

Archiving and publication of research data

Dörte Bange
Research data management

UNIVERSITY LIBRARY



Universität Regensburg



What is research data?

No unique definition, but usually referred to as all the (digital) data collected/observed/created/... in the course of a research project by means of

- instrument measurements
- experimental observations
- survey results and interview transcripts
- simulations, models
- software
- ...

Research data management – why care about?

Some **funding organizations** require a data management plan when starting the project and sharing of the data at the end.

- European Commission – Horizon 2020:
Open Research Data Pilot
 - Certain research areas, including biotechnology – voluntary participation for other areas
 - Research data should be made accessible – “as open as possible, as closed as necessary”
 - Might become standard in future framework programmes...?
(see Open Access Pilot for publications in FP7)

Research data management – why care about?

Some **funding organizations** require a data management plan when starting the project and sharing of the data at the end.


- German Research Foundation (DFG):
Guidelines on the handling of research data, Sep 2015
 - Proposals should include considerations on data management planning.
 - Research data should be made accessible for re-use as soon as possible.
 - Good scientific practice: Archive your research data for at least 10 years.

Research data management – why care about?

When done properly from the beginning, data management can

- save you time in the end and
- minimize the risk of data loss.

CASH REWARD
for returning my lost backpack



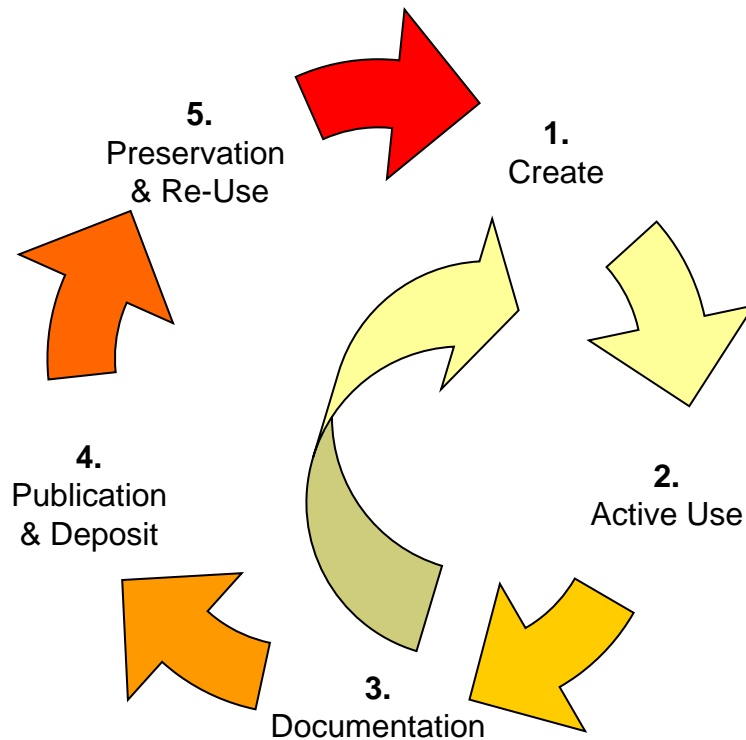
- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is VERY important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on: [REDACTED]

Thank you!!

What does research management mean?

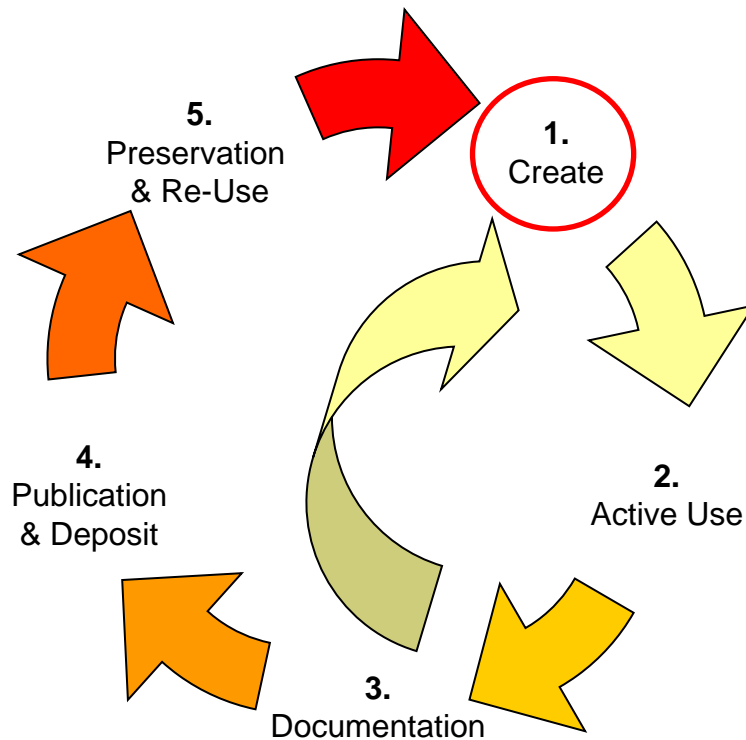


It means thinking about the steps of the “research data life cycle”.

Write your thoughts down
→ that’s your data management plan (DMP)!

- The DMP can and should be updated in the course of your project.
- Might be useful for structuring your research project/your PhD project.
- Templates and checklists

What does research management mean?



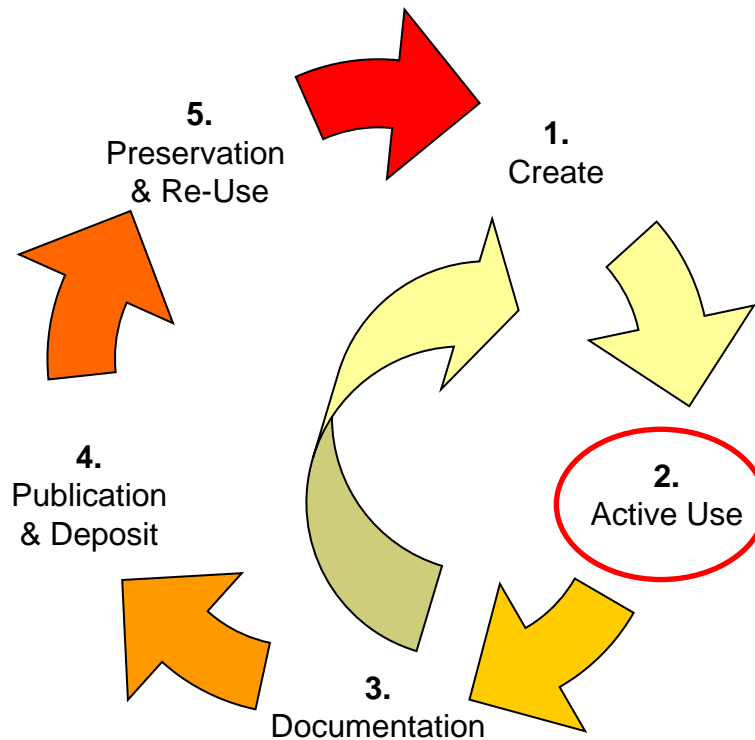
What type of data will you produce?

What types of file format?

How easy is it to create or reproduce?

What software will you use?

What does research management mean?



Is your data safe?

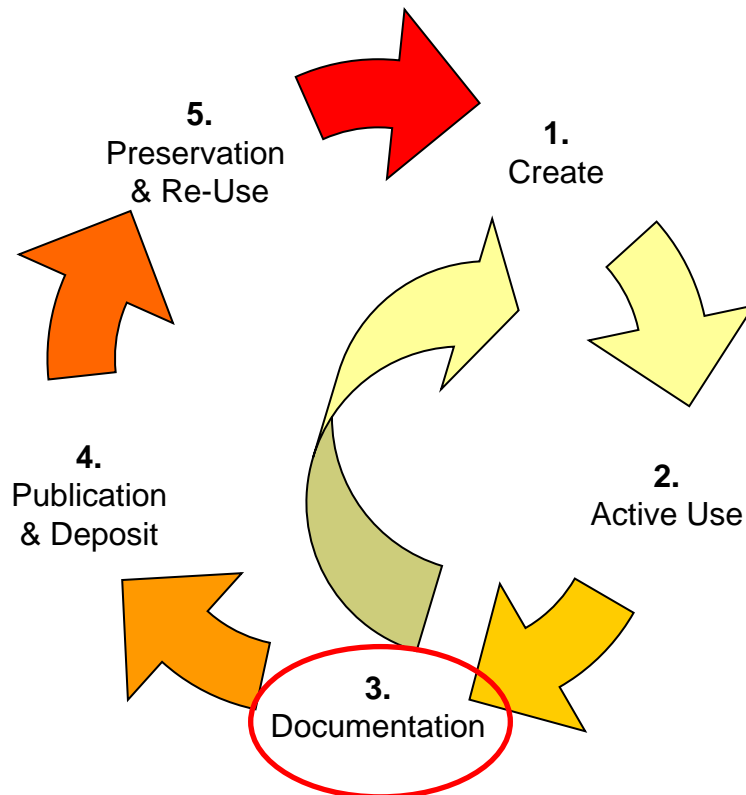
- Backup
- File encryption, if necessary

Is your data organised?

Can you find your data?

- Clear directory structure
- File version control
- File naming conventions

What does research management mean?



Do you still understand your older work?

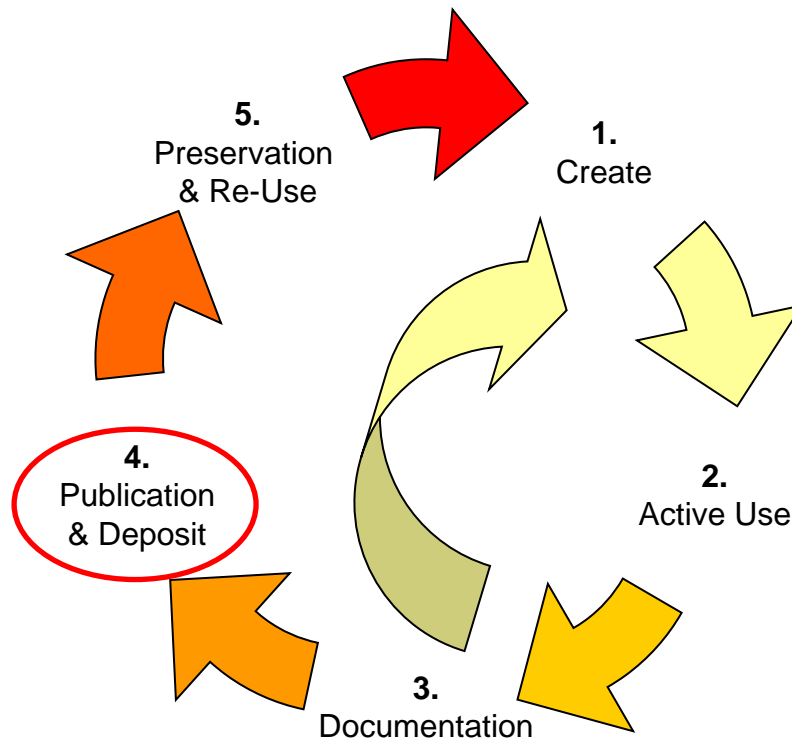
- Experimental protocols
- Lab notebooks
- Equipment/software used
- 'readme' files

Is the file structure/naming understandable to others?

Which data will be kept?

Which data can be discarded?

What does research management mean?



Are you *expected* to share your data?

Are you *allowed* to share your data?

Define the core data set of the project.

Which data will be included in your publication/thesis?

Share your data – requirements

Comply with funding requirements.

Some journals require sharing the data underlying the published results, e.g. PLOS journals:

- “PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.”*
- See also Barsh, Gregory S. et al. (2015): PLOS Genetics Data Sharing Policy: In Pursuit of Functional Utility.
DOI: [10.1371/journal.pgen.1005716](https://doi.org/10.1371/journal.pgen.1005716)

* PLOS Data Policy: <http://journals.plos.org/plosgenetics/s/data-availability> (including FAQs for Genetics Submissions)

Share your data – restrictions

Data cannot be made publicly available

- if it contains personal data / data which allows to potentially identify individuals and if there is no informed consent to public sharing, or
- if there are privacy requirements from commercial partners, or
- for other legal or ethical reasons.

→ Anonymization

→ Public release of aggregate information

→ Deposition in a controlled access repository

Share your data – where?

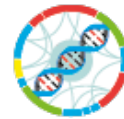
Data in supporting information files

Research data repository (*recommended*)

- Disciplinary repositories – search by
 - re3data.org (**registry of research data repositories**)
 - publisher's recommendations, e.g.
 - [PLOS Genetics](http://plos.org)
 - [Scientific Data](http://www.nature.com/scientificdata) (Nature Publishing Group)
- Institutional repositories – open to all research fields
 - University of Regensburg Publication Server
<http://epub.ur.de>

Share your data – where?

For microarray and sequencing data, there is a number of specialized repositories:



European Variation Archive



GenBank

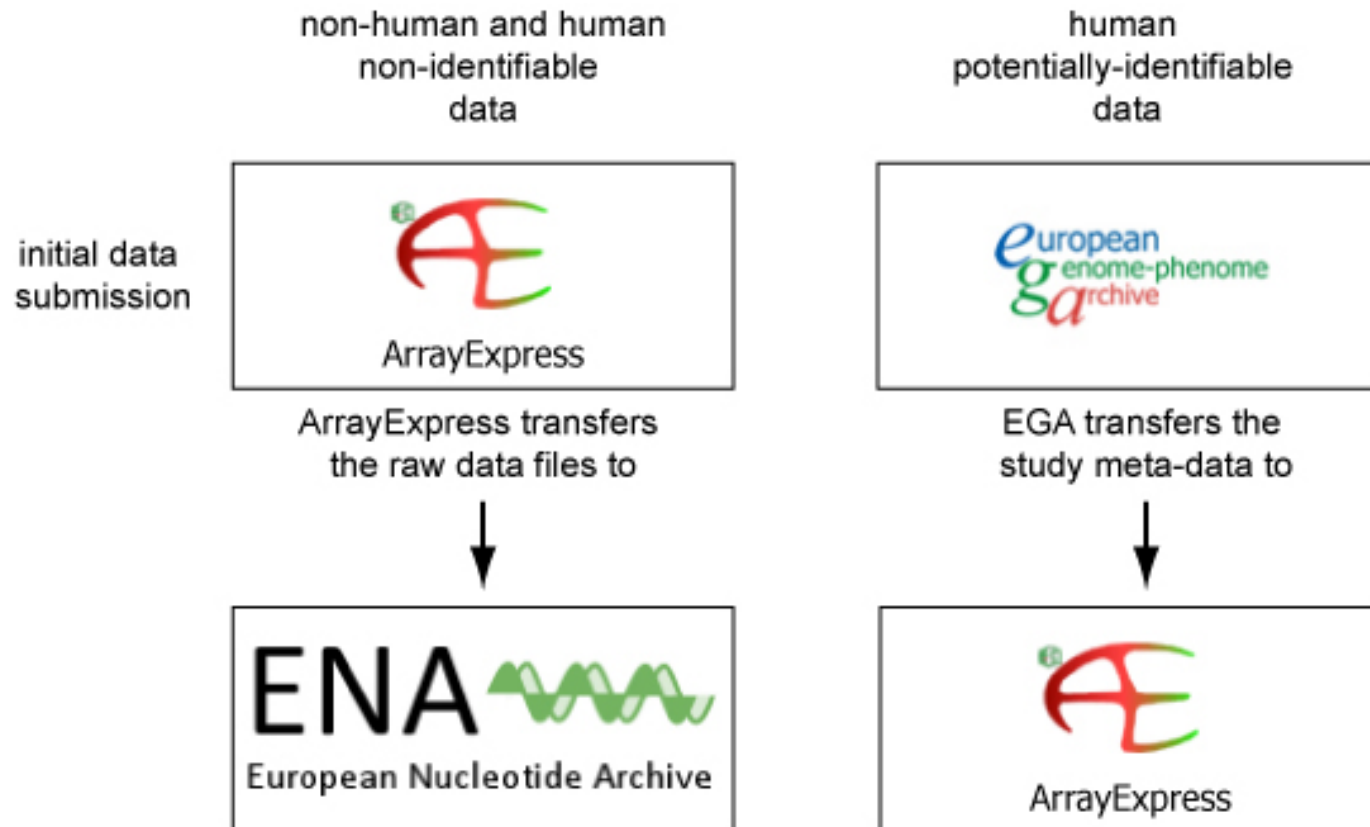
Genome RNAi



ArrayExpress

See also the European Bioinformatics Institute's services:
<http://www.ebi.ac.uk/services>

Share your data – where?



Share your data – where?

Unstructured data, associated analyses, experimental-control data, software scripts, etc. may be deposited on the **University of Regensburg Publication Server**. There you

- get a persistent identifier (DOI), thus
 - you can make your deposit part of a publication, and
 - you can get cited;

Share your data – where?

Frequency-resolved dynamic functional connectivity and scale-invariant connectivity-state behavior

DOI to cite this document: [10.5283/epub.32642](https://doi.org/10.5283/epub.32642)

Goldhacker, Markus (2015) *Frequency-resolved dynamic functional connectivity and scale-invariant connectivity-state behavior*. [Dataset]



License: [Creative Commons Attribution 4.0](#)

Plain Text – Additional Metadata

README.txt, in which the folder structure is explained

[Download \(3kB\)](#)



License: [Creative Commons Attribution 4.0](#)

ZIP Archive – Data

rs-fMRI data

[Download \(11GB\)](#)



License: [Creative Commons Attribution 4.0](#)

ZIP Archive – Data

simulated data

[Download \(13GB\)](#)

Abstract

Investigating temporal variability of functional connectivity is an emerging field in connectomics. Entering dynamic functional connectivity by applying sliding window

Share your data – where?

Unstructured data, associated analyses, experimental-control data, software scripts, etc. may be deposited on the **University of Regensburg Publication Server**. There you

- get a persistent identifier (DOI), thus
 - you can make your deposit part of a publication, and
 - you can get cited;
- can licence your data using Creative Commons licences, so others know they must give you credit;

Share your data – where?

Frequency-resolved dynamic functional connectivity and scale-invariant connectivity-state behavior

DOI to cite this document: [10.5283/epub.32642](https://doi.org/10.5283/epub.32642)

Goldhacker, Markus (2015) *Frequency-resolved dynamic functional connectivity and scale-invariant connectivity-state behavior*. [Dataset]



License: [Creative Commons Attribution 4.0](#)

Plain Text – ~~Additional Metadata~~

README.txt, in which the folder structure is explained

[Download \(3kB\)](#)



License: [Creative Commons Attribution 4.0](#)

ZIP Archive – Data

rs-fMRI data

[Download \(11GB\)](#)



License: [Creative Commons Attribution 4.0](#)

ZIP Archive – Data

simulated data

[Download \(13GB\)](#)

Abstract

Investigating temporal variability of functional connectivity is an emerging field in connectomics. Entering dynamic functional connectivity by applying sliding window


Share your data – where?

Unstructured data, associated analyses, experimental-control data, software scripts, etc. may be deposited on the **University of Regensburg Publication Server**. There you

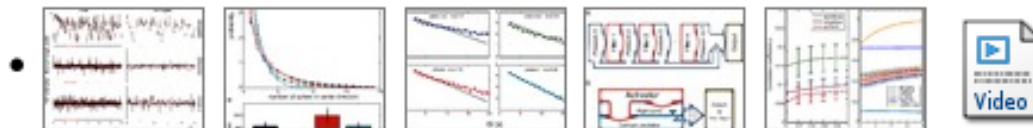
- get a persistent identifier (DOI), thus
 - you can make your deposit part of a publication, and
 - you can get cited;
- can licence your data using Creative Commons licences, so others know they must give you credit;
- can link your datasets to the publication based on them;

Share your data – where?


Linked items

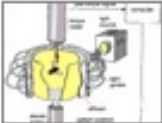
- 
 • Maye, A. and Hsieh, C. and Sugihara, G. and Brembs, Björn (2007)

Order in spontaneous behavior. PLoS ONE 2 (5), e443.



Maye, Alexander and Hsieh, Chih-hao and Sugihara, George and Brembs, Björn (2007) *Order in Spontaneous Behavior.* [Image] **[Currently Displayed]**

- 
 • Maye, Alexander and Hsieh, Chih-hao and Sugihara, George and Brembs, Björn (2007) *Alternative models conceptualizing the open-loop experiment.* [Image]

- 
 • Maye, Alexander and Hsieh, Chih-hao and Sugihara, George and Brembs, Björn (2007) *Flight simulator set-up.* [Image]



Share your data – where?

Unstructured data, associated analyses, experimental-control data, software scripts, etc. may be deposited on the **University of Regensburg Publication Server**. There you

- get a persistent identifier (DOI), thus
 - you can make your deposit part of a publication, and
 - you can get cited;
- can licence your data using Creative Commons licences, so others know they must give you credit;
- can link your datasets to the publication based on them;
- can set embargo periods or control access for some files.

Share your data – where?

Data archive of JETP Letters 97, 41–44

DOI to cite this document: [10.5283/epub.31490](https://doi.org/10.5283/epub.31490)

Kvon, Z. D. and Kozlov, D. A. and Danilov, Sergey and Zoth, Christina and Vierling, P. and Stachel, Sebastian and Belkov, Vassilij and Bakarov, A. K. and Dmitriev, D. V. and Toropov, A. I. and Ganichev, Sergey (2015) *Data archive of JETP Letters 97, 41–44*. [Dataset]




Other – Data

Data for figures in Origin format

[Download \(498kB\)](#) – Repository staff only

Request a copy

Linked items

- [Kvon, Z. D. and Kozlov, D. A. and Danilov, Sergey and Zoth, Christina and Vierling, P. and Stachel, Sebastian and Belkov, Vassilij and Bakarov, A. K. and Dmitriev, D. V. and Toropov, A. I. and Ganichev, Sergey \(2013\) *Terahertz Radiation-Induced Magnetoresistance Oscillations of a High-Density and High-Mobility Two-Dimensional Electron Gas*. JETP Letters 97 \(1\), pp. 41–44.](#)
-  Kvon, Z. D. and Kozlov, D. A. and Danilov, Sergey and Zoth, Christina and Vierling, P. and Stachel, Sebastian and Belkov, Vassilij and Bakarov, A. K. and Dmitriev, D. V. and Toropov, A. I. and Ganichev, Sergey (2015) *Data archive of JETP Letters 97, 41–44*. [Dataset] **[Currently Displayed]**

Share your data – how?

Your data can only be found and re-used if it contains sufficient **metadata**.

- Descriptive metadata: title, description, keywords, author, ...
- Administrative metadata: rights management, file formats, ...

Use discipline-specific standards, such as

- MIAME (Minimum Information About a Microarray Experiment)
- MINSEQE (Minimum Information about a high-throughput SEQuencing Experiment)
- Gene Ontology

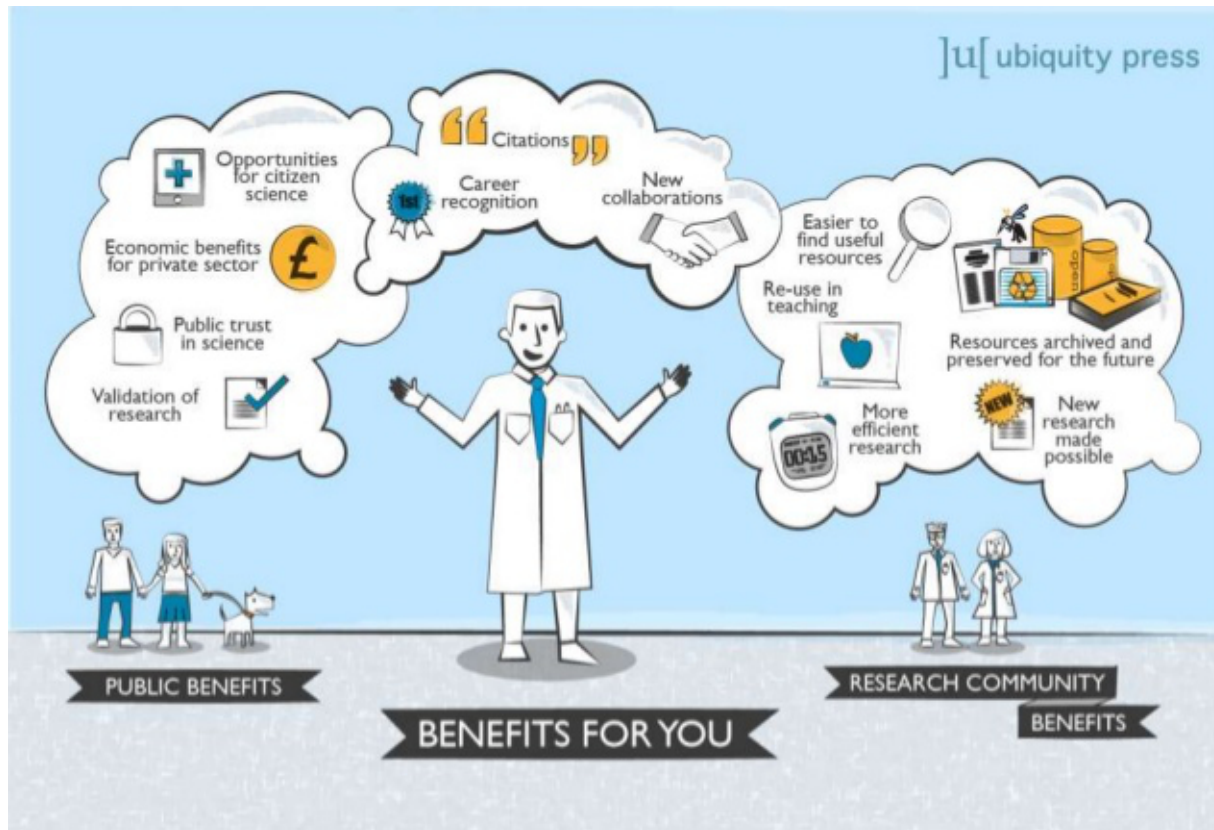
Share your data – how?

Publish your data under a free **licence**, e.g. a Creative Commons licence, which allows re-use of the data.

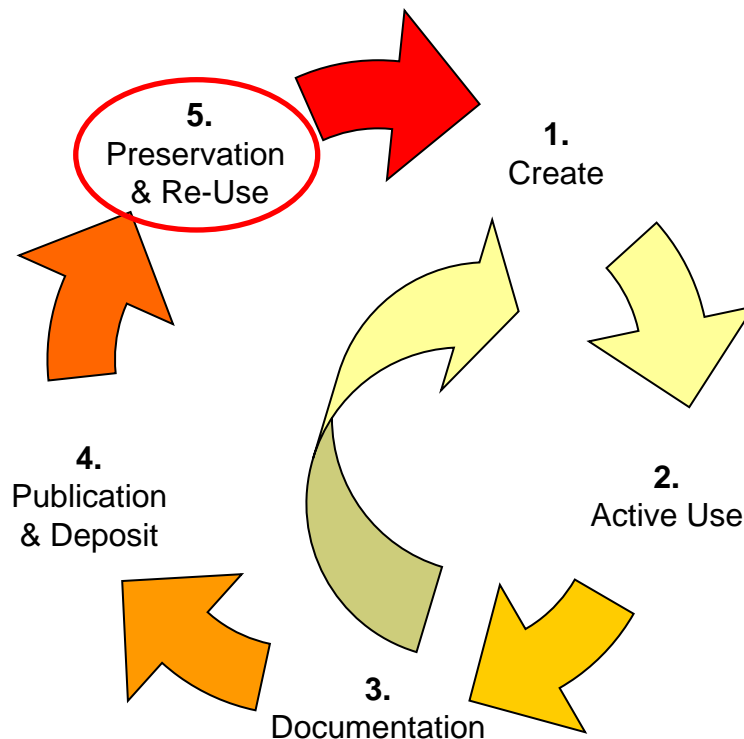
For research data, we recommend a Creative Commons Attribution licence (CC BY). This helps others finding data they can use, and makes sure you get cited.



Share your data – why?



What does research management mean?



Which data do you need to keep?

- All versions? Just final version?
- Should discarded data be destroyed?

What are the re-processing costs?

- Keep only software and protocol/methodology information.

Are there tools/software needed to process/visualise the data?

- Archive these with your data.



Universität Regensburg

Dörte Bange
Research data management
UNIVERSITY LIBRARY

Research data team

Dörte Bange
Central library, ZB 453
Phone: 1645

Dr. Gernot Deinzer
Mathematics building, M 202
Phone: 2759

E-mail: daten@ur.de

Thank you for your attention!

Acknowledgements

In parts based on

1. Pink, Catherine/Cope, Jez (2012): Managing your research data. University of Bath Research Data Management training for researchers. Online available at <http://de.slideshare.net/jezcope/university-of-bath-research-data-management-training-for-researchers> (accessed 2016-03-01). Licenced under CC BY 4.0
2. Hole, Brian (2015): Preparing Data for (Open) Publication. Online available at <http://de.slideshare.net/brianhole/preparing-data-for-open-publication> (accessed 2016-03-03). Licenced under CC BY 4.0