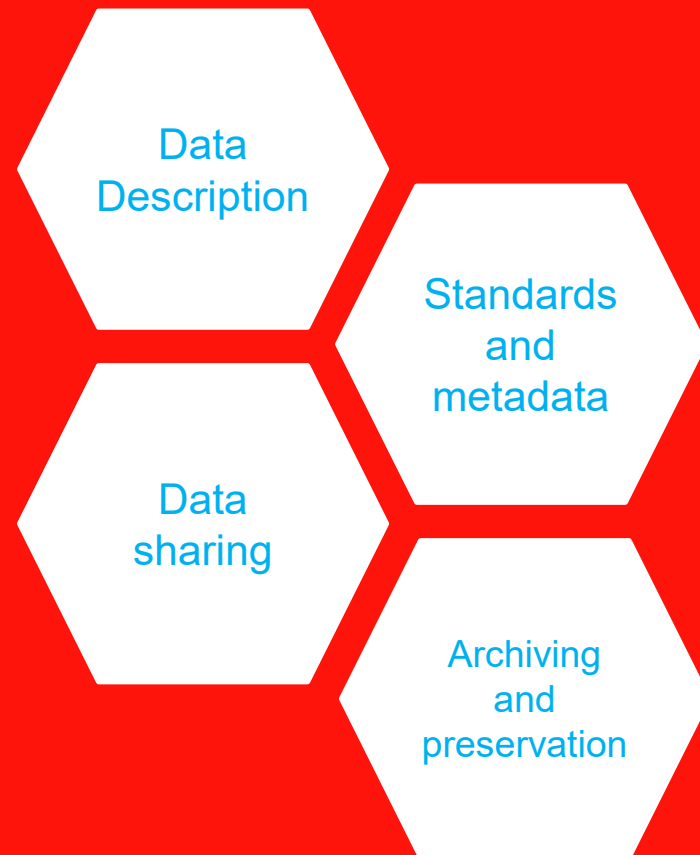
for sharing data



THIS WORK IS LICENSED UNDER
<http://creativecommons.org/licenses/by-sa/3.0/>

foter

Evaluating a DMP



Check (1/3)

- Full description of the data to be produced?
Statistics about the size, quantity and duration help reviewers get a proper sense of scale
- Sharing considered in the consent and licence agreements if third-party data being reused or working with human subjects?
- Will sufficient metadata and documentation be provided to allow others to find, understand the reuse the data?
- Is the choice of file format explained, so it is clear that appropriate decisions have been made?

Venkataraman, S. (2018, November). RDM, Open Research and DMP presentations and associated files. Zenodo. <https://doi.org/10.5281/zenodo.1489929>



Check (2/3)

- Is it clear which data will be shared and with whom?
- How will the data be shared? E.g. deposit in a repository (re3data)
- If an embargo period is planned, is that in line with norms for that discipline?
- Will the data be deposited in a suitable community database, repository or archive?
- Is it clear which data should be preserved and for how long?
- Are there any costs associated with preservation, and if so, how will these be covered?

Venkataraman, S. (2018, November). RDM, Open Research and DMP presentations and associated files. Zenodo. <https://doi.org/10.5281/zenodo.1489929>



Check (3/3)

- Is the plan appropriate?
e.g. adopting relevant standards; practices in line with norms for that field; use of university storage, subject repositories
- Seem feasible to implement?
- Sufficient detailed information provided?
- Advice sought where needed?
- Restrictions and costs properly justified?

Venkataraman, S. (2018, November). RDM, Open Research and DMP presentations and associated files. Zenodo. <http://doi.org/10.5281/zenodo.1489929>



Specific information

“we will use suitable formats to ensure that our data can be preserved and sustained over the long term”

- Which standards? Name them
- Does the team know which are suitable?
- Does the chosen repository have preferences?
- Be clear, specific and detailed
- Justify decisions

“we will provide **MP3 audio files for online dissemination**. While this is **not an open format**, it is well-established and the most **widely supported**. High-resolution **WAV files** will be used for the **archival master recordings**”.

Venkataraman, S. (2018, November). RDM, Open Research and DMP presentations and associated files. Zenodo. <http://doi.org/10.5281/zenodo.1489929>



Easy to evaluate

“Online resource development will cost 21,000 €”

VS

“Online resource development, 60 days at 350 €”

- Don't make reviewers dig around for information
- Be consistent in what you say
- Stick to page limits, follow the template (if provided)
- Beware of blanket copy/paste

Venkataraman, S. (2018, November). RDM, Open Research and DMP presentations and associated files. Zenodo. <http://doi.org/10.5281/zenodo.1489929>





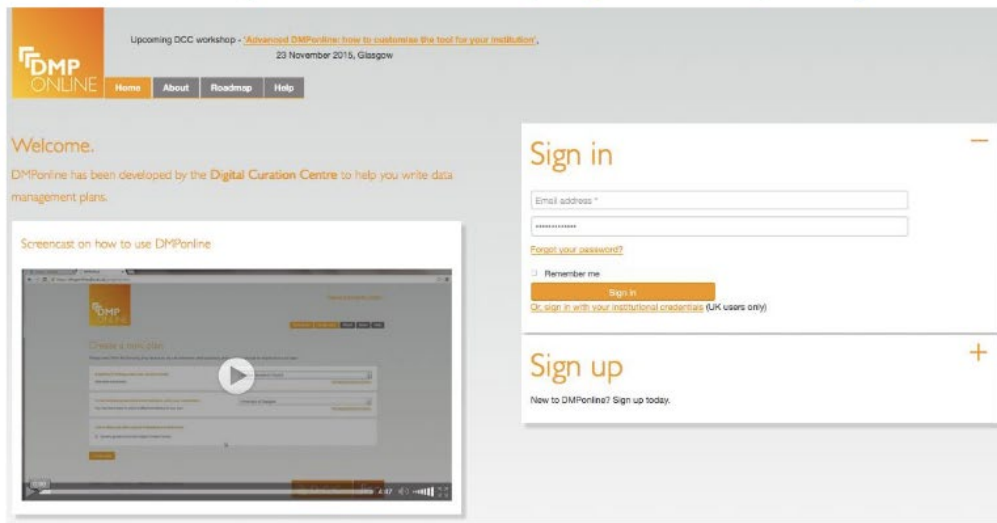
An example

DMPonline



A web-based tool to help researchers write DMPs

Includes templates for funders, organisations, disciplines



<https://dmponline.dcc.ac.uk>



An example

How the tool works

Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, [select here to write a generic DMP](#) based on the most common themes.

If applying for funding, select your research funder.
Otherwise leave blank.

European Commission (Horizon 2020) [Not applicable/not listed.](#)

To see institutional questions and/or guidance, select your organisation.
You may leave blank or select a different organisation to your own.

University of Glasgow [Not applicable/not listed.](#)

Tick to select any other sources of guidance you wish to see.

- DCC guidance
- EUDAT
- School of Humanities
- Computing

Create plan

Click to write a generic DMP

Or choose your funder to get their specific template

Pick your uni to add local guidance and to get their template if no funder applies

Choose any additional optional guidance



An example

Plan details: summary

This plan is based on:

Funder | Economic and Social Research Council

Answer questions

Export

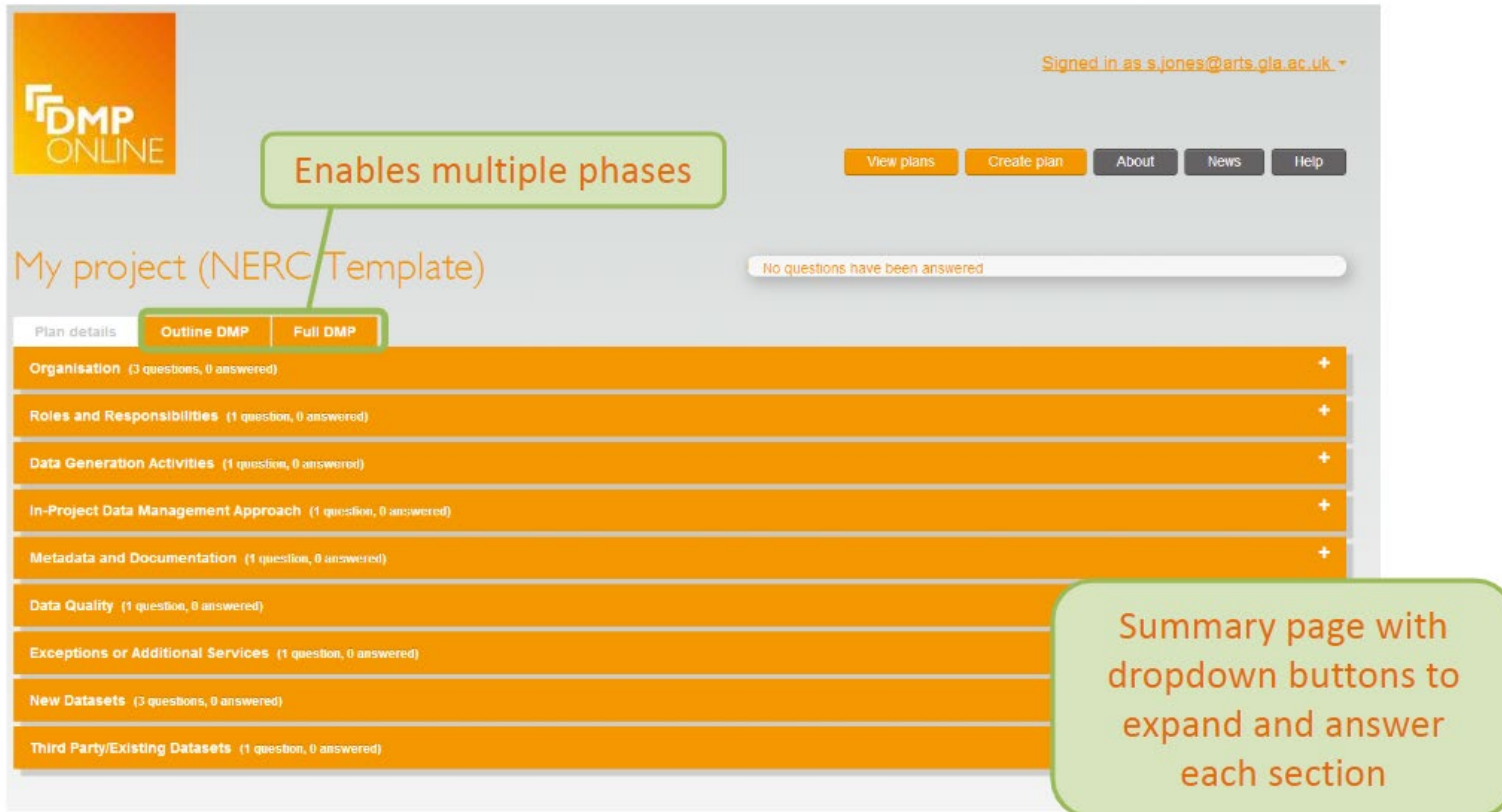
The ESRC requires that all applicants seeking ESRC funding include a statement on data sharing in the relevant section of the Je-S application form. If data sharing is not possible, the applicant must present a strong argument to justify their case.

Sections	Questions
Existing data	<ul style="list-style-type: none">- An explanation of the existing data sources that will be used by the research project (with references)- An analysis of the gaps identified between the currently available and required data for the research
Information on the data that will be produced	<ul style="list-style-type: none">- Methodologies for data collection- Data volume and data type, e.g. qualitative or quantitative data- Data quality, formats, standards documentation and metadata
Planned quality assurance and back-up procedures (security/storage)	<ul style="list-style-type: none">- Quality Assurance- Back-Up
Management and archiving of collected data	<ul style="list-style-type: none">- Plans for management and archiving of collected data
Overcoming data sharing difficulties	<ul style="list-style-type: none">- Expected difficulties in data sharing, along with causes and possible measures to overcome these difficulties.
Consent, confidentiality, anonymisation and other ethical considerations	<ul style="list-style-type: none">- Explicit mention of consent, confidentiality, anonymisation and other ethical considerations
Copyright and intellectual property ownership of the data	<ul style="list-style-type: none">- Copyright and intellectual property ownership of the data
Responsibilities for data management and curation	<ul style="list-style-type: none">- Responsibilities for data management and curation

Summary of the sections and questions in your DMP

An example

Overview of sections in a DMP



The screenshot shows the DMP ONLINE interface for a project titled "My project (NERC Template)". The user is signed in as s.jones@arts.gla.ac.uk. The interface includes a navigation menu with "View plans", "Create plan", "About", "News", and "Help". A callout box points to the "Outline DMP" and "Full DMP" tabs, stating "Enables multiple phases". The main content area displays a list of sections, each with a dropdown arrow and a status indicator (e.g., "3 questions, 0 answered"). A second callout box points to the list of sections, stating "Summary page with dropdown buttons to expand and answer each section".

Enables multiple phases

Summary page with dropdown buttons to expand and answer each section

Remind to...

- Organise your research data, you will thank yourself later I promise!
- Take the time to think about what kind of data you are producing and processing.
- Design your data folder structure, decide on a file naming routine, think which file format you are using to save/share, document everything as you go along, backup your data.
- Stick to your system! (but remain flexible)





Sources and other useful references

RESEARCH DATA MANAGEMENT

- Venkataraman, S. (2018, November). RDM, Open Research and DMP presentations and associated files. Zenodo.
<http://doi.org/10.5281/zenodo.1489929>
- Science Europe- RDM guide
<https://scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/>
- Research data management (RDM) open training materials
<https://zenodo.org/communities/dcc-rdm-training-materials/?page=1&size=20>
- OpenAIRE Research Data Management Hand Book
<https://www.openaire.eu/rdm-handbook>
- Research Data Management Adventure (serious game)
<https://rdm-games.gitlab.io/rdm-adventure/>

WHAT IS DATA & FAIR DATA

- What counts as research data? (University of Bristol)
<https://data.blogs.bristol.ac.uk/bootcamp/data/>
- Data Management Expert Guide – Research Data (CESSDA ERIC)
<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/1.-Plan/Research-data>
- Research Data in context (MANTRA)
<https://mantra.edina.ac.uk/researchdataexplained/>
- What is data? (University of Leicester)
<https://www2.le.ac.uk/services/research-data/rdm/what-is-rdm/research-data>

- Zenodo-FAIR principles
<https://about.zenodo.org/principles/>
- "A love letter to your future self": What scientists need to know about FAIR data
<https://www.natureindex.com/news-blog/what-scientists-need-to-know-about-fair-data>
- FAIRsFAIR Europe
<https://www.fairsfair.eu/>
- GO FAIR
<https://www.go-fair.org/>

DATA ORGANISATION

- Prepare and manage data (UK Data Service)
<https://www.ukdataservice.ac.uk/manage-data>
- Towards a Standardized Research Folder Structure
<https://genr.eu/wp/towards-a-standardized-research-folder-structure/>
- UK Data Service recommended file formats
<https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>
- Creating data – Choosing Formats (University of Cambridge)
<https://www.data.cam.ac.uk/data-management-guide/creating-your-data/choosing-formats>
- Naming things
<https://speakerdeck.com/jennybc/how-to-name-files>



Sources and other useful references

BEST PRACTICES

- 'I ain't afraid of no myth' – busting the myths on data sharing
<https://blog.wellcomeopenresearch.org/2020/08/20/i-aint-afraid-of-no-myth-busting-the-myths-on-data-sharing/>
- Guide de bonnes pratiques sur la gestion des données de la recherche
<https://mi-gt-donnees.pages.math.unistra.fr/guide/00-introduction.html>

DATA MANAGEMENT PLAN

- DMP Online (Digital Curation Centre)
<https://dmponline.dcc.ac.uk/>
- Registry of Research Data Repositories – re3data
<https://re3data.org>
- How guides – develop a data plan
<http://www.dcc.ac.uk/resources/how-guides/develop-data-plan#f3>
- Data Management Plans (TUDelft)
<https://www.tudelft.nl/en/library/research-data-management/r/plan/data-management-plans>
- Guidelines on FAIR Data Management in Horizon 2020
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- Data Management costs
<https://www.tudelft.nl/en/library/research-data-management/r/plan/data-management-costs>

- What will it cost to manage and share my data?
<https://www.openaire.eu/what-will-it-cost-to-manage-and-share-my-data>
- RDM costs estimation tool
<https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>

OPEN SOURCE & OPEN LICENSING

- Open source software and open data: open licensing of software and data for public policy analysis and for collaborative research
<https://doi.org/10.5281/zenodo.4537157>

OPEN SCIENCE

- Open Science Hand Book
<https://doi.org/10.5281/zenodo.1212496>
- Open Data Hand Book
<https://opendatahandbook.org/guide/en/why-open-data/>
- Open Science, Open Data, Open Source. 21st century research skills for the life science
<http://doi.org/10.5281/zenodo.1015288>
- Open Scientist Handbook
<https://openscientist.pubpub.org/>
- Winning Horizon 2020 with Open Science
<https://doi.org/10.5281/zenodo.12247>

Questions?
