

# Data Management and Data Management Plans

Dr. Tomasz Miksa

TU Wien & SBA Research

[tmiksa@sba-research.at](mailto:tmiksa@sba-research.at)

These slides are available online: [10.5281/zenodo.1068223](https://zenodo.org/doi/10.5281/zenodo.1068223)

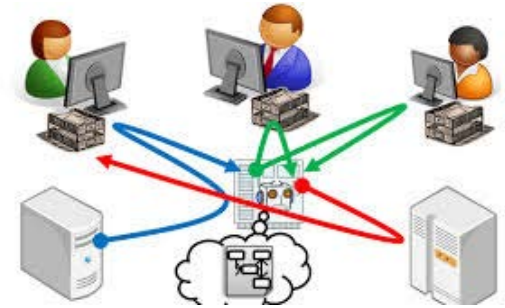
# Agenda

- Why do we need Data Management Plans (DMPs)?
- What is a DMP?
- How to create a DMP?
- Tips for writing DMPs?
- What is the future?

# **WHY DO WE NEED DATA MANAGEMENT PLANS?**

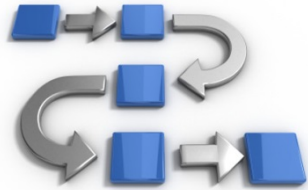
# e-Science and Research Infrastructures

- Scientists exchange
  - data
  - services
  - computational power
- Collaborate to solve challenges
  - DNA sequencing
    - climate change
    - tsunami forecasting
  - Earth Observation
    - climate change
    - tsunami forecasting
  - Large Hadron Collider



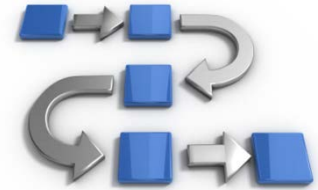


# Reproducibility



Original experiment

- Is the result correct?
- Do the results match?
- Can I reuse her workflow?



Re-executed experiment



# Reproducibility

- Reproducibility requires
  - well documented research workflows
  - precise information on the experiment's environment [1] [2]
  
- Studies show very low reproducibility in
  - medicine
  - economy
  - computer science

[1] <https://doi.org/10.1016/j.jbi.2016.10.011>

[2] <https://doi.org/10.1371/journal.pone.0038234>

# Reproducibility

## Neuroanatomical studies

### ■ FreeSurfer Software

- cortical thickness and volume of neuroanatomical structures

### ■ Different

- FreeSurfer Versions

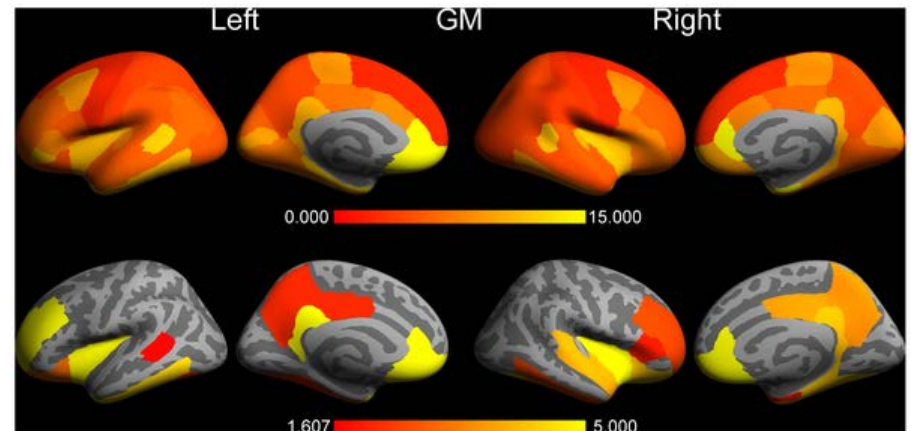
- v4.3.1, v4.5.0, v5.0.0

- Workstation

- Mac, Hewlett-Packard

- Operating system version

- OSX 10.5, OSX 10.6

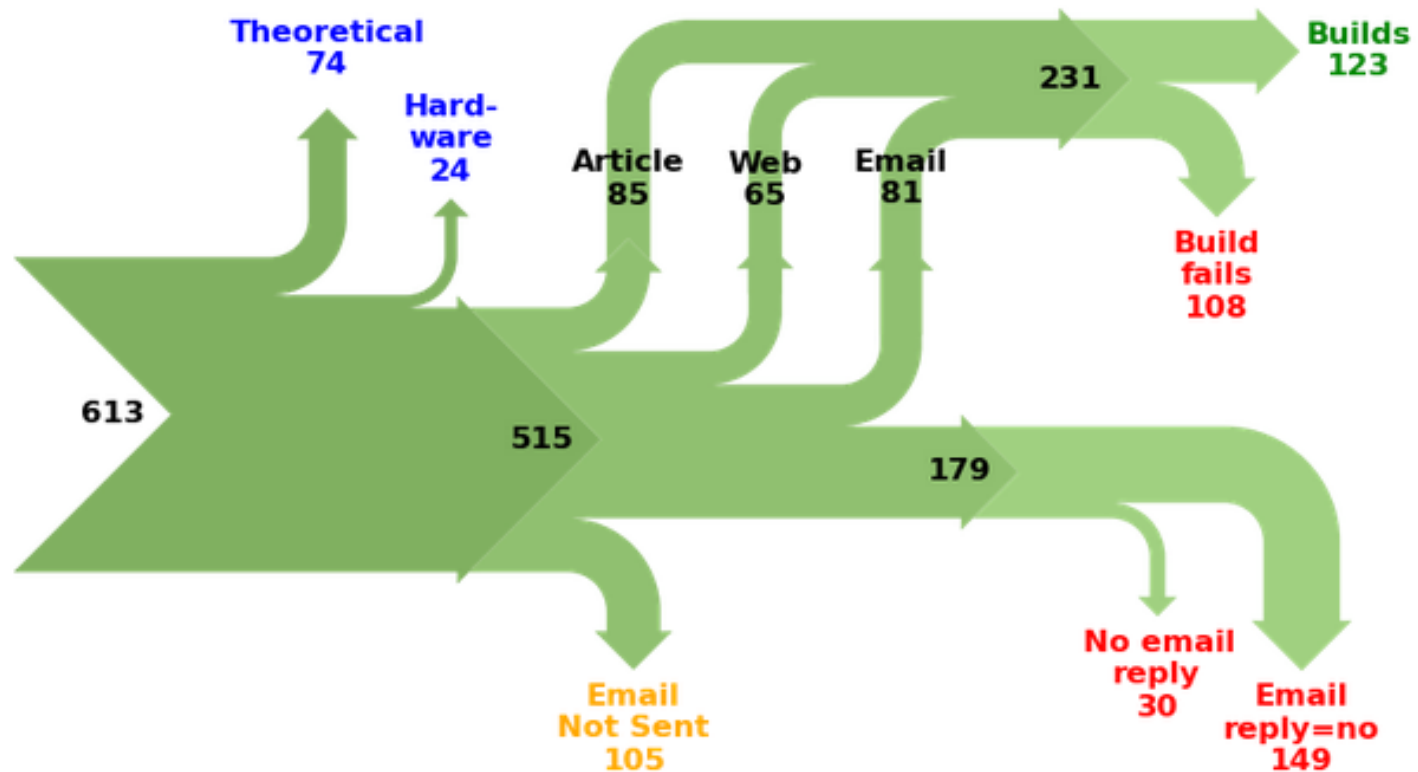


E. Gronenschild, P. Habets, H. I. L. Jacobs, R. Mengelers, N. Rozendaal, J. van Os, and M. Marcelis, "The effects of freesurfer version, workstation type, and macintosh operating system version on anatomical volume and cortical thickness measurements," 2012.



# Reproducibility Computer Science

- 613 papers in 8 ACM conferences



C. Collberg and T. Proebsting, "Measuring reproducibility in computer systems research," 2014. [Online]. Available: <http://reproducibility.cs.arizona.edu/tr.pdf>

# Reproducibility

## Computer Science

- E-mail replies from authors
  - Wrong version
  - Code will be available soon
  - Programmer left
  - Bad backup practices
  - Commercial code
  - Proprietary academic code
  - Intellectual property
  - No intention to release
  - ...

# Variety of solutions

- To improve reproducibility and data management many solutions were proposed
  - **open access** to scientific publications and data
  - research **data repositories** to host the data
  - **data citation** to reference the datasets
  - **DATA MANAGEMENT PLANS**



# WHAT IS A DATA MANAGEMENT PLAN?

# Data Management Plan

- DMP is a formal document
- It outlines what you will do with your data **during** and **after** you complete your research
- It ensures your data is safe for the **present** and the **future**

[ from University of Virginia Library]

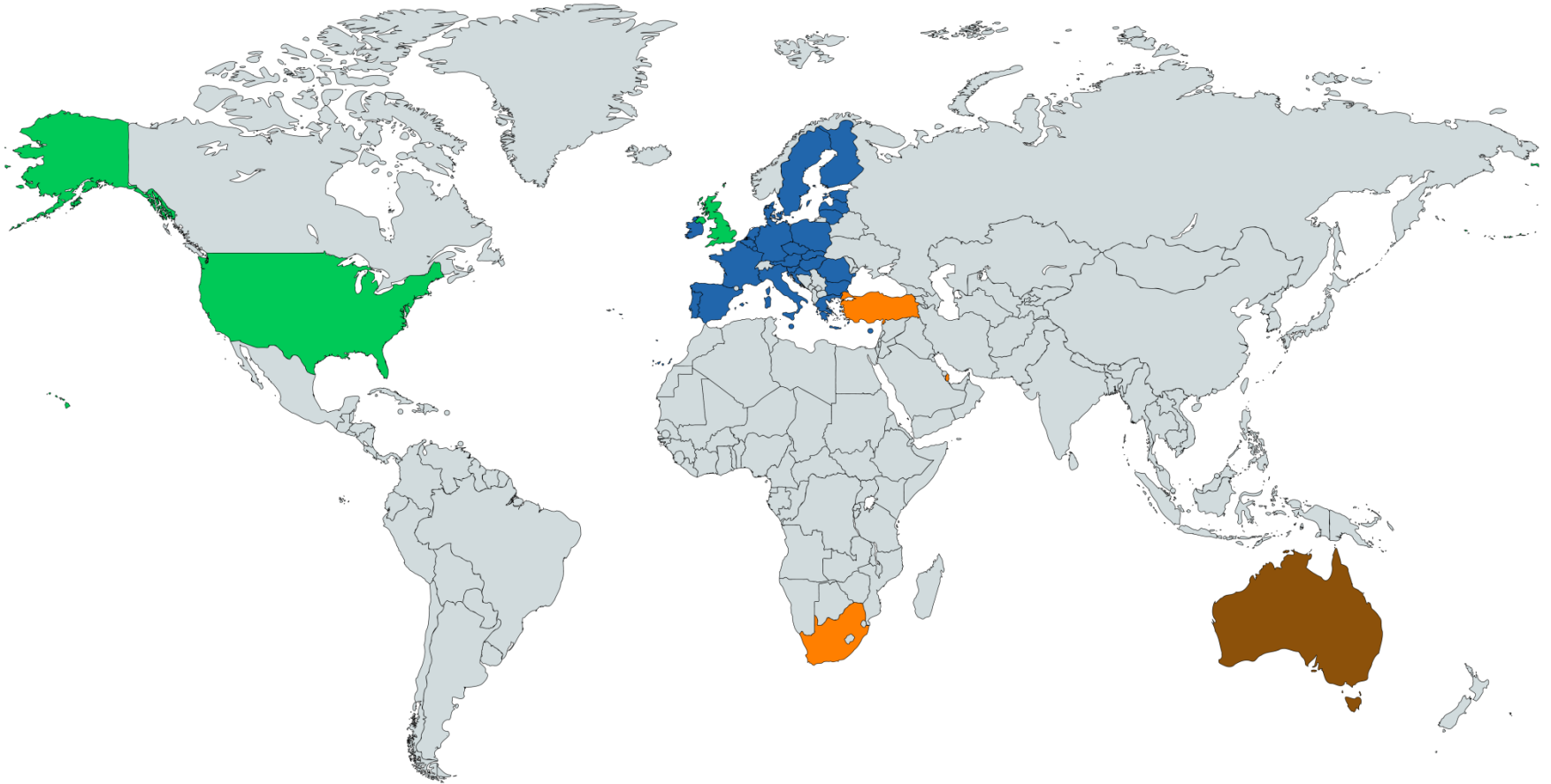


# DMP is an awareness tool!

- DMP makes you think
  - what data you will use and where you get it from
  - what infrastructure, software, licenses are needed
  - what will be the output of your research
  - how you will share your research outputs
- DMP helps you organise yourself better
- DMP can be useful for ethics committee
- DMP can reveal how solid your research methodology is
  - is it a 'fishing expedition'?



# DMPs worldwide



Created with mapchart.net ©



# EC Horizon 2020

## Open Research Data Pilot

- Open access to publications is default
- Open access to research data is default from 2017
  - **NOT** all data must be released
  - data needed to validate scientific publications
  - other data on a voluntary basis
- Opt-out possibility
  - No impact on the proposal evaluation
  - At every stage
- Data management costs can be claimed





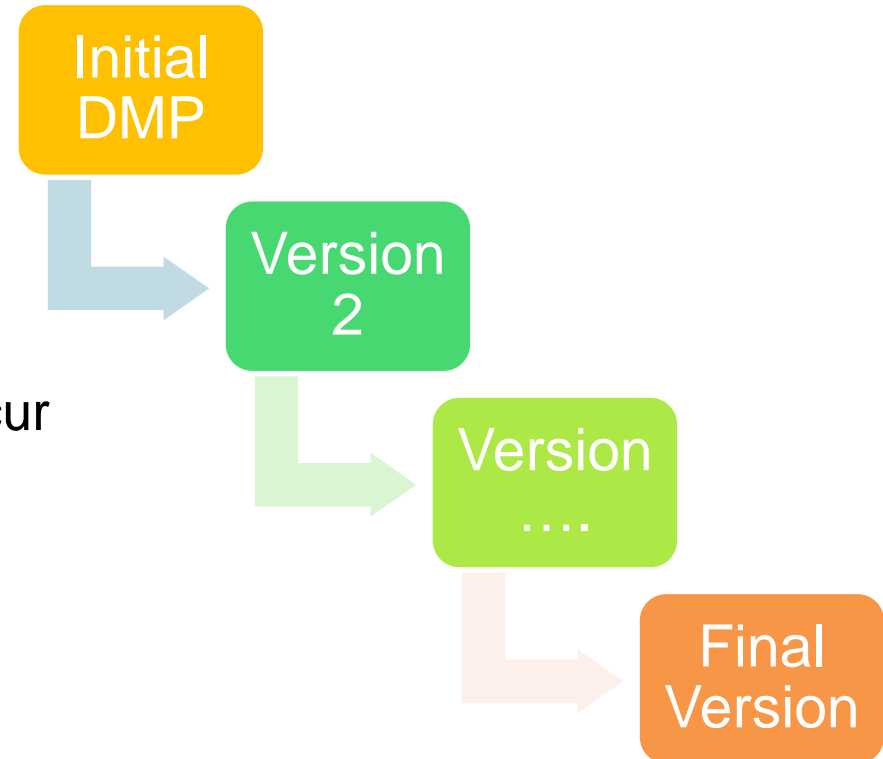
# EC Horizon 2020 DMP Template

- Template is recommended but not required
  - 6 sections
  - 31 questions
  - Follows FAIR principles
    - Data must be **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable

DMP component	Issues to be addressed
1. Data summary	<ul style="list-style-type: none"> <li>• State the purpose of the data collection/generation</li> <li>• Explain the relation to the objectives of the project</li> <li>• Specify the types and formats of data generated/collected</li> <li>• Specify if existing data is being re-used (if any)</li> <li>• Specify the origin of the data</li> <li>• State the expected size of the data (if known)</li> <li>• Outline the data utility: to whom will it be useful</li> </ul>
2. FAIR Data	

# EC Horizon 2020 DMP versions

- DMP is a living document
- First version
  - within the first 6 months
- Updated versions
  - when significant changes occur
    - new datasets
    - changes in policies
  - periodic reporting
    - project reviews
  - end of project



# HOW TO CREATE A DMP?



# How to create a DMP?

- By answering questions from a checklist
- Software tools for filling out DMPs
  - users choose appropriate funders template
  - relevant questions and guidance is presented
  - facilitate co-working
  - results can be exported to PDF
- Usually mixed approach works best
  - check guidance in the tool and create your own document



# DCC Checklist

- Synthesis of
  - funder requirements
  - institutional guidelines
  - good practice
- Contains 8 sections
  - the extent to which they need to be covered depends on a kind of research
- Resulting DMP
  - between a few paragraphs to a few pages long

<b>Checklist for a Data Management Plan, v4.0</b>	
<p><b>Please cite as:</b> DCC. (2013). <i>Checklist for a Data Management Plan</i>. v.4.0. Edinburgh: Digital Curation Centre. Available online: <a href="http://www.dcc.ac.uk/resources/data-management-plans">http://www.dcc.ac.uk/resources/data-management-plans</a></p>	
DCC Checklist	DCC Guidance and questions to consider
<b>Administrative Data</b>	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	<p><b>Questions to consider:</b></p> <ul style="list-style-type: none"> <li>- What is the nature of your research project?</li> <li>- What research questions are you addressing?</li> <li>- For what purpose are the data being collected or created?</li> </ul> <p><b>Guidance:</b> Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.</p>
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID <a href="http://orcid.org/">http://orcid.org/</a>
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	<p><b>Questions to consider:</b></p> <ul style="list-style-type: none"> <li>- Are there any existing procedures that you will base your approach on?</li> <li>- Does your department/group have data management guidelines?</li> <li>- Does your institution have a data protection or security policy that you will follow?</li> <li>- Does your institution have a Research Data Management (RDM) policy?</li> <li>- Does your funder have a Research Data Management policy?</li> <li>- Are there any formal standards that you will adopt?</li> </ul> <p><b>Guidance:</b> List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.</p>
<b>Data Collection</b>	
What data will you collect or create?	<p><b>Questions to consider:</b></p> <ul style="list-style-type: none"> <li>- What type, format and volume of data?</li> <li>- Do your chosen formats and software enable sharing and long-term access to the data?</li> <li>- Are there any existing data that you can reuse?</li> </ul> <p><b>Guidance:</b> Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.</p>
How will the data be collected or created?	<p><b>Questions to Consider:</b></p> <ul style="list-style-type: none"> <li>- What standards or methodologies will you use?</li> <li>- How will you structure and name your folders and files?</li> <li>- How will you handle versioning?</li> <li>- What quality assurance processes will you adopt?</li> </ul> <p><b>Guidance:</b> Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning</p>

# DMP in Horizon 2020

- DMP is a project deliverable
- DMP is a written document
- DMP Template
  - recommended but not required
  - contains auxiliary questions
  - [Template for the Data Management Plan](#)

## 1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

What types and formats of data will the project generate/collect?

Will you re-use any existing data and how?

What is the origin of the data?

What is the expected size of the data?

To whom might it be useful ('data utility')?

## 2. FAIR data

### 2.1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

### 2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

What methods or software tools are needed to access the data?

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)?

How will the identity of the person accessing the data be ascertained?

### 2.3. Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

# PLOS – Ten Simple Rules

PERSPECTIVE

## Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener\*

College of University Libraries &amp; Learning Sciences, University of New Mexico, Albuquerque, New Mexico, United States of America

\* [william.michener@gmail.com](mailto:william.michener@gmail.com)


### Introduction

Research papers and data products are key outcomes of the science enterprise. Governmental, nongovernmental, and private foundation sponsors of research are increasingly recognizing the value of research data. As a result, most funders now require that sufficiently detailed data management plans be submitted as part of a research proposal. A data management plan (DMP) is a document that describes how you will treat your data during a project and what happens with the data after the project ends. Such plans typically cover all or portions of the data life cycle—from data discovery, collection, and organization (e.g., spreadsheets, databases), through quality assurance/quality control, documentation (e.g., data types, laboratory methods) and use of the data, to data preservation and sharing with others (e.g., data policies and dissemination approaches). [Fig 1](#) illustrates the relationship between hypothetical research and data life cycles and highlights the links to the rules presented in this paper. The DMP undergoes peer review and is used in part to evaluate a project's merit. Plans also document the data management activities associated with funded projects and may be revisited during performance reviews.

Earlier articles in the Ten Simple Rules series of *PLOS Computational Biology* provided guidance on getting grants [1], writing research papers [2], presenting research findings [3], and caring for scientific data [4]. Here, I present ten simple rules that can help guide the process of creating an effective plan for managing research data—the basis for the project's findings, research papers, and data products. I focus on the principles and practices that will result in a DMP that can be easily understood by others and put to use by your research team. Moreover, following the ten simple rules will help ensure that your data are safe and sharable and that your project maximizes the funder's return on investment.

### Rule 1: Determine the Research Sponsor Requirements

Research communities typically develop their own standard methods and approaches for managing and disseminating data. Likewise, research sponsors often have very specific DMP expectations. For instance, the Wellcome Trust, the Gordon and Betty Moore Foundation (GBMF), the United States National Institutes of Health (NIH), and the US National Science Foundation (NSF) all fund computational biology research but differ markedly in their DMP requirements. The GBMF, for instance, requires that potential grantees develop a comprehensive DMP in

CrossMark  
click for updates

 OPEN ACCESS

**Citation:** Michener WK (2015) Ten Simple Rules for Creating a Good Data Management Plan. *PLoS Comput Biol* 11(10): e1004525. doi:10.1371/journal.pcbi.1004525

**Editor:** Philip E. Bourne, National Institutes of Health, UNITED STATES

**Published:** October 22, 2015

**Copyright:** © 2015 William K. Michener. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by NSF IIA-1301346, IIA-1329470, and ACI-1430508 (<http://nsf.gov>). The funders had no role in study design, data collection and analysis, decision to publish, or

<http://dx.doi.org/10.1371/journal.pcbi.1004525>

dmponline.dcc.ac.uk/plans/new

# DMP ONLINE

[View plans](#) [Create plan](#) [About](#) [Future plans](#) [Help](#) [Change language](#)

## Create a new plan

Before you get started, we need to ask a few questions to set you up with the best DMP template for your needs.

### What research project are you planning?

Project title

If applying for funding, state the title exactly as in the proposal.

### Primary research organisation

Select the primary research organisation responsible

My research organisation is not on the list or no research organisation is associated with this plan

### Funding organisation

Select the funding organisation

No funder associated with this plan

[Create Plan](#) <https://dmponline.dcc.ac.uk>





- [View plans](#)
- [Create plan](#)
- [About](#)
- [Future plans](#)
- [Help](#)
- [Change language](#)

## FFG Webinar Horizon 2020 Example

0/71 questions answered

- [Plan details](#)
- [Initial DMP](#)
- [Detailed DMP](#)
- [Final review DMP](#)
- [Share](#)
- [Export](#)

- 1. **Data summary** (1 question, 0 answered) +
- 2. **FAIR data** (4 questions, 0 answered) +
- 3. **Allocation of resources** (1 question, 0 answered) +
- 4. **Data security** (1 question, 0 answered) +
- 5. **Ethical aspects** (1 question, 0 answered) +
- 6. **Other** (1 question, 0 answered) +

Export



View plans Create plan About Future plans Help Change language ▾

## FFG Webinar Horizon 2020 Example

0/71 questions answered

Plan details **Initial DMP** Detailed DMP Final review DMP Share Export

### 1. Data summary (1 question, 0 answered) +

### 2. FAIR data (4 questions, 0 answered) -

In general terms, your research data should be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

#### 2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

**B** *I*

Save

Guidance Share note

#### EC Guidance -

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

## FFG Webinar Horizon 2020 Example

Plan details Initial DMP Detailed DMP Final review DMP Share **Export**

From here you can download your plan in various formats. This may be useful if you need to submit your plan as part of a grant application. Select what format you wish to use and click to 'Export'.

### Initial DMP

Format

pdf ▾

### Export Settings (Using default PDF formatting values)

### File Name

File Name

### Included Elements

#### Details

- Plan Name
- Plan ID
- Grant Number
- Principal Investigator / Researcher
- Plan Data Contact
- Description
- Funder
- Institution

#### Sections

- 1. Data summary  
Provide a summary of the data addressing the fol
- 2. FAIR data
  - 2.1 Making data findable, including provisions for
  - 2.2 Making data openly accessible: Specify which
  - 2.3 Making data interoperable: Assess the interop

### FFG Webinar Horizon 2020 Example

**Plan Name** Horizon 2020 DMP - FFG Webinar Horizon 2020 Example  
**Plan ID** -  
**Grant Number** -  
**Principal Investigator / Researcher** Tomasz Miksa  
**Plan Data Contact** miksa@ifs.tuwien.ac.at  
**Plan Description** -  
**Funder** European Commission (Horizon 2020)  
**Institution** Other  
**Your ORCID** -

#### 1. Data summary

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

Question not answered.

#### 2. FAIR data

##### 2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Question not answered.

##### 2.2 Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

Question not answered.

# WHAT SHOULD I WRITE?

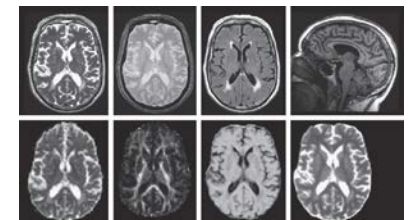
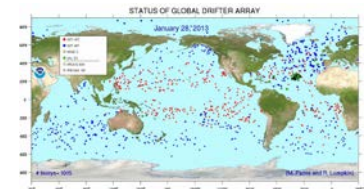
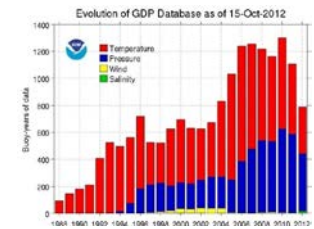
# Most templates require

1. Data Summary
2. FAIR data
3. Allocation of resources
4. Data Security
5. Ethical aspects
6. Other issues



# What is data?

- Instrument measurements
- Experimental observations
- Still images, video and audio
- Text documents, spreadsheets, databases
- Quantitative data (e.g. survey data)
- Survey results & interview transcripts
- Simulation data, models & software
- Slides, artefacts, specimens, samples
- Questionnaires
- Sketches, diaries, lab notebooks ...



# Data Summary

- Type
  - text, spreadsheets, software, models, images, movies, audio, patient records, etc.
- Source
  - human observation, laboratory, field instruments, experiments, simulations, compilations, etc.
- Volume
  - total volume of data, number of files, etc.
- Data and file formats
  - non-proprietary formats
  - used within community

## Data Summary - example

Every two days, we will subsample *E. affinis* populations growing under our treatment conditions. We will use a microscope to identify the life stage and sex of the subsampled individuals. We will **document the information first in a laboratory notebook and then copy the data into an Excel spreadsheet**. The Excel spreadsheet will be saved as a comma separated value **(.csv) file**.

From DataOne – *E. affinis* DMP example



# FAIR Principles

- **Findable**
  - the data is available online, likely in a repository
  - contains metadata that facilitates search
- **Accessible**
  - access conditions are specified (e.g. no login)
  - software needed to interpret data is known
- **Interoperable**
  - Follow standards and domain specific conventions
- **Reusable**
  - clear license and documentation
  - 'sum of the three other rules'
- There is no clear distinction between principles
  - e.g. metadata supports all of them

# Standards and Metadata



# Metadata – Atlas Of Living Australia

## NatureShare - 2380\_Gymnorhina\_tibicen

HumanObservation of *Cracticus tibicen* | Australian Magpie recorded on 2011-04-17T12:32:00+1000

Flag an issue Contact curator

Dataset
Event
Taxonomy
Geospatial
Images
Date quality tests (1 4 21 13 48 0)
Additional political boundaries information
Environmental sampling for this location

### Location of record



### Images



Photographer: Russell Best

### Dataset

Data resource	NatureShare
Catalogue number	2380_Gymnorhina_tibicen
Basis of record	Human observation
Observer	Best, R. Russell Supplied as 'Russell Best'
Rights	CC BY 2.5 AU
More details	<a href="http://natureshare.org.au/observation/2380/">http://natureshare.org.au/observation/2380/</a>
Photographer	Russell Best
Rightholder	Russell Best via NatureShare
Occurrence remarks	Tags: Female
Occurrence status	present
Abcd identification qualifier	Not provided

### Event

Record date	[date not supplied] Supplied date '2011-04-17T12:32:00+1000'
Event remarks	Photo date/time used.

### Taxonomy

Scientific name	<i>Cracticus tibicen</i> Supplied scientific name 'Gymnorhina tibicen'
Taxon rank	Species
Common name	Australian Magpie
Kingdom	Animalia
Phylum	Chordata
Class	Aves
Order	Passeriformes
Family	Artamidae
Genus	<i>Cracticus</i>
Species	<i>Cracticus tibicen</i>

<http://biocache.ala.org.au/occurrences/544b0271-5f04-47ab-9d8b-0dbe3b5f59d7>

# Metadata – Atlas Of Living Australia

## Dataset

<b>Data resource</b>	NatureShare
<b>Catalogue number</b>	2380_Gymnorhina_tibicen
<b>Basis of record</b>	Human observation
<b>Observer</b>	Best, R. Russell <i>Supplied as "Russell Best"</i>
<b>Rights</b>	CC BY 2.5 AU
<b>More details</b>	<a href="http://natureshare.org.au/observation/2380/">http://natureshare.org.au/observation/2380/</a>
<b>Photographer</b>	Russell Best
<b>Rightsholder</b>	Russell Best via NatureShare
<b>Occurrence remarks</b>	Tags: Female
<b>Occurrence status</b>	present
<b>Abcd identification qualifier</b>	Not provided

# Standards and metadata

- Metadata
  - helps to understand and interpret data
  - provides details about experiment setup
    - who, when, in which conditions, tools, versions, etc.
  - helps identify and discover new data
- Use community standards to enable interoperability
  - <http://www.dcc.ac.uk/resources/metadata-standards>
  - <http://rd-alliance.github.io/metadata-directory/standards/>

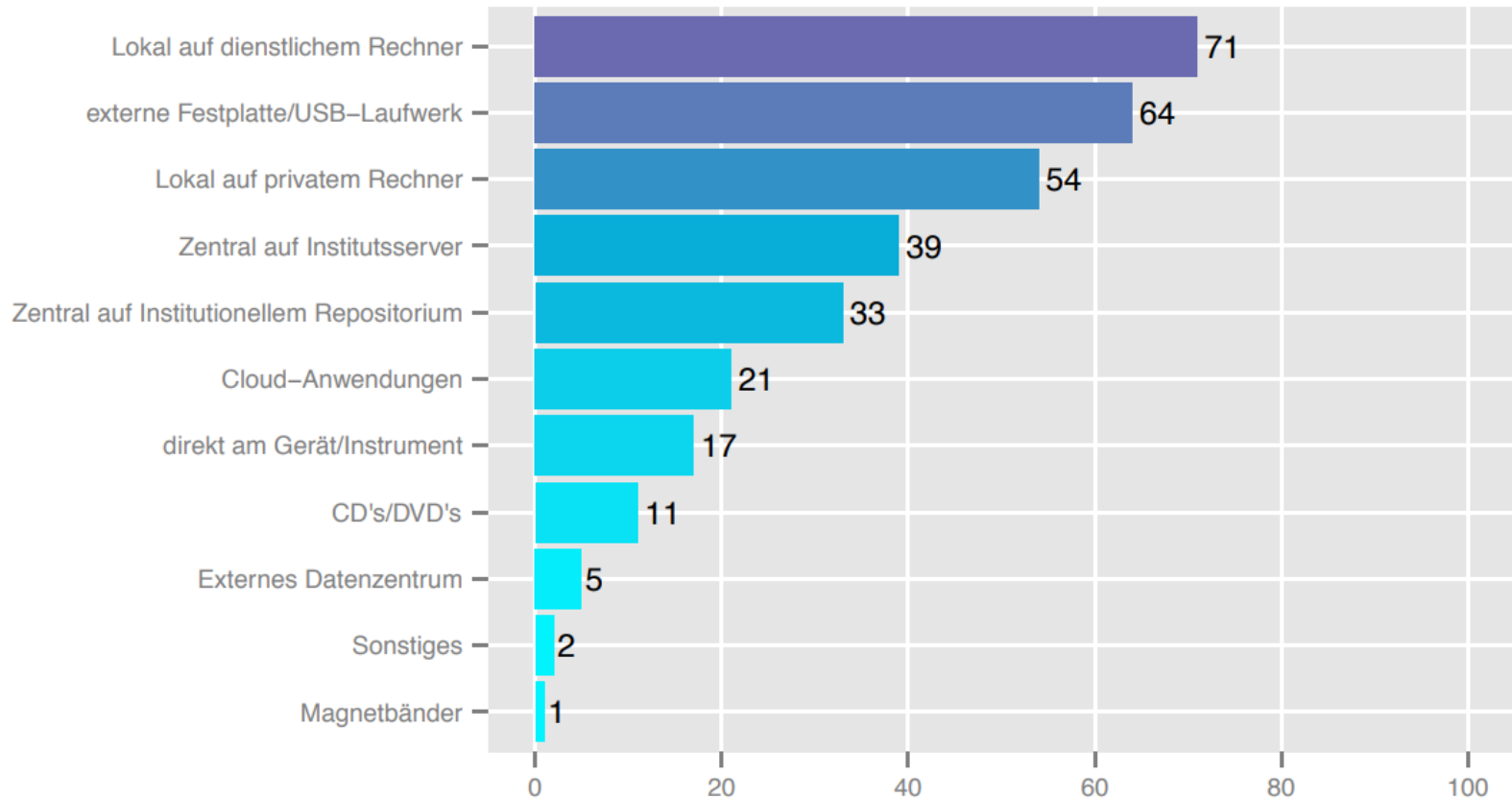
## Standards and metadata

We will first document our metadata by taking careful notes in the laboratory notebook that refer to specific data files and **describe all columns, units, abbreviations, and missing value identifiers**. These notes will be transcribed into a **.txt document that will be stored with the data file**. After all of the data are collected, we will then use EML (Ecological Metadata Language) to digitize our metadata. **EML is one of the accepted formats used in ecology**, and works well for the types of data we will be producing. We will create these metadata using Morpho software, available through KNB. The metadata will fully describe the data files and the context of the measurements.

From DataOne – E. affinis DMP example

# Managing data during research

## Wo speichern Sie normalerweise Ihre Forschungsdaten ab?



Anzahl der Antworten, Skalierung in %

e-infrastructures  
austria



<http://phaidra.univie.ac.at/o:407513>

# Managing data during research

- If you loose your data there will be nothing to share!
- Recreating or recollecting data can be
  - impossible
    - e.g. observational data
  - too expensive
    - e.g. cost of computational power
- How do you manage data during the project?
  - file naming convention
  - versioning
  - backups
  - should the access be restricted?
  - who is responsible?





# Data sharing

With collaborators while research is active



Data are mutable

(Open) data sharing



Data Repository

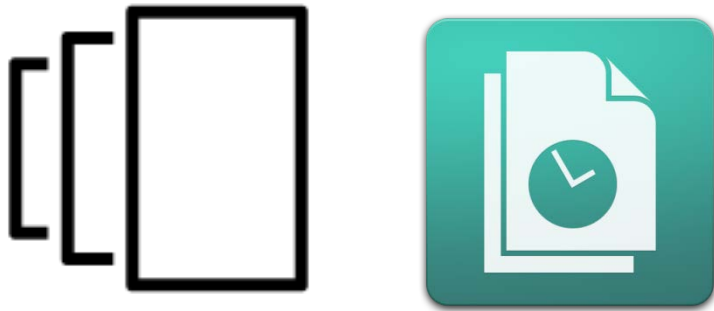


zenodo

Data are stable, searchable, citable, clearly licensed

# Backup vs preservation

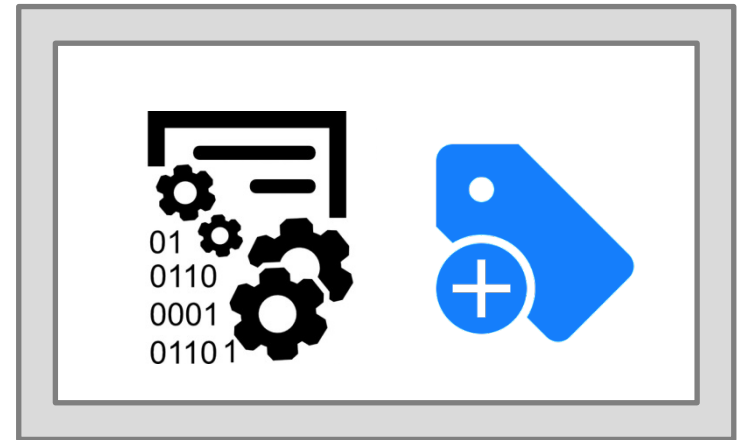
Storing and backing up files while research is active



Likely to be on a networked filestore or hard drive

Easy to change or delete

Archiving or preserving data in the long-term



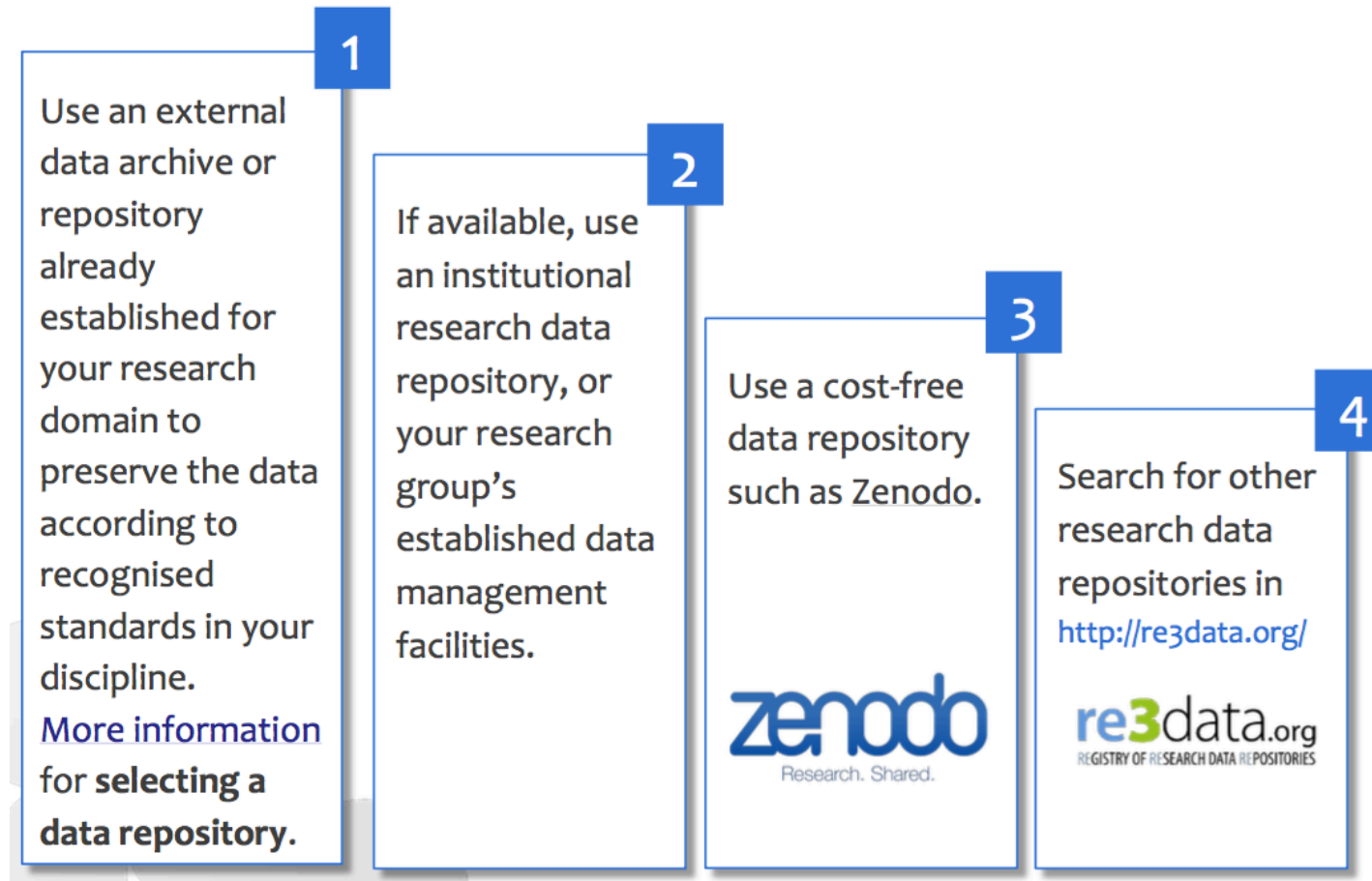
Likely to be deposited in a digital repository

Safeguarded and preserved

# Archiving and preservation

- Which data will be shared?
  - What has to be kept?
  - What can't be recreated?
  - What is potentially useful to others?
  - What has scientific, cultural or historical value?
  - What legally must be destroyed?
- Where will the data be deposited?
  - not all of the data must be shared in the same way
- Are there any embargo periods?
- For how long?
- What is the cost and who will pay for it?
- Which license to use?

# Where to find a repository?

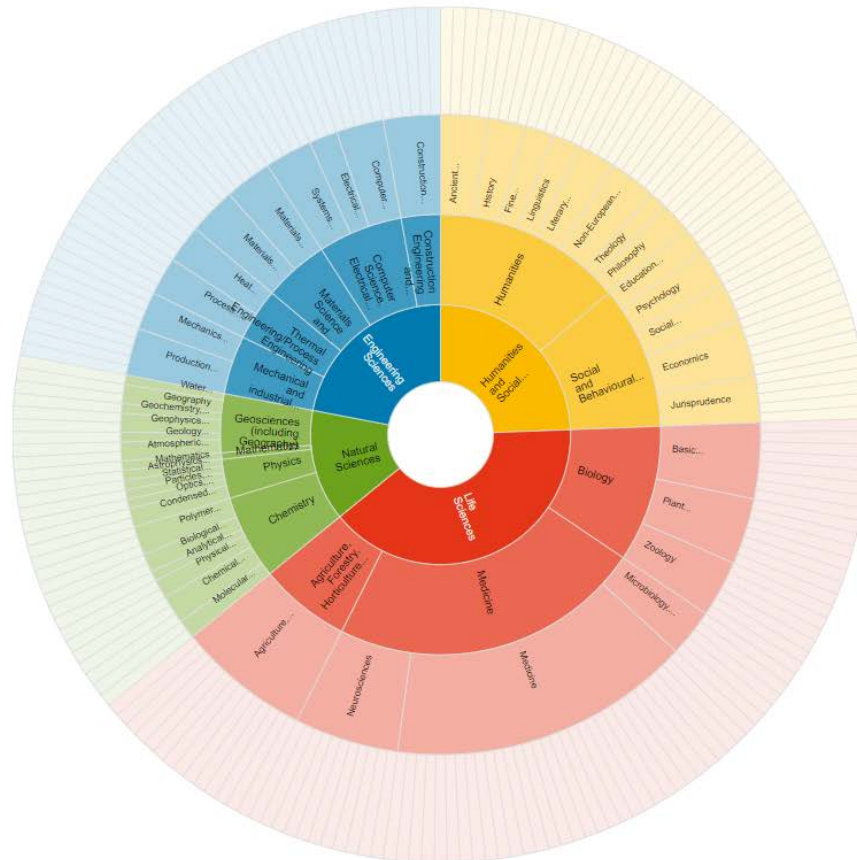


- More information: <https://www.openaire.eu/opendatapilot-repository>
- Zenodo: <http://www.zenodo.org>
- Re3data.org: <http://www.re3data.org>

## Browse by subject

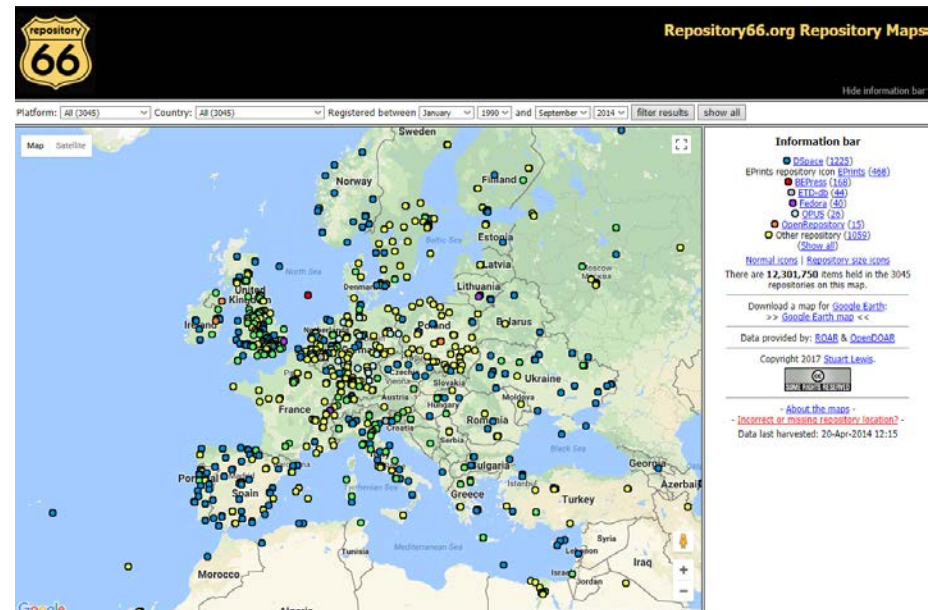
Graphical  Text

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



# Repository registries

- Directory of Open Access Repositories – DOAR
  - <http://www.opendoar.org/>
- Registry of Open Access Repositories – ROAR
  - <http://roar.eprints.org/>
- Projection of DOAR and ROAR onto google maps
  - <http://maps.repository66.org>



# Repositories in Austria - examples

The screenshot shows the Phaidra website interface. At the top, it features the University of Vienna logo and navigation links. A search bar is prominently displayed. Below the search bar, there is a 'Featured collections' section with several items, each with a thumbnail image and a brief description. For example, one collection is titled 'E-Books on Demand' and another is 'Digitales Forschungsarchiv Byzanz'. A 'Contact' sidebar on the right provides contact information for phaidra@univie.ac.at and lists social media links for Facebook, Twitter, LinkedIn, and YouTube.

<https://phaidra.univie.ac.at>

The screenshot displays the CCCA Data Server website. The header includes the CCCA logo (Climate Change Centre Austria) and navigation links for 'Log In', 'Register', and 'Contact'. Below the header, there are tabs for 'Datasets', 'Organizations', 'Groups', 'Categories', and 'About'. The main content area features a 'Welcome to the CCCA Data Server' section with a descriptive paragraph. To the right, there is a 'Search data' box with a search input field containing 'E.g. environment' and a search button. Below the search box, 'Popular keywords' are listed: 'CCCA Dataset', 'glacier', and 'scaled distribution'. A 'CCCA Data Server statistics' box shows the following data: 1898 resources in, 146 datapackages, 33 organizations, and 6 groups. A map of Austria is shown with a color-coded temperature scale, titled 'Daily Mean Near-Surface Air Temperature'. At the bottom, there is a 'ZAMG' logo and a news item dated 05.10.2017 about 'Austrian PANGAEA data harvested'.

<https://data.ccca.ac.at>

# Licenses

- Horizon 2020 guidelines point to CC-BY or CC-0



- DCC How-to guide helps you to license data

[www.dcc.ac.uk/resources/how-guides/license-research-data](http://www.dcc.ac.uk/resources/how-guides/license-research-data)

- EUDAT licensing wizard help you pick licence for data & software

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>



# Data sharing and preservation

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](#)

# Persistent Identifiers

- Digital Object Identifier (DOI)
  - Uniquely identify objects
  - handle system
    - DOI assigned once
    - Physical location of data can change



10.5281/zenodo.1068223

- ORCID

- Unique user ID

ORCID

Connecting Research  
and Researchers

0000-0002-4929-7875

# ORCID Example

Search
English

[EDIT YOUR RECORD](#)
[ABOUT ORCID](#)
[CONTACT US](#)
[HELP](#)

Connecting Research and Researchers

4,115,029 ORCID iDs and counting. [See more...](#)

**Daniel Mietchen**

**ORCID ID**  
<https://orcid.org/0000-0001-9488-1870>

[Print view](#)

**Also known as**  
D. Mietchen, Mietchen, Daniel, Mietchen, D., EvoMRI, D Mietchen, Mietchen D, Mietchen-D

**Country**  
Germany

**Keywords**  
open science, open data, open access, magnetic resonance microscopy, evolution, biodiversity, social machines, vocal learning

**Websites**  
[Twitter](#)  
[Wikidata](#), [Wikipedia et al.](#)  
[GitHub](#)  
[Open Science Q & A](#)  
[Scholia](#)

**Other IDs**  
Scopus Author ID: 7801384320  
ResearcherID: A-7748-2009

**Employment (2)**
Sort

National Center for Biotechnology Information: Bethesda, MD, United States  
2015-03-01 to present | Intramural researcher (Computational Biology Branch)

Source: Daniel Mietchen

Museum für Naturkunde - Leibniz-Institut für Evolutions- und Biodiversitätsforschung: Berlin, Berlin, Germany  
2013-08-16 to 2015-02-28 | Researcher (Digital World)

Source: Daniel Mietchen

**Works (64)**
Sort

Machine-actionable data management plans (maDMPs)  
Research Ideas and Outcomes  
2017-04-05 | journal-article  
DOI: [10.3897/rio.3.e13086](#)

Source: CrossRef Metadata Search Preferred source

Progress in promoting data sharing in public health emergencies  
Bulletin of the World Health Organization  
2017-04-01 | journal-article  
DOI: [10.2471/blt.17.192096](#)

Source: CrossRef Metadata Search Preferred source

Strategies and guidelines for scholarly publishing of

# TIPS FOR WRITING DMPS

# Tips for writing DMPs

- DMPs vary across scientific domains
- No good or bad answers – rationale is important
- DMP can reveal how solid your research is
- Seek advice - consult and collaborate
- Discuss any technical issues with the IT personnel
- When answering questions from checklists write coherent text
- Be specific when referring to tools and standards
- Assign responsibilities and name responsible personnel

# Tips for writing DMPs

- Think about things early...
  - Negotiation on licenses and consent agreement may preclude later sharing if not careful
  - Useful to consider data issues at the consortium negotiation stage to make sure potential issues are identified and sorted asap
  - Manage your data correctly from the very beginning
    - backups, file naming conventions, access restrictions, metadata collection
  - Plan your budget

**Decisions made early on affect what you can do later**

Machine-actionable Data Management Plans

**FUTURE**

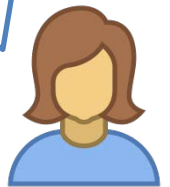
# Data Management Plans



How to discover these tools?  
Which one do I need to use?  
Why do I have to provide the same  
information again?

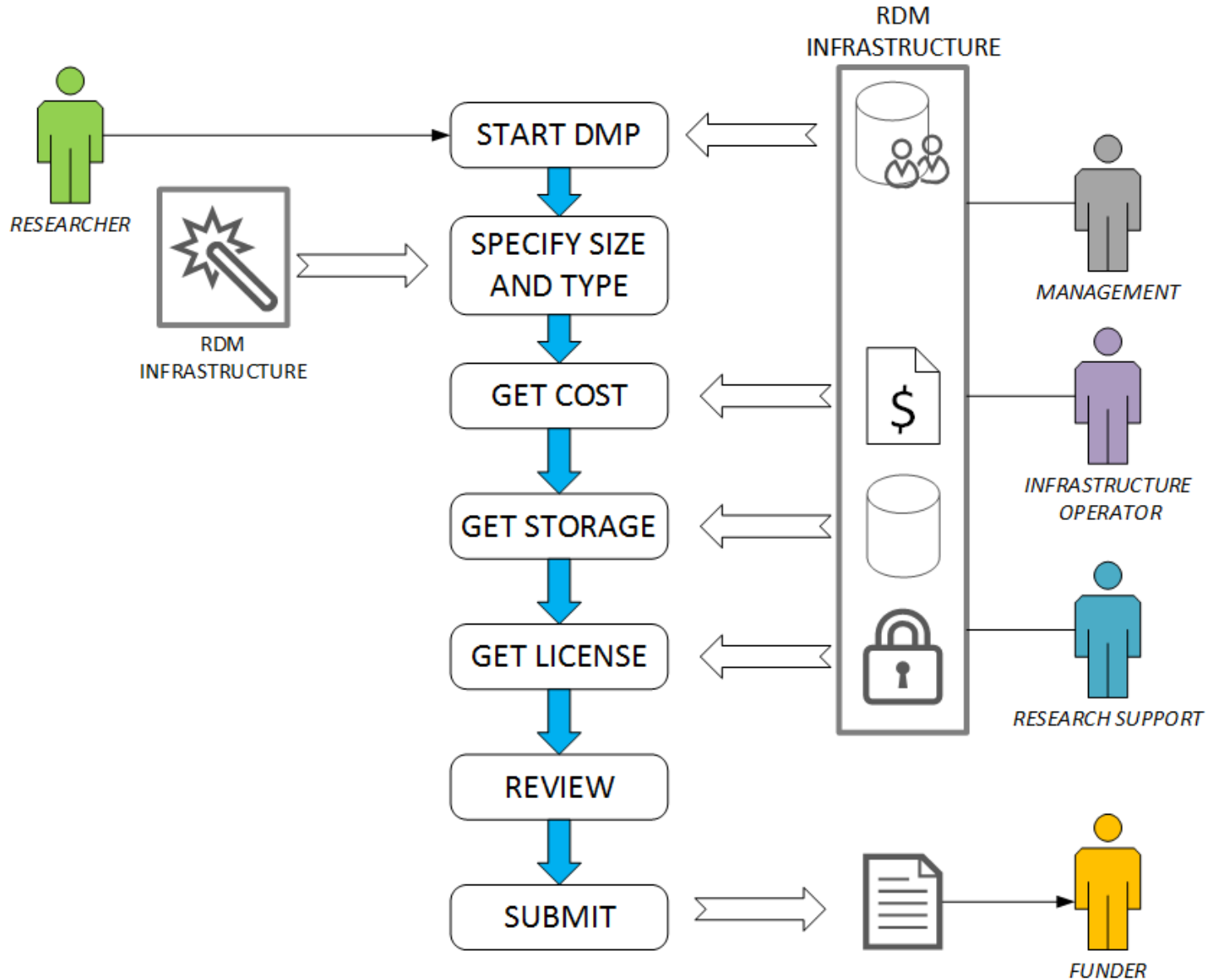


Why haven't they consulted us before?  
Who is going to pay for this?  
We don't have enough people for that!





# Automated Data Management Workflow



# Current activities

- RDA DMP Common Standards working group
  - goal is to define common data model for DMPs
  - <https://www.rd-alliance.org/groups/dmp-common-standards-wg>
- Join open consultation on GitHub
  - <https://github.com/RDA-DMP-Common/user-stories>



# CONCLUSIONS

# Conclusions

- DMPs are NOT meant to be yet another paper work imposed on researchers!
- DMPs are an awareness tool
- DMPs help you plan your project
- DMPs help in making data FAIR
- Future: machine-actionable DMPs
  - automate data management

# Useful resources

- Most links can be found on specific slides
- Managing and sharing data by UK Data Archive
  - <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>
- DMP Online
  - <https://dmponline.dcc.ac.uk>
- Ten Simple Rules
  - <http://dx.doi.org/10.1371/journal.pcbi.1004525>
- DMP Checklist
  - [http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP\\_Checklist\\_2013.pdf](http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf)
- EUDAT webinars on data management
  - <https://eudat.eu/training/research-data-management>
- DMP Schulungsunterlagen e-Infrastructures Austria
  - <http://phaidra.univie.ac.at/o:459770>

**Thank you! Any questions?**

[tmiksa@sba-research.org](mailto:tmiksa@sba-research.org)

**10.5281/zenodo.1068223**

**Acknowledgements:**

Thanks to EUDAT, DANS and DCC for reuse of slides, and to the OpenMinTeD and CAPSELLA projects for sharing their Data Management Plans