



InnoRenew CoE

Livade 6, 6310 Izola/Isola, Slovenia, T: +386 40 282 944, E: coe@innorenew.eu, www.innorenew.eu



Research Data Management (RDM)

Dr. Ana Slavec, InnoRenew CoE

<https://zenodo.org/record/4727133>

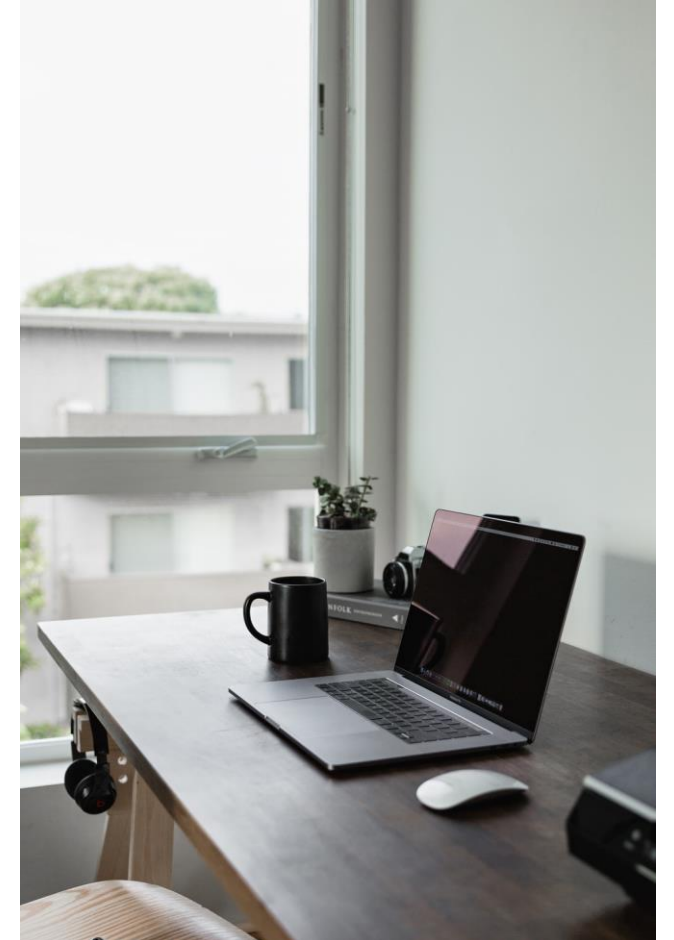


Ice-breaking questions

1. What type of digital content do you generate in your research?
2. What formats do you use to save your data?
3. Where do you store your data?

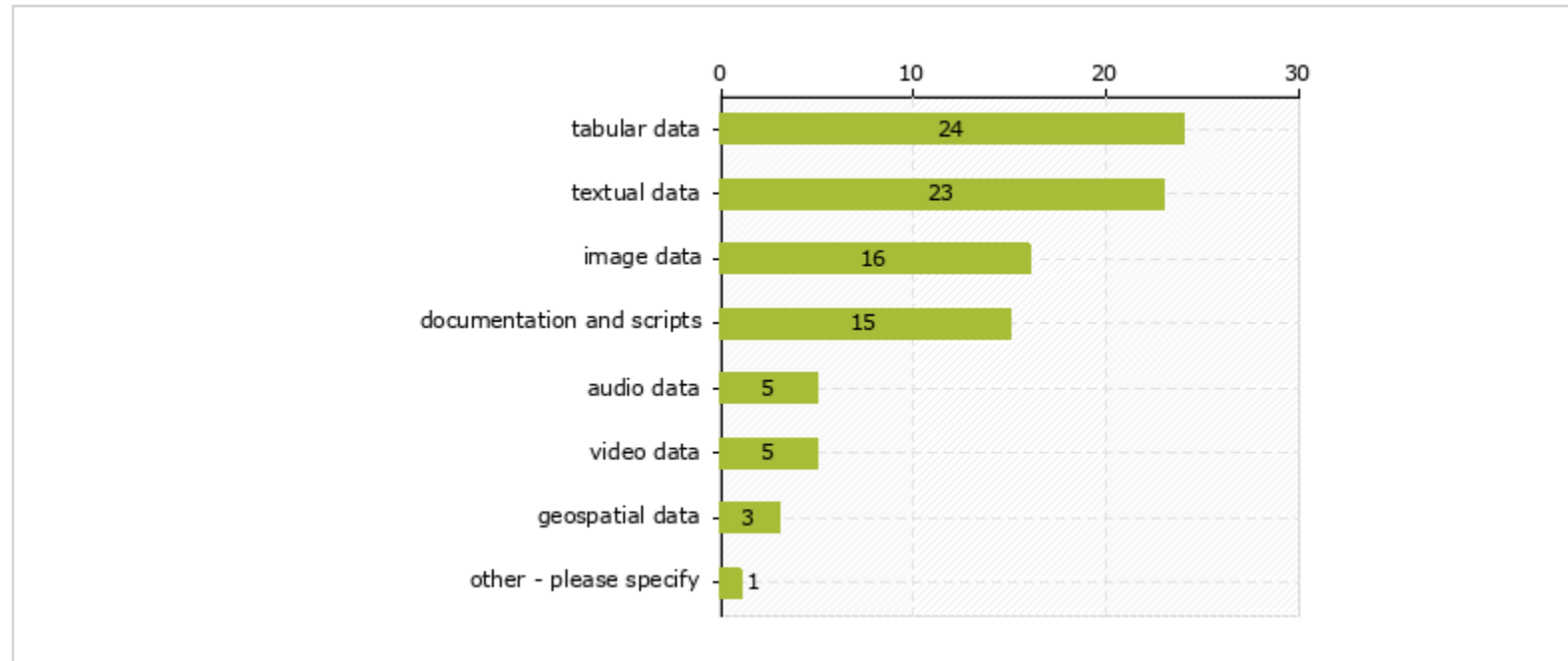
<https://www.1ka.si/rdm>

>>Results

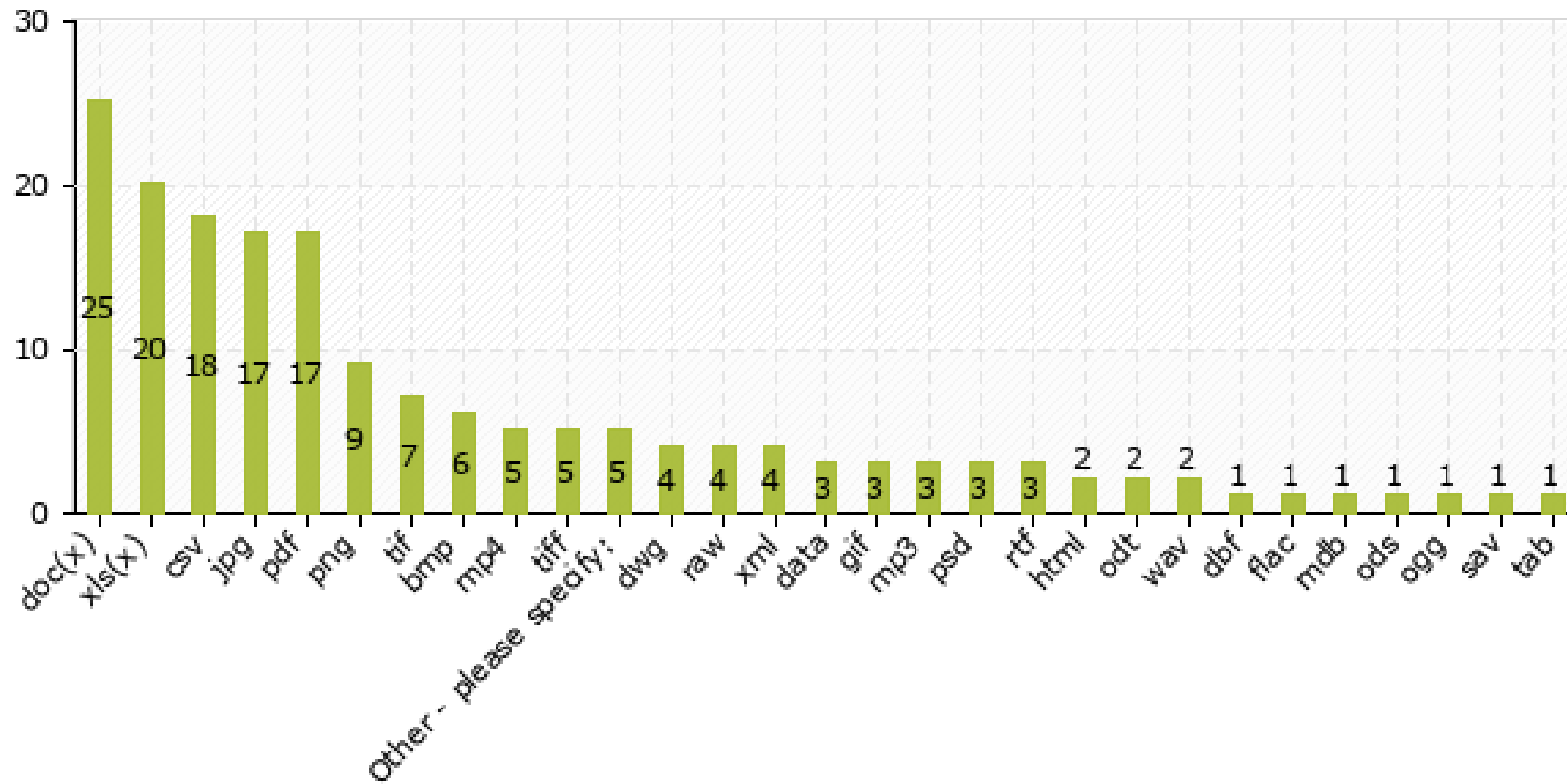




What type of digital content do you generate in your research? (2021, n=28)

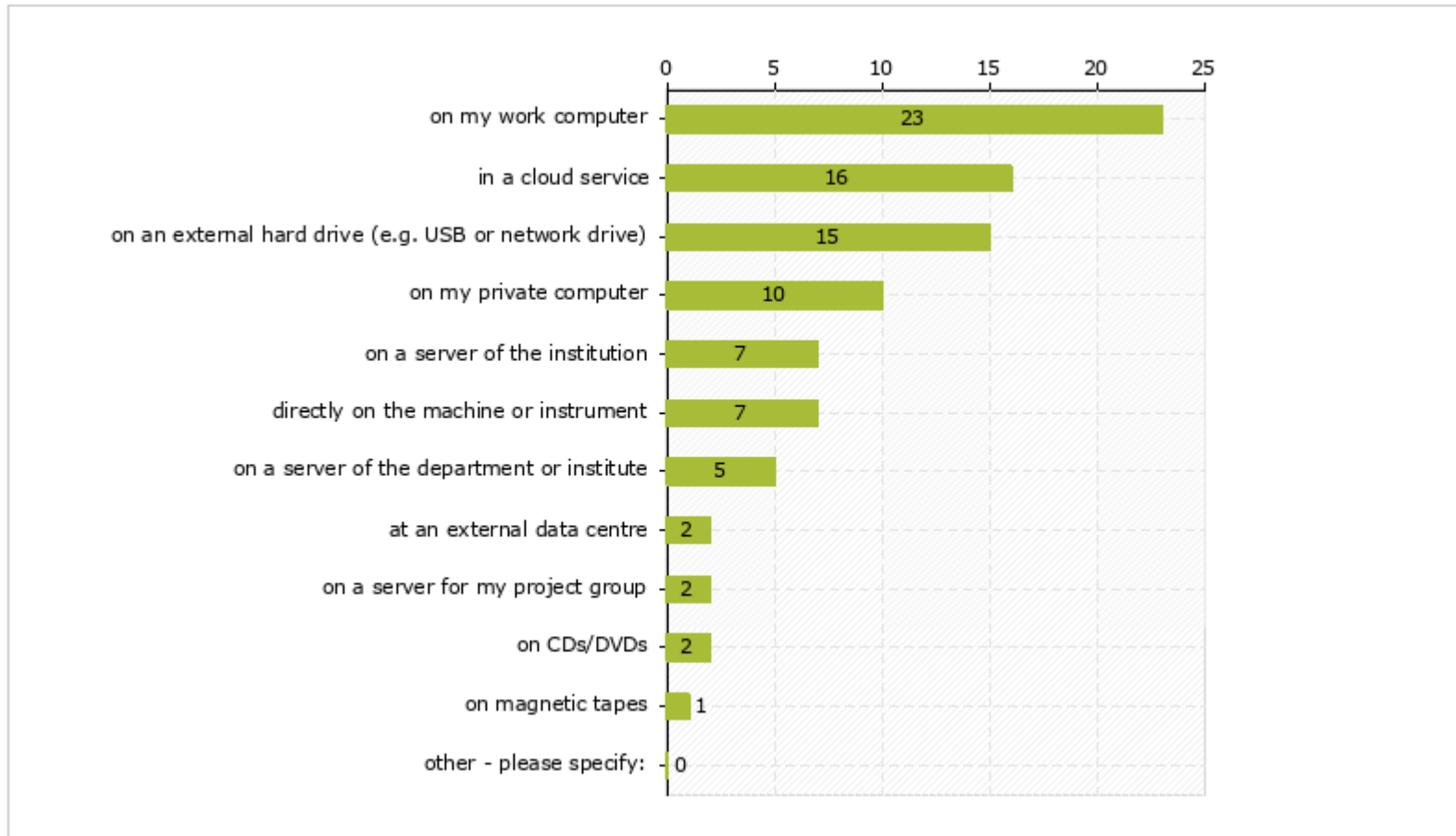


What formats you use to save your data? (2021, n=28)





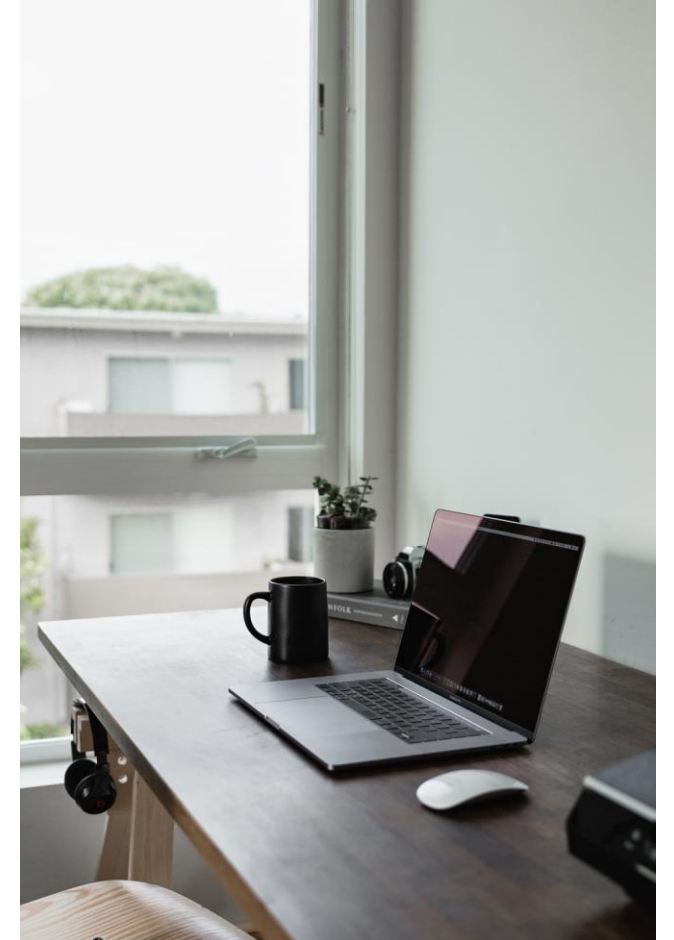
Where do you store your research data? (2021, n=28)





Outline

- What is RDM and why should that interest us?
- What are the RDM challenges in engineering?
- How to prepare a Research Data management plan?
- What are the recommended file formats?
- What are the FAIR principles?
- How to select the most appropriate repository?
- Where can I learn more about RDM?



Why should this interest you?

1. Ethics:

RDM is part of the responsible conduct of research, i.e. the practice of scientific investigation with integrity. Well-managed and accessible data allows others to validate and replicate findings.

2. Requirements imposed by funders and publishers:

Open Science is one of the pillars of the Horizon Europe, the new EU framework programme for research and innovation

3. Saves times and resources (in the long run):

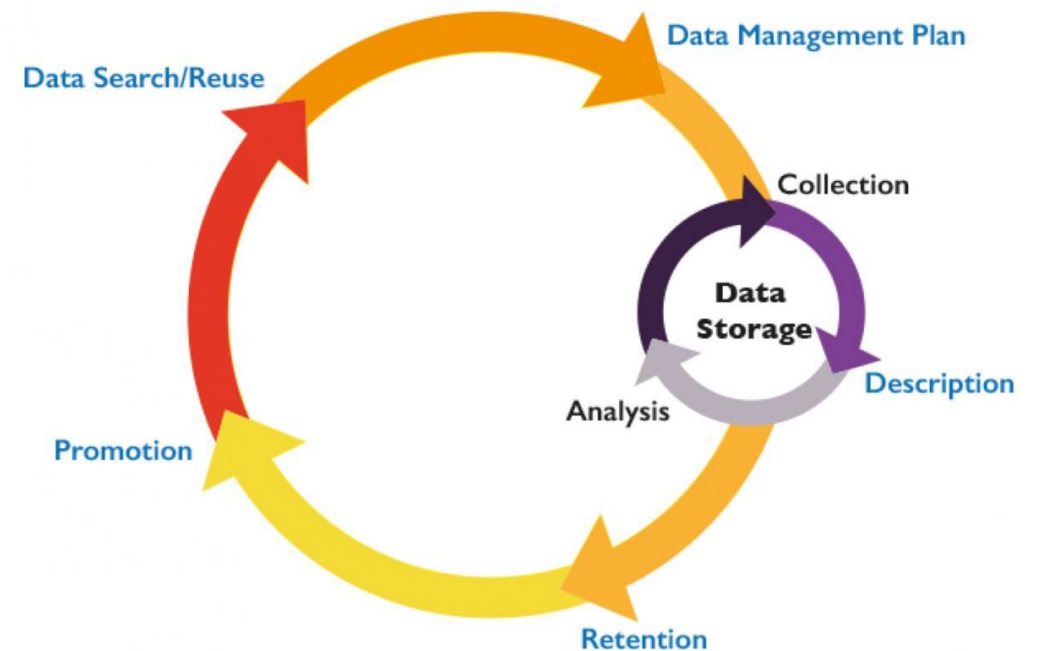
You are the most likely re-user of your own data.



What is Research Data Management?

The Research Data Management Lifecycle

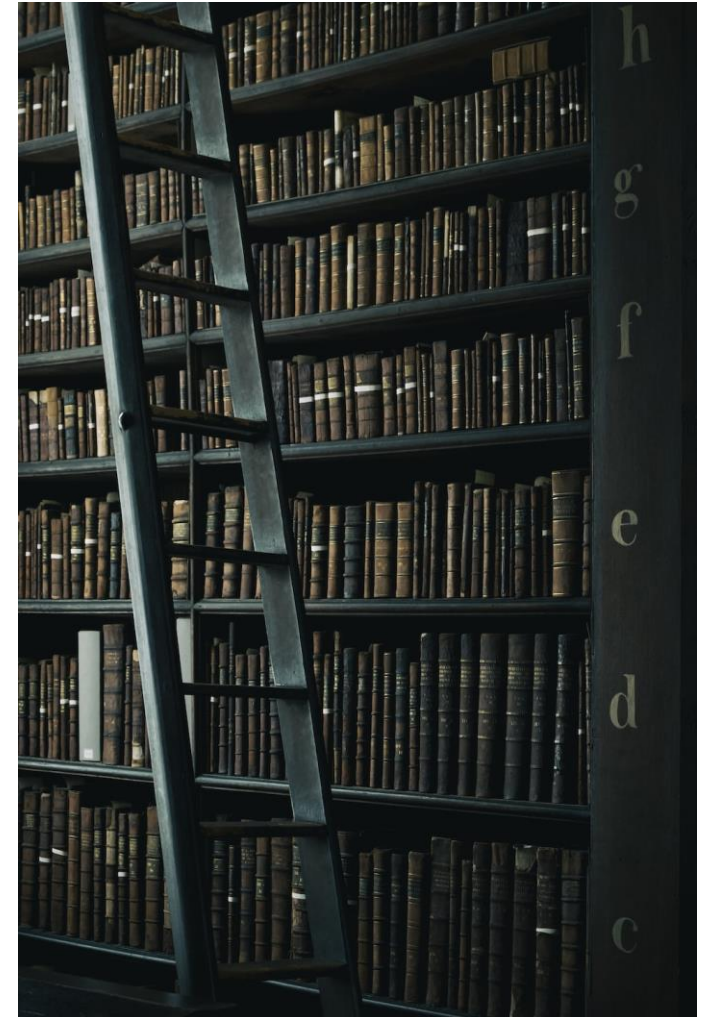
- RDM describes the organisation, storage, preservation, and sharing of data collected and used in a research project.
- It involves decision about how data will be preserved and shared after the project is completed.



Source: University of California, Santa Cruz, [Data Management LibGuide](#)

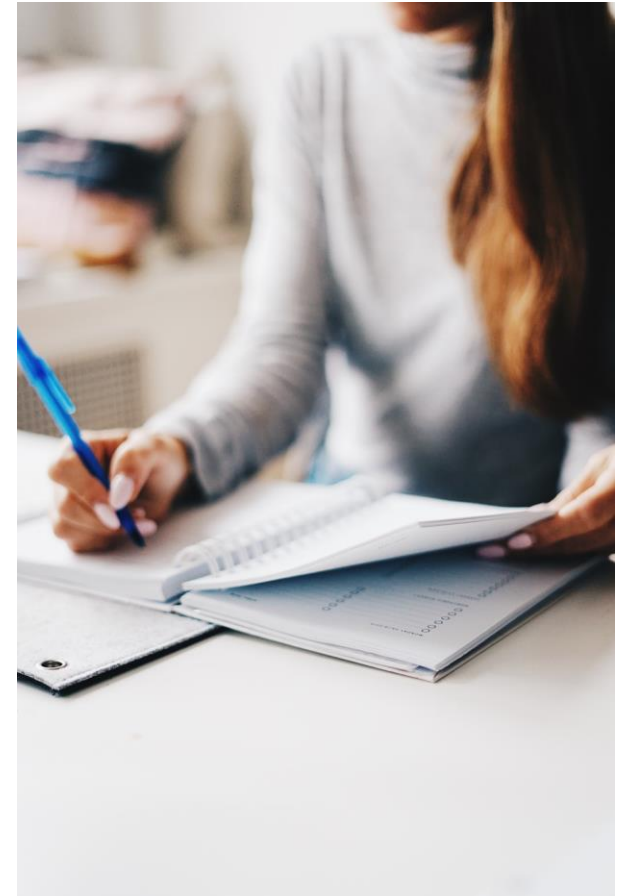
Data search/reuse

- Official statistics (national statistical offices, Eurostat...)
- International NGOs (United nations, World bank...)
- Research data repositories (general purpose, institutional, domain specific)
- Google dataset search
- ...



Data management plan (DMP)

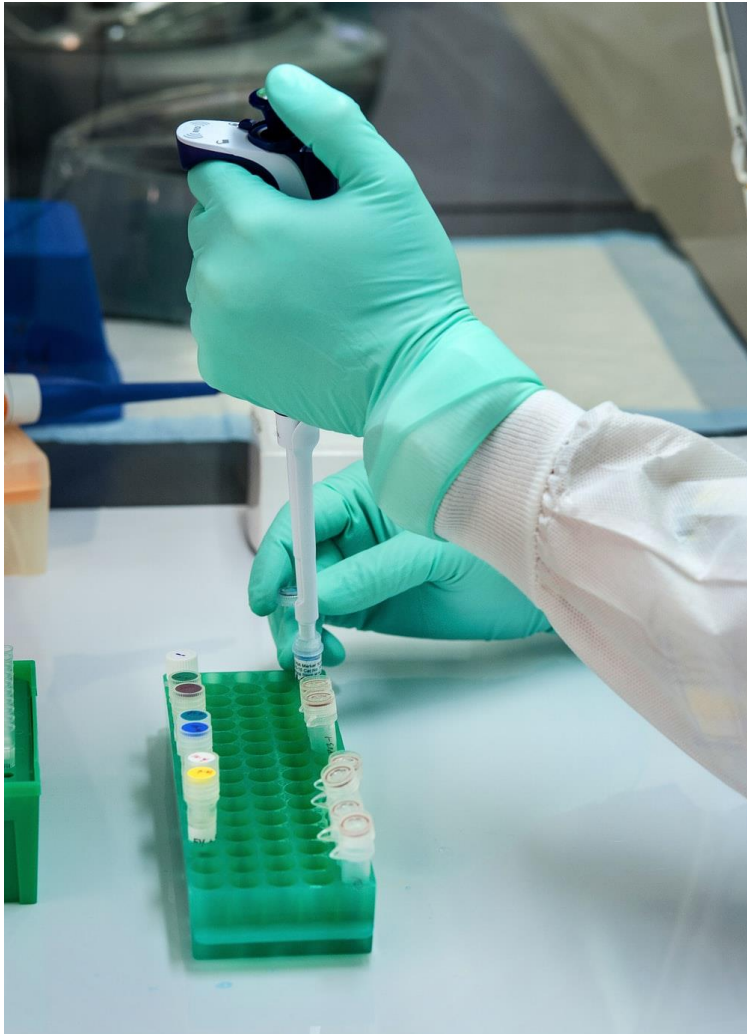
- A formal document that outlines how data are to be handled during a research project, and after the project is completed
- What should a DMP include?
 - Description of data to be collected/created
 - Standards/methodologies for data collection and management
 - Ethics and intellectual property right
 - Plans for data sharing and access
 - Strategy for long-term preservation
- Useful resources: [Digital Curation Centre](#)



DMP template – questions to respond

1. What data will you collect or created? How?
2. What documentation and metadata will accompany the data?
3. How will you manage any ethical and legal issues?
4. How will the data be stored and backed up during research?
5. How will you manage access and security?
6. How will you share the data? Are any restrictions required?
7. Who will be responsible for data management? What resources will you require to deliver your plan?





1. Data collection

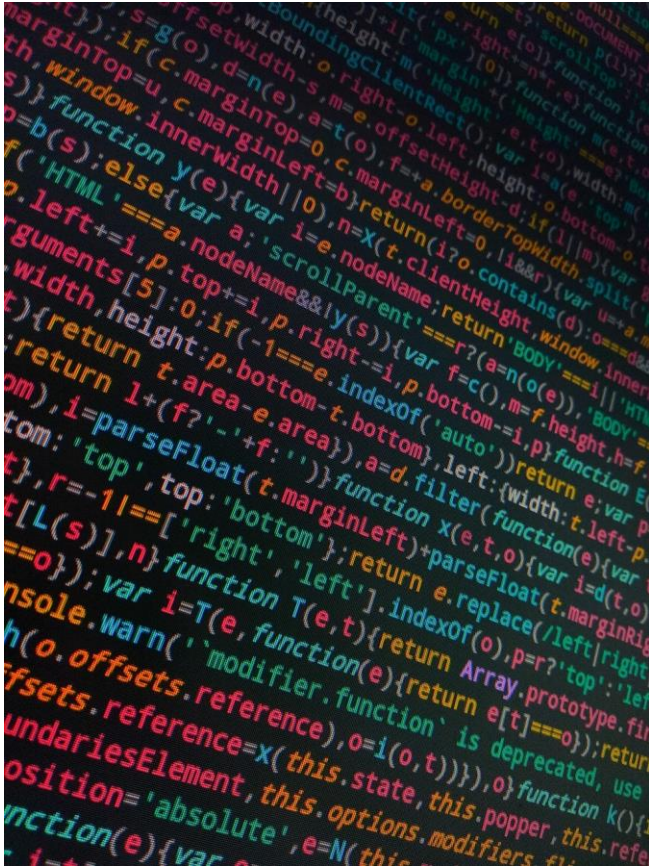
- Type, format and volume of data
 - Formats and software
- Reuse of existing data (secondary data)
 - Standards and methodologies
 - Structure and name of files
 - Versioning
- Quality assurance processes



Type of data	Recommended formats	Acceptable formats
Tablular data	csv, .tab, .por, .xml	.txt, xls, .dbf, .ods, .sav, .dta, .mdb
Geospatial data	.shp, .shx, .dbf, .prj, .sbx, .sbn, .tif, .tfw, .dwg, .gml	.mdb, .mif, .kml, .ai, dxf, .svg
Textual data	.rtf, .txt, .xml	.html, .doc
Image data	.tif	.jpg, .gif, .tif, .tiff, .raw, .psd, .bmp, .png, .pdf
Audio data	.flac	.mp3, .aif, .wav
Video data	.mp4, .ogv, .ogg, .mj2	.avchd
Documentation and scripts	.rtf, .pdf, .xhtml, .htm, .odt	.txt, .doc, .xls, .xml

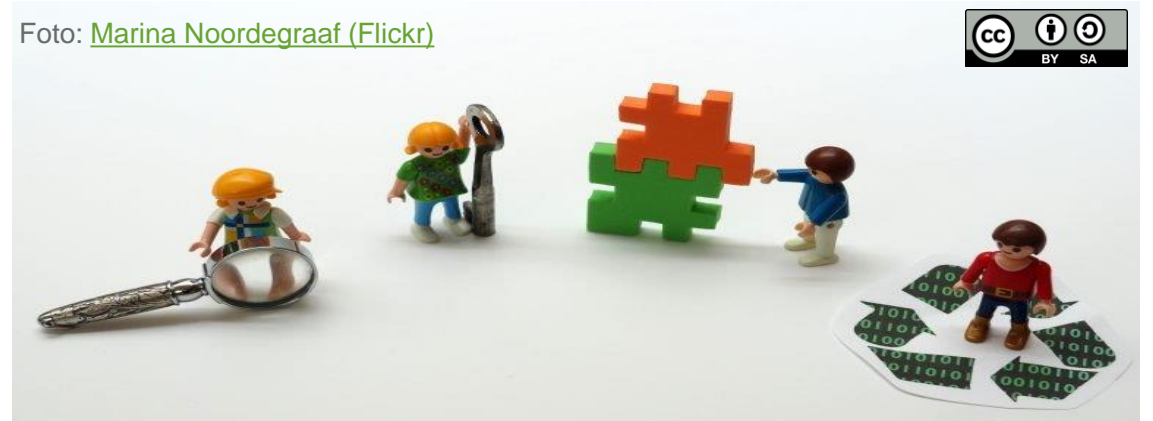
Source: [UK Data Service](#)

2. Documentation and metadata



- Metadata is „data about data“ (Examples: persistent identifier such as DOI, publication date, title, authors, description, keywords, licence, funding, related identifiers, etc.)
- Documentation may also include details on the methodology used, analytical and procedural information, definition of variables, vocabularies, units of measurement, assumptions made, and the format and file type of the data
- Existing community metadata standards: General (e.g. Dublin Core) or discipline specific (e.g. DDI);
See [RDA Metadata directory](#)

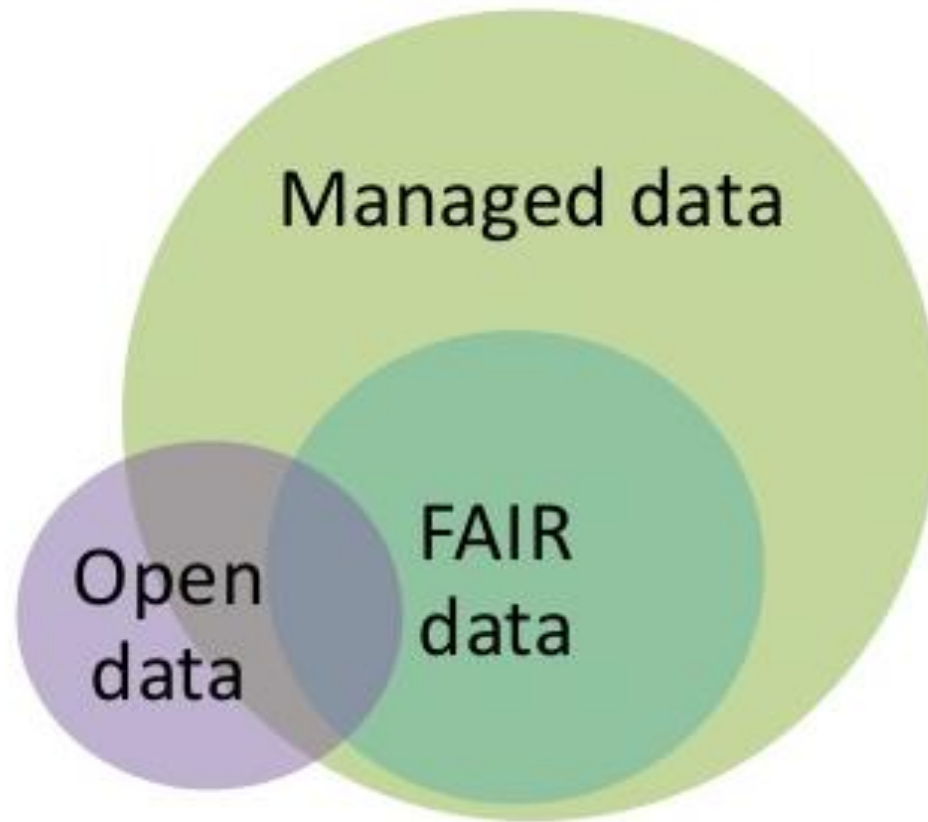
The FAIR data principles



- **Findable:** metadata and data easy to find for both humans and computers
 - **Accessible:** users need to know how can they be accessed, possibly including authentication and authorisation
 - **Interoperable:** data can be integrated with other data in applications or workflows for analysis, storage, and processing
 - **Reusable:** metadata and data should be well-described so that they can be used in different settings
-



Open vs. FAIR data

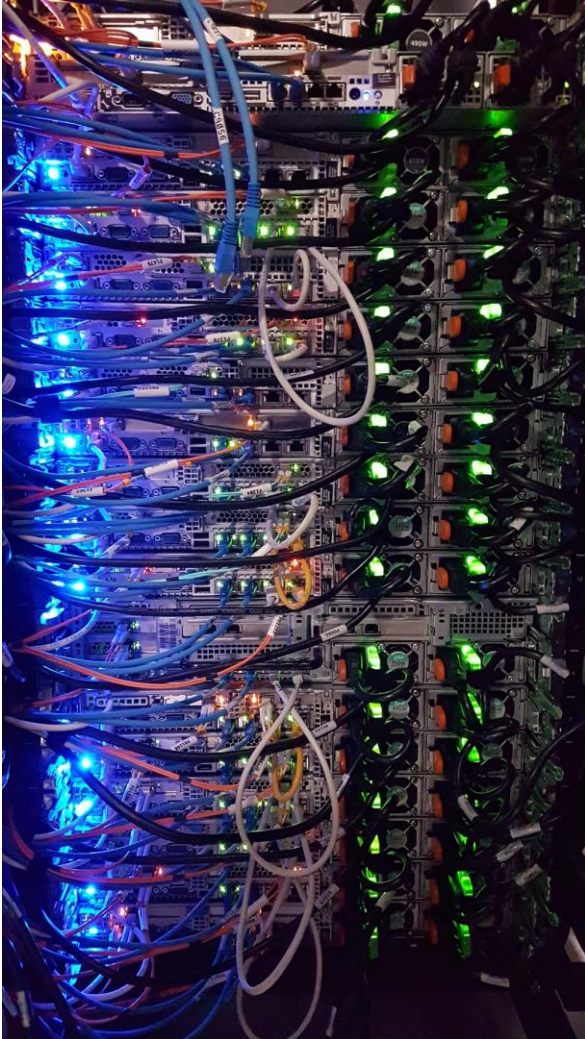


Jones, S. 2018. Open data, FAIR data and RDM: the ugly duckling. Available at: [Zenodo](#).

3. Ethics and legal compliance

- Review of research plans involving sensitive research on human subjects are submitted to the ethical committee
- Informed consent sought for data collection, processing and long-term preservation
 - Removal, aggregation, pseudo anonymisation, or anonymisation of direct and indirect identifiers in data files
 - Restriction of access do the data in cases, when anonymisation would hinder the reusability of data
- Compliance with General Data Protection Regulation (GDPR)
 - Confidential information and trade secrets

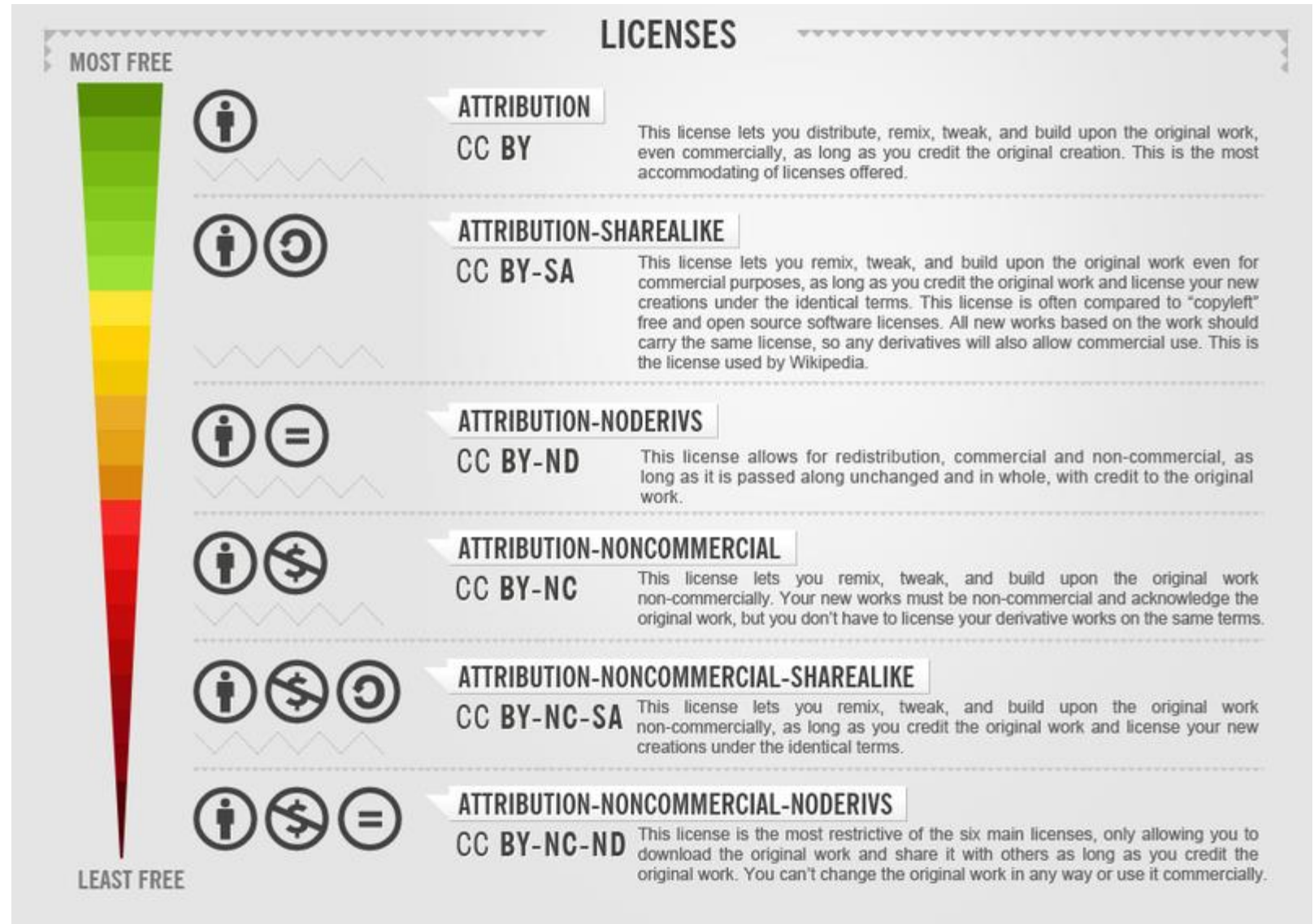




4. Storage and backup

- Copyright and intellectual property rights (Consortium agreements)
- Licences for reuse (e.g. [Creative commons](#))
 - Restriction on reuse of third-party data
- Data sharing restrictions (embargo periods)
 - Security measures and standards for confidential data

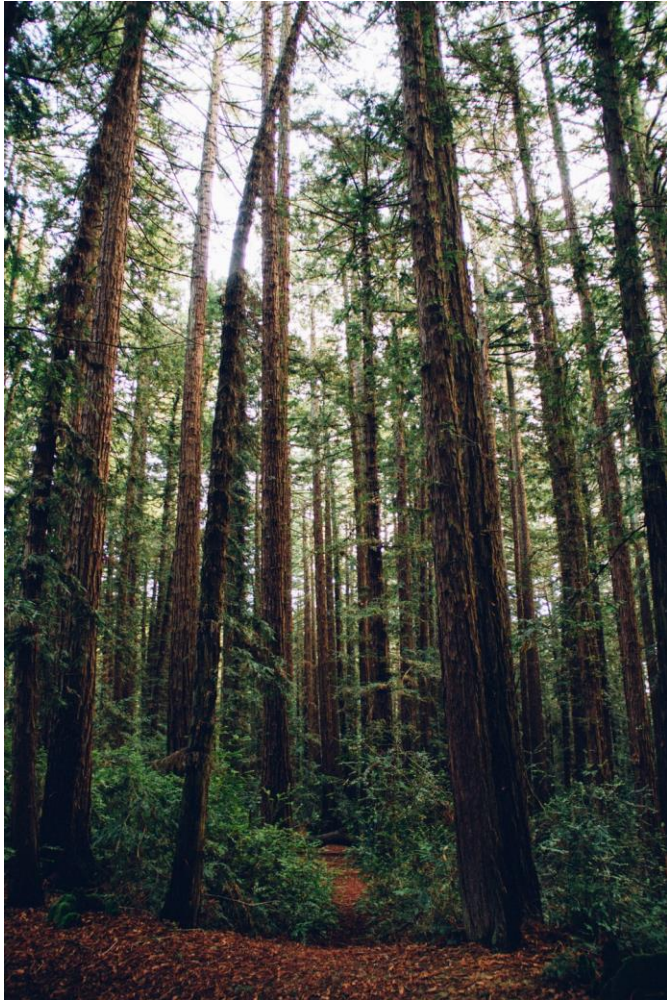
Publication Licences for reuse (e.g. Creative commons)



Source: [Wikipedia](#)

5. Selection and preservation

- What data must be retained/destroyed for contractual, legal, or regulatory purposes
 - Foreseeable research uses for the data (validation of research findings, conduct of new studies, teaching)
 - Length of retention and preservation
 - Repository or archive for data to be held
- Costs (repository charges, time and effort to prepare the data for sharing/preservation)





Research data repositories

- General purpose repositories (e.g. Zenodo, Figshare)
- Institutional data repositories
- Domain specific repositories
- Registry of research data repositories:





Example of domain-specific repository



The Materials Data Facility (MDF)

A simple way to publish, discover, and access materials datasets

Actions

- Publish
- Discover
- Contribute

Quick Links

- MDF Forge
- MDF Connect Client
- MDF Examples
- MDF Slides



Publish Data

1

Sign up and Join

Sign up for free Globus account using existing credentials. Join group for access to data publication capabilities.

- Create a Free Account
- Join this Group

2

Collect

Collect the data into your preferred file structure, preferably in openly accessible formats. Feel free to nest files as deeply as necessary for your use case, our indexers will find them!

- Set up an endpoint

3

Publish

Follow instructions in the form to publish your dataset.

- Publish Your Data
- Python Tools

General purpose and institutional repositories

- + Simple process
- + Fast publication
- The metadata scheme is too general
- Data is not reusable without rich metadata and documentation

Domain specific repositories

- + Field specific metadata standards
- Lack of repositories for certain fields
- Time consuming to prepare
- Requires domain specific skills



20/04/2018





6. Data Sharing

- Where, how, and to whom?
- Conditions for data sharing (data sharing agreements)
 - When will you make data available
 - Persistent identifiers
 - Outline of expected restrictions due to confidentiality, lack of consent agreements or intellectual property rights
- Non-disclosure agreements for protection of confidential data



7. Responsibilities and resources

- Roles and responsibilities for data capture, metadata production, data quality, storage and back-up, data archiving and data sharing.
- Determine if specialist expertise is required (data stewards)
- Determine if additional hardware or software is needed

Key data challenges in engineering

- A lot of research is done in collaboration with industry which is reluctant to share data
- Fear of losing competitive advantage
- Fear of data being misused and misinterpreted
- Lack of metadata standards and ontologies (for certain engineering fields)
- Lack of domain-specific repositories (for certain engineering fields)



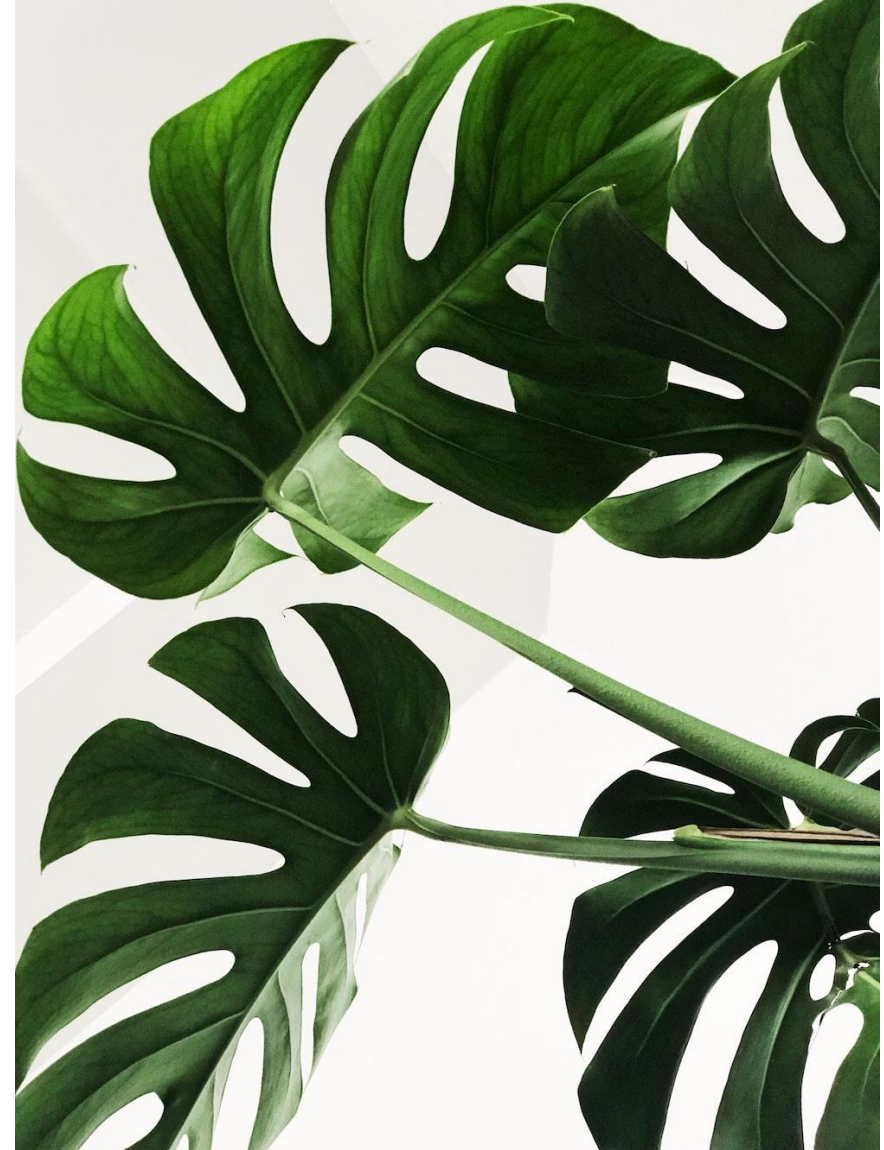


DMP examples

- Material science:
 - Hong, X. 2016. [Effects of mine waste materials in the north-central Mojave dessert.](#)
 - Sergan, V. 2014. [Orientational order induced by a polymer network in the isotropic phase of liquid crystal.](#)
 - Dahle, S. et. al. 2019. [Data Management Plan for the MSCA project PlasmaSolution.](#)
 - Social science:
 - Slavec, A. 2019. [Using questionnaires to measure attitudes and behaviours of building users: Data Management plan.](#)
-

Organisations, resources

- [Research Data Alliance](#)
- [CODATA](#)
- [Digital Curation Centre](#)
- [UK Data Service](#)
- [GO FAIR project](#)
- [FOSTER project](#)
- [Open Science MOOC](#)
- [European Open Science Cloud](#)
- ...



Where to learn more?

<https://rd-alliance.org/>



RESEARCH DATA ALLIANCE

„RDA is building the social and technical bridges that enable open sharing of data to achieve its vision of researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.“



RDA in Czech Republic

Contact persons: **Tomas Mildorf** (University of West Bohemia), **Karel Charvat** (Czech Centre for Science and Society) and **Hana Kubuíčková** (Plan4All)



RESEARCH DATA ALLIANCE
EUROPE





Questions?





Let's stay in contact



Ana Slavec

InnoRenew CoE

ana.slavec@innorenew.eu

@aslavec
