



The What, Why and How of Data Management Planning

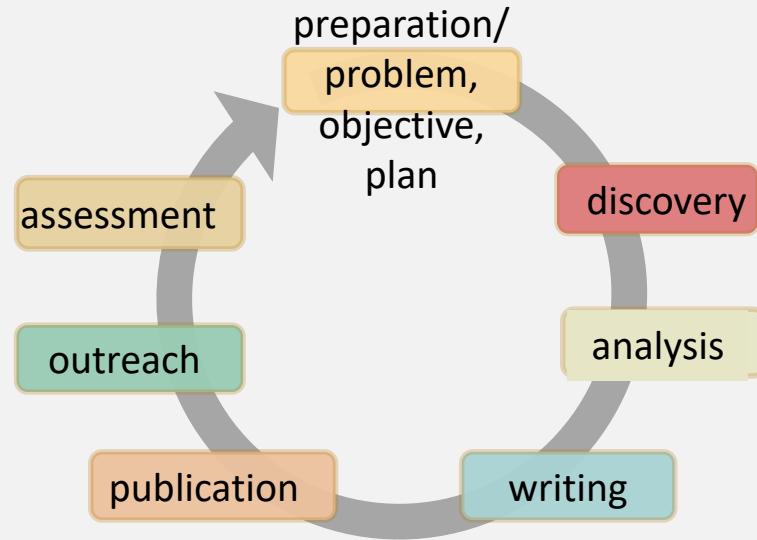
Iryna Kuchma, EIFL Open Access Programme Manager, @irynakuchma

Seminar for young researchers in SSH on Open Science and Data Management, EKT

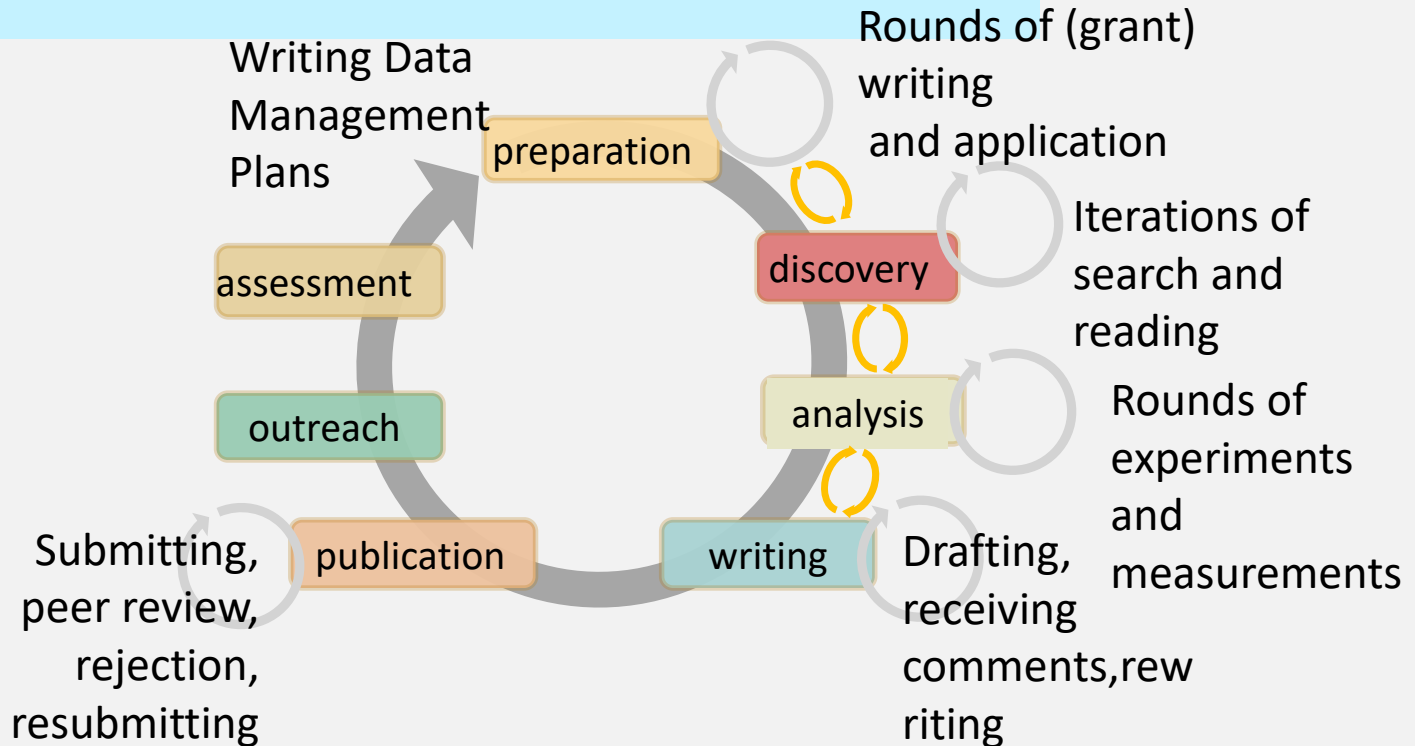
5th June 2018



A model of the research workflow



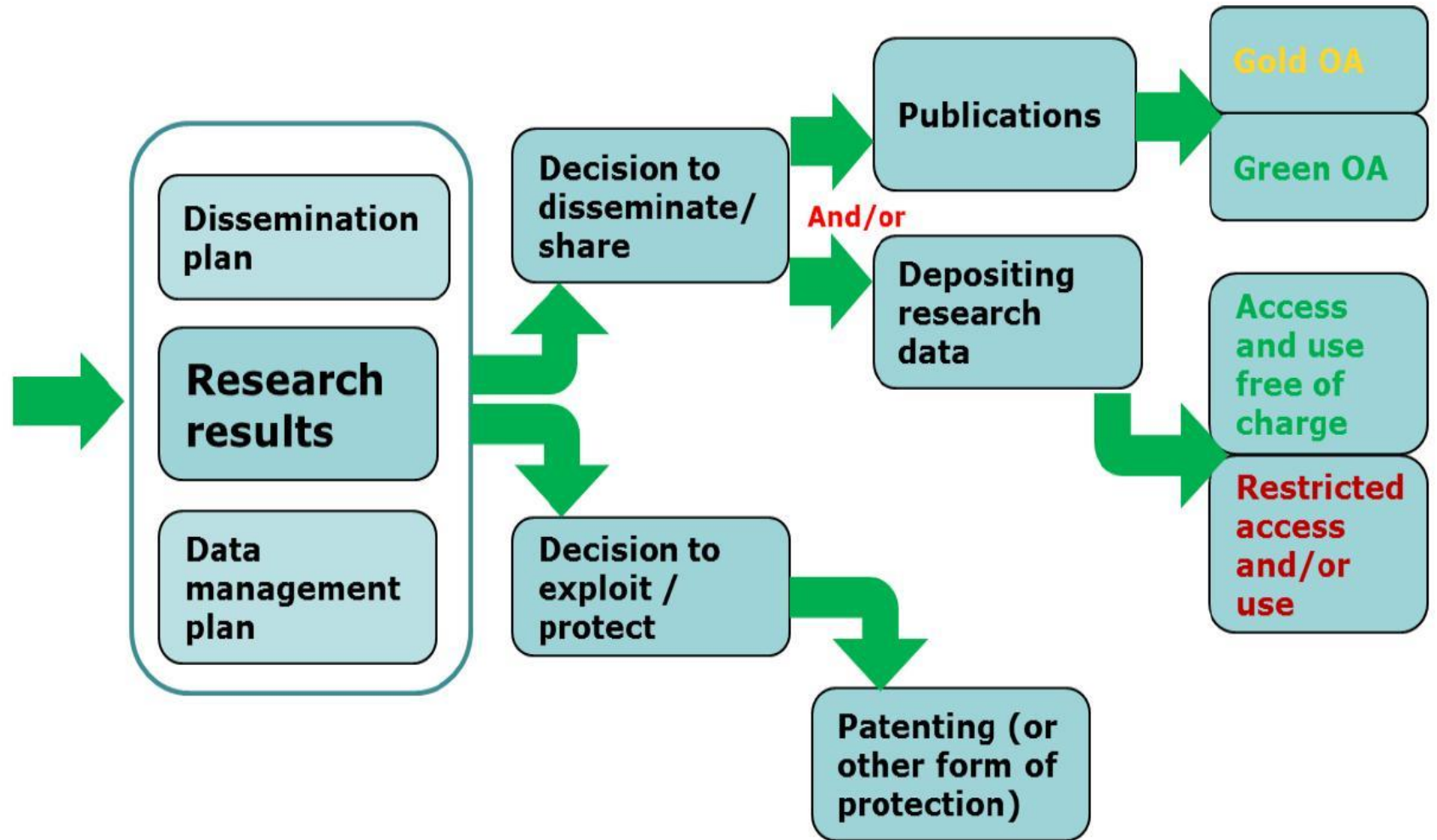
A model of the research workflow



Disciplinary variety and Open Science

	ARTS & HUMANITIES	SOCIAL SCIENCE	LIFE/HEALTH	PHYSICAL SCIENCES
Research types	often exploratory research	often confirmatory research	often confirmatory research	often confirmatory research?
Data	often textual data	also qualitative data, sometimes sensitive data	sensitive patient data / big datasets	big datasets
Publ. Types	books, chapters, articles	mostly articles and chapters	mostly articles, (syst.) reviews	preprints, conf papers, articles
Collaboration	typically 1	typically 1-4	typically 3-10	typically 3-many
Languages	native language & some English	English, some native languages	English	English
Funding	small scale funding	small & medium scale funding	large scale funding	large scale funding
Review	double blind	double + single blind	single blind	single blind

Research



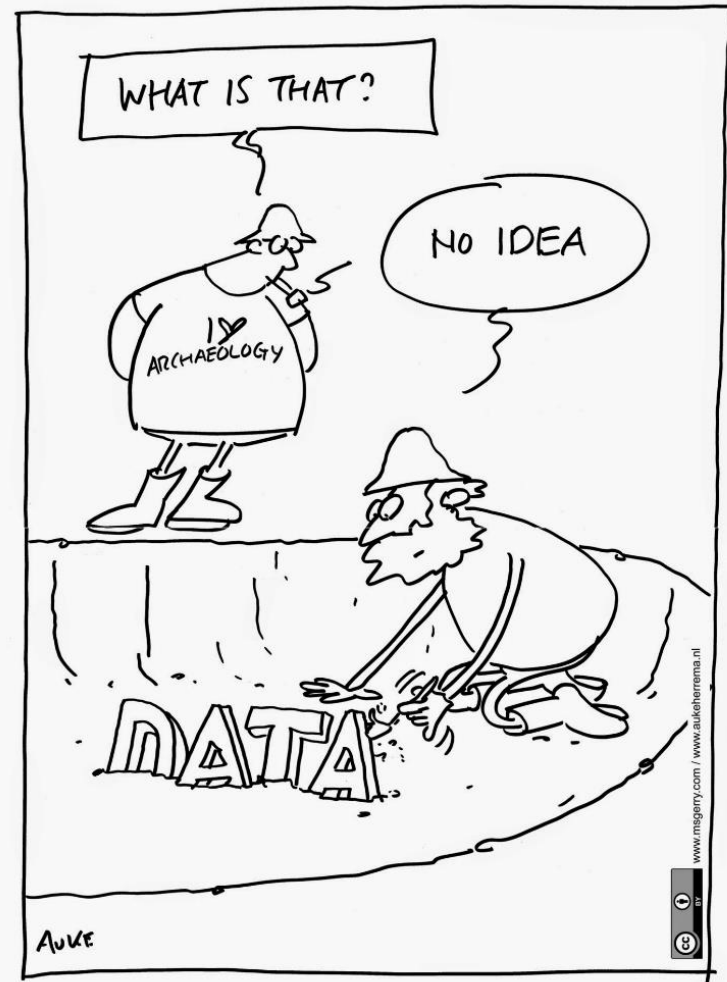


FOSTER

Defining your data

Describe your data (e.g. type,
format, volume)

A faint, larger dandelion seed head graphic is located in the bottom left corner of the slide, partially overlapping the text.



DATA FOR FUTURE GENERATIONS

What are research data?

The University of Leeds describes research data as 'any information that has been collected, observed, generated or created to validate original research findings'. Research data can include things like...



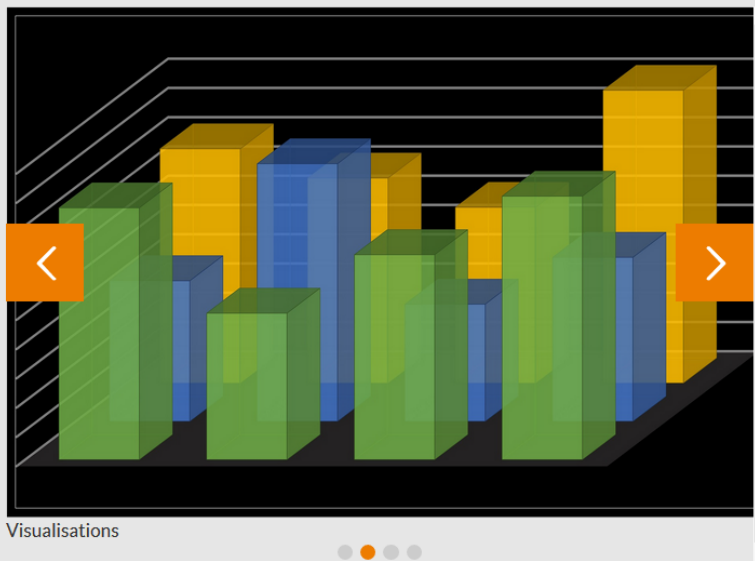
Raw data

...raw data

Raw data are those which are captured from instruments and sensors such as telescopes, smart phones, and satellites.

What are research data?

The [University of Leeds](#) describes research data as 'any information that has been collected, observed, generated or created to validate original research findings'. Research data can include things like...



...visualisations, models, and algorithms

Researchers also generate digital resources such as models and algorithms to help them analyse, visualise and present raw data in a meaningful way.

What are research data?

The [University of Leeds](#) describes research data as 'any information that has been collected, observed, generated or created to validate original research findings'. Research data can include things like...



Digital images, audio and video files

...images, audio, and video files

Remember that digital images are data too. This is also true of any audio files or videos captured during the course of research such as taped interviews.

What are research data?

The University of Leeds describes research data as 'any information that has been collected, observed, generated or created to validate original research findings'. Research data can include things like...



Data can be anything and everything!

...just about anything!

Essentially, research data can be just about anything that researchers produce or work with during the course of their research.

Definition of Open Data

Open Data are online, free of cost, accessible data that can be used, reused and distributed provided that the data source is attributed.

Tip - when training use 5 Star Open Data Model to help explain FAIR



make your stuff available on the Web (whatever format) under an open license



make it available as structured data (e.g., Excel instead of image scan of a table)



make it available in a non-proprietary open format (e.g., CSV instead of Excel)

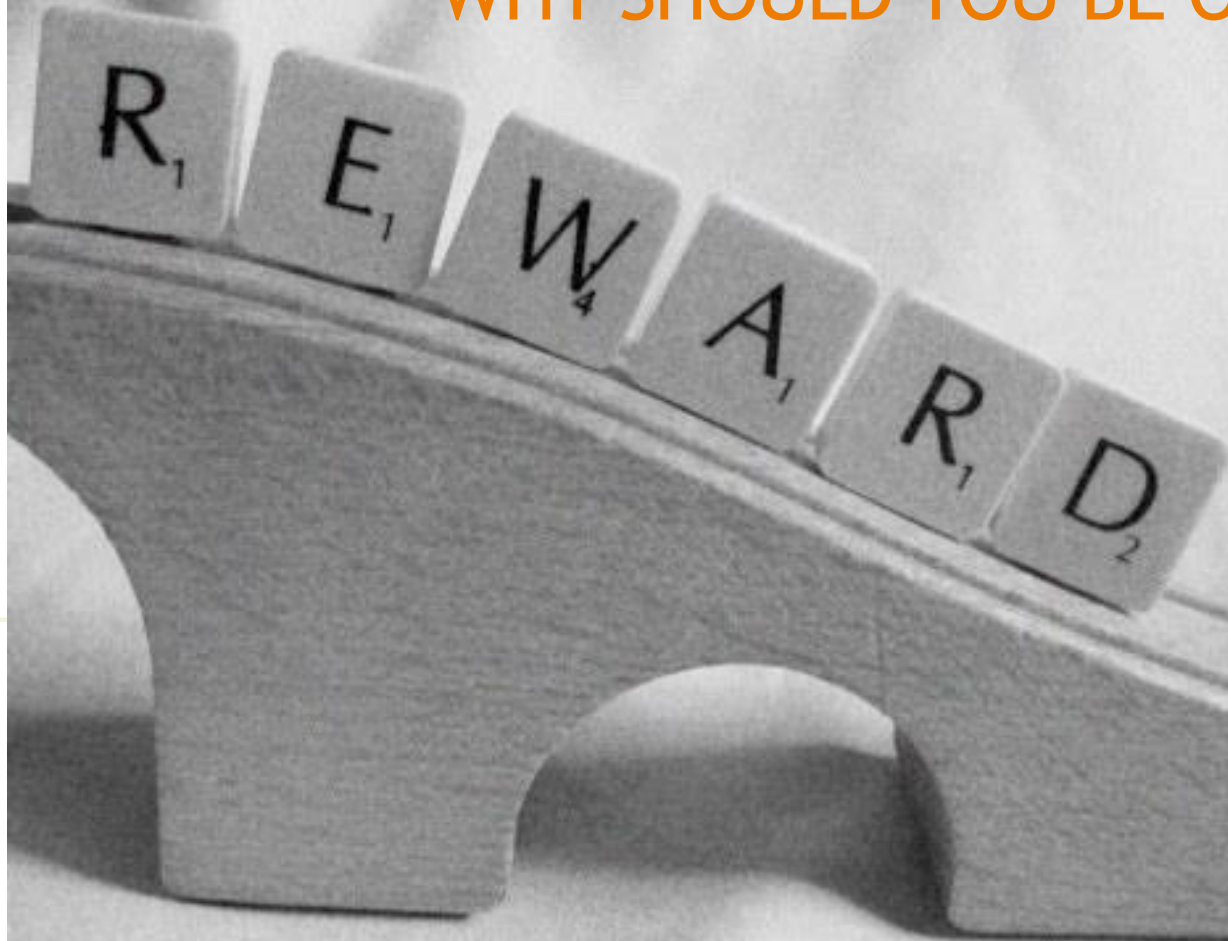


use URIs to denote things, so that people can point at your stuff



link your data to other data to provide context

WHY SHOULD YOU BE OPEN?





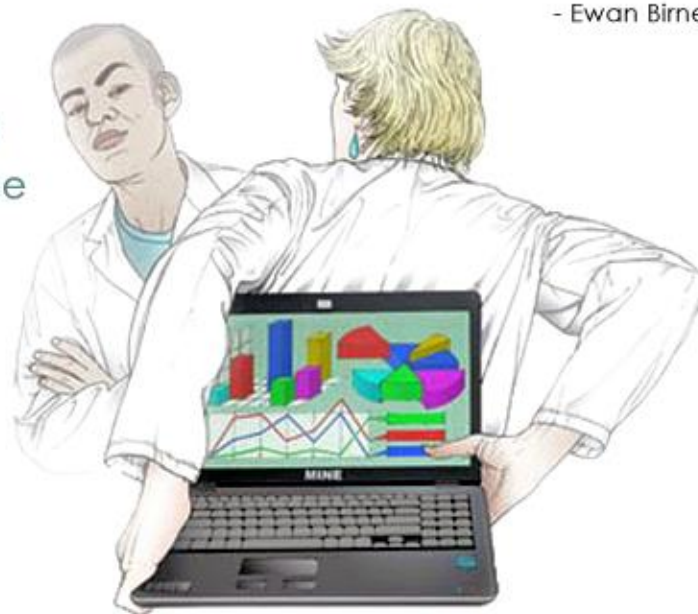
PUBLICATIONS AND DATA

It's part of good research practice

"It was *never* acceptable to publish papers without making data available."

- Ewan Birney

#OpenData
#OpenScience

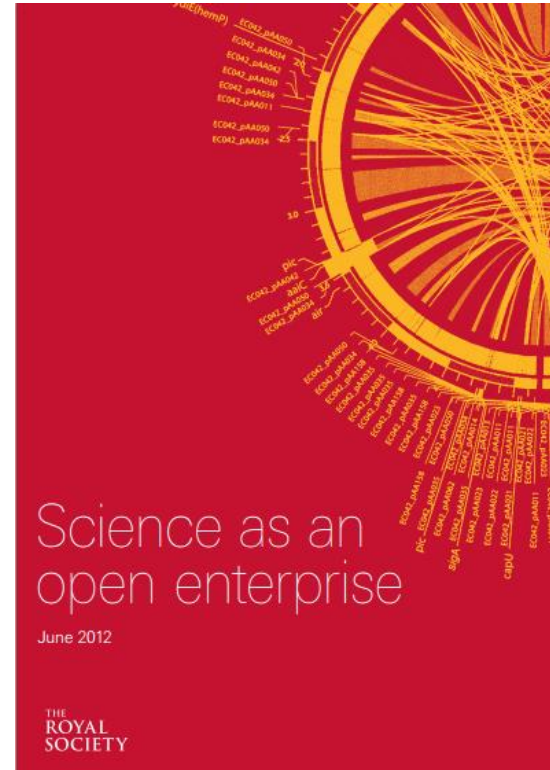


Original image via doi:10.1038/461145a. "Research cannot flourish if data are not preserved and made accessible. Data management should be woven into every course in science." - *Nature* 461, 145

Science as an open enterprise

“Much of the remarkable growth of scientific understanding in recent centuries is due to open practices; open communication and deliberation sit at the heart of scientific practice.”

Royal Society report calls for ‘intelligent openness’ whereby data are accessible, intelligible, assessable and usable.



<https://royalsociety.org/policy/projects/science-public-enterprise/Report>

Cut down on academic fraud

nature International weekly journal of science Login

[nature news home](#) [news archive](#) [specials](#) [opinion](#) [features](#) [news blog](#) [nature journal](#)

[comments on this story](#)
Published online 1 November 2011 | *Nature* **479**, 15 (2011) | doi:10.1038/479015a
[Updated](#) online: 1 November 2011
[Updated](#) online: 8 December 2011

News

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology papers contain faked data.

Even Callaway

When colleagues called the work of Dutch psychologist Diederik Stapel too good to be true, they meant it as a compliment. But a preliminary investigative report (go.nature.com/tamp5c) released on 31 October gives literal meaning to the phrase, detailing years of data manipulation and blatant fabrication by the prominent Tilburg University researcher.

"We have some 30 papers in peer-reviewed journals where we are actually sure that they are fake, and there are more to come," says Pim Levelt, chair of the committee that investigated Stapel's work at the university.

Stapel's eye-catching studies on aspects of social behaviour such as power and stereotyping garnered wide press coverage. For example, in a recent *Science* paper (which the investigation has not identified as fraudulent), Stapel reported that untidy environments encouraged discrimination ([Science 332, 251-253; 2011](#)).



Dutch psychologist Diederik Stapel.
Persbureau van Eindhoven

Related stories

- [Seven days: 9-15 September 2011](#)
14 September 2011
- [Chaos promotes stereotyping](#)
07 April 2011

Naturejobs

Tenure-Track Faculty Positions (Assistant / Associate / Full Professor) Yale University, Department of Genetics
Yale University School of Medicine

Assistant Professor
Harvard Medical School

- [More science jobs](#)
- [Post a job for free](#)

Resources

- [PDF Format](#)
- [Send to a Friend](#)
- [Reprints & Permissions](#)
- [RSS Feeds](#)

external links

- [Tilburg University](#)
- [Interim investigation report](#)

Stories by subject

- [Brain and behaviour](#)
- [Lab life](#)

Stories by keywords

- [Diederik Stapel](#)
- [Tilburg University](#)
- [Academic fraud](#)
- [Retractions](#)
- [Social psychology](#)

This article elsewhere

- [Blogs linking to this article](#)

[Add to Digg](#)
[Add to Facebook](#)
[Add to Newsvine](#)
[Add to Del.icio.us](#)
[Add to Twitter](#)

Validation of results

“It was a mistake in a spreadsheet that could have been easily overlooked: a few rows left out of an equation to average the values in a column.

The spreadsheet was used to draw the conclusion of an influential 2010 economics paper: that public debt of more than 90% of GDP slows down growth. This conclusion was later cited by the International Monetary Fund and the UK Treasury to justify programmes of austerity that have arguably led to riots, poverty and lost jobs.”

www.guardian.co.uk/politics/2013/apr/18/uncovered-error-george-osborne-austerity

The error that could subvert George Osborne's austerity programme

The theories on which the chancellor based his cuts policies have been shown to be based on an embarrassing mistake

Charles Arthur and Phillip Inman
The Guardian, Thursday 18 April 2013 21.10 BST



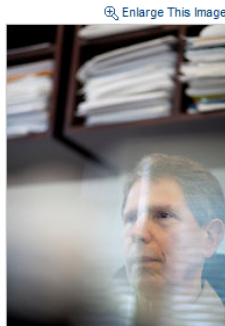
George Osborne says that Ken Rogoff, the man whose economic error has been uncovered, has strongly influenced his thinking. Photograph: Stefan Wermuth/PA

More scientific breakthroughs

Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA
Published: August 12, 2010

In 2003, a group of scientists and executives from the [National Institutes of Health](#), the [Food and Drug Administration](#), the drug and medical-imaging industries, universities and nonprofit groups joined in a project that experts say had no precedent: a collaborative effort to find the biological markers that show the progression of [Alzheimer's disease](#) in the human brain.



Now, the effort is bearing fruit with a wealth of recent scientific papers on the early diagnosis of Alzheimer's using methods like PET scans and tests of spinal fluid. More than 100 studies are under way to test drugs that might slow or stop the disease.

And the collaboration is already serving as a model for similar efforts against [Parkinson's disease](#). A \$40 million project to look for biomarkers for Parkinson's, sponsored by the [Michael J. Fox Foundation](#), plans to enroll 600 study subjects in the United States and Europe.

“It was unbelievable. Its not science the way most of us have practiced in our careers. But we all realised that we would never get biomarkers unless all of us parked our egos and intellectual property noses outside the door and agreed that all of our data would be public immediately.”

Dr John Trojanowski, University of Pennsylvania

www.nytimes.com/2010/08/13/health/research/13alzheimer.html?pagewanted=all&_r=0

A citation advantage

A study that analysed the citation counts of 10,555 papers on gene expression studies that created microarray data, showed:

“studies that made data available in a public repository received 9% more citations than similar studies for which the data was not made available”



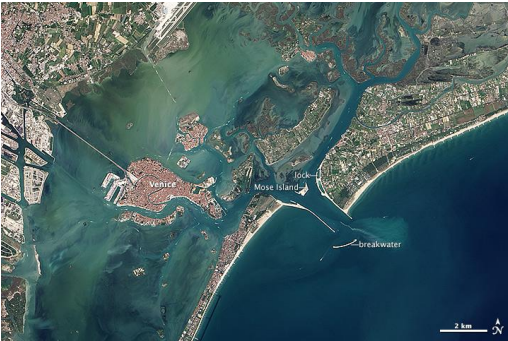
Data reuse and the open data citation advantage,
Piwowar, H. & Vision, T. <https://peerj.com/articles/175>

Increased use and economic benefit

The case of NASA Landsat satellite imagery of the Earth's surface:

Up to 2008

- Sold through the US Geological Survey for US\$600 per scene
- Sales of 19,000 scenes per year
- Annual revenue of \$11.4 million



Since 2009

- Freely available over the internet
- Google Earth now uses the images
- Transmission of 2,100,000 scenes per year.
- Estimated to have created value for the environmental management industry of \$935 million, with direct benefit of more than \$100 million per year to the US economy
- Has stimulated the development of applications from a large number of companies worldwide

<http://earthobservatory.nasa.gov/IOTD/view.php?id=83394&src=ve>



Case study - benefits to researchers in the Arts and Humanities

This video by DANS Data Archiving features Dutch historian Martijn Kleppe who explains why he chose to share his photo database with other researchers, and quantitative data analyst Manfred te Grotenhuis talks about some of the treasures in data archives that are waiting to be discovered.



BE PART OF THE NEW ERA OF OPEN SCIENCE



reach more
people,
have greater
impact



avoid
duplication
of efforts



preserve data
for future
researchers



simplify final
Horizon 2020
reporting
thanks to an
up-to-date DMP



Looking after your data

Explain how you will manage your data, noting particular concerns or issues (e.g. storage and backup, data structuring, versioning, documentation)

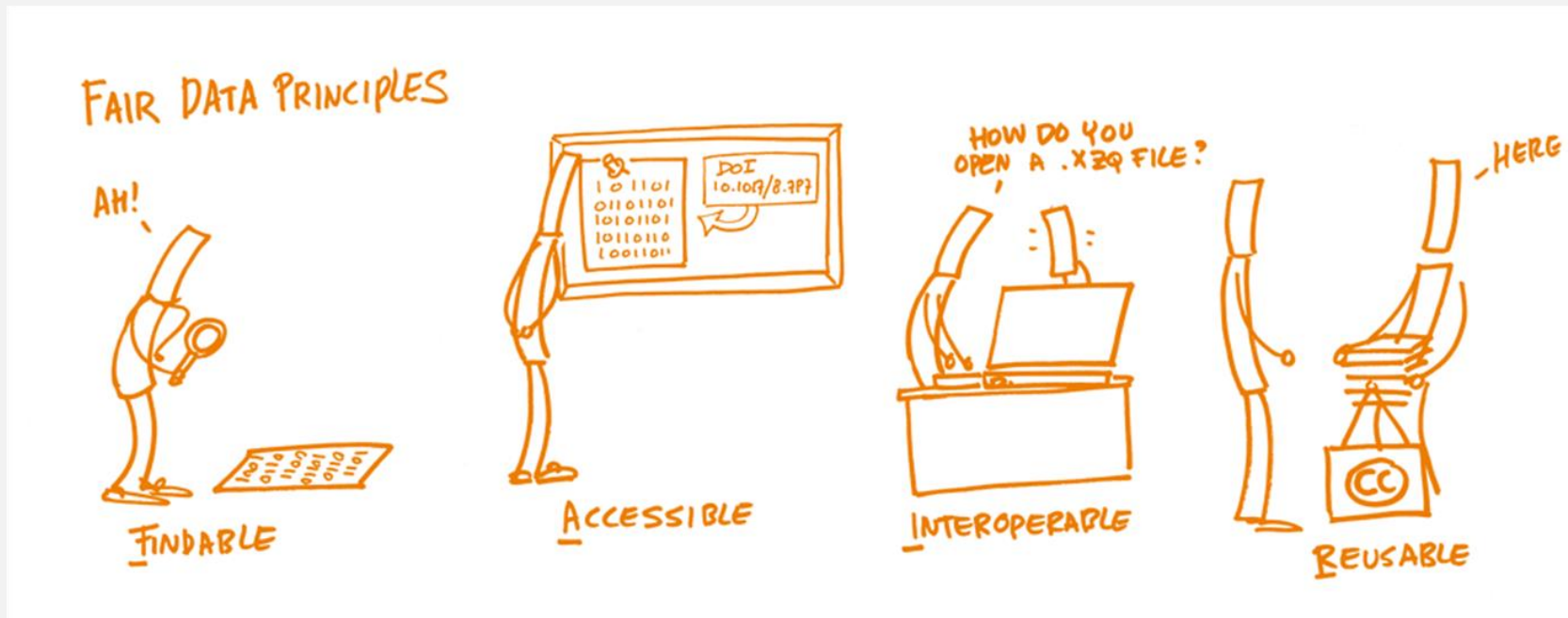


Sharing your data

Explain which data will be shared and how
(e.g. via repository, under what licence)

Misconception #1:

My web page is a FAIR way to share my data.



Better options for open data

- Domain repository (first choice)
- General repository (Figshare, Zenodo, Dryad)
- Institutional repository
- Data journal
- Journal supplementary material



re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

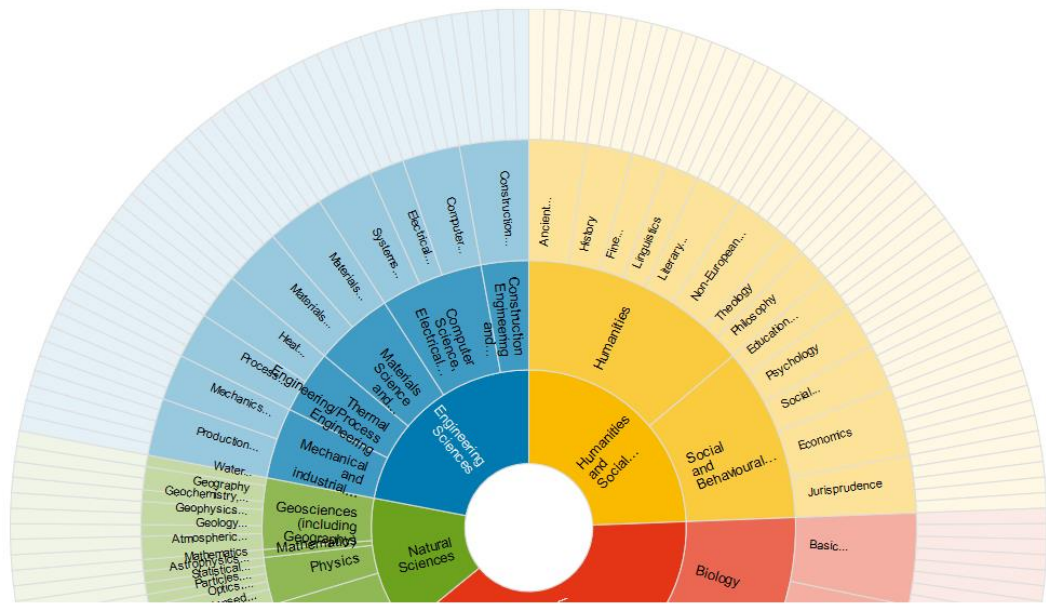
 

Browse by subject

Graphical

Text

click to zoom into subjects or to select a bottommost subject in the hierarchy as filter for the re3data search page
 ctrl + click on a top subject to select it as filter



Browse by subject

[Graphical](#)[Text](#)

A. Humanities and Social Sciences

a. Humanities

I. Ancient Cultures

1. Prehistory
2. Classical Philology
3. Ancient History
4. Classical Archaeology
5. Egyptology and Ancient Near Eastern Studies

II. History

1. Medieval History
2. Early Modern History
3. Modern and Current History
4. History of Science

III. Fine Arts, Music, Theatre and Media Studies

1. Art History
2. Musicology
3. Theatre and Media Studies

IV. Linguistics

1. General and Applied Linguistics
2. Individual Linguistics
3. Typology, Non-European Languages, Historical Linguistics

V. Literary Studies

1. Medieval German Literature
2. Modern German Literature
3. European and American Literature
4. General and Comparative Literature and Cultural Studies

Browse by content type

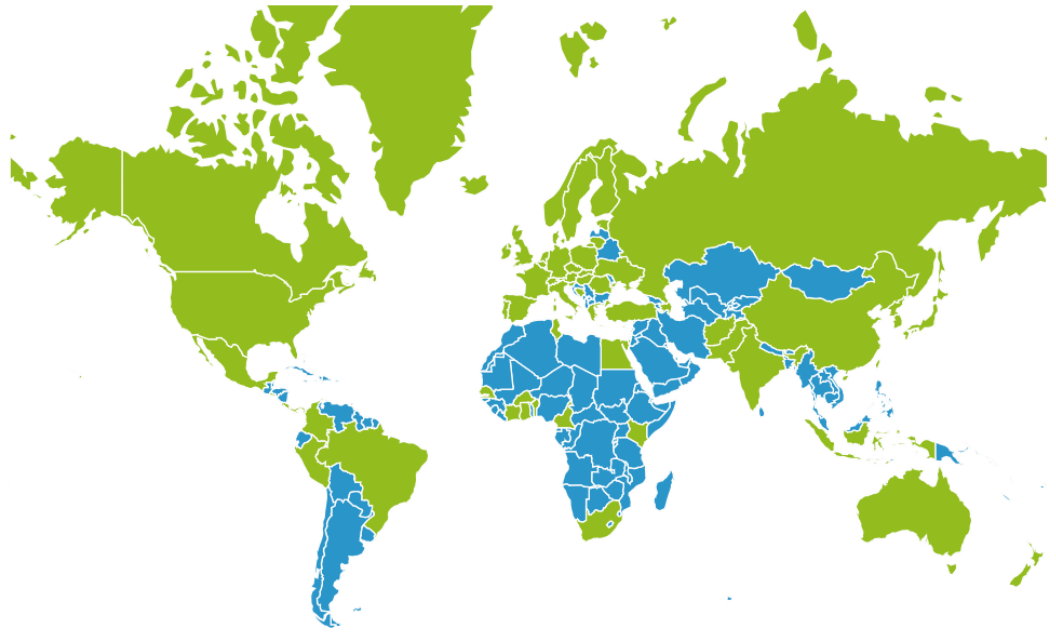
- [Archived data](#)
- [Audiovisual data](#)
- [Configuration data](#)
- [Databases](#)
- [Images](#)
- [Networkbased data](#)
- [Plain text](#)
- [Raw data](#)
- [Scientific and statistical data formats](#)
- [Software applications](#)
- [Source code](#)
- [Standard office documents](#)
- [Structured graphics](#)
- [Structured text](#)
- [other](#)



Browse by country

Graphical

Text



Browse by country

Graphical

Text

International

-  Afghanistan
-  Australia
-  Austria
-  Azerbaijan
-  Belgium
-  Benin
-  Burkina Faso
-  Bosnia and Herzegovina
-  Brazil
-  Canada
-  Switzerland
-  China
-  Cote d'Ivoire
-  Cameroon
-  Colombia
-  Costa Rica
-  Cyprus
-  Czech Republic
-  Germany
-  Denmark
-  European Union
-  Egypt
-  Spain
-  Estonia

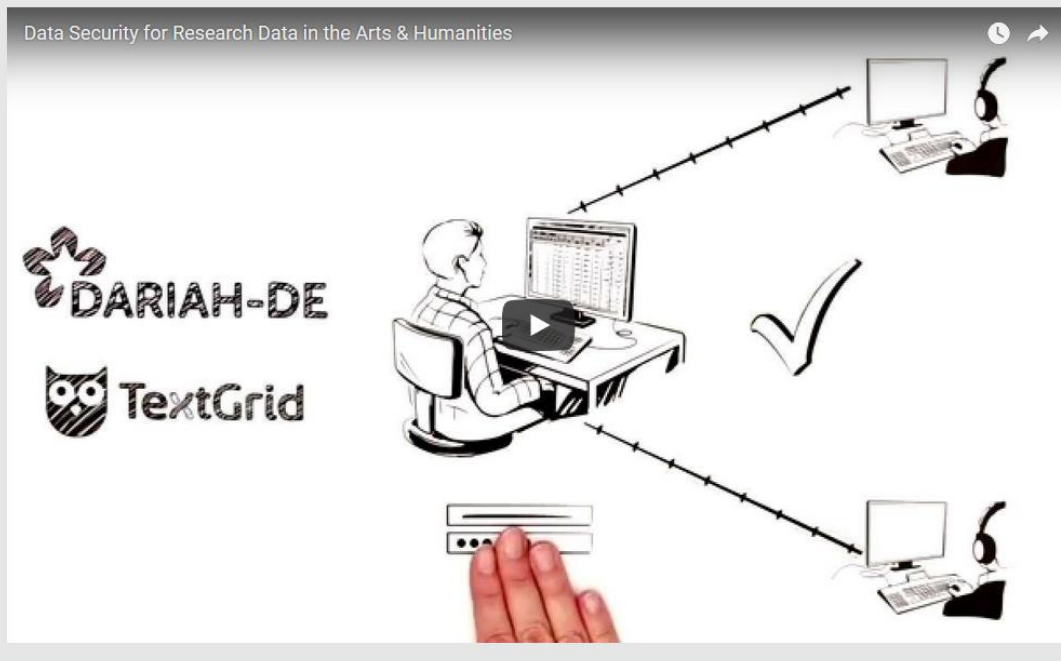
There are ways to share sensitive data too

- Open metadata
- Data brokers and data access committees
- Safe havens
- Institutional data archive/vault



Secure access using the DARIAH-DE Repository

This video explains how DARIAH-DE provides secure access to data for Arts and Humanities researchers.



Tip - some repository decisions are tricky

- There may be a preferred repository that the funder expects
- Data from multidisciplinary studies may not have an obvious home
- Data types and volumes will also need to be taken into account

Misconception #2:

I don't need to decide now if I want to share.
I can wait and see what I want to do at the
end of my project.



Open Data doesn't just happen - data management planning helps!

- What data will be created (format, types, volume...)
- Standards and methodologies to be used (incl. metadata)
- How ethics and Intellectual Property will be addressed
- Plans for data sharing and access
- Strategy for long-term preservation

A DMP is a plan to share!



Tip - use existing tools and guidance to help write their plans



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

Data management planning tools - DMPonline

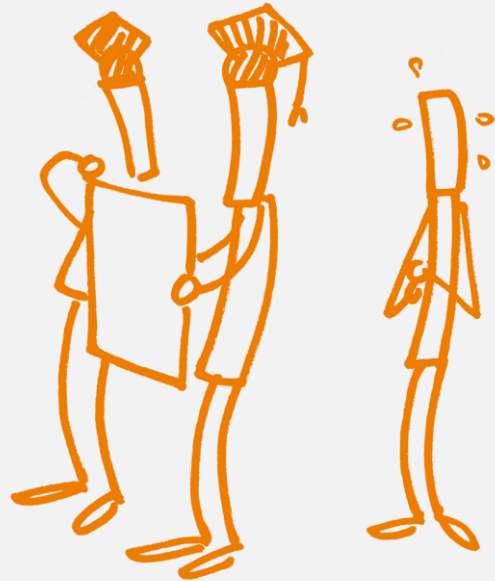
DMPonline is a freely available tool that helps research teams to write data management plans that meet funding body requirements. DMPonline was jointly developed by the Digital Curation Centre (DCC) and the University of California Curation Center (UC3). The tool contains a number of templates that represent the requirements of different funding bodies across Europe. Users are asked three questions at the outset to determine the appropriate template to display (e.g. the Economic and Social Research Council (ESRC) template when applying for an ESRC grant). Using tools like DMPonline takes the guesswork out of writing your data management plan by providing you with the specific set of questions that individual funding bodies want you to answer. The tool also provides users with general guidance - and where provided, institutional guidance - to make sure that your answers are realistic and implementable.

For more information on data management plans and tips on writing them, check out the [DCC website](#).

The screenshot shows the DMPonline website interface. At the top is an orange navigation bar with the DMPonline logo and menu items: Home, Public DMPs, Funder requirements, and Help. A language dropdown is on the right. The main content area has a 'Welcome' section with a brief description of the tool and a link to join the international community. Below this are four statistics: 17,622 Users, 203 Organisations, 23,083 Plans, and 89 Countries, each with a corresponding icon. A sign-in form is on the right, featuring fields for email and password, a 'Remember email' checkbox, and buttons for 'Sign in' and 'Sign in with institutional credentials (UK only)'. A 'Forgot password?' link is also present.

Misconception #3:

If I share my data early, I'll be scooped!



Pre-registration timestamps your work

Register Your Project



Open Science Framework

A registration on OSF creates a frozen, time-stamped version of a project that cannot be edited or deleted. The *original project* can still be edited, while the registered version cannot. You might create a registration to capture a snapshot of your project at certain points in time - such as right before data collection begins, when you submit a manuscript for peer review, or upon completion of a project.

Registrations can be made public immediately or embargoed for up to 4 years. Registrations cannot be deleted, but they can be withdrawn. [Withdrawing a registration](#) removes the content of the registration but leaves behind basic metadata, like registration title, contributors, and a reason for the withdrawal (not required).

<http://help.osf.io/m/registrations/l/524205-register-your-project>

Tips - share preprints too

- Early feedback on methods and initial findings
- Time to correct and mistakes before publishing
- Recognition for your ideas by peers

Misconception #4:

I have to keep and share everything.



Deciding which data need to be kept after the project ends

Five steps to follow

- ① **Could** this data be re-used
- ② **Must** it be kept as evidence or for legal reasons
- ③ **Should** it be kept for its potential value
- ④ **Consider costs** – do benefits outweigh cost?
- ⑤ **Evaluate criteria** to decide what to keep

5 steps to decide what data to keep

www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep

Tip - link data to other outputs for context (reuse)

Open Data



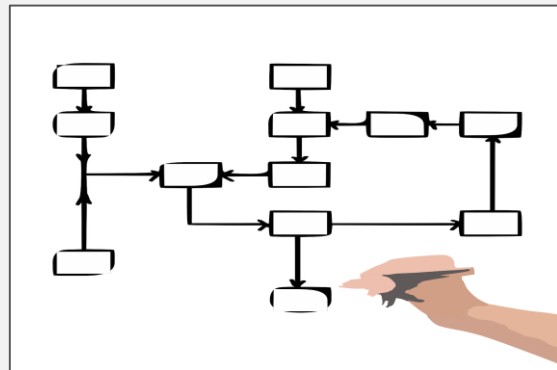
To support validation and facilitate reuse

Open Code



Software created to analyse and/or visualise the data

Open Workflows



What steps were taken and in what order?

Consider who else has a say about sharing data

- Collaborators
- Research participants
- Commercial partners
- Data repository



Exercise: barriers to data sharing

In groups of 2-3, consider any barriers to sharing data.

If there are any specific issues for your discipline please feel free to note these.

10 minutes plus feedback

How to make data open?



<https://okfn.org>

1. Choose your dataset(s)
 - What can you open? You may need to revisit this step if you encounter problems later.
2. Apply an open license
 - Determine what IP exists. Apply a suitable licence e.g. CC-BY
3. Make the data available
 - Provide the data in a suitable format. Use repositories.
4. Make it discoverable
 - Post on the web, register in catalogues...



WHAT IS A DMP & WHY WRITE ONE?

Data Management Plans (DMP)

A DMP is a brief plan to define:

- how the data will be created
- how it will be documented
- who will be able to access it
- where it will be stored
- who will back it up
- whether (and how) it will be shared & preserved

DMPs are often submitted as part of grant applications, but are useful whenever researchers are creating data.



European
Commission



OPEN RESEARCH DATA IN HORIZON 2020

**Jean-François Dechamp
& Daniel Spichtinger**

European Commission
Directorate-General for Research &
Innovation

RESEARCH DATA - OPEN BY DEFAULT



Making data FAIR

Findable - Assign persistent IDs, provide rich metadata, register in a searchable resource,...

Accessible - Retrievable by their ID using a standard protocol, metadata remain accessible even if data aren't...

Interoperable - Use formal, broadly applicable languages, use standard vocabularies, qualified references...

Reusable - Rich, accurate metadata, clear licences, provenance, use of community standards

www.force11.org/group/fairgroup/fairprinciples

Why do we need FAIR data?

Barend Mons, Professor of BioSemantics at the Human Genetics Department of Leiden University Medical Center explains why we need FAIR data in this short video 'Social machines & FAIR data' (in Dutch with English subtitles).



Managing and Sharing Research Data

https://www.fosteropenscience.eu/learning/managing-and-sharing-research-data


FOSTER

Assessing the FAIRness of your data

FAIR self-assessment tool

This [self-assessment tool](#) was developed by the [Australian ANDS-Nectar-RDS initiative](#) and enables you to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness. It was developed primarily for research support staff but can be used by anyone.

ANDS, Nectar and RDS are supported by the Australian Government through the [National Collaborative Research Infrastructure Strategy](#) program.



FAIR data self-assessment tool

The new FAIR data self-assessment tool enables you to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable). This self-assessment tool has been designed predominantly for data librarians and IT staff, but could be used by anyone.

[Check out the tool](#)

ANDS-Nectar-RDS FAIR self-assessment tool

<https://www.fosteropenscience.eu/learning/managing-and-sharing-research-data>

European Research Infrastructure to support FAIR data

There are a number of existing and emerging research infrastructures being developed through support from the European Commission and other European funding bodies and governments. Consider making use of these to help you to manage and share your data. To find out more about the range of research infrastructures available for your discipline, check out the [European Strategy Forum on Research Infrastructures \(ESFRI\) Roadmap](#).



The poster features a dark blue background with a network of white lines and dots. Three large, concentric white circles are arranged horizontally across the middle. The text is in white and yellow. At the top left is the 'PL' logo. The title 'STRATEGY REPORT ON RESEARCH INFRASTRUCTURES' is in large white letters. Below the circles are three columns of text describing online navigation, local consultation, and downloading for printing.

PL ESFRI ROADMAP 2016

STRATEGY REPORT ON RESEARCH INFRASTRUCTURES

Navigate online
You can read on-line, select cross-links, search for information by Country and Science domain

Download for local consultation
You can download an Interactive PDF for local reading and browsing

Download for printing
You can download a Print-Ready PDF

Case study - making Social Science research data FAIR using the CESDDA ERIC infrastructure

This video introduces [CESDDA ERIC](#) - a European Research Infrastructure for the Social Sciences.



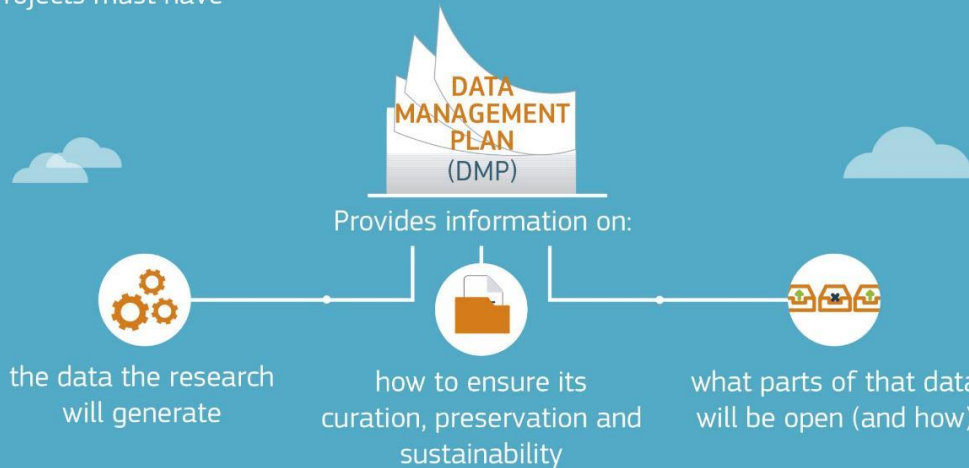
Case study - making Arts and Humanities research data FAIR using the DARIAH-DE infrastructure

This brief overview of [DARIAH-DE](#) research infrastructure shows you how their tools and resources can help you to share and store Arts and Humanities data and work collaboratively.

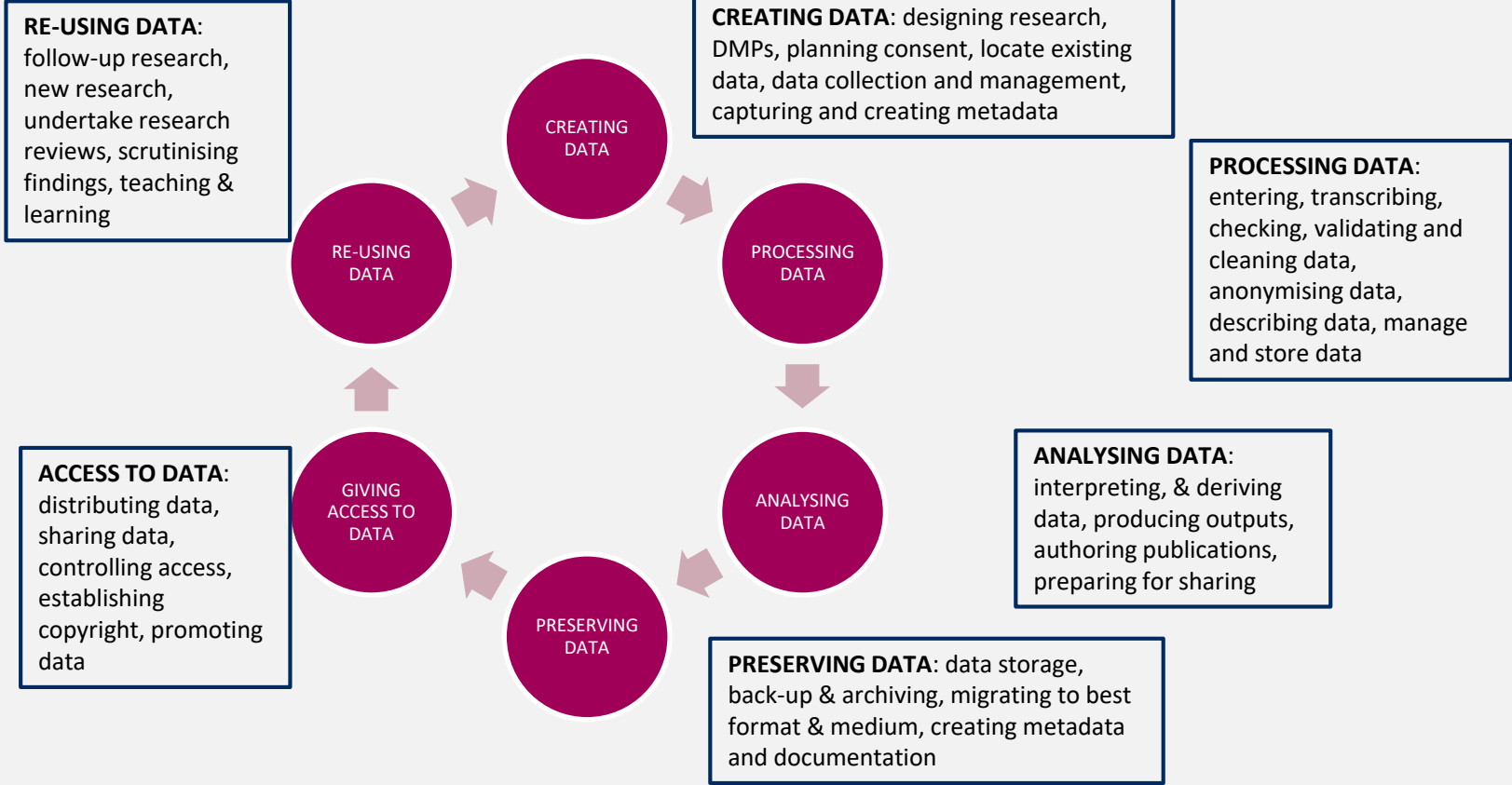


RESEARCH DATA - OPEN BY DEFAULT

Projects must have



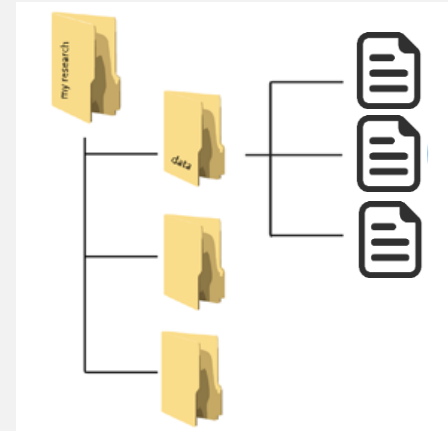
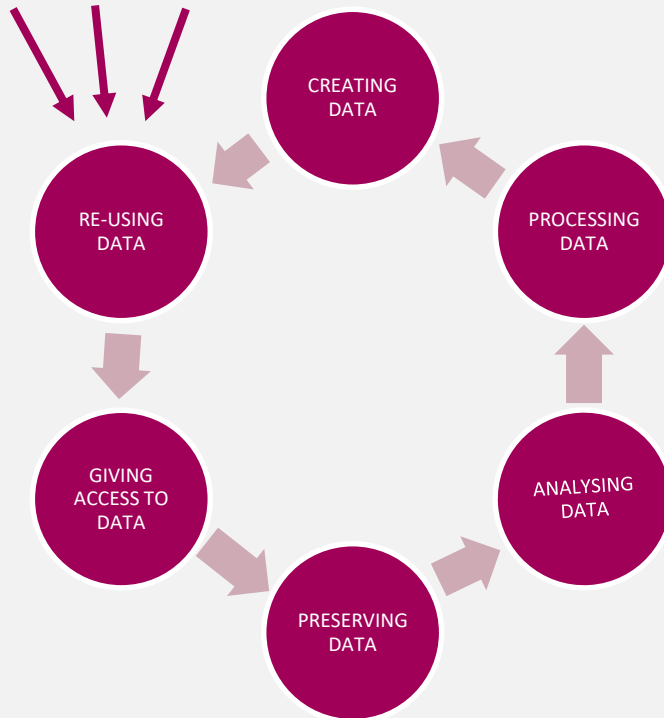
Research data lifecycle



Ref: UK Data Archive: <http://www.data-archive.ac.uk/create-manage/life-cycle>

Planning trick 1: think backwards

What data organisation would a re-user like?



Data organisation

Meaningful file names

Below are tips on meaningful and consistent file names. Read more in '[Choosing a file name](#)'.⁽²⁾

- ❑ Make sure to use consistent file names. When you use a date in the file name, choose a notation (for instance, YYYYMMDD of yymmdd).
- ❑ Do not use strange characters like ?\!@*%{[<> in the file name.
- ❑ Use traceable file names, such as Project_Instrument_locatie_YYYYMMDD.ext.
- ❑ Make sure to only use each file once in the folder structure. If you store a file in more than one place, several versions of the same file can unwillingly be created.
- ❑ See also [version management](#).

It is good practice to note the file naming and its meaning in a readme.txt.

Even if a researcher is well underway with his project consistent file naming is still an option by using a [bulk file rename utility](#).⁽³⁾ It is important, however, to check if this bulk renamer delivers on its promises.



white_data_20140708.csv



blue_data_20140708.docx



red_data_20140708.R



red_data_20140708_v02.R

File naming and version management

Expert Tour Guide on Data Management

1. Plan

2. Organise & Document

Designing a data file structure

Organisation of variables

File naming and folder structure

Documentation and metadata

Adapt your DMP: part 2

Sources and further reading

3. Process

4. Store

5. Protect

6. Archive & Publish

TIP: Batch renaming of automatically generated files



Batch renaming is organising research data files and folders in a consistent and automated way with software tools (also known as mass file renaming, bulk renaming).

Batch renaming software exists for most operating systems. See the accordion for examples.

+ Batch renaming tools

It may be useful to rename files in a batch when:

- Images from digital cameras are automatically assigned base filenames consisting of sequential numbers;
- Proprietary software or instrumentation generate crude, default or multiple filenames;
- Files are transferred from a system that supports spaces and/or non-English characters in filenames to one that doesn't (or vice versa). Batch renaming software can be used to substitute such characters with acceptable ones.

Expert Tour Guide on Data Management

1. Plan

2. Organise & Document

Designing a data file structure

Organisation of variables

File naming and folder structure

Documentation and metadata

Adapt your DMP: part 2

Sources and further reading

3. Process

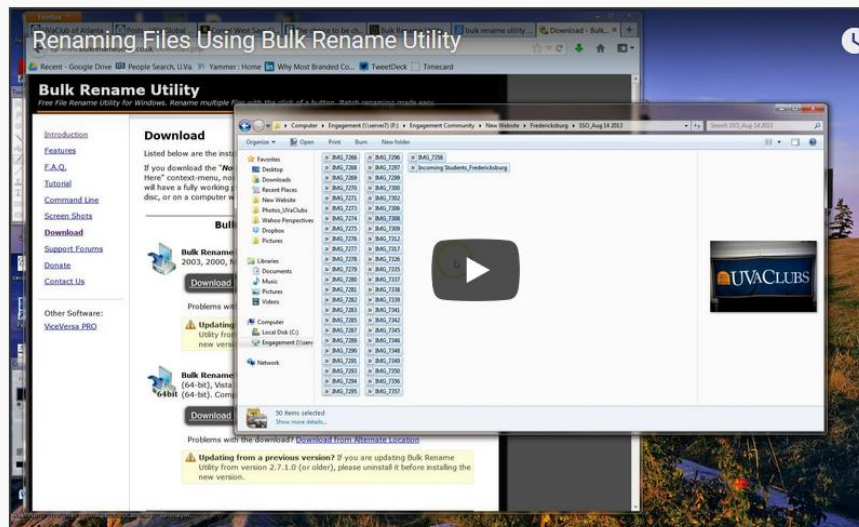
4. Store

5. Protect

6. Archive & Publish

How to ... use Bulk Rename Utility

Follow the steps in the video to use Bulk Rename Utility to batch rename your files.



« Previous

Next »

<https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/2.-Organise-Document/File-naming-and-folder-structure>



In this section

Funding Opportunities

› Research Funding

› Research Funding Guide

Email response templates

Monitoring, ROS and
Researchfish

Panel Outcomes

Subject Coverage

Independent Research
Organisations

Museums and Galleries

International Funding

[Home](#) > [Funding](#) > [Research Funding](#) > [Research Funding Guide](#)
> [Attachments](#) > [Technical plan](#)

Technical plan

Naming convention: [PI Surname] TechP

Before reading this section, please see the [Case for Support Guidance](#) regarding a Technical Summary.

A Technical Plan should be no more than four pages long and provided for all applications where digital outputs or digital technologies are an essential part to the planned research outcomes. A digital output or digital technology is defined as an activity which involves the creation, gathering, collecting and/or processing of digital information. For present purposes digital technologies do not include conventional software such as word processing packages and [ICT](#) activities such as email.

Please read this guidance carefully and consider its definitions within the context of your own research proposal.

The purpose of the Technical Plan is to demonstrate to the AHRC that technical provisions within a research proposal have been adequately addressed in terms of:

(a) Delivering the planned digital output or the digital technology from a practical and

Naming convention: [PI Surname]
TechP

Before reading this section, please see the [Case for Support Guidance](#) regarding a Technical Summary.

A Technical Plan should be no more than four pages long and provided for all applications where digital outputs or digital technologies are an essential part to the planned research outcomes. A digital output or digital technology is defined as an activity which involves the creation, gathering, collecting and/or processing of digital information. For present purposes digital technologies do not include conventional software such as word processing packages and [ICT](#) activities such as email.

Please read this guidance carefully and consider its definitions within the context of your own research proposal.

The purpose of the Technical Plan is to

Why manage data?

NON PECUNIAE INVESTIGATIONIS CURATORE

SED VITAE FACIMUS PROGRAMMAS DATORUM

(Not for the research funder, but for life we make data management plans)

PROCURATIONIS

- Make your research easier
- Stop yourself drowning in irrelevant stuff
- Save data for later
- Avoid accusations of fraud or bad science
- Write a data paper
- Share your data for re-use
- Get credit for it

Why bother writing a DMP?

Some researchers are a bit sceptical about the value of writing a data management plan and feel it is just another administrative burden. There are lots of good reasons for writing a data management plan. Here are just a few of them.



Research Integrity

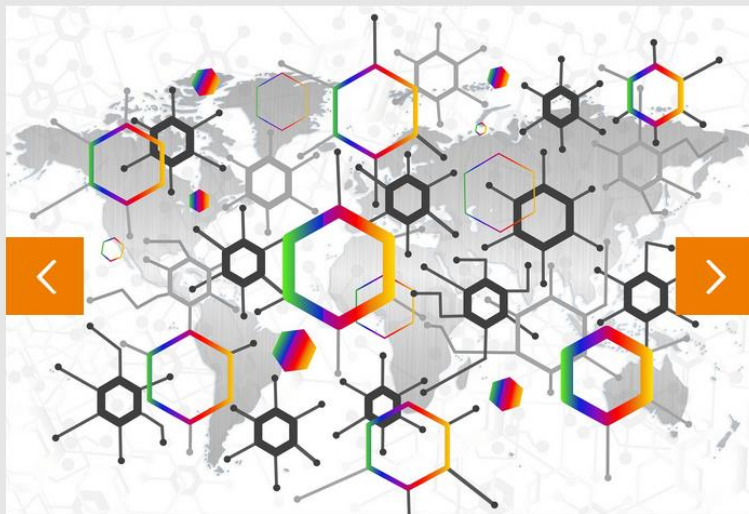
To support research integrity

Data management planning helps:

- avoid accusations of fraud or bad science as you can produce evidence for findings and enable validation
- support good research practice
- meet funding body requirements

Why bother writing a DMP?

Some researchers are a bit sceptical about the value of writing a data management plan and feel it is just another administrative burden. There are lots of good reasons for writing a data management plan. Here are just a few of them.



Sharing

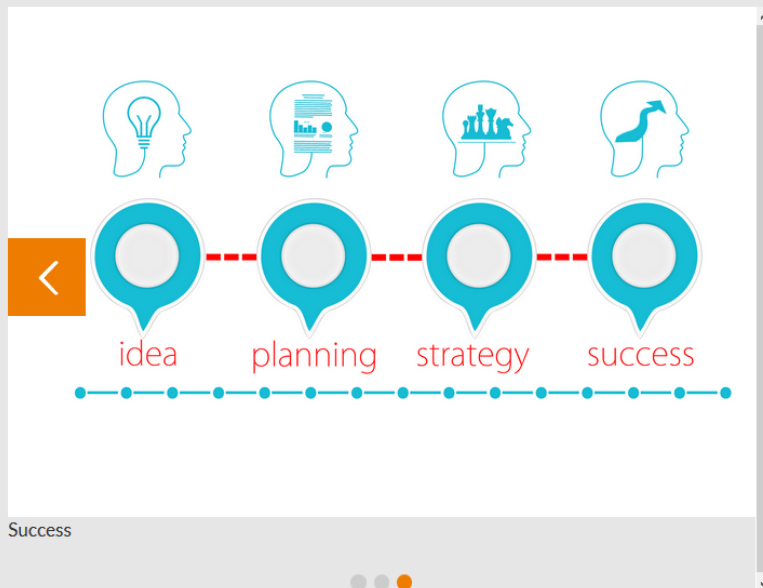
Improves your potential to share

Data management planning supports responsible sharing which helps to:

- enable others to reuse and build on your research
- achieve greater impact and foster new collaborations
- promote innovation and allow research in your field to advance more quickly

Why bother writing a DMP?

Some researchers are a bit sceptical about the value of writing a data management plan and feel it is just another administrative burden. There are lots of good reasons for writing a data management plan. Here are just a few of them.

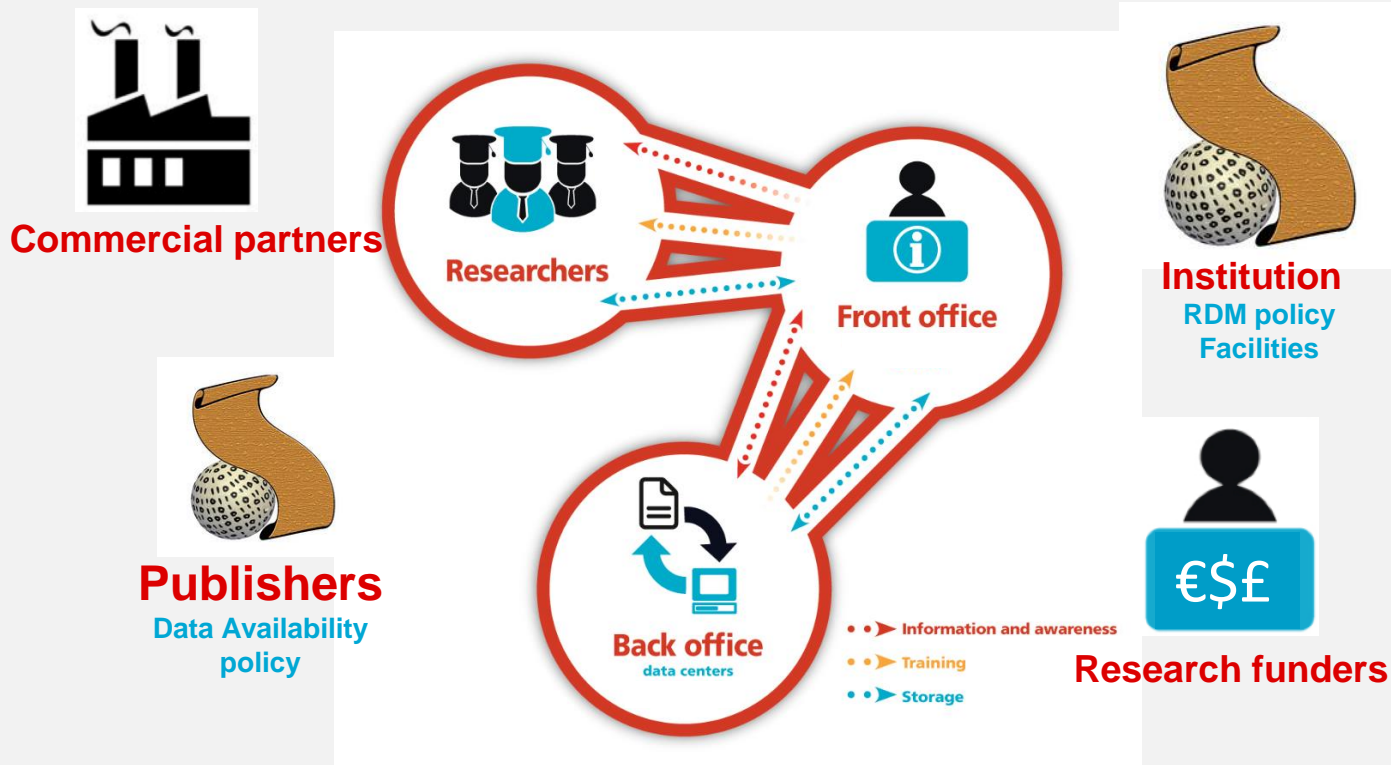


Direct benefits for you!

Data management planning provides benefits for you too! Good data management planning:

- avoids you drowning in irrelevant data
- helps you to know which versions of your data are most up-to-date
- helps you to understand and reuse your own data in the future
- gets you higher citation rates

Planning trick 2: include RDM stakeholders



<https://www.openaire.eu/briefpaper-rdm-infonoads>

Responsibilities in RDM

- **The principal investigator** – ultimately responsible for the data and for data management
- **Researchers, research assistants and/or data managers** – involved in day-to-day data management
- **The institution's management** – draft and enforce data policies; raise data awareness
- **The institution's research office consisting of library, IT and legal services** – provide external data, tools, secure storage and access; expertise on rights management and ethics, data citation, metadata, access and licenses, funder requirements; raise data awareness
- **Research funders** – encourage good data practices; invest in data infrastructure; raise data awareness
- **Project partners** in academic and other research institutions as well as commercial partners
- **Academic publishers** – impose requirements on the availability of data underlying submitted and/or published papers; provide identifiers to cite papers and link to related data
- **Research data repositories** – preserve data long term; provide persistent identifiers and data discovery service

So, who decides whether your research data should be open, closed, or somewhere in between?

Researchers have a key role to play in deciding what data can be shared but it is important to note that they are not the only stakeholders involved in making this decision.

- + Research participants
- + Research collaborators
- + Research data infrastructure
- + Research data repositories
- + Secondary data reusers

— Research participants

If your research will involve working with human subjects, you will need to ensure that you obtain informed consent. Informed consent should let research participants know about any plans you have for sharing their data (i.e., within your research team or more widely) as well as any plans you have relating to the longer-term retention of their data to support reuse.

Speak to your Ethics Team and make sure that any plans for reuse are built into consent forms you will use. If you plan to make use of a standard consent form available from your institution, be sure to read it through first to make sure it does not contain any blanket statements about not sharing data or promises to destroy the data at the end of the project.

Be sure to consider any data cleaning and/or anonymisation procedures that will need to be carried out to facilitate sharing early on. Bear in mind that these actions can be very costly so be sure to request sufficient budget for these activities in your grant proposals.

For more information on obtaining informed consent, please see our course on Data Protection and Ethics from the main list of courses.

— Research collaborators

If you are involved in a collaborative research project - either with other academic institutions, industry partners, or citizen science - you will need to make sure that your partners agree to data sharing. This should be clarified during the idea stage of your project and built into any consortium agreements that are developed to govern your project. Your Grants or Contracts Team should be able to help you. Bear in mind that your partners may only agree to sharing particular datasets and request that others are kept confidential. It is a good idea to make any such restrictions clear in your research proposal. You should also agree at which point during the life of the project the selected datasets will be shared and with whom, and document these decisions so that all partners have a clear idea of what will happen and when.



– Research data infrastructure

As you are planning your research, consider which research data infrastructure you will use to make the data accessible and who will need access. It can be tempting to make use of Google docs or Dropbox to easily share data with collaborators during the active stage of the project. However, be sure to consider whether you will be working with any personally or commercially sensitive data. If so, these are not the best options for secure data sharing. It is always a good idea to spend some time with your Information Services Team to find out what data sharing infrastructure they can support in-house. In most cases, managed infrastructure provided by your institution will provide better security and back-up and can often be made accessible to collaborators from other institutions. Be sure to discuss the scale and format(s) of the data you anticipate sharing and also consider how frequently access will be required (daily, monthly, annually). These aspects will inform the best storage and sharing options for your data during the life of your project and beyond.



<https://www.fosteropenscience.eu/learning/managing-and-sharing-research-data>

– Research data repositories

When selecting a repository for those data selected for longer-term retention and sharing, be sure to check that the repository meets your needs. For instance, if your data will only be shared with a specific research community you will need to ensure that the repository can provide a means of allowing researchers to request access and to be authenticated.

Most data repositories have policies outlining any limits relating to the size of data deposit or restrictions on formats they will accept. Be aware of any normalisation processes that are undertaken by the repository (i.e., when deposited data are migrated to preferred formats). In many cases, normalisation can affect the usability of the data. For example, if an Excel spreadsheet that was presented in a publication is saved as a PDF, it will be available as a record of what was presented in the article but will lose much of the functionality needed to support validation and reuse. For instance, any formulas applied to individual cells within the spreadsheet will be lost. It will also mean that reuse of the data is more difficult and would require re-keying the data into a new spreadsheet.



<https://www.fosteropenscience.eu/learning/managing-and-sharing-research-data>

– Secondary data reusers

If you have concerns that your data will be misused and this is putting you off sharing, bear in mind that you can still make the data accessible while requesting that potential reusers accept general terms and conditions regarding fair use. For example, you may request that researchers wishing to reuse the data tick a box indicating that they will not attempt to identify anonymised participants. The [UK Data Service End User License \(EUL\)](#) provides an example of how this approach is put into practice for a national data archive.



Developing a DMP helps to avoid data management nightmares!

This short video provides some pointers on things that should be included in your DMP to help you avoid problems during your research. The video was created by NYU Health Sciences Library's Karen Hanson, Kevin Read, and Alisa Surkis.



A DMP is about 'keeping' data



- Storing data < > archiving data
 - Archived data < > findable data
 - Findable < > accessible
 - Accessible < > understandable
 - Understandable < > usable
-
- A USB stick is not safe
 - A persistent ID is essential but no guarantee for usability
 - Data in a proprietary format is not sustainable

How to deal with data and context?

Versioning, back-up, storage and archiving

– During the project and in the long term

Ethics, consent forms, legal access

Security and technical access

Usage licences



What should be preserved and shared?

- The **data** needed to validate results in scientific publications (minimally!).
- The associated **metadata**: the dataset's creator, title, year of publication, repository, identifier etc.
 - Follow a metadata standard in your line of work, or a generic standard, e.g. Dublin Core or DataCite, and be FAIR.
 - The repository will assign a persistent ID to the dataset: important for discovering and citing the data.

What should be preserved and shared? (2)

- **Documentation**: code books, lab journals, informed consent forms - domain-dependent, and important for understanding the data and combining them with other data sources.
- **Software**, hardware, tools, syntax queries, machine configurations - domain-dependent, and important for using the data. (Alternative: information about the software etc.)

Basically, everything that is needed to replicate a study should be available. Plus everything that is potentially useful for others.

Data description examples

The final dataset will include self-reported demographic and behavioural data from interviews with the subjects and laboratory data from urine specimens provided.

From [NIH data sharing statements](#)

Metadata examples

Metadata will be tagged in XML using the [Data Documentation Initiative \(DDI\) format](#). The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and [all information necessary for a secondary analyst](#) to use the data accurately and effectively.

From [ICPSR Framework for Creating a DMP](#)



REUSABLE DATA

Use standards of
your domain

General

- Dublin Core (DC)
- Datacite metadata schema
- Metadata Object Description Schema (MODS)

METADATA
STANDARDS

Archives/Repositories

- Datastar minimD-space metadata
- um Metadata


Humanities

- Text Encoding Initiative (TEI)
- Visual Resources Association Core (VRA)

Social Science

- Data Documentation Initiative (DDI)

Contact us

 **D|C|C** because good research needs good data **Search**

Home | Digital curation | About us | News | Events | Resources | Training | Projects | Community | Tailored support

<http://www.dcc.ac.uk/resources/subject-areas/social-science-humanities>

Home > Resources > Subject Areas > Social Science Humanities

Social Science & Humanities

Archaeology General Architecture **Economics** Historical and
Philosophical Studies Law **Social Policy** Heritage Studies Anthropology
Human and Social Geography **Statistics** Health Policy Music
Planning (Urban, Rural and Regional) Politics History by Area
Sociology Rural and Regional) Planning (Urban Creative art and design **Demography**
History Building Conservation Multi-disciplinary

Metadata Standards

DDI - Data Documentation Initiative

An international standard for describing data from the social, behavioral, and economic sciences. Expressed in XML, the DDI metadata specification supports the entire research data life cycle.

MIDAS-Heritage

A British cultural heritage standard for recording information on buildings, archaeological sites, shipwrecks, parks and gardens, battlefields, areas of interest and artefacts.

OAI-ORE - Open Archives Initiative Object Reuse and Exchange

Defines standards for the description and exchange of aggregations of Web resources.

QuDEX - Qualitative Data Exchange Format

A qualitative data exchange model for the archiving and interchange of data.

SDMX - Statistical Data and Metadata Exchange

A set of common technical and statistical standards and guidelines to be used for the efficient exchange and sharing of statistical data and metadata.

DMPonline



DMPonline is the DCC's data management planning tool. In addition to the questions included in the DCC's Checklist for a Data Management Plan, it also contains useful guidance on how to prepare a data management plan and carry it through to execution.

[Read more](#)



Extensions

[CARARE metadata schema](#)

An application profile of the [MIDAS Heritage](#) standard intended for delivering metadata to the CARARE service environment about an organisation's online collections, monument inventory database and digital objects.

[CESSDA MLI - Council of European Social Science Data Archives Minimum Level of Information](#)

A common base profile of DDI for use by the member archives of CESSDA.

[GSMES/TS - GEneric Statistical MESsage for Time Series](#)

An extension of SDMX used to exchange statistical data and metadata.

Tools

[DDI on Rails](#)

Server-side software for building a data portal, with a particular focus on survey datasets. It uses DDI to provide access to the data at the level of concepts and variables. For an example of it in use, see the [SOEPinfo data portal](#).

[DDI Tools](#)

The Data Documentaion Initiative website's list of tools to implement the [DDI](#) standard.

[FISH Interoperability Toolkit](#)

A suite of tools using the [MIDAS Heritage](#) metadata standard to facilitate the process of moving information between the wide variety of information systems used to record the historic environment.

[Istat SDMX Framework Project](#)

A suite of tools for managing data and metadata in [SDMX](#).

[SDMX Editor](#)

A simple tool for managing and accessing statistical metadata, using the [SDMX](#) framework.

[SDMX Mapping Assistant](#)

A tool to facilitate the mapping between the structural metadata provided by an [SDMX-ML](#) Data Structure Definition and those that reside in a database of a dissemination environment.

[SDMX Tool Repository](#)

A list of software tools supporting the [SDMX](#) standard.

Use Cases

[CESSDA Catalogue](#)

Provides a seamless interface to datasets from social science data archives across Europe using the CESSDA MLI profile of DDI.

[DDI Projects](#)

The Data Documentation Initiative website's list of projects adopting or encouraging DDI as a standard.

[DDI Use Case Literature](#)

Links to DDI 3 use case papers, which were the outcomes of a workshop held at the Schloss Dagstuhl - Leibniz Center for Informatics in Wadern, Germany, November 2-6, 2009.

[English Heritage Listed Buildings System](#)

A case study of the use of the MIDAS XML Monument schema as a vehicle for storing data exported from a major heritage sector information system, the English Heritage Listed Building System (LBS).

[Eurostat](#)

The statistical office of the European Union, which implements SDMX in a number of projects.

[ICPSR - Inter-university Consortium for Political and Social Research](#)

A data archive providing leadership and training in data access, curation, and methods of analysis for the social science research community; all metadata conforms to the DDI standard.

[Open Archives Initiative](#)

Develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content.

[SDMX Implementations](#)

The SDMX website's list of current and planned practical implementations from national and international organisations.

[The Centre for Digital Music Research Data Repository](#)

Used by researchers at C4DM to share their research data with their colleagues and others in the digital music research community, this repository uses the DataCite metadata schema to describe its holdings.

[The Institution for Social and Policy Studies \(ISPS\) Data Archive](#)

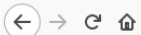
Provides members of the scholarly community with access to files associated with scholarly studies for the purpose of replication, for all studies conducted by ISPS-affiliated researchers. ISPS metadata records conform to DDI requirements and include a minimal set of Dublin Core metadata elements.

[UK ADS - UK Archeology Data Service](#)

The ADS collects, catalogues, manages, preserves, and encourages re-use of digital resources created by archaeologists. It promotes good practice in the use of digital data in archaeology, provides technical advice to the research community, and supports the deployment of digital technologies. Its catalogue records are based on Dublin Core.

[UKDA - UK Data Archive](#)

Curator of the largest collection of digital data in the social sciences and humanities in the United Kingdom, the archive uses DDI as the basis for its catalogue records.











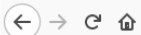
Metadata



<http://rd-alliance.github.io/metadata-directory/subjects>

Arts and Humanities Edit

- [Archaeology](#)  Edit
- [Creative art and design](#)  Edit
- [Heritage Studies](#)  Edit
- [Historical and Philosophical Studies](#)  Edit
- [History by Area](#)  Edit
- [History](#)  Edit
- [Law](#)  Edit
- [Music](#)  Edit



Social and Behavioral Sciences

[Edit](#)

- [Anthropology](#) [Edit](#)
- [Demography](#) [Edit](#)
- [Economics](#) [Edit](#)
- [Geography](#) [Edit](#)
- [Health Policy](#) [Edit](#)
- [Human and Social Geography](#) [Edit](#)
- [Planning \(Urban, Rural and Regional\)](#) [Edit](#)
- [Politics](#) [Edit](#)
- [Sociology](#) [Edit](#)

<http://rd-alliance.github.io/metadata-directory/subjects>

Data sharing examples

The videos will be made available [via the bristol.ac.uk website](#) (both as streaming media and downloads) HD and SD versions will be provided to accommodate those with lower bandwidth. Videos will also be made available [via Vimeo](#), a platform that is already well used by research students at Bristol. [Appropriate metadata will also be provided](#) to the existing Vimeo standard.

All video will also be available [for download and re-editing by third parties](#). To facilitate this [Creative Commons](#) licenses will be assigned to each item. In order to ensure this usage is possible, the [required permissions will be gathered](#) from participants (using a suitable release form) before recording commences.

From [University of Bristol Kitchen Cosmology DMP](#)

We will make the data and associated documentation available to users under a [data-sharing agreement](#) that provides for: (1) a commitment to using the data [only for research purposes](#) and not to identify any individual participant; (2) a commitment to [securing the data](#) using appropriate computer technology; and (3) a commitment to [destroying or returning the data after analyses](#) are completed.

From [NIH data sharing statements](#)

Examples restrictions

Because the STDs being studied are reportable diseases, we will be **collecting identifying information**. Even though the final dataset will be stripped of identifiers prior to release for sharing, we believe that there **remains the possibility of deductive disclosure of subjects** with unusual characteristics. Thus, we will make the data and associated documentation available to users **only under a data-sharing agreement**.

From [NIH data sharing statements](#)

Examples restrictions (2)

1. Share data **privately within 1 year.**

Data will be held in Private Repository, but metadata will be public

2. Release data to **public within 2 years.**

Encouraged after one year to release data for public access.

3. **Request, in writing, data privacy up to 4 years.**

Extensions beyond 3 years will only be granted for compelling cases.

4. Consult with creators of private CZO datasets prior to use.

*It is required to **seek consent before using private data** they can access*

From [Boulder Creek Critical Zone Observatory DMP](#)

Archiving examples

The investigators will **work with staff at the UKDA** to determine **what to archive and how long** the deposited data should be retained. Future long-term use of the data will be ensured by **placing a copy of the data into the repository**.

From [ICPSR Framework for Creating a DMP](#)

Data will be provided in **file formats considered appropriate for long-term access**, as recommended by the UK Data Service. For example, SPSS Portal format and tab-delimited text for qualitative tabular data and RTF and PDF/A for interview transcripts. Appropriate **documentation necessary** to understand the data will also be provided. Anonymised data will be held for **a minimum of 10 years** following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule. Biological samples (output 3) will be **deposited with the UK BioBank** for future use.

From [Writing a Wellcome Trust Data Management and Sharing Plan](#)

Sharing data: what is meant?

With collaborators
while research is active



Data are mutable

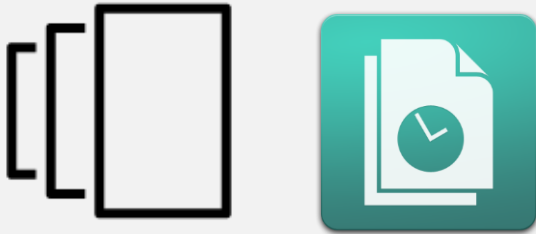
(Open) data sharing



Data are stable,
searchable, citable,
clearly licensed

Storing data: what is meant?

Storing and backing up
files while research is
active

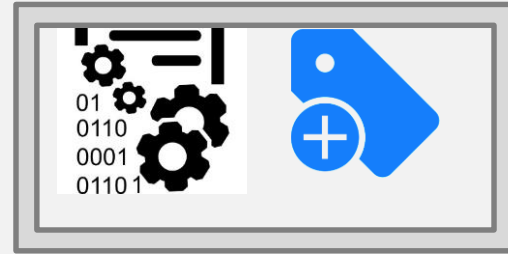


Likely to be on a
networked filestore or
hard drive

Easy to change or
delete



Archiving or
preserving data in
the long-term



Likely to be
deposited in a digital
repository
safeguarded and
preserved

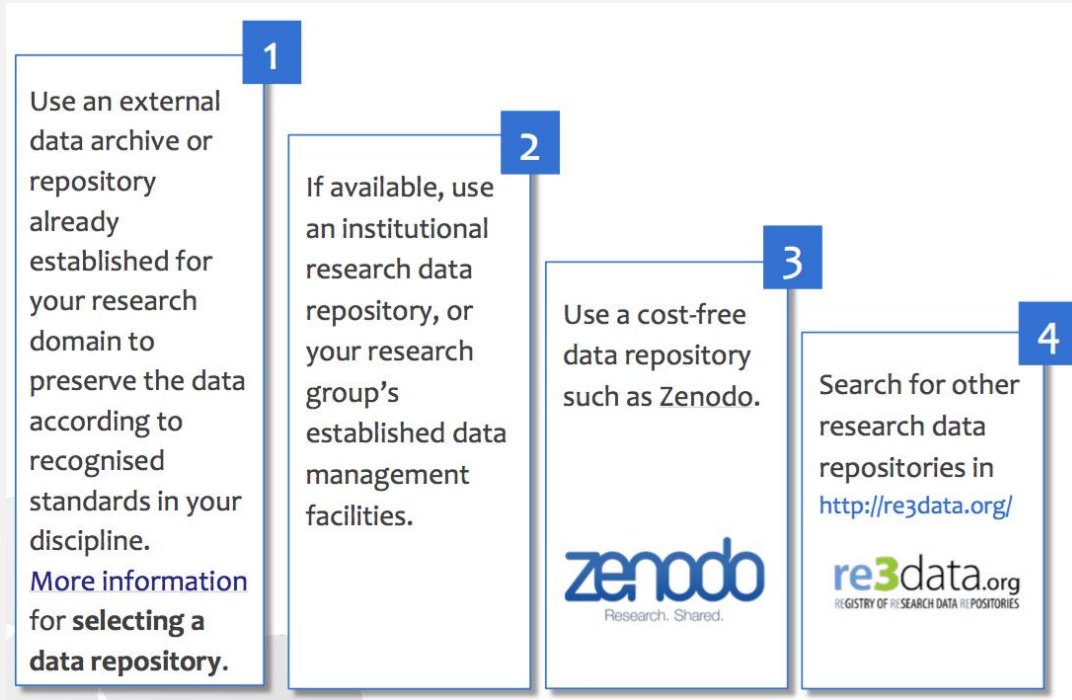
Archiving, repositories, ehm?

Select a data repository that will preserve your data, metadata and possibly tools in the long term.

It is advisable to **contact the repository of your choice when writing the first version of your DMP.**

Repositories may offer guidelines for sustainable data formats and metadata standards, as well as support for dealing with sensitive data and licensing.

Where to find a repository?



More information: <https://www.openaire.eu/opendatapilot-repository>

Zenodo: <http://www.zenodo.org>

Re3data.org: <http://www.re3data.org>

How to select a repository?

Main criteria for choosing a data repository:

Certification as a ‘Trustworthy Digital Repository’, with an explicit ambition to keep the data available in the long term.

- Three common certification standards for TDRs:



Data Seal of Approval: <http://datasealofapproval.org/en>

nestor seal: http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html

ISO 16363: <http://www.iso16363.org>

How to select a repository? (2)

- Matches your particular data needs: e.g. formats accepted; mixture of Open and Restricted Access.
- Provides guidance on how to cite the data that has been deposited.
- Gives your submitted dataset a persistent and globally unique identifier: for sustainable citations - both for data and publications - and to link back to particular researchers and grants. www.openaire.eu/opendatapilot-repository

Zenodo (OpenAIRE/CERN repository)

(All) Research. Shared.
— your one stop research shop!

All research outputs from across all fields of science are welcome! Zenodo accept any file format as well as both positive and negative results. However, we do promote peer-reviewed openly accessible research, and we curate your upload before putting it on the front-

Citeable. Discoverable.
— be found!

Zenodo assigns all publicly available uploads a Digital

Community Collections
— create your own repository

Zenodo allows you to create your own collection and accept or reject all uploads to it. Creating a space for your next workshop or project have never been easier. Plus, everything is citeable and discoverable.



Safe
— more than just a drop box!

Your research output is stored safely for the future in same cloud infrastructure as research data from CERN's Large Hadron Collider using a CERN's battle-tested repository software INVENIO used by some of the world's largest repositories such as INSPIRE HEP and CERN Document Server.

Reporting
— tell your funding agency!

Zenodo is integrated into reporting lines for research funded by the European Commission via OpenAIRE. Just upload your research on Zenodo and we will take care of the reporting for you. We plan to extend with further funding agencies in the future so stay tuned!

Flexible Licensing
— not everything is under Creative Commons

Zenodo encourage you to share your research as openly as possible to maximize use and re-use of your research results. However, we also acknowledge that one size does not fit all, and therefore allow for uploading under a multitude of different licenses and access levels*.

* You are responsible for respecting applicable copyright and license conditions for the files you upload.



www.zenodo.org

Zenodo Repository

Multiple data types

- Publications
- Long tail of research data

Citable data (DOI)

Links to funding, pubs, data, software



H2020: Option to gather, preserve and share project's scientific output

Get started!

Make your first upload - all research outputs from across all fields of research are welcome.

[New Upload](#)[About](#)[About](#)
[Contact](#)
[Policies](#)[Resources](#)[Features](#)
[FAQ](#)[Developers](#)[REST API](#)
[OAI-PMH](#)[Contribute](#)[GitHub](#)
[Donate](#)[Funded by](#)

Delete

Save

Publish

New upload

Instructions: (i) Upload minimum one file or fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.

Files

Choose files

Start upload

Drag and drop files here

— or —

Choose files

(minimum 1 file required, max 100 MB per dataset - contact us for larger datasets)

Upload type

required



zenodo

Search



Upload

Communities

Log in

Sign up

Recent uploads

May 18, 2018 (v2)

Dataset

Open Access

View

Historical climate model output of ECHAM5-wiso from 1871-2011 at T106 resolution

Nathan J. Steiger

Historical climate model simulation of the isotope-enabled ECHAM5-wiso model from the years 1871 to 2011 at T106 (1 degree) resolution. The model code was provided by Martin Werner of AWI. The simulations were designed and run by Nathan Steiger on the Yellowstone supercomputer. The boundary...

Uploaded on May 18, 2018

1 more version(s) exist for this record

May 11, 2018 (v4)

Dataset

Open Access

View

A global network of biomedical relationships derived from text

Percha, Bethany; Altman, Russ B.

This repository contains labeled, weighted networks of chemical-gene, gene-gene, gene-disease, and chemical-disease relationships based on single sentences in PubMed abstracts. All raw dependency paths are provided in addition to the labeled relationships. PART I: Connects dependency paths to...

Uploaded on May 12, 2018

3 more version(s) exist for this record

May 8, 2018 (v1)

Presentation

Open Access

View

Software Citation and the Wikidata Ecosystem

Thornton, Katherine

Slides for a presentation to the Force2017 Software Citation Implementation Working Group describing software and file format data in Wikidata and data validation using ShEx.

Uploaded on May 8, 2018

April 13, 2018 (v1.0.1)

Software

Open Access

View

Zenodo now supports DOI versioning!



Read more about it, in our newest blog post.

Using GitHub?

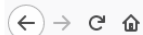


Just Log in with your GitHub account and click here to start preserving your repositories.

Zenodo in a nutshell

- **Research. Shared.** – all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** – uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** – create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- **Funding** – identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- **Flexible licensing** – because not everything is under Creative Commons.
- **Safe** – your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

Read more about Zenodo and its [features](#)



May 18, 2018

Dataset **Open Access**

Historical climate model output of ECHAM5-wiso from 1871-2011 at T106 resolution

Nathan J. Steiger

Historical climate model simulation of the isotope-enabled ECHAM5-wiso model from the years 1871 to 2011 at T106 (1 degree) resolution. The model code was provided by Martin Werner of AWI. The simulations were designed and run by Nathan Steiger on the Yellowstone supercomputer. The boundary conditions were interpolated HadISST fields. The simulations also included updated fractionation factors (an option within the ECHAM5-wiso Fortran code). All variables here are at monthly resolution in netcdf format. Please [contact](#) Nathan Steiger if you have any questions about the simulation. In addition to the data citation, please also cite the following reference for where the data were first published: Steiger, N.J., E.J. Steig, S.G. Dee, G.H. Roe, and G.J. Hakim, (2017): *Climate reconstruction using data assimilation of water isotope ratios from ice cores*. Journal of Geophysical Research: Atmospheres, doi:10.1002/2016JD026011.

Standard variables include: ECHAM5 T106 orography, 2 m temperature, surface pressure, mean sea level pressure, vertically integrated water vapor, total precipitation, evaporation, soil moisture, relative humidity, specific humidity, atmospheric stream function at 200 hPa, geopotential height at 500 hPa, windspeed at 10 m, and u-velocity wind at 200 hPa. Moisture variables are given at the surface (lowest atmospheric level).

Isotope variables include: d18O and dD of total precipitation, d18O and dD of evaporation, d18O and dD of snow fall, d18O and dD of seasonal snow cover, d18O and dD of snow on glaciers, d18O and dD of soil moisture, and specific humidity of water isotopes.

Version 2 includes additional model output.

Files (9.0 GB)

Name	Size	
d18O_glaciersnow_mon_1871_2011.nc	346.5 MB	Download
md5:dbdb2e897cddb2fe05128ed58359db7		
d18O_precip_mon_1871_2011.nc	346.5 MB	Download
md5:759f8bfbf0a6001589a76c6196dd73fb		

Indexed in

OpenAIRE

Publication date:

May 18, 2018

DOI:DOI [10.5281/zenodo.1249604](https://doi.org/10.5281/zenodo.1249604)**Keyword(s):**

climate model ECHAM5-wiso water isotopes
 historical climate simulation paleoclimate
 climate reconstruction proxy system model

Communities:

Zenodo

License (for files):[Creative Commons Attribution 4.0](#)**Versions**Version 2 [10.5281/zenodo.1249604](https://zenodo.org/record/1249604) May 18, 2018Version 1 [10.5281/zenodo.1041751](https://zenodo.org/record/1041751) Nov 3, 2017

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.1041750](https://doi.org/10.5281/zenodo.1041750). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

md5:759f8bfbf0a6001589a76c6196dd73fb ⓘ

d180_snowcover_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:af7cf212d2a6f47aa985511d82c1fe08 ⓘ

d180_snowfall_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:cdfb16a647988603a76ce3754244ee2d ⓘ

d180_soilmoisture_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:63d0dea69592feefdc53f0c7dae7de60 ⓘ

dD_glaciersnow_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:6d7258a1b820ce8422f524bf1ef753dd ⓘ

dD_precip_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:341c0b3c8425126628cd217b9cc16df1 ⓘ

dD_snowcover_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:dffac99ee9df4483c41a69aa48b24093 ⓘ

dD_snowfall_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:68a8e3e48e5efa8cc38454bc15ea105d ⓘ

dD_soilmoisture_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:e005504641e9e945a686fbfb0afd963a ⓘ

evap_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:455111aeb272e69a62dbbf7d6980f05e ⓘ

orography_ECHAM5_T106.nc

414.9 kB

[Download](#)

md5:300c2eda2898b30a808b48957fbd072f ⓘ

precip_mon_1871_2011.nc

346.5 MB

[Download](#)

md5:a6b7c7a347ad061ddf31888cf0a132b8 ⓘ

os_mon_1871_2011.nc

346.5 MB

[Download](#)

Share



Cite as

Nathan J. Steiger. (2018). Historical climate model output of ECHAM5-wiso from 1871-2011 at T106 resolution [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1249604>

Export

[BibTeX](#) [CSL](#) [DataCite](#) [Dublin Core](#) [JSON](#)
[JSON-LD](#) [MARCXML](#) [Mendeley](#)

Expert Tour Guide on Data Management

1. Plan

2. Organise & Document

3. Process

4. Store

5. Protect

Ethics and data
protection

Ethical review
process

Processing
personal data

Diversity in data
protection

Informed consent

Anonymisation

Copyright

Diversity in

Adapt your DMP:

In the case studies in the tabs, you can identify the potential copyright issues and state how you would address these in practice.

ARCHIVED DATA

DATA IN THE PUBLIC DOMAIN

SURVEY QUESTIONS

INTERVIEWS

Case Study 1 – Copyright of Archived Data

A researcher uses [International Social Survey Programme](#) (ISSP, n.d.) data obtained from ZACAT/GESIS - Leibniz Institute for the Social Sciences in Germany. These data are freely available to registered users. The researcher incorporates some of the ISSP data within a database containing his own research data. Can this database be deposited with another archive?

+ Click for the answer

2. Organise & Document

3. Process

4. Store

5. Protect

Ethics and data protection

Ethical review process

Processing personal data

Diversity in data protection

Informed consent

Anonymisation

Copyright

Diversity in

Adapt your DMP:

Click for the answer

Although the ISSP data are available for free to all researchers, this does not mean that the data can be published in another archive and made available to others. The data can be incorporated into a database and used for personal analysis. But, before this dataset can be deposited with another archive, permission must be sought from the owner of the original data.

[« Previous](#)

[Next »](#)

<https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/5.-Protect/Copyright>

LICENCE



LINKS

[TRAINING AND OTHER EVENTS](#)

[CESSDA TRAINING BLOG](#)

CONTACT

training@cessda.net

The Expert Tour Guide on Data

Expert Tour Guide on Data Management

1. Plan

2. Organise & Document

3. Process

4. Store

5. Protect

Ethics and data
protection

Ethical review
process

Processing
personal data

Diversity in data
protection

Informed consent

Anonymisation

Copyright

Diversity in
copyright

yright

Case studies

In the case studies in the tabs, you can identify the potential copyright issues and state how you would address these in practice.

[ARCHIVED DATA](#)

[DATA IN THE PUBLIC DOMAIN](#)

[SURVEY QUESTIONS](#)

[INTERVIEWS](#)

Case Study 2 – Copyright of Data in the Public Domain

A researcher studies how health issues around obesity are reported in the media in the last 10 years. Freely available newspaper websites and library sources are used to obtain articles on this topic. Articles or excerpts are copied into a database and coded according to various criteria for content analysis. (i) Can the researcher use such public data without breaching copyright? (ii) Can the database be archived and shared with other researchers?

+ Click for the answer

Document

3. Process

4. Store

5. Protect

Ethics and data protection

Ethical review process

Processing personal data

Diversity in data protection

Informed consent

Anonymisation

Copyright

Diversity in copyright

Adapt your DMP:

Click for the answer

Even though the articles are publicly available, they are still under copyright. Whilst such information can be used for personal research purposes (e.g. in the UK this would be under the broad exemption of 'fair dealing'), the articles cannot be archived, unless permission is obtained from the newspapers; otherwise this would breach copyright.

[« Previous](#)

[Next »](#)

<https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/5.-Protect/Copyright>

LICENCE



The Expert Tour Guide on Data Management by CESSDA ERIC is

LINKS

[TRAINING AND OTHER EVENTS](#)

[CESSDA TRAINING BLOG](#)

CONTACT

training@cessda.net

The Expert Tour Guide on Data Management is created for CESSDA ERIC by ADP, AUSSDA, CSDA, DANS.

Expert Tour Guide on Data Management

1. Plan

2. Organise & Document

3. Process

4. Store

5. Protect

Ethics and data
protection

Ethical review
process

Processing
personal data

Diversity in data
protection

Informed consent

Anonymisation

Copyright

Diversity in
copyright

Adapt your DMP:

In the case studies in the tabs, you can identify the potential copyright issues and state how you would address these in practice.

ARCHIVED DATA

DATA IN THE PUBLIC DOMAIN

SURVEY QUESTIONS

INTERVIEWS

Case Study 3 – Copyright of Survey Questions

A researcher wishes to reuse a set of questions from an existing survey questionnaire, to compare results between the newly proposed survey and the original.

+ Click for the answer

« Previous

Next »

Expert Tour Guide on Data Management

1. Plan

2. Organise & Document

3. Process

4. Store

5. Protect

Ethics and data
protection

Ethical review
process

Processing
personal data

Diversity in data
protection

Informed consent

Anonymisation

Copyright

Diversity in
copyright

Adapt your DMP:

Click for the answer

The survey questions and instruments will be copyright protected, with copyright residing with the organisation who commissioned, designed or conducted the survey (unless the original creator/owner transfers all ownership rights). The researcher needs to contact the copyright holder directly for permission to reproduce the questionnaire text for any new use. Some questionnaires will contain measurement scales, batteries of questions or classifications. These instruments are again copyrighted. Therefore, to reproduce them the researcher will need permission.

[« Previous](#)

[Next »](#)

<https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/5.-Protect/Copyright>

Expert Tour Guide on Data Management

1. Plan

2. Organise & Document

3. Process

4. Store

5. Protect

Ethics and data
protection

Ethical review
process

Processing
personal data

Diversity in data
protection

Informed consent

Anonymisation

Copyright

Diversity in
copyright

Adapt your DMP:

ARCHIVED DATA

DATA IN THE PUBLIC DOMAIN

SURVEY QUESTIONS

INTERVIEWS

Case Study 4 – Copyright of Interviews with Stay-at-Home Parents

A researcher interviews various stay-at-home parents about their careers and produces audio recordings and near verbatim transcripts herself. The researcher analyses this material and offers it to a data archive. The researcher did not get signed copyright transfers for the interviewees' words. What are the rights issues surrounding this offer of data?

+ Click for the answer

« Previous

Next »

<https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/5.-Protect/Copyright>

Expert Tour Guide on Data Management

1. Plan

2. Organise & Document

3. Process

4. Store

5. Protect

Ethics and data
protection

Ethical review
process

Processing
personal data

Diversity in data
protection

Informed consent

Anonymisation

Copyright

Diversity in
copyright

Adapt your DMP:

Click for the answer

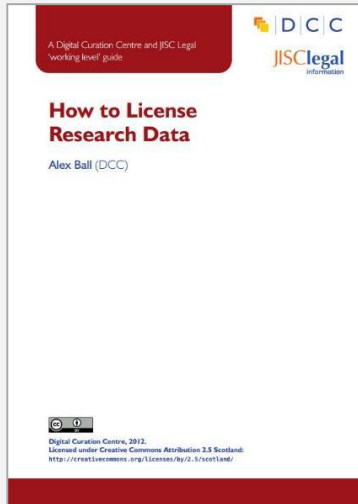
In this case, the stay-at-home parents hold copyright in their own recorded words, whilst the researcher holds copyright over the transcribed interviews. Quoting large extracts of the data, either in publications or by archiving the transcripts, would breach the copyright of the interviewees in their recorded words. If the researcher wants to publish large extracts of data, or archive the transcripts, they need to request permission to do so from the interviewees or request that the interviewee transfers the copyright of the interview content to the researcher, which could be achieved through the use of a Recording Agreement.

« Previous

Next »

<https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/5.-Protect/Copyright>

Licensing research data



Horizon 2020 Open Access guidelines point to:



This DCC guide outlines the pros and cons of each approach and gives practical advice on how to implement your licence

CREATIVE COMMONS LIMITATIONS



NC Non-Commercial

commercial?

What counts as



ND No Derivatives

Severely restricts use

These clauses are not open licenses

www.dcc.ac.uk/resources/how-guides/license-research-data

EUDAT licensing tool

Answer questions to determine which licence(s) are appropriate to use

Do you own copyright and similar rights in your dataset and all its constitutive parts?

Do you allow others to make commercial use of you data?

Creative Commons Attribution (CC-BY)
This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

Public Domain Dedication (CC Zero)
CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

<http://ufal.github.io/public-license-selector>

Research Ethics and Legal Compliance: Informed Consent and Data Licensing | FOSTER - Mozilla Firefox

https://www.fosteropenscience.eu/content/research-ethics-and-legal-compliance-informed-consent-an 90% Search

Research Ethics and Legal Compliance: Informed Consent and Data Licensing

CESSDA Training at the Data Archive for the Social Sciences GESIS - Leibniz Institute for the Social Sciences @CESSDA_Data

Preview



Download PDF
Download EPUB
Authors: Sebastian Netscher
Publication year: 2015
Language: English (EN)
Level of knowledge: Advanced: apply
Usage rights:



Topics

- Approach
- Legal Issues
- Research Data Management
- Ethics

Audience

- PHD Students
- Researchers and Students

<https://www.fosteropenscience.eu/content/research-ethics-and-legal-compliance-informed-consent-and-data-licensing>



Examples of DMP questions and answers

Expert Tour Guide on Data Management

1. Plan

Benefits of data
management

Research data

Data in social
sciences

FAIR data

European diversity

**Adapt your DMP:
Part 1**

Sources and
further reading

2. Organise & Document

3. Process

4. Store

5. Protect

6. Archive & Publish

For inspiration of filled in DMPs look at some example DMPs we prepared. Both DMPs are based on a fictional research project with a basis in reality. For each topic of the DMP, there are example questions and answers where applicable. The examples are not country specific. Some of the information is generic.



Qualitative data

During this project, in-depth interviews with teachers in primary school will be held. The project has just started.

Click the link to view and download the DMP:

[DMPQuestionsQualitativeData.pdf](#) (165 KB)



Quantitative data

The project concerns a survey which is conducted in order to identify how the evolution of society affects attitudes and behaviour. The project is still running.

Click the link to view and download the DMP:

[DMPQuestionsQuantitativeData.pdf](#) (205 KB)

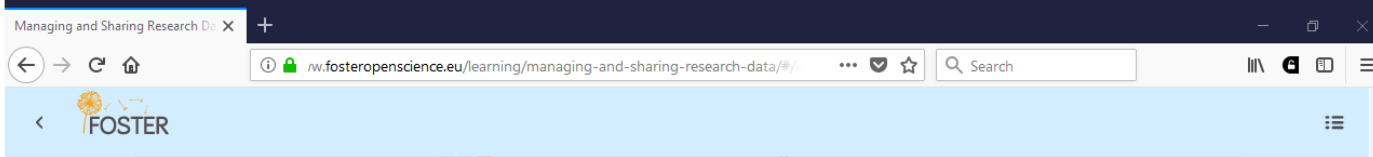
Discipline specific examples of real DMPs

Here are a few examples of data management plans (DMPs) that have been produced by projects in the Arts and Humanities, Social Sciences, and Life Sciences. For more examples of DMPs, please see the [DCC website](#).



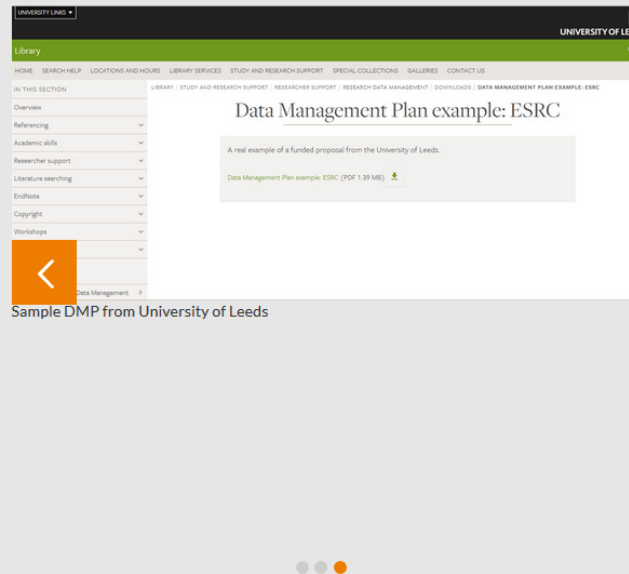
A DMP from the Arts and Humanities

The H2020 supported [FREME](#) project is an open framework of e-services for multilingual and semantic enrichment of digital content project. The project has shared version one of their [DMP](#) as a public deliverable.



Discipline specific examples of real DMPs

Here are a few examples of data management plans (DMPs) that have been produced by projects in the Arts and Humanities, Social Sciences, and Life Sciences. For more examples of DMPs, please see the [DCC website](#).



A DMP from the Social Sciences

This [DMP](#) was developed and shared by PI Andrea Holomotz, University of Leeds for her Economic and Social Research Council (ESRC) funded project 'Realist Evaluation of Adapted Sex Offender Treatment Programs for Men with Intellectual Disability'. The DMP covers the management of quantitative and qualitative data, including audio interviews. The plan discusses ethical issues around a very sensitive dataset, including using consent and anonymisation to generate data which can be shared and reused. It was shared by the University of Leeds.

Guidelines on DMPs

How to develop a DMP www.dcc.ac.uk/resources/how-guides/develop-data-plan

RDM brochure and template

https://dans.knaw.nl/en/about/organisation-and-policy/information-material?set_language=en

OpenAIRE guidelines www.openaire.eu/opendatapilot-dmp

ICPSR framework for a DMP

www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html

Other resources

Where to keep research data <http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data/where-keep-research-data>

Five steps to decide what data to keep

<http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>

Re3data <http://www.re3data.org/>

Figshare <https://figshare.com/>

Genbank <https://www.ncbi.nlm.nih.gov/genbank/>

How to write a lay summary <http://www.dcc.ac.uk/resources/how-guides/write-lay-summary>

Lay summaries <https://www.bhf.org.uk/research/information-for-researchers/how-to-apply/lay-summaries>

With thanks to

Joy Davidson, University of Glasgow

Bianca Kramer & Jeroen Bosman, Utrecht

University Library

Marjan Grootveld, DANS

Sarah Jones, DCC

Acknowledgements:

Jonathan Rans, DCC

Thanks to DANS and DCC for reuse of slide



Thank you! Questions?

Facebook: @fosteropenscience

Twitter: @fosterscience

Youtube: [FOSTER Open Science](#)