



How FAIR is your data?

An introduction to research data management

Roman Gerlach

Research Data Management Helpdesk
Michael-Stifel-Center Jena

Outline

1. Introduction / problem statement
2. How to do professional research data management?
3. Documenting your data/work
4. Quality control and assurance
5. Data preservation and publication
6. Data management planning



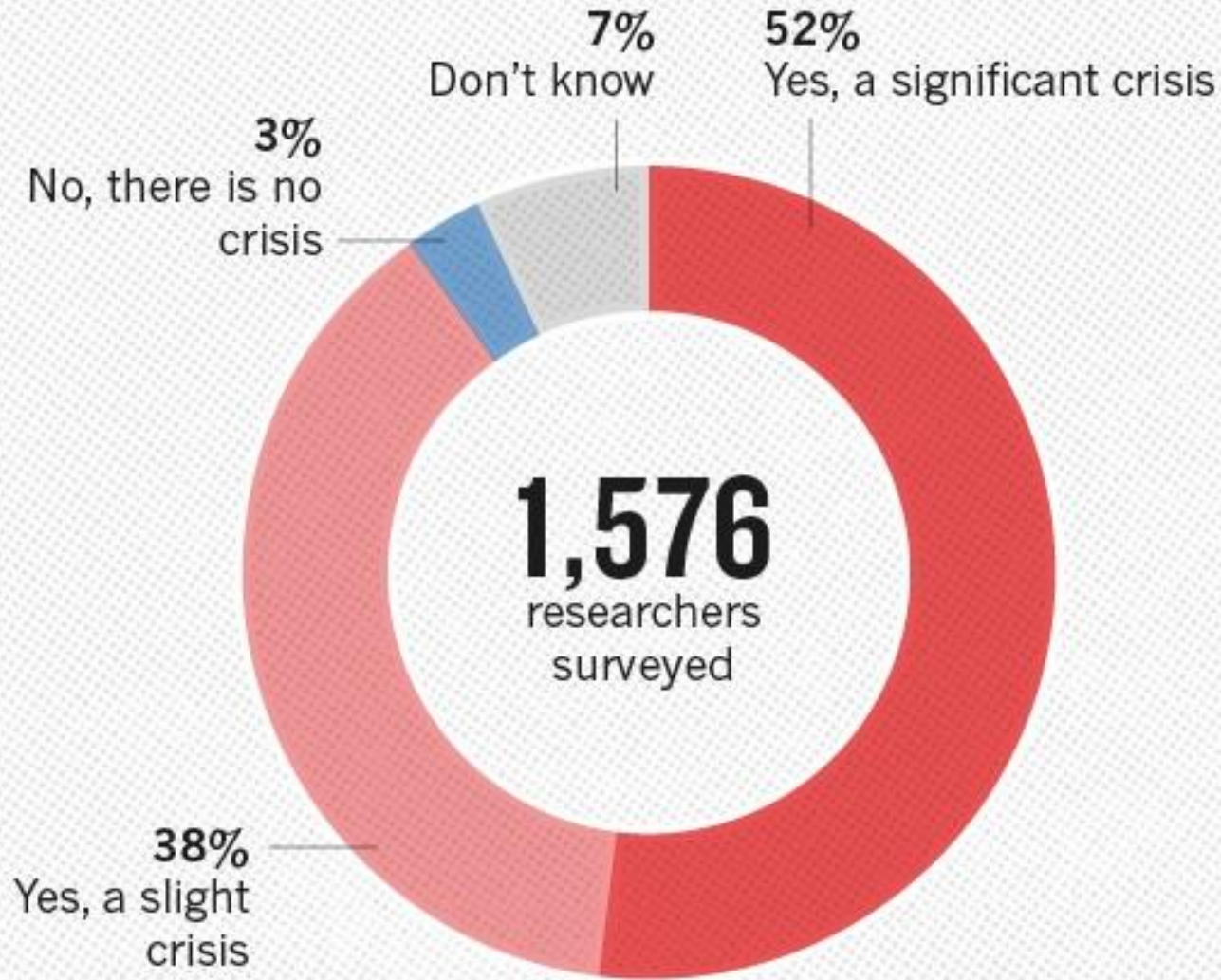
Data Sharing and Management Snafu in 3 Short Acts (Higher Quality)

19.020 Aufrufe

113 5 TEILEN SPEICHERN ...

https://www.youtube.com/watch?v=66oNv_DJuPc

IS THERE A REPRODUCIBILITY CRISIS?



©nature

Question

What would you do to ensure others are able to reproduce your research findings?

The three pillars of reproducibility

- Documentation
- Quality Control & Assurance
- Preservation & Publication

Data Policies

By international organisations

By funding agencies

By institutions

By journals

Example: EU

The screenshot displays the 'RESEARCH & INNOVATION Participant Portal H2020 Online Manual' interface. The top navigation bar includes the European Commission logo and the title 'RESEARCH & INNOVATION Participant Portal H2020 Online Manual'. The sidebar on the left contains a search bar and a list of navigation links: 'H2020 Online Manual', 'My Area - User account & roles' (with sub-links for EU Login, Roles & access rights, and Terms and Conditions of Use), 'Grants', 'Applying for funding' (with sub-links for Find a call, Horizon 2020 structure and budget, What you need to know about Horizon 2020 calls, and Find partners or apply as individual), 'Financial viability check', 'Data update', 'Certifications', 'Submit a proposal', 'Get prepared', and 'Electronic proposal'. The main content area is titled 'Data management' and includes a breadcrumb trail: '> H2020 Online Manual > Cross-cutting issues > Open access & Data management >'. Below the breadcrumb trail are two buttons: 'Open access' and 'Data management'. The 'Data management' button is highlighted. The main content area contains a section titled 'Background - Extension of the Open Research Data Pilot in Horizon 2020' with the following text: 'Please note the distinction between open access to scientific peer-reviewed **publications** and open access to research **data**:' followed by a bulleted list: '• **publications** – open access is an *obligation* in Horizon 2020.' and '• **data** – the Commission is running a flexible pilot which has been *extended* and is described below.' Below the list is a link: 'See also the Guidelines: [Open access to publications and research data in Horizon 2020](#).' A red box highlights a warning message: '⚠ Participating in the ORD Pilot does not necessarily mean opening up all your research data. Rather, the ORD pilot follows the principle "**as open as possible, as closed as necessary**" and focuses on encouraging sound data management as an essential part of research best practice.'

Example: DFG

Guidelines on the Handling Research Data (2015):

2. Accessibility

Assuming that the publication of research data from a DFG-funded project does not conflict with the rights of third parties (in particular data protection or copyright), research data should be made available as soon as possible. Data should be made accessible at a stage of processing that allows it to be usefully reused by third parties (raw data or structured data). To make sure this is the case, it must be ensured that access to the data is still guaranteed when, through publication, the rights of use relating to research data are transferred to a third party, usually a publishing house.

3. Long-term archiving

In accordance with the rules of good scientific practice, research data should be archived in the researcher's own institution or an appropriate nationwide infrastructure for at least 10 years.

More rules in Guidelines for proposals.

FAIR Principles

Findable

Accessible

Interoperable

Re-usable

<https://www.force11.org/group/fairgroup/fairprinciples>

FAIR Principles

- **'Findable'**, i.e. discoverable with metadata, identifiable and locatable by means of a standard identification mechanism;
- **'Accessible'**, i.e. always available and obtainable;
- **'Interoperable'**, i.e. both syntactically parseable and semantically understandable, allowing data exchange and reuse between researchers, institutions, organisations or countries; and
- **'Reusable'**, i.e. sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible and the least cumbersome integration with other data sources.

https://www.slideshare.net/OpenAIRE_eu/horizon-2020-open-research-data-pilot-jeanclaude-burgelman-dg-rtd-european-commission-8th-openaire-workshop

Publishers Perspective



nature publishing group

„[...] authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications.“



“PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.“

NPG (2013). Availability of data and materials. Retrieved from <http://www.nature.com/authors/policies/availability.html>
PLOS (2014). PLOS Editorial and Publishing Policies. Retrieved from <http://www.plosone.org/static/policies.action>

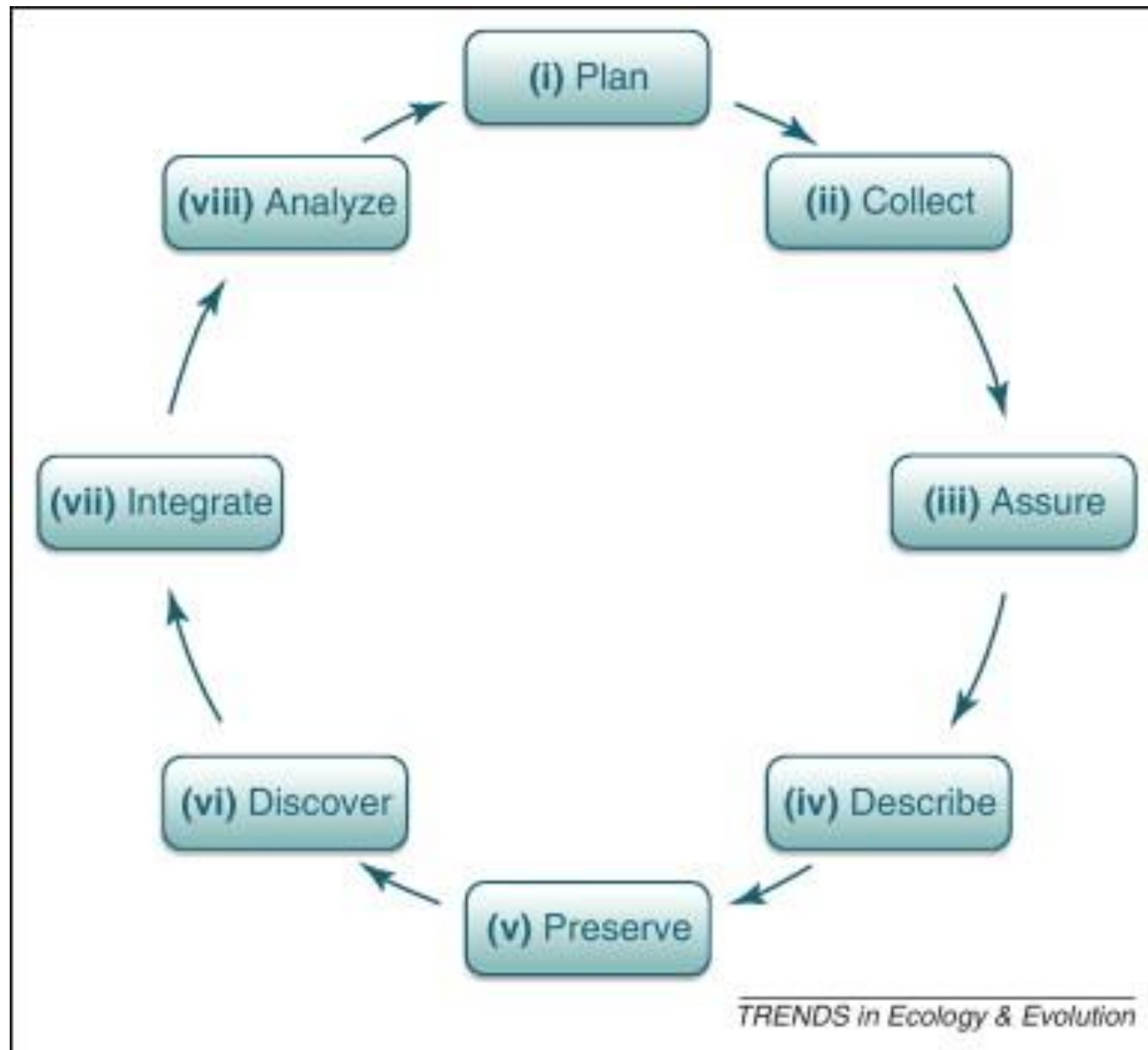


How to do professional research data management?

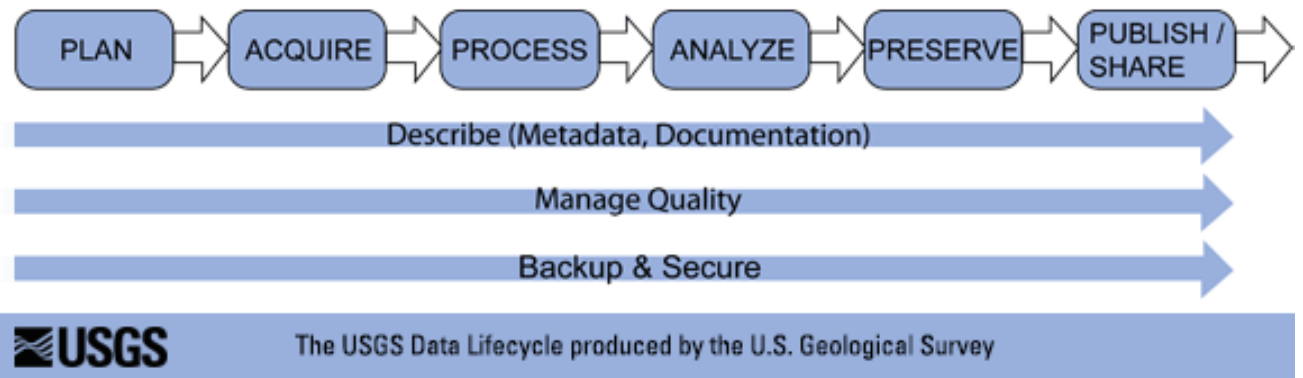
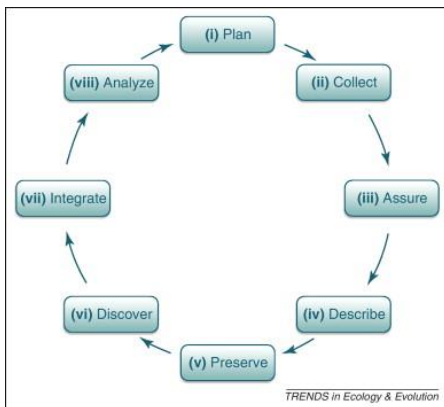
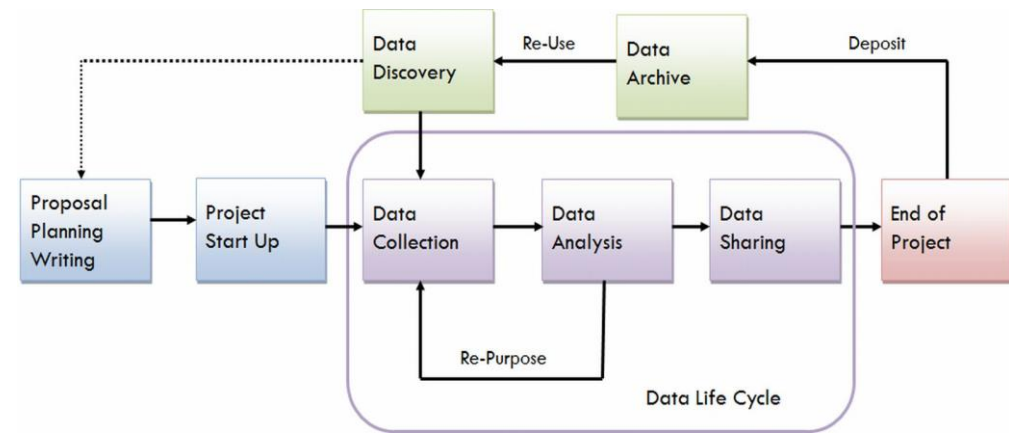
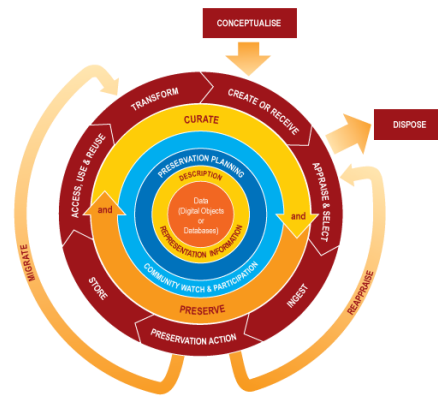
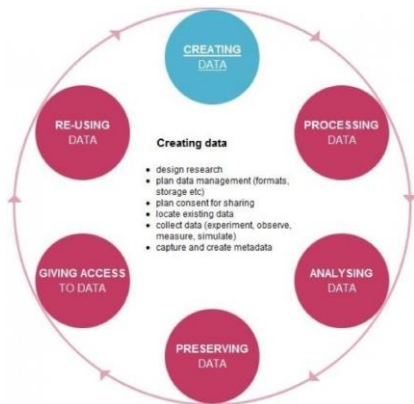
Recap: How does science work?



Data Life Cycle concept



There are many lifecycles...



Documenting your work/data

What is Metadata?

Metadata is: Data 'reporting'

- **WHO** created the data?
- **WHAT** is the content of the data?
- **WHEN** were the data created?
- **WHERE** is it geographically?
- **HOW** were the data developed?
- **WHY** were the data developed?



Photo by Michelle Chang. All Rights Reserved

Metadata collection

Field/Lab book



human readable

Metadata / Lineage

```
<?xml version="1.0"?>
<dwr:DarwinRecordSet
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://rs.tdwg.org/dwc/dwcrecord/ http://rs.tdwg.org/dwc/xsd/tdwg_dwc_classes.xsd"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
  xmlns:dwr="http://rs.tdwg.org/dwc/dwcrecord/">
  <dcterms:Location>
    <dwc:locationID>http://guid.mvz.org/sites/arg/127</dwc:locationID>
    <dwc:country>Argentina</dwc:country>
    <dwc:countryCode>AR</dwc:countryCode>
    <dwc:stateProvince>Neuquén</dwc:stateProvince>
    <dwc:locality>25 km al NNE de Bariloche por Ruta 40 (=237)</dwc:locality>
  </dcterms:Location>
  <dwc:Occurrence>
    <dcterms:type>PhysicalObject</dcterms:type>
    <dcterms:modified>2009-02-12T12:43:31</dcterms:modified>
    <dcterms:rightsHolder>Museum of Vertebrate Zoology</dcterms:rightsHolder>
    <dcterms:rights>Creative Commons License</dcterms:rights>
  </dwc:Occurrence>
</dwr:DarwinRecordSet>
```

human AND machine readable
+ searchable
+ interconnected (linked data)
+ standardized

How to create quality metadata?

Tips for Writing Quality Metadata

- Do not use jargon
- Define technical terms and acronyms:
 - CA, LA, GPS, GIS : what do these mean?
- Clearly state data limitations
 - E.g., data set omissions, completeness of data
 - Express considerations for appropriate re-use of the data
- Use “none” or “unknown” meaningfully
 - None usually means that you knew about data and nothing existed (e.g., a “0” cubic feet per second discharge value)
 - Unknown means that you don’t know whether that data existed or not (e.g., a null value)



Tips for Writing Quality Metadata

A Clear Choice: Which title is better?

Rivers



CC image by dolfi on
Flickr

OR

*Greater Yellowstone Rivers from 1:126,700 U.S. Forest Service Visitor
Maps (1961-1983)*

Greater Yellowstone (where) Rivers (what) from 1:126,700 (scale) U.S.
Forest Service (who) Visitor Maps (1961-1983) (when)

Tips for Writing Quality Metadata

Use standards whenever possible:

ISO 639 Code for the representation of the **names of languages**.

e.g. eng – English; ger/deu – German; fre/fra - French

<http://www.loc.gov/standards/iso639-2/langhome.html>

ISO 3166 Codes for the representation of **names of countries**.

e.g. DE – Germany; IR – Iran; CN – China; HR - Croatia

<http://www.oasis-open.org/cover/country3166.html>

ISO 8601 Codes for the representation of **date and time**.

e.g. 2014-06-19T13:15:30Z; 2014-06-19T13:16:30+01:00

http://en.wikipedia.org/wiki/ISO_8601

Tips for Writing Quality Metadata

- Remember: a computer will read your metadata
- Do not use symbols that could be misinterpreted: Examples:
! @ # % { } | / \ < > ~
- Don't use tabs, indents, or line feeds/carriage returns
- When copying and pasting from other sources, use a text editor (e.g., Notepad) to eliminate hidden characters

The 20-Year Rule

- The metadata accompanying a data set should be written for a user 20 years into the future--what does that investigator need to know to use the data?
- Prepare the data and documentation for a user who is unfamiliar with your project, methods, and observations

The three pillars of reproducibility

- Documentation 
- Quality Control & Assurance
- Preservation & Publication

Quality Control & Assurance

Data quality

What does quality assurance mean?

- to make sure others can understand your data (i.e. your conclusions)
- to enable others to assess **fitness for use**

Reference: A.D. Chapman, 2005,
Principles of Data Quality

Data quality

How to assure data quality?

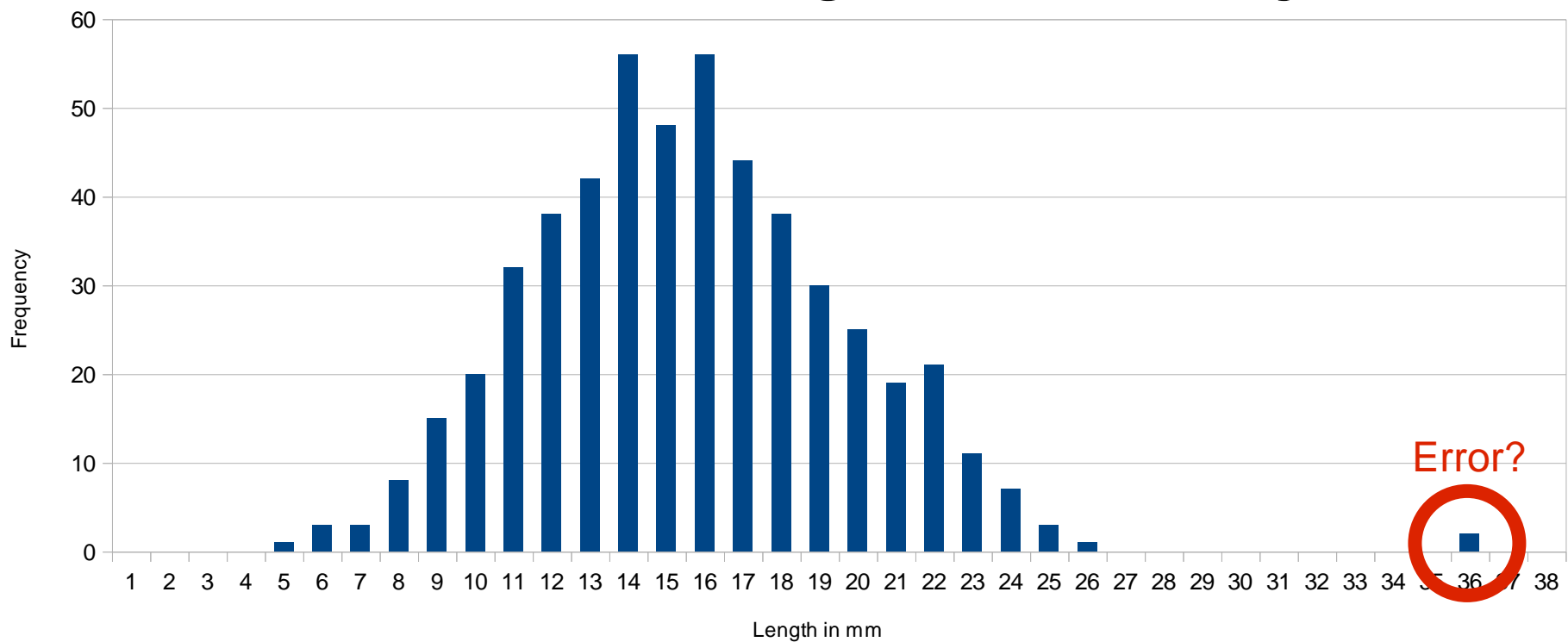
- Verification and review
- Ensure consistency, accuracy, completeness, reliability, ...
- providing context information = documentation, provenance

Reference: A.D. Chapman, 2005,
Principles of Data Quality

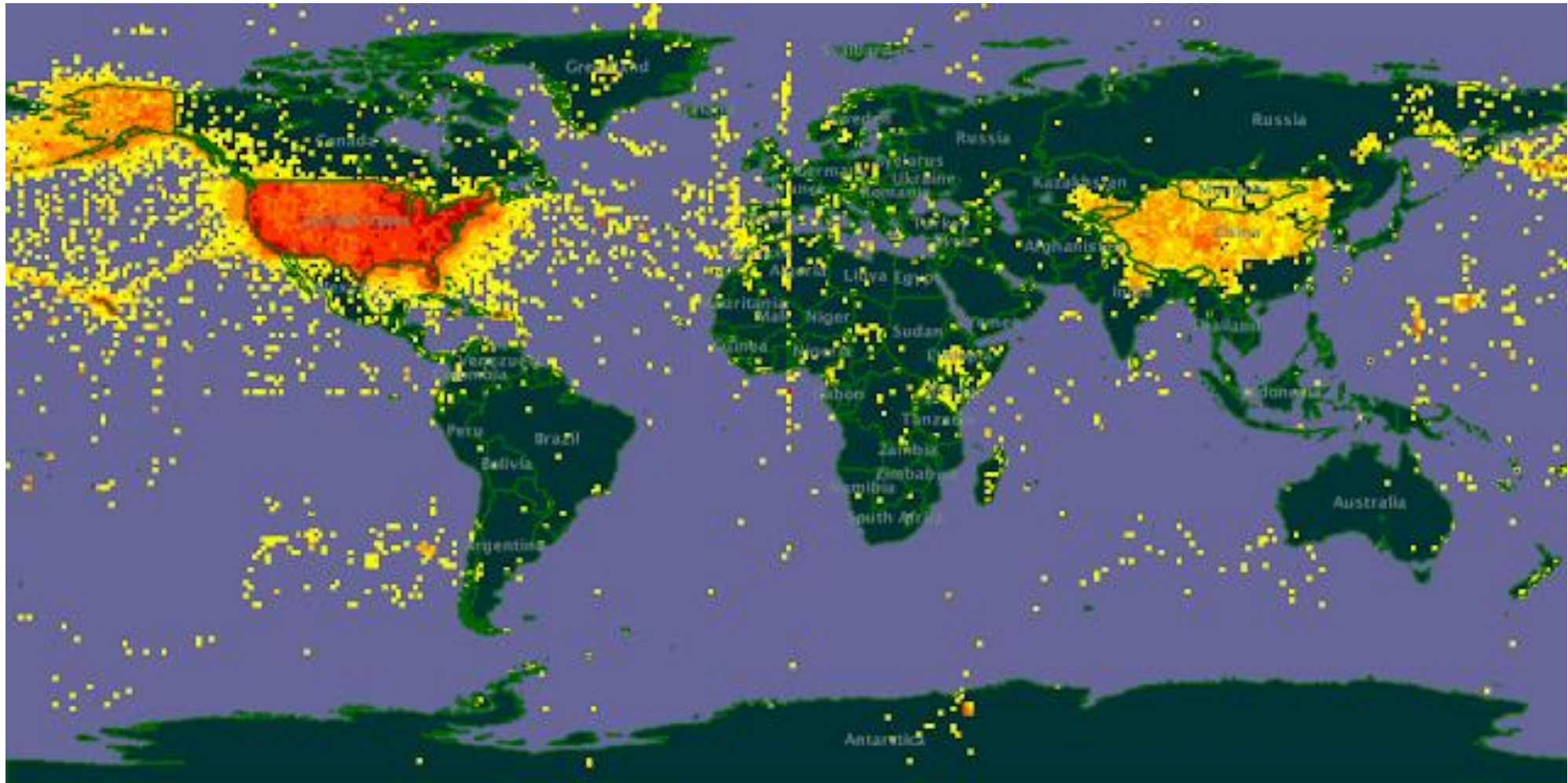
Error detection

Visualize data values

Tab. 1: Overall Nose's Length of *Nasobema lyricum*



Visualization



Place geographic data on a map to ensure that geographic coordinates are correct.

Tool: OpenRefine

Google refine *A power tool for working with messy data.*

Create Project « Start Over Configure Parsing Options Project name 00_smallaphiddataset.csv Create Project »

Open Project
Import Project

Excerpt from <http://lter.kbs.msu.edu/datasets/52> with added errors

	Column1	Column2	Column3	Column4	Column5
1.	sample_date	state	station	species	sex count
2.					
3.					
4.	25.10.13	IA	Nashua	Rhopalosiphum.sp.	male 1
5.	25.10.13	IA	McNay	Rhopalosiphum.padi	F 4
6.	25.10.13	IA	Nashua	Rhopalosiphum.insertum	F 1
7.	25.10.13	IA	McNay	Rhopalosiphum.insertum	F 1
8.	25.10.13	IA	McNay	Anoecia.cornicola	F 2
9.	25.10.13	IA	McNay	Hysteronura.setariae	F 1
10.	25.10.13	IA	McNay	Dysaphis.plantaginea	F 2
11.	25.10.13	IA	McNay	Nearctaphis.crataegifoliae	F 1
12.	25.10.13	IA	McNay	Hyadaphis.foeniculi	F 1
13.	18.10.13	IA	McNay	Myzus.lythri	F 5
14.	18.10.13	IA	McNay	Dysaphis.plantaginea	F 7
15.	18.10.13	IA	Nashua	Rhopalosiphum.rufiabdominale	F 1
16.	18.10.13	IA	McNay	Rhopalosiphum.rufiabdominale	F 1
17.	18.10.13	IA	Ames	Anoecia.cornicola	F 1
18.	18.10.13	IA	McNay	Anoecia.cornicola	F 2
19.	18.10.13	IA	McNay	Anoecia.corni	F 1
20.	18.10.13	IA	Ames	Nearctaphis.bakeri	F 1
21.	18.10.13	IA	McNay	Anoecia.oenotherae	F 1
22.	18.10.13	IA	McNay	Nearctaphis.crataegifoliae	F 3
23.	18.10.13	IA	McNay	Nearctaphis.crataegifoliae	M 1
24.	18.10.13	IA	McNay	Hyadaphis.foeniculi	F 3

Parse data as

Character encoding

Update Preview

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

RDF/N3 files

XML files

Open Document Format spreadsheets (.ods)

RDF/XML files

Columns are separated by

☐ commas (CSV)

☐ tabs (TSV)

☒ custom ;

Escape special characters with \

☐ Ignore first 0 line(s) at beginning of file

☒ Parse next 1 line(s) as column headers

☐ Discard initial 0 row(s) of data

☐ Load at most 0 row(s) of data

☒ Parse cell text into numbers, dates, ...

☒ Quotation marks are used to enclose cells containing column separators

☒ Store blank rows

☒ Store blank cells as nulls

☐ Store file source (file names, URLs) in each row


















Version 2.5 [r2407]

Help
About

File Naming

A STORY TOLD IN FILE NAMES:

Location:  C:\user\research\data

Filename	Date Modified	Size	Type
 data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
 data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
 data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
 data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
 data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
 data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
 data_2010.05.29_aaarrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file
 data_2010.05.29_#\$\$@*&!!.dat	2:40 AM 5/29/2010	0 KB	DAT file
 data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
 data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
 data_2010.05.29_woohoo!!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
 data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
 analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
 ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
 Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
 JUNK...	2:45 PM 5/29/2010		Folder
 data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file

File Naming

Use descriptive file names

- Unique
- Reflect contents
- ASCII characters only
- Avoid spaces

Bad: Mydata.xls

2001_data.csv
best version.txt

Better: bigfoot_agro_2000_gpp.tiff

Project
Name

Site
name

Year

What was measured
(gross primary productivity)

File
Format

File Naming

- Be **consistent** throughout your project directories/files
- Keep it **short**!
- A good format for **date designations** is YYYYMMDD_...
- use "001, 002, ...010, 011, etc." instead of "1, 2, ...10, 11 ... 100, 101, etc."

Folder Structure

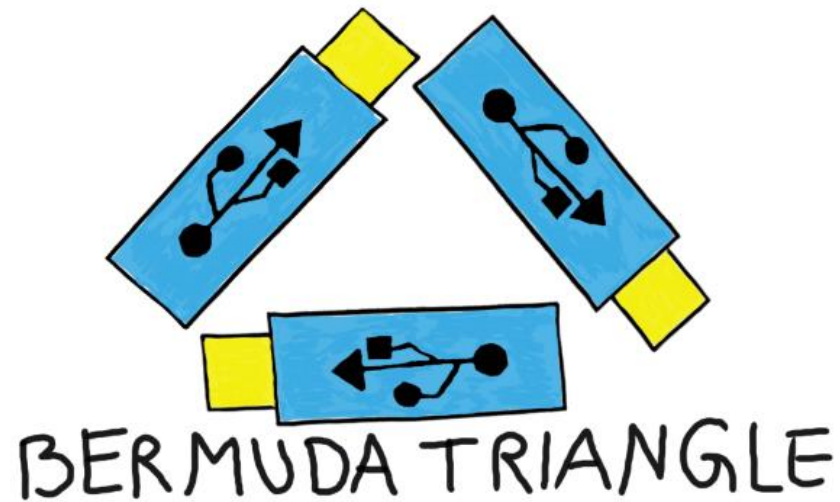
The **primary goals** of the folder structure are to

- Provide **predictable locations** for data files and other types of resources (e.g. code)
- Provide distinction between "**original**", "**finalized**" datasets and datasets that are still in production
- Provide a **space for archiving** outdated datasets
- for each directory include a **readme.txt** file that explains your naming format along with any abbreviations or codes you have used.

The three pillars of reproducibility


- Documentation ✓
- Quality Control & Assurance ✓
- Preservation & Publication

Data Preservation



Data Preservation

Or: How to make sure that data remains available after your project ends, after you leave the university, after your boss retires,...



Really available, not
just stored on some
tape in some dusty
basement....

Data Preservation

Or: How to make sure that data remains
available after your project ends, after you
leave the university, after your boss
retires,...

SPACE

McMoon's: The Former McDonald's Where NASA Digitizes Old Moon Photos



Matt Novak

8/19/13 12:59pm • Filed to: SPACE ✓

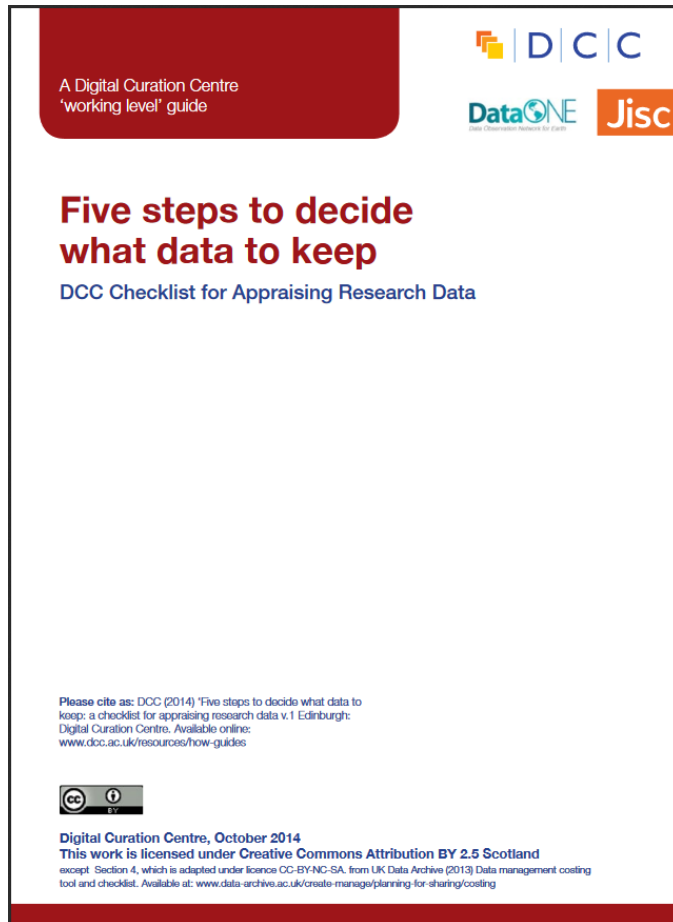
9

2



<http://www.moonviews.com/>

Five steps to decide what data to keep



What data and for how long?

Step 1. Identify **purposes** that the data could fulfill

Step 2. Identify data that **must** be kept

Step 3. Identify data that **should** be kept

Step 4. Weigh up the **costs**

Step 5. **Complete** the data appraisal

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

Identify data that **must** be kept

- Are there Research Data Policy reasons to keep it?
- Do regulations require the data to be available?
- Are there other legal or contractual reasons?
- Does it contain personal data relevant to the reuse purpose?

Identify data that **should** be kept

- Is it good enough?
- Is there likely to be a demand?
- How difficult is it to replicate?
- Do any barriers to further use exist?
- Is it the only copy?

Data appraisal

1. data that cannot be recreated, because it contains unique observations/measurements in space and time



Data appraisal

2. data that can only be reproduced with enormous effort
(e.g. super computer calculations)



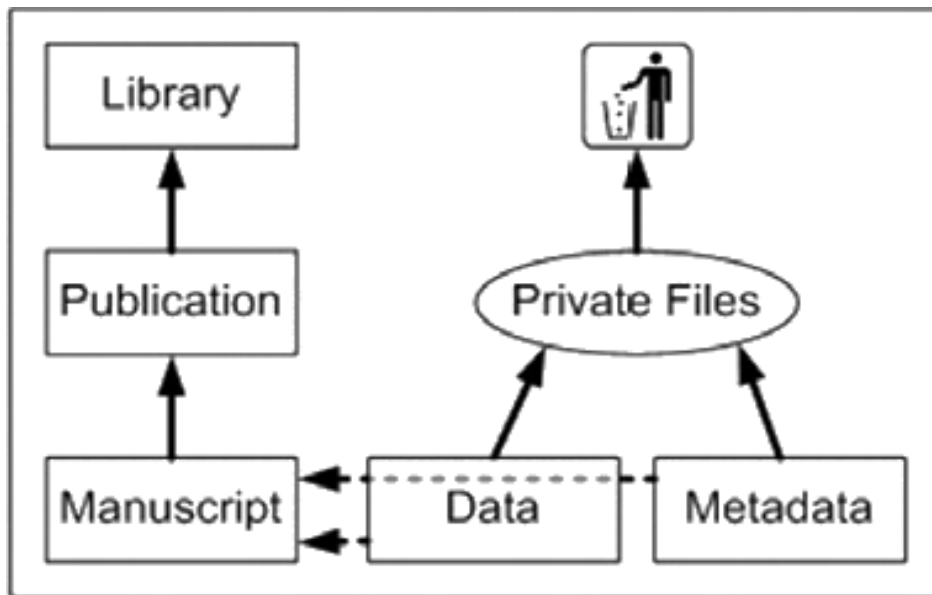
Long-term archiving

Challenges:

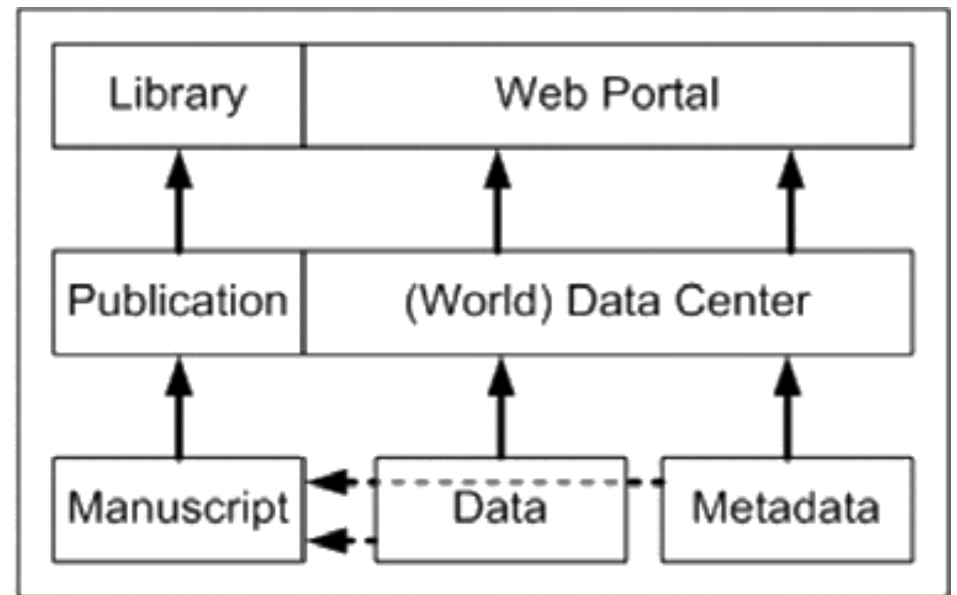
- File formats
- Migration (Hardware, Software, Format, ...)
- Data retirement/disposal
- Continuous funding

Data Publication

Data Publication today!?



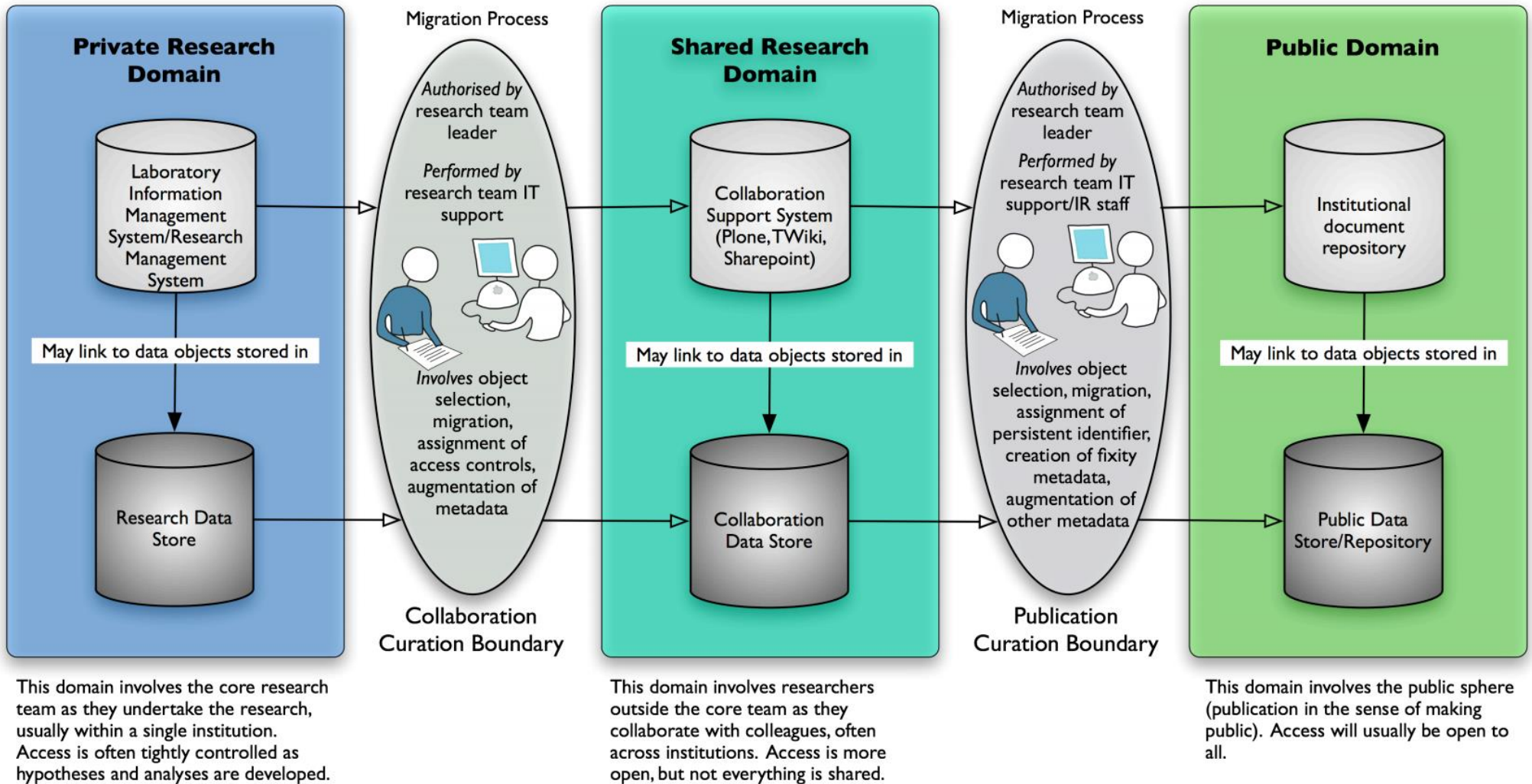
Common way



Preferred way

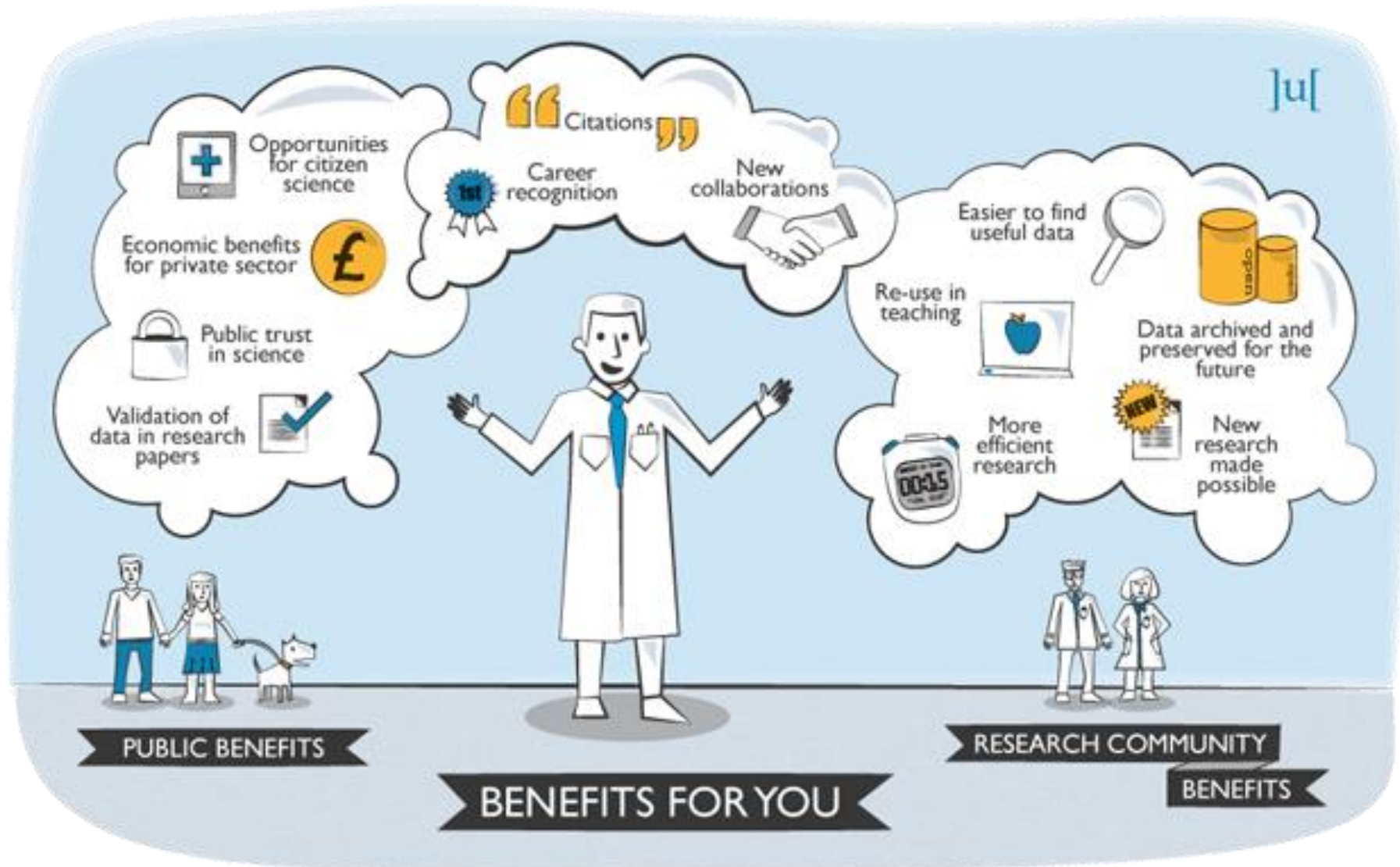
Rümpel 2011:26

Data Domains



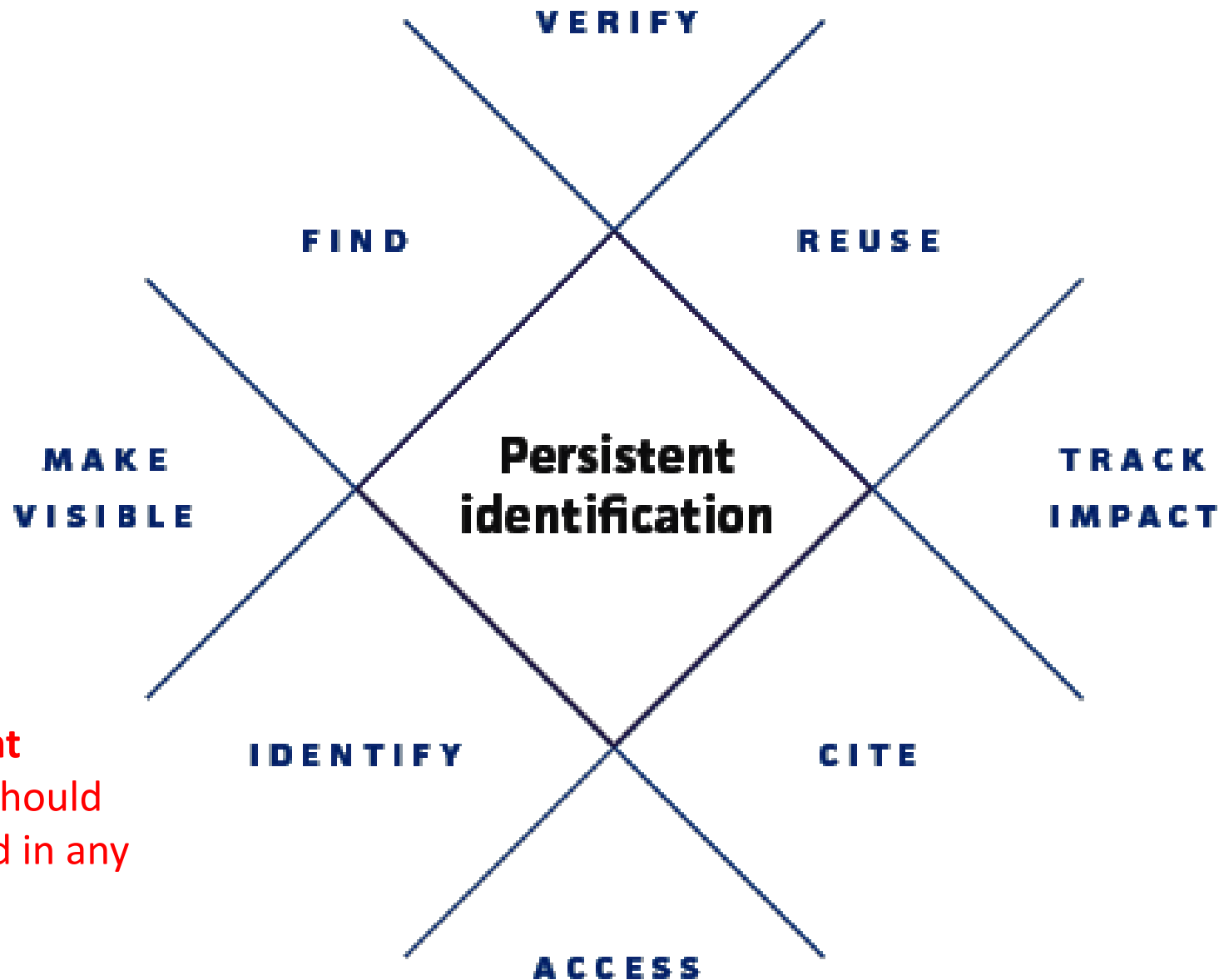
Version 1.4, <http://andrew.treloar.net/>, 07Dec07

Benefits of data publication



How to find something after years in the internet?

Persistent Identifier

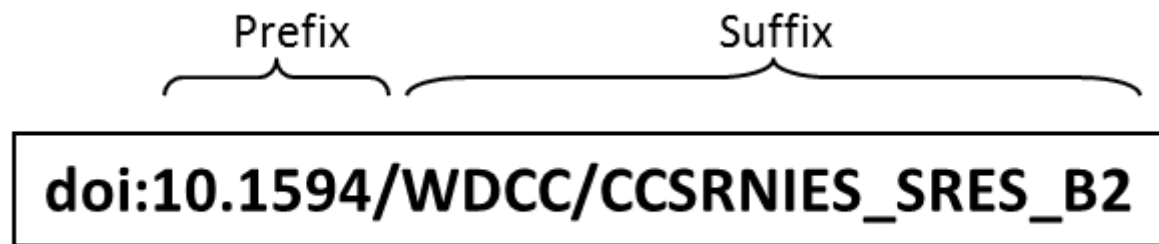


A persistent identifier should be included in any citation

PID Example: DOI




- **DOI** (Digital Object Identifier)
 - Globally unique, alphanumeric string assigned by a registration agency



- DOI name should **not** contain any changeable attributes of the object (e.g. physical location, ownership); attributes are encoded in the metadata, not in the DOI name
- A DOI should always resolve to a **landing page**

PID Example:

Researcher identifier: **ORCID** (Open Researcher & Contributor ID)

- Central registry of unique identifiers for individual researchers to address author name ambiguity
- Transparent linking mechanism between ORCID and other author ID schemes
- Example:  orcid.org/0000-0001-5104-4247

Where can I publish my data?

- Journal paper supplement
- Data Journal
- Data Repository

Journal

Journal = Paper + PID + [supplement]

The screenshot displays the Science Magazine website interface. At the top, the 'Science' logo is prominent, followed by navigation links like 'AAAS.ORG', 'FEEDBACK', 'HELP', and 'LIBRARIANS'. A search bar is located on the right. Below the header, a red navigation bar contains links for 'NEWS', 'SCIENCE JOURNALS', 'CAREERS', 'BLOGS & COMMUNITIES', 'MULTIMEDIA', and 'COLLECTIONS'. A secondary navigation bar includes 'Science Home', 'Current Issue', 'Previous Issues', 'Science Express', 'Science Products', 'My Science', and 'About the Journal'. The main content area shows the breadcrumb trail: 'Home > Science Magazine > Pereira et al., 339 (6117): 277-278 > Supporting Online Material'. On the left, an 'Article Views' sidebar lists options: 'Summary', 'Full Text', 'Full Text (PDF)', 'Figures Only', and 'Supplementary Materials'. The main article information includes the date 'Science 18 January 2013:', volume and page numbers 'vol. 339 no. 6117 pp. 277-278', and the DOI 'DOI: 10.1126/science.1229931'. The article title is 'Essential Biodiversity Variables', followed by a list of authors. A section titled 'Materials/Methods, Supplementary Text, Tables, Figures, and/or References' contains links for 'Download Supplement', 'Supplementary Text', 'Fig. S1', and 'Full References'. Two blue circles highlight the DOI and the supplementary materials section, with arrows pointing to labels on the right.

Science 18 January 2013:
vol. 339 no. 6117 pp. 277-278
DOI: 10.1126/science.1229931

Essential Biodiversity Variables

H. M. Pereira, S. Ferrier, M. Walters, G. N. Geller, R. H. G. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. M. Butchart, A. C. Cardoso, N. C. Coops, E. Dulloo, D. P. Faith, J. Freyhof, R. D. Gregory, C. Heip, R. Höft, G. Hurtt, W. Jetz, D. S. Karp, M. A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B. Reyers, R. Sayre, J. P. W. Scharlemann, S. N. Stuart, E. Turak, M. Walpole, M. Wegmann

Materials/Methods, Supplementary Text, Tables, Figures, and/or References

Download Supplement
Supplementary Text
Fig. S1
Full References

DOI

Supplement

Data Journal

Data Journal = Paper + Datasets + PID

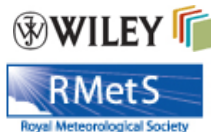


An 18-yr long (1993–2011) snow and meteorological dataset from a mid-altitude mountain site (Col de Porte, France, 1325 m alt.) for driving and evaluating snowpack models

S. Morin, Y. Lejeune, B. Lesaffre, J.-M. Panel, D. Poncet, P. David, and M. Sudul

Page(s) 13-21

▣ [Abstract](#) ▣ [Final Revised Paper](#) (PDF, 2559 KB) ▣ [Supplement](#) (14997 KB) ▣ [Discussion Paper](#) (ESSDD)



ONLINE LIBRARY
Geoscience Data Journal

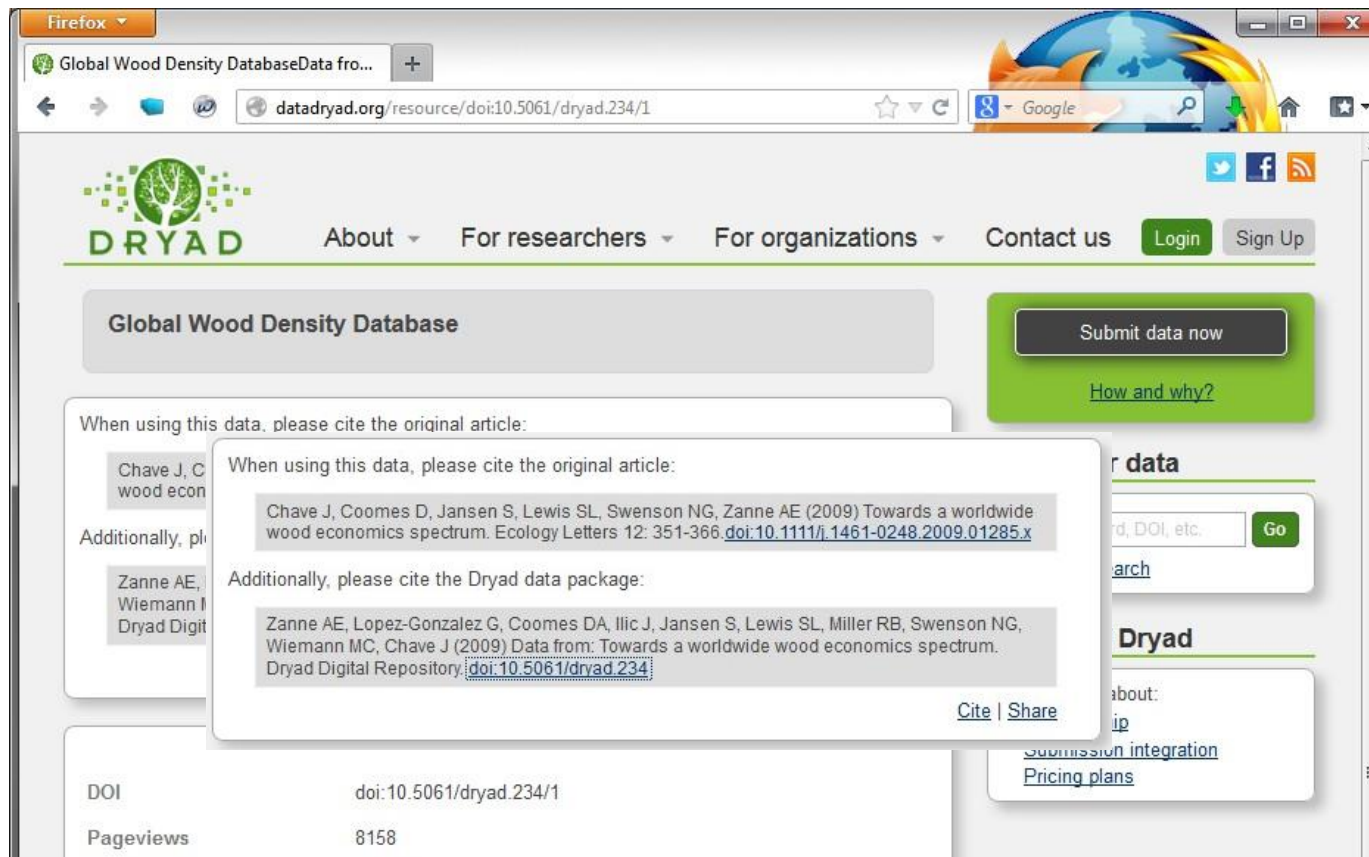


List of Data Journals:

<http://www.researchdata.uni-jena.de/Datenpublikation.html>

Data repository

Data repository = Datasets + Metadata + PID



Types of Repositories

- Domain specific repositories
- Generic/institutional repositories

How to find a data repository?



Home Search Browse Suggest FAQ About Schema API Contact Imprint

Search for Repositories (1223 Reviewed Repositories)

Subject: Content Type: Country (of the responsible institutions):

☒ Biology

☐ Certificates ☒ Open Access ☐ Persistent Identifier

421 results (filtered) (1 – 25)

« 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 »

Sort by weight

3TU.Datacentrum

3TU.DC



Subjects: **Biology** Agriculture, Forestry, Horticulture and Veterinary Medicine Agriculture, Forestry, Horticulture and Veterinary Medicine
Analytical Chemistry, Method Development (Chemistry) Basic Biological and Medical Research Bioinformatics and Theoretical Biology

How to select a repository

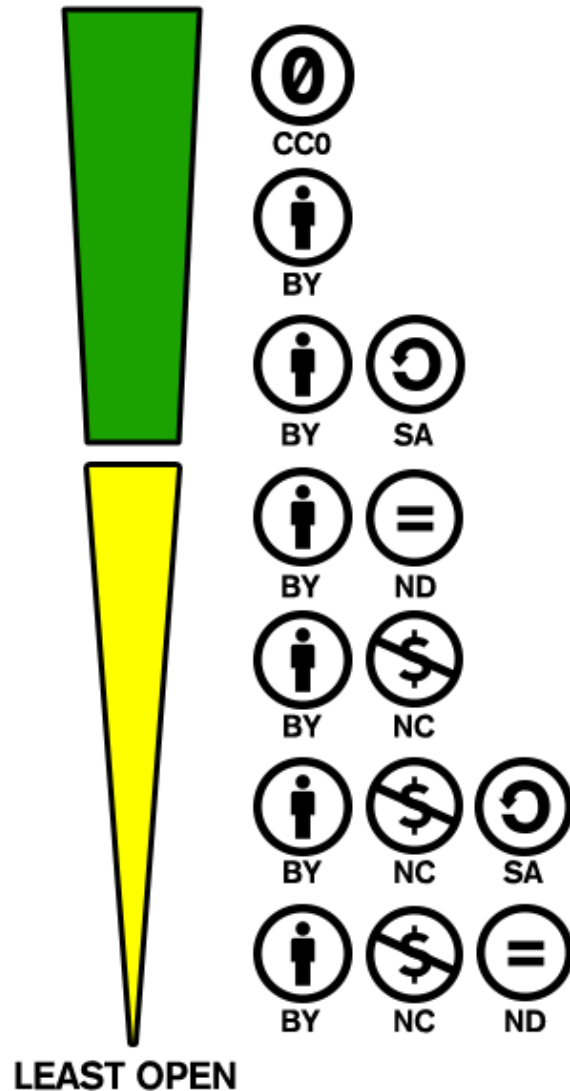
Criteria:

- Long-term availability (Certification?)
- Provides PIDs (e.g. DOI, URN, ARK)
- License (e.g. Open Access, Restricted)
- 1st domain, 2nd generic
- Data review process?



Licences: e.g. Creative Commons

MOST OPEN



Only version 4.0 of
CC0

CC-BY

CC-SA

conform to the Open Definition (i.e.
Open Access)

*Summary: Knowledge is open if
anyone is free to access, use,
modify, and share it — subject, at
most, to measures that preserve
provenance and openness.*

<http://opendefinition.org/od/2.1/en/>

Data Management Planning

1. How will the data be managed during the project?

Data Storage

- Is all data **digital**?
- What data **volume** will be produced?
More than 500GB?
- Do you have a **naming convention**?
- What **software** and **file formats** will you use?

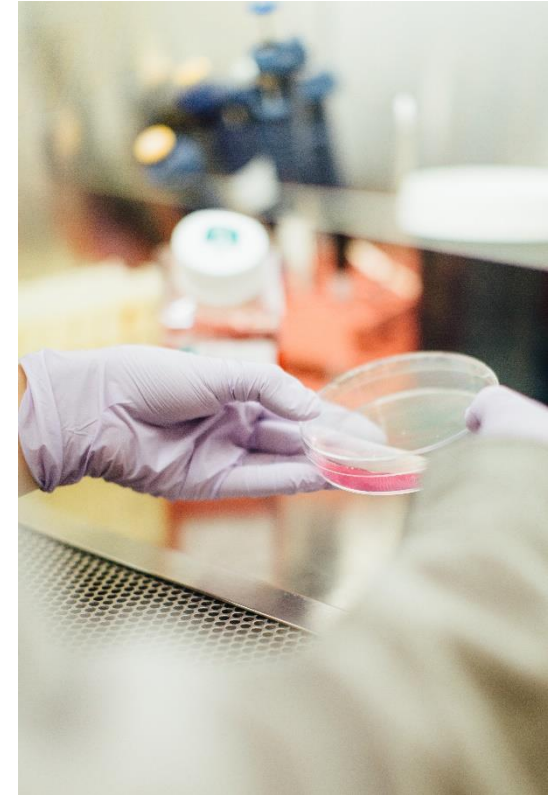


Photo by Drew Hays on Unsplash-206414

Data description and documentation

- What standards will be used to record the data?
- What additional information is required?
 - Collection methods
 - Context information (metadata)
- Where will the metadata be located



Photo by Stefan Stefancik on Unsplash-257625

Data Access & Security

- Version control (e.g. SVN, Git)
- Back-ups (frequency, media)
- Security & protection from manipulation/theft
- Who will be responsible?

2. How will the data be managed **after** the project?

Plans for longterm archiving

- Where will the data be archived?
- For how long?
- What data should be archived?
- What data can be deleted?



Photo by Sanwal Deen on Unsplash-93466

Data Sharing & Publication

- How can I share my data? Isn't my Website enough?
- Where can I publish my data (e.g. data journal, research data repository)?
- Where can I get persistent identifiers (PID) for my data (e.g. DOI, URN)?
- What data objects get a PID (i.e. data granularity)?

Other Aspects of Data Management Planning

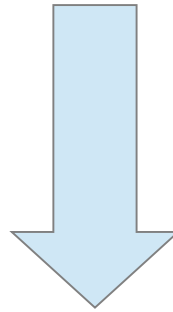
Legal & Ethical Aspects

- Ownership, copyright, intellectual property, licenses
- Are there ethical issues with the data (e.g. privacy, confidentiality, cultural sensitivity)
- Do I need written agreements from probands/patients?
- Do I need to anonymize the data?
- Do we need a data policy for our project?

Planning cost

- **Data collection** cost (e.g. data purchase, hardware, labour, usage of equipment)
- **Data infrastructure** cost (i.e. storage, computing power)
- **Labour** cost (e.g. data manager, system admin, Hiwi's)
- **Publication** cost (e.g. Open Access, archiving)

To ensure that no step of research data lifecycle is
forgotten



Research Data Management Plan
(DMP)

What is a DMP?

- (formal) document that outlines how you will handle your data both during your project, and after the project is completed
- It's a living document

Data Management Plan for Post-Graduate Research Projects

Researcher:	
Project Title:	
Project Duration:	
Project Context:	
1. What Data will be Produced?	
2. How will the Data be Documented and Described?	
3. Has a 'File Structure/Naming Form' been completed?	
4. Deposition of E-Thesis:	
5. What are the plans for data sharing and access after submission of the thesis?	
6. What are the plans for long-term archiving of the digital data supporting the thesis?	
Signed:	Version:
Date Created:	Date Amended:

For your project

1. Make a plan! (incl. back-up & sharing)
2. Document your processing steps
3. Make sure your data comes with metadata
4. Implement consistent file naming (+ versioning)
5. Prepare your data/project so that others can understand and use it

Thanks for attending!



Research Data Management Helpdesk



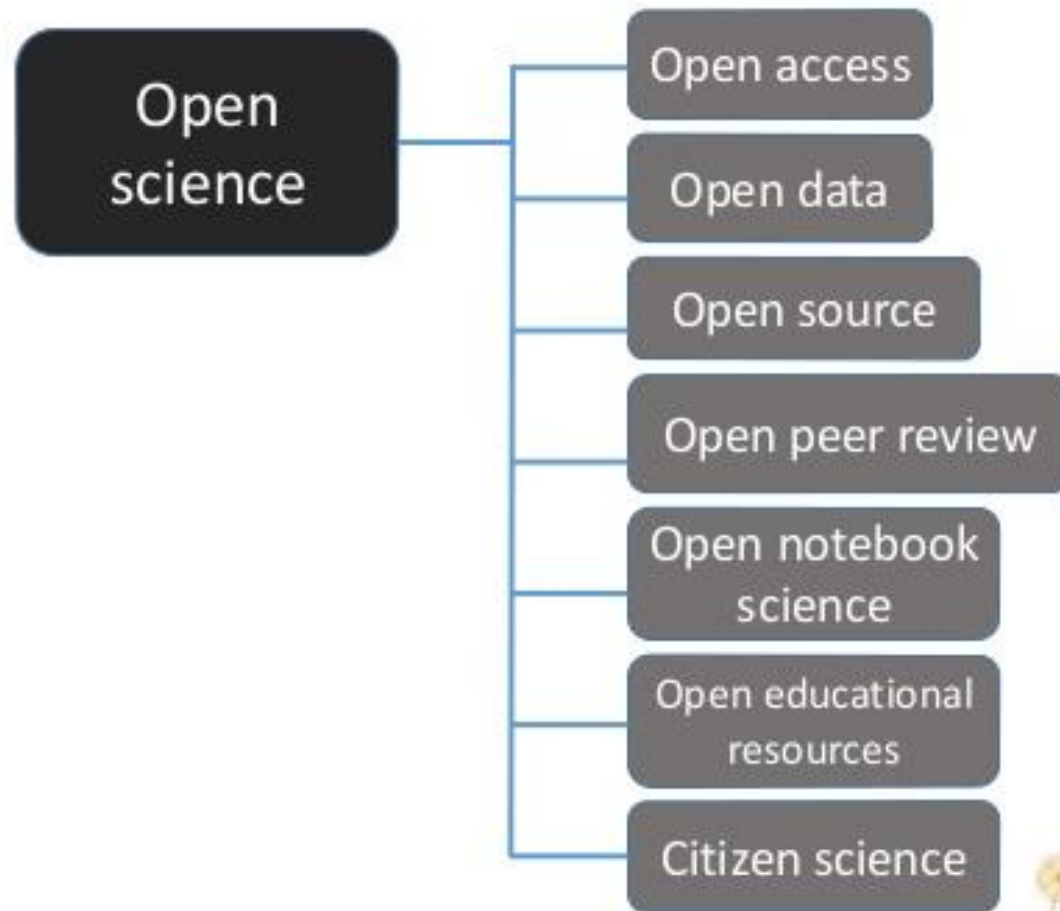
FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

www.researchdata.uni-jena.de
researchdata@uni-jena.de

+49-(0)3641-9-48-968
+49-(0)3641-9-46-363

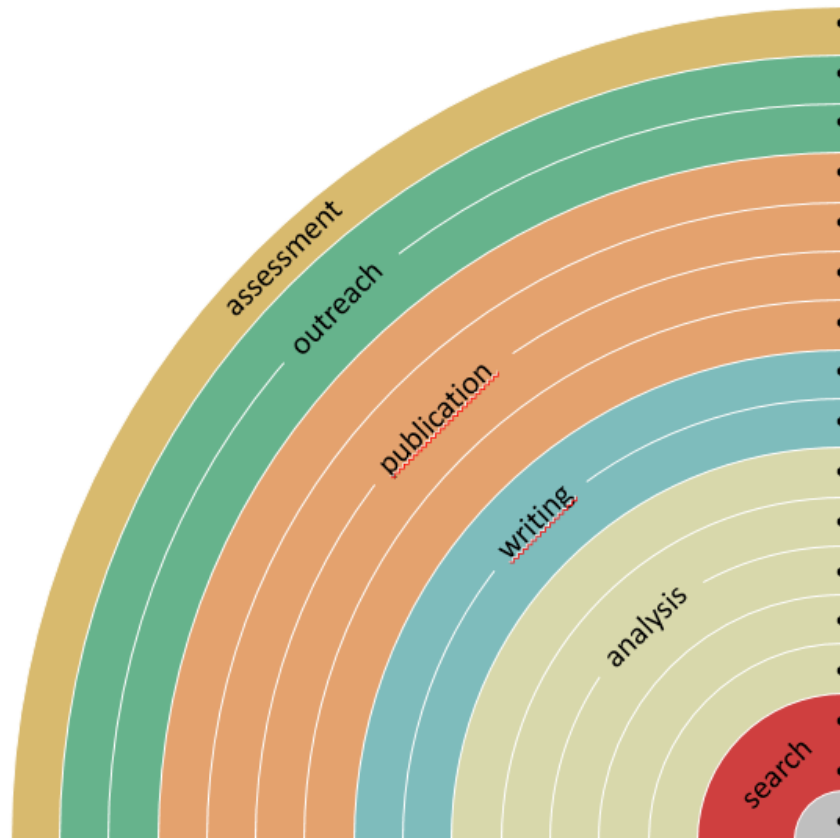
Backup slides

Open Science



Source: <https://image.slidesharecdn.com/potsdam2015-lsu-160111213959/95/open-science-open-data-foster-training-potsdam-6-638.jpg>

You can make your workflow more open by ...



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO



 Bianca Kramer & Jeroen Bosman <https://101innovations.wordpress.com>

[DOI: 10.5281/zenodo.1147025](https://doi.org/10.5281/zenodo.1147025)