17.06.2019 | DHI Paris

Conference: Teaching History in the Digital Age – international Perspectives (DHI Paris)

https://dhdhi.hypotheses.org/5896

Is Your Research Future Proof?
Data Management
Techniques & Tools for Digital Historians

Ulrike Wuttke

(v_1.0)

DOI: http://doi.org/10.5281/zenodo.3247014

Ulrike Wuttke, University of Applied Sciences Potsdam / RDMO

@UWuttke | CC-BY 4.0 | RDMO | @RDMOrganiser

Gefördert durch DFG Deutsche Forschungsgemeinschaft

TABLE OF CONTENTS

1) Warm Up
2) Code of Conduct
3) Rationales and Benefits of the Workshop
4) Background: Open Science & Open Scholarship
5) Research Data in Humanities and Heritage Science
6) Key Concepts: Open Data, FAIR Principles, Research Data management, Research Data life Cycle
7) RDMO
8) Good Research Data Management Practices
9) Wrap Up
10) Further Learning (Resources)

01

WARM UP

Who has an ORCID?

https://orcid.org/

Who has already published something OA besides articles, books?

- Data?
- Software code?
- …

Source Picture: Afficheur par Joly by Spiessens, CC BY SA 3.0 https://commons.wikimedia.org/wiki/File:Afficheur_par_Joly.jpg

Who has already written a data management plan?

Source pictures: Left ‘Adam’s Creation Sistine Chapel ceiling’ by Jörg Bittner Unna, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=46496746, Right ‘Hand’ by Ulrike Wuttke, CC BY 4.0

02

CODE OF CONDUCT

Workshop Code of Conduct

- **Respect** for each other
- There are **no stupid questions**
- We are all **experts**
- **Connect** with each other
- **Share** with the world

03

RATIONALES AND BENEFITS OF THE WORKSHOP

Learning objectives

Participants…

- will be able to **explain what is (their) humanities research data**
- understand **Open Principles** and the Open Science vision
- can summarize the **FAIR principles** in a Humanities context
- understand **importance of FAIR RDM for their own research**
- understand how **writing a DMP can help their research**
- know the **topic areas of a DMP**
- have acquired a **basic understanding** how to address them
- know the **data management tool RDMO** and have used it
- know where to look for **further resources**

✓ **Access to workshop materials** (slides) via **Zenodo**

Source Picture: From “The Open Science Training Handbook” v.1.0, 2018, https://book.fosteropenscience.eu/

04

BACKGROUND: OPEN SCIENCE & OPEN SCHOLARSHIP

What is Open Science?

- Open Science is a **paradigm change**
  - **2nd Scientific Revolution** > from traditional journal system (17th c. = print paradigm) to **new digital forms and methods** that make full use of the possibilities of the internet
  - Dissemination of scientific knowledge **as wide as possible, free of charge to all users, and accessible online**
  - Umbrella term for different practices aimed at making research more **accessible** and **transparent**
- Alternative terms: Open Research, Open Scholarship, e-Science, Science 2.0

Why Open Science?

- Scientific knowledge is a product of **social collaboration** and its ownership belongs to the **community** (sociological argument)
- Scientific outputs generated by public research are a **public good** that everyone should be able to use at no cost (economical argument) (UN Sustainable Development Goals)

Read more: https://www.fosteropenscience.eu/content/what-open-science-introduction

Advantages of Open Science for research(ers)

Picture: open by velkr0 CC BY 2.0, https://flic.kr/p/mzqM

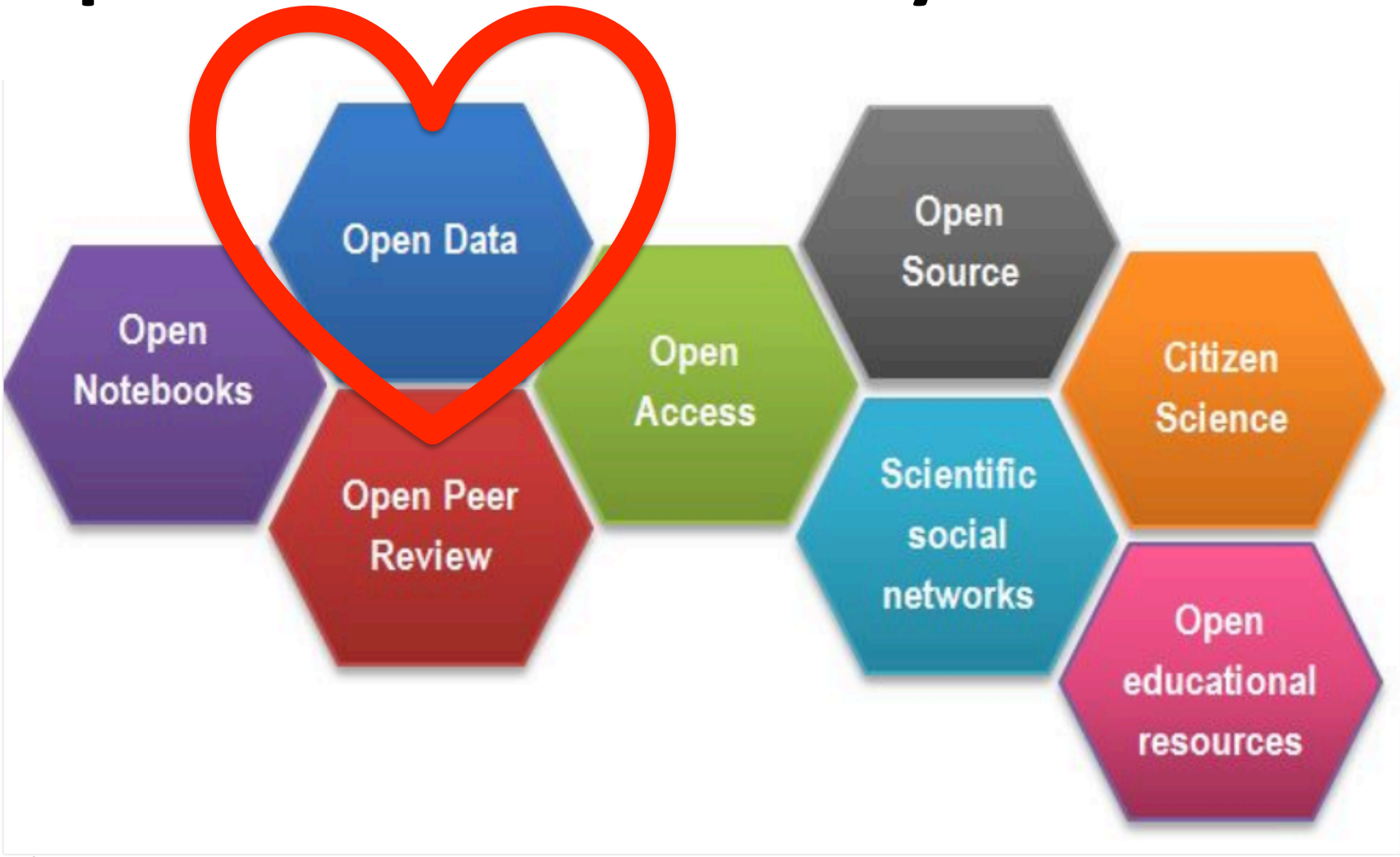
- Higher **transparency** of research methods and evaluation
- Higher **reproducibility** of research findings
- Researchers and research institutions **save money and time**
- Higher (societal) **impact** of research(ers)

➢ Open Science gets research(ers) **out of the Ivory Towers**!

➢ Open Science is **Good Scientific Practice**

Source Picture: https://zenodo.org/record/1285575#.W09yZH59jOR (Melanie Imming, John Tennant, CC0)

Open Science has many facets



Source Picture:
https://www.fosteropenscience.eu/content/what-open-science-introduction, CC BY 4.0

05

RESEARCH DATA IN HUMANITIES AND HERITAGE SCIENCE

eHumanities and eHeritage Research

What is it about?

- **Digital transformation** (digital workflows & digital methods) led to a rise of **quantitative** and **statistical methods** in the Humanities and Cultural Heritage

Opportunities

- New scholarly methods, research activities, and objects transform and broaden the Humanities and CH > **Digital Humanities** (DH) and **eHeritage**

Challenges

- Research processes dominated by **traditional paradigms**
- **Issues of Access** (copyright and license issues)
- **Sustainability** (data loss)
- Lack of documentation and standardization

➢ eHumanities and eHeritage are based on **accessible, correct, authorative, well structured data**

➢ **Interoperability** (machine actionability) & **Reusability** (culture of sharing)

What is Data, Anyway?

Do Humanities and Cultural Heritage researchers have data?

- Yes, a lot, but they don't tend to use the word data (but **sources** etc.)
- **Research data is data that is used and produced in research processes** such as digitization, study of sources, experiments, measurements, interviews, and surveys

What is Data, Anyway?

- **Examples for Humanities data**: primary sources (texts, pictures), secondary sources, theoretical texts, digital tools (software), annotations, bibliographies etc.
- Most **sources are research data** and their management has in fact always been part of the scientific process; digitization only adds complexity
- There are **digitized sources** and **born digital sources**
- **Various formats and types** (pictures, texts, multimedia, measurements, etc.)

Are Humanities and Cultural Heritage Data Special?

- Yes and No!
- Humanities are a very **broad research discipline**, many specific research contexts, but also increasingly interdisciplinary research
- Humanities research lives from **enrichment** of data (layers of interpretation)
- Problematic to distinguish between **primary data (raw data) and secondary data**
- Issues with **ownership** of the data (cultural heritage institutions, publishers)

➢ **Many issues and solutions apply beyond Humanities and Heritage Science!**

It can get pretty complex, though…

An **information unit** consists of - e.g. in the case of an **interview**:

- **Audio file**
- **Interview transcript** (digital text file)
- Discussion guide or **questionnaire**, which explains the methodological approach (comprehensibility of the results of the study)
- **Project explanation** as well as the **declaration of consent** of the interviewee (compliance with the legal provisions of the Federal and State Data Protection Act)
- **Codebook** (documents the development categories and variables)
- **Documentation** of the procedure for anonymization and pseudonymization
- **Indexing information** (metadata) (guarantees citation ability of the interview and its findability)

Based on Gisela Minn und Marina Lemaire (2017): Forschungsdatenmanagement in den Geisteswissenschaften. Eine Planungshilfe für die Erarbeitung eines digitalen Forschungskonzepts und die Erstellung eines Datenmanagementplans (Universität Trier eSciences Working Papers, Nr. 03), Trier, p. 10 <urn:nbn:de:hbz:385–10715>

Playful Exercise 1

WHAT ARE YOUR RESEARCH DATA?

What are YOUR Research Data?

- In your discipline?
- In your current project?
- In past projects?

✓ Form groups
✓ Discuss and note results on sticky note
✓ Bring sticky notes to front

06

KEY CONCEPTS: OPEN DATA, FAIR PRINCIPLES, RESEARCH DATA MANAGEMENT, RESEARCH DATA LIFE CYCLE

Key Concept: Open Data

What is it about?

- Open Data = (research) data that is **freely available online for (re)use and republish** for everyone provided that the data source is attributed:

  „Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.“

  Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003)

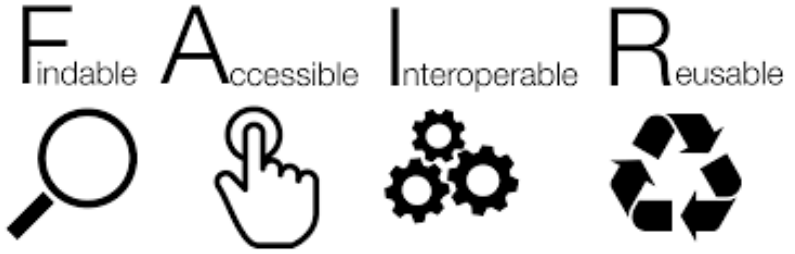
- Ideal: Data with no restrictions from copyright, patents, or other control mechanisms > transparent results
- However: “**as open as possible, as closed as necessary**”

Key Concept: Open Data

What does Open Data involve?

- Sharing is not giving away, to work in an open environment **benefits all, especially the data sharer**
  - **reach as many people as possible**
  - **be cited more often**
  - **build cooperation**
  - **etc.**
- Poses challenges, e.g. **interoperability** and **documentation**
- Some aspects are **discipline specific** > e. g. Humanities
- Essential: **Data Management Planning**

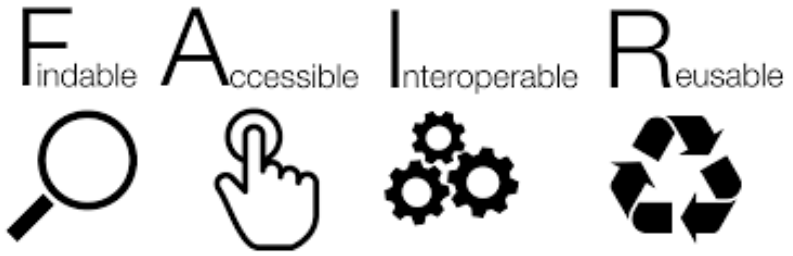
Key Concept: The FAIR Principles



- FAIR Guiding Principles for scientific data management and stewardship
- Baseline understanding for the value sharing data can deliver and the baseline requirements for doing so
- Developed by FORCE 11 [1]
  - **F**indable
  - **A**ccessible
  - **I**nteroperable
  - **R**eusable
- Note: **Not all FAIR Data is Open Data** (e. g. sensitive data) and not all **Open Data is FAIR**

[1] https://www.force11.org/group/fairgroup/fairprinciples

Key Concept: The FAIR Principles



- **F**indable = Data and Metadata are **easy to find** for both humans and computers
- **A**ccessible = Humans and computers can **readily access** or download datasets
- **I**nteroperable = Data from different datasets can be **prepared to be combined** or exchanged
- **R**eusable = Published data can be **easily combined** or **replicated** in future research

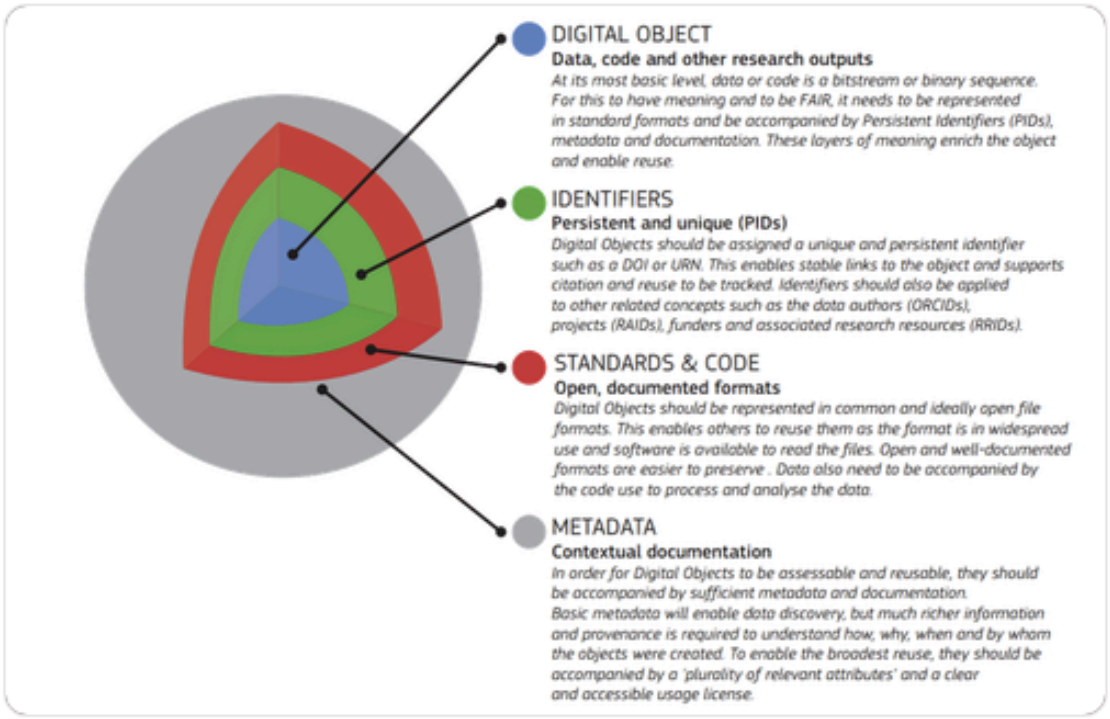
Compare EPFL Library Research Data Management FAST Guides:
https://researchdata.epfl.ch/wp-content/uploads/EPFL_Library_RDM_FastGuide_All.pdf

Simon Hodson
@simonhodson99

Folgen

A model for FAIR digital objects - 'FAIR should be applied broadly to all objects (including metadata, identifiers, software and DMPs) that are essential to the practice of research' #FAIRdata #EOSCStakeholdersForum

Tweet übersetzen



https://twitter.com/simonhodson99/status/1065606239135170560

Key concept: Research Data Management

"Research Data Management describes the process to **curate** (or manage) research data along the **research data lifecycle** and includes various activities such as **planning, producing, selection, analysis, archiving, and preparation for reuse**. Because data are very heterogeneous, **discipline and data specific solutions** can be required."

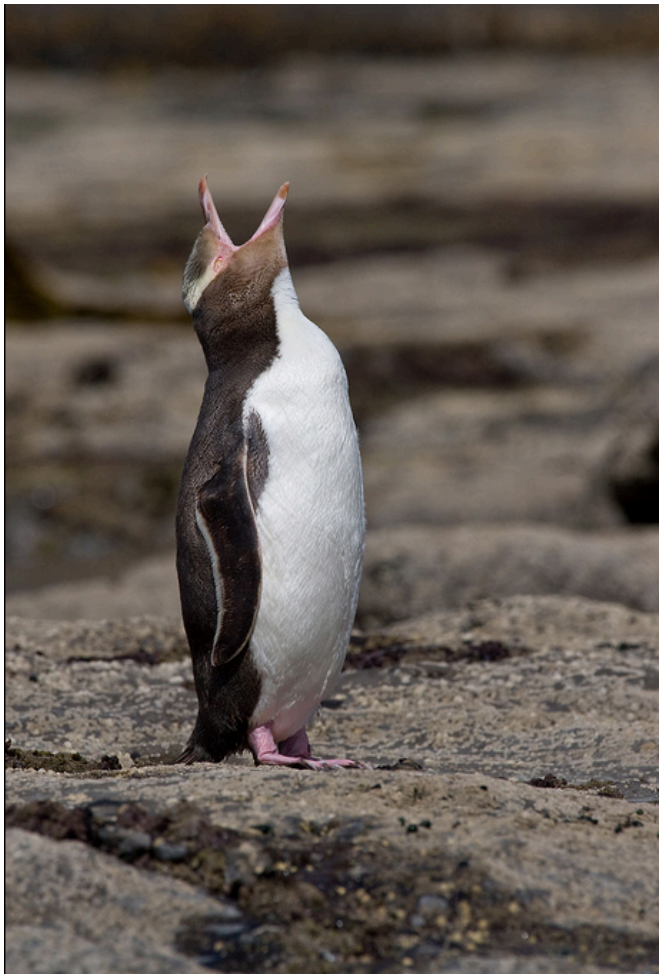
Translated (UW) from: AG Forschungsdaten der Schwerpunktinitiative "Digitale Information" der Allianz der deutschen Wissenschaftsorganisationen, Forschungsdatenmanagement: Eine Handreichung, 2018, p. 4. Online: http://doi.org/10.2312/allianzoa.029

Key concept: Research Data Management

- **Documentation and Selection of Research Data for Publication** (are there **valid** reasons to not publish the data?)
- **Apply FAIR Principles** to publication of Research Data

RDM: The Early Bird Catches the Worm!

- Research Data Management is **Good Scientific Practice** and a basic research skill
- Make a **Research Data Management Plan:** It's not just a document or an afterthought, but an **action plan to facilitate FAIR data** and **to ensure your research integrity** and **transparency**



Picture: A Yellow-eyed Penguin (Megadyptes antipodes) in the Curio Bay, New Zealand by Christian Mehlführer CC-BY 2.5
https://commons.wikimedia.org/wiki/File:Yellow-eyed_Penguin_crying_MC.jpg

Research Data Management

=

basic research skill

Sarah Jones
@sjDCC

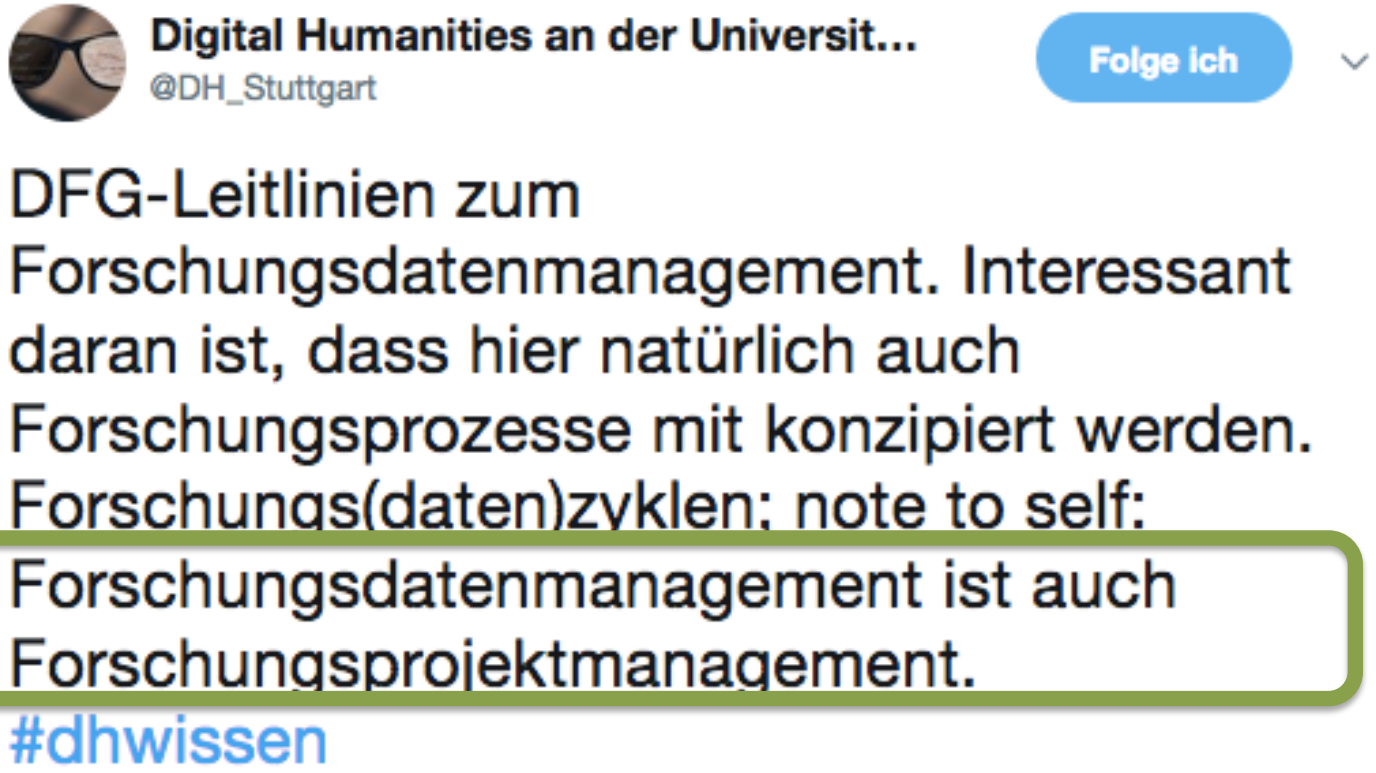
Folge ich

In the #FAIRdata EG report we argue that all researchers need a foundational level of data skills. They don't need to become data scientist or data stewards (unless they choose to) but do need to understand the main principles of data management #EOSCStakeholdersForum

Monsense and more... @barendmons
#EOSCStakeholdersForum agree with Francoise that researchers who do proper data stewardship should be rewarded, but the future scientist should be 'consciously imcompetent on data stewardship and hire professionals in the team.

https://twitter.com/sjDCC/status/1065569309106282496

Research Data Management = Research Project Management



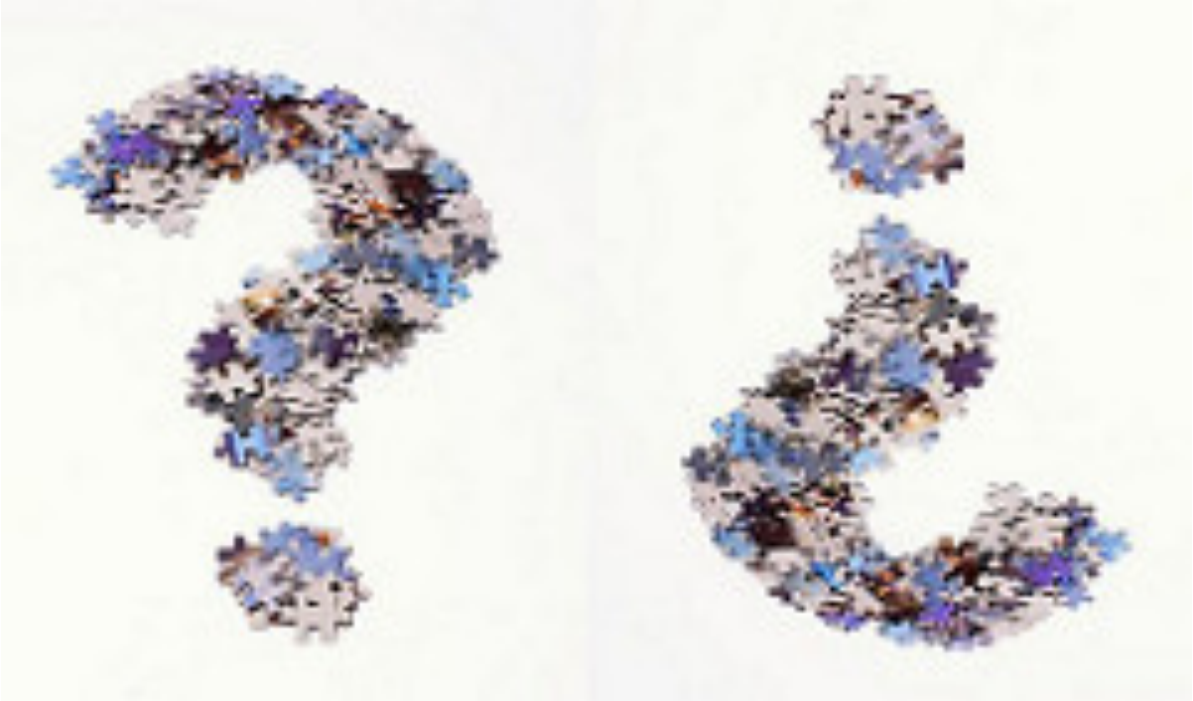
https://twitter.com/DH_Stuttgart/status/1057677832330055683

Playful Exercise 2

RESEARCH DATA LIFECYCLE

What is the correct order of the **Research Data Lifecycle**?

✓ Discuss in groups and use the cards (3 min.)



Key Concept: Research Data Lifecycle

Research Data Lifecycle from https://www.ukdataservice.ac.uk/manage-data/lifecycle

07

BASIC PRINCIPLES OF RESEARCH DATA MANAGEMENT (IN HUMANITIES AND HERITAGE SCIENCE)

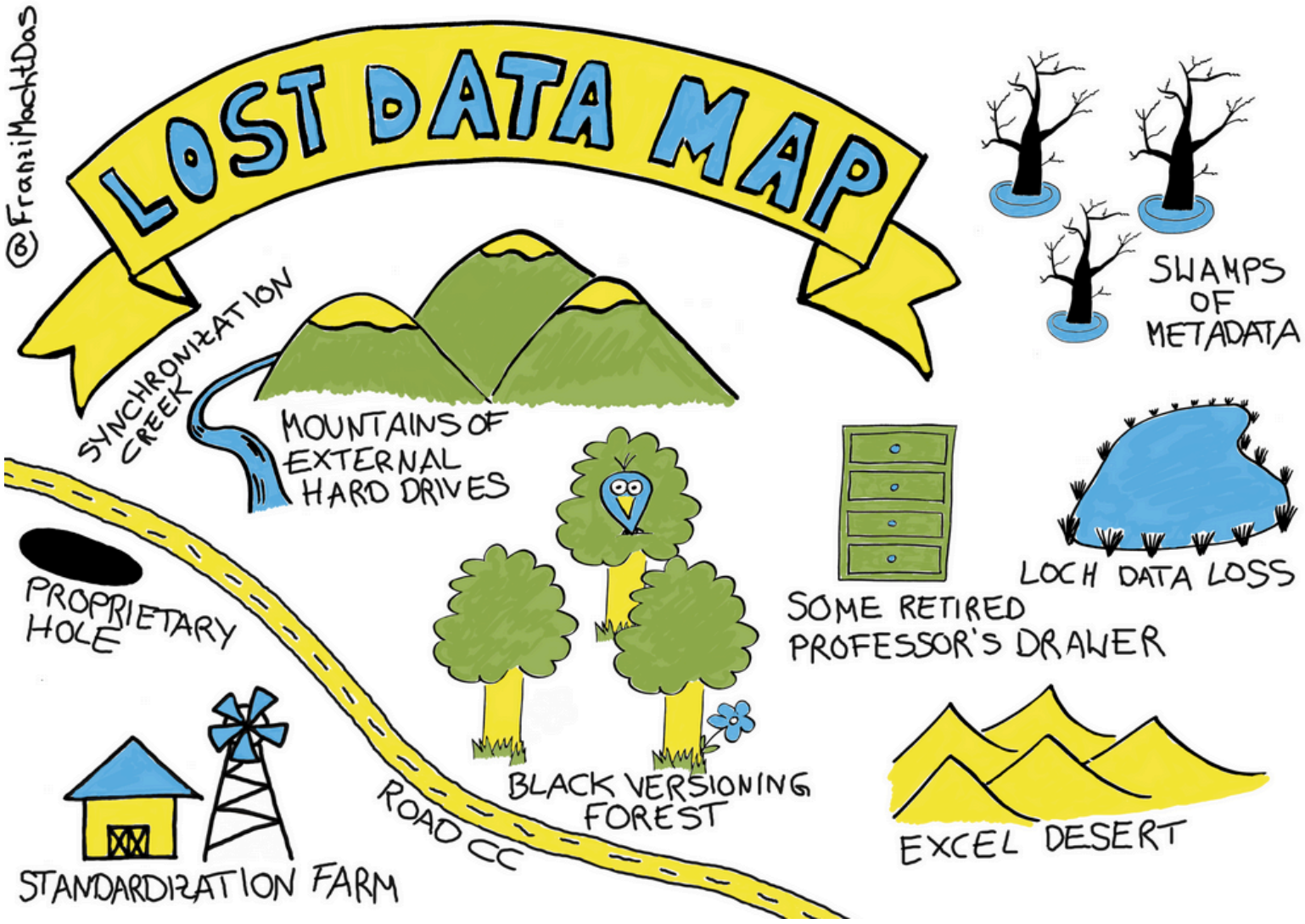
Why Research Data Management?

Source video: https://rdmpromotion.rbind.io/blog/2019/01/24/video-pre-release/ Julien Colomb (CC-BY 4.0)

Why would I want or need to manage, improve or open up my data?

- **Create opportunities for using and reusing data**, for collaborating, informing and increasing the impact of the work (contemporary issues, interdisciplinary research, engaging broader society) > **Publication of research data**
- Fulfill **funder requirements** on national and international level (e.g. European Commission) = Research Data Management and Open Science
- Be prepared for **Research Data Policies** (institutional, journals)

Good Data Management is good for you!



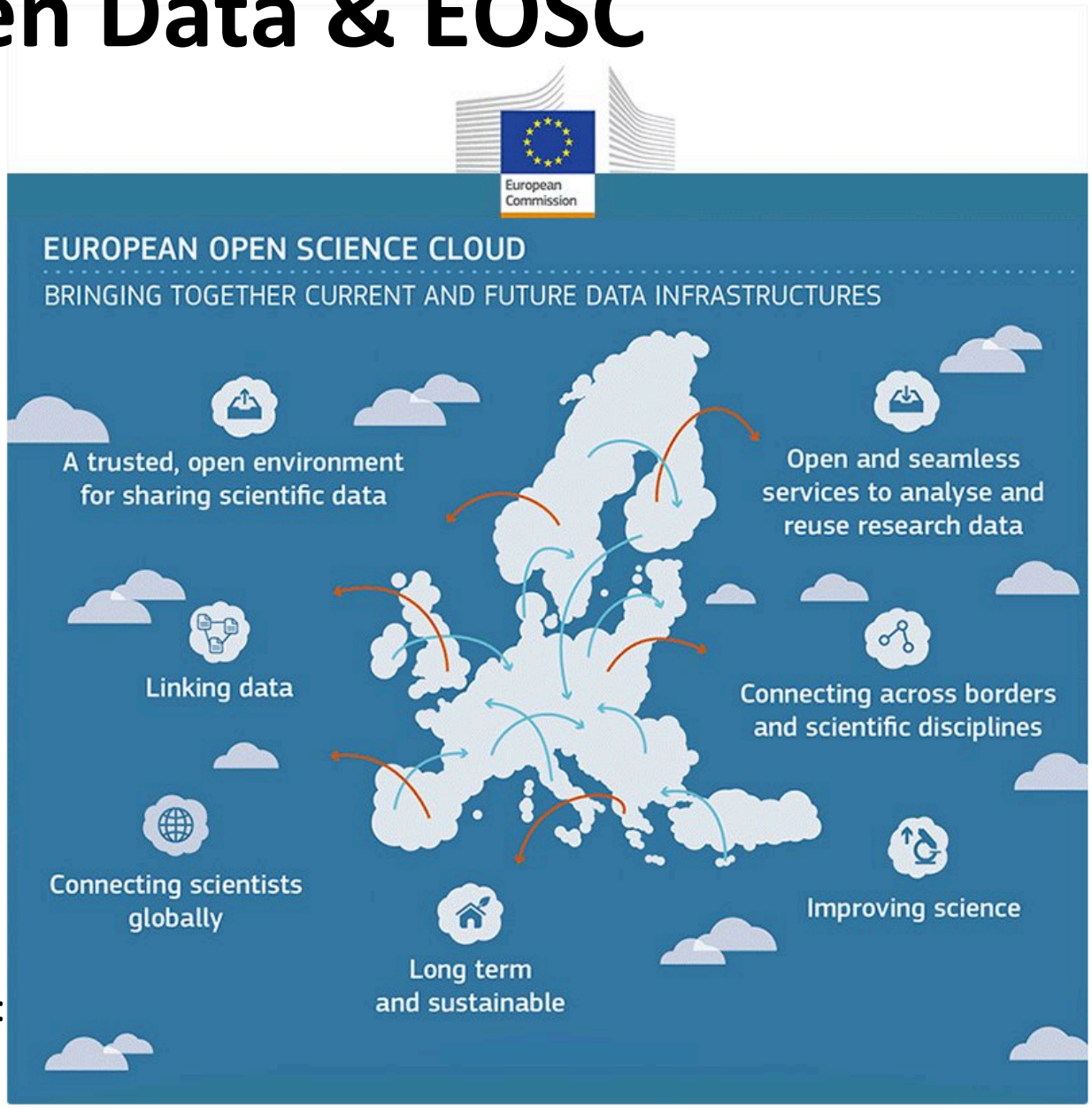
DORA
(San Francisco Declaration on Research Assessment)

For the purposes of research assessment, ***consider the value and impact of all research outputs (including datasets and software)*** *in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.*

https://sfdora.org/

Vision: Open Data & EOSC

"a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines"
https://www.eosc-portal.eu/about/eosc



European Open Science Cloud: https://www.eosc-portal.eu/

Source Picture: https://library.ktu.edu/news/launch-of-the-european-open-science-cloud-eosc/

Vision: Open Data & EOSC

Screenshot from Video: https://youtu.be/1e0pahL9wr4

Future proof Research Data Management: Let's go!

- Main obstacles to Research Data Management and sharing are **cultural**
- Data skills are necessary, but not main obstacle and **beginnings of Research data management are easy to grasp**

Picture: Road Sign by Free Images (www.inkmedia), CC BY 2.0 https://flic.kr/p/JoVNhU

What is a Research Data Management Plan (DMP)?

- A DMP is a document that contains **information about handling, organising, documenting and enhancing research data**, and enabling their sustainability and sharing for a research project
- A DMP describes and analyzes **workflows** along the Research Data Lifecycle
- A DMP can be a **few paragraphs short up to several pages long** (first draft usually short)

The first step is always the hardest...

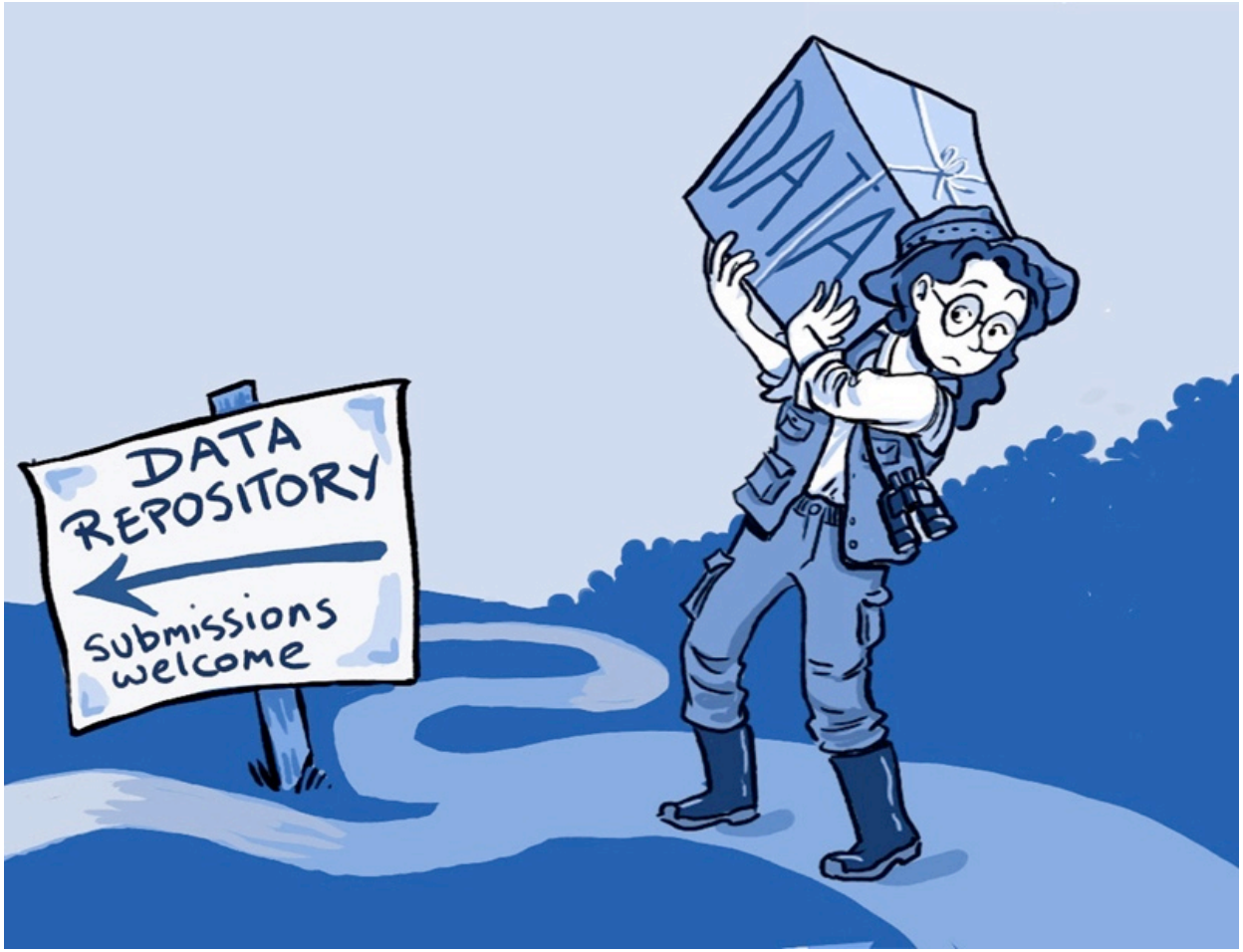
Topics in a DMP (here: DCC Template):

- Data Collection
- Data Documentation and Metadata
- Ethics and Legal Compliance
- Storage and Backup
- Selection and Preservation
- Data Sharing
- Responsibilities and Resources

MORE in Session 2!

RDM? You are not alone!

☺



(Credit Ainsley Seago, PLoS Biology)
http://www.kavlifoundation.org/science-spotlights/breaking-down-data-barriers-neuroscience#.XD5P2fx7nOR

RDM? You are not alone!

- RDM is **team work**!
- Use **tools** for **Data Management Planning,** e. g.
  - RDMO (https://rdmorganiser.github.io/en/)
  - DCC DMPOnline (https://dmponline.dcc.ac.uk/)
- Make **use** of **infrastructural support** (research infrastructures, cultural heritage institutions, libraries, data centres)
  - e. g. **DARIAH-EU Helpdesk**: https://www.dariah.eu/helpdesk/
- Ecosystem of **digital research infrastructures**, cultural heritage institutions, libraries, data centers, etc.
- RDM costs can be **eligible for reimbursement** (e.g. H2020)

➢ Ask your **library** and **research data manager**!

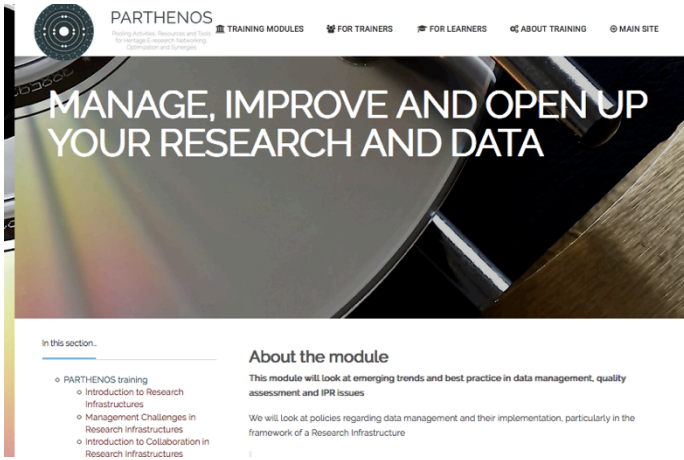
Learn Research Data Management in the PARTHENOS Training Suite

Module "Manage, Improve and Open Up Your Research Data"

- Intermediate level

- Emerging trends and best practice in Data Management, Quality Assessment, Intellectual Property Rights

- e.g. FAIR Principles, Data Management Planning, Open Data, Open Access, Open Science, etc.



- http://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/

Learn Research Data Management in the PARTHENOS Training Suite

Webinar: "How to work together successfully with eHumanities and eHeritage research infrastructures: The Devil is in the Details"
Trainers: Marie Puren (Inria) and Klaus Illmayer (OEAW)

- Beginners' to intermediate level
- Research lifecycle "Plan Research Project"
- FAIR Principles
- Standards (PARTHENOS Standardization Survival Kit – SSK)

- http://training.parthenos-project.eu/sample-page/ehumanities-eheritage-webinar-series/webinar-work-with-research-infrastructures/

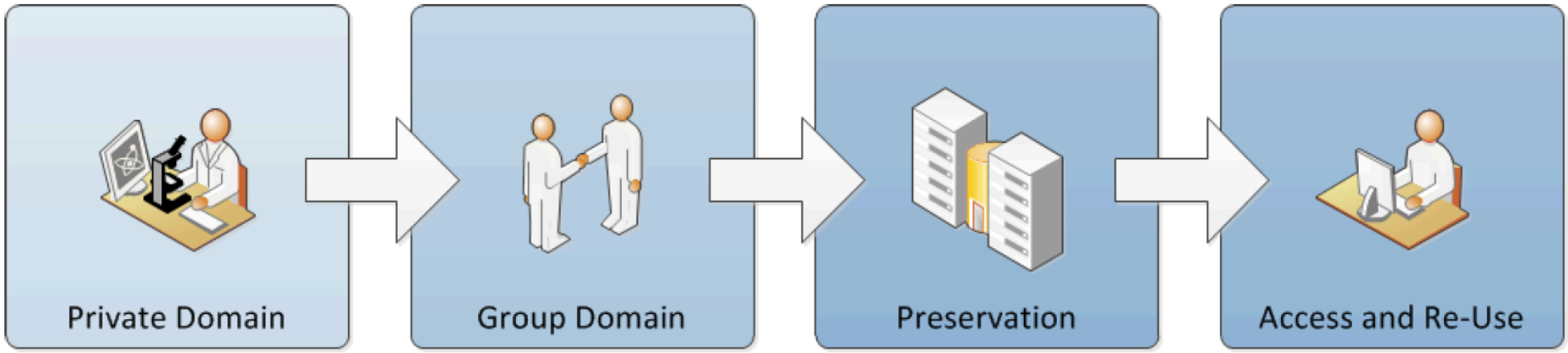
You will find more RDM Resources at the end of this slide deck!

Playful Exercise 3

DOMAIN MODEL FOR RESEARCH DATA

What about your data?

Discuss in groups (5 min.)

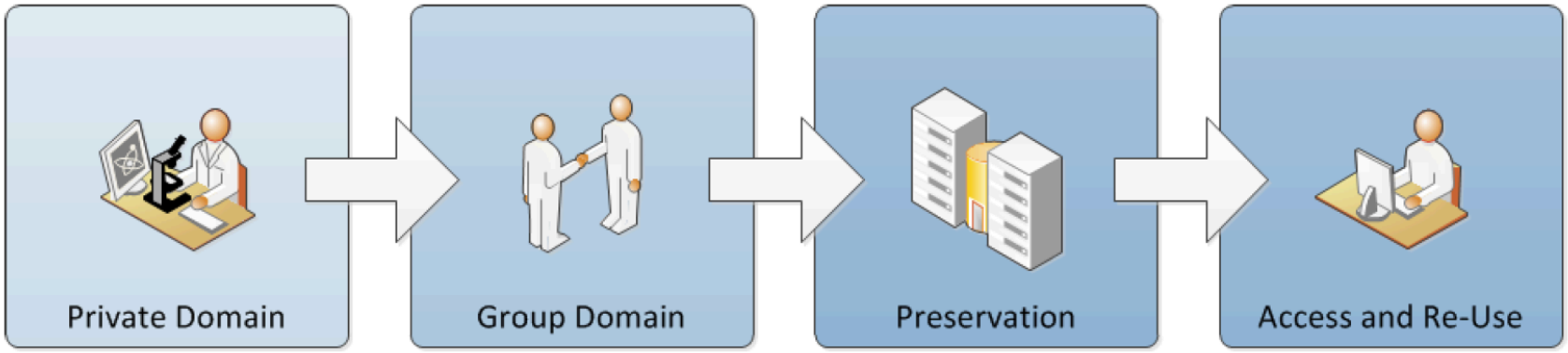


- Which data do you produce & use?
- Which of your data need to be kept?
- Which of your data could be published OA?
- Have you ever reused data? What are your experiences? Which documentation is needed?

& Publication

- ✓ Form groups
- ✓ Discuss and note results on paper
- ✓ One person summarizes results (1-2 sentences)

Domain model for research data



& Publication

Effort for **data management** ⟶

Complexity of **metadata** ⟶

Treloar, A., D. Groenewegen, and C. Harboe-Ree (2007), The Data Curation Continuum - Managing Data Objects in Institutional Repositories, http://dx.doi.org/10.1045/september2007-treloar

DFG-Projekt RADIESCHEN (2013): Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur. Organisation und Struktur. http://dx.doi.org/10.2312/RADIESCHEN_005

Source Slide: Jochen Klar. (2018, November). Create a data management plan (with RDMO). Zenodo. http://doi.org/10.5281/zenodo.1493342

Question Time & Outlook to Session 2

Break

Picture: Manchots empereurs tobogannent by Samuel Blanc
https://commons.wikimedia.org/wiki/Spheniscidae#/media/File:Manchots_empereurs_tobogannent.JPG, CC BY SA 3.0

Playful Exercise 4

KEYWORD SALAD

Keyword Salad!

- ✓ Form Groups
- ✓ Explain the Keywords on the cards to each other

Picture: Pop: appearing from Nowhere by Gord Webster https://www.flickr.com/photos/thievingjoker/352538051, CC BY SA 2.0

This session:
Treasure Hunt for your own DMP

- ✓ Use the DCC Print Template or RDMO and additional materials to collect information for your own project etc.

Picture: Lissa R “Lopsided Rainbow and a pot of Gold”, CC BY-NC 2.0 Generic
https://www.flickr.com/photos/lepetitcirque/5829795680

07

RDMO

Recap: Research Data Management Planning

- Often you will need a written and agreed **Data Management Plan** (DMP), esp. in case of external funding
- To help DMP, many funding agencies provide a **model or template for a DMP**
- DMP may seem an intimidating (or even unwelcome task), but in the end, it is just a tool for **thinking systematically through your research process from a "data perspective"**
- DMP helps you to **maximize research value** (high quality research data and research excellence) and **prevents unpleasant surprises** at the close of your project (and data loss!)

Tools for Data Management Planning

Why use a tool for a DMP?

- Structured (discipline) questions and additional features

Tools:

- **RDMO** Germany (https://rdmorganiser.github.io/en/)
- **DMPOnline** U.K. (https://dmponline.dcc.ac.uk/)
- **DMPTool** U.S. (https://dmptool.org)

Today:

- **RDMO** (Research Data Management Organiser)

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

RDMO in a nutshell

RDMO

- Developed by AIP, FHP, and KIT
- Funded by DFG
- **Organiser** instead of a plan: *Actionable* DMP
- **Local** instead of central (protect sensitive information)

AIP Leibniz-Institut für Astrophysik Potsdam

FH;P Fachhochschule Potsdam University of Applied Sciences

KIT Karlsruher Institut für Technologie

RDMO in a nutshell

- **RDMO supports RDM over the whole lifetime of the project**
- Engage all stakeholders
- Collect all necessary information for a sustainable RDM
- Guided questionnaire
- Controlled vocabularies
- Checkboxes, dropdowns, sliders
- Export in different formats using predefined views
- Identify follow up tasks
- Interfaces to other software, APIs
- Open Source

MORE about RDMO: Jochen Klar. (2018, November). Create a data management plan (with RDMO). Zenodo. http://doi.org/10.5281/zenodo.1493342

Playful Exercise 5

TRY OUT RDMO

Try Yourself RDMO!

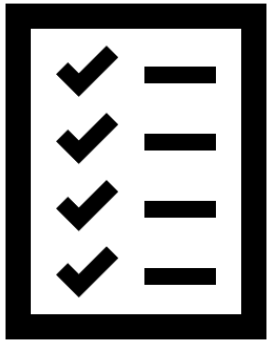
RDMO

- ✓ Go to the RDMO demo instance: https://rdmo.aip.de
- ✓ Login
- ✓ Look around
- ✓ Locate DCC template and test it out
- ✓ If you like it: Use it for the **Treasure Hunt**

08

GOOD RESEARCH DATA MANAGEMENT PRACTICES

Recap: Topics in a DMP (here: DCC Template):



- Data Collection
- Data Documentation and Metadata
- Ethics and Legal Compliance
- Storage and Backup
- Selection and Preservation
- Data Sharing
- Responsibilities and Resources

Data Collection

- Think about **naming conventions** and the **order** of your files
- **Use open** or **standard formats (archive format** will maybe be different from **working format)**
- Creating a **structure** that is FUTURE proof, will save you a lot of time afterwards and act as **booster for your individual research** and especially for **team efficiency**
- Use tools to **find research data** (e.g. re3data: https://www.re3data.org/)
- Get acquainted with **Data (and Software) Citation**

Data Documentation and Metadata

- Metadata are used to describe and organize data: formal description and content description
- Metadata are a love letter to the future
- Use **Standards** for Metadata to enhance interoperability (disciplinary and generic)
- **Transparent documentation** includes project description, aims, methods, data cleaning, versioning, etc. (=Your DMP!)

Picture, Otto the Open Access Otter, by Katja Diederichs

Data Documentation and Metadata

➽ **Relevant Standards:**

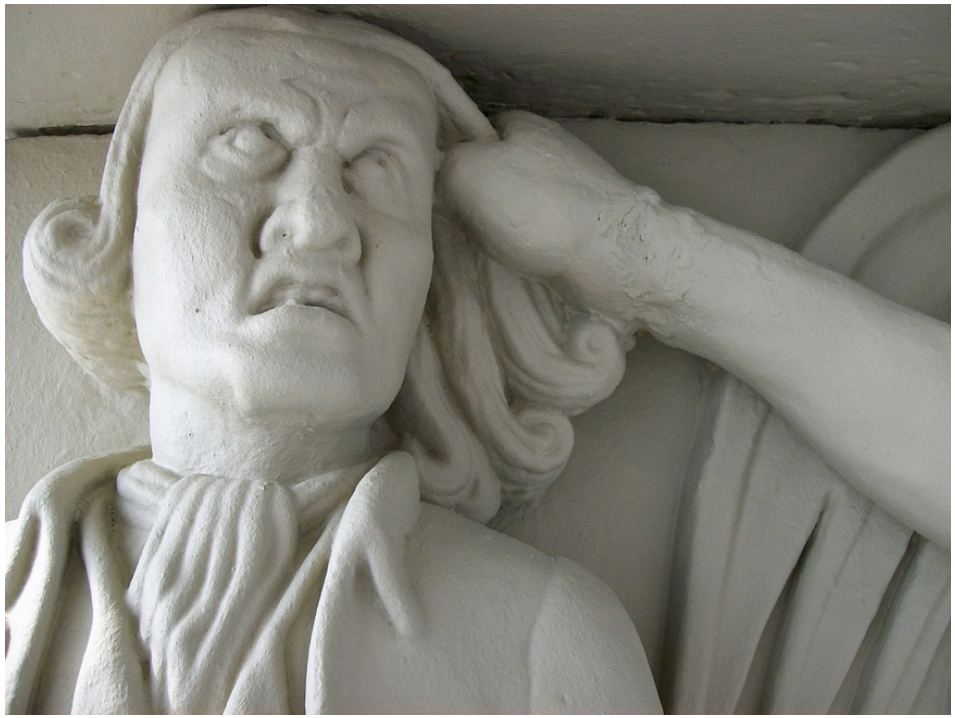
- **TEI** (Text Encoding Initiative): www.tei-c.org
- **CEI** (Charter Encoding Initiative): http://www.cei.lmu.de/index.php
- **MEI** (Music Encoding Initiative): https://music-encoding.org/
- **CMDI** (Language Resources, CLARIN):
- **IIIF** (International Image Interoperability Framework): https://iiif.io/
- **EAD** (Encoded Archival Description, for finding aids): https://www.loc.gov/ead/
- **Dublin Core** (description of digital documents): http://dublincore.org/

✓ **BARTOC: Basel Register of Thesauri, Ontologies & Classifications:** https://bartoc.org/

✓ **PARTHENOS Standardization Survival Kit (SSK):** https://ssk-application.parthenos.d4science.org/ssk/#/

Ethics and Legal Compliance

- If you are dealing with **personal or sensitive data** you need to take extra measures for data safety
- Get acknowledged with **GDPR** (General Data Protection Regulation) and **IPR** (Intellectual Property Right)
- **If in doubt:** Talk to your institutional responsible



Tony Grist, The Devil's Fireplace, detail: the lawyer CCO
https://commons.wikimedia.org/wiki/File:The_Devil%27s_Fireplace,_detail-_the_lawyer.jpg

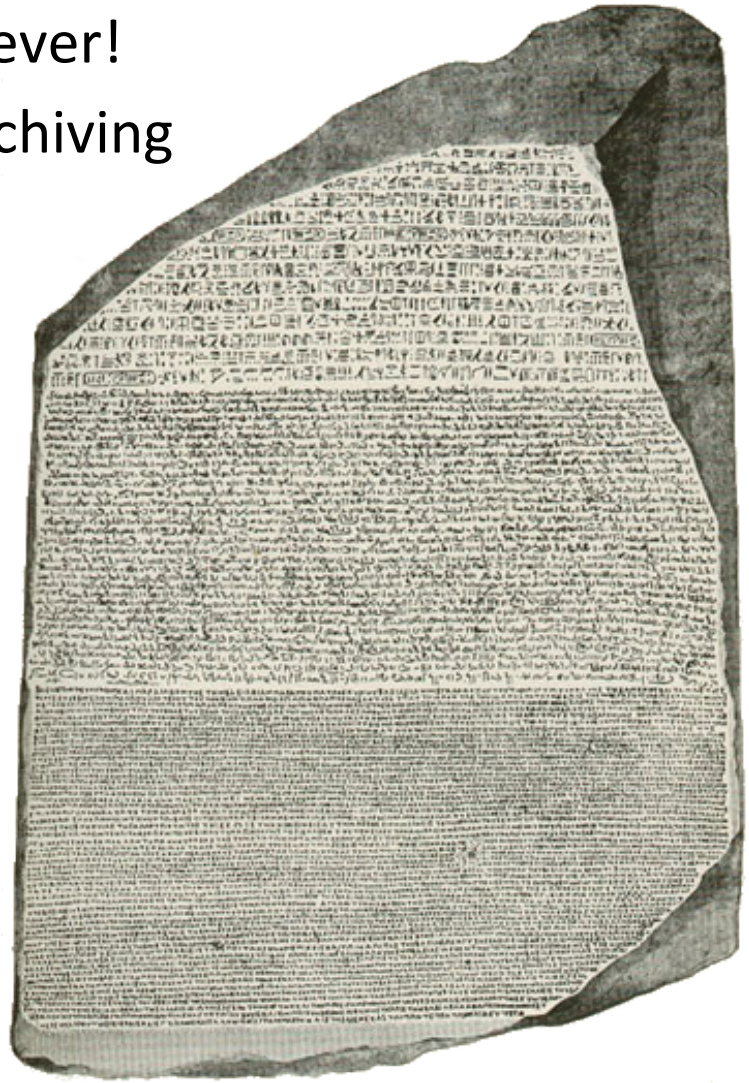
Storage and Backup

- Most storage media have only **limited time span** (PC, mobile storage media)
- Not all storage media are “safe”
- Some storage media are **prone to easy loss**
- **Institutional storage** vs. external storage (cloud, e.g. Drop-Box, safety issues, often addressed in institutional policies)
- **Back Up** needs to be planned, done, and checked!

Photo by Kev Seto on Unsplash

Selection and Preservation

- Not everything has to be kept forever!
- Not all formats are suitable for archiving
- Formulate your **requirements** for long term preservation (volume, certificate, costs, access rights, sustainabity)



Rosetta Stone by Unknown, CCO
https://commons.wikimedia.org/wiki/Rosetta_Stone#/media/File:RosettaStone.png

Data Sharing & Publishing

➽ Dropbox, your website, Research Gate/Academia.edu are not OA repositories!

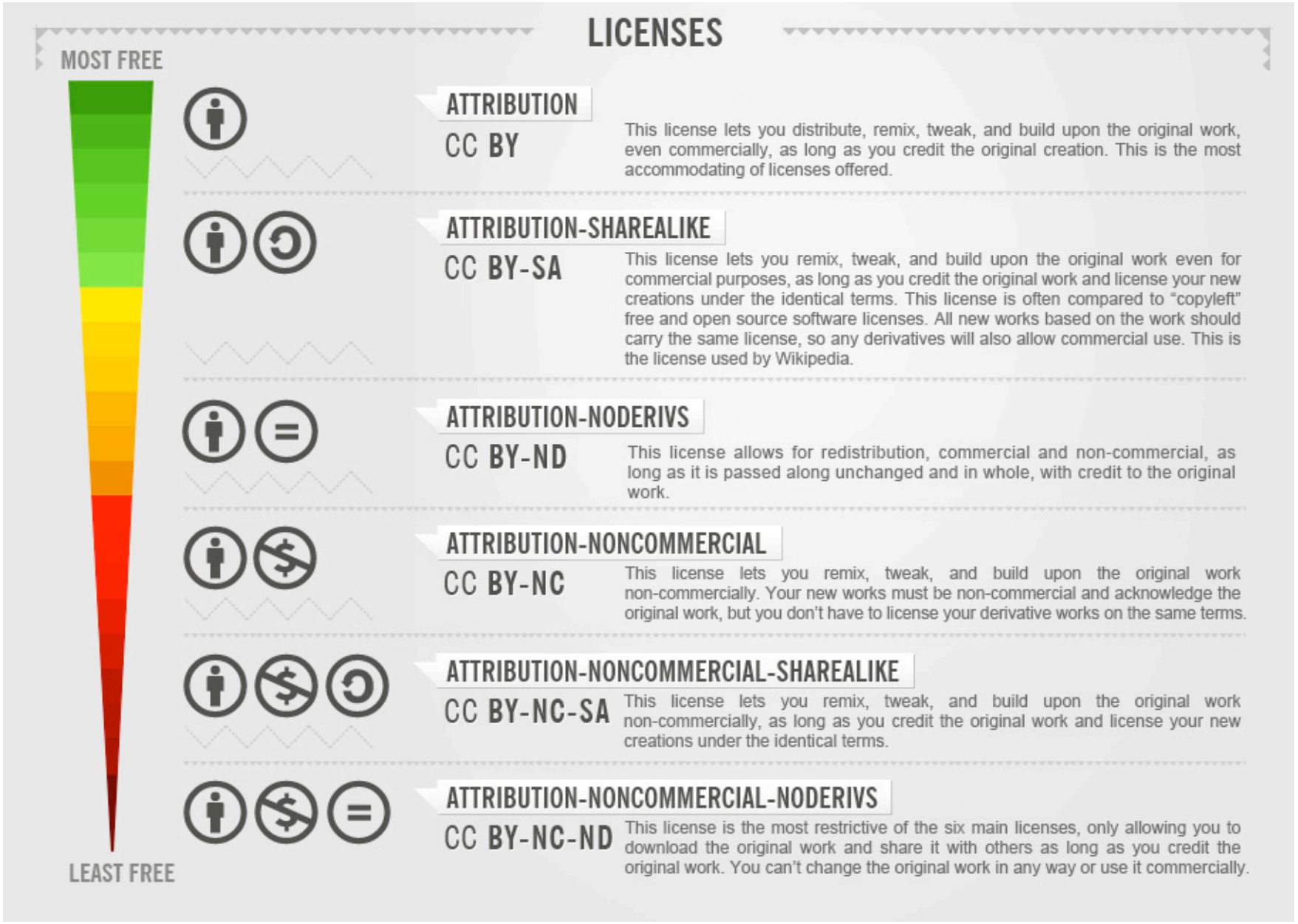
➽ **Data Publication** (Data Supplement to Article, Repository, Data Journals, e. g.
Research Data Journal for the Humanities and Social Sciences

➽ Make use of discipline specific, institutional or European **repositories** to deposit data/publications, generic **Zenodo**, for **humanities** e.g.:

- **HAL**, **DARIAH-DE**, **CLARIN**, **GESIS**

➽ Use tools to **register research data and/or** to **find a repository: re3data** & **Directory of Open Access Repositories**

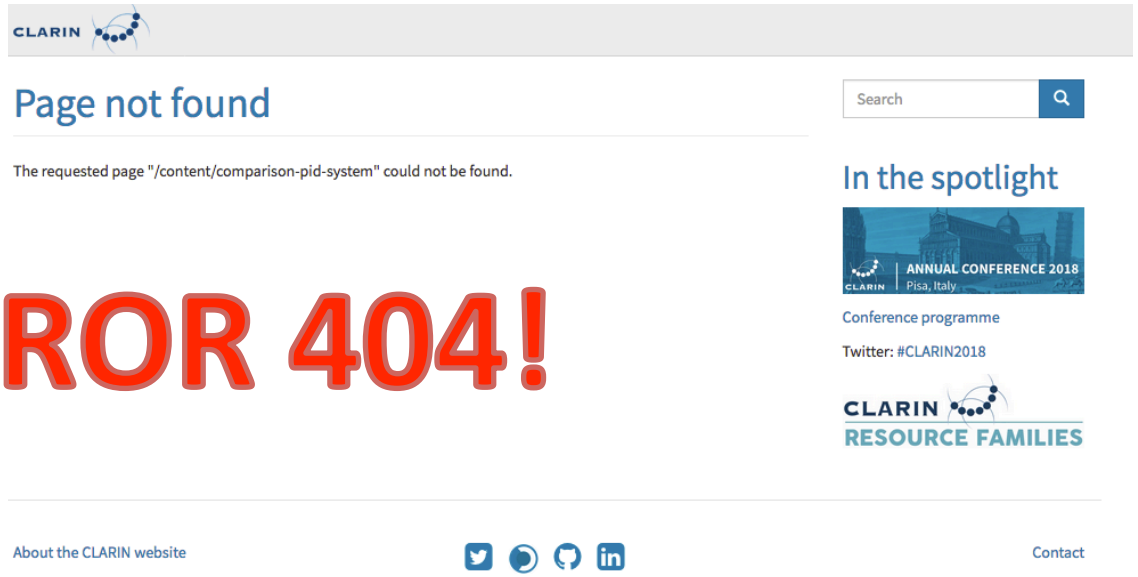
➽ Use **free licences** (e. g. Creative Commons Licenses)



"CC License Freedom Scale Chart" by Foter, https://commons.wikimedia.org/wiki/File:CC_License_Freedom_Scale_Chart.png, This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

Data Sharing & Publishing

- ➽ Additional value of **Persistent Identifiers (e.g. DOI and ORCID)**
- ✓ **Slayer of the Error 404 message & Champion of Linked Open Data**
- PIDs organisations: **Crossref**, **DataCite** and **ORCID**
- example ORCID: https://orcid.org/0000-0002-8217-4025



AVOID ERROR 404!

https://www.clarin.eu/content/comparison-pid-system

Responsibilities and Resources

- **Who** will be responsible for Data Management?
- Plan **resources** during and after the project
- **Review** and **update** your Data Management Plan regularly

Photo by Helloquence on Unsplash

The FAIR Principles and the Humanities

- **F**indable = DESCRIBE (rich Metadata!), e.g.
  - Persistent Identifiers
  - Appropriate metadata schema
  - Make metadata available even if data are not
- **A**ccessible = OPEN, e.g.
  - Trustworthy Repository
  - State Accessibility (open licences if possible)
- **I**nteroperable = LINK, e.g.
  - Establish well documented machine actionable APIs
  - Well documented and defined vocabularies and metadata models
  - Future-proof file formats
- **R**eusable = PUBLISH, e.g.
  - License for reuse
  - Documentation

Sources:

- EPFL Library Research Data Management FAST Guides: https://researchdata.epfl.ch/wp-content/uploads/EPFL_Library_RDM_FastGuide_All.pdf
- PARTHENOS Guidelines to FAIRify data management: https://doi.org/10.5281/zenodo.2668479

Playful Exercise 6

DATA CITATION

If you like it, cite it!

Data Citation:

How do I cite source materials with a CC-License?

- ✓ What do you think of this citation?
- ✓ What would you change?
- ✓ How do you find relevant information?
- ✓ Discuss in groups and collect results!

Data Citation:

How do I cite source materials with a CC-License?

Steps (for pictures, adapt for other materials):

- ✓ Identify source image (Google Reverse Picture Search, etc.)
- ✓ Check reuse information (license) > Are you allowed to use it?
- ✓ Cite (if possible) creator & title
- ✓ Include link to source image and license information
- ✓ For pictures from Wikipedia/Wikimedia you may use: https://www.lizenzhinweisgenerator.de

- **Basic citation (suggestion): Wilhelm Busch [Public domain], via Wikimedia Commons, Source: https://commons.wikimedia.org/wiki/File:L%C3%A4mpel.jpg**

- **More extensive citation (suggestion): Wilhelm Busch, Lehrer Lämpel (aus Max und Moritz), marked as Public Domain (CCO 1.0), via Wikimedia Commons, Source: https://commons.wikimedia.org/wiki/File:L%C3%A4mpel.jpg**

Question Time

09

WRAP UP

Your Data Is Powerful. If You Make It FAIR!

Start Today!

What's in there for You? Benefits of Open Research Practices and RDM for Researchers

- Open Publications get **more citations** and gain **higher media attention**
- Higher chances for **research collaborations**
- Better **job and funding opportunities**
- Higher (team) **effectiveness** and **sustainability**
- Stand in for your open ideals

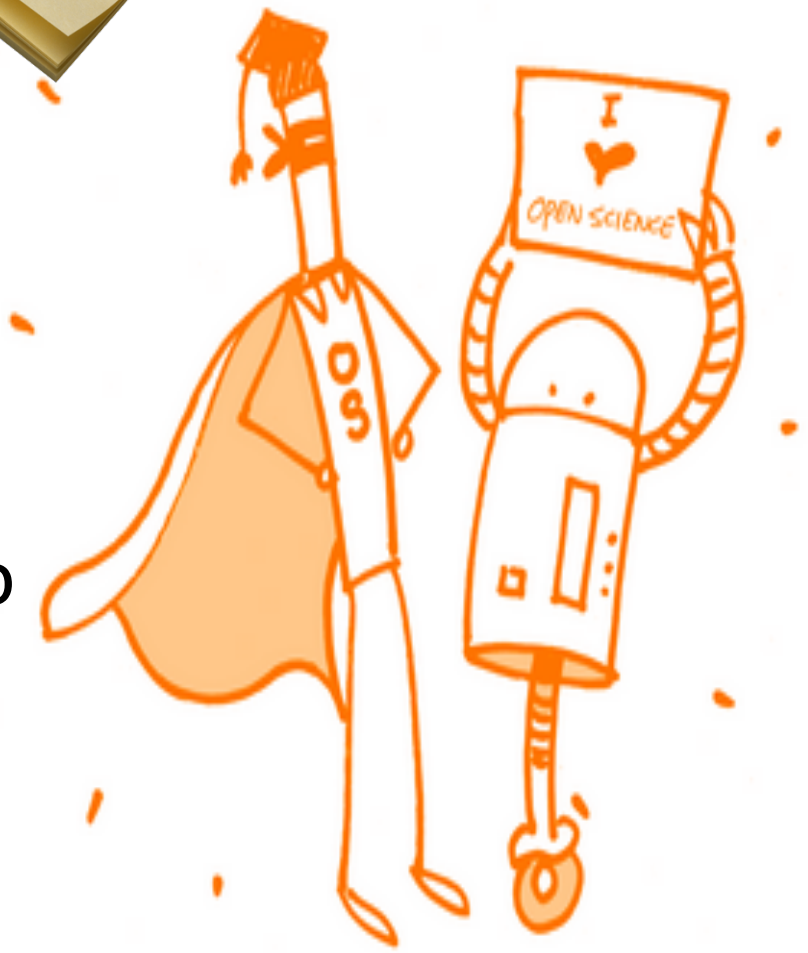
Picture: https://zenodo.org/record/1285575#.W09yZH59jOR
(Melanie Imming, John Tennant, CC0)

- ✓ **RDM is an integral part of Open Science and of Good Scientific Conduct and has many benefits**
- ✓ **Practice Open Access to Data and RDM early and be prepared for the future!**

Read more? McKiernan, E. C. et al. (2016). How open science helps researchers succeed. *eLife*, *5*, e16800. http://doi.org/10.7554/eLife.16800

Your Next Steps

- Your own Data Management Plan!
- Publish Data!
- Start a discussion about Open Access to Research Data at Your institution!



Source Picture: From “The Open Science Training Handbook” v.1.0, 2018, https://book.fosteropenscience.eu/

The End! Feedback!

Picture: Michal Klajban CC BY-SA 4.0
https://upload.wikimedia.org/wikipedia/commons/7/7c/A_pier_at_a_campsite_during_sunset%2C_Sidney_Spit_%28part_of_Gulf_Islands_National_Park_Reserve%29%2C_Sidney_Island%2C_British_Columbia%2C_Canada_20.jpg

Trainer CONTACT:

Dr. Ulrike Wuttke

University of Applied Sciences Potsdam (FHP)

wuttke@fh-potsdam.de

@UWuttke

RDMO:

- **Project Website:** https://rdmorganiser.github.io/
- **Try Out:** https://rdmo.aip.de/
- **Open Source Code:** https://github.com/rdmorganiser
- **Twitter:** https://twitter.com/rdmorganiser
- **Public Mailing-List:** https://www.listserv.dfn.de/sympa/info/rdmo

10

FURTHER LEARNING: OPEN SCIENCE / RESEARCH DATA MANAGEMENT / WORK FLOWS / SERVICES

Further Learning

Open Science in General:

- **FOSTER Open Science Module**
  - https://www.fosteropenscience.eu/learning/what-is-open-science
- **Open Science MOOC (under development)**
  - https://opensciencemooc.github.io/site/
- **TU Delft Open Science MOOC (started October 30, 2018)**
  - https://online-learning.tudelft.nl/courses/open-science-sharing-your-research-with-the-world/
- **Innovations in Scholarly Communication (Bianca Kramer & Jeroen Bosman)**
  - https://101innovations.wordpress.com/
- **Helmholtz Open Science Webinars**
  - https://os.helmholtz.de/bewusstsein-schaerfen/workshops/webinare/
- **European Union Open Science Resources**
  - https://ec.europa.eu/research/openscience/index.cfm

Further Learning

FAIR Principles and Open Access to Data

- **Wilkinson, Mark D. et al. 2016, The FAIR Guiding Principles for Scientific Data Management and Stewardship, in: *Scientific Data*, Nr. 3.**

➢ https://doi.org/10.1038/sdata.2016.18

- **Explanation of FAIR principles by Swiss National Science Foundation (SNF) (eng.)**

➢ http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

- **Explanation of FAIR principles in German (TIB Blog, Angelika Kraft)**

➢ https://blogs.tib.eu/wp/tib/2017/09/12/die-fair-data-prinzipien-fuer-forschungsdaten/

- **Book: Mons, Barend, *Data Stewardship for Open Science: Implementing FAIR Principles*, 2018**
- **Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities**

➢ https://openaccess.mpg.de/Berliner-Erklaerung

Further Learning

- **Carrara, Wendy et al., Open Data Goldbook for Data Managers and Data Holders, European Commission, 2018 (CC BY)**
  - https://www.europeandataportal.eu/sites/default/files/goldbook.pdf
- **European Data Portal Open Data Training Companion**
  - https://www.europeandataportal.eu/en/resources/training-companion
- **Plan S and Coalition S**
  - https://www.coalition-s.org/
- **DARIAH‘s position on PlanS**
  - https://www.dariah.eu/2018/10/25/towards-a-planhss-dariahs-position-on-plans/
- **FORCE11 Guidelines for Data Citation**
  - https://www.force11.org/datacitationprinciples
- **Foster Course: Use Open Data in Teaching**
  - https://www.fosteropenscience.eu/node/2656

Further Learning

Research Data Management

- **PARTHENOS Module "Manage, Improve and Open Up Your Research Data" (eHeritage and eHumanities)**

➢ http://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/

- **FOSTER Module on Data Management**

➢ https://www.fosteropenscience.eu/node/2328

- **Ulrike Wuttke. (2018, November). Introduction to Humanities Research Data Management. Zenodo.**

➢ http://doi.org/10.5281/zenodo.1491250

- **PARTHENOS Submodule "Research Impact"**

➢ http://training.parthenos-project.eu/sample-page/intro-to-ri/research-impact/

- **OSODOS Open Science Training Handbook (Open Science, Open Data, Open Source)**

➢ http://osodos.org; https://pfern.github.io/OSODOS/gitbook/

Further Learning

- **Research Data Management Promotional Material**
  - https://rdmpromotion.rbind.io/
- **PARTHENOS Guidelines to FAIRify data management:**
  - https://doi.org/10.5281/zenodo.2668479
- Holmstrand, K.F., den Boer, S.P.A., Vlachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.) (2019). **Research Data Management** (eLearning course, *Danish National Forum for Data Management*).
  - https://vidensportal.deic.dk/RDMelearn (doi: 10.11581/dtu:00000047)
- **EPFL Library Research Data Management Fast Guides**
  - https://researchdata.epfl.ch/wp-content/uploads/EPFL_Library_RDM_FastGuide_All.pdf

Further Learning

Licensing / Legal Aspects

- **Kreutzer, Open Content – A Practical Guide to Using Creative Commons Licenses, 2014**

➢ https://irights.info/wp-content/uploads/2014/11/Open_Content_A_Practical_Guide_to_Using_Open_Content_Licences_web.pdf

- **ARDC, Research Data Rights Management Guide** (ARDC Guides), September 2018

➢ https://www.ands.org.au/guides/research-data-rights-management

- **CLARIN-D Language Resources Legal Issues Bibliography**

➢ https://www.clarin-d.net/de/legal-issues-bibleography

Networks and Organizations

- **Open Knowledge Foundation**
- https://okfn.org/
- **Research Data Alliance (RDA)**
- https://www.rd-alliance.org/
- **Generation R** (Open Science Discourse Platform)
- http://genr.eu
- **GO FAIR Initiative**
- https://www.go-fair.org/
- **Collections as Data**
- https://collectionsasdata.github.io/

Source Slide nr. 23 of: Stefan Schmunk, & Steven Krauwer. (2018, March). Slides from "e-Humanities and e-Heritage Research Infrastructures: Beyond tools" (PARTHENOS eHumanities and eHeritage Webinar, Thursday, 22.02.2018, 11:00 – 12:00 A.M. CET). Zenodo. http://doi.org/10.5281/zenodo.1203335

European RIs for SSH & CH

RIs set up under the auspices of ESFRI, each based on national consortia of universities, libraries, museums, archives etc.:

cessda Consortium of European Social Sciences Data Archives

CLARIN

DARIAH-EU

E-RIHS EUROPEAN RESEARCH INFRASTRUCTURE FOR HERITAGE SCIENCE

European Social Survey

SHARE SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE

In addition a number of past or ongoing EC supported Infrastructure Projects, such as

EHRI EUROPEAN HOLOCAUST RESEARCH INFRASTRUCTURE

ARIADNE www.ariadne-infrastructure.eu

CENDARI COLLABORATIVE EUROPEAN DIGITAL & ARCHIVAL RESOURCE INFRASTRUCTURE

CHARISMA

PARTHENOS Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

DCH RP DIGITAL CULTURAL HERITAGE ROADMAP FOR PRESERVATION

DMP Tools

- **DMPonline**
  - https://dmponline.dcc.ac.uk/
- **RDMO**
  - https://rdmorganiser.github.io/

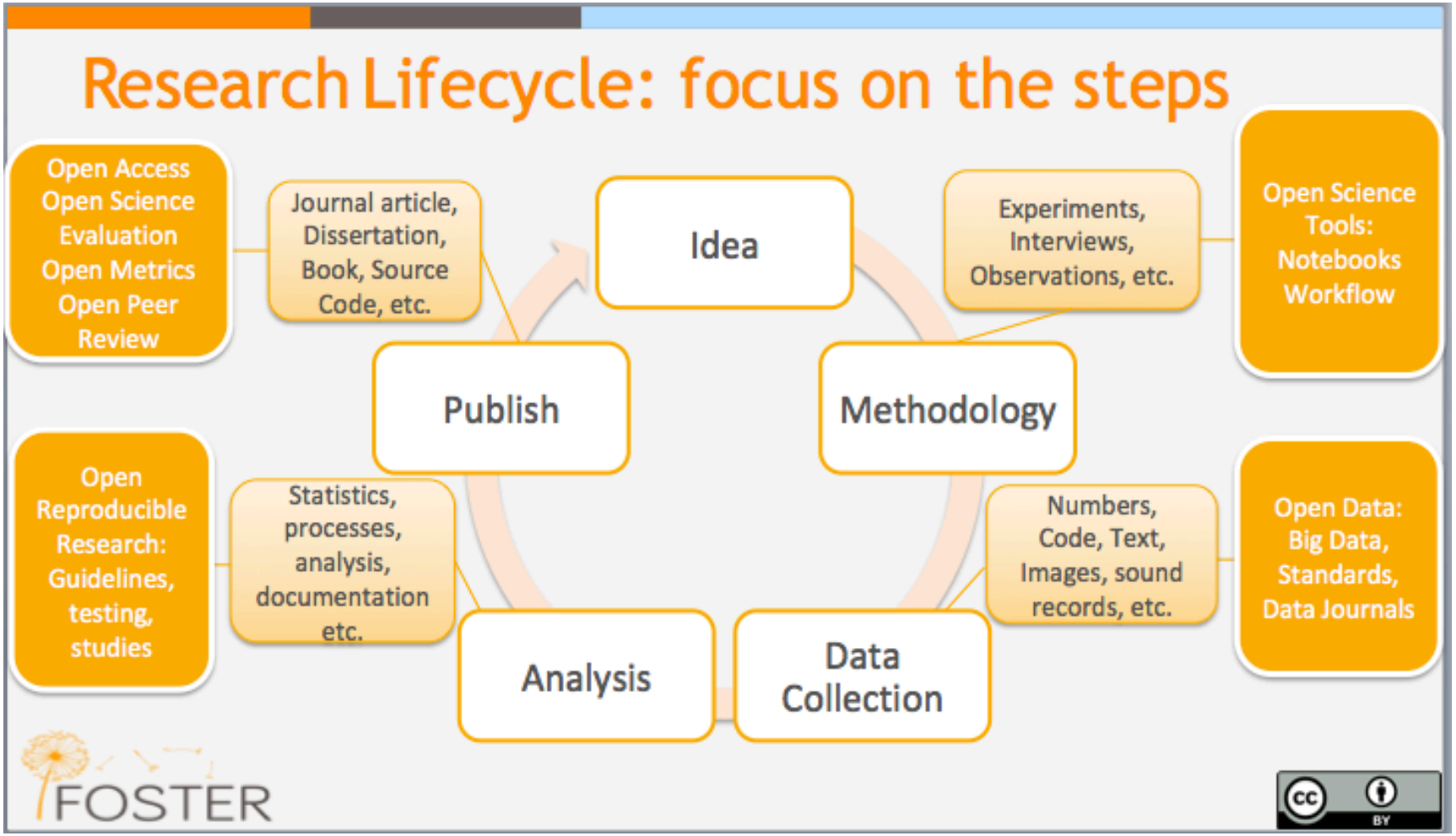
DMP Examples

- **NEH** (National Endowment for the Humanities) makes example grants available, including DMPs
  - ➢ under the various programs e.g. https://www.neh.gov/grants/odh/digital-humanities-advancement-grants
- **DMPOnline** Public DMPs
  - ➢ https://dmponline.dcc.ac.uk/public_plans

Advantages of Open Access

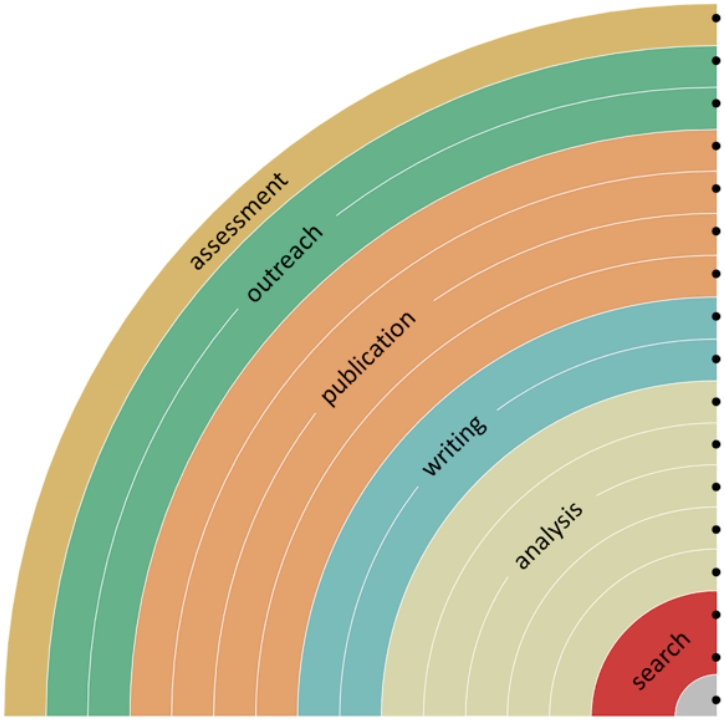
Source: https://kib.ki.se/en/publish-analyse/open-access

Open scholarly practices that can make your research more visible



Source: Slide from presentation “Open Science What, why and best practices in open research” Nancy Pontika, Foster Open Science Bootcamp Barcelona, 18.04.2018

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

DOI: 10.5281/zenodo.1147025

Data Sharing & Publishing

Publishing high-quality Open Data requires some effort. The W3C Foundation has created a basic model for Open Data with regard to quality: the 5-Star Open Data model. The 5 stages of Open Data are:

★	Make your stuff available on the web (whatever format) under an open licence
★★	Make it available as structured data (e.g. Excel instead of image scan of a table)
★★★	Use non-proprietary formats (e.g. CSV instead of Excel)
★★★★	Use URIs to denote things, so that people can point at your stuff
★★★★★	Link your data to other data to provide context

Table 2: Descriptions of all stages of the 5-star Open Data Model

Source picture: Carrara, Wendy et al., Open Data Goldbook for Data Managers and Data Holders, European Commission, 2018 (CC BY), p. 50. Download: https://www.europeandataportal.eu/sites/default/files/goldbook.pdf

The FAIR Principles (1/2)

- **F**indability :
  - F1. (Meta)data are assigned a globally unique and persistent identifier
  - F2. Data are described with rich metadata
  - F3. Metadata clearly and explicitly include the identifier of the data they describe
  - F4. (Meta)data are registered or indexed in a searchable resource
- **A**ccessibility
  - A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
    - A1.1 The protocol is open, free, and universally implementable
    - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
  - A2. Metadata are accessible, even when the data are no longer available

Source: https://www.force11.org/group/fairgroup/fairprinciples

The FAIR Principles (2/2)

- **I**nteroperability
  - I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
  - I2. (Meta)data use vocabularies that follow FAIR principles
  - I3. (Meta)data include qualified references to other (meta)data
- **R**euse
  - R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
    - R1.1. (Meta)data are released with a clear and accessible data usage license
    - R1.2. (Meta)data are associated with detailed provenance
    - R1.3. (Meta)data meet domain-relevant community standards

Source: https://www.force11.org/group/fairgroup/fairprinciples

The GO FAIR Initiative

GO **FAIR** initiative - practical implementation of the European Open Science Cloud (EOSC):

"… guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data."

Source: https://www.go-fair.org/fair-principles/