

Breakfast with Data Stewards

Marie Šafner Matúš Drexler



The profile and role of group data stewards: YOU

- Awareness about data management procedures in your group
- Point-of-contact for group members struggling with DM (formats, metadata, software, DMP, storage, backup, retention, sharing)
- If guidance needed, the data steward will get in touch with us



The profile and role of group data stewards: YOU

- Awareness about data management procedures in your group
- Point-of-contact for group members struggling with DM (formats, metadata, software, DMP, storage, backup, retention, sharing)
- If guidance needed, the data steward will get in touch with us



Replication Crisis! -> Open Science

News Feature | Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

Monya Baker

Nature 533, 452–454 (2016) Cite this article

A questionnaire on various aspects of reproducibility.

224k Accesses 2416 Citations 5169 Altmetric Metrics









Is open science going to prevent error/fraud in research?

No, but it has the potential to lower it and give researchers building on an existing dataset opportunity to save time attempting to replicate the results! (if irreproducible)

Replication Crisis! -> Open Science

News Feature Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

Monya Baker

Nature 533, 452–454 (2016) Cite this article

A questionnaire on various aspects of reproducibility.

224k Accesses | 2416 Citations | 5169 Altmetric | Metrics





Is open science going to prevent error/fraud in research?

No, but it has the potential to lower it and give researchers building on an existing dataset opportunity to save time attempting to replicate the results! (if irreproducible)

Open Data x FAIR Data

Open data are not necessarily FAIR and vice versa – "AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY"

Funding agencies require your data to be open -> Act No. 130/2002 Coll. on the Support of R&D from public funds

What does it mean?

• You must be able to provide all your publicly funded project data on request.

Is it FAIR?

• No. For the data to be FAIR, you must deposit your data in a repository.

Can I deposit my data without opening it?

• Yes.



What specifically makes the data FAIR?

Findable

Machine-readable rich indexed metadata, PIDs, searchable resource

Accessible

License, CC0 metadata, software and version used

Interoperable

Rich metadata, defined or standard vocabularies, ontologies, data capture methodology

Reusable

License, rich metadata, provenance, links to other datasets, papers, source code, data quality measures, open, long-term viable formats





Think about your group's data workflow

Would you describe your workflow as FAIR?

• FINDABILITY

Can you find data from your group's project (published/unpublished) 2 years ago?

• INTERPRETABILITY

Do you understand your or someone else's data a year ago?

• INTEROPERABILITY

Can you share your 1-year-old data to a colleague without lack of comprehension?

ACCESSIBLE

Are your group's 8-years old data openable and legible without losing its original format?

FAIR Data Self-Assessment Tool

Australian Research Data Commons

Use this handy, researcher-friendly tool to discover how <u>findable, accessible, interoperable and reusable (FAIR)</u> your research dataset is and get practical tips on how to enhance its FAIRness.

https://ardc.edu.au/resource/fair-data-self-assessmenttool/



Considerations throughout the data life cycle



The ultimate Group Data Management Plan

- The file-data organization system
- Storage-backup-retention system
- Readme project/study/experiment
- Metadata system
- ELN synchronized use in the group (annotations, grouping)
- Data cleaning, curation
- Use graphs to capture the essence and ease the understanding
- Could it help? ____







PhD Bootcamp

- The freshers need to be stressed the value of good data management
- Workshop on DM as part of their induction process
- Every group should have an established system by September to test efficiency of the workshop
- Content will be shared will you

What is the biggest trouble with the students?

 An idea? → Send me an email with your suggestion(s) on what the students should be introduced to.



Data Repositories

Data repositories – which one to choose?

1 st

Domain-specific

Rich discipline-specific metadata (often a developed own standard)

Support, navigate, or help to shape the best practice in the field

Data automation, visualization, data quality check, peer review



OpenAIRE



🔁 Crossref

Commons

Shared features to look out for in all the domain-specific, institutional, and general repos DOIs, versioning, licenses, rich metadata (availability after dataset no longer available), long-term preservation/backup policy, data storage location, searchable

NATIONAL METADATA D

Dataset

Search

Preprint repos

isc

ChemRxiv

bioRχiv

Repository searches – domain-specificity

registry of research data repositories

Data submission

EMBL-EBI home > Services > Data submission

Use this data submission wizard to find the right archive for your data in a few simple steps.

🕂 EMBL-EBI home

Services

🕅 Research

🎄 Train

https://www.ebi.ac.uk/submission/

1 What type	of data do you have?	
DNA/RNA sequ	uence Expression data Protein data Structures Systems	
Chemical biolo	ogy Ontologies Images Multi-omics or other cross-domain study	/
Other biologica	al research data	

Why submit data to an archive?

- Submission of primary data and derived information to public data repositories is an essential step in the scientific process.
- Through submission, the scientific community is fed the raw materials for the building and maintenance of the complete and up-to-date data sets that support searches and analysis on
- the latest sequences, structures and molecular profiles of living systems.

The CAS institutional repo

- Paper and <u>data repository</u>
- GLs, Pls, data stewards, curators
- 50GB per dataset deposition (may be more zipped up files each 50GB How to amounting altogether up to 200GB), unlimited depositions (if you'd need more one-off increase of up to 200GB can be negotiated) deposit my
 - No limitation to data formats, open formats advised

Raw data (proprietary + open) + secondary data + metadata (=cleared detailed information)

ASEP

REPOSITORY

- Correct license needs to be selected as that is unchangeable (CC BY 4.0) do not violate funder requirements!
- ORCID, DOI, handle
- Links with papers

data?

- Soon: pre-prints, official policy, NMA, Google Dataset Search, applying for CoreTrustSeal certification
- General metadata curator the ASEP administrator Jana Prochazkova please preregister (link in the email)

Synthetic chemistry repo



- Synthetic reactions and compounds with supportive data from spectroscopic, chromatographic techniques, mass spectroscopy, microscopy, electrochemistry
- FAIR repository, part of the NFDI4Chem (NMRXiv)
- ORCID sign in
- Interactive, user friendly, automated interface for data entry (ChemDraw, CAS, clear directions as to what should be entered) ChemConverter - Converts chemical data into standardized formats (SMILES, InChI, MOL) to ensure compatibility with other databases
- More detailed search (incl. reaction type in the user profile interface) -> is expected to be
 included in the forefront of the repo in 2nd development phase
- Workflow: Properties -> Analyses (chemical methods ontology) -> QC & Curation (1H NMR, 13C NMR, MS, IR) -> References -> Results
- Quality check
- Peer review
- Workshop (2-stage workshop)
- Downside: 10-12 years guarantee, long-term preservation strategy cannot be formed at the moment, but exit strategy in place (the data would be accessible on a Radar with DOI for 20-30 years), but efforts to change the way NFDI is lead to a community effort, also down to interest in the repos)



- Plans
- DSW instance soon at IOCB (maybe within a week)
- ASEP data repository information on intraweb
- Storage guide
- Chemotion training



Menti: 2254 4877

ELN testing





•IOCB continues discusions with CDD Vault and eLabNext

•Both platforms align with the institute's needs and offer relevant features

•Further assessment may be required for full implementation

Report z testovania elektronických laboratórnych dennikov pre ÚOCHB AV ČR

1. Úvod

Elektronizácia vedeckých dát je dnes nevyhnutným krokom pre moderný výskum. Dodržiavanie FAIR princípov (Findable, Accessible, Interoperable, Reusable) pomáha lepšie organizovať a spravovať experimentálne dáta, čo znamená jednoduchšie vyhľadávanie, vyššiu reprodukovateľnosť výsledkov a efektívnejšiu spoluprácu medzi členmi tímov či





Pros	Cons	Pros	Cons
Simple and intuitive	Basic reaction scheme	Molecule-based data management	Complex solution
Mnova visualization	Minimal support for additional data	API integration (export/import)	Mnova visualization
	Limited filter options	Bulk data import	
	Limited export options	Robust and versatile protocol system	
	No bulk import	Empirical calculation	
	Very slow development	Suitable for collaboration	
	No integration options	MS Office, PDF, TXT files integration	

Request system

Time plan:

- April 2025 IOCB board progress presentation
- May 2025 IOCB Core facilities progress presentation
- ~Jun 2025 IOCB Presentation
- 1st September 2025 Request system released

IOCB PRAGUE EQUEST MANAGEMENT	Se	ervio	es					
Requests			Id T	7 Name	7	Grou	ip 了	Cate
Services				Eleme	ntarv	Analy	/tical	Com
Projects		>	1	analys	sis	labor	atories	cher anal
Inventory						A	dia al	
Files		~	2	Weigh	ing	labo	ratories	Sup
Reports								
ADMINISTRATION User Accounts Organization Structure		The sta For lab	e servic rting fro reques oratory	e enables om 0.1 mg, ts requirin staff is ne	precise with an g smalle cessary.	weighing accurac r amoun	∣ of subst y of 1 or (ts, consu	ances, t).5 micro Itation w
Settings		An	nounce	ments				
		20.	11. 202	4 Stanis	slava Ma	tějková	In Dec	ember, p
		>	3	Small molec analys	ule is	Mass Spec	trometry	Mas
		~~	<	12	3 4	4 5	> >	» [10
DJ Help								

Join at menti.com | use code 2254 4877

材 Mentimeter

What do YOU want to see delivered/in progress the most?

5 responses

	asep	ELN	- The polici of lab notebooks and electronic lab notebooks. Some
	Have ELN as soon as possible Have data storage from IOCB for daily data	- How to handel paper notebookes and notes to analysis - what are the data	scientist would never turn out from paper Data resposibility of core facilities
storage and backup	storage and backup	resposibility of corr facilities	



