TRAINING: Data Science and AI for Medicine Training School 2026
              Day 2: RDM

SPEAKER: Lea Kabjesz, Maryam Khodaei Dolouei, Toralf Kirsten

AGENDA

- **Theory and Terms**
  - FAIR-Principles

- **Medical Data**

- **Practice**
  - Data Types & Principles in RDM

- **MII Academy**

Sneak Preview ;-)

Basics

What does it actually mean?

Research Data Management (RDM)

Legal Studies

copyright   licensing

Ethics

consent   privacy

transparency

research integrity

IT & Cybersecurity

access control   storage

data protection   backup

RDM

Data Science

data quality   curation

metadata   analysis/ AI readiness

Project Management

planning   documentation   workflows

Principles in RDM

FAIR data should be...

FINDABLE — easy to discover
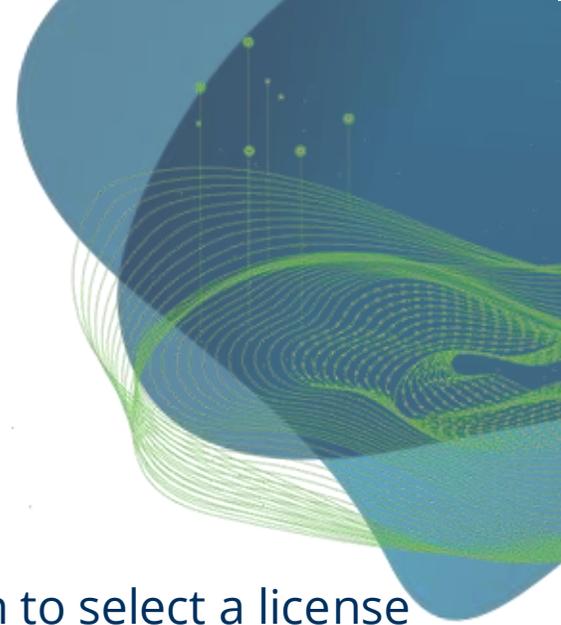ACCESSIBLE — easy to retrieve
INTEROPERABLE — easy to combine with other data or tools
REUSABLE — easy to use again (by people + computers)

➡ As **open** as possible, as **closed** as necessary!

Guidelines

1) Register datasets in public repositories
   - use trusted repositories with clearly defined access conditions and option to select a license
   - choose a repo that assigns a **DOI** (**= permanent** string **uniquely** identifying a digital object)

https://doi.org/10.1038/s41392-020-00243-2

DOI

URL

Everything is data but <u>where</u> can I upload it?

Data Repositories

for code

for images

for posters, presentations

for DL models, datasets

→ next talks

special considerations
for medical data

Guidelines
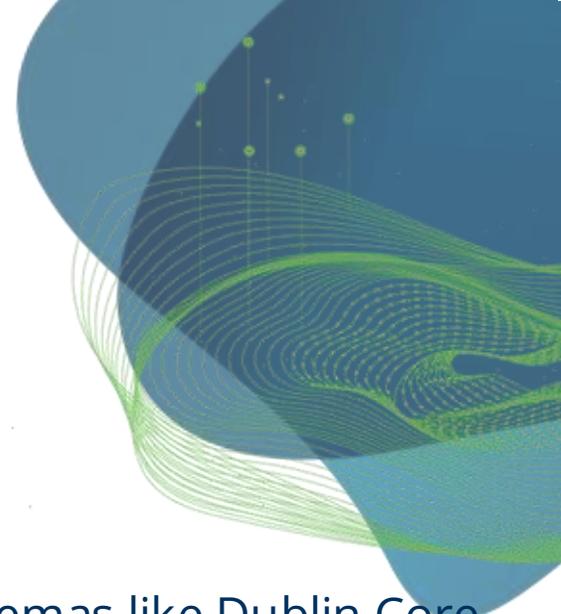
2) Provide documentation & metadata (= data about data)
   - describe context, data structure, variables, methods using metadata schemas like Dublin Core
   - use tools like ReadMe files, Jupyter Notebooks

→ watch out for open standards, file formats (e.g. CSV, TXT, JSON, XML, PDF) & schemas!
→ watch out for consistent file naming & version control!

All of this is possible...

... but please stick to this:

or this

INCENTIVES!

Increased visibility.

Higher citation of data and publications.

More recognition in the field.

Easier collaboration.

Fostering new discoveries.

Saves time.

RDM TOOLS

OMERO

= An **open-source** software for **image data** management

Riccardo Massei

https://zenodo.org/records/16880913

Tutorials & Help

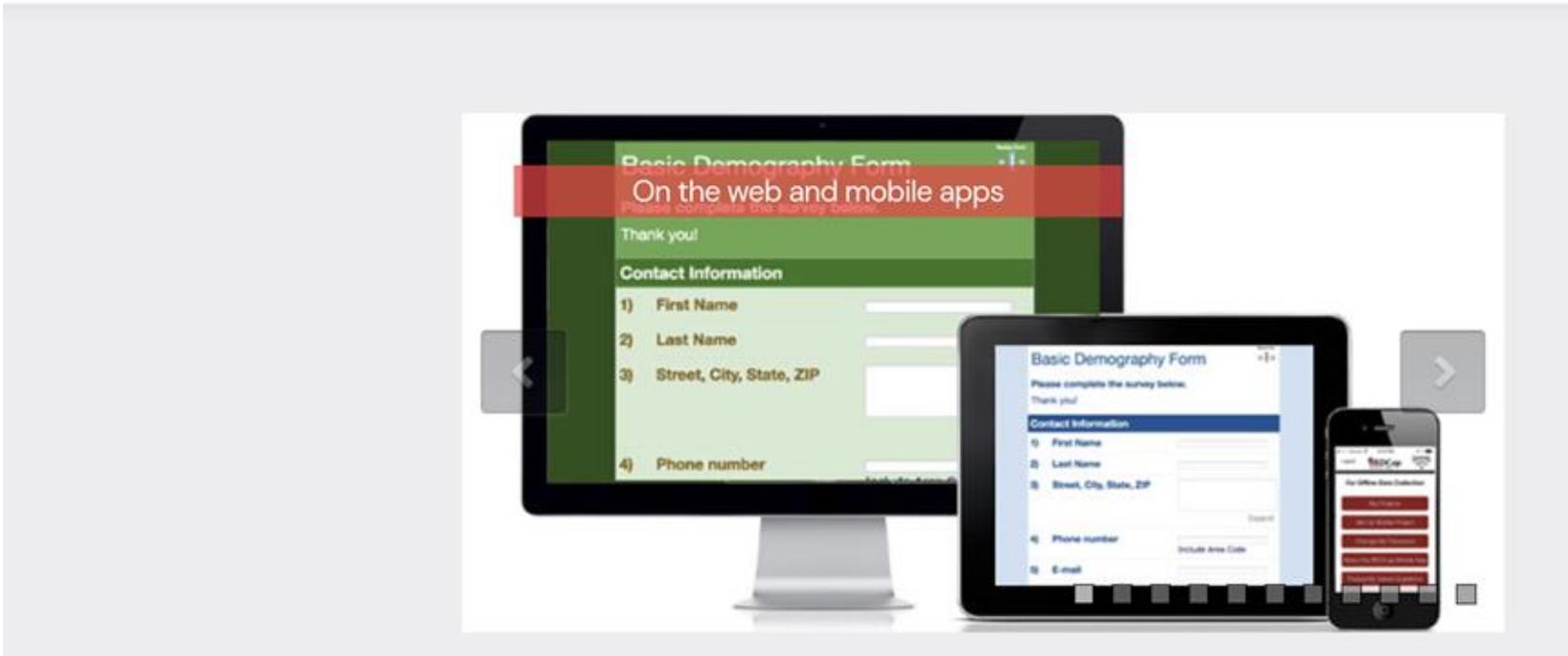https://omero-guides.readthedocs.io/projects/introduction/en/latest/data-management.html
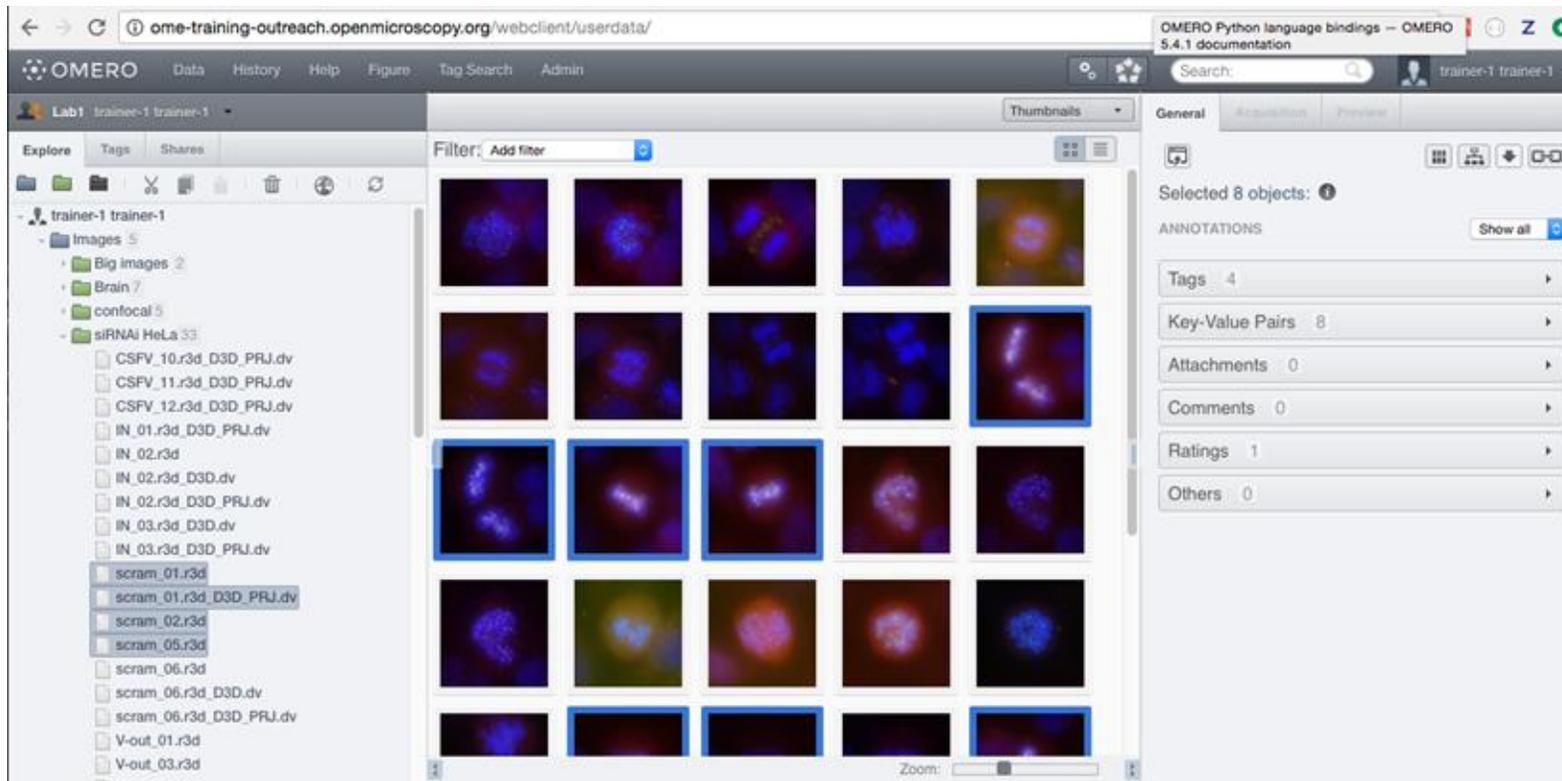
REDCap

= A **secure** web application for building and managing **survey data**

Christof Meigen

Contact & Help

Further Reading

Guides

- RDM-Guide (Uni Leipzig):
  https://www.uni-leipzig.de/fileadmin/ul/Dokumente/190424_Grunds%C3%A4tze_Forschungsdatenmanagement.pdf

- Medical RDM-Guide (DFG):
  https://zenodo.org/records/4912388#.YPazF5dxfcu

Contact Points

- RDM-Services (Uni Leipzig)
  https://www.uni-leipzig.de/forschung/forschungsservice/forschungsdatenmanagement

- RDM-Services (Saxony)
  https://saxfdm.de/

- RDM-Services (Germany)
  https://www.nfdi.de/consortia/

Take Home Messages

If you want to make your data *findable*:

- use **unique identifiers (DOIs), rich metadata, and searchable repositories**.

If you want to make your data *accessible*:

- use **open <u>or controlled</u>-access repositories and <u>clear</u> retrieval protocols**.

If you want to make your data *interoperable*:

- use **standard formats, vocabularies, and ontologies compatible with other datasets**.

If you want to make your data *reusable*:

- use **detailed documentation, clear licensing, sufficient metadata for others to understand and use it**.

Keep FAIR in mind!

Finding Data Repositories

- re3data.org
- for storing data but also useful for finding new data

Filter for discipline
+ DOI assignment!

Registering data in a Repo ≠ Backup!

Archiving vs. **Backup**

long-term preservation for **10 years** (DFG).

data **recovery** after **loss** or **corruption**.
usually **three copies** stored in **separate** locations.

Acknowledgments