Data Publishing

Carolin Odebrecht
Humboldt-Universität zu Berlin, Computer and Media Service
https://orcid.org/0000-0003-4887-7701

Anja Herwig
Humboldt-Universität zu Berlin, University Library
https://orcid.org/0000-0003-1703-3979

Welcome!

Presentation slides and materials:
Link to HU-Box will be shared in the chat and afterwards again via e-mail

Questions:

Either in the chat or via the blue hand below the list of participants

Website Research Data Management:

https://hu.berlin/dataman

Who are you?

What are your expectations for this workshop?

Your way to data publication

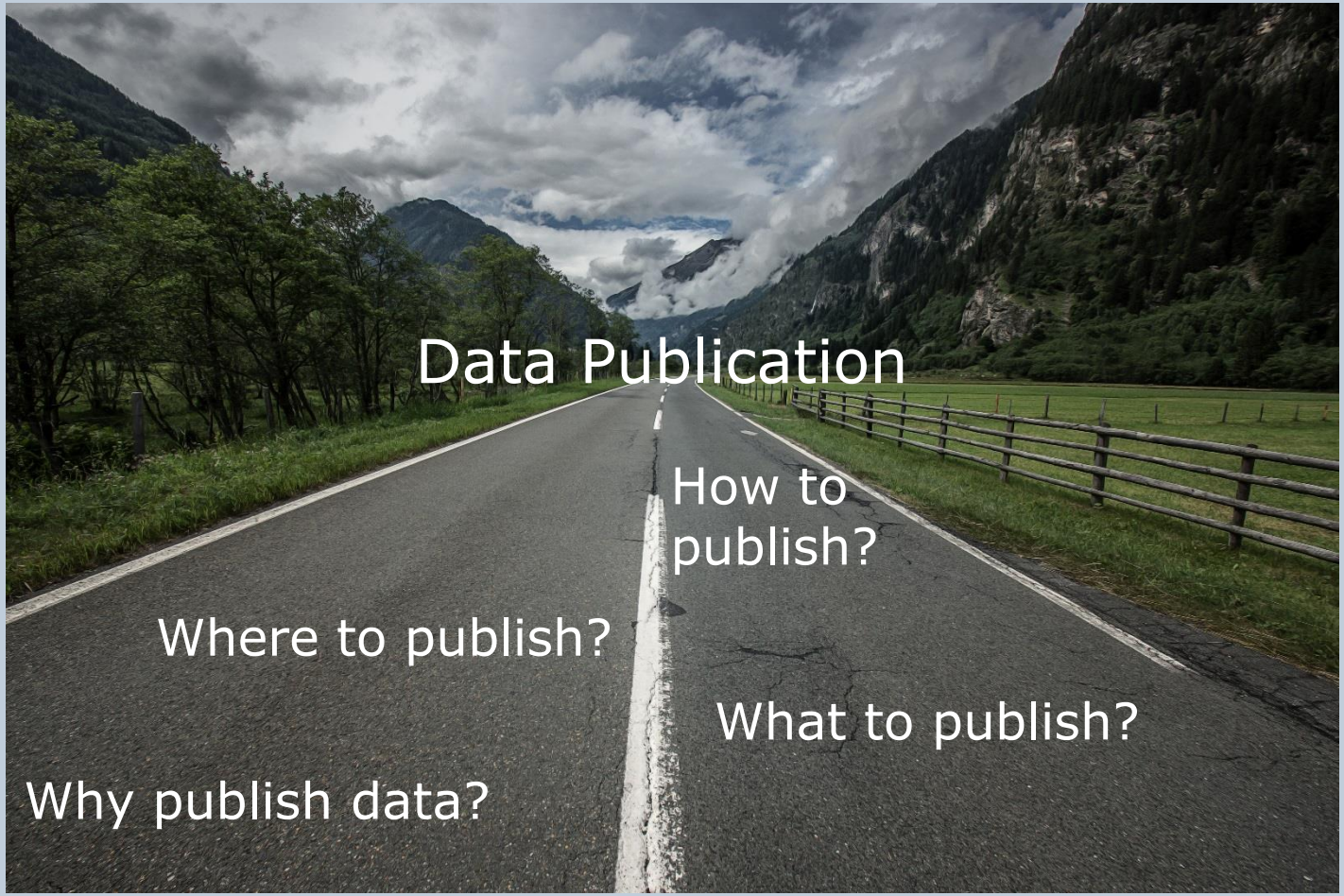


Photo: CC0

Publisher policies – example Wiley

„We encourage authors of articles published in our journals to share their research data including, but not limited to: raw data, processed data, software, algorithms, protocols, methods, materials."

Four standardized data sharing policies:

1. Encourages data sharing
2. Expects data sharing
3. Mandates data sharing
4. Mandates data sharing and peer reviews data

Source: https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html

Research funders

European Commission Horizon 2020

German Research Foundation

Wellcome Trust

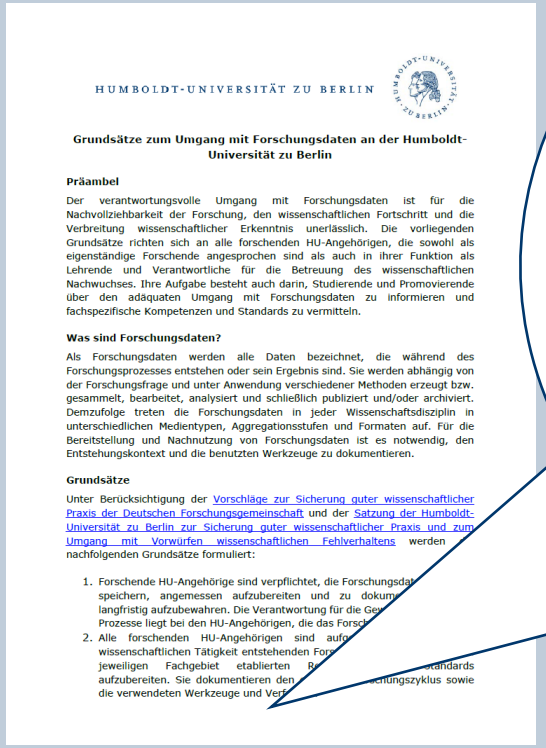
National Science Foundation

Federal Ministry of Education and Research

Volkswagen Foundation

Research Data Policy

Policy



HUMBOLDT-UNIVERSITÄT ZU BERLIN

Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin

Präambel

Der verantwortungsvolle Umgang mit Forschungsdaten ist für die Nachvollziehbarkeit der Forschung, den wissenschaftlichen Fortschritt und die Verbreitung wissenschaftlicher Erkenntnis unerlässlich. Die vorliegenden Grundsätze richten sich an alle forschenden HU-Angehörigen, die sowohl als eigenständige Forschende angesprochen sind als auch in ihrer Funktion als Lehrende und Verantwortliche für die Betreuung des wissenschaftlichen Nachwuchses. Ihre Aufgabe besteht auch darin, Studierende und Promovierende über den adäquaten Umgang mit Forschungsdaten zu informieren und fachspezifische Kompetenzen und Standards zu vermitteln.

Was sind Forschungsdaten?

Als Forschungsdaten werden alle Daten bezeichnet, die während des Forschungsprozesses entstehen oder sein Ergebnis sind. Sie werden abhängig von der Forschungsfrage und unter Anwendung verschiedener Methoden erzeugt bzw. gesammelt, bearbeitet, analysiert und schließlich publiziert und/oder archiviert. Demzufolge treten die Forschungsdaten in jeder Wissenschaftsdisziplin in unterschiedlichen Medientypen, Aggregationsstufen und Formaten auf. Für die Bereitstellung und Nachnutzung von Forschungsdaten ist es notwendig, den Entstehungskontext und die benutzten Werkzeuge zu dokumentieren.

Grundsätze

Unter Berücksichtigung der Vorschläge zur Sicherung guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft und der Satzung der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens werden nachfolgenden Grundsätze formuliert:

1. Forschende HU-Angehörige sind verpflichtet, die Forschungsdat
   speichern, angemessen aufzubereiten und zu dokum
   langfristig aufzubewahren. Die Verantwortung für die Ge
   Prozesse liegt bei den HU-Angehörigen, die das Forsch
2. Alle forschenden HU-Angehörigen sind aufg
   wissenschaftlichen Tätigkeit entstehenden For
   jeweiligen Fachgebiet etablierten R
   aufzubereiten. Sie dokumentieren den
   die verwendeten Werkzeuge und Verf

„HU researchers should take responsibility for deciding at what time and on what legal terms research data may be accessed."

"Research data underlying scholarly publications should be archived for the long-term and/or published in an appropriate trustworthy data archive or repository."

https://hu.berlin/RDM-policy

Your way to data publication

Photo: CC0

Data selection – some criteria

- ✓Data underlying a publication
- ✓Data necessary to understand methodological approach (verification)
- ✓Milestone version
- ✓Uniqueness/of interest to others
- ✓Costs
- ✓Rights
- ✓Data quality
- ✓Documentation
- ✓Technical maintenance

Exercise: Which data will I publish?

Think about your research data. Which files would you like to publish?

- Write down which data/files you would like to publish.
- Also think about what kind of data would be needed for your own dissertation to be understood properly.

Time: 2 minutes

Rights

- Data protection and personal rights
  → Anonymization or pseudonomization
  → Informed consent (analysis + data publication)
  → Support from data protection officers
  → Information at
  https://www.forschungsdaten-bildung.de

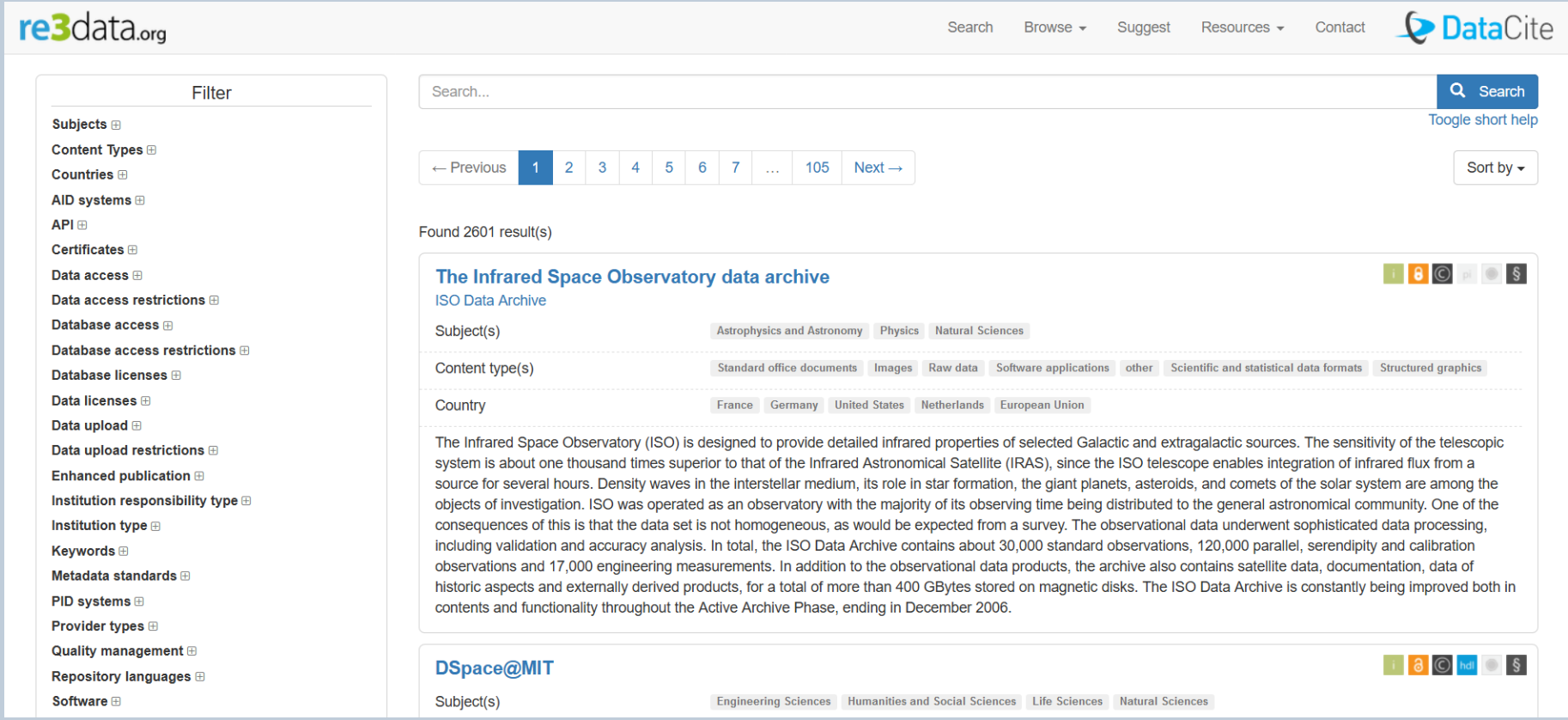
- Copyright law
  → Licenses
  → Contracts

- Exploitation rights from publisher or university

Your way to data publication

Photo: CC0

Finding a data publisher



https://www.re3data.org

How to choose a data repository?

- Does the repository have a certificate (e. g. Data Seal of Approval)?
- Are persistent identifiers assigned (e. g. DOI, handle)?
- How is access to the data (open, restricted, inaccessible)?
- Are the terms of use and license of the data mentioned by the repository?

If you do not find an appropriate data repository…

Generic repositories:

- Zenodo, https://zenodo.org
- Figshare, https://figshare.com
- DRYAD (life sciences), http://datadryad.org

Institutional repositories:

- edoc publication server (HU Berlin), https://edoc.hu-berlin.de
- DepositOnce (TU Berlin), https://depositonce.tu-berlin.de
- Refubium (FU Berlin & Charité), https://refubium.fu-berlin.de

Data Journals

A data paper documents and describes research data to facilitate dissemination and re-use. It informs about data collection, features and potential reuse.

Examples:

- Scientific data
  http://www.nature.com/sdata
- Data in Brief
  http://www.journals.elsevier.com/data-in-brief
- Data
  http://www.mdpi.com/journal/data

Your way to data publication

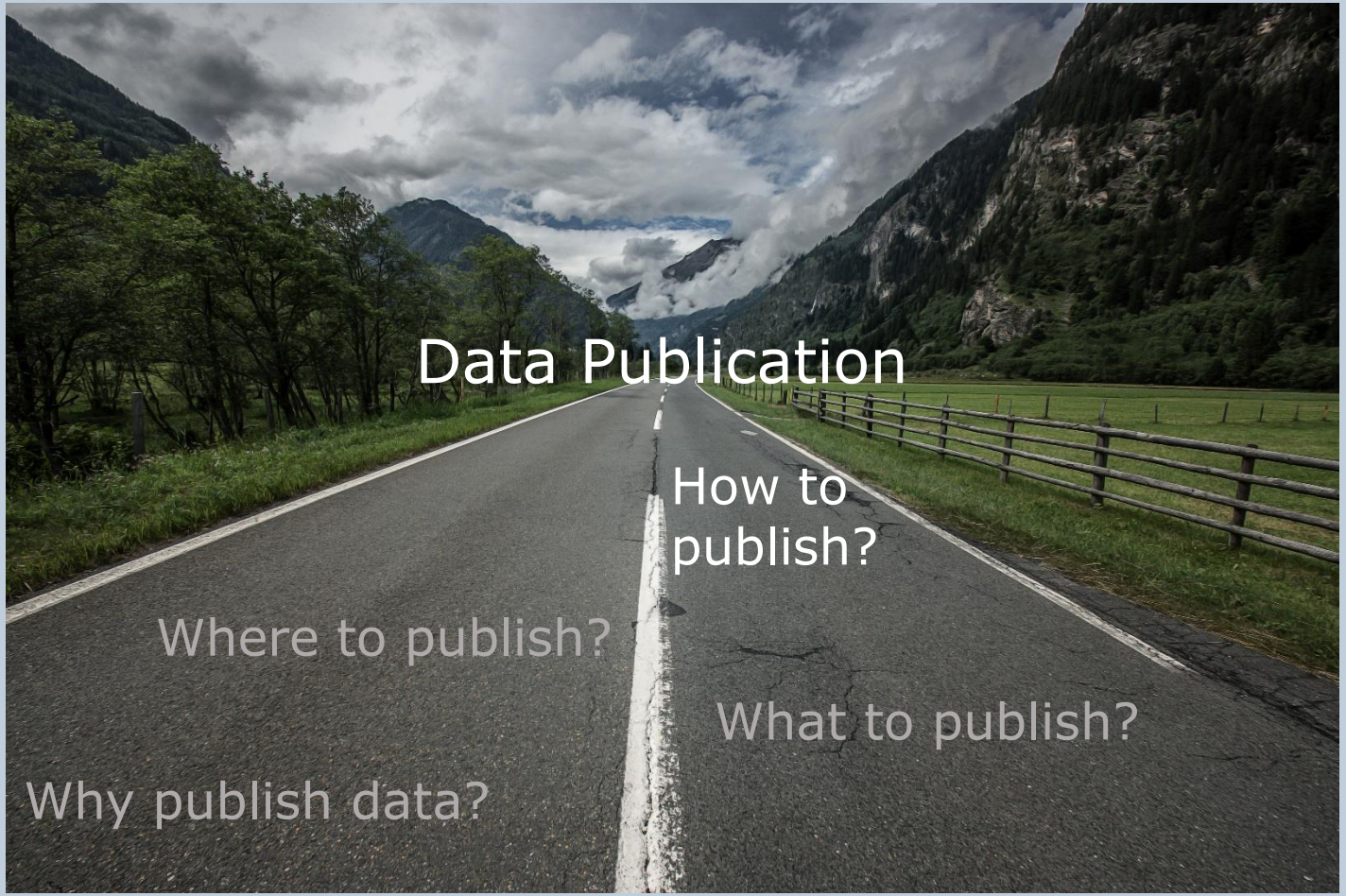


Photo: CC0

Metadata

- Serve primarily to **find** the data (e.g. principal investigator, time, location)
- Many disciplines have their own standards
- A discipline-specific **overview** of metadata standards can be found at: http://rd-alliance.github.io/metadata-directory/subjects/
- Help: Subject librarians of the University Library

Metadata

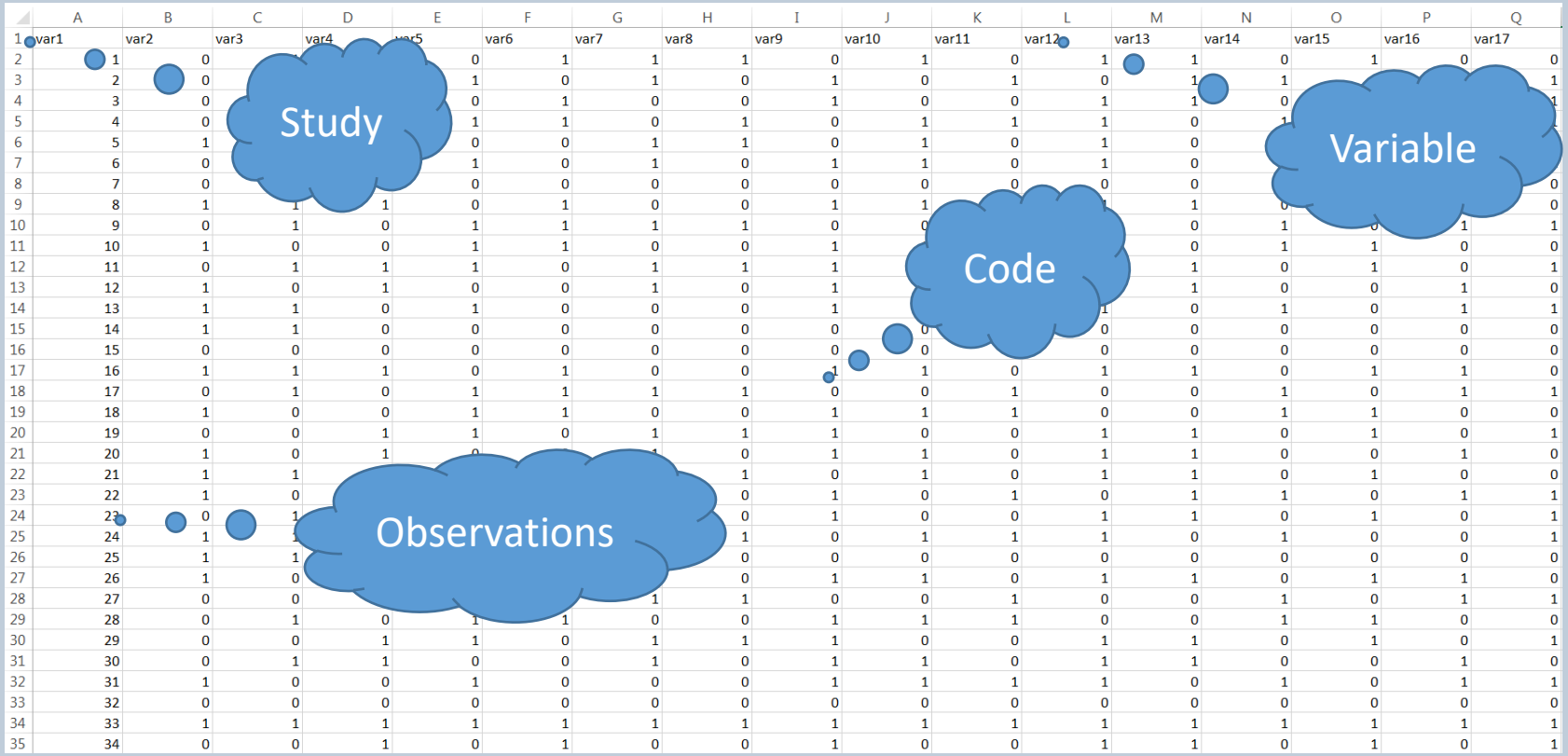
- Title
- Author/Primary researcher
- Identifier
- Keywords/Topic
- Dates
- Funder
- Language
- File format
- Unit
- Method
- Sources
- Place
- Rights
- File name(s) and relation to other files

Assigning keywords

- Thesauri and classifications are documentation languages used to describe the content of objects (e. g. research data)
- This makes it easier to find the data
- There are already specialized classifications and thesauri for many disciplines

Overview: Basel Register of Thesauri, Ontologies & Classifications
http://www.bartoc.org

Why an additional documentation?



Readme

- Documents the research process and the data
- Includes:
  - Abstract
  - Hypotheses
  - Information on the collection of data (methods, units, time, place, devices)
  - Measures for data cleansing (deletion of outliers, weighting)
  - Structure of the data and its relationships to each other
  - Explanation of variables, labels and codes
  - Differences between different versions
  - Information on access and terms of use

Data preparation

- Do you have all the data necessary?
- Is it the right version?
- Are file names and properties of files logic?
- Are file formats appropriate (depends on repository)?
- Do you have all necessary metadata?
- Have you prepared a documentation in form of a readme?

Licenses for research data

Open licenses

- Creative Commons
  CC0, CC BY, CC BY-SA
- Open Data Commons
  Public Domain Dedication and License (PDDL), Open Database License (Odbl), Attribution License (ODC-By)
- GNU Free Documentation License (GFDL)
- Free Digital Peer Publishing License (f-DPPL)

Restrictive licenses

- Creative Commons
  CC BY-NC, CC BY-ND,
  CC BY-NC-ND, CC BY-NC-SA
- Digital Peer Publishing License (DPPL), Modular Digital Peer Publishing License (m-DPPL)

Information about software licenses:
https://opensource.org/licenses

Creative Commons

Attribution

ShareAlike

NonCommercial

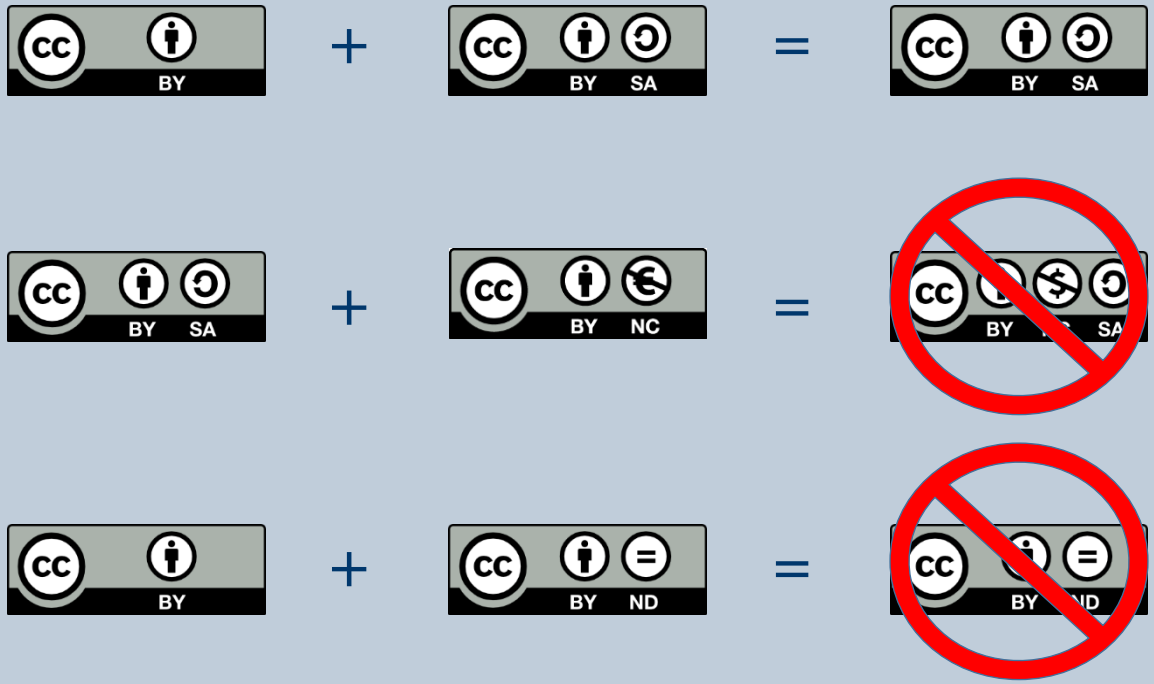
NoDerivatives

Public Domain (CC0)

No known copyright

Source: https://creativecommons.org/share-your-work/licensing-types-examples/

Activity: License matching



Source: https://creativecommons.org/licenses/

Why do you need persistent identifiers?

Researcher
(ORCID, author ID)

is provider of

is author of

belongs to

Research data
(e.g. DOI, handle)

Publication
(e.g. DOI, URN)

10 things to know about ORCID

1. Stands for Open Researcher and Contributor ID
2. (alpha-)numeric 16-digit code
3. Unambiguous scientific identity (beyond name change or typing errors)
4. Used by journals, research funders and institutions as an authoritative file
5. Is maintained by the researcher
6. Lasts longer than an email address
7. ORCID creation takes about 30 seconds
8. Operated by non-profit initiative
9. Continuous growth (February 14, 2022: 13.397.934 ORCIDs)
10. Connection to Web of Science, zenodo, DataCite and others

Questions?

Data publishing with edoc

https://edoc.hu-berlin.de

Let’s make a test data publication!

Photograph by
Anja Herwig

Data citation (FORCE11 recommendation)

Author(s) (Publication year): Title of research data. Data repository or archive. Version. Global persistent identifier (preferably as a link)

Example:

Jane Doe, John Doe (2015): *Successfully citing research data*. Humboldt-Universität zu Berlin. Version 1.0. http://doi.org/10.17172/this-is-an-example-so-do-not-click

Please note:
Citation practices can vary between subject areas and publishers!

For software please add [software: source code] to the title.

Questions?

Feedback

Rate this workshop and its content!

Please answer the questions in the Zoom survey.

Many thanks for your interest!

Carolin Odebrecht, research data management coordinator

Anja Herwig, data librarian

forschungsdaten@cms.hu-berlin.de

Further information on research data management:
https://hu.berlin/dataman

Follow us on Twitter @DatawomenHUB