Universität Zürich[UZH]

swissuniversities

b UNIVERSITÄT BERN

Information Event: Open Research Data in Veterinary Medicine (ORDVET)

December 8, 2023

Vetsuisse Faculty

https://padlet.com/ORDVET/open_research_data_vetsuisse

ACCREDITED BY EAEVE/FVE

Universität Bern | Universität Zürich

vetsuisse-fakultät

Program

1st Part 10:15 – 12:30

Talk	Speaker	Time
Host	E. Dhein	
Introduction	Prof. A. Hehl	10:15-10:20
Survey results	Dr. E. Dhein	10:20-10:30
Basics ORD	Dr. M. Röthlisberger (OSS ZH)	10:30-10:50
Research data management	Dr. O. Churakova (OSS BE)	10:50-11:10
Reproducibility	Dr. E. Furrer (CRS)	11:10-11:30
Data protection	S. Marazza (CCdigitallaw, USI)	11:30-12:10
General Q&A	all	12:10-12:30

2nd Part 13:30 – 15:00

Talk	Speaker	Time
Host	T. Leeb	
Data sharing in genetics	Prof. T. Leeb	13:30-13:45
Swiss digital pathology initiative	Prof. I. Zlobec	13:45-14:00
Survey results	Dr. E. Dhein	14:00-14:05
Introduction BIMS	Prof. F. Guscetti	14:05-14:20
Use case BIMS	Dr. E. Dhein	14:20-14:40
General Q&A	all	14:40-15:00

ORDVET survey - results (part 1)

Elena Dhein

Institute of Veterinary Pathology, Vetsuisse Faculty, University of Zurich

Postdoctoral researcher

Universität Zürich[UZH]

swissuniversities

b UNIVERSITÄT BERN

Open Research Data in Veterinary Medicine (ORDVET) project

Elena Dhein	Institute of Veterinary Pathology, Vetsuisse Faculty Zurich
Franco Guscetti	Institute of Veterinary Pathology, Vetsuisse Faculty Zurich
Tosso Leeb	Institute of Genetics, Vetsuisse Faculty Bern
Sven Rottenberg	Institute of Animal Pathology, Vetsuisse Faculty Bern

Project aims:

1. **Education**
   - Raise awareness for ORD practices and responsible data management at the Vetsuisse Faculty
   - Building networks within and outside the Vetsuisse Faculty

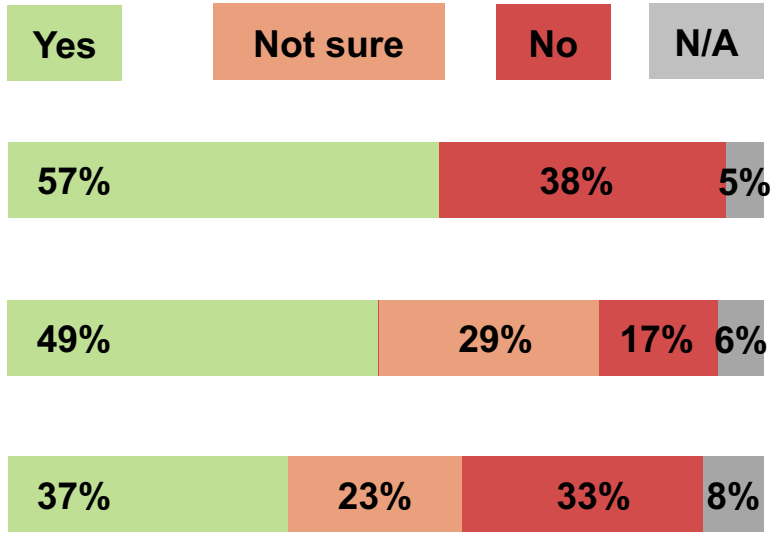
2. **Standardization of (meta)data**
   - Evaluation and concept development of the integration of terminology/coding system(s) in the process of (routine) diagnostics

Survey: Status quo at the Vetsuisse Faculty

- Online survey (1st July 2023 – 15th September 2023)
- Distributed among the employees at the Vetsuisse Faculty
- Up to 19 questions
- 212 participants (167 total answers, 45 partial answers)
- 2 parts:
  1. Open Research Data (general)
  2. Data standardization/vocabularies

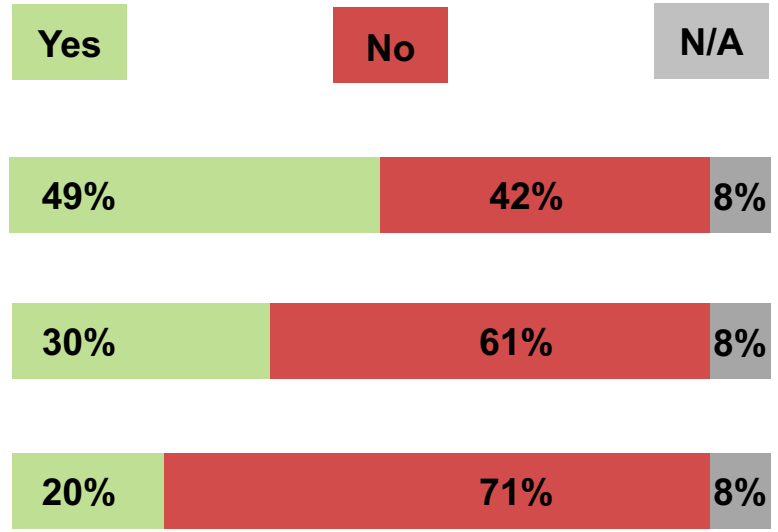
Status quo at the Vetsuisse Faculty

General



Status quo at the Vetsuisse Faculty

Have you ever published a dataset of research data or do you plan to do so in the future?



Status quo at the Vetsuisse Faculty

What obstacles do you see in publishing datasets containing research data? (%)

Options	Yes	No	N/A
Data protection regulations	**50**	35	14
Time required	**47**	39	14
Copyrights	**44**	42	14
Lack of standards	39	47	14
Lack of appropriate infrastructure	26	59	14
Lack of incentives	24	62	14
Expenses	19	67	14
I don't see any obstacles.	8	78	14

Other: Lack of data protection, ”data stealing”, lack of experience (metadata and curation), ”paper mills that produce fantasy papers”, lack of infrastructure, conflict with pressure to publish, not clear in which repository to find what

Status quo at the Vetsuisse Faculty

What form of support would you need to publish datasets containing research data? (%)

Options	Yes	No	N/A
Clarification about data protection regulations	**52**	33	15
Provision of uniform standards	**51**	34	15
Clarification about copyrights	50	35	15
Provision of suitable **infrastructure**	48	37	15

Other: Financial/technical support, information on repositories, more time for research, faculty competence center

Status quo at the Vetsuisse Faculty

Standards:

- University standards
- Specifications in data and file formats, data structures (minimal datasets, metadata, sensitive data)
- Repositories for specific data types
- Guidelines on nomenclature, controlled vocabulary
- Data security
- Application examples

Status quo at the Vetsuisse Faculty

Infrastructures:

– Suitable platform for data upload/deposition:
  - Institutional vs. discipline-specific
  - Generation of DOIs
  - Easy to use/user-friendly
  - Linked to existing systems, in conformity with established journals
  - No costs
  - Long-term storage possible
  - Large data volumes
  - Copyright and legal framework compliant
  - Flexible embargo solution

– Contact for help/support:
  - Preparation, uploading, formatting, curation, anonymizing/pseudonymizing data → Padlet: Link collection

Status quo at the Vetsuisse Faculty

Thank you! ☺

Part two will follow at 2 pm…

Basics in Open Research Data

Melanie Röthlisberger

Open Science Services, University Library, University of Zurich

Data Expert

Library
University of Zurich

University of Zurich[UZH]

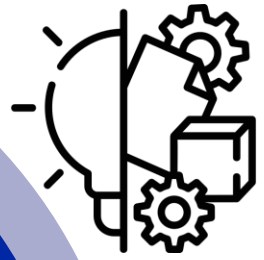
Basics of Open Research Data

Dr. Melanie Röthlisberger
Open Science Services UZH

data@ub.uzh.ch

Open educational resources

Share posters & presentations



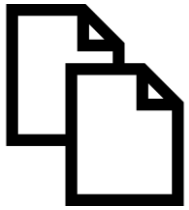
Open methodologies

Shared reference libraries (e.g. zotero)

Preregistration

Open/FAIR data

Open code



Open Access

Open peer review

Open licenses (CC0/CC-BY)

Open/FAIR data

Open Source

Open code

Open protocols

Use actionable formats

What is Open Data?

Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.

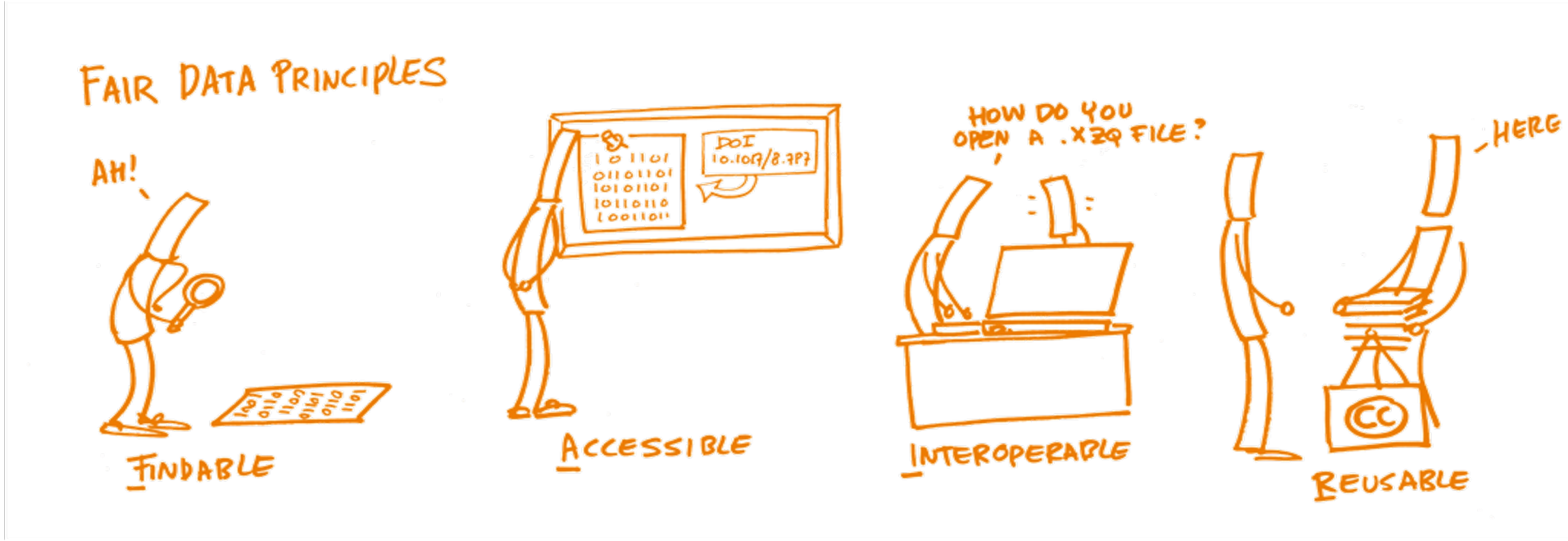
The Open Data Handbook
http://opendatahandbook.org/guide/en/what-is-open-data/

What is Open Data?

*Open data is data that can be freely used, **re-used** and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.*

The Open Data Handbook
http://opendatahandbook.org/guide/en/what-is-open-data/

Reusable data



- ✓ Data documentation / metadata
- ✓ Persistent identifiers

- ✓ Open metadata
- ✓ Access regulations (FAIR ≠ Open)

- ✓ Open formats
- ✓ Controlled vocabularies

- ✓ Data documentation / metadata
- ✓ licenses

Source: H. Brinken and T. Ross-Hellauer, Open Science Training Handbook (FOSTER Open Science, 2018)

Data documentation

Bad

Better...

Best
(Rich, Structured)

Nutrition Facts*

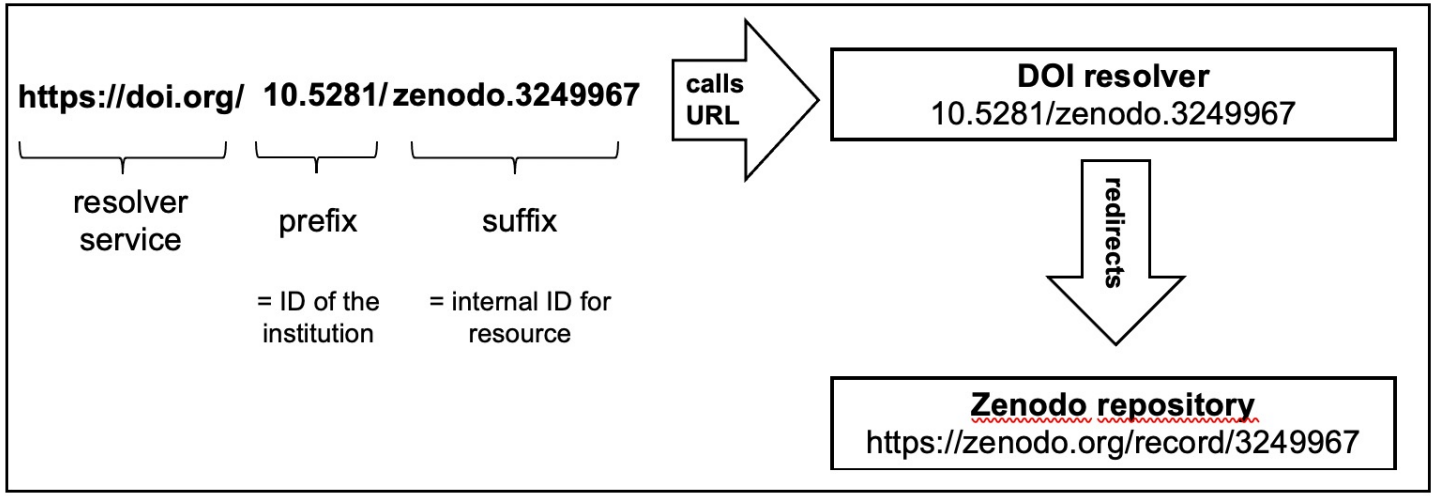
Amount Per Serving (serving size) = 1/2 cup (120mL) condensed soup

Calories 100	Dietary Fiber 2g
Fat Calories 50	Sugars 1g
Total Fat 6g	Protein 1g
Sat. Fat 1g	Potassium 60mg
Trans Fat 0g	
Polyunsat. Fat 4g	**% Daily Values****
Monounsat. Fat 1g	Vitamin A 0%
Cholesterol 5mg	Vitamin C 0%
Sodium 870mg	Calcium 0%
Total Carb. 9g	Iron 0%

* The nutrition information contained in this list of Nutrition Facts is based on our current data. However, because the data may change from time to time, this information may not always be identical to the nutritional label information of products on shelf.

** % Daily Values (DV) are based on a 2,000 calorie diet.

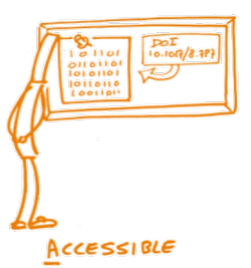
Persistent identifiers



PIDs in a nutshell:

PIDs today are often expressed as **URLs**, and the registry indicates where that URL should ultimately resolve. That PID will always point to the correct item even if the item's location changes.

Access control



Open Access

All users can access the data without ide ation.

Access fo rs

The data can only be cred users. They contain no direct ide tain indirect identifiers.

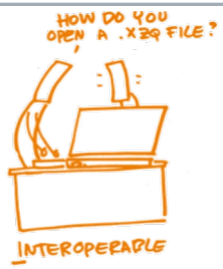
tricted Access

Access is restricted a can only be granted upon request. This access level is useful for the most sensitive data.

Sensitive Data:
"As open as possible, as "restricted" as necessary"

optional embargo

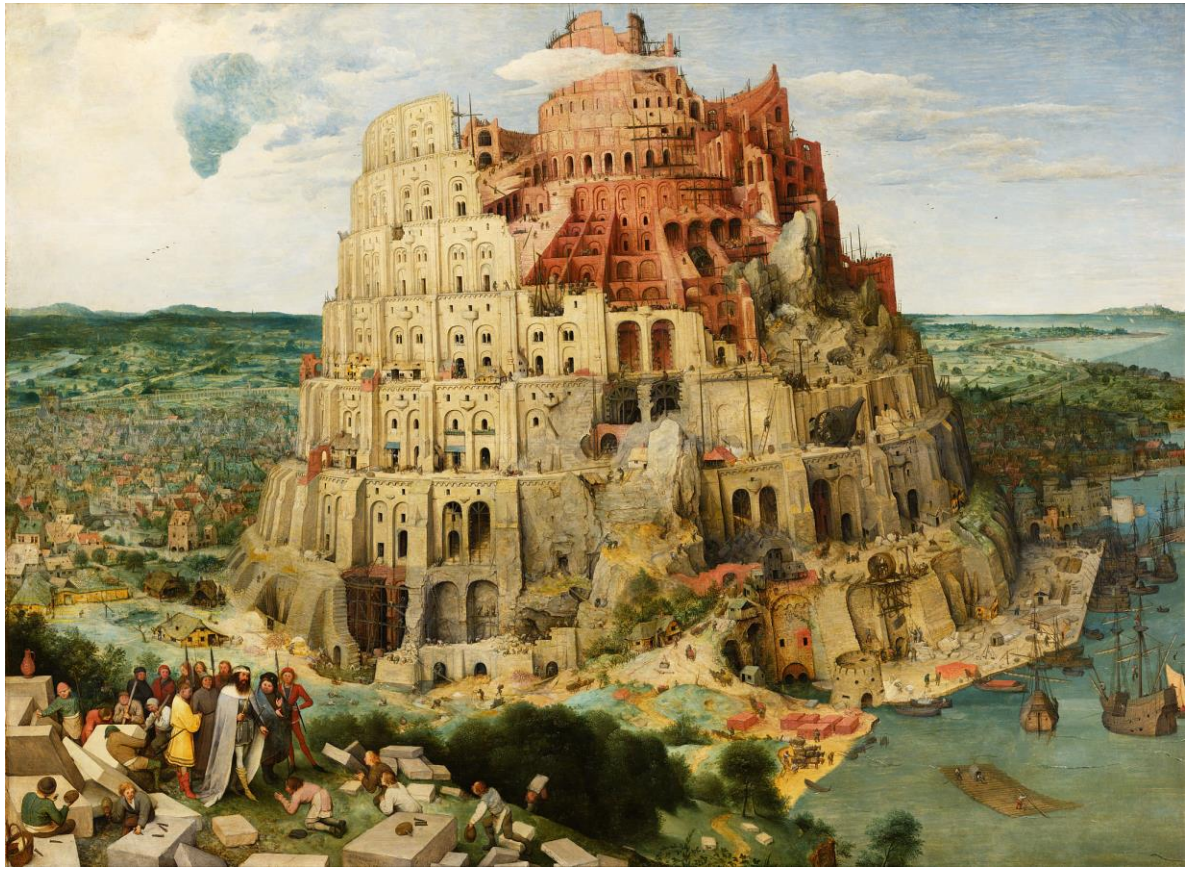
Open file formats



Data type	Good	Acceptable	Unsuitable
Tabular data	.csv / .hdf5	.txt / .html / .tex / .por	
Tabular data with few metadata	.csv / .tab / .ods / SQL	.xml if appropriate DTD / .xlsx	.xls / .xlsb
Text	.pdf / .txt / .odt / .odm / .tex / .md / .htm / .xml	.pptx / .pdf with embedded forms / .rtf	.doc / .ppt
Code	.m / .R / .py / .iypnb / .rstudio / .rmd / NetCDF	.sdd	.mat / .rdata
Images	.tif / .tiff (uncompressed) / .png / .svg / .jpeg	.jpg / .jp2 / .tif / .tiff (compressed) / .pdf / .gif / .bmp	.indd / .ait / .psd
Audio data	.flac / .wav / .ogg	.mp3 / .mp4 / .aif	
Video data	.mp4 / .mj2 / .avi / .mkv	.ogm / .webm	.wmv / .mov
Geographical data	NetCDF, tabular GIS attribute data, .shp / .shx / .dbf / .prj / .sbx / .sbn / PostGIS / .tif / .tfw / GeoJSON	.mdb / .mif	
General data	.xml / .json / .rdf		

Source: EPFL Fastguide on open file formats: https://infoscience.epfl.ch/record/265349/files/04_File_Formats_EPFL_Library_RDM_FastGuide.pdf?ln=en

Controlled vocabularies



FAIRsharing.org
standards, databases, policies

This image was first published on Flickr. Original image by Classic Art Wallpapers. Uploaded by Ibolya Horvath, published on 12 December 2016, published with Creative Commons Attribution-ShareAlike. Available at: https://www.worldhistory.org/image/6123/the-tower-of-babel/

Copyright

Copyright protects works of intellectual (human) **creation** that are original and have **individual character**.

- texts
- movies
- presentations
- databases
- computer programs (software)
- photographs
- technical drawings
- interviews

By **licensing** your data, you can clearly communicate permissions to reusers of your data. A license is a **legal instrument** for a rights holder to **permit** a second party to do things that would otherwise infringe on the rights held.

Where to share data: FAIR-compliant repositories

institutional repositories:

E.g. SWISSUbase, ETH Research Collection

general purpose repositories:

E.g. Zenodo, EUDAT, Dryad, Harvard Dataverse

discipline-specific repositories:

Ask colleagues, or search re3data.org. Recommended wherever possible.

commercial repositories:

E.g. Figshare, Mendeley Data

The ORD landscape in Switzerland

swissuniversities

National ORD Strategy:

- Support researchers and research communities
- Develop, promote and maintain infrastructures
- Develop skills and best practices
- Build systemic and supportive conditions

Schweizerischer Nationalfonds

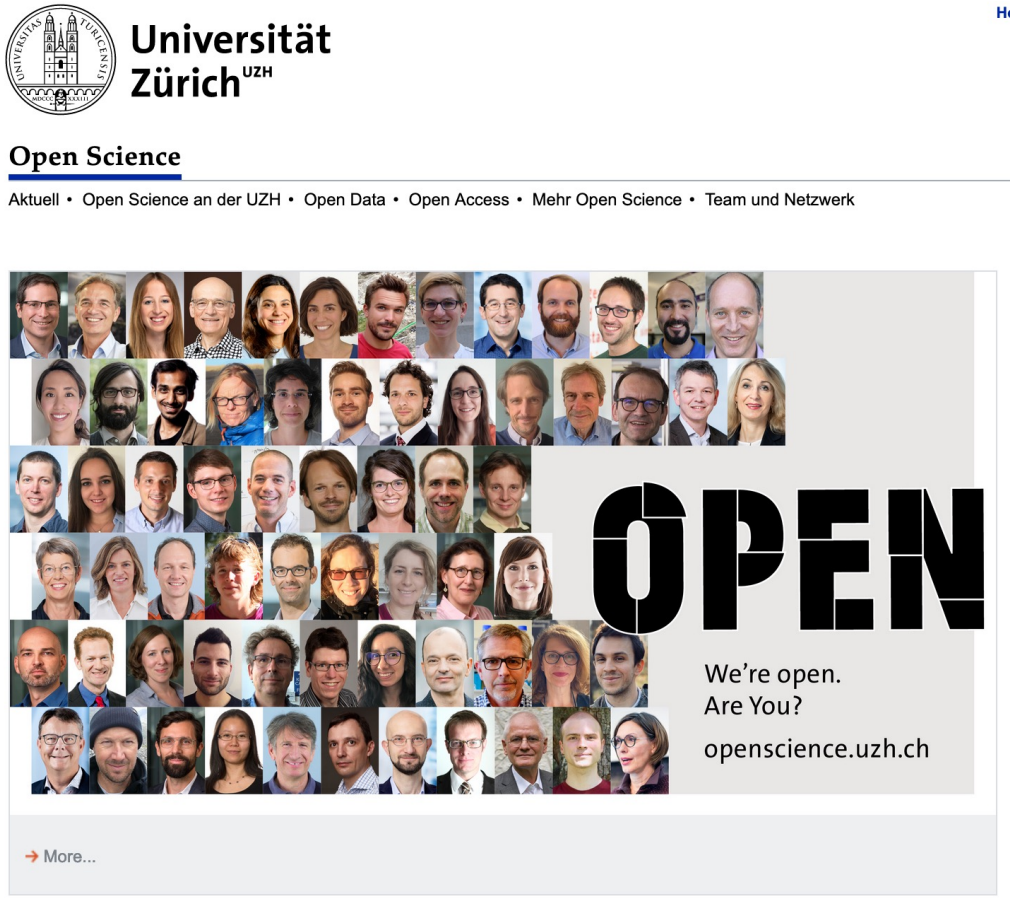
ORD requirements:

- store the research data they have produced
- share these data with other researchers, unless bound by legal, ethical, etc. clauses,
- deposit their data and metadata onto existing public repositories

Universität Zürich[UZH]

UNIVERSITÄT BERN

UZH recommendations



- Data (at least the metadata) that form the basis for a publication should be openly accessible.
- Data should be made available as early as possible, but no later than at the time of first publication.
- Data should comply with the FAIR principles and the rules of good scientific and legal practice.
- The UZH expects its researchers to use existing open data if they are available in good quality.

Training & skills at UZH

UB-courses for BA, MA and PhDs

- Open Access Basics
- Making your data FAIR
- Open Data Basics
- Data management planning

Courses by the CRS

- Good Research Practice (GRC)
- 5 Steps to Good Data Science Practice in R (STS)

Courses for Postdocs and WiMi

- Open Up and Share
- Writing your DMP for the SNSF
- Publishing sensitive data

More support at UZH

- Open Research Data (UZH library): https://www.ub.uzh.ch/de/wissenschaftlich-arbeiten/mit-daten-arbeiten.html
- Data protection: https://www.rud.uzh.ch/en/angebot/datenschutzrecht.html
- Legal aspects: https://www.rud.uzh.ch/en.html
- Trainings by 3R Tierwohl: https://www.tierschutz.uzh.ch/en/Training-and-education.html

Thank you

data@ub.uzh.ch

Please cite as: Melanie Röthlisberger. 2023. Basics of Open Research Data. ORDVET Info Event December 2023. Universität Zürich.

Research Data Management

Olga Churakova

Open Science RDM, University Library, University of Bern

Data Steward (Medicine, Vetsuisse, Insel Hospital)

Open Research Data
Research Data Management Life Cycle

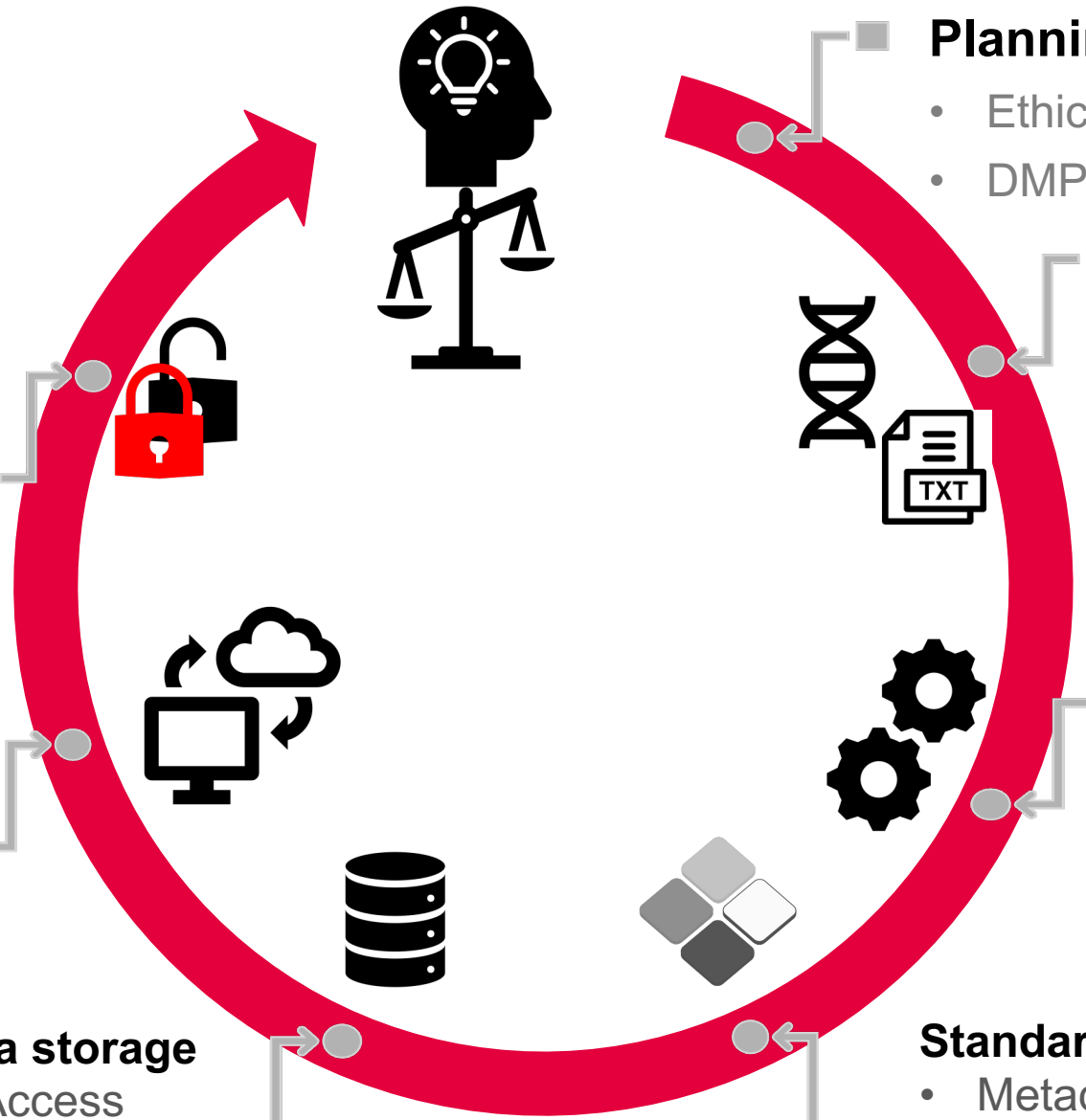
Dr. habil. Olga Churakova

Data Steward: Medicine, Vetsuisse, Insel Hospital

Open Science RDM, University Library of Bern

Image: © UB UniBE

Research Data Management Life Cycle



Planning
- Ethics
- DMP

Data collection
- Type
- Format
- Size
- File convention
- Description
- Documentation

Data processing
- Tools
- Software
- Codes

Standards
- Metadata
- Controlled Vocabularies

Data storage
- Access
- Timeline

Data Sharing
- Distribution
- Reuse

Data Publishing
- Access
- License
- Monitor
- DMP final updates

Planning: Before Data Collection

- Code of conduct for scientific integrity

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

- New Federal Act on Data Protection nFADP
- Animal Welfare Act

Kanton Bern
Canton de Berne

- Cantonal Data Protection Act

Planning: Before Data Collection

- University of Bern Act

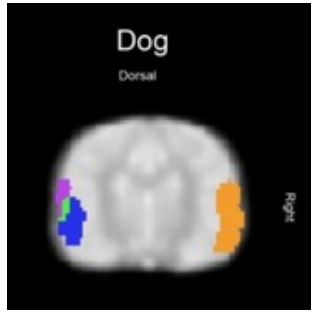
- Animal Welfare Office

- Data Protection Officer (DPO)

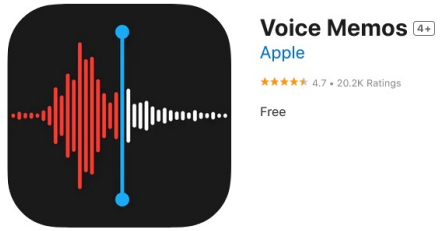
- Information security and data protection (ISDP)
- IT-Department UniBE

Data Collection

Data Type and File Format



https://doi.org/10.3390/ani12010108



Define data type

- Survey
- fMRI image
- Genomic data
- Video
- Voice recording

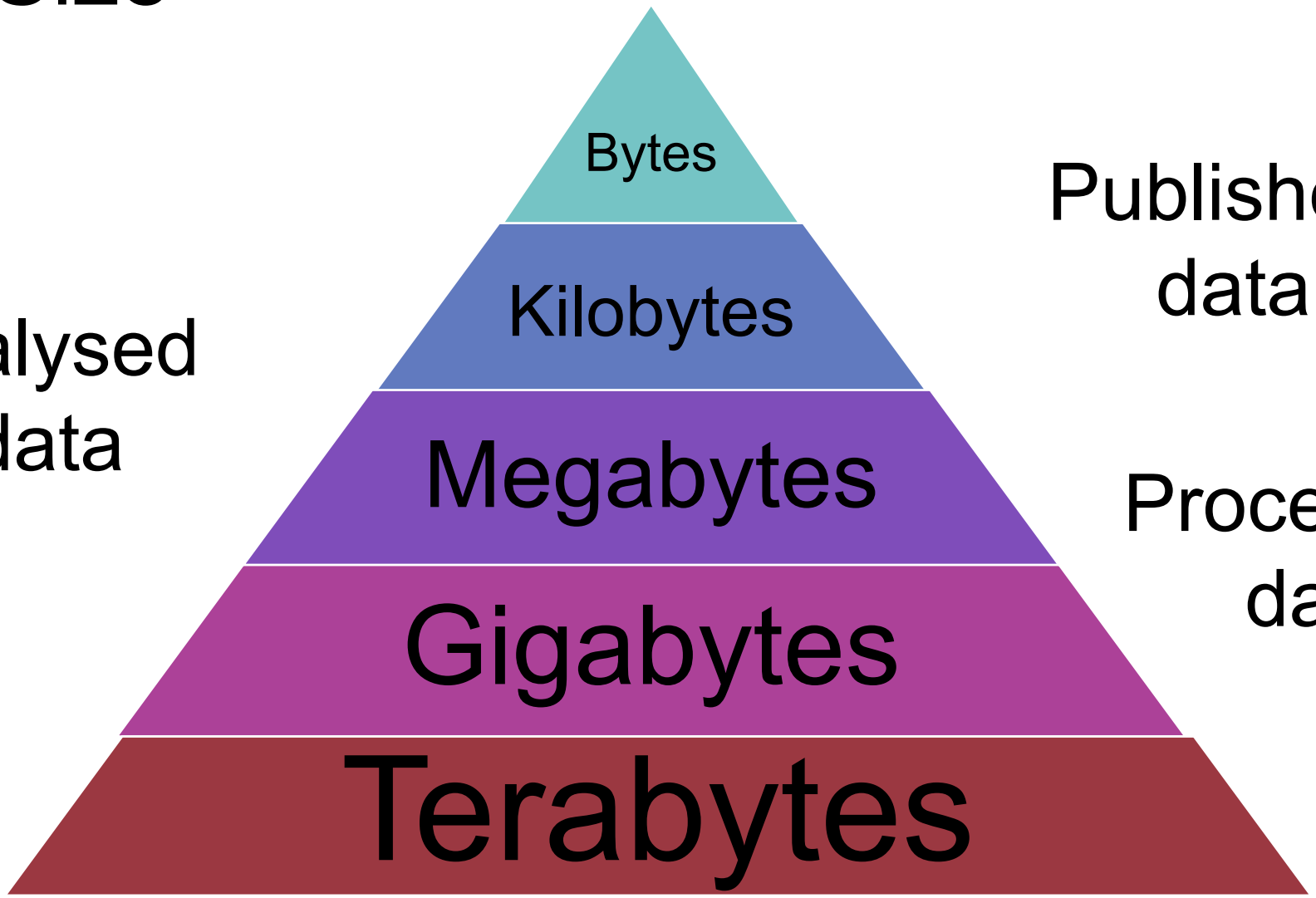
File format

- .csv
- .nifti
- .PDF/A
- .txt
- .wav

BORIS file formats UnIBE (EN)
ETH Zürich (EN)

Data Collection

Data Size



Documentation
File-level: ReadMe and Codebook

ReadMe File:

- **Who** created the data?
- **What** is the content of the data?
- **Why** were the data developed?
- **Where** is it geographically located (coordinates)?
- **When** were the data created?
- **How** were the data developed?

Readme_Template_EN.txt (3KB)

Codebook:

- Variables name
- Labels
- Codes
- Units
- Missing values

Use:

- Tabular (raws, columns)
- Statistical data

Documentation Tools

Searchable
Protocol, sample, data set

Automated

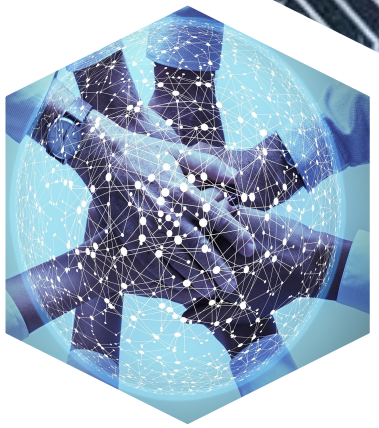
Report generation, manuscript writing, or inventory tracking

Consistent

Results are accurate and reproduceable

Collaborative

Knowledge exchange among researchers and stakeholders, regulatory bodies, and funders



Traceable

Track experiments, samples, protocols, and results

https://www.dlcm.ch/resources/dlcm-eln

Documentation Tools

Documentation tools

- Study protocols, documents exchange
- Laboratory Information Management System (OpenBIS)
- https://github.com/openbis
- https://github.com/

Recommendation on research data documentation from the Open Science Team (PDF, 141KB)

Documentation Tools

Computational Notebooks

Jupyter Notebook

https://github.com/

Open-source web application

- Create and share documents containing live code, equations, visualizations and narrative text
- Uses include:
  - Comments in scripts
  - Data cleaning and transformation
  - Numerical simulation and statistical modeling
  - Data visualization
  - Create tutorials and interactive manuals

Description from: Jupyter procect, https://jupyter.org/documentation;
J.M.Perkel: Why Jupyter is data scientists' computational notebook of choice DOI: 10.1038/d41586-018-07196-1

Metadata

Dublin Core: Key Metadata Elements

Contributor	Title		Creator	Date
Description	Type		Rights	Relation
Publisher	Format		Language	Subject
Identifier	Coverage		Source	

Agnew, 2005

Metadata
BORIS Portal Example

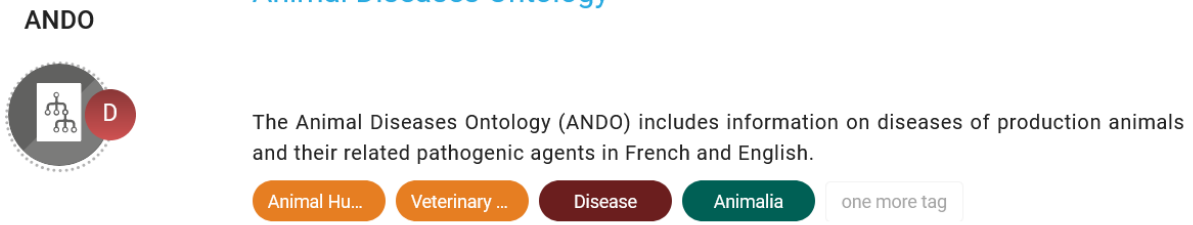
Based on Dublin Core elements

Title:	Daily suicides and ambient temperature data in Switzerland
Contributor(s):	Bär, Séverine Vicedo Cabrera, Ana Maria Bundo, Marvin Müller, Thomas de Schrijver, Evan
Affiliations:	Institute of Social and Preventive Medicine Institute of Social and Preventive Medicine Institute of Social and Preventive Medicine Institute of Social and Preventive Medicine
Contact:	Vicedo Cabrera, Ana Maria
Data Availability:	Open
Keyword(s):	epidemiology;public health;climate
Subject(s):	600 - Technology > 610 - Medicine & health
Description:	The dataset consists of daily number of suicides per Canton in Switzerland between 1995 and 2016 by age, sex and method of suicide. It also includes daily mean temperature per Canton.
URI:	http://hdl.handle.net/20.500.12422/71
DOI:	https://doi.org/10.48620/38
Rights:	cc-nc
Type:	Dataset
Appears in Collections:	Research Data

https://doi.org/10.48620/38

Metadata
Controlled Vocabularies

"Controlled vocabularies provide a clearly defined terminology to catalog and retrieve information with control of synonyms and variant terms."
(RDM services, KU Leuven)



- Data Curation Center (DCC) metadata standards
- FAIR standards Animal Diseases Ontology
- Veterinary Extension SNOMED-CT
- ICD-11
- Vet-ICD-O-canine-1

Standards

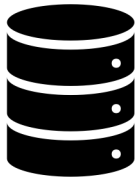
Harmonised Practices

EU Data Standardisation Strategy
[European Veterinary Big Data strategy 2022-2027]

ISO Standards

- 11.220 Veterinary medicine including equipment specific to veterinary medicine

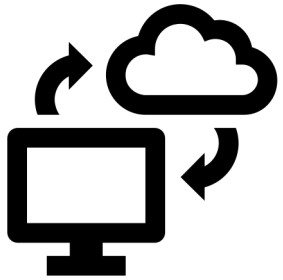
Data storage & backup



- Back-up Campus: daily automated
- Back-up Hard drive (manual, no Internet access)
- IT-Department UniBE

Data sharing

?



SWITCHdrive

VeraCrypt 1.26.7

Idrix

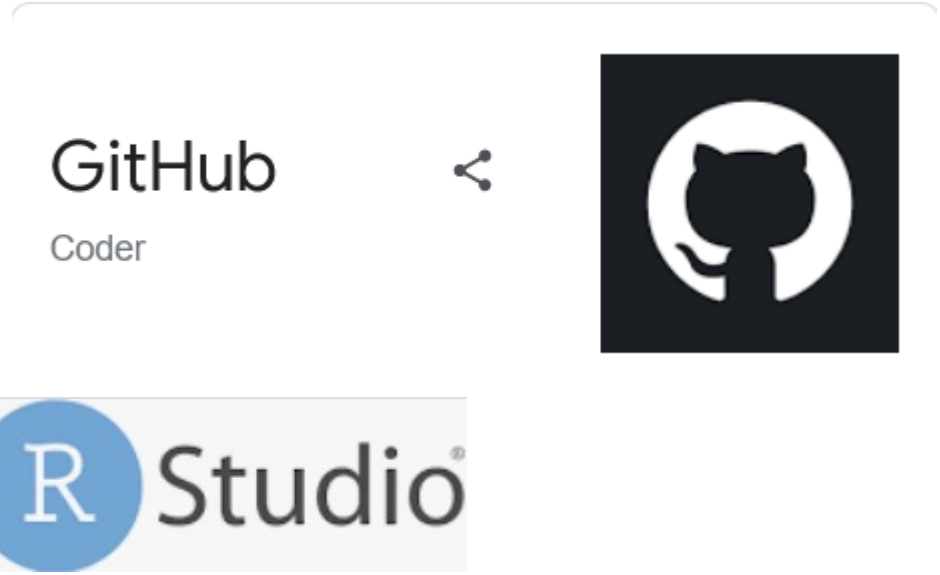
[Cryptomator]

[VeraCrypt]

Data Preparation for Publishing

Computer Codes and Software

- Github Link and Versions Control
- Software version used
- Details of where the software can be accessed
- Computer Scripts, Codes

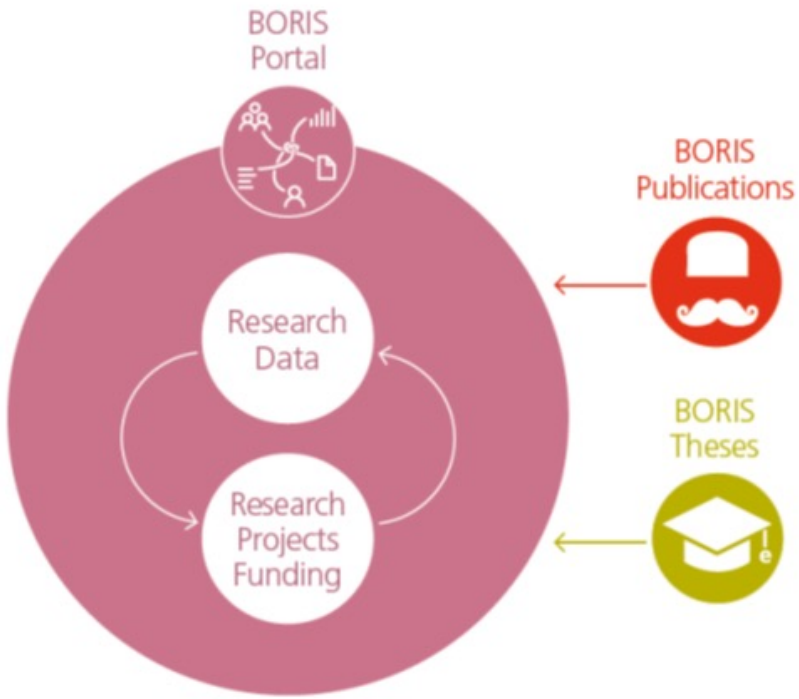


UniBE Recommendation: Open-Source-Software (PDF, 68KB)

Open Research Data

Institutional Repository – Overview

BORIS Portal Research Data Repository



https://boris-portal.unibe

- Institutional research data repository @ UniBE
- Digital Object Identifier (DOI)
- Metadata description (Dublin Core)
- Metadata stored under CC0 and permanently
- Data documentation upload
- Managing data access
- Licenses
- Policy for long-term preservation (**10 years**)
- For clinical studies permitted (anonymized data only)

Access, Distribution, or Reuse Considerations

Open

Open access to dataset

Embargoed

Enter date after which dataset will be released.

Restricted

Upload dataset and grant access on request.

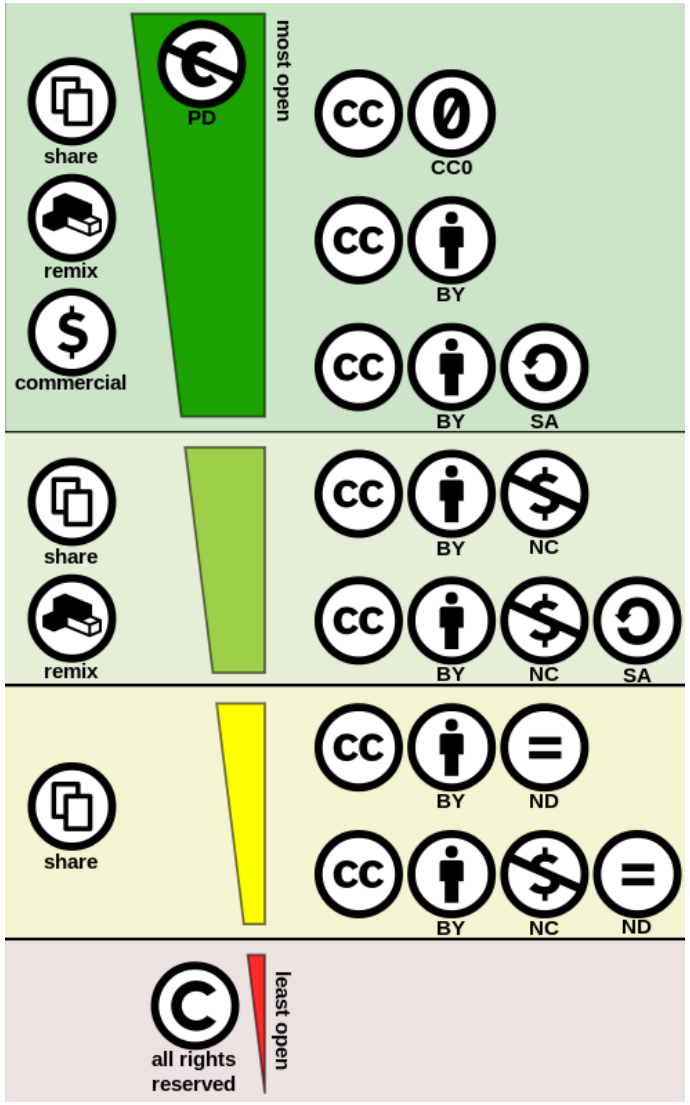
Closed

No data upload, but meta-data should be entered to verify existence of dataset.

Data Transfer and Use Agreement

- recommended for data that cannot be shared openly
- individually define re-use conditions for dataset
- share DTA along with data for others to download and sign

Data, Computer Code, Software Publication



To reuse data, take into account:

- Licenses
- Ownership (including joint ownership)
- Data transfer and use agreement

- If there are **NO** ethical, legal or contractual issues
- Metadata
- Supplementary material

- Some forms of processed data (e.g., elaborate visualisation)
- Data will be commercially exploited (patents)

CC Licenses; Software licenses: https://choosealicense.com/

Data Publishing
Summary

- Check policies
- Select data
- Sensitive data→Metadata

- www.re3data.org
  SNSF repositories list

- File naming & structuring
- Data de-identification
- Documentation
- Data format
- Software source code

- Choose access
- Choose a license
- Check list

- Data availability statement
- Digital Object Identifier
- Link your articles, datasets, code

Open Research Europe

Support
Training

How to publish & license research data

How to plan and update Data Management Plan

How to store research data

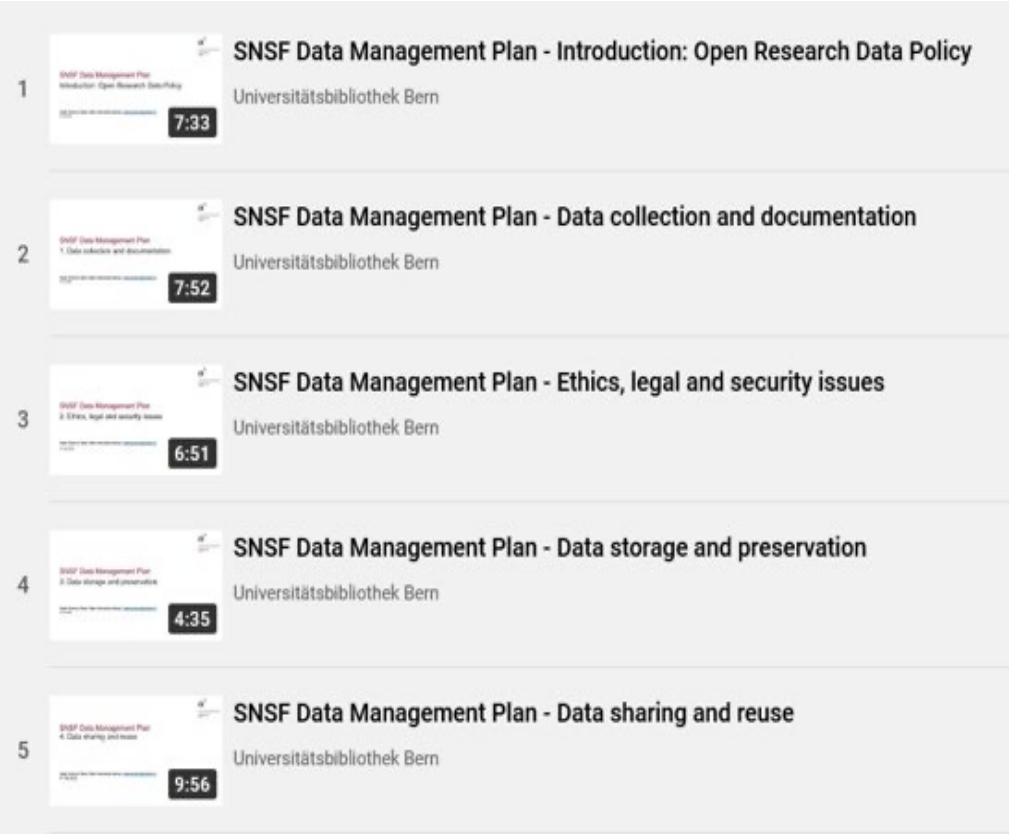
How to ethically manage research data

How to efficiently document research data?

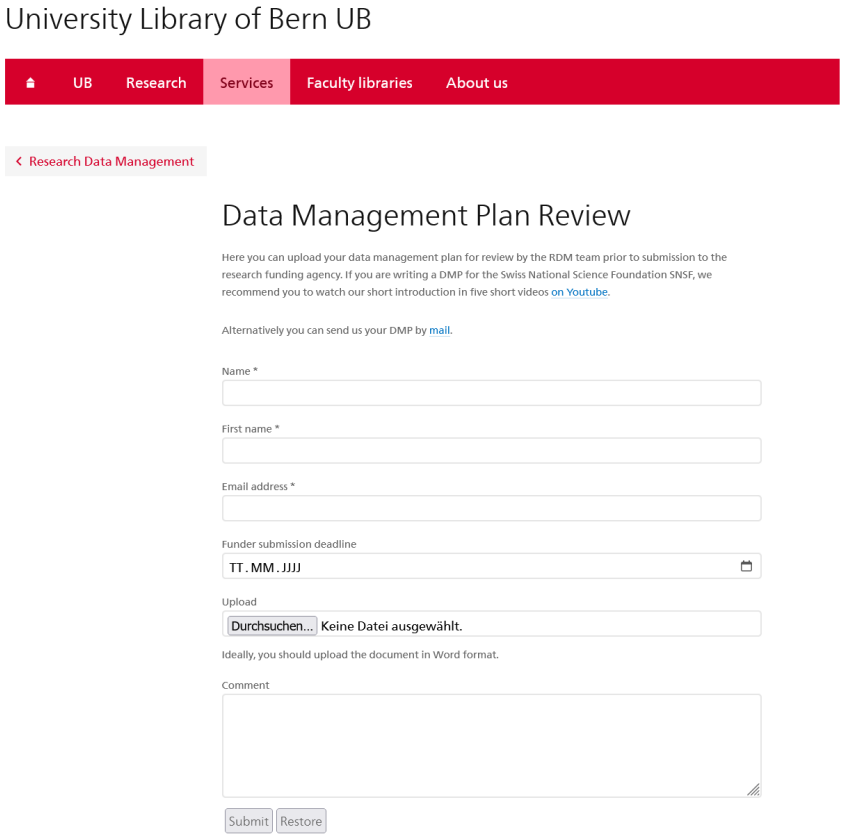
How to collect and organize data and files, formats, names, version control

@ openscience@unibe.ch

Support
Data Management Plan Review



Video modules YouTube



Data Management Plan review online
Feedback within 1-3 working days

Open Science
Newsletter

- Funder's news
- Training sessions and courses
- Developments in Open Access and Research Data

To subscribe:

https://www.unibe.ch/ub/osnews

Contact

Dr. habil. Olga Churakova

E-Mail: olga.churakova@unibe.ch; openscience@unibe.ch

Data Steward: Medicine, Veterinary Medicine, Insel

Research Data Management Support

Thank you

for your attention

Open Science Team
openscience@unibe.ch

Reproducibility projects (CRS)

Eva Furrer

Epidemiology, Biostatistics and Prevention Institute, University of Zurich

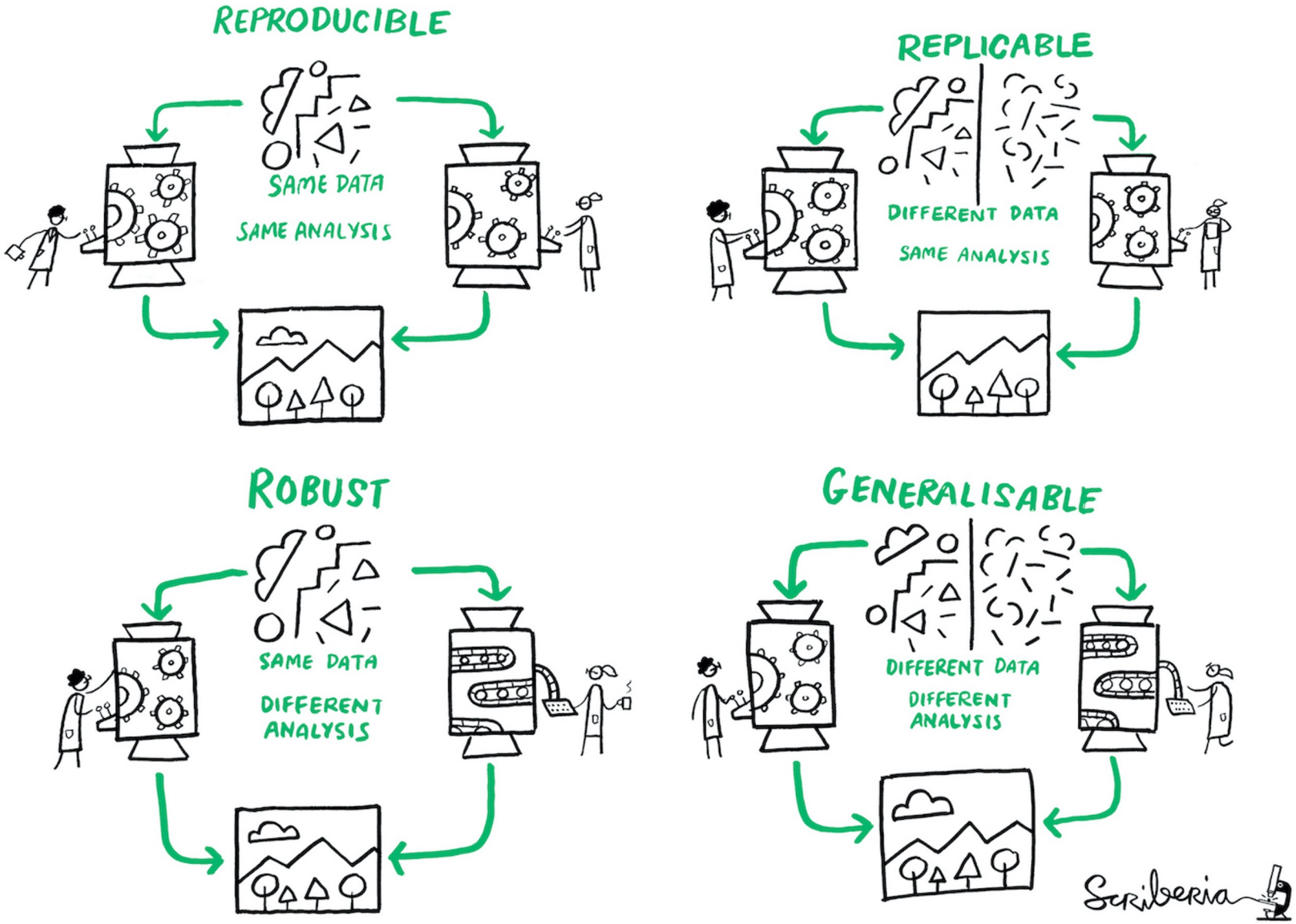
Managing director at the Center for Reproducible Science

Reproducibility - where to start?

ORDVET Information event, December 8, 2023

Eva Furrer, Center for Reproducible Science, University of Zurich

Let’s start with a definition



Why and how?

Markowetz *Genome Biology* (2015) 16:274
DOI 10.1186/s13059-015-0850-7

Genome Biology

COMMENT Open Access

Five selfish reasons to work reproducibly

Florian Markowetz

Abstract

And so, my fellow scientists: ask not what you can do for reproducibility; ask what reproducibility can do for you! Here, I present five reasons why working reproducibly pays off in the long run and is in the self-interest of every ambitious, career-oriented scientist.

Keywords: Reproducibility, Scientific career

Patterns

CellPress
OPEN ACCESS

Perspective

Reproducibility Starts from You Today

Yasemin Turkyilmaz-van der Velden,[1,*] Nicolas Dintzner,[1] and Marta Teperek[1]
[1]Delft University of Technology, Mekelweg 5, 2628 CD Delft, the Netherlands
*Correspondence: y.turkyilmaz-vandervelden@tudelft.nl
https://doi.org/10.1016/j.patter.2020.100099

THE BIGGER PICTURE In recent years, discussions about the reproducibility of scientific experiments have bloomed everywhere but have left researchers with either very high-level and unachievable goals or lost in a sea of recommendations. However, there are many small, simple steps that any researcher can take to improve the reproducibility of their results. Crucially, improving the reproducibility of one's own research workflows offers numerous selfish benefits: not only making research more efficient (less time wasted!) but also increasing research impact and reach.

This article is aimed at researchers and offers several simple recommendations that can result in incremental improvements to the reproducibility of research results.

1 2 3 4 5 **Concept:** Basic principles of a new data science output observed and reported

https://doi.org/10.1186/s13059-015-0850-7 https://doi.org/10.1016/j.patter.2020.100099

Five selfish reasons to work reproducibly

Reason 1: reproducibility helps to avoid disaster

⇒ detailed record saves time later

Reason 2: reproducibility makes it easier to write papers

⇒ update results automatically when data change

Reason 3: reproducibility helps reviewers see it your way

⇒ reviewers have access to complete analysis

Reason 4: reproducibility enables continuity of your work

⇒ continuity for collaborators and future self

Reason 5: reproducibility helps to build your reputation

⇒ honest and careful researcher

Planning For Reproducibility Should Start At The Very Beginning Of The Project

- Carefully Design Your Research Project
- Write Up and Publish the Design of Your Study as a ‘‘Registered Report’’
- Start a Data Management Plan for Your Project
- Plan for Computational Reproducibility
- Ask for Help

Planning For Reproducibility Should Start At The Very Beginning Of The Project

- Carefully Design Your Research Project
- Write Up and Publish the Design of Your Study as a ‘‘Registered Report’’
- Start a Data Management Plan for Your Project
- Plan for Computational Reproducibility
- Ask for Help

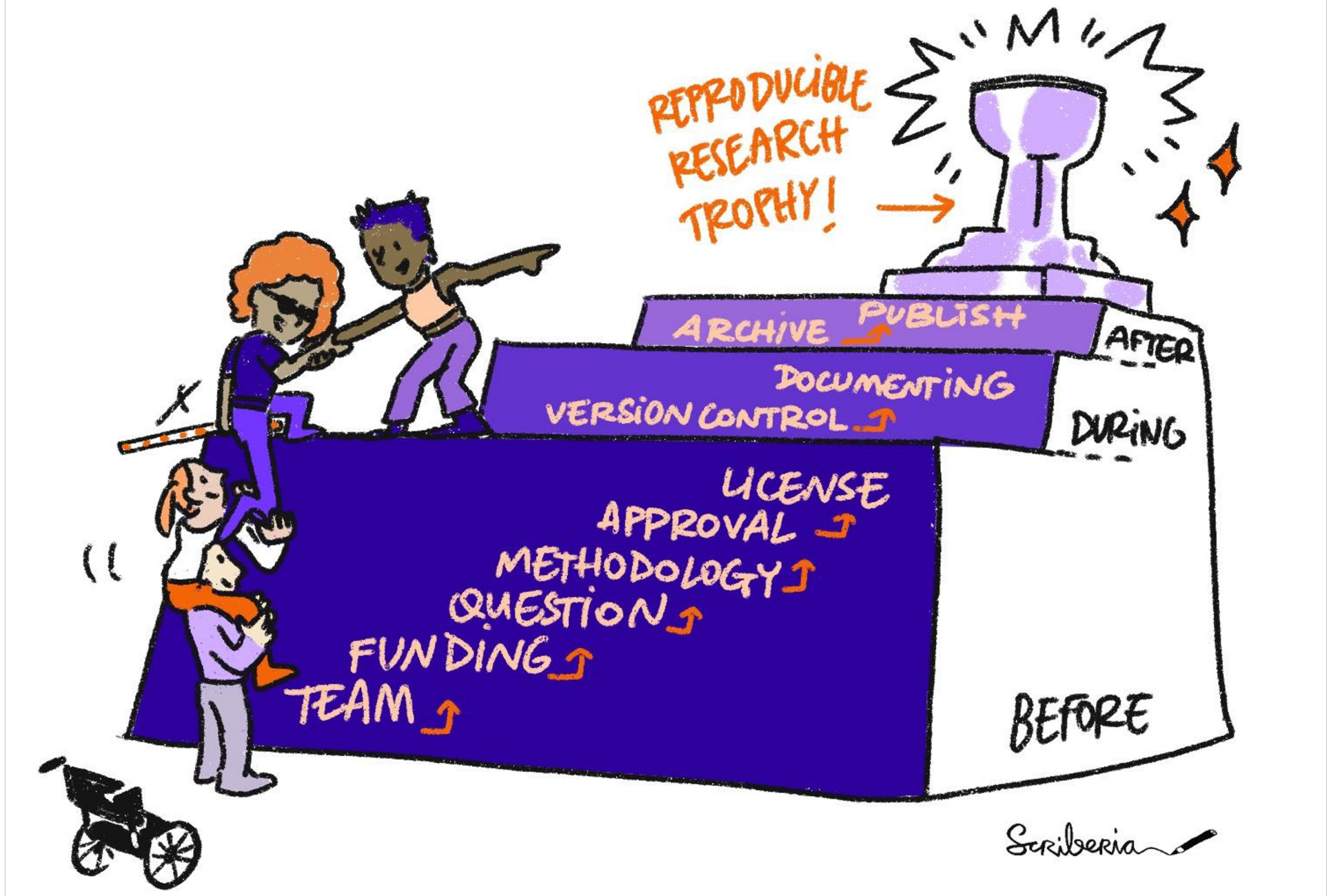


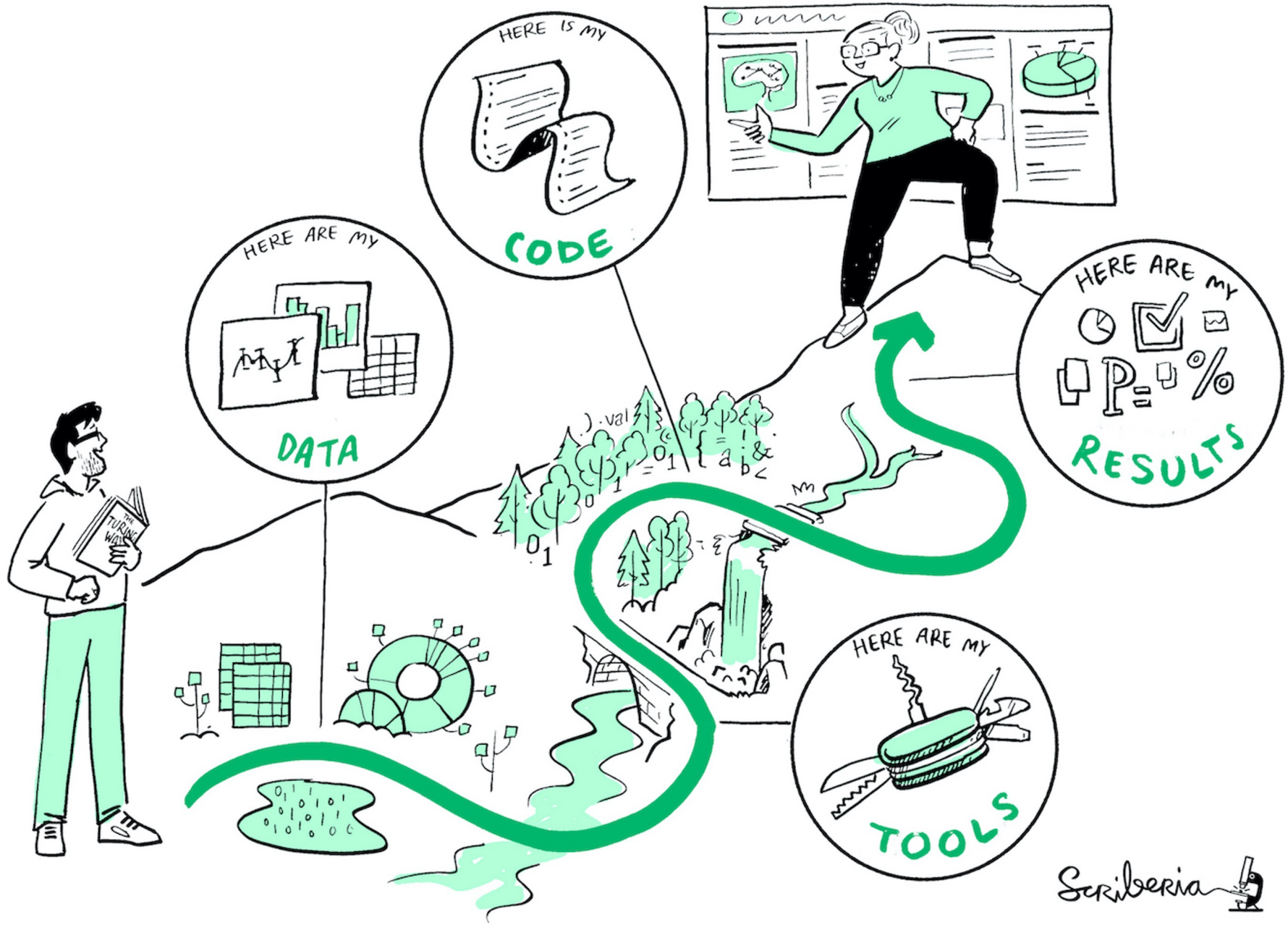
Image credit: Illustration by David Parkins for Nature https://www.nature.com/articles/d41586-019-02674-6

During Your Project: Documentation And Version Control Is Essential For Reproducibility

- Document Your Experimental Work
- Document Your Computational Work
- Use Version Control

During Your Project: Documentation And Version Control Is Essential For Reproducibility

- Document Your Experimental Work
- Document Your Computational Work
- Use Version Control



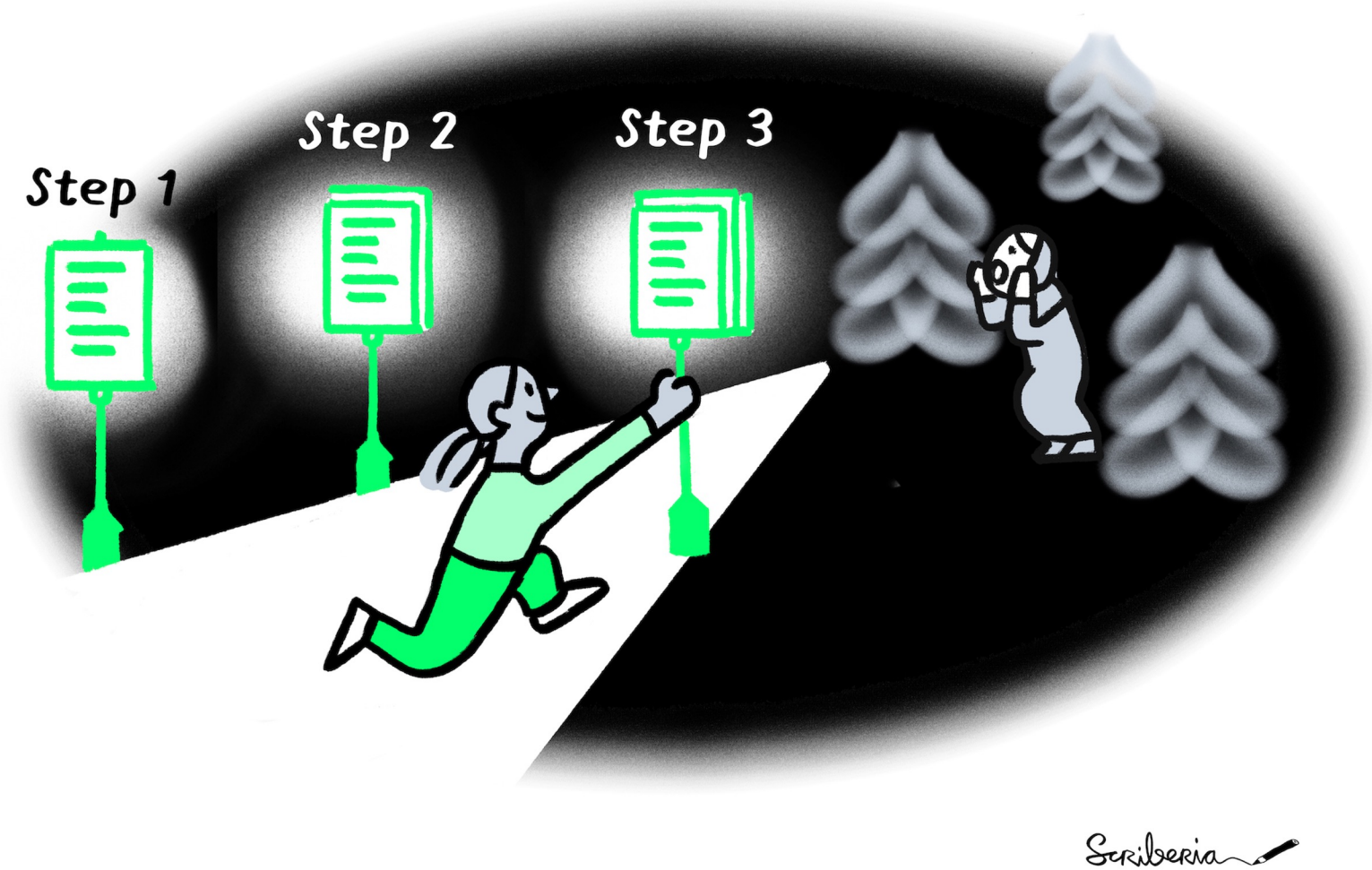
After Your Project: How To Share Your Work So That Others Can Reproduce It

- Archive Your Output
- Increase Your Citations
- Get a License
- Share Your Protocols
- Share Your Analysis Scripts and Research Software

After Your Project: How To Share Your Work So That Others Can Reproduce It

- Archive Your Output
- Increase Your Citations
- Get a License
- Share Your Protocols
- Share Your Analysis Scripts and Research Software

DOCUMENTATION



Reproducibility Project: Cancer Biology

https://www.cos.io/rpcb?hsLang=en
https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology

Overview RP:CB

- Acquire evidence about the replication of preclinical research in cancer biology by repeating selected experiments from 53 high-impact papers published 2010-2012*

- Selection: 400 most cited papers from both Scopus and Web of Science using a specific search string for 2010, 2011, and 2012 (Errington, 2014)

- Due to challenges: project only completed 50/193 planned experiments to repeat (26%)

Overview RP:CB

Three papers in the e-life collection

➔ Investigating the replicability of preclinical cancer biology

➔ Reproducibility in Cancer Biology: Challenges for assessing replicability in preclinical cancer biology

➔ Experiments from unfinished Registered Reports in the Reproducibility Project: Cancer Biology

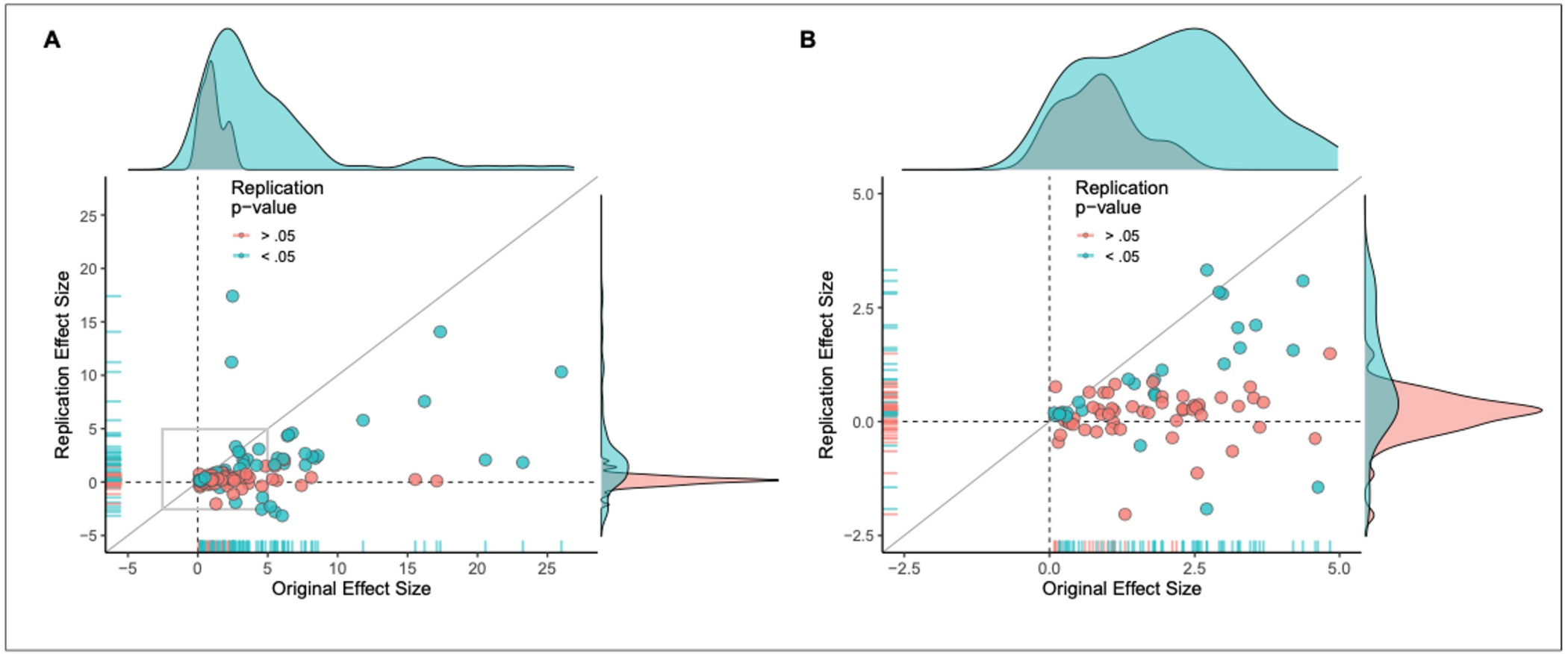


Figure 2. Replication effect sizes compared with original effect sizes. (**A**) Graph in which each circle represents an effect for which an SMD effect size could be computed for both the original effect and the replication (n = 110). Blue circles indicate effects for which $p < 0.05$ in the replication, and red circles indicate $p > 0.05$. Two effects for which the original effects size was >80 are not shown. The median effect size in the replications was 85% smaller than the median effect size in the original experiments, and 97% of replication effect sizes were smaller than original effect sizes (below the gray diagonal line). (**B**) An expanded view of panel A for effect sizes < 5 (gray outline in panel A). SMD: standardized mean difference.

Figure 1. Barriers to conducting replications – by experiment. During the design phase of the project the 193 experiments selected for replication were coded according to six criteria: availability and sharing of data; reporting of statistical analysis (i.e., did the paper describe the tests used in statistical analysis?; if such tests were not used, did the paper report on biological variation (e.g., graph reporting error bars) or representative images?); availability and sharing of analytic code; did the original authors offer to share key reagents?; what level of protocol clarifications were needed from the original authors?; how helpful were the responses to those requests? The 29 Registered Reports published by the project included protocols for 87 experiments, and these experiments were coded according to three criteria: were reagents shared by the original authors?; did the replication authors have to make modifications to the protocol?; were these modifications implemented? A total of 50 experiments were completed.

Center for Reproducible Science CRS@UZH

Improve overall reproducibility and quality of empirical research

- Good research practice courses
- Workshops
- Lectures

Promote original research in reproducibility and methodology

- Methodology related to reproducibility
- Replication studies
- Meta-research

Image credits: Eva Furrer, Davide Bonazzi/Salzmanart for Science https://www.science.org/doi/10.1126/science.361.6408.1178

ReproducibiliTea

Journal club dedicated to topics related to reproducibility, statistics in data analysis, open science, research quality and good research practices across fields, biomedicine, social sciences, computer science. etc.

Happens around the world: https://reproducibilitea.org/

Diverse program every semester:

University of Zurich

University of Basel

University of Geneva

Swiss Reproducibility Network

Peer-led consortium that aims to promote and ensure **rigorous research practices** in Switzerland by

- establishing appropriate training activities
- designing and evaluating research improvement efforts
- disseminating best practice
- working with stakeholders to coordinate efforts
- aligning with international networks

SwissRN aims for broad disciplinary representation and an intensive interdisciplinary dialogue.

https://www.swissrn.org

License

Please cite as: Eva Furrer, Reproducibility - where to start? ORDVET Information event, December 8, 2023, University of Zurich, December 8, 2023

Data protection in veterinary medicine

Suzanna Marazza

CCdigitallaw, Università della Svizzera italiana

Legal consultant

Data protection ORDVET - UZH

8 December 2023

Suzanna Marazza
CCdigitalLaw - Università della Svizzera italiana

Program

- ❑ Who **owns** data?
- ❑ **Privacy** and **Data protection**
- ❑ When do I need to consider **Data protection laws**?
- ❑ What are **personal**, **sensitive** and **anonymised data**?
- ❑ How do I **lawfully process** personal data?
- ❑ What is an **informed consent** and when is it needed?

Data **ownership**: who owns data?

What categories of data do we know?

- **Non-personal / commercial data**

  → Not protected as such by law

- **Personal data / personal sensitive data**

  → Protection of privacy in the digital environment

What categories of data do we know?

- **Human data** (data resulting from emails, spreadsheets, presentations, images, ect.)

  → Intellectual property righs (copyright) over original expression

- **Machine-generated data** (data automatically generated by a computer process, application or other mechanism)

  → Not protected

What categories of data do we know?

- **Compilation or aggregation of data**
  - → Might be protected by copyright
    - → If originality in **selection** or **arrangement** of the data or materials
    - → But copyright protection is limited to the particular selection or arrangement.
  - → EU: *sui generis* protection aimed at rewarding investment (Directive 96/9/EC)

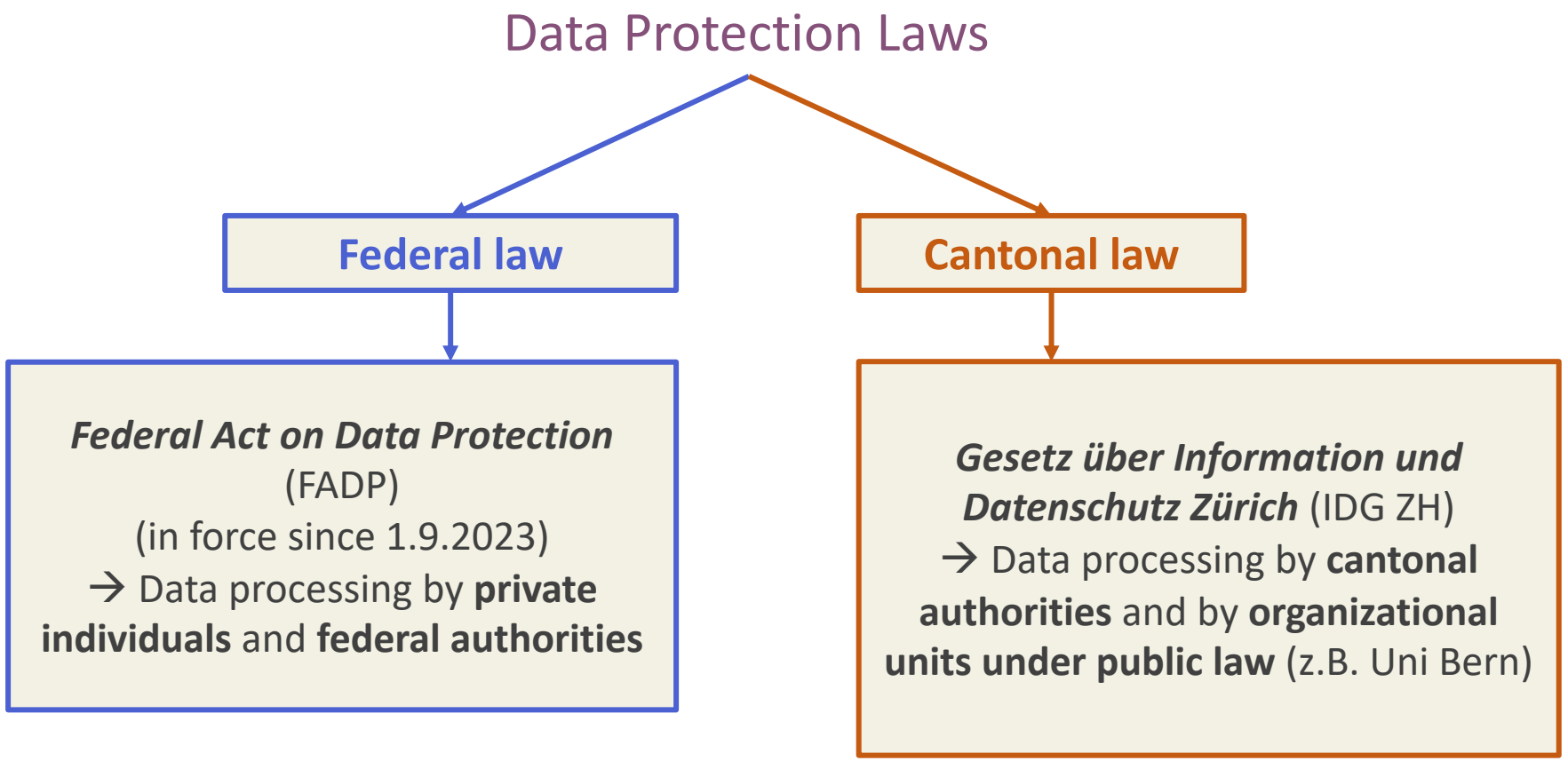
Data ownership

- There is no ownership as such over data.

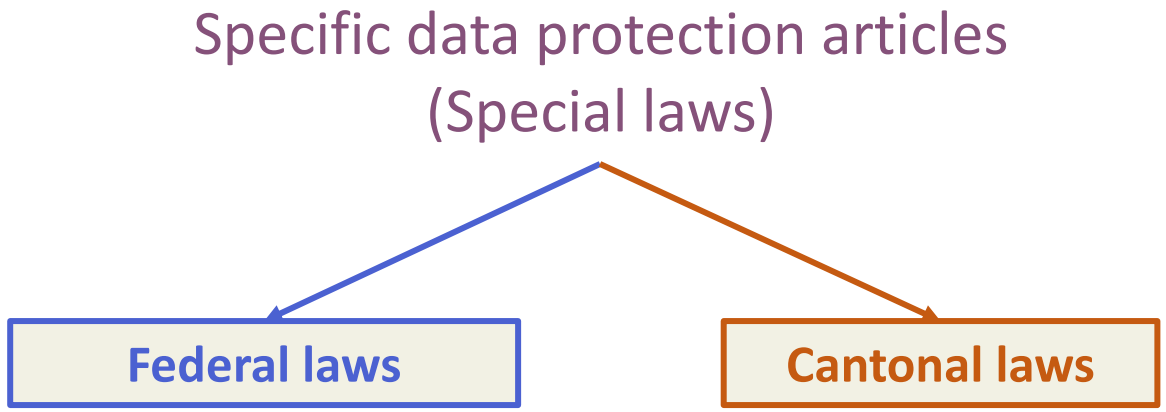
→ It is possible to agree privately on the use of data, considering the respective legal obligations.

Privacy and Personal data

Swiss legal acts



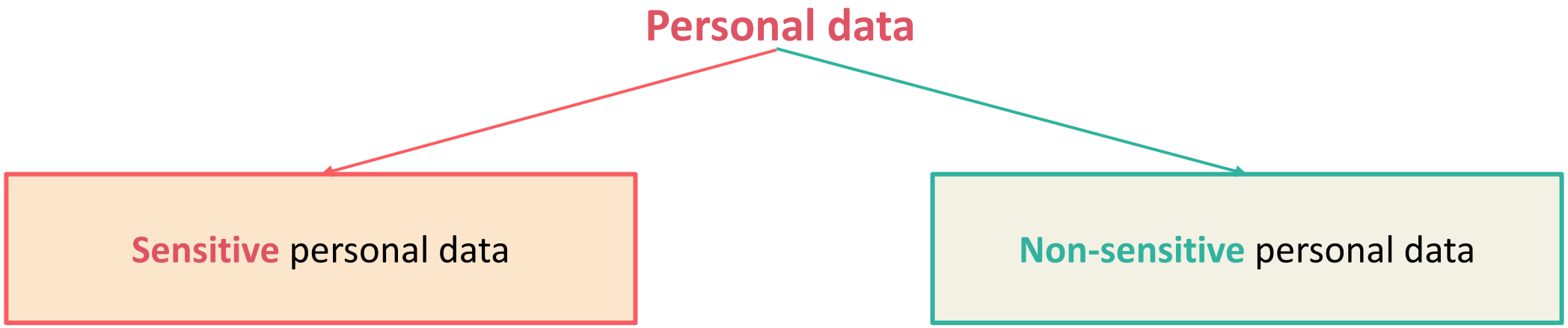
Swiss legal acts

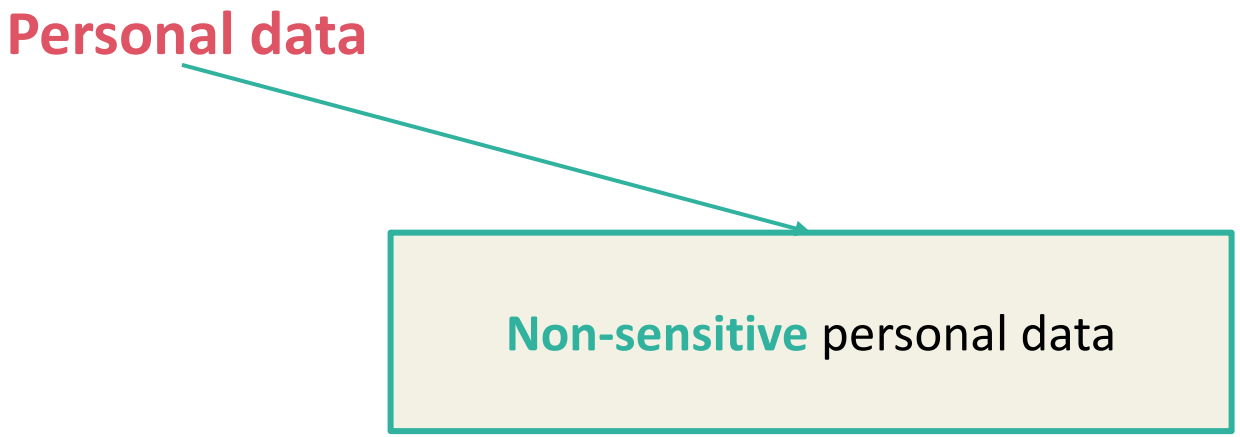


E.g. *Federal Act on Research involving Human Beings HRA*, *Federal Act on Health Insurance*

E.g. (Cantonal) *Police Act*

Personal data


Personal data
Sensitive personal data
Non-sensitive personal data



Personal data: **any** information relating to an **identified** or **identifiable** natural **person**

Personal data

Sensitive personal data

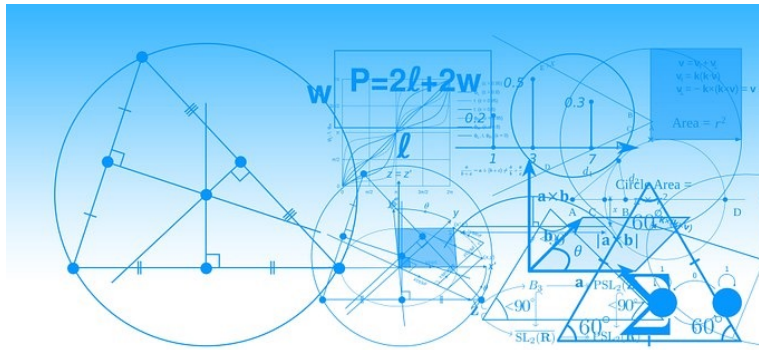
Information about:

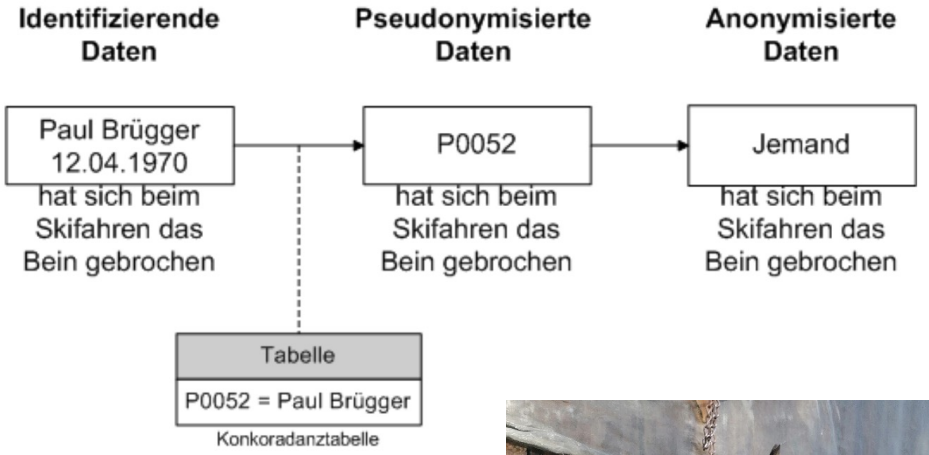
- Religious, philosophical, political or trade union-related views or activities;
- Health, the private sphere or affiliation to a race or ethnicity;
- Administrative and criminal proceedings or sanctions;
- Social assistance measures;
- Genetic data;
- Biometric data that uniquely identifies a natural person.

Personal data

- First name
- Last name
- Address
- Phone / fax
- E-Mail
- Customer number

Non-personal data

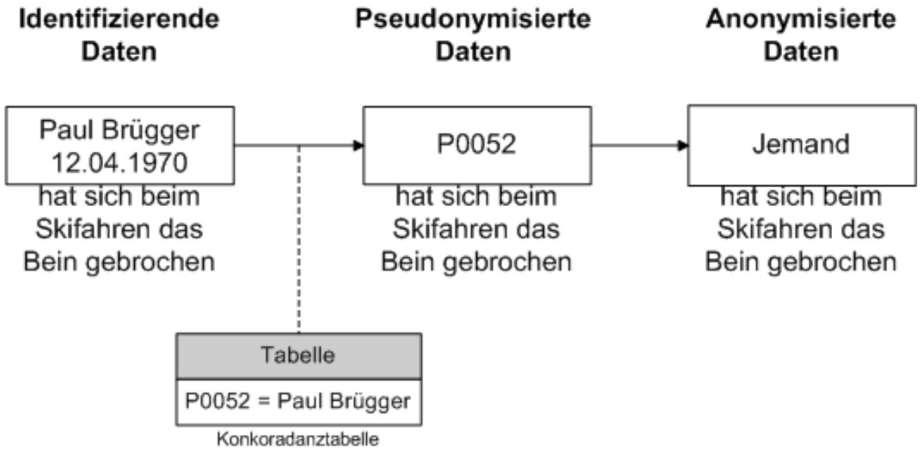




- Facts of nature
- Calculations
- Anonymised data

→ Data protection laws do NOT apply!

Anonymised vs. pseudonymized data



- **Anonymised** data

  → identification of individual persons is **impossible** or possible only with disproportionate effort.

- **Pseudonymised** data

  → identification of individual persons is **possible** with the use of a **key** or a **set of rules** to conceal identifying data.

Processing personal data

Processing personal Data

any handling of personal data, irrespective of the means and procedures used, in particular the **collection**, **storage**, keeping, **use**, **modification**, **disclosure**, archiving, deletion or destruction of data;

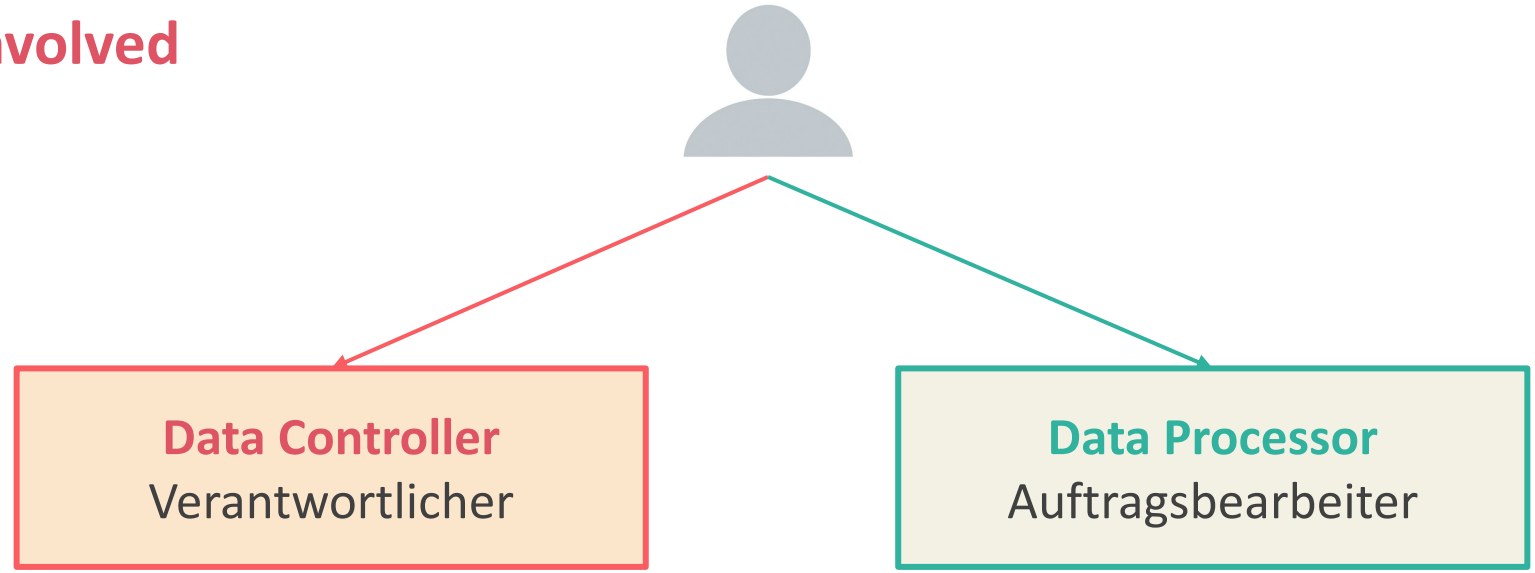
Data protection legislation

- Protects the **privacy of individuals** by protecting their data;
- Protects the data of all **living individuals** regardless of their relationship to the data processor → e.g. employee, student, website visitor, supplier, job applicant;
- Protection goes beyond maintaining confidentiality

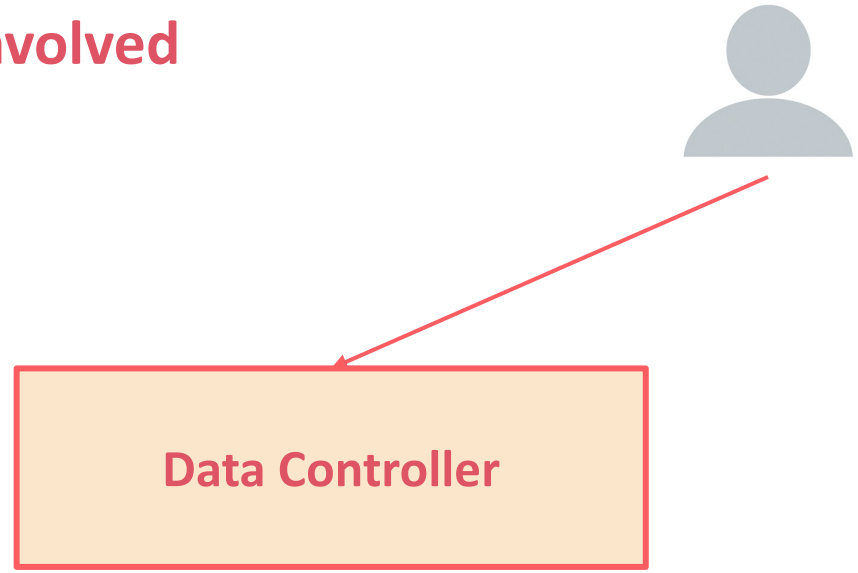
Data protection legislation

- Provides for obligations to be imposed on data controller and data processor (**privacy by default** and **privacy by design**)
- The data must be protected by **organizational** and **security measures** that are appropriate to the risk

Actors involved

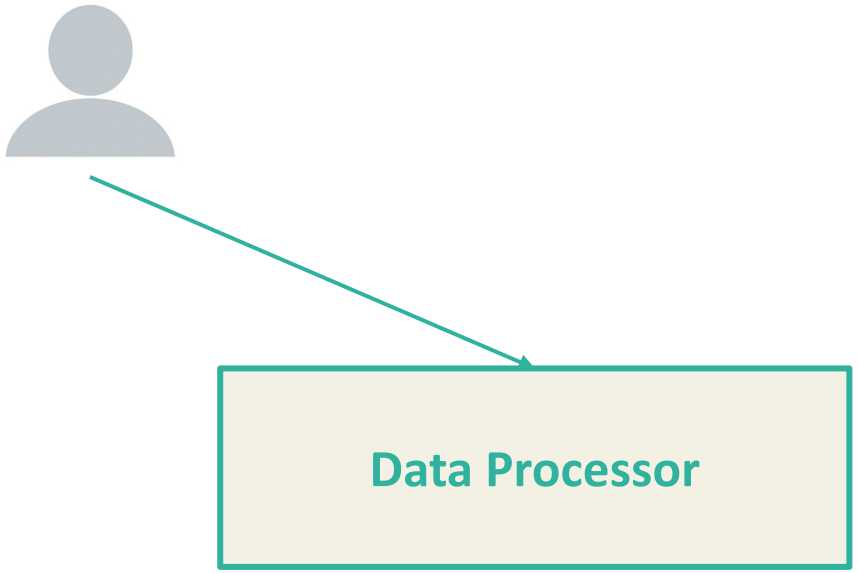


Actors involved



A private person who or federal body which, alone or jointly with others, **determines the purpose and the means** of processing personal data.
The controller must ensure that the data controller is able to guarantee **data security** and has the **duty to inform** data subjects about the processing of personal data.
(e.g. UZH or the project manager of a specific project)

Actors involved



A private person or federal body that **processes** personal data **on behalf of the controller** (e.g. the veterinarian who collects data from clients and the researcher who uses the data in the context of research.)

Duty to inform and informed consent

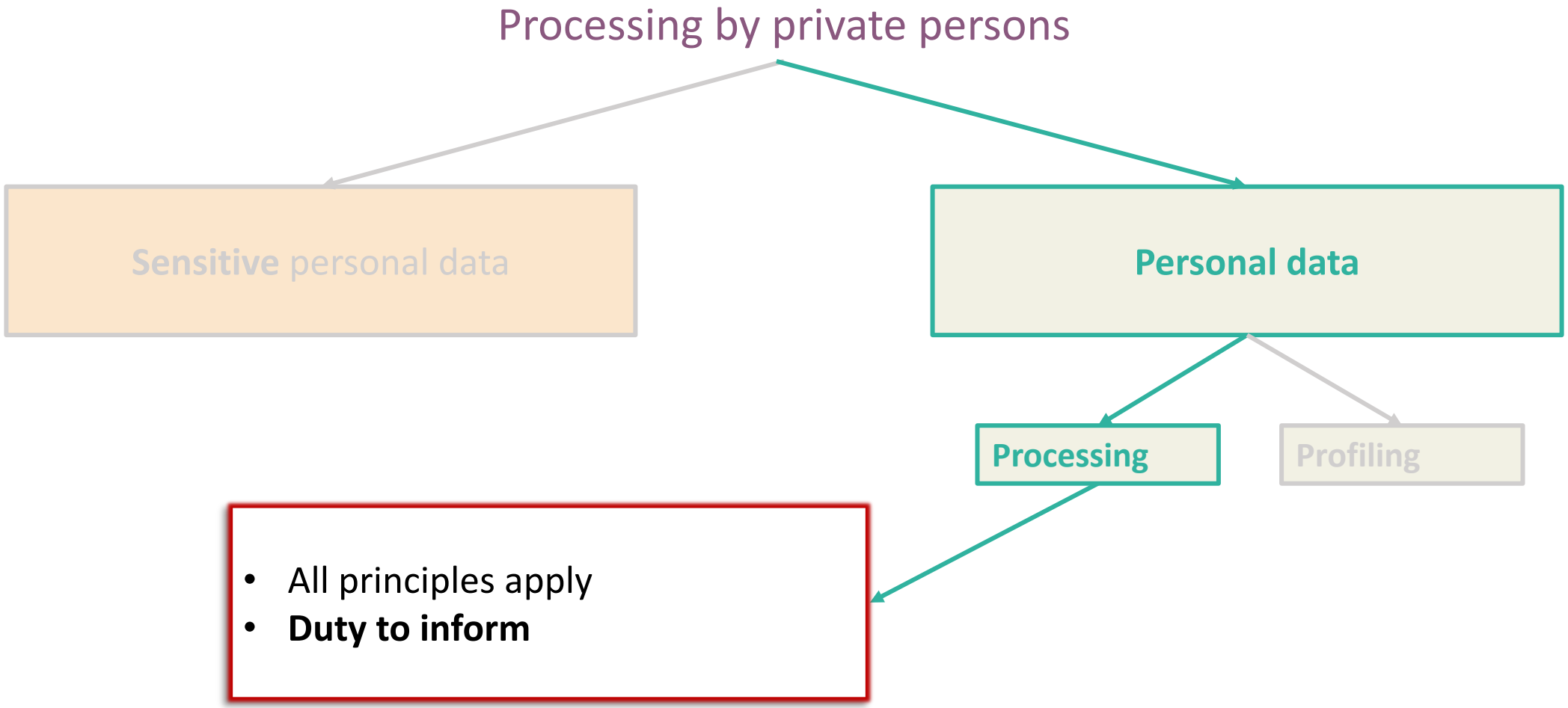
A legitimate ground is necessary to process personal data

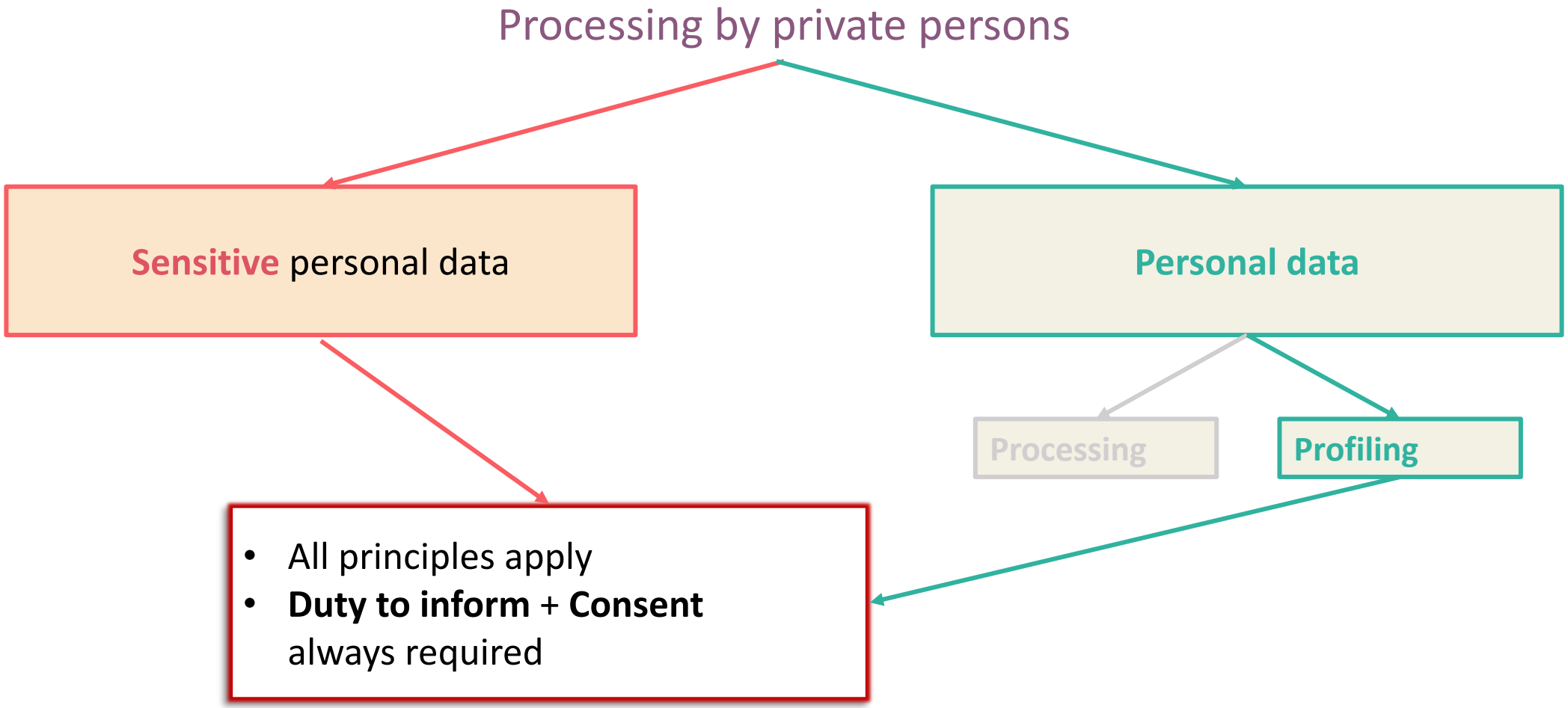
→ **Legal basis** (law)

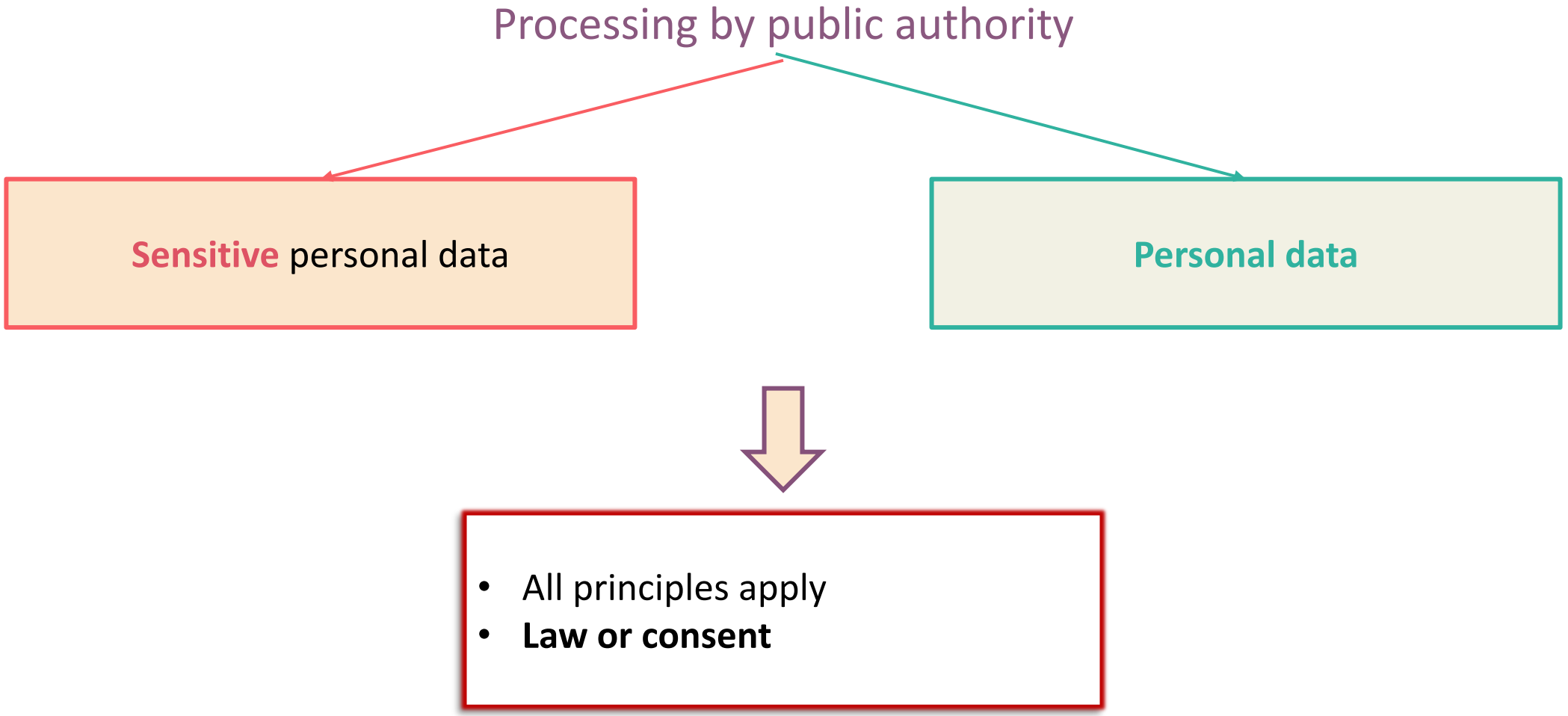
→ **Consent**

→ Overriding public or private interest

→ The person has published their personal data themselves and does not object to its use.


Processing by private persons
Sensitive personal data
Personal data
Processing
Profiling
All principles apply
Duty to inform


Processing by private persons
Sensitive personal data
Personal data
Processing
Profiling
• All principles apply
• **Duty to inform** + **Consent** always required


Processing by public authority
Sensitive personal data
Personal data
All principles apply
Law or consent

Data subject's rights

- Duty to inform
- Right of access
- Right to object to the use of personal data

Duty to inform

- No formal requirement, but "in an appropriate manner" (on a website? correspondence?) and you must be able to provide proof!
- At least the following information:
  - the **identity of the data controller**;
  - the **purpose** of the data processing;
  - the categories of data **recipients** (Art. 9 FDPA, e.g. cloud), if data transfer is planned;
  - When will personal data be **deleted** or anonymised.

Principles

- Personal data must be processed **lawfully**.

  → There is either a law or consent.

- Processing must be carried out in **good faith**.

  → If I provide information about the way in which I intend to use data, I must tell the truth.

- The processing must be **proportionate**.

  → I only collect the data that I need for the respective purpose.

Principles

- Personal data may only be processed for a **specific purpose**.

  → I can only use the data for the purpose for which I collected it.

- The data processor must be **transparent**.

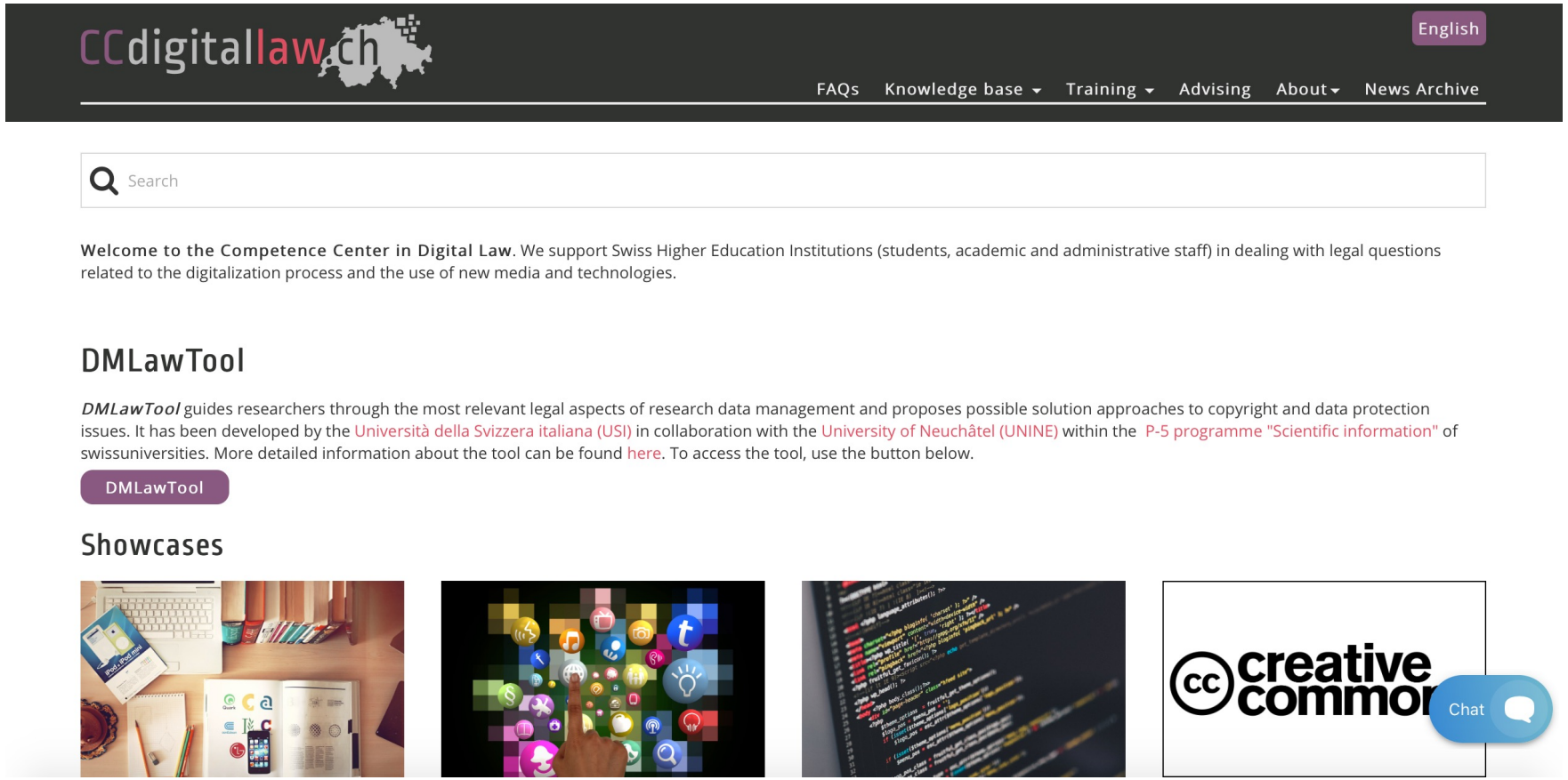
  → Information obligation = privacy policy

- If the data processor processes data, they must take all necessary **security measures**...

  → ...to prevent data from being lost or someone breaking into the system and stealing it.

Thank you for your attention!

www.ccdigitallaw.ch



Questions & Answers

Part two of the information event will start at 13:30