RDMO

Gefördert durch DFG Deutsche Forschungsgemeinschaft

FH;P Fachhochschule Potsdam University of Applied Sciences

Make your digital research more sustainable and visible:

Data Sharing and Data Management Techniques & Tools for Digital Medievalists

Dr. Ulrike Wuttke

Carmen Annual Meeting 2019,
University of Prague, 05.09.2019
DOI: 10.5281/zenodo.3405322

FH Potsdam, FB Informationswissenschaften, RDMO
Twitter: @uwuttke
ORCiD: https://orcid.org/0000-0002-8217-4025

TABLE OF CONTENTS

01

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

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

02

GETTING TO KNOW EACH OTHER

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

03

RATIONALES AND BENEFITS OF THE WORKSHOP

LEARNING OBJECTIVES

Participants…

- Can name opportunities and challenges of digital research practices
- Can describe basic concepts related to humanities research data
- Can identify good practices to make their research more sustainable and visible

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

04

HUMANITIES RESEARCH DATA (Key Concepts)

Data Management & Data Sharing: Why does it matter?

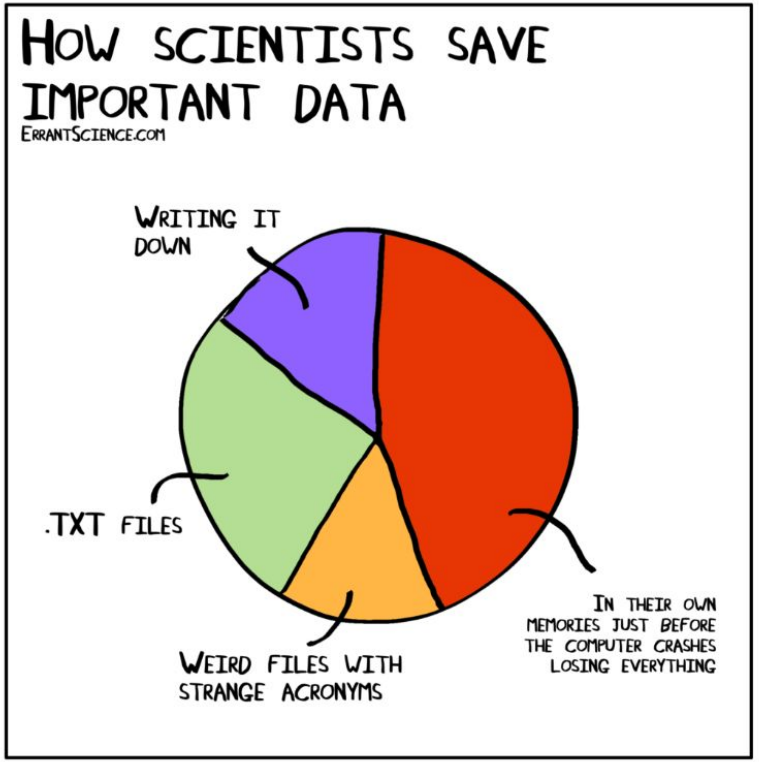
- Computers, the internet, and big data, have led to a rise of **digital quantitative** and **statistical methods** in Humanities and Cultural Heritage Studies
- Digital workflows & methods transform & broaden the humanities (esp. Digital Humanities), but …
  are based on **accessible, correct, authoritative, well structured data**
- **General challenges**, esp. barriers to **interoperability** (machine actionability) and **reuse:**
  - Data loss, findability and access (Closed Data), lack of documentation and standardization
- **Individual digital project (self) management** helps to avoid problems that can **compromise** your research process and outcomes

Want to know more?

☛ PARTHENOS Training Module **"Manage, Improve and Open Up your Research and Data"**

Looks familiar?

- 90% of all digital research data is lost (estimation 2011) [1]



Picture by Matthew Partridge, source: https://errantscience.com/blog/2019/06/24/cartoon-gallery-safety-unicorn-office-hours-and-gt/, CC-BY-NC

[1] Source: Stefan Winkler-Nees (from the Deutschen Forschungsgemeinschaft-DFG), Vorwort, In: Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (ed.): *Handbuch Forschungsdatenmanagement*, Bad Honnef 2011, p. 5.

What is Research Data?

- **Research data** are produced in and used in scientific processes
- **Sources are research data** and their management has always been part of the scientific process; digitization adds **complexity**
- **Humanities data** are primary **sources** (texts, pictures, etc.), secondary sources, theoretical texts, digital tools (software), annotations, bibliographies etc. (analogue and digital)
- **Digital data come in various formats and types** (pictures, texts, multimedia, measurements, etc.), **digitized** and **born digital**
- **Digital data** are **fragile**

PLAYFUL EXERCISE 1

Magic Data Glasses

What are YOUR Research Data?

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

➢ Form pairs of two
➢ Discuss (take notes of examples)
➢ Collection by trainer (acclamation)

05

MAKE YOUR RESEARCH MORE SUSTAINABLE AND VISIBLE (Good Practices)

SUSTAINABLE

Picture: Unknown, Braunau Stadtpfarrkirche Bäckeraltar Marientod 01, marked as public domain, more details on Wikimedia Commons

Sustainability of digital research projects

- To ensure **technical and intellectual sustainability** you need to understand that data (just as projects) have a life cycle (data life cycle) and adopt some (often simple) techniques throughout it in order to not face unpleasant surprises and ensure **data quality** and **research integrity** (Good Scientific Practice
- Often there are **institutional, national and international research infrastructures and experts** who work for them (e. g. data managers, DH specialists) and who will be able to support you in questions related to sustainability (e. g. information, counselling, tools, data repositories)

Key Skill: Research Data Management (RDM)

- Process to **curate** (or manage) research data along the **research data lifecycle** (e. g. planning, producing, selection, analysis, archiving, and preparation for reuse)
- RDM = Digital Project Management
- Detailed in **Data Management Plan (DMP)**
- Phase of **Planning your Research** is **crucial**:
  - Will you **share your data** and **how**?
  - Do **comparable projects** exist (reuse, good practices)?

Tools for Research Data Management

- **Why use a tool for a Data Management Plan (DMP)?**
  - Structured (discipline specific) questions and additional features
- DMPonline (DCC; UK)
- RDMO (AIP, FHP, KIT; Germany)
  - Organiser instead of a plan
  - Website: https://rdmorganiser.github.io
  - Demo Instance: https://rdmo.aip.de/

Good Practices to start with...

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Data Collection:

- Reuse data! Use tools to **find research data** (e. g. re3data) and clarify the rights to use the data (e. g. licenses)
- Create **machine readable** data! Use services and tools for optimal results (e. g. Transcribus for recognition of handwritten texts, Optical Character Recognition (OCR)). Ask your library.
- Get **organized**! Think of filename conventions and folder structures and how you will work together as a team efficiently.

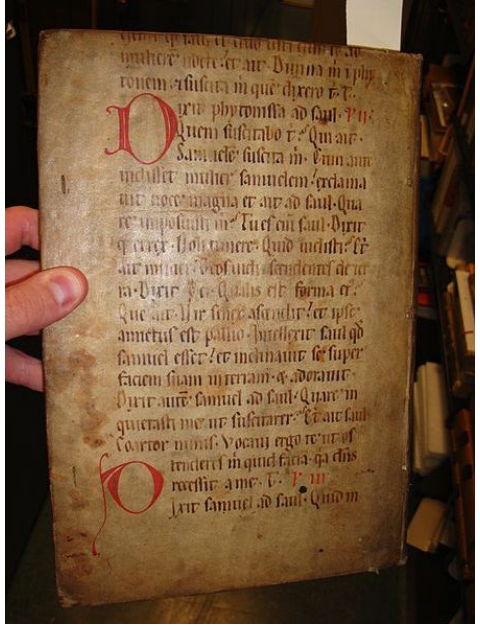
Good Practices to start with...

- Consider which **formats** are appropriate for your project (open formats, standard formats, proprietary formats).

Data Analysis

- Be wise in the **choice of your tools**! Consider pro's and con's of developing your own solutions (e. g. Open Source) against costs and risks of using commercial solutions (e. g. black boxing, workflow and vendor lock-in).

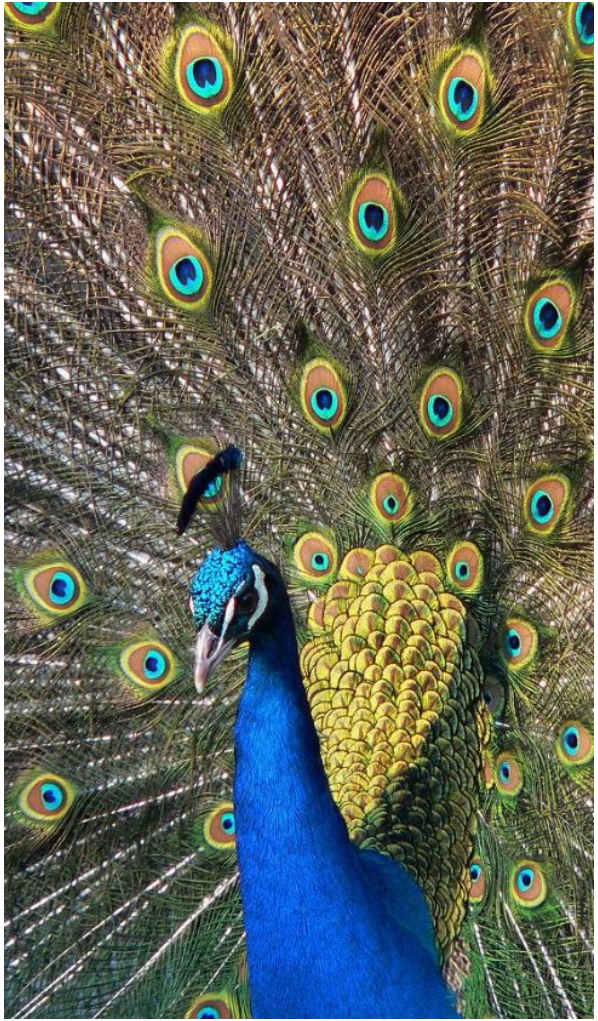
Good Practices to start with...



Data Storage and Backup

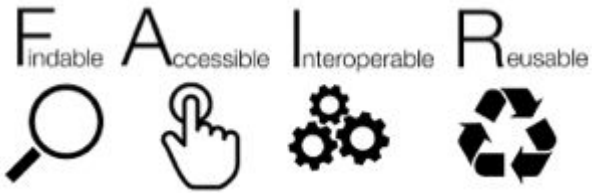
- Consider **timeliness of digital projects** (formats, programming languages, tools etc.)
- Your **archive format** can be different from working format (e. g. DOC vs. PDF/A)
- Chose **safe storage options** (institutional vs. Dropbox, USB-stick etc.)
- Carry out **Back-Ups** on a regular basis and test them!
- **Crucial question: Which data** to make future proof and **how**? Consider value of the data and resources needed to archive them. Develop **selection criteria** (e. g. political value, ethical value, intellectual value).
    - Rule of thumb: Always archive data on which publications are based!

VISIBLE (and FAIR)



Picture: BS Thurner Hof, Pfau imponierend, size, CC BY-SA 3.0

The FAIR Principles



Baseline understanding for the value sharing data can deliver and the baseline requirements for doing so

- **F**indable (e. g. DOI)
- **A**ccessible (e. g. open licenses) ☛ 
- **I**nteroperable (e. g. standards)
- **R**eusable (documentation, metadata)

FAIR for **humans** & **machines**! FAIR is not = Open!

Want to know more?

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

Standards, Metadata, Documentation

- Metadata are a love letter to the future
- Metadata are used to describe and organize data (formal description and contents description)
- Metadata standards enhance interoperability
- Transparent and sufficient **project documentation** and **abundant metadata** greatly enhances value of the data

Picture, Otto the Open Access Otter, by Katja Diederichs, CC-BY

Some relevant standards...

- **TEI** (Text Encoding Initiative)
- **CEI** (Charter Encoding Initiative)
- **MEI** (Music Encoding Initiative)
- **CMDI** (Language Resources, CLARIN)
- **IIIF** (International Image Interoperability Framework)
- **EAD** (Encoded Archival Description, for finding aids)
- **Dublin Core** (description of digital documents)

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

☛ **PARTHENOS Standardization Survival Kit (SSK):**

https://ssk-application.parthenos.d4science.org/ssk/#/

Make your research (data) visible!

- Publish for consulting, not only for reading, **link data with publications** and **online research profile** (website, blog) **& identity** (ORCiD)**!**
- **Data publication**:
  - **Data supplement** to article
  - **Data set publication** in a trusted (Open Access) repository (DOI!),
    - Generic: Zenodo, GitHub (Software), Wikidata (structured data)
    - **Humanities**: HAL**,** DARIAH-DE**,** CLARIN**,** GESIS
    - Institutional repositories
  - **Data Paper** in Data Journal, e. g. Research Data Journal for the Humanities and Social Sciences
- Tools to **find a repository:**
  - re3data & Directory of Open Access Repositories,
  - look out for CoreTrustSeal or Dini Certificate

Make data count!

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/

PLAYFUL EXERCISE 2

Challenges and Needs

Photo by Helloquence on Unsplash

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/

Your Data Is Powerful, If You Make It FAIR! Start Today!

RDMO

Gefördert durch DFG Deutsche Forschungsgemeinschaft

FH;P Fachhochschule Potsdam University of Applied Sciences

Trainer CONTACT:

Dr. Ulrike Wuttke

University of Applied Sciences Potsdam (FHP)

E-Mail: wuttke@fh-potsdam.de

Twitter: @UWuttke

Learning Resources

- Wuttke, Ulrike. Is Your Research Future Proof? Data Management Techniques & Tools for Digital Historians. 2019. DOI: http://doi.org/10.5281/zenodo.3247014
- Wuttke, Ulrike. Introduction to Humanities Research Data Management. 2018. DOI: http://doi.org/10.5281/zenodo.1491250
- PARTHENOS Module "Manage, Improve and Open Up Your Research Data" (eHeritage and eHumanities)
- PARTHENOS Guidelines to FAIRify data management, DOI: https://doi.org/10.5281/zenodo.2668479
- Edmond, Jennifer, & Tóth-Czifra, Erzsébet. Open Data for Humanists, A Pragmatic Guide. 2018. DOI: http://doi.org/10.5281/zenodo.2657248
- Website: Innovations in Scholarly Communication (B. Kramer & J. Bosman)
- Digital Preservation Coalition (esp. Knowledge Base)
- 5*Open Data (website with examples and benefits)

5-star Open Data Model (Tim Berners-Lee)

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