NFDI4Chem

ENHANCE
YOUR
DATA.

Overview of research data management in chemistry

Workshop topics

- Introduction
- Research data management basics
- FAIR principles
- Data life cycle
- NFDI4Chem

NFDI4Chem

ENHANCE
YOUR
DATA.

Introduction

NFDI4Chem: Involvement of the community

Local Helpdesks

Newsletters

Website

Social Media

RDM Ambassadors

Best Practice

Flagship Labs

Requirement Analysis

Workshops

Training

Tutorials

Roadshows

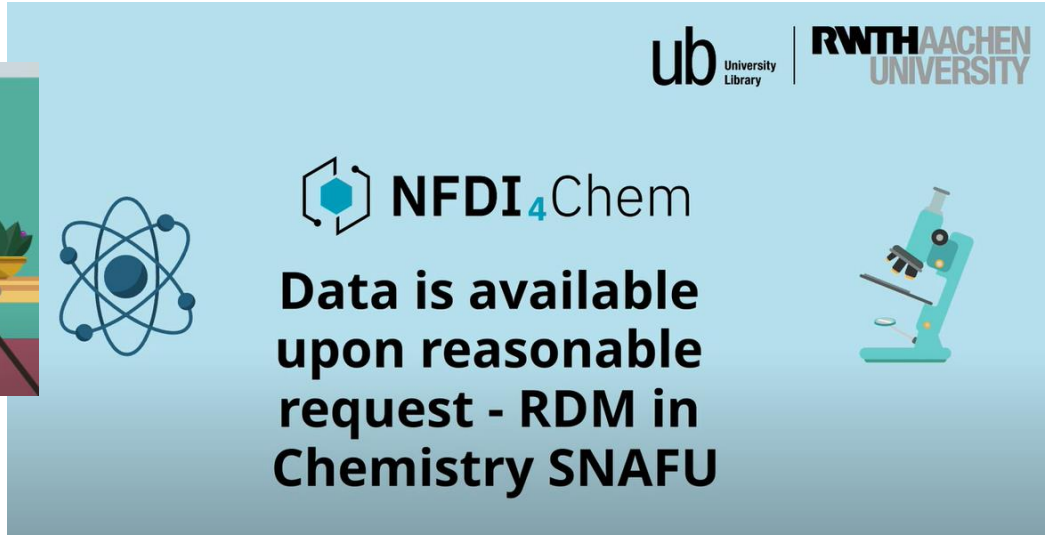
Teaching Materials

Curricular Embedding

Motivation

- NFDI4Chem - Data is available upon reasonable request - RDM in Chemistry SNAFU by UB RWTH Aachen University

https://www.youtube.com/watch?v=OA0GcvacjqI



Workshop motivation

I am attending this workshop to learn more about

a) Research data management in general

b) A specific research data management topic

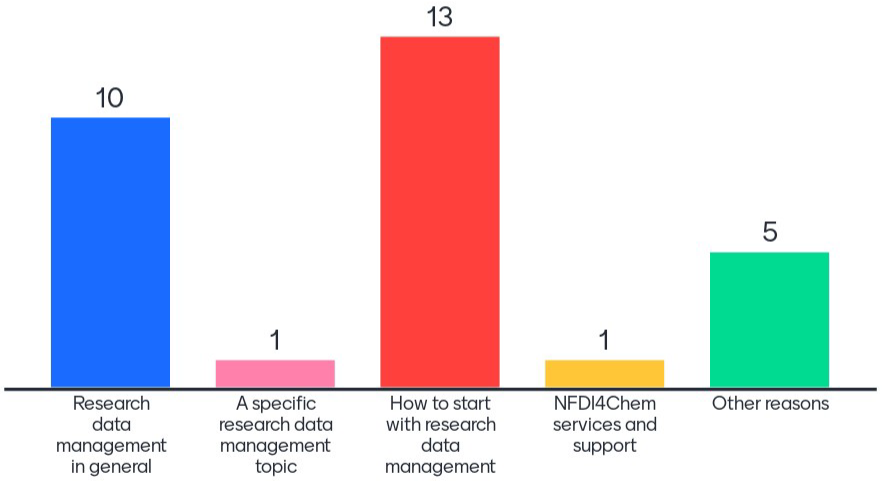
c) How to start with research data management

d) NFDI4Chem services and support

e) Other reasons

I am attending this workshop to learn more about

Mentimeter



21

NFDI4Chem

ENHANCE
YOUR
DATA.

RDM basics

Research data

- Difficult to define
- Discipline-specific definitions

DFG:

"Research data include **measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections or samples** that are **created, developed or evaluated** in scientific work. **Methodological test procedures** such as **questionnaires, software** and **simulations** can also represent central results of scientific research and should therefore also be included under the term research data. "

Research data in chemistry

electrochemical data

spectroscopy data

simulations

laboratory notebooks

database content

chromatographic data

models
algorithms

…everything, that is result and target of your scientific work

topographical data

interferometric data

experimental data

crystallographic data

ellipsometric data

software

microscopy data

Research data management

- Activity of working with research data throughout the research process
- Including all aspects from data collection, to data storage and backup, through to data sharing
- One of the essential areas of responsible conduct of research

"Although handling research data can be challenging, managing your data effectively will not only help your **research to be robust and replicable**, but can help you to **anticipate potential problems** that can occur during the research process, and will **ensure that your research meets the requirements set out by research funders and publishers**."

https://mantra.ed.ac.uk/researchdataincontext/

Research data management – Why?

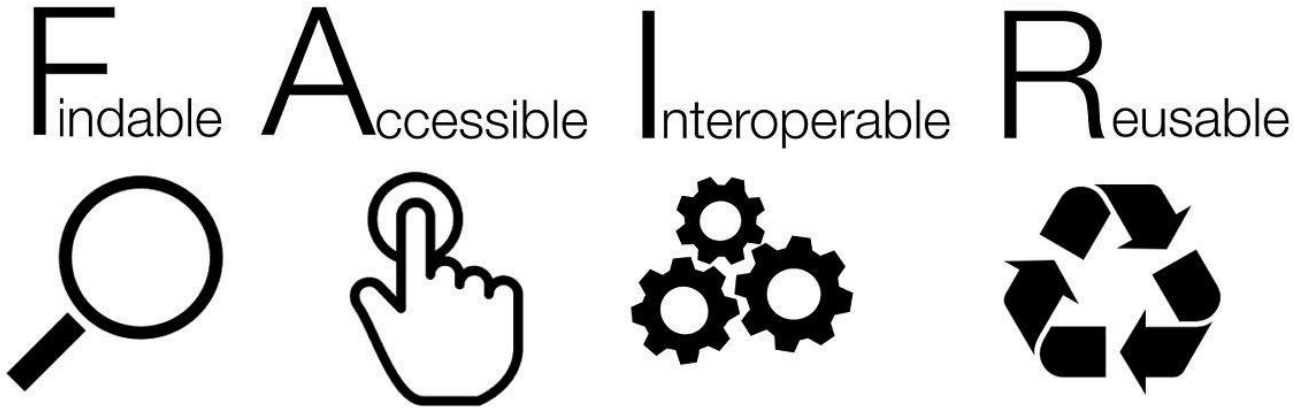
- Data security and prevention of data loss
- Verifiability, reproducibility and transparency of research results
- Reduction of scientific errors
- Faster retrieval of data and information
- Long-term availability of research data
- Data re-use in new research projects
- **Required in guidelines and institutional policies on handling research data**
- **Requirement of third-party funders and science organisations**

RDM basics

FAIR principles

FAIR principles

- According to the DFG‘s new guidelines on good research practice, research data must be FAIR!



https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

FAIR principles in detail

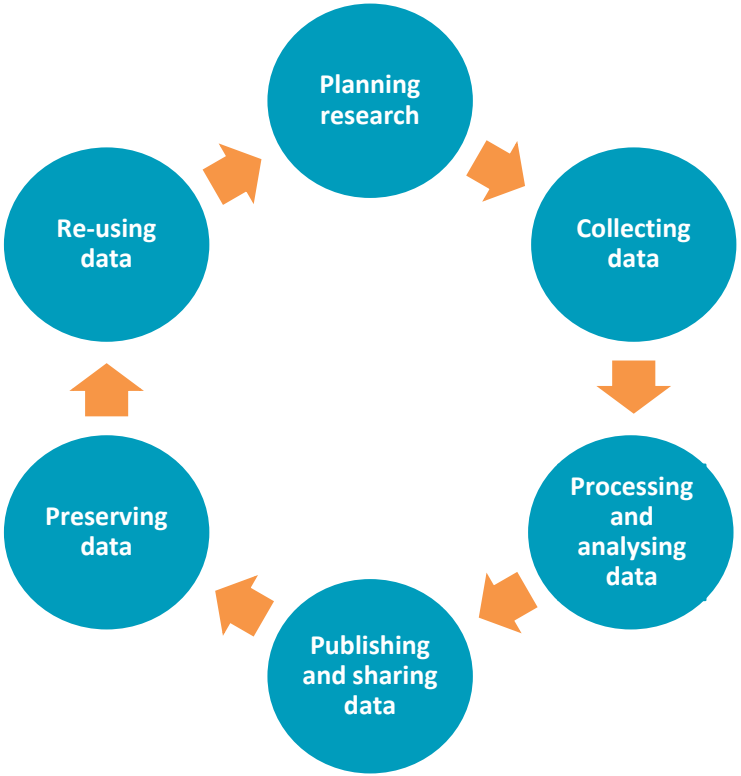
FINDABLE	ACCESSIBLE
F1. **(Meta)data** are assigned a globally unique and persistent identifier F2. Data are described with rich **metadata** (defined by R1 below) F3. **Metadata** clearly and explicitly include the identifier of the data they describe F4. **(Meta)data** are registered or indexed in a searchable resource	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
INTEROPERABLE	**REUSABLE**
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**	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

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

RDM basics

Data life cycle

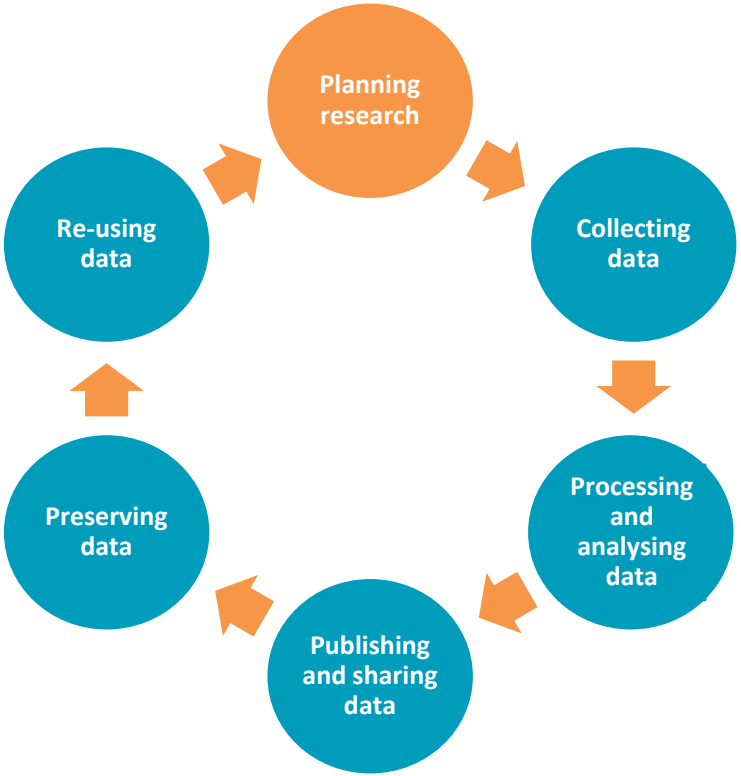
Data life cycle



- A key concept in research data management
- Describes the lifespan of the data and beyond
- Based on various phases
- Different approaches to the same model depending on the institution, the funder, …

https://knowledgebase.nfdi4chem.de/knowledge_base/docs/topics/data_life_cycle/

Planning research



- Research design
- Planning data management (formats, storage locations, ...)
- Create an initial **data management plan**
- Determine responsibilities
- Locate existing data
- Clarify authorship and data ownership
- Coordinate access conditions, prepare consent procedures

Planning research

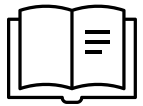
Data management plan

What is a Data Management Plan (DMP)?

Description of the handling of research data during and after a research project

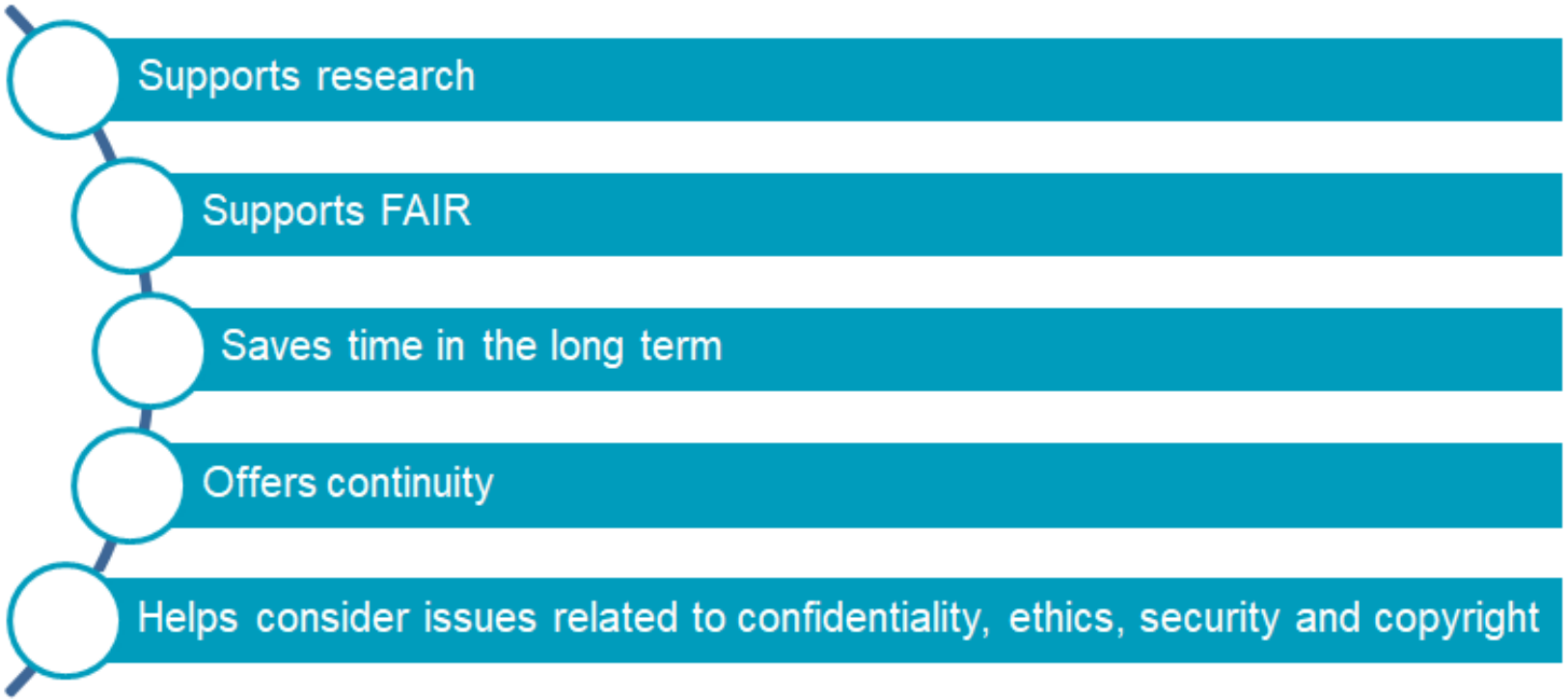
A DMP is a formal and at the same time a living document

- **WHAT** data goes into a project (reuse) and comes out of it (potential reuse)?
- **HOW** does the team take care of the data?
- **WHO** is allowed to do **WHAT** with the data **WHEN**?



(source: https://www.openaire.eu/blogs/why-is-this-a-good-data-management-plan)

Benefits of a DMP



Requirements of Research Funders – Example: DFG

Checklist

A checklist must be submitted as part of the proposal

1. Data description
2. Documentation and data quality
3. Storage and technical archiving the project
4. Legal obligations and conditions
5. Data exchange and long-term accessibility
6. Responsibility and resources

Released December 2021

Link to DFG checklist: de/en

Planning research

Good Research Practice and Funders

Funding organisation guidelines

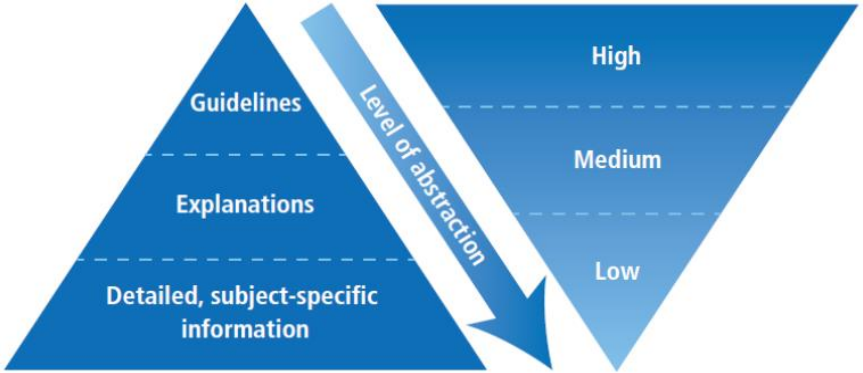
There are different formal requirements depending on the individual research funding organisation! Before you apply, find out exactly what is relevant in your case.

Overview:

https://www.forschungsdaten.info/themen/informieren-und-planen/foerderrichtlinien/#c492536

DFG Guidelines for Safeguarding Good Research Practice (2019)

- Fundamental revision of the recommendations from 1998
- Transition extended until 31st July 2023
- Modifications:
  - Recommendations
  - Multidimensional approach
  - Codex with 19 guidelines
  - 11 guidelines on the research process
  - RDM is relevant in 8 of these 11 guidelines



All research institutes must implement these guidelines in a legally-binding manner in order to be eligible to receive DFG funding.

Guidelines for Safeguarding Good Research Practice. Code of Conduct: https://doi.org/10.5281/zenodo.3923602

General GRP-Guidelines (DFG)

- Guideline 7: Cross-phase quality assurance
- Guideline 10: Legal and ethical frameworks, usage rights
- Guideline 11: Methods and standards
- Guideline 12: Documentation
- Guideline 13: Providing public access to research results
- Guideline 14: Authorship
- Guideline 15: Publication medium
- Guideline 17: Archiving

Example: Guideline 12 Documentation

- Documentation of **all information** relevant to the production of a research result (in accordance with existing recommendations and guidelines)
- **Selection of results must be avoided!**
- **Documentation** and **research results must not be manipulated**; they are protected as effectively as possible against manipulation.

Relevance in terms of research data:

- Documentation of necessary information to understand the research (results)
- Information on research data used or generated, the methods, evaluation and analysis steps, the development of the hypothesis and citations
- Possibility of replication
- Documentation of the source code in the development of research software

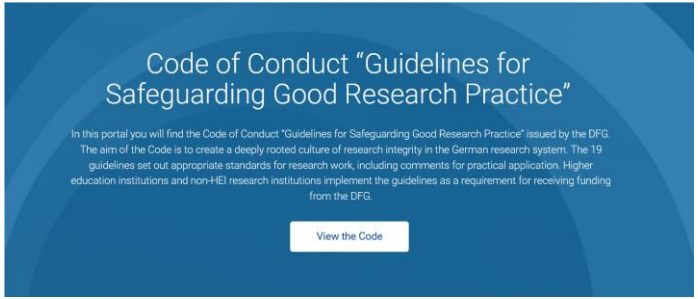
Guidelines for Safeguarding Good Research Practice. Code of Conduct: https://doi.org/10.5281/zenodo.3923602

Research Integrity in Chemistry

Portal: **https://wissenschaftliche-integritaet.de/en**

for subject-specific information

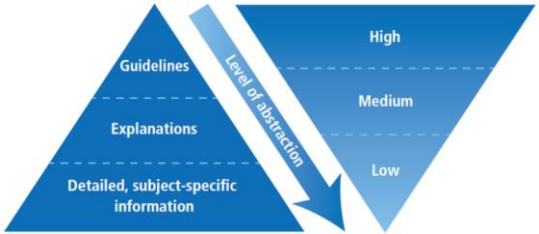
- Articles available in German and English



Exemplary results for chemistry:

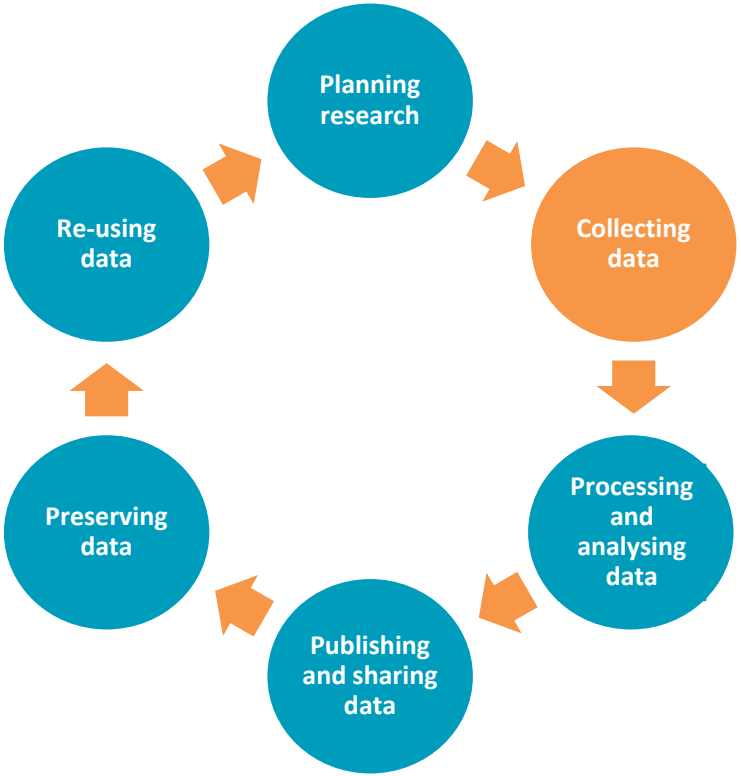
- Documentation of research results in experimental chemistry
- Quality assurance in experimental chemistry
- Use of chemistry-specific repositories
- Handling research software - Case studies
- Further links to performance dimensions and evaluation criteria
- Author order in physics and chemistry
- Electronic laboratory journal and repository in chemistry
- Ethical principles in chemistry

Research Integrity



https://wissenschaftliche-integritaet.de/

Collecting data



- Perform experiments, measurements, simulations, observations...
- Collect and create **metadata**
- **Document und describe data**
- Enter, digitize, transcribe and translate data
- Check, validate and clean data
- **Save and manage data**

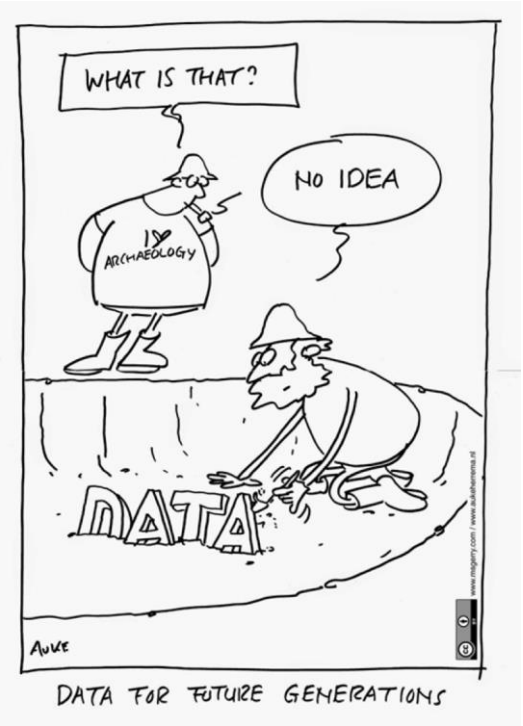
Collecting data

Metadata

Metadata

What is metadata and why is it important?

- Data that describes data
- Makes datasets searchable (and findable)
- Makes datasets understandable and FAIR
- Machine and human-readable
- Standardization is ongoing



Source: https://www.openaire.eu/rdm-handbook

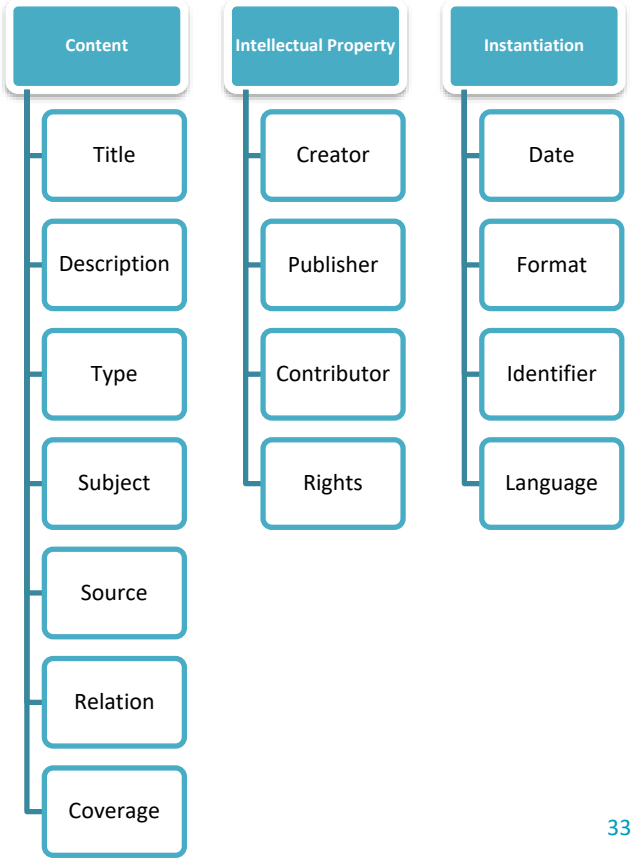
Metadata Schema

- A Metadata scheme determines and structures the metadata elements.
- Pairs of **metadata element - value**
- Defines input type
- Sets restrictions, such as the use of (controlled) vocabulary or required fields

Dublin Core

International Data Exchange format

- 22 elements –
  15 with an ISO certificate
- Refinements and encoding schemes for subject specific applications
  - http://www.dublincore.org/
  - http://wiki.dublincore.org/index.php/User_Guide



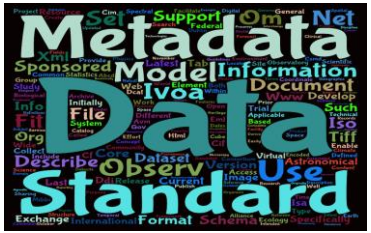
Data Cite

Defines core metadata for research data and it is community driven

DataCite term	SRD 78 (Atomic Spec)	comments
Identifier		lacking - but required - looking for a DOI or
Creator	Alexander Kramida, Yuri Ralchenko, and Joseph Reader, Edward B. Saloman	took the 'active' names from website - can also add affiliation and identifier for each name - DataCite allows more than one and recognizes ordering, DC wants "an entity"
Title	NIST Atomic Spectra Database - SRD 78	took name and appended SRD #, like EDI
Publisher	Atomic Spectroscopy Group, Physical Measurment Laboratory, National Institute of Standards and Technology	If it needs to be reduced - could take group and lab away and leave NIST
PublicationYear	2015	
Subject	Atomic spectra, atomic ground state, atomic ionization energy, atomic transition probability, atomic energy levels	EDI entry lists all element names - might be overkill for this purpose
Contributor		Could list "past contributors" or "students contributing to data entry" - how far to go?
Date		Could give more specific date here if that makes sense
Language	en	
ResourceType	Dataset/Atomic Spectra	"Dataset" comes from list, "Atomic Spectra" is free text
AlternateIdentifier	http://www.nist.gov/pml/data/asd.cfm	Should add property alternateIdentifierType with value URI?
GeoLocation		
RelatedIdentifier	http://www.nist.gov/pml/pubs/atspec/index.cfm, http://www.nist.gov/pml/data/asd_contents.cfm, http://physics.nist.gov/PhysRefData/ASD/Html/help.html, http://www.nist.gov/pml/data/asbib/index.cfm	These are resources listed on the homepage (Intro to atomic spectroscopy, intro to ASD contents, help, bibliography)
Size		
Format		
Version	5	
Rights	http://www.nist.gov/data/license.cfm	
Description	This database provides access and search capability for NIST critically evaluated data on atomic energy levels, wavelengths, and transition probabilities that are reasonably up-to-date. The Atomic Spectroscopy Data Center has carried out these critical compilations. The Data Center is located in the Physical Measurement Laboratory at the National Institute of Standards and Technology (NIST).	

Source: https://datacite.org/ & https://bit.ly/3f3DZgR

Find standards, controlled vocabulary, …



- FAIRSharing.org
  https://fairsharing.org/
- NFDI4Chem – Knowledge Base
  https://knowledgebase.nfdi4chem.de/knowledge_base/docs/topics/format_standards/
- Metadata Directory of the Research Data Alliance
  http://rd-alliance.github.io/metadata-directory/

Collecting data

Electronic lab notebook (ELN)

Electronic lab notebook

Do you use an electronic lab notebook (ELN)?

a) Yes, I use an ELN

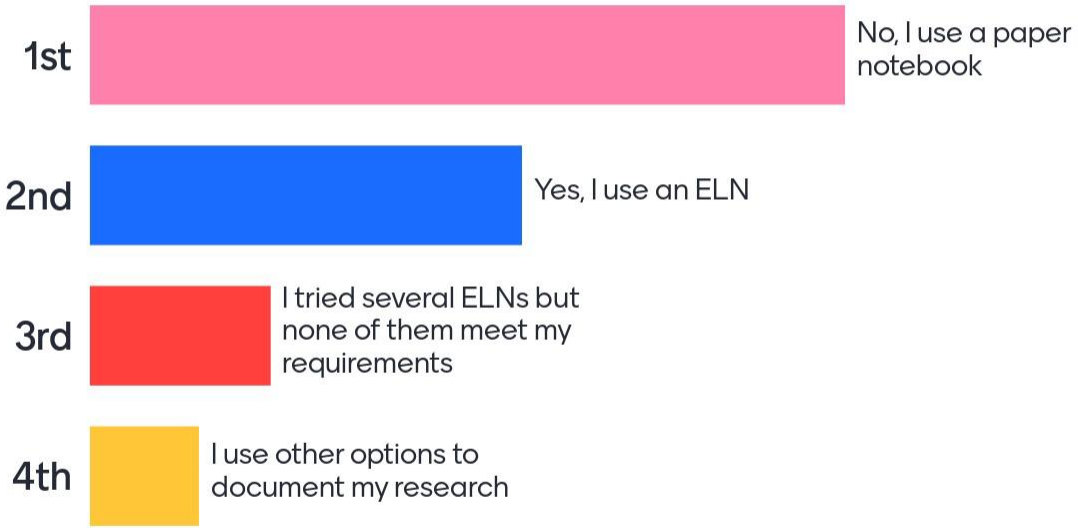
b) No, I use a paper notebook

c) I tried several ELNs but none of them meet my requirements

d) I use other options to document my research

Do you use an electronic lab notebook (ELN)?

Mentimeter



What is an ELN?

Simple System: Blank Sheet

- Enter text
- Add notes
- Add files as attachments (e.g. images, tables)
- Sharing
- Searching

e.g. Evernote, GoogleDrive, Dropbox, MS Sharepoint

Electronic Lab Notebook (ELN)

+ Structured metadata in human and machine-readable formats
+ Discipline-specific functions/ editors
+ Rights management
+ Audit trail
+ API (Application Programming Interface)

e.g. Labfolder, RSpace, eLabFTW, Labguru

Laboratory Information Management System (LIMS)

+ Sample management
+ Instrument integration
+ Electronic signatures
+ Reporting or statistics modules

e.g. Benchling, Starlims, Limesophy

Source: Lindstädt, Adam, MB MED ELN-Wegweiser: Elektronsiche Laborbücher im Kontext von FDM und GWP, RWTH Aachen FDM netzwerktreffen 10.11.2021

Advantages of an ELN

Avoid Data Loss

- Linking experimental descriptions to collected data (analog and digital)
- Secure data storage, backups

Standardised Documentation

- Structured and standardised collection of metadata
- Generation of interoperable (meta)data

Knowledge Management

- Data are findable
- Data are accessible
- Data are available, even after change of personnel!

Publication

- Data provision for publication of research results
- Simple transfer of data to repositories

A plethora of available ELNs

labfolder

sciformation

eLabFTW

RSpace

PerkinElmer
For the Better

BioELN

SignalsNotebook

openBIS

Chemotion

SciNote

… and many more

epam
Indigo ELN

How to introduce an ELN

Needs assessment:

- Analysing current situation (budget, IT resources, software environment)
- Definition of important features
- ELN concept (generic, discipline-specific)
- Drawing on experiences of other research institutions

Testing the selected products:

- Demo versions or free trial access for individual users
- Testing no more than 2 – 3 ELNs
- In-depth testing using real-life use cases from the lab

Introducing the chosen ELN:

- Run training courses, training material
- Designate contact persons from the test team
- Continuous mentoring

Source: ZB MED ELN-Guide, 2021, https://dx.doi.org/10.4126/FRL01-006425772

How to introduce an ELN

Needs assessment:

- Analysing current situation (budget, IT resources, software environment)
- Definition of important features
- ELN concept (generic, discipline-specific)
- Drawing on experience

Testi

- De
- Tes
- In-

Introduc **ELN:**

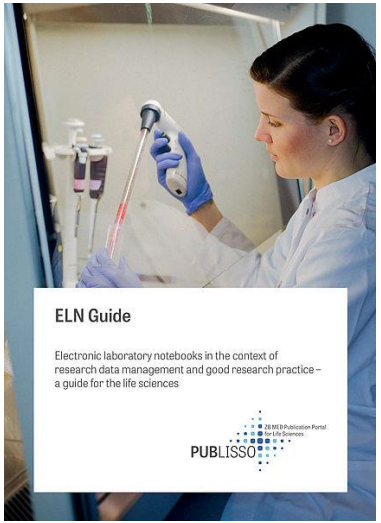
- Run training courses, training material
- Designate contact persons from the test team
- Continuous mentoring

Motivate by demonstrating the benefits!

Source: ZB MED ELN-Guide, 2021, https://dx.doi.org/10.4126/FRL01-006425772

Further reading

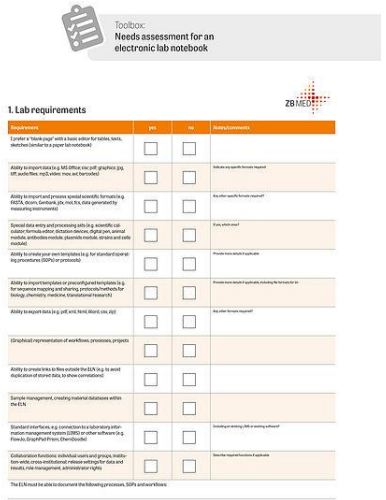
ZB MED Documenting research data: Electronic Lab(oratory) Notebooks



ELN Guide with best practices

Product	Last update	Values from provider query	Values from external sources	Cloud usage (Software as a Service)	Local use (On-premises)	Audit Trail	FDA CFR 21 part 11	Electronic signature	Witnessing functions	Electronic time stamps	Versioning	pdf	html	jpeg	MS Office	Data Base	csv/tsv	json	xml
				/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Arxlab	2019-12	Y	N	Y	N	Y	Y	Y	Y	Y	Y	N	N	N	N	Y	N	Y	N
Benchling	2018-11	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
Biovia	2019-12	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y
eLABJournal	2019-12	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
eLabFTW	2019-11	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	Y	Y	N
IDBS	2018-11	N	Y	Y	N	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N
LabArchives	2019-12	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
Labcollector	2019-12	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	Y
Labfolder	2019-11	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N
Labguru	2018-11	N	Y	Y	Y	Y	N	Y	N	Y	N	Y	N	N	N	N	N	N	Y
LabWare	2019-11	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Limsophy	2019-12	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
NuGenesis	2019-12	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
openBIS	2019-11	Y	N	N	Y	Y	N	N	Y	Y	N	N	Y	N	Y	N	Y	N	Y
RSpace	2019-11	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	N	Y
Kategorie				**Providing model**		**Security of evidence**						**File-Export**							

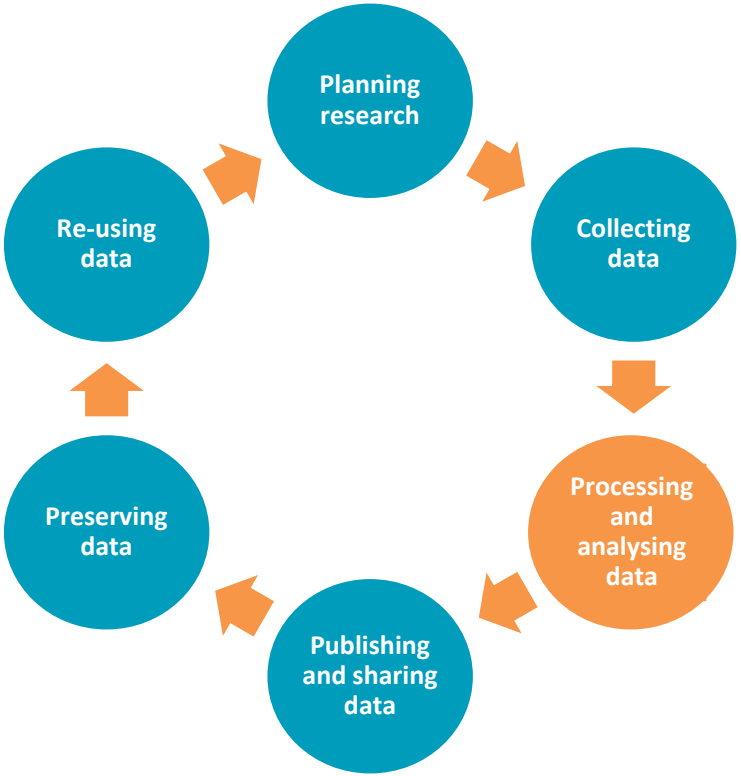
PUBLISSO ELN filter



Toolbox Needs Assessment

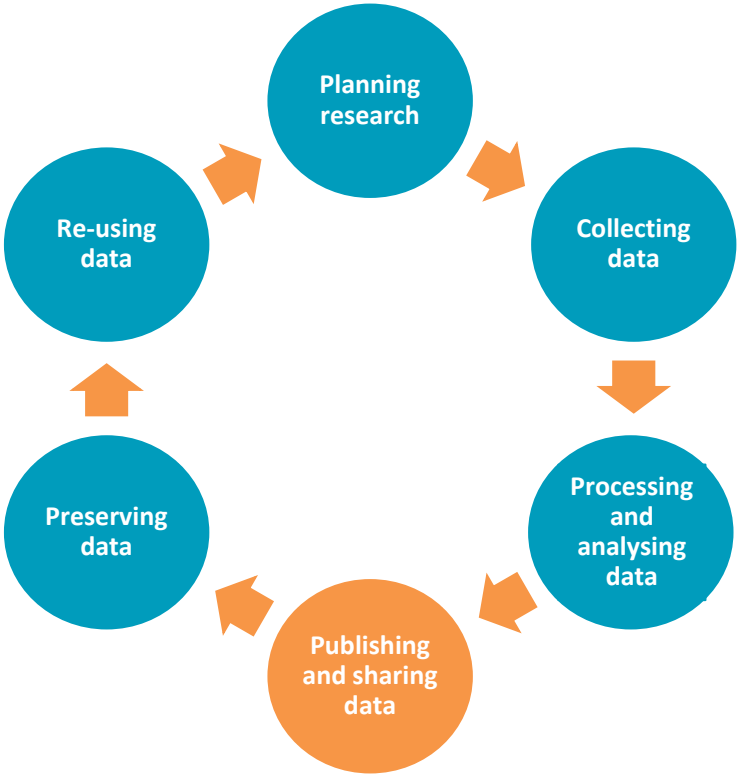
Or check out the eln-finder: https://eln-finder.ulb.tu-darmstadt.de/home

Processing and analyzing data



- Interpret data
- Use the data in scientific publications
- Backing-up data and preparing it for data storage
- Enabling data exchange during the project

Publishing and sharing data



- Share, disseminate, publish data
- Making data known and findable (catalogues)
- Making data citable (DOI)
- Issue licences
- Access control (if necessary)

Reasons for Data Publication

- Advantages
  - Better (re-)use of your research data
  - Increased visibility, openness, transparency and accountability

    "As open as possible, as closed as necessary"
  - Clearly citable with persistent identifiers (e.g. DOI)
  - Long-term availability
  - Data production as an independent scientific result (e.g. Data Citation Index)
- Requirements
  - Of funders
  - Of scientific journals

Data Publication

How do you publish your data?

a) Research article + supplement

b) Article in data journal

c) Repository

d) Website (personal or institutional)

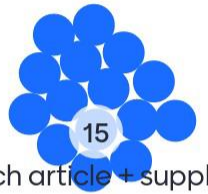
e) I only ever publish data if specifically required by journal or reviewers

f) I never publish data

g) Other options

How do you publish your data?

Mentimeter



Research article + supplement

0

Article in data journal

Repository

Website (personal or institutional)

0

If required by journal/ reviewers

0

I never publish data

0

Other options

16

How can I publish my data?

Supplement of an article in scientific journal

Data Journal

GIGAScience Data

(GIGA)n SCIENCE

RMetS Geoscience Data Journal

Nuclear Data Sheets

DATA SCIENCE Journal

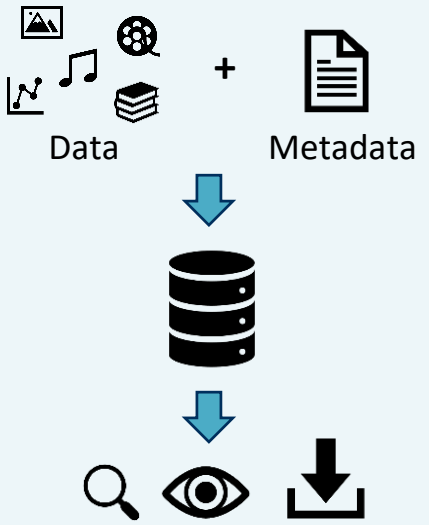
Biodiversity Data Journal

Journal of chemical & engineering data

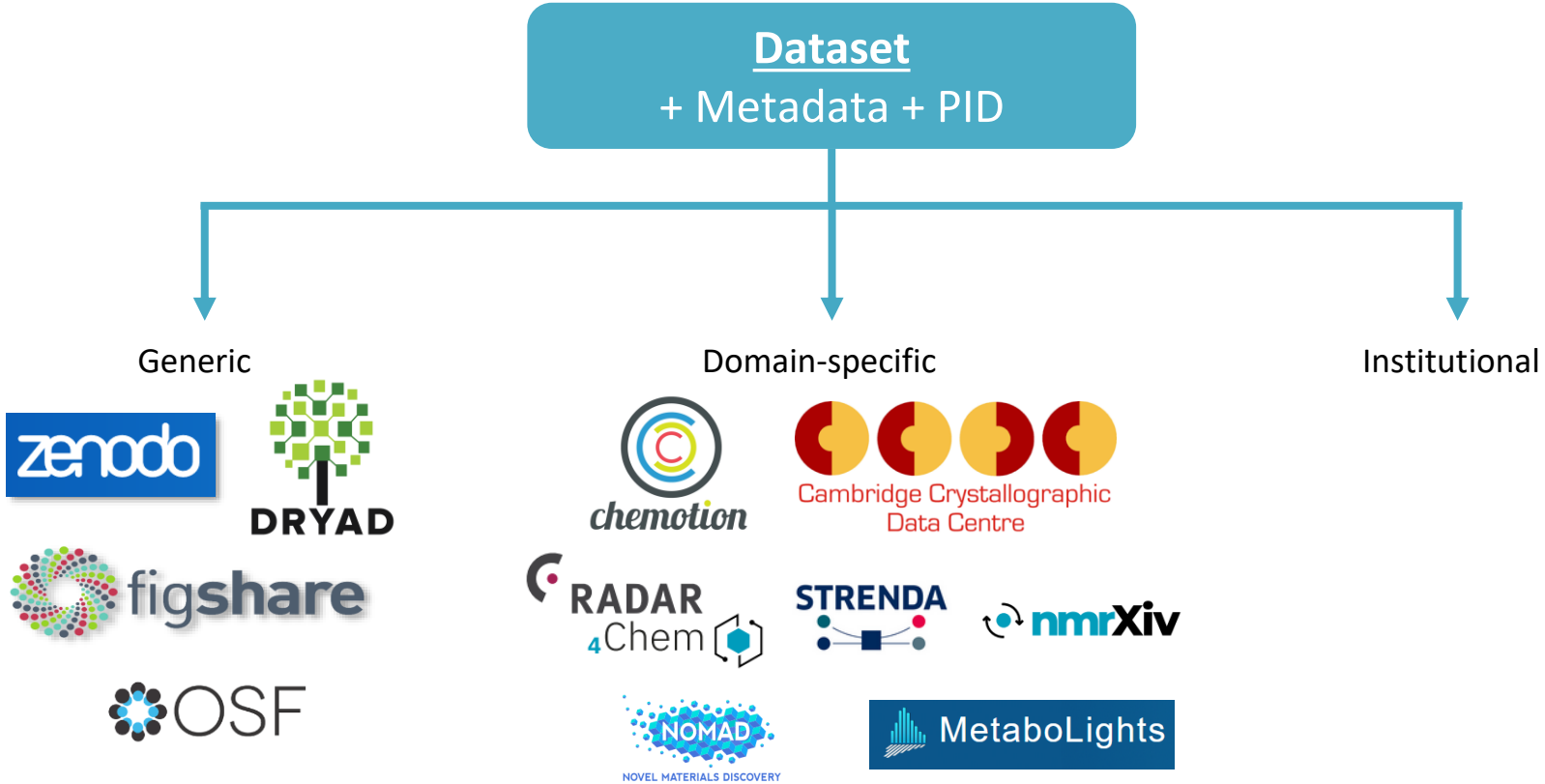
WILEY ONLINE LIBRARY

SCIENTIFIC DATA

Data Repository



Data Repositories



Further reading: Publishing

- Persistent identifiers (PIDs)

  https://knowledgebase.nfdi4chem.de/knowledge_base/docs/topics/pid/

- Data availability statements

  https://knowledgebase.nfdi4chem.de/knowledge_base/docs/topics/data_availability_statement/

- Best Practice

  https://knowledgebase.nfdi4chem.de/knowledge_base/docs/topics/best_practice/

Further reading: Repositories

- Repositories

  https://knowledgebase.nfdi4chem.de/knowledge_base/docs/topics/repositories/

- How to choose the right repository

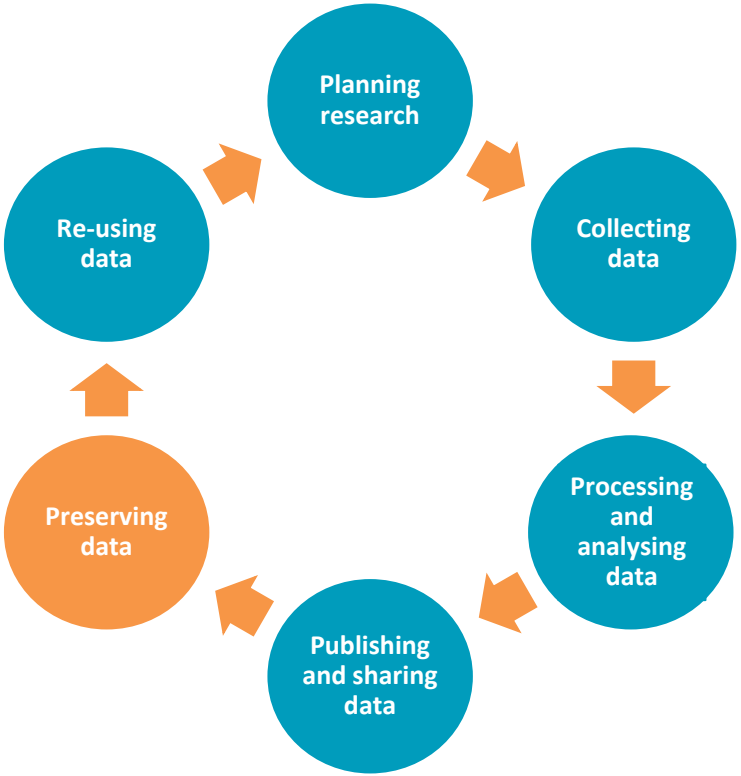
  https://knowledgebase.nfdi4chem.de/knowledge_base/docs/data/choose_repository/

- Repository choosing tools

  re3data.org

  fairsharing.org

Preserving data



Preserving ≠ Not deleting

Aim: Keep data, metadata and documentation safe, available and re-usable in the long term

Risks:

- Corruption of data or storage medium
- Outdated file formats
- Lack of metadata and/or documentation
- Data not findable or accessible

Different types of data preservation

- Access for the data producer
- **Active data**
- Data kept short-term
- Purpose: Protection and discovery

Backup

- Access for the data producer
- **Final Data**
- Data kept long-term
- Purpose: Preservation of information

Archiving

- Open access
- **Final Data**
- Data persistence depends on the publishing institution
- Purpose: Provision of data for reuse

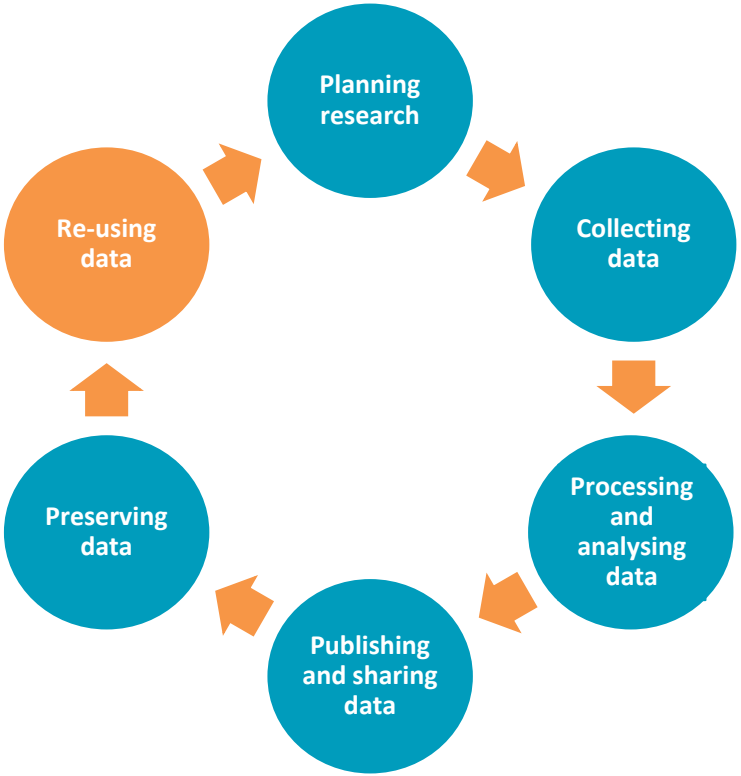
Publication

Steps of data preservation

- What to keep? – Data selection
  - Define selection criteria
- Where to preserve data? - Suitable location or medium
  - External data repository or archive versus institutional infrastructure
- Prepare data and files for the preservation
  - Organized files and suitable file formats
  - Including metadata, documentation, access rights and conditions
- Perform periodic checks of the preserved data

https://www.youtube.com/watch?v=UaiRAI-fwmw

Re-using data

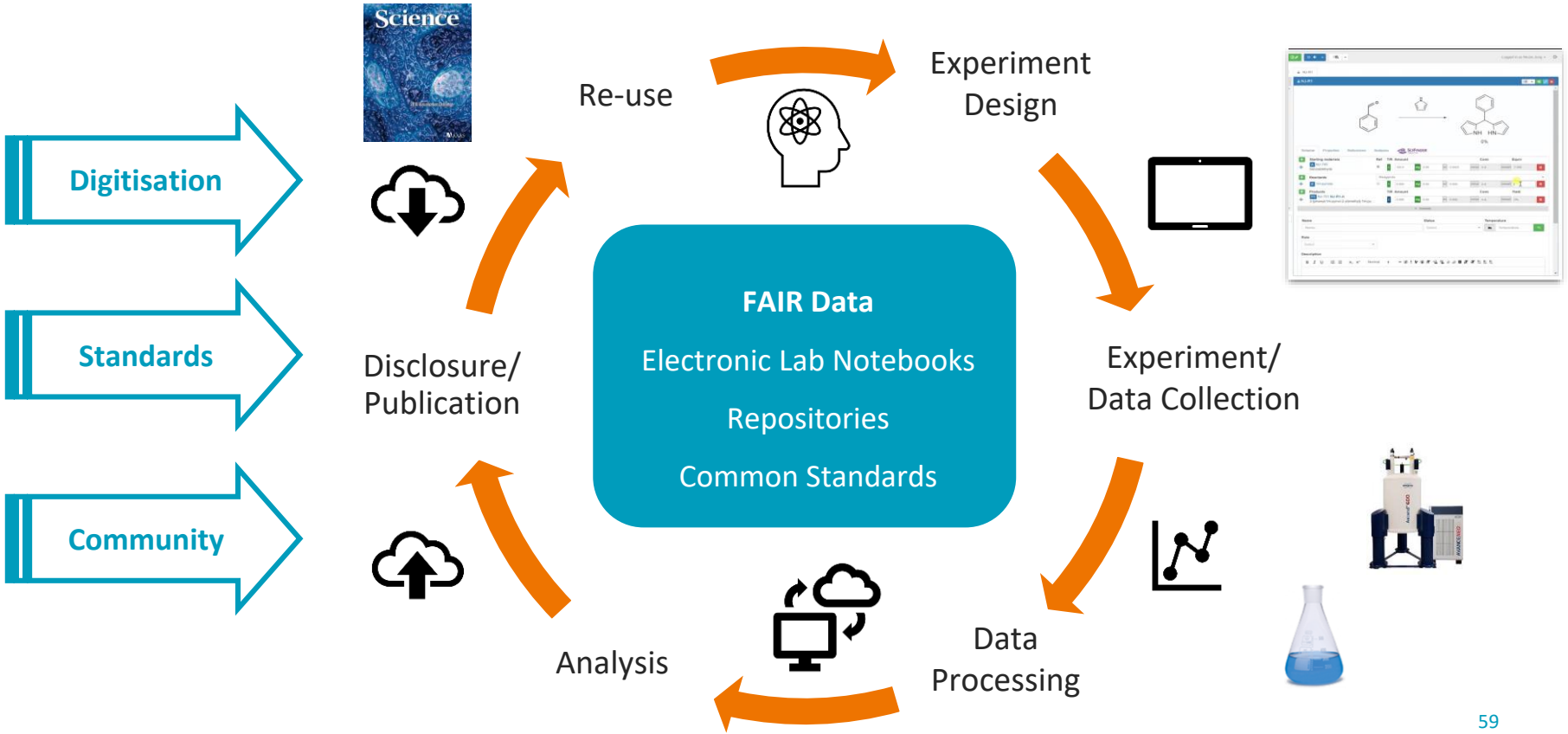


What is the potential of your research data?

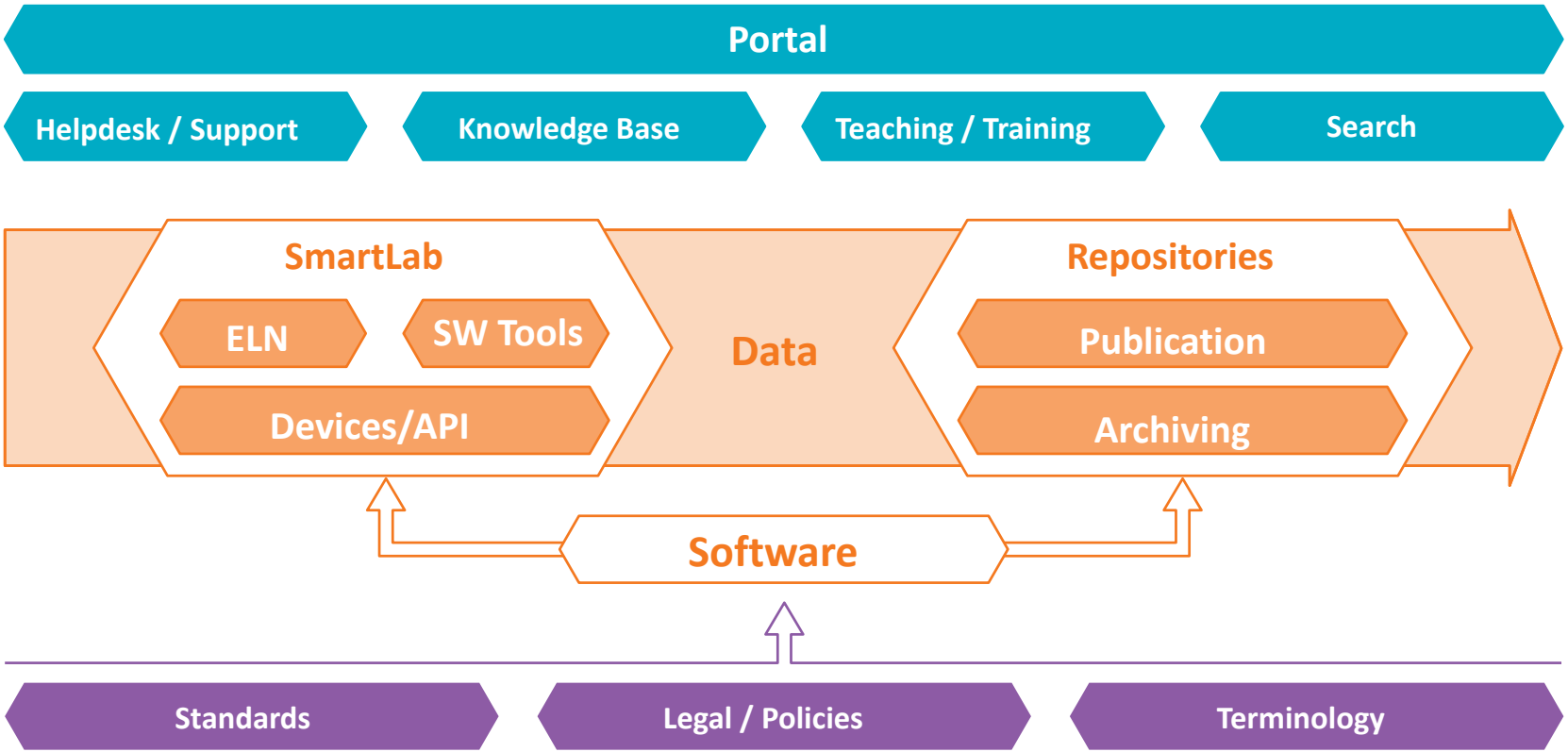
- Conduct further research with the data
- Put data into new contexts, using data in an interdisciplinary way
- Big Data applications
- Review, critique and discuss research findings
- Teaching and learning
- Citing research data

NFDI4Chem

NFDI4Chem: Vision



NFDI4Chem: Strategy



Services

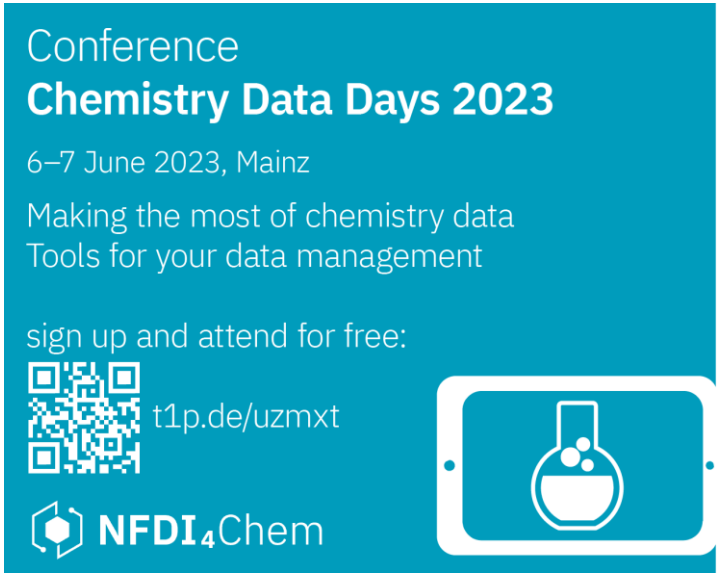
- Helpdesk
  helpdesk@nfdi4chem.de

- Knowledge Base for RDM in chemistry
  https://knowledgebase.nfdi4chem.de/knowledge_base/

- Events such as workshops, Q&A, Stammtisch
  https://www.nfdi4chem.de/index.php/events/

Chemistry Data Days 2023

- 6th and 7th of June in Mainz
- Free (excluding travel & accommodation)
- Learn about tools and resources for data management in chemistry (E.g. ELNs)
- Hear from invited speakers how they have made the most of their data
- In cooperation with GDCh and JCF who will be hosting sessions on career entry for young chemists



In cooperation with:

GDCh Gesellschaft Deutscher Chemiker
Partner of NFDI4Chem
GDCh JCF

Register here

Thank you for your attention!

Question & Answer Session

NFDI4Chem
ENHANCE
YOUR
DATA.