

Archiving and **D**ata **S**haring

Dr. Marieke Heers
November 4th, 2021

FORS webinar series



Presentation outline:

1. About data sharing and open research data
2. Why archive and share your data?
3. What and how to share?
4. The FAIR principles
5. Where to archive and share your data?
6. Sharing data with FORS

*About data sharing
and open research
data*

The new research environment

- Push towards open research and open data
- Requests for more transparency in social science research as many published results are not reproducible (e.g., the replication crisis in psychology)
- Researchers are asked to make their data and other research materials available, by funders, journals, and their peers

Open data

Who supports the idea of Open data?



Who practices open data?



Data sharing

Making research data available to other researchers at the end of a project unless there are ethical, legal or contractual reasons not to do so.

Why archive and share your data

Advantages of data sharing (1)

- Additional use of existing data.
- Allows for transparency, reproducibility, and verification of research findings.
- Ensures long-term preservation of data.
- Meets the requirements of some scientific funders and journals.

Benefits for researchers

- Increases the quality of your data and research.
- Makes research work and results more visible.
- Increases the number of citations of scientific articles for which research data are also published.
- Encourages new collaborations and new avenues of research.

What to share?

What data to share?

- Data that are publicly funded
- Data underlying publications
- Data of interest for further use



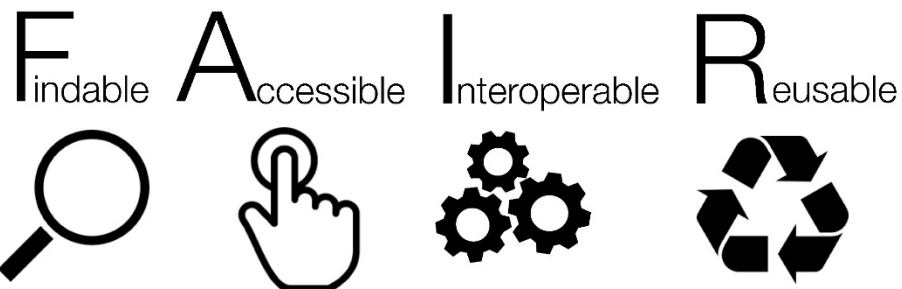
Re-use potential of data

- Amplified analysis: comparison or combination of several datasets
- For carrying out longitudinal research
- Matched analysis: re-use of data together with new data collection
- Further analysis: in-depth analysis of a question or aspect of the data that has not been fully analyzed in the primary study
- Teaching methods
- Working on non-reproducible materials
- Getting access to hard-to-reach populations

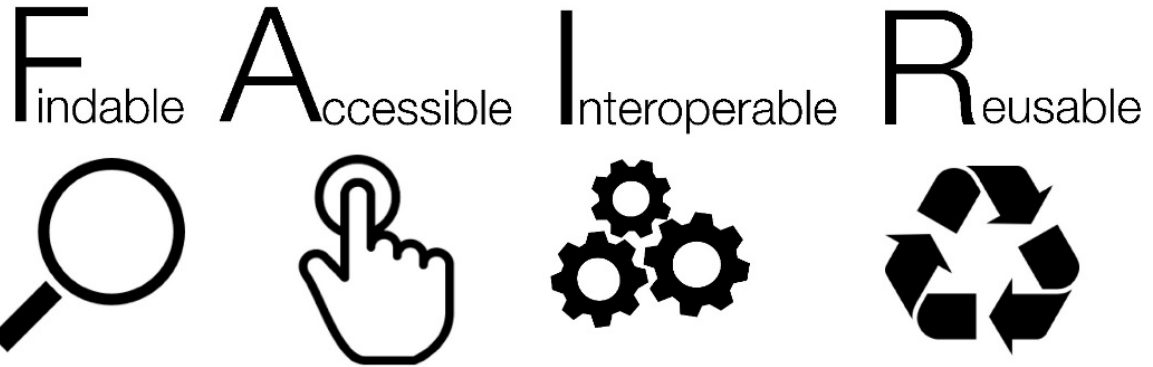
Requirements for data sharing

- Adequate documentation
- Clean data
- Valid informed consent
- Sufficient anonymisation in accordance with the informed consent
- Definition and implementation of access conditions
- Intellectual property considerations

- FAIR-principles

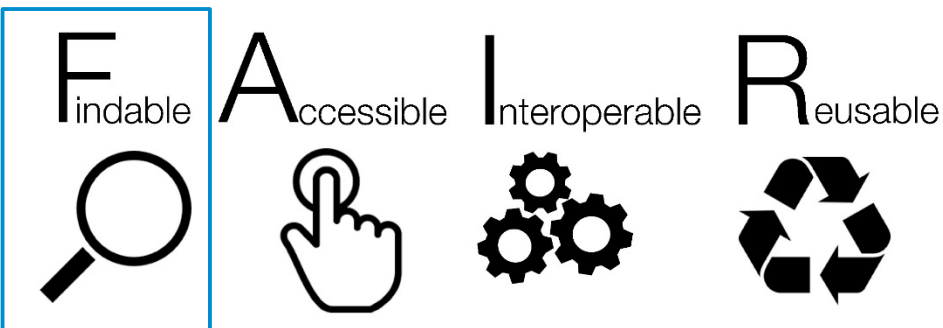


The FAIR principles



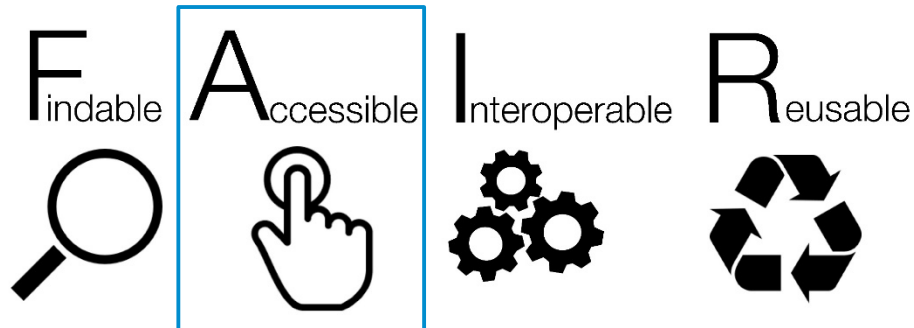
Findable

- The data can be found by both humans and machines. This can be achieved by:
 - Providing meaningful machine-actionable **metadata**
 - Referencing the data with a unique and **persistent identifier**



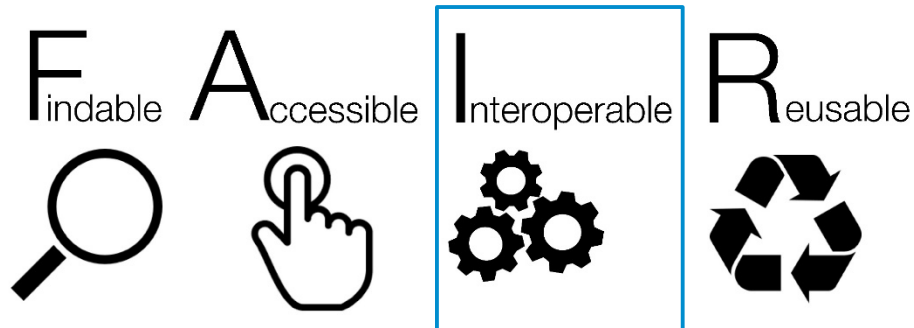
Accessible

- The data are available in long-term storage and can be made accessible using standard technical procedures. This requires providing information on **access conditions**.
- In this sense: open data is about ‘accessibility’ rather than ‘openness’.



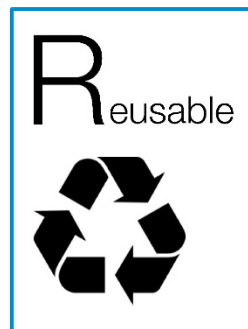
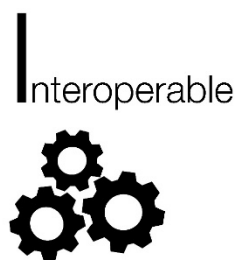
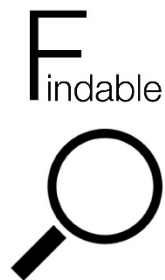
Interoperable

- The data can be exchanged and used across different applications and systems currently and in the future.
- The data can be integrated with other data from the same research field or data from other research fields.
- Interoperability is made possible by:
 - Using open **file formats**
 - Using **metadata** standards and controlled vocabularies



Reusable

- The data can be re-used for validating the results of the original study or to generate new research.
- Reusability is made possible by:
 - Providing strong **documentation** about the data and the context of creation
 - Including clear terms and conditions on how the data may be accessed and reused (licenses)



Applying the FAIR principles

Key features of FAIR data

- Documentation
- Metadata
- File formats
- Access to data
- Persistent identifiers

Key features of FAIR data

- **Documentation**
- Metadata
- File formats
- Access to data
- Persistent identifiers

Feature 1: Documentation

- Definition: any information that serves as a record of a research project and that renders data usable and meaningful.
- Potential audiences
 - the members of the original research team
 - secondary data users
- In the social sciences, documentation describes
 - the context of the study
 - the research process
 - how the data were collected and manipulated
 - the structure and the content of the data

What to document?

- Documentation should justify the needed effort.
- In deciding what and how much to document, try to answer:
 - How long will the project last? Might you need the data in the future?
 - Will the data be shared with others?
 - Who might need the information?
 - For what purposes is the information needed?
 - What might happen if the information was not available in the future – to you, to your team, to others?
 - What resources are available for the work of documentation?

Two levels of documentation

- Project-level documentation
- Data-level documentation

Project-level documentation

- The purpose of the study
- The context of data collection
- The content of the dataset
- The sampling
- Data collection methods
- Instrument construction
- Information on confidentiality
- Consent and anonymisation strategy
- Quality assurance procedures
- Important decisions taken within the research team
- Information on access and use conditions

Data-level documentation – quantitative data

- Names, labels and descriptions for variables
- Values for response categories
- Attribution of missing values
- Explanation of codes and classification schemes used
- Derived data created after collection, with code, algorithm or command file used
- Weighting variables and their application

IDNO	D16	D17	D14a	D1_1r	D1_2r	D1_3r	D2_01	D2_02	D2_03	Langue	paper	regling	Clim_q1a_p	Clim_q1b_p
5	1	27	14	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
8	2	35	6	21.00	-5.00	-5.00	13	2	15	1	.00	2	-8	-8
9	2	38	11	22.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
10	2	30	12	7.00	1.00	-5.00	2	14	14	1	.00	2	-8	-8
14	1	28	5	1.00	-5.00	-5.00	2	3	15	1	.00	2	-8	-8
16	2	35	5	23.00	1.00	-5.00	2	13	5	1	.00	2	-8	-8
19	2	45	13	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
23	1	77	5	1.00	-5.00	-5.00	2	14	14	1	.00	2	-8	-8
25	1	38	17	7.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
31	1	43	13	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
34	1	67	15	1.00	-5.00	-5.00	2	5	3	1	1.00	2	1	4
41	1	65	12	1.00	-5.00	-5.00	2	14	14	1	1.00	2	2	1
42	2	40	16	1.00	-5.00	-5.00	2	3	15	3	1.00	2	4	2
43	1	35	12	1.00	-5.00	-5.00	2	14	14	1	1.00	2	4	1
44	2	20	8	-2.00	1.00	-5.00	2	14	14	1	1.00	2	2	1
59	1	81	11	1.00	-5.00	-5.00	2	14	14	1	.00	2	-8	-8
60	1	48	10	1.00	-5.00	-5.00	1	2	15	1	1.00	2	2	4
62	1	60	5	2.00	-5.00	-5.00	2	14	14	1	1.00	2	1	1
64	2	36	3	5.00	1.00	-5.00	2	14	14	1	1.00	2	2	4
68	1	29	12	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	4
70	2	31	5	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
72	2	46	17	1.00	-5.00	-5.00	1	14	14	1	.00	2	-8	-8
76	2	33	8	22.00	-5.00	-5.00	2	14	14	1	1.00	2	1	3
79	1	60	13	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
80	2	54	5	1.00	-5.00	-5.00	2	1	7	1	.00	2	-8	-8
95	1	63	5	1.00	-5.00	-5.00	2	1	6	1	1.00	2	1	2
97	1	20	3	3.00	-5.00	-5.00	2	8	15	1	.00	2	-8	-8
99	1	34	5	1.00	-5.00	-5.00	2	14	14	1	.00	2	-8	-8
100	2	45	8	1.00	-5.00	-5.00	2	14	14	1	.00	2	-8	-8
101	2	49	5	1.00	-5.00	-5.00	2	14	14	1	1.00	2	-1	-1
103	1	55	14	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
110	2	29	16	7.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
112	1	45	3	3.00	-5.00	-5.00	2	8	7	1	.00	2	-8	-8
113	2	51	16	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	2
123	2	54	5	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	-1
126	2	54	5	1.00	-5.00	-5.00	2	14	14	1	1.00	2	2	1
127	1	19	11	1.00	-5.00	-5.00	2	14	14	1	1.00	2	2	1
130	2	31	16	-2.00	2.00	1.00	5	2	15	1	1.00	2	2	1
131	2	60	5	1.00	-5.00	-5.00	2	14	14	1	1.00	2	1	4
140	2	88	15	1.00	-5.00	-5.00	2	14	14	1	.00	2	-8	-8
142	2	83	6	1.00	-5.00	-5.00	2	14	14	1	.00	2	-8	-8
144	1	51	15	1.00	-5.00	-5.00	2	14	14	1	1.00	2	2	1
150	1	39	14	2.00	-5.00	-5.00	2	3	15	1	1.00	2	2	1

Data-level documentation – qualitative data

- Qualitative data must be documented if they are to be used properly.
- Crucial elements include:
 - Interview setting (participants, location, time of day, etc.)
 - Criteria and layout for transcriptions of interviews
 - Background description of respondents

Practical recommendations

- Documentation should be part of the research project from the beginning and be a constant component of the project
- Rules on documentation should be fixed from the start of the project and updated on a regular basis

Key features of FAIR data

- Documentation
- **Metadata**
- File formats
- Access to data
- Persistent identifiers

Feature 2: Metadata

- Definition: Metadata is data about the data.
- It applies to the level of the project and the level of the data.

Metadata for the social sciences

- Descriptors that facilitate cataloguing data and data discovery
- Unlike documentation, metadata are a formally agreed set of standards often with controlled fields and vocabularies
- Machine-readable metadata help to explain the purpose, origin, methods, time, location, terms of use, and access conditions of research data.
- Some fields may be mandatory (title, principal investigator), others recommended (language), and some optional (embargo period).
- There are different standards. Within the social, behavioral and economic sciences the metadata standard of the Data Documentation Initiative (DDI) is mainly used.

Key features of FAIR data

- Documentation
- Metadata
- **File formats**
- Access to data
- Persistent identifiers

Feature 3: File formats

- File formats determine how data can be used.
- File format choice depends on the research phase.
- It is important to decide what file formats to use for data collection, data processing, data archiving, and long-term preservation.
- Following discipline-specific standards is generally the way to go.
 - How widespread are these standards?
 - To what extent will they allow data processing by others?

Be careful with proprietary file formats!

Key features of FAIR data

- Documentation
- Metadata
- File formats
- **Access to data**
- Persistent identifiers

Feature 4: Access to data

- You determine who you make your data available to, how you provide access, and under which conditions.
- Most repositories provide for different access conditions, including:
 - **Open access**: data can be accessed by anyone
 - **Access upon registration**: data can only be accessed by registered users
 - **Restricted access**: access is limited and can only be granted upon request
 - Access after an **embargo** period

Key features of FAIR data

- Documentation
- Metadata
- File formats
- Access to data
- **Persistent identifiers**

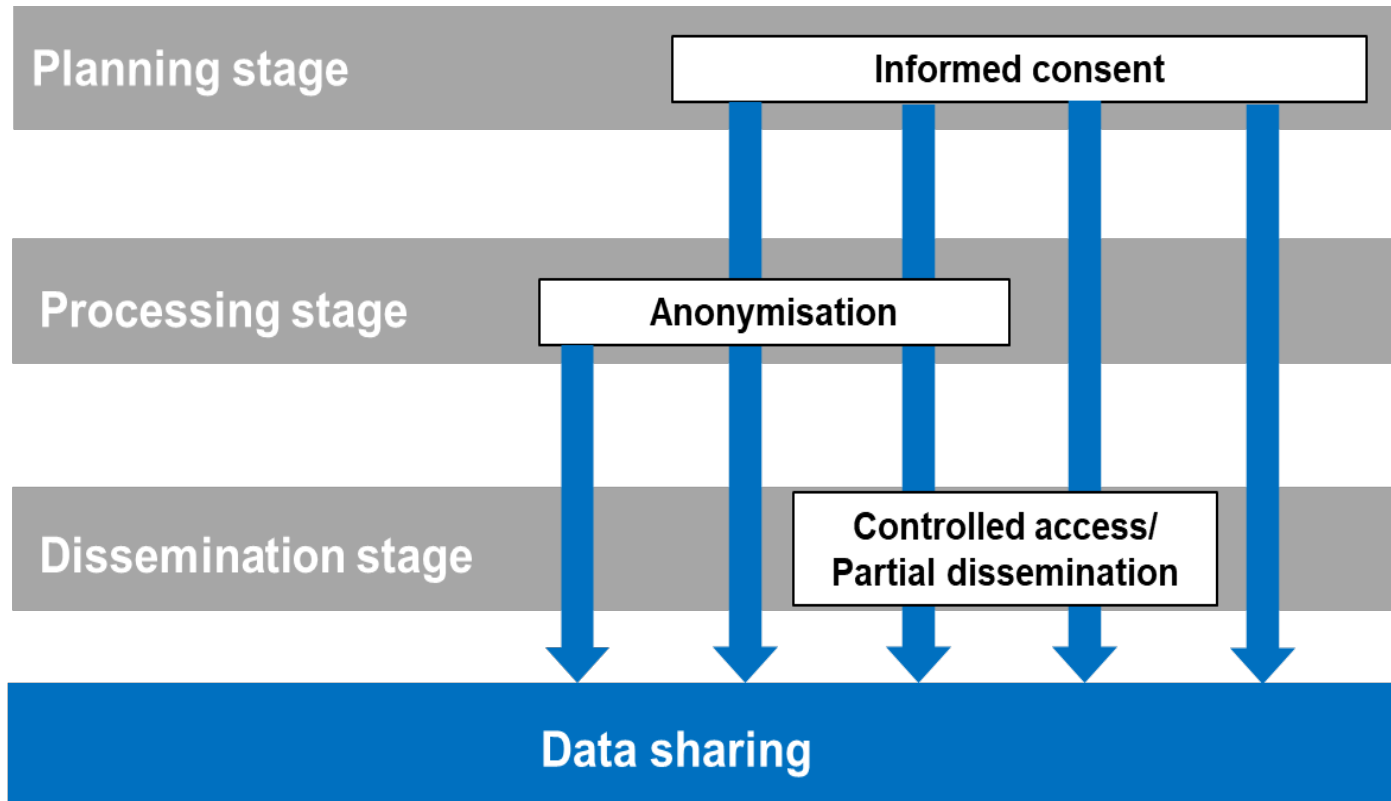
Feature 5: Persistent identifiers

- To make your data accessible and easy to find, provide your data and metadata with a persistent identifier (PID).
- A persistent identifier is a long-lasting numeric link to your dataset.
- A commonly used persistent identifier is a DOI (Digital Object Identifier).
- Adding a DOI to your data will allow others to easily find and access your data, as well as cite the data and give you credit for collecting them.

Consent for data sharing

Informed consent for sharing

On top of asking research participants their consent for taking part in the study, they should also consent for their data to be archived and shared.



*Where to archive
and share your
data?*

Where to share data?

- FAIR and non-commercial repository
- Different solutions, some better than others
 - (trusted) domain-specific repository
 - institutional repository
 - general purpose repository (e.g., Figshare)
 - project archives
 - journal supplementary material services

Where to share data?

- Privilege domain-specific repositories or institutional repositories
- Make sure the chosen repository offers:
 - long-term preservation
 - persistent identifiers
 - visibility of your data
 - checking of data and documentation (quality control)
 - catalogue for discovery
 - dissemination capacity
 - access control

A full list of repositories can be found on [Re3data.org](https://re3data.org)

Benefits of sharing your data through an archive

- Key practices to make data FAIR will be taken care of
 - (Documentation)
 - Metadata
 - File formats
 - Access to data
 - Persistent identifiers
- Contracts for depositors and users
- Promotion and dissemination of the data

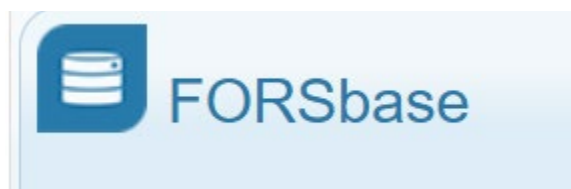
Sharing data with FORS

Sharing your data with FORS

Services and support with respect to the following:

- Documentation
- Metadata
- File formats
- Access to data
- Persistent identifiers

SWISSUbase



Webinar on SWISSUBase beginning of 2022

Some advantages of sharing with FORS

Go with a trusted Swiss solution

- ✓ 100% Swiss, in line with Swiss legal requirements, data on Swiss servers, Swiss quality
- ✓ FAIR repository recognized by the Swiss National Science Foundation
- ✓ Curated by archive experts
- ✓ International certification of quality (CoreTrustSeal)

Protect your rights and keep control over your data

- ✓ Enjoy long-term preservation of your data – not limited in time
- ✓ Access control options (your prior approval, embargo, restrictions on use)
- ✓ Protection of depositors' rights (binding deposit and user contracts)
- ✓ Access statistics on data downloads and the use of your data

Increase your visibility

- ✓ Get free persistent identifiers (DOIs) for all published datasets to increase your citation rates
- ✓ Get assistance with documenting your data according to international, field-specific standards
- ✓ Benefit from the visibility, outreach, and international exposure of FORSbase, the main data repository for the social sciences in Switzerland

Take advantage of our personalized services throughout the research lifecycle

- ✓ We help you find and access the data and research information you are looking for
- ✓ We assist with the deposit and sharing of your data
- ✓ We provide data management consulting
- ✓ Our services are free of charge

Some advantages of sharing with FORS

Go with a trusted Swiss solution

- ✓ 100% Swiss, in line with Swiss legal requirements, data on Swiss servers, Swiss quality
- ✓ FAIR repository recognized by the Swiss National Science Foundation
- ✓ Curated by archive experts
- ✓ International certification of quality (CoreTrustSeal)

Protect your data over your lifetime

- ✓ Enjoy long-term access, not limited by time or space
- ✓ Access control and restriction options
- ✓ Protection of user confidentiality
- ✓ Access statistics on data downloads and the use of your data

Increase your visibility

Go with a trusted Swiss solution

- ✓ 100% Swiss, in line with Swiss legal requirements, data on Swiss servers, Swiss quality
- ✓ FAIR repository recognized by the Swiss National Science Foundation
- ✓ Curated by archive experts
- ✓ International certification of quality (CoreTrustSeal)

Some advantages of sharing with FORS

Go with a trusted Swiss solution

- ✓ 100% Swiss, in line with Swiss legal requirements, data on Swiss servers
- ✓ FAIR repository according to the Science Foundation of Switzerland
- ✓ Curated by an international team
- ✓ International data standards

Protect you over your data

- ✓ Enjoy long-term access, not limited in time
- ✓ Access control, no restrictions on data reuse
- ✓ Protection of user contract
- ✓ Access statistics on data downloads and the use of your data

Increase your visibility

- ✓ Get free persistent identifiers (DOIs) for all published datasets

Increase your visibility

- ✓ Get free persistent identifiers (DOIs) for all published datasets to increase your citation rates
- ✓ Get assistance with documenting your data according to international, field-specific standards
- ✓ Benefit from the visibility, outreach, and international exposure of FORSbase, the main data repository for the social sciences in Switzerland

according

international
repository for

standardized
data lifecycle

research

your data

Some advantages of sharing with FORS

Go with a trusted Swiss solution

- ✓ 100% Swiss in line with Swiss legal requirements on Swiss servers
- ✓ FAIR repository Science Foundation Europe
- ✓ Curated by
- ✓ International

Protect your rights over your data

- ✓ Enjoy long-term preservation – not limited
- ✓ Access control options (your prior approval, embargo, restrictions on use)
- ✓ Protection of depositors' rights (binding deposit and user contracts)
- ✓ Access statistics on data downloads and the use of your data

Increase your visibility

- ✓ Get free persistent identifiers (DOI) for all published

Protect your rights and keep control over your data

- ✓ Enjoy long-term preservation of your data – not limited in time
- ✓ Access control options (your prior approval, embargo, restrictions on use)
- ✓ Protection of depositors' rights (binding deposit and user contracts)
- ✓ Access statistics on data downloads and the use of your data

a according

ternational
ository for

alized
ch lifecycle

research

your data

Some advantages of sharing with FORS

Go with a trusted Swiss solution

- ✓ 100% Swiss - in line with Swiss legal requirements, data on Swiss
- ✓ FAIR re Science
- ✓ Curatec
- ✓ Internal

Protect over yo

- ✓ Enjoy Ic not limi
- ✓ Access restricti
- ✓ Protect user co
- ✓ Access statistics on data downloads and the use of your data

Increase your visibility

- ✓ Get free persistent identifiers (DOI) for all published

Take advantage of our personalized services throughout the research lifecycle

- ✓ We help you find and access the data and research information you are looking for
- ✓ We assist with the deposit and sharing of your data
- ✓ We provide data management consulting
- ✓ Our services are free of charge

Resources

Main resources

UK data service: <https://ukdataservice.ac.uk/learning-hub/research-data-management/>

Competence Center in Digital Law: <https://ccdigitallaw.ch>

CESSDA expert guide: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>

University of Portsmouth:
<https://library.port.ac.uk/researchdata.html>

How to FAIR: <https://www.howtofair.dk/what-is-fair/>

Thank you!

rdmservices@fors.unil.ch

Questions?