

**15.11.2018** | University of Luxemburg

Luxemburg Open Science Forum 2018, 14.-16.11.2018

<https://openscience2018.uni.lu/>



PARTHENOS

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



Fachhochschule Potsdam  
University of  
Applied Sciences

# Introduction to Humanities Research Data Management

Ulrike Wuttke & Jochen Klar

*Morning Session (Ulrike Wuttke)*

(v\_1.0)

DOI: 10.5281/zenodo.1491250

[Ulrike Wuttke](#), University of Applied Sciences Potsdam / PARTHENOS

@PARTHENOS\_EU @UWuttke | [CC-BY 4.0](#) | [PARTHENOS](#)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654119

Unless otherwise stated the content of these slides is under the license CC-BY 4.0

# TABLE OF CONTENTS

- 1) Warm Up: Introduce yourself
- 2) Housekeeping/Code of Conduct
- 3) Rationales and Benefits of the Workshop
- 4) Open Science / Open Scholarship
- 5) Research Data in Humanities and Heritage Science
- 6) Basic principles of Research Data Management
- 7) The individual parts of a Data Management Plan
- 8) Further Learning

01

**WARM UP**

# Introduce yourself (Twitter Style)

Example 1:

- Ulrike Wuttke (@UWuttke)
- Medieval Studies / Information Science
- University of Applied Sciences Potsdam (FHP)
- Teach you how to make your data future proof and FAIR



**Ulrike Wuttke** @UWuttke · 1 Std.

Do you want to learn how to make your humanities research data future proof, #FAIR and learn more about #openscience? Join my session with @jochenklar at #LuxOSF tomorrow! We prepared loads of fun (with Penguins!) @fosterscience @OpenAIRE\_eu @Parthenos\_EU @rdmorganiser @uni\_lu

Tweet übersetzen



**Now It's Your Turn!**

🗨️ ↻ 1 ❤️ 6 📺

# Introduce yourself (Twitter Style)

- Name (Twitter handle)
- Discipline
- Function/Affiliation
- Motivation/Expectation for this Workshop

02

# **HOUSEKEEPING / CODE OF CONDUCT**

# Our Code of Conduct

- Respect for each other
- There are no stupid questions
- Everyone is an expert in something
- Connect
- Share

03

# **RATIONALES AND BENEFITS OF THE WORKSHOP**

- Two sessions
  - Morning Session: Theory (Ulrike Wuttke)
  - Afternoon Session: Hands-On (Jochen Klar)
- Access to materials (Slides) via Zenodo

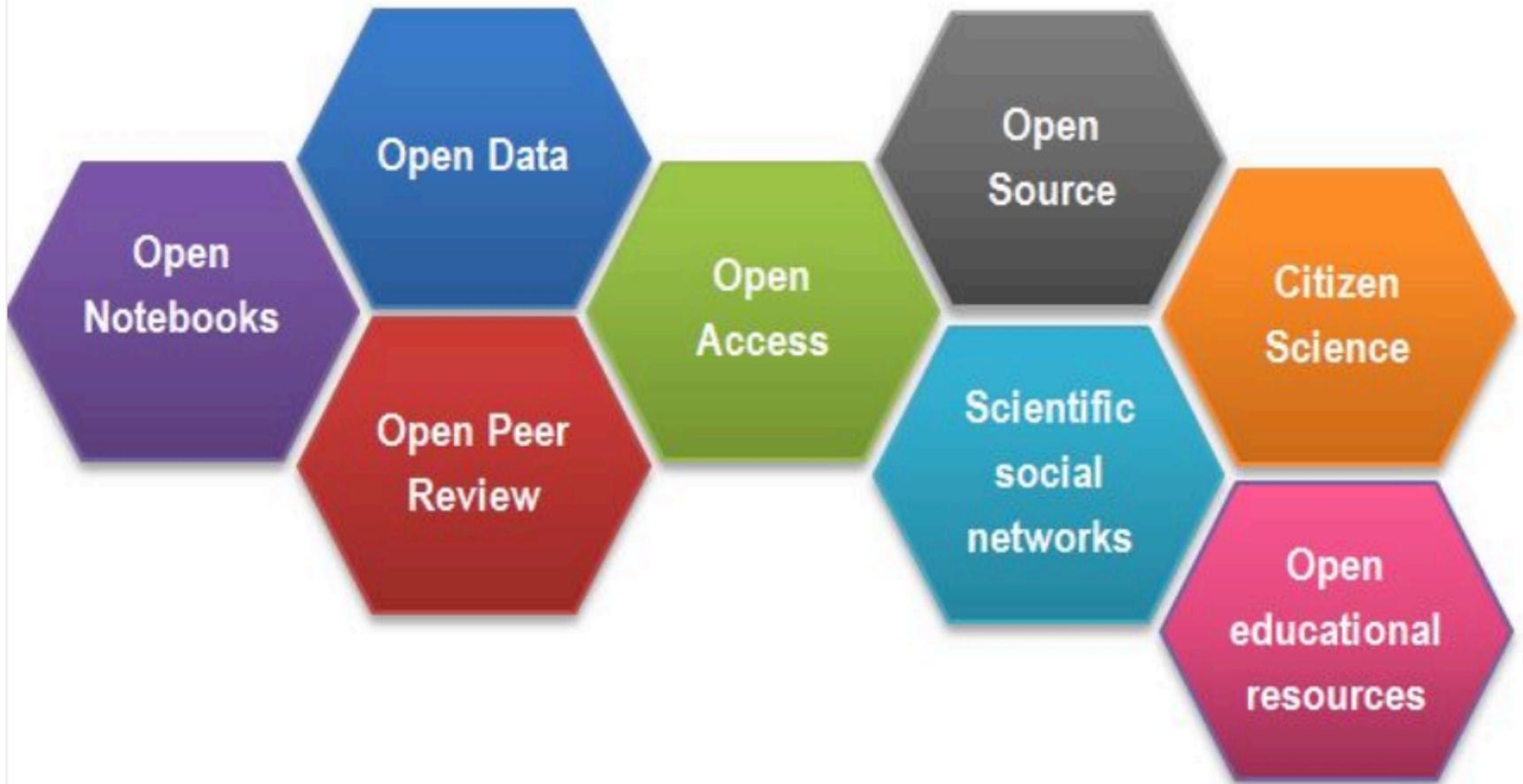
04

**OPEN SCIENCE – OPEN  
SCHOLARSHIP**

# What is “Open Science”?

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

# Open Science has many facets



# 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)

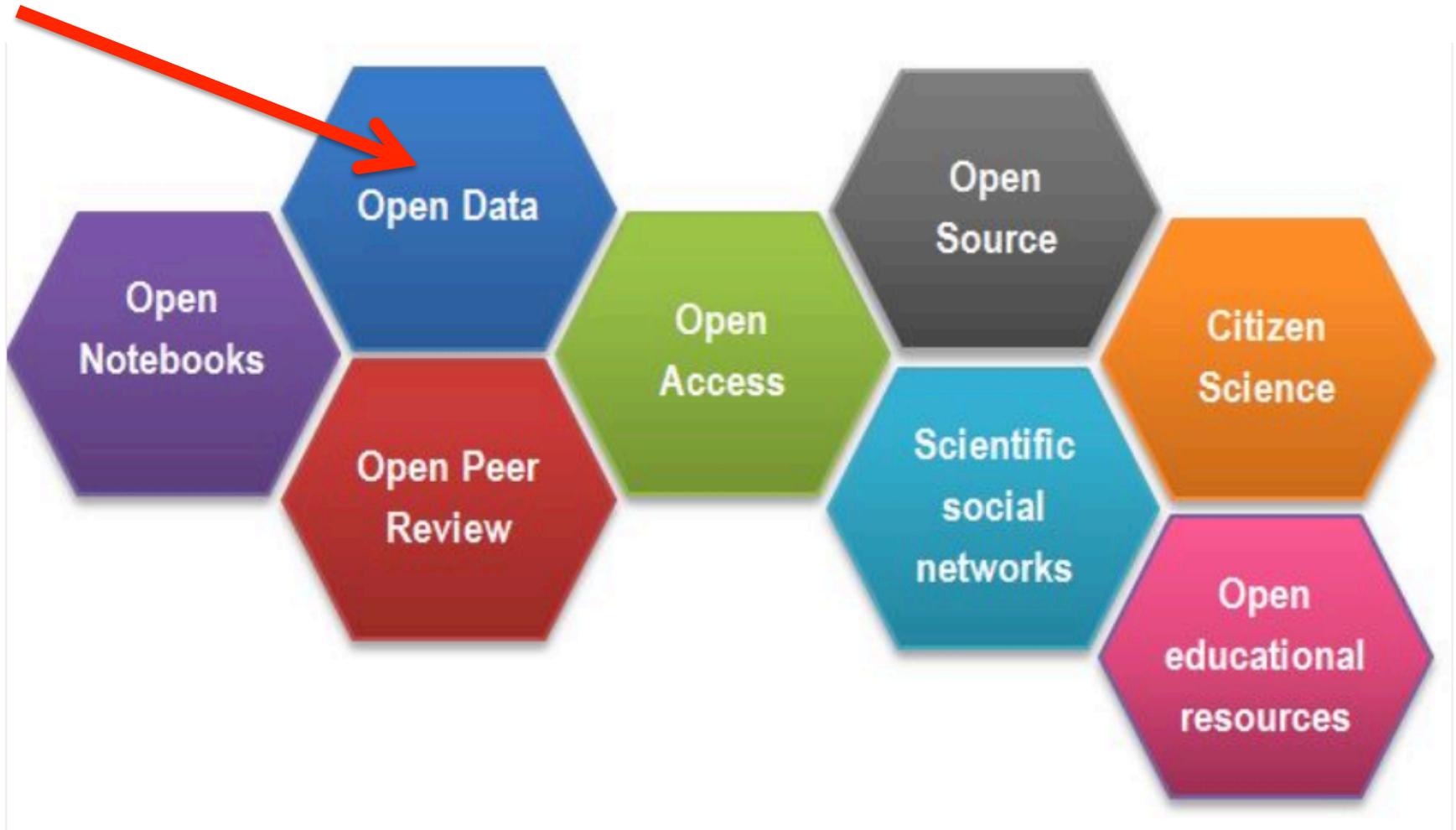
# Advantages of Open Science for research(ers)

- Higher **reproducibility** of research findings
- Higher **transparency** of research methods and evaluation
- Researchers and research institutions **save money and time**
- Higher (societal) **impact** of research(ers)
- Open Science gets research(ers) **out of the Ivory Towers!**



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

# Open Science has many facets



# Key Concept: Open Data

What is it about?

- Open Data = (research) data that is freely available for (re)use and republish for everyone
- Ideal: Data with no restrictions from copyright, patents, or other control mechanisms > transparent results
- However: “as open as possible, as closed as necessary”

What does Open Data involve?

- Sharing is not giving away, to work in an open environment benefits all of us
- Poses challenges, is discipline specific > e.g. Humanities
- Essential: Data Management Planning

05

**RESEARCH DATA IN HUMANITIES  
AND HERITAGE SCIENCE**

# eHumanities and eHeritage Research

## What is it about?

- Computers, the internet, and big data, led to a rise of **quantitative** and **statistical methods** in the Humanities and CH
- digital workflows & digital methods

## 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
- Access (copyright and license issues)
- Sustainability (Data loss)
- Lack of documentation and standardization
- **Interoperability** (machine actionability) and **Reuse** (culture of sharing)
- eHumanities and eHeritage are based on accessible, correct, authoritative, well structured data

# 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
- Research Data are data that are produced in and used in scientific 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, etc.
- most “sources” are research data and their management has in fact always been part of the scientific process, digitization only adds complexity
- digitized sources, 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)
- But: Many issues and solutions apply to the broader field (and beyond Humanities and Heritage Science!)

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

- Opening up the data could lead to many **opportunities** for using and reusing it, for collaborating, informing and increasing the impact of the work (contemporary issues, interdisciplinary research, engaging broader society)
- **Funder requirements** on national and international level (e.g. European Commission) = Research Data Management and Open Science
- **Research Data Policies** (institutional, journals)

# Good Data Management is good for you!



**IMC\_Leeds**

@IMC\_Leeds

71% of you got this right. Quite a few pairs of sunglasses, but we have a \*lot\* of USB sticks! Let us know if you think you may have lost yours on campus.

**IMC\_Leeds** @IMC\_Leeds

What do you think was the most common item of lost property handed in at #IMC2018?  
Get in touch if you're missing something!

5:04pm · 11 Jul 2018 · Twitter Web Client

Link to original tweet: [https://twitter.com/IMC\\_Leeds/status/1017062144280588290](https://twitter.com/IMC_Leeds/status/1017062144280588290)



**Bastien Conan**

@Bastien\_Conan

Folgen



Fantastic talk from [@martateperek](#): “the main obstacles to data management and sharing are cultural, not technological” [#scidata18](#)

Tweet übersetzen

11:49 - 14. Nov. 2018

7 Retweets 18 „Gefällt mir“-Angaben



[https://twitter.com/Bastien\\_Conan/status/1062658855203753984](https://twitter.com/Bastien_Conan/status/1062658855203753984)

# Future proof Research Data Management: Let's go!

- 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.



Picture: Road Sign by Free Images  
([www.inkmedia.com](http://www.inkmedia.com)), CC BY 2.0  
<https://flic.kr/p/JoVNHU>

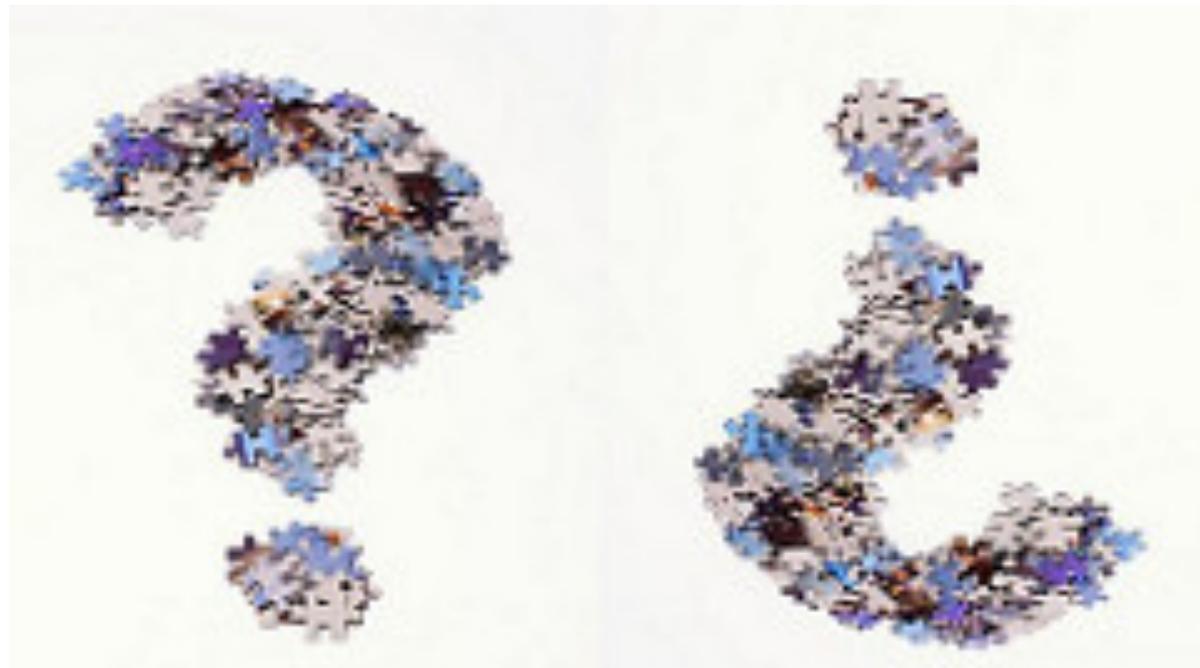
Translated (UW) from: AG Forschungsdaten der Schwerpunktinitiative "Digitale Information" der Allianz der deutschen Wissenschaftsorganisationen, Forschungsdatenmanagement: Eine Handreichung, 2018, p. 4. Online: <http://doi.org/10.2312/allianz0a.029> (CC BY 4.0)

# Playful Exercise 1

## **RESEARCH DATA LIFECYCLE**

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

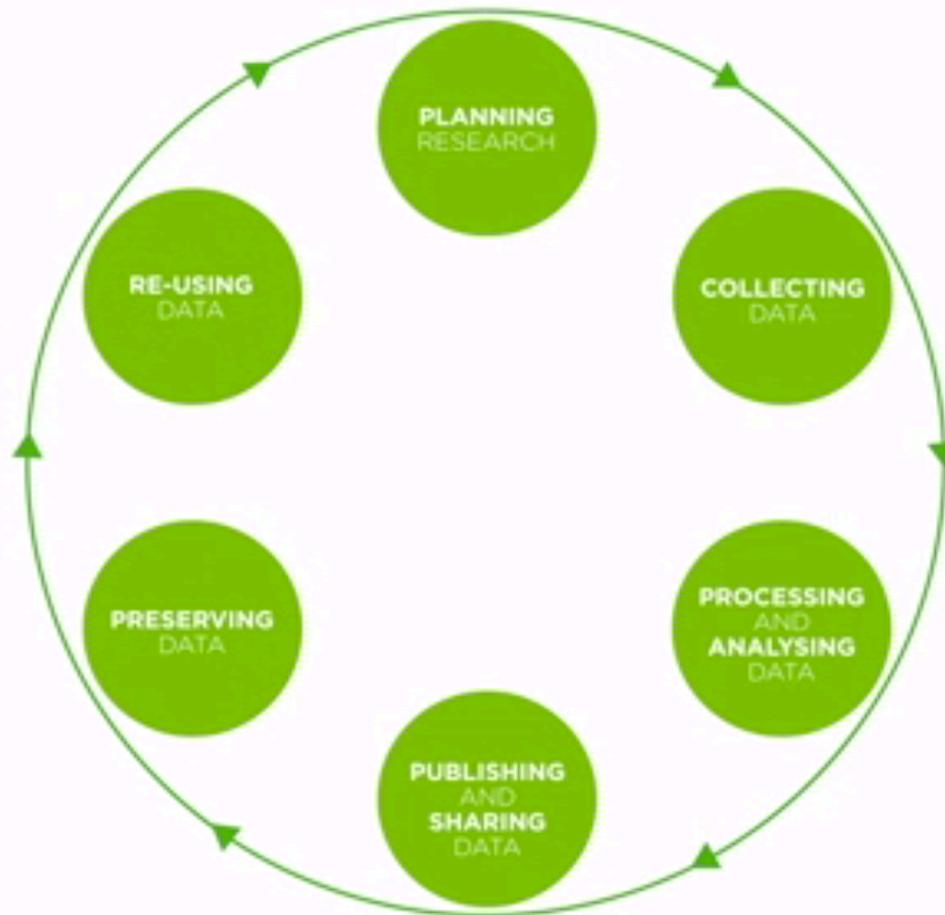
- Discuss in groups and use the cards. (3 min.)
- Online Version: Try the Drag-and-Drop Exercise:  
<http://training.parthenos-project.eu/sample-page/ehumanities-eheritage-webinar-series/webinar-work-with-research-infrastructures/wrap-up-materials/>



This exercise is adapted from: Biernacka, K.; Dolzycka, D.; Helbig, K.; Buchholz, P. 2018. Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement. DOI: 10.5281/zenodo.1215377 (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>

Picture: Questions by Roland O'Daniel CC BY-SA 2.0 (<https://creativecommons.org/licenses/by-nc-sa/2.0/>) from <https://flic.kr/p/7Ww83p>

# Solution: Research Data Lifecycle



## Research Data Lifecycle

from <https://www.ukdataservice.ac.uk/manage-data/lifecycle>

06

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

# Theory and Practice of Data Management: 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!)

# What is a Research Data Management Plan?

- DMP = 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

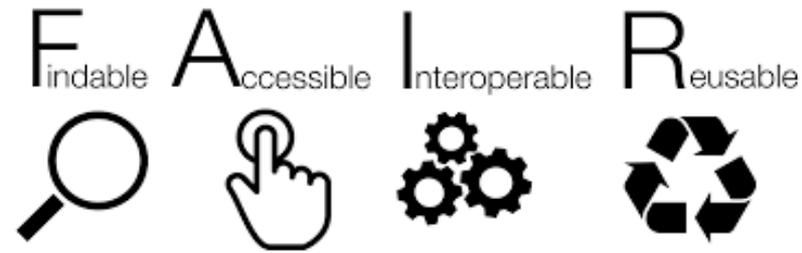
# What is a Research Data Management Plan?

- Basic Principles of RDM:
  - FAIR
  - Open (where possible)
  - The early bird catches the worm!  
RDM is not an afterthought!
  - DMP as living document



Picture: Rocks at [Vlychada Beach](#) in [Exomytis, Santorini](#), Greece, by Dietmar Rabich, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=63225571>

# The FAIR Principles



- The 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 (<https://www.force11.org/group/fairgroup/fairprinciples>)
  - Findable
  - Accessible
  - Interoperable
  - Reusable
- Note: Not all FAIR Data is Open Data

# The FAIR Principles

- 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.”

# It can get pretty complex, though...

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

- the **audio file** of the interview
- the **interview transcript** in the form of a digital text file
- the discussion guide or **questionnaire**, which explains the methodological approach and is necessary for the comprehensibility of the results of the study.
- the **project explanation** as well as the **declaration of consent** of the interviewee, which documents compliance with the legal provisions of the Federal and State Data Protection Act
- the **codebook**, which e.g. documents the development categories and variables used
- the **documentation** of the procedure for anonymization and pseudonymization
- the indexing information (**metadata**), which guarantees the citation ability of the interview and its findability

# What are YOUR Research Data?



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

This exercise is adapted from: Biernacka, K.; Dolzycka, D.; Helbig, K.; Buchholz, P. 2018. Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement. DOI: 10.5281/zenodo.1215377 (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>

Picture: Thinking statues taken by Rui Fernandes, CC-BY 2.0 (<https://creativecommons.org/licenses/by/2.0/>), <https://flic.kr/p/8WpM2U>

# Break



Picture: Manchots empereurs tobogannent by Samuel Blanc

[https://commons.wikimedia.org/wiki/Spheniscidae#/media/File:Manchots\\_empereurs\\_tobogannent.JPG](https://commons.wikimedia.org/wiki/Spheniscidae#/media/File:Manchots_empereurs_tobogannent.JPG), CC BY SA 3.0

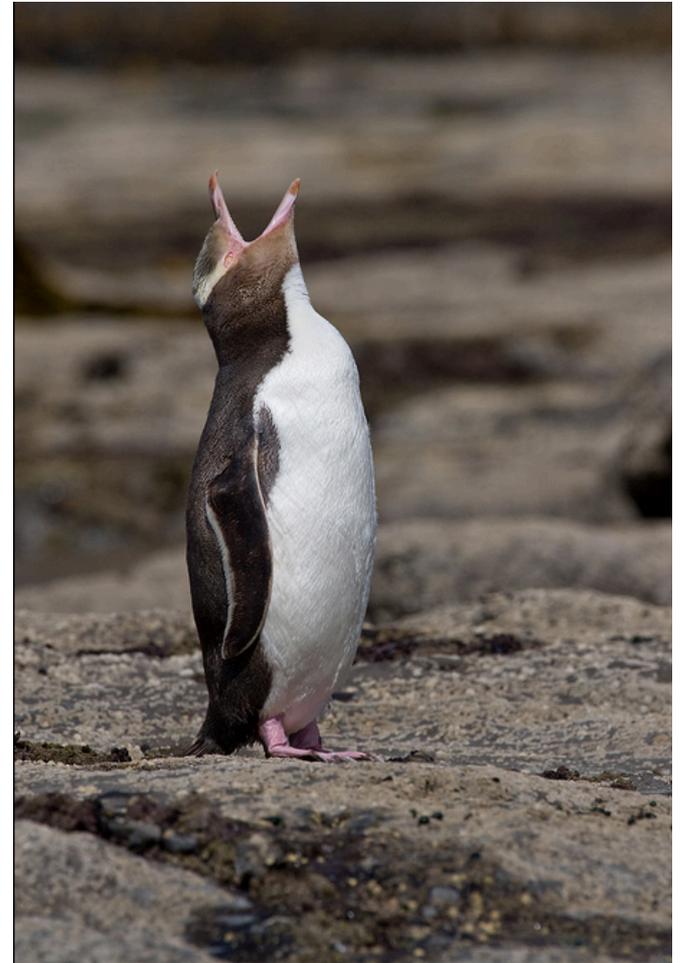
07

# **THE INDIVIDUAL PARTS OF A DATA MANAGEMENT PLAN**

# 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



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](https://commons.wikimedia.org/wiki/File:Yellow-eyed_Penguin_crying_MC.jpg)

# Data Collection

What data will you collect or create?

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

How will the data be collected or created?

- What standards or methodologies will you use?
- How will you structure and name your folders and files?
- How will you handle versioning?
- What quality assurance processes will you adopt?

# Data Collection

- **Avoid automatically generated filenames**, e.g. of digital cameras
- Think about **naming conventions** and the **order** of your files
- **Use open or standard formats**
- Remember Wordstar?!? **Archive format** will maybe be different from **working format**
- Thinking about these matter first and 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/>)
- If you like it, cite it! **Get acquainted with Data (and Software) Citation**

# Data Documentation and Metadata

What documentation and metadata will accompany the data?

- What information is needed for the data to be to be read and interpreted in the future?
- How will you capture / create this documentation and metadata?
- What metadata standards will you use and why?



# Data Documentation and Metadata

➔ R1.3. (Meta)data meet **domain-relevant community standards**

➔ Metadata are “data about data”

➔ Metadata are a love letter to the future

➔ Metadata are used to describe and organize data: formal description and content description

➔ Standards for Metadata to enhance interoperability (disciplinary and generic)

➔ **Transparent Documentation** includes project description, aims, methods, data cleaning, versioning, etc. (= your detailed data management plan)



# Data Documentation and Metadata

## ➤ FAIR Humanities!

- **TEI** (Text Encoding Initiative): [www.tei-c.org](http://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/>

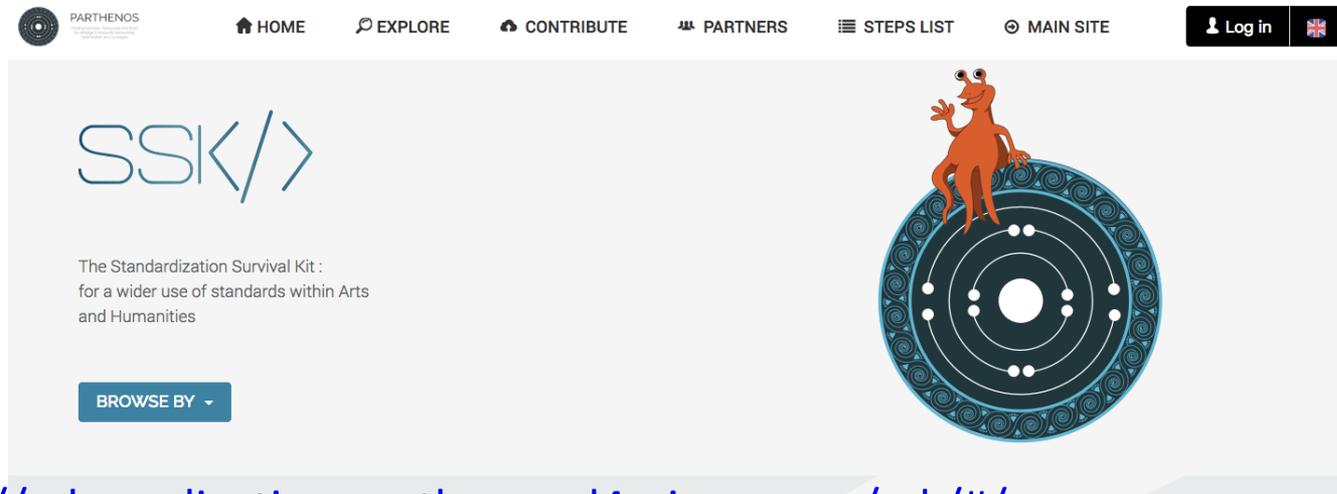
# Data Documentation and Metadata

- ✓ Avoid disambiguity
- ✓ Easy tagging
- ✓ Independent from spelling
- ➔ Linked data and unique identifiers: Use **authority files** (Normdaten)!
- ➔ BARTOC (Basel Register of Thesauri, Ontologies & Classifications): <https://bartoc.org/>
  - **GND** (German National Library)
  - **ORCID** (Open Researcher and Contributor ID)
  - **GeoNames**
  - **Wikidata**
  - **Getty Union List of Artist Names**
  - **VIAF** (Virtual International Authority File)

# Data Documentation and Metadata

## Standardization Survival Kit (SSK)

- Overlay platform developed by PARTHENOS dedicated to promoting a wider use of **standards** (TEI, Dublin Core, etc.) within the Arts and Humanities
- Aims:
  - Designed to support researchers in selecting and using the appropriate standards for their particular disciplines and work flows
  - Documentation of existing standards by providing reference materials
  - Foster the adoption of standards
  - Communication with research communities



# Ethics and Legal Compliance

How will you manage any ethical issues?

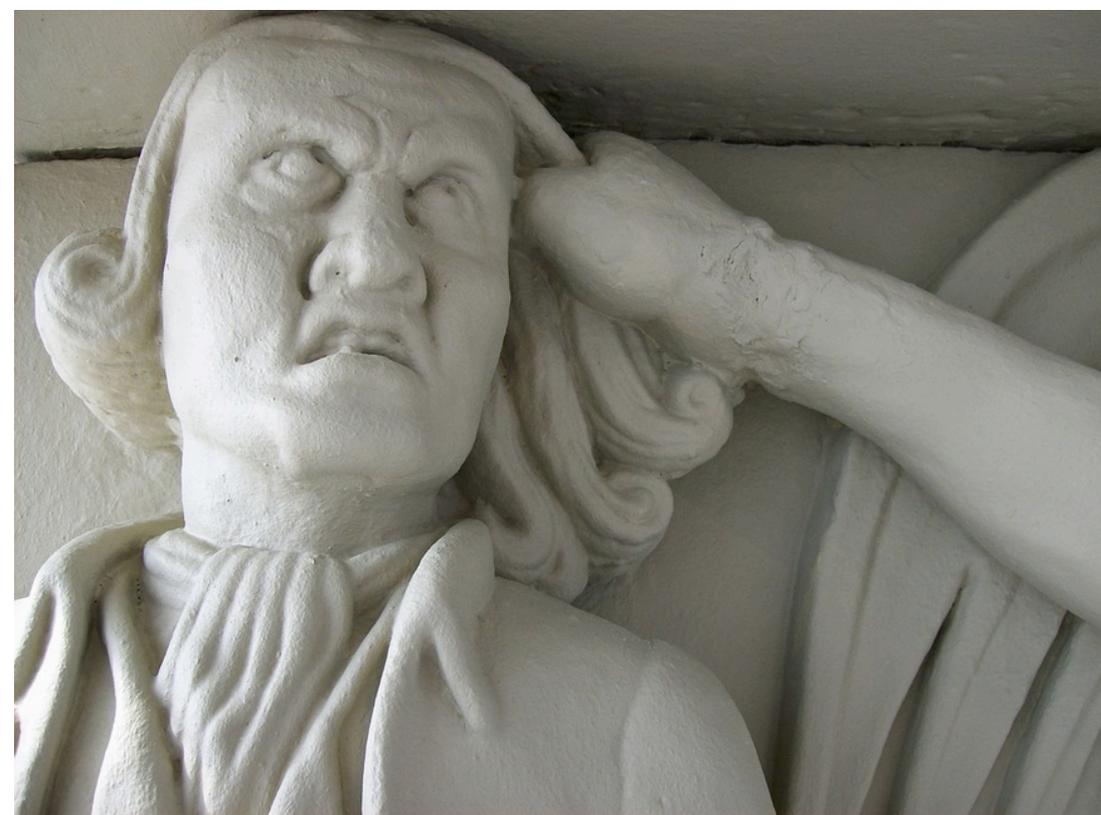
- Have you gained consent for data preservation and sharing?
- How will you protect the identity of participants if required? e.g. via anonymisation
- How will sensitive data be handled to ensure it is stored and transferred securely?

How will you manage copyright and Intellectual Property Rights (IPR) issues?

- Who owns the data?
- How will the data be licensed for reuse?
- Are there any restrictions on the reuse of third-party data?
- Will data sharing be postponed / restricted e.g. to publish or seek patents?

# 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)
- ➔ **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](https://commons.wikimedia.org/wiki/File:The_Devil%27s_Fireplace_detail_the_lawyer.jpg)

# Storage and Backup

How will the data be stored and backed up during the research?

- Do you have sufficient storage or will you need to include charges for additional services?
- How will the data be backed up?
- Who will be responsible for backup and recovery?
- How will the data be recovered in the event of an incident?

How will you manage access and security?

- What are the risks to data security and how will these be managed?
- How will you control access to keep the data secure?
- How will you ensure that collaborators can access your data securely?
- If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems?

# 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!

# Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

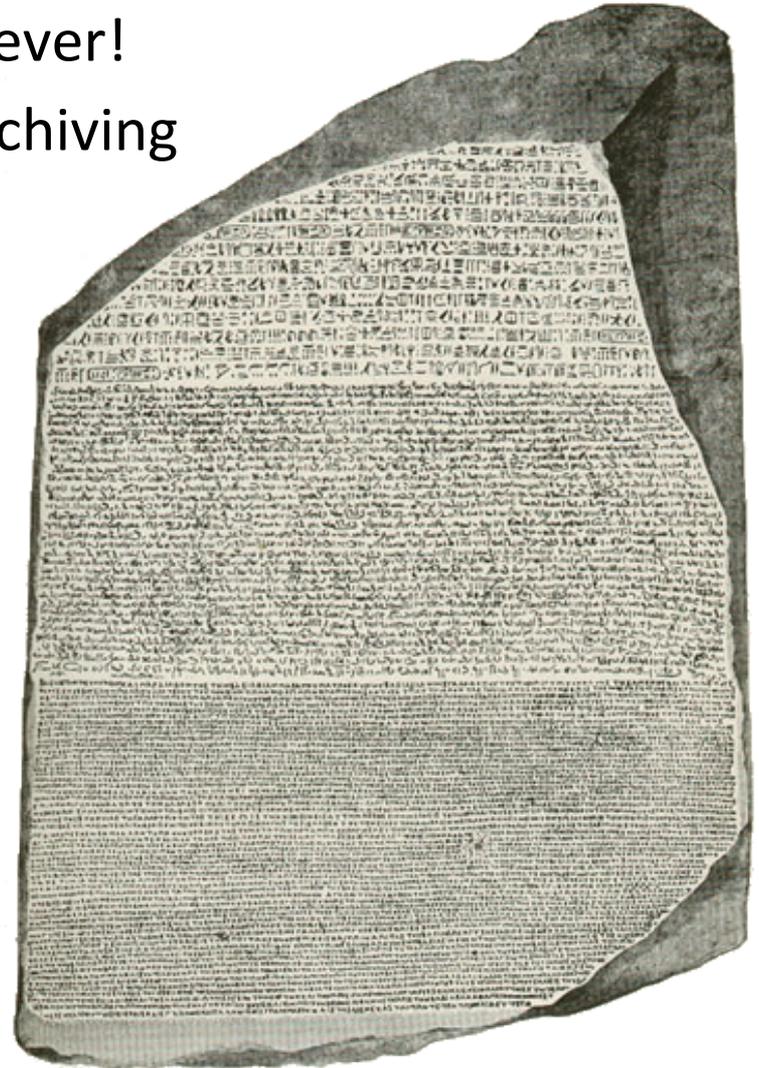
- What data must be retained/destroyed for contractual, legal, or regulatory purposes?
- How will you decide what other data to keep?
- What are the foreseeable research uses for the data?
- How long will the data be retained and preserved?

What is the long-term preservation plan for the dataset?

- Where e.g. in which repository or archive will the data be held?
- What costs if any will your selected data repository or archive charge?
- Have you costed in time and effort to prepare the data for sharing / preservation?

# 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, sustainability)



Rosetta Stone by Unknown, CCO

[https://commons.wikimedia.org/wiki/Rosetta\\_Stone#/media/File:RosettaStone.png](https://commons.wikimedia.org/wiki/Rosetta_Stone#/media/File:RosettaStone.png)

# Data Sharing

How will you share the data?

- How will potential users find out about your data?
- With whom will you share the data, and under what conditions?
- Will you share data via a **repository**, handle requests directly or use another mechanism?
- When will you make the data available?
- Will you pursue getting a persistent identifier for your data?

Are any restrictions on data sharing required?

- What action will you take to overcome or minimise restrictions?
- For how long do you need exclusive use of the data and why?
- Will a data sharing agreement (or equivalent) be required?

# Data Sharing

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

➔ **Do it right: Data Publication**

➔ **Data Journals** (Data Paper), **Data Supplementaries** to articles

– Research Data Journal for the Humanities and Social Sciences:

<https://brill.com/view/journals/rdj/rdj-overview.xml?rskey=2G8kx3&result=1>

– Journal of Open Archaeology Data: <http://openarchaeologydata.metajnl.com/>

➔ Make use of discipline specific, institutional or European **repositories** to deposit data/publications (e.g. Zenodo: <https://zenodo.org/>)

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

➔ Use tools to **register research data** (e.g. re3data:

<https://www.re3data.org/>) and to find a repository, for humanities data e.g.:

• DARIAH EU (<https://hal.archives-ouvertes.fr/>,

<https://de.dariah.eu/en/repository>)

• CLARIN (<https://www.clarin.eu/content/repositories>)

• GESIS ([www.gesis.org](http://www.gesis.org))

• IANUS ([www.ianus-fdz.de](http://www.ianus-fdz.de))

# Data Sharing

## ➤ Open Access Publishing

- Directory of Open Access Journals: <https://doaj.org>)
- Directory of Open Access Books: <https://www.doabooks.org/>)
- Directory of Open Access Repositories:  
<http://v2.sherpa.ac.uk/opensoar/>

## ➤ Publisher copyright policies & self-archiving:

<http://www.sherpa.ac.uk/romeo/index.php>

# Data Sharing

- ➔ Additional value of **Persistent Identifiers (e.g. DOI and ORCID)**
- ✓ Slayer of the Error 404 message & Champion of linked open data
- ✓ long-lasting, unambiguous reference to a digital object (journal article, dataset, scientific sample, artwork, PhD thesis, publication or person)
- PID takes you to a metadata record that contains information about a digital object or person (its current location for access or download)
- PIDs are stable: metadata of PID record can be updated (e.g. new location)
- PIDs organisations: **Crossref, DataCite** and **ORCID**
- example ORCID: <https://orcid.org/0000-0002-8217-4025>

# Responsibilities and Resources

Who will be responsible for data management?

- Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?
- Who will be responsible for each data management activity?
- How will responsibilities be split across partner sites in collaborative research projects?
- Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners?

What resources will you require to deliver your plan?

- Is additional specialist expertise (or training for existing staff) required?
- Do you require hardware or software which is additional or exceptional to existing institutional provision?
- Will charges be applied by data repositories?

# Responsibilities and Resources

- Plan resources during and after the project
- Review and update your data management plan regularly

# Luckily: You are not alone!

- RDM is **team** work!
- Use tools for **Data Management Planning** (e.g. DCC DMPOnline (<https://dmponline.dcc.ac.uk/>), RDMO (<https://rdmorganiser.github.io/en/>))
- Make use of **infrastructural support** (research infrastructures, cultural heritage institutions, libraries, data centres)
- Ecosystem of **digital research infrastructures**, cultural heritage institutions, libraries, data centers, etc.
- Ask your library and research data manager!

# PARTHENOS

- PARTHENOS is a Horizon 2020 project (European Commission, 12 million EUR)
- Aim: strengthen the cohesion of Heritage related E-research
- Running time: 1 May 2015 - 30 April 2019 + 6 months extension
- PARTHENOS has 16 partners from 9 European countries, including the two humanistic research infrastructures CLARIN ERIC and DARIAH ERIC
- PARTHENOS Coordinator: PIN Scrl - Educational and Scientific Services for the University of Florence, Italy



# Research Data Management in the PARTHENOS Training Suite

## Module “Introduction to Research Infrastructures”

- Beginners’ level
- Main purpose: essential concepts underpinning a Research Infrastructure



### RDM topics:

**Interoperability:** e.g. What is Data?, What is Metadata?, What are Standards? What Are Knowledge Representation Systems and 'Ontologies'?

**Sustainability:** e.g. Data Lifecycle and Curation, Persistent Identifiers, Intellectual Property Rights and Licensing

### Methods and tools

- <http://training.parthenos-project.eu/sample-page/intro-to-ri/>

# 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/>

# 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/>

# European RIs for SSH & CH

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



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



# (Institutional) Infrastructure



- In your discipline?
- In your (local) institution/ university?
- On national/ European level?
- In your current project?
- In past projects?

This exercise is adapted from: Biernacka, K.; Dolzycka, D.; Helbig, K.; Buchholz, P. 2018. Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement. DOI: 10.5281/zenodo.1215377 (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>

Picture: Thinking statues taken by Rui Fernandes, CC-BY 2.0

(<https://creativecommons.org/licenses/by/2.0/>), <https://flic.kr/p/8WpM2U>

# What's in there for you?

## Benefits of Open Research Practices 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 ideals

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

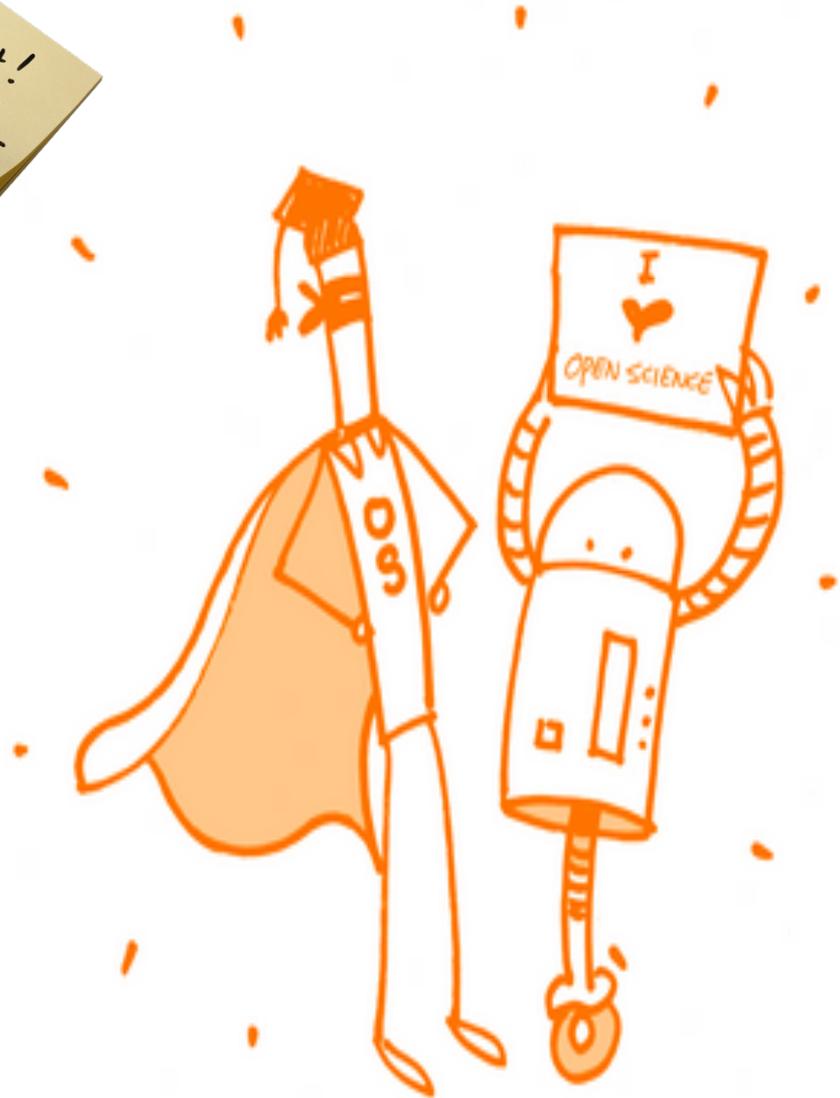


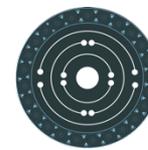
Picture: <https://zenodo.org/record/1285575#.W09yZH59jOR>  
(Melanie Imming, John Tennant, CC0)

# Next Session



- Your own Data Management Plan!
- Think of a “project” you will use!





CONTACT:

**Dr. Ulrike Wuttke**

University of Applied Sciences Potsdam (FHP)

[wuttke@fh-potsdam.de](mailto:wuttke@fh-potsdam.de)

@UWuttke

[www.parthenos-project.eu](http://www.parthenos-project.eu)

<http://training.parthenos-project.eu>

## Never miss an update! PARTHENOS Social Media Kanäle:

➤ **Twitter:** [https://twitter.com/Parthenos\\_EU](https://twitter.com/Parthenos_EU)

➤ **Facebook:** <https://www.facebook.com/PARTHENOSproject/>

➤ **YouTube Channel:**

[https://www.youtube.com/channel/UCnKJnFo\\_IFfoAI3VH51t1hw](https://www.youtube.com/channel/UCnKJnFo_IFfoAI3VH51t1hw)

08

**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 (starting 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

## 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/>
- **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/>
- **Research Data Management Promotional Material**
  - <https://rdmpromotion.rbind.io/>

# Further Learning

## FAIR Principles

- 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 of Swiss National Science Foundation SNF
  - [http://www.snf.ch/SiteCollectionDocuments/FAIR\\_principles\\_translation\\_SNSF\\_logo.pdf](http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf)
- Mons, Barend, *Data Stewardship for Open Science: Implementing FAIR Principles*, 2018

# 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](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**
  - <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/>

# 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](https://dmponline.dcc.ac.uk/public_plans)

# DMP Example

**Project "Oberdeutscher Spiegel Historiae Digital" (Fachhochschule Potsdam und Staatsbibliothek zu Berlin) U. Wuttke (FICTIONAL EXAMPLE!)**

General background

- Aim is to merge the handwritten tradition of the Oberdeutscher Spiegel Historiae as diplomatic transcriptions of the handwritten tradition with digital copies of the historical witnesses (2 manuscripts and one fragment) in a web-based presentation interface as a **digital edition**
- Funder DFG (German National Research Funder)
- **Research Data Policies:** "Richtlinie zur Sicherung guter wissenschaftlicher Praxis an der Fachhochschule Potsdam", no institutional research data guideline, DFG Guidelines are applicable
- DFG Richtlinie Gute Wissenschaftliche Praxis: store research data at least for 10 years

# DMP Example: Data Collection

Data is generated in the following formats and intended for archiving:

- 1) image data (master) of the handwritten transmission (**TIFF** uncompressed, specifications of the DFG Practice Guidelines on Digitization),
- 2) text data of the edition (**XML: TEI P5**) (possibly DTABf),
- 3) presentation environment: presentation of the full text (**XSLT, CSS**) and the image files (**JPEG**, not intended for archiving)
- 4) A naming convention will be developed.

# DMP Example: Data Documentation and Metadata

- Digital copies of the **image data** are provided with descriptive, administrative and technical metadata
- The structural metadata for the indexing of images (**TEI**) will be converted into **METS** according to the specifications for the DFG Viewer.
- The final text of the edition is saved in Unicode and coded in an **XML-based markup of the TEI (DTABf)**.
- In the markup descriptive, administrative and technical metadata, including coding of the essential structural elements and structure levels of the text, are assigned
- **Project-specific** XML elements and attribute-value pairs are documented in the **TEI header**
- The indexing is based on relevant disciplinary standards and reference models and, wherever possible, uses published **norm data** (e.g. the GND of the German National Library)

# DMP Example: Ethics and Legal Compliance

- No personal data
- Free from third party rights
- The generated data shall be made available for scientific research under the following conditions:
  - Image data: CCO,
  - text data: CCBY 4.0,
  - metadata: CCO.
- A data transfer agreement between the data provider and the infrastructure is planned.

# DMP Example: Storage and Backup

## Data management during the project:

- 1) Image data (TIFF, JPEG): institutional server (with institutional backup scheme, once a month),
  - 2) Text data: institutional server (with institutional backup scheme, once a month),
  - 3) Presentation data (XSLT): institutional server (with institutional backup scheme, once a month).
- The research data manager of the project (N.N.) ensures that only **authorized persons** have access to the data through the following measures: Development of a **project-related research data policy**, which is presented to all (relevant) employees of the project partners for information (protocol of signature).

# DMP Example: Selection and Preservation

- The data for subsequent use and archiving will be selected as follows:
  - 1) all master files of the digital manuscripts (incl. metadata),
  - 2) the final text data (incl. metadata),
  - 3) the editing guidelines,
  - 4) the code of the web presentation (incl. metadata),
  - 5) the documentation of the presentation (incl. description of the software, interfaces (APIs) and application profiles)
- The digital objects are subjected to a quality assurance process before being handed over. Data volume for archiving: not more than 1 TB.
- It is planned that a technically appropriate and certified infrastructure institution (e.g. the CLARIN-D Centre Berlin-Brandenburgische Akademie der Wissenschaften, **Data Seal of Approval**) will take over the storage and dissemination of the data and make them available to the scientific community, the reviewers and, if necessary, the public according to good scientific practice (i.e. at least 10 years).

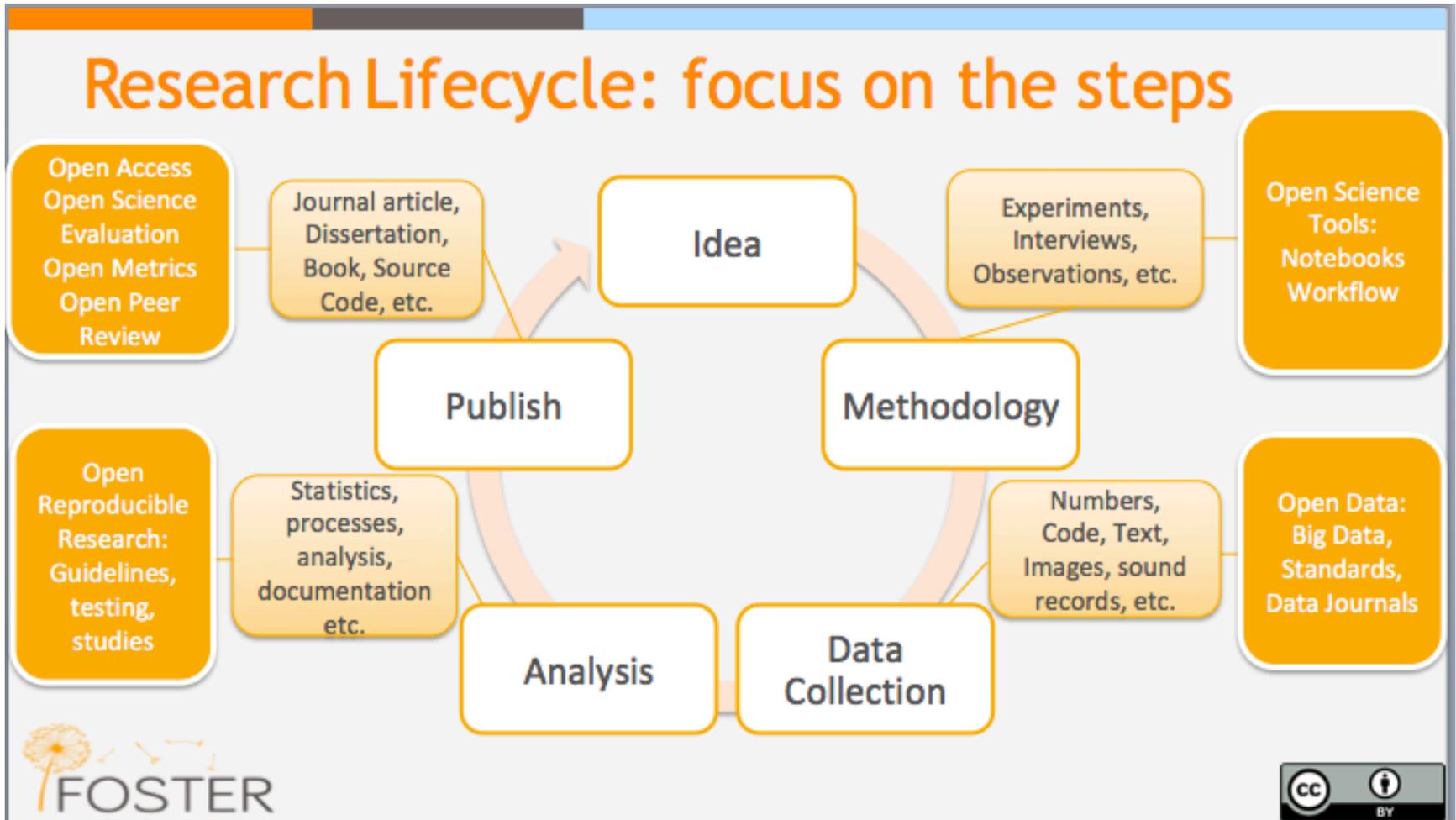
# DMP Example: Data Sharing

- The following groups of persons should have long-term access to the data:
  - 1) project members of the "Oberdeutscher Spiegel Historiae Digital" project,
  - 2) reviewers appointed by the funders and participating institutions,
  - 3) Data should be re-used by humanities scholars, especially literary scholars and historians.
- For the image data: CCO
- For the text data: CCBY 4.0
- For the metadata: CCO
- **Provision of the data:** OAI (standardized description in XML based Dublin Core Collections Application Profile and METS/TEI).
- The infrastructure partner includes the metadata descriptions of the research data in the search catalogue, so that the data can be easily found.

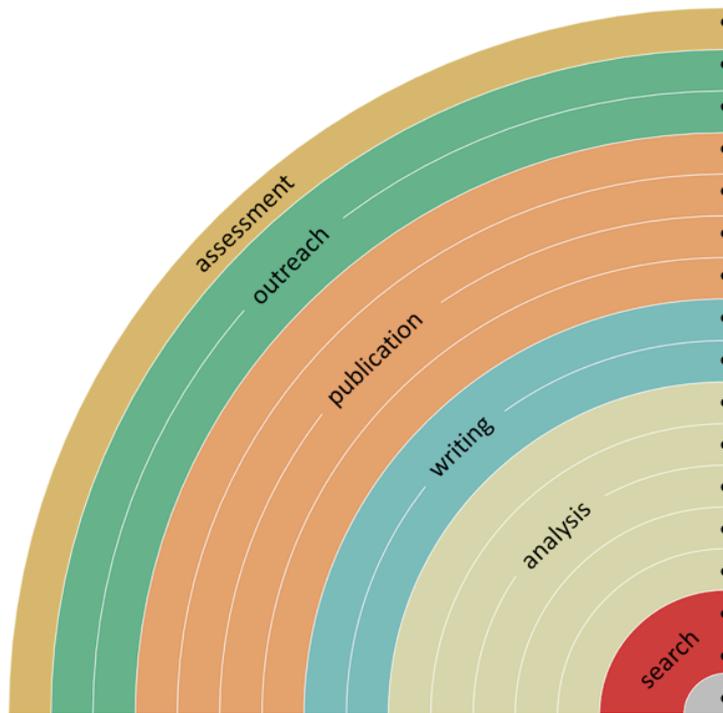
# DMP Example: Responsibilities and Resources

- Budget: The following expenses are expected for data management:
  - Creation of the metadata description: x FTE over x years
  - For the storage of the amount of data (budget data handling) and the associated data backup an effort of xx Euro is calculated.
- Time frame: The exact date of the data transfer has to be specified. The data should be made generally available to the scientific community and the public for re-use as soon as possible after transfer.

# Open scholarly practices that can make your research more visible



# 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



# Open Science Workflow (exemplary)



Source Picture: <https://101innovations.wordpress.com/workflows> (Bianca Kramer and Jeroen Bosman)

# Advantages of Open Access



# The FAIR Principles (1/2)

- **Findability** :
  - 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
- **Accessibility**
  - 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

# The FAIR Principles (2/2)

- **Interoperability**
  - 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
- **Reuse**
  - 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**