Active and Sustainable Research Data Management in project consortia and research groups – The RETAIN Fellowship Program

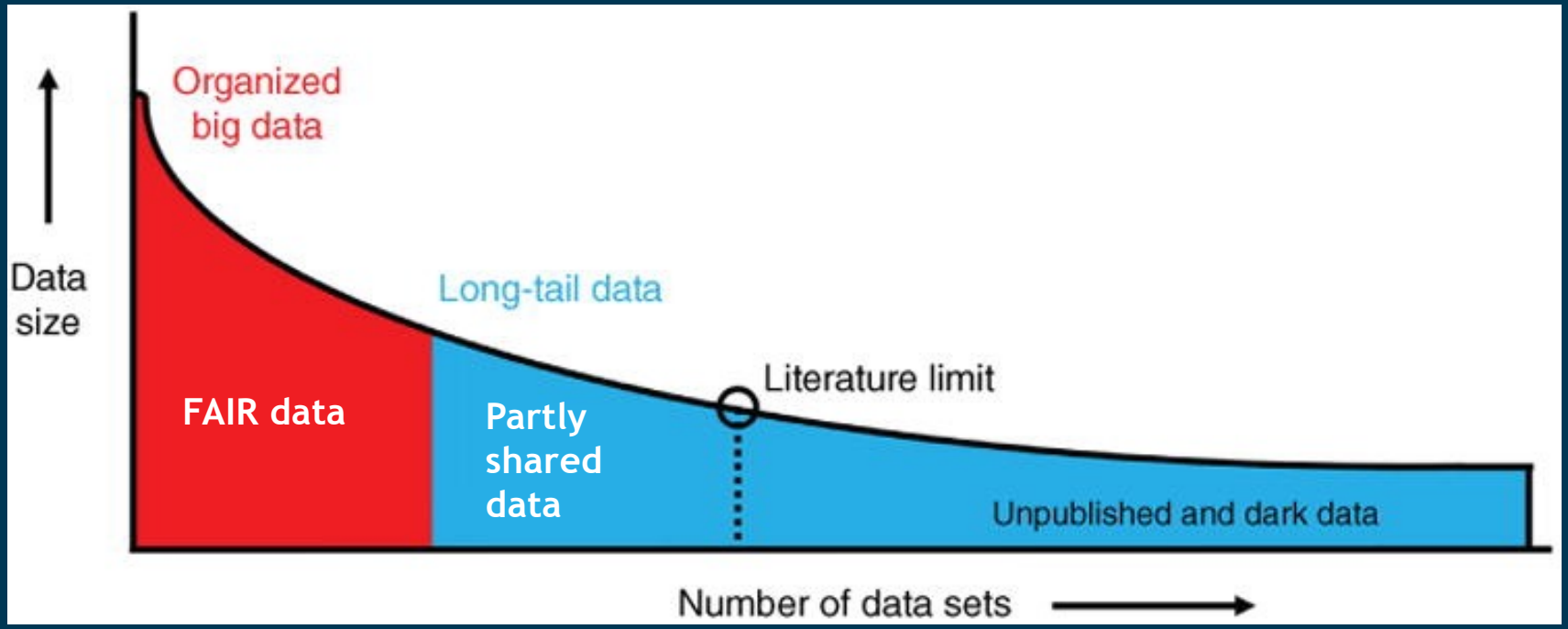
René Bernard, PhD
Cluster of Excellence NeuroCure
BIH QUEST Center for Responsible Research
rene.bernard@charite.de

BIH QUEST Center for Responsible Research

NeuroCure Cluster of Excellence

BIH Berlin Institute of Health @Charité

Datafication in Biomedicine (applicable to Neuroscience)



Ferguson et al., 2014, *Nature Neuroscience*, MacLeod et al., 2014, *Lancet*

Fact: Data are not shared

Data availability

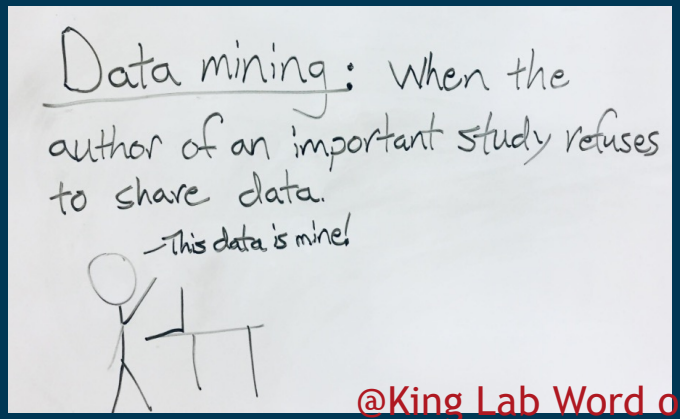
The data that support the findings in this study are available from the corresponding author upon reasonable request. DNA constructs are available via Addgene at

Data availability

Data can be made available upon request.

Data Availability Statement

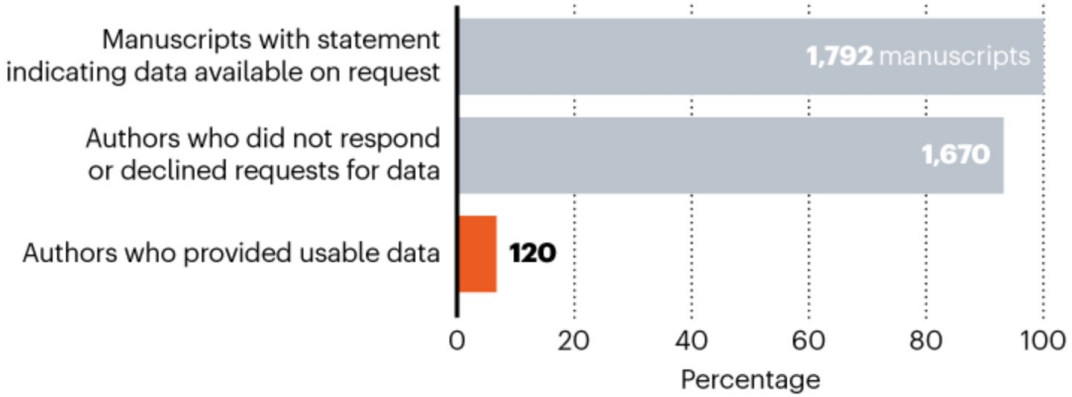
All data are contained within the article.



@King Lab Word of the Day

DATA-SHARING BEHAVIOUR

Of almost 1,800 manuscripts for which the authors stated they were willing to share their data, more than 90% of corresponding authors either declined or did not respond to requests for data. Only about 7% of authors actually handed over data.



©nature

Source: Livia Puljak et al

https://www.nature.com/articles/d41586-022-01692-1

Why is that ? – What good science meant historically

- **Publications: report summarized results only**
- **Methods: narrative and selective**
- **Data analysis flows: in lab notebooks, personal files, saved and archived locally**

→ Reward mechanisms were designed accordingly

- **The redefinition of countable scientific output: paradigm shift (still met with some resistance)**
- **RDM asks from researchers to:**
  - Externalize hidden knowledge
  - formalize decisions that were never documented
  - anticipate future users (even in their work environment) they will never meet

Complexity in the Neurosciences - Subfields & Subdisciplines

- Molecular
- Developmental
- Clinical
- Cellular
- Sensory
- Systems
- Social
- Evolutionary
- Computational
- Behavioral

Many interdisciplinary approaches and techniques in combination

Many data have a spatial and temporal component

- Neurogenetics
- Neuropharmacology
- Neurophysiology
- Neuroanatomy
- Neuroendocrinology
- Neuroinformatics
- Neuropsychology

This complexity is a big hurdle in effective RDM

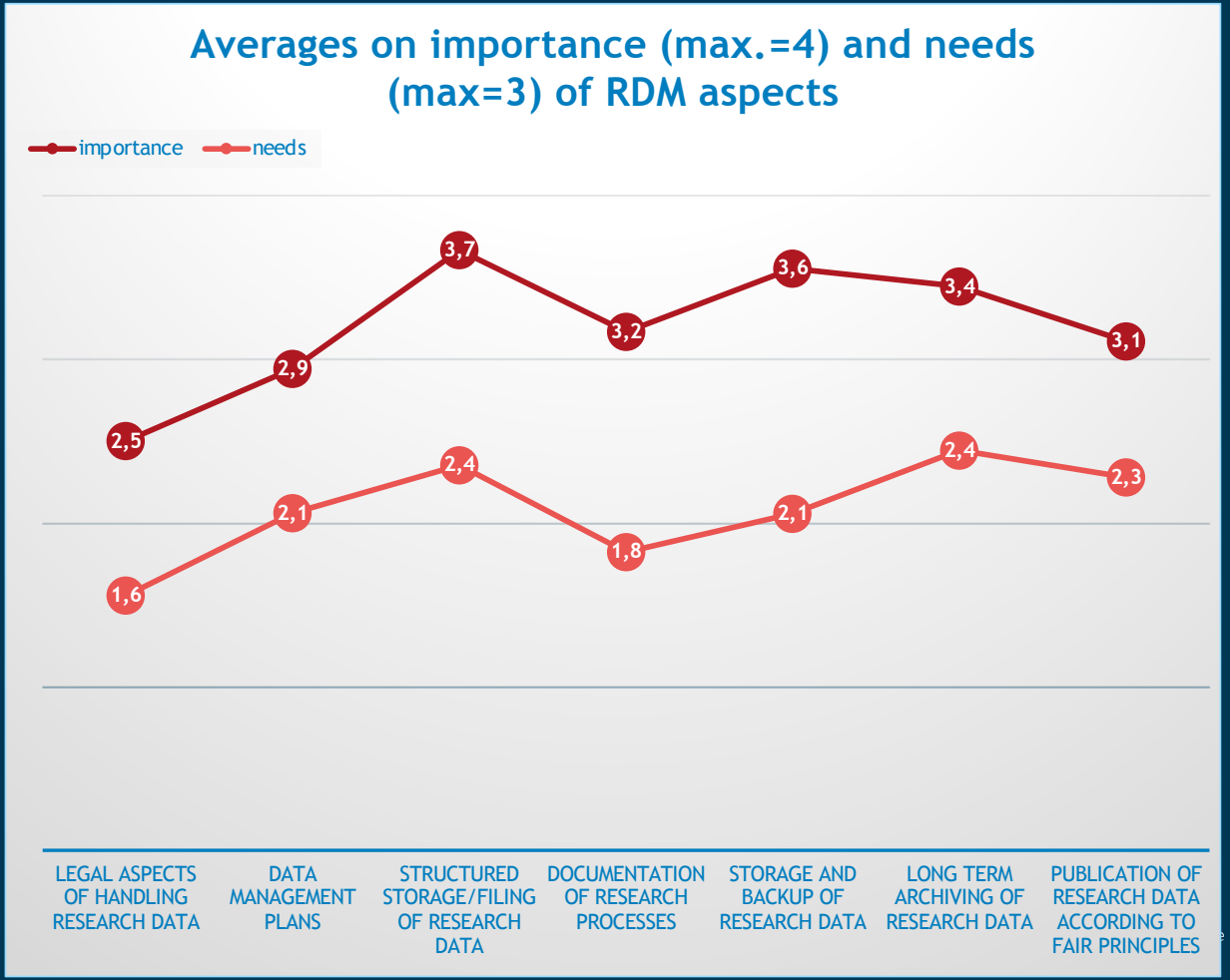
Research Data Management Questionnaire among NeuroCure PIs and Group Leaders

Researcher's perspective:

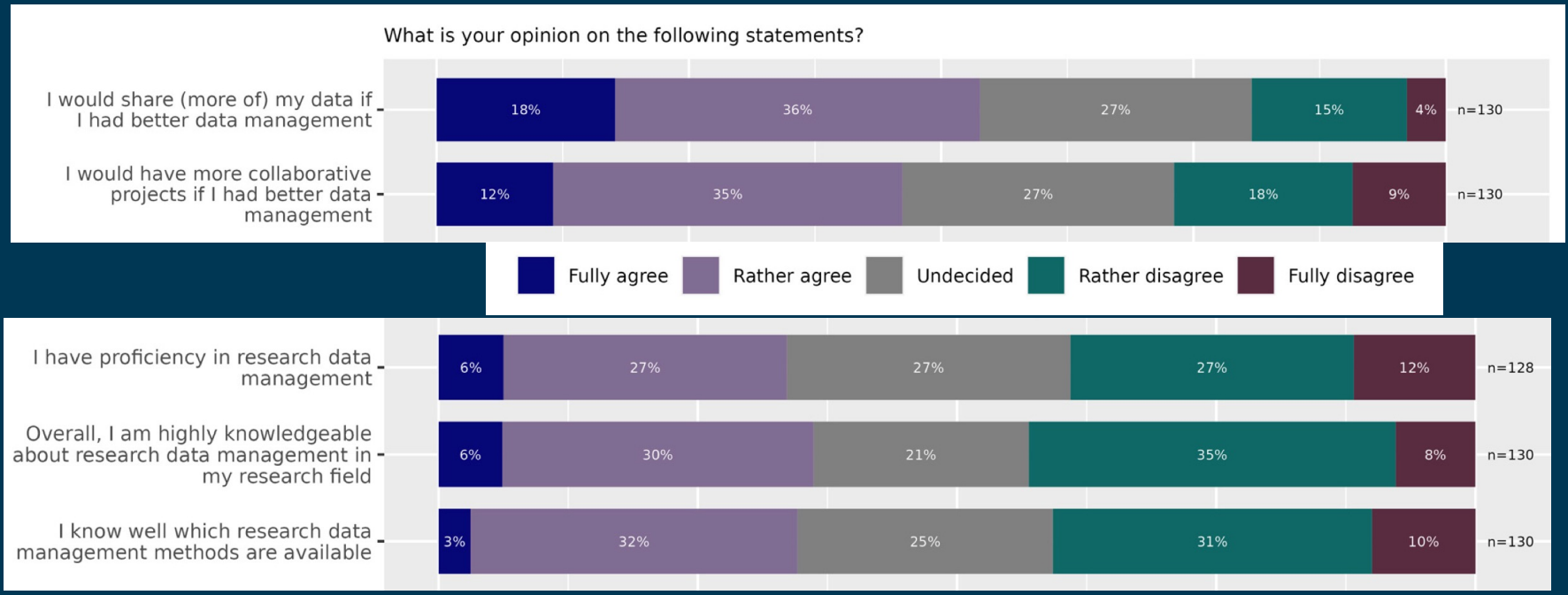
- Structured storage and filing
- Long term achiving and access

are most important - reflect mainly the use and access to data

But documentation is rated lower



Research Data Management and Data Sharing for Reproducible Research—Results of a Community Survey of the German National Research Data Infrastructure Initiative Neuroscience



Awareness of potential benefits, yet there is a lack of knowledge and best practices

eNeuro 7 February 2023, 10 (2) ENEURO.0215-22.2023; https://doi.org/10.1523/ENEURO.0215-22.2023

BIH Berlin Institute of Health at Charité

Neuroscientists need support to build a lasting RD infrastructure

From the abstract:

and opinions about RDM. The German neuroscience community perceives barriers with respect to RDM and data sharing mainly linked to (1) lack of data and metadata standards, (2) lack of community adopted provenance tracking methods, (3) lack of secure and privacy preserving research infrastructure for sensitive data, (4) lack of RDM literacy, and (5) lack of resources (time, personnel, money) for proper RDM. However, an overwhelming majority of community members (91%) indicated that they would be willing to share their data with other researchers and are interested to increase their RDM skills. Taking advantage of this willingness

- **NFDI did not fund the NFDI-Neuro consortium**
- **NFDI4health, NFDI4bioimaging, GHGA**
- **What can we do as Cluster of Excellence ?**

eNeuro 7 February 2023, 10 (2) ENEURO.0215-22.2023; https://doi.org/10.1523/ENEURO.0215-22.2023

VOS coordinator actions can help but are limited

- Counseling for DMPs
- Course (ReproducibiliTeach) that teaches basic RDM best practices, data dictionaries, and data repository uploads

BUT: Visitation and document reviews are barriers and generally not welcome
Therefore: A review and specific counseling from within a group is almost impossible
Often domain-specific knowledge is required

There is a clear limit what an outside person is able to do → true changes in RDM infrastructure must be initiated from within a research environment

Shaping the boundaries of RETAIN

- **Meeting with Uli Dirnagl, Dietmar Schmitz and me (May 2022)**
  - Protected time model for researchers implementing RDM or financial reward
  - Ideally suited for almost finished PhDs or early postdocs
  - One year fully funded position
  - Condition: must improve RDM for all group members in a sustainable fashion
- **Idea received approval by the board of directors (Jan/Feb 2023)**
  - VOS: Generation of fellowship call and application forms

External evidence of fellowship for RDM

ELIXIR-UK FAIR Data Stewardship Training Fellowship (est. 2021) – 24 fellows

- Fellows receive training are supported in creating learning materials and offering local training in their fields of FAIR data expertise **(life sciences)**

Skills4EOSC Fellowship Programme (2023/2024) – 9 fellows

- sponsors short secondments for data professionals, with “institutional capacity building” as an explicit objective

Both programs are aimed at “data professionals” or “data stewards” and provide learning pathways and training material.

A research data management (RDM) community for ELIXIR (https://doi.org/10.12688/f1000research.146301.1)

Overview of RETAIN Program (Research Data Management Implementation in the Neurosciences

- Goal: Develop a sustainable RDM infrastructure in a neuroscientific research environment
- Who: last PhD student / postdoc student receives a fellowship for 2 years that covers 50% of salary
- Applicant shows interest in RDM, data literacy, and willingness to pursue skills further with specialized education
- Written proposals require statement of support from the PI for RDM implementation
- Fellowship program is overseen by VOS Coordinator (incl. progress tracking)
- Proposals and Progress Reports need approval from the NeuroCure Board of Directors

Programm Benefits

- Provides sufficient time to develop and implement best-fitting solutions for an entire research team in their environment
- Researchers don’t have to stop their research (50%)
- Fellow’s expertise and skills qualify them for data manager and data steward positions in the future
- Show-case labs serve as blueprints for other NeuroCure groups and beyond
- RETAIN is a proof-of-concept program but if successful – it is adaptable and scalable
- VOS coordinator guides and consults fellows

RETAIN application package

1. Description of RD handling in the research environment (lab standards, conventions, organizational structures for storage, archiving; workflow, if they exist)

2. Table with: data types, metadata, data processing, data archiving, use of ELN, data access for collaborators, data sharing, legal obligations

3. If you were to re-use data created by others, which properties and metadata should a dataset have to qualify as reusable?

4. Description of the ideal RD solution achievable in the research environment

5. For applications information: general project outline, workplan, and resources

6. Commitment of support by applicants PI

7. CV of applicant

BIH Berlin Institute of Health at Charité

Timeline of RETAIN fellowships

- Call for applications: 1.6.2023 – 31.7.2023
- Review & Evaluations: 15.8.23
- Proposal Presentations to BOD: 30.9.23
- Start of fellowships: 1.1.24
- VOS Progress Meetings: 25.4.24, 25.9.24, 05.03.25, 26.05.25, 05.11.25
- BOD progress report: 30.11.24
- BOD final report: 12.12.25
- End of fellowships: 31.12.25
- Independent evaluations and follow up monitoring in 2026

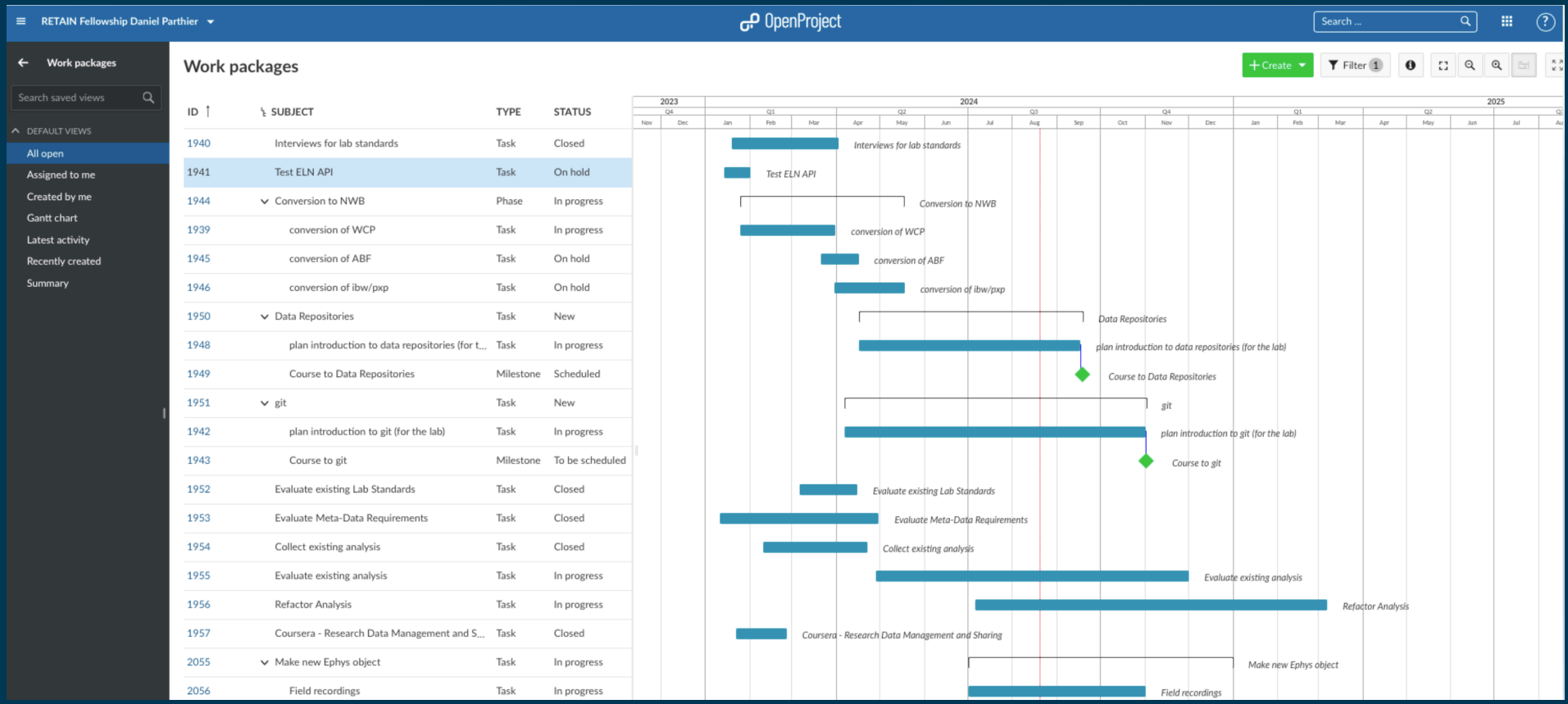
BIH QUEST
Center for Responsible Research

BIH Berlin Institute of Health @Charité

OpenProject for low-threshold progress tracking

- Fellow break up their projects into phases, tasks, and milestones
- Fellows' first exposure to Project Management Software
- Fellows update OpenProject → email notification
- Transparency increased; potential problems are detected early
- Read access: VOS coordinator, other fellow
- Questions/clarifications: MS Teams communication channel
- Joint progress meetings every 5 months

OpenProject for low-threshold RETAIN progress tracking



BIH Berlin Institute of Health at Charité

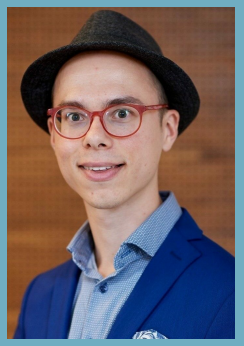
RETAIN fellow

- Daniel Parthier, PhD in the lab of Dietmar Schmitz (preclinical)- Charité
- Electrophysiology, confocal/STED imaging and behavior
- 3 ephys recording setups (different makers, different software, different output formats)
- Synchronized behavioral data and electrophysiological recordings
- Project metadata + study documentation integration with FAIR data
- Engagement with data steward community - DSgG meeting 10/24

BIH QUEST
Center for Responsible Research

BIH Berlin Institute of Health @Charité

RETAIN fellow



- Jojo Vanhoecke, PhD student – PI: Andrea Kühn (clinical)- Charité
- Deep brain stimulation/sensing device in patients with Parkinson
- human clinical data (e.g. behavioral, clinical, and routine clinical care data), sometimes ”event data” or continuous (home) recordings
- Construct a novel data container analogous to the Brain Imaging Data Structure (BIDS)
- Share metadata with publications and speed up data transfer agreements by prepared templates
- Develop DMP for most common recording setups

BIH QUEST
Center for Responsible Research

BIH Berlin Institute of Health @Charité

Preliminary learnings from the RETAIN fellowship

- Believe: RDM measures just need to be prescribed
- Lack of RDM knowledge is the main barrier for success
- Even with active implementors - one barrier remains:

  The human factor

BIH QUEST
Center for Responsible Research

BIH Berlin Institute of Health @ Charité

From an internal survey by Daniel Parthier:

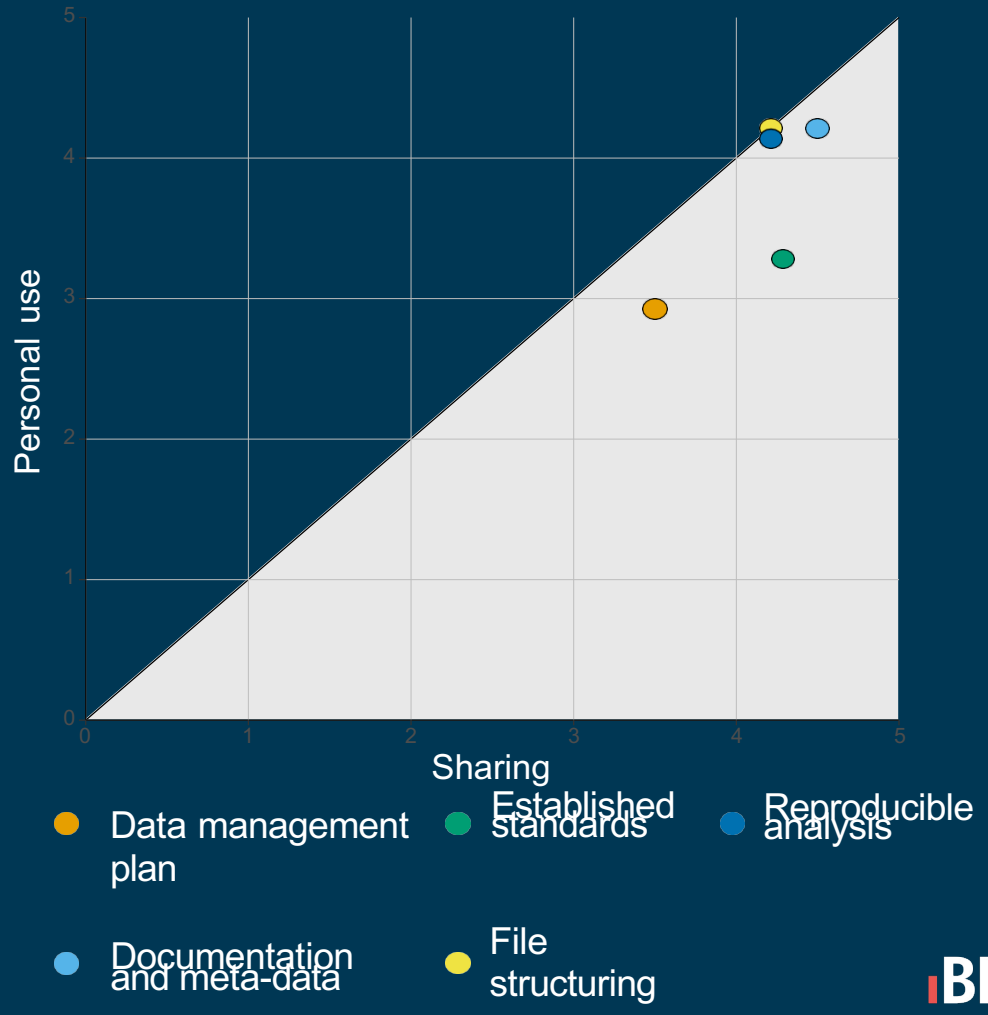
1-on-1 talks with subset of lab
in-vitro electrophysiology

Identify concrete needs

What do we need?

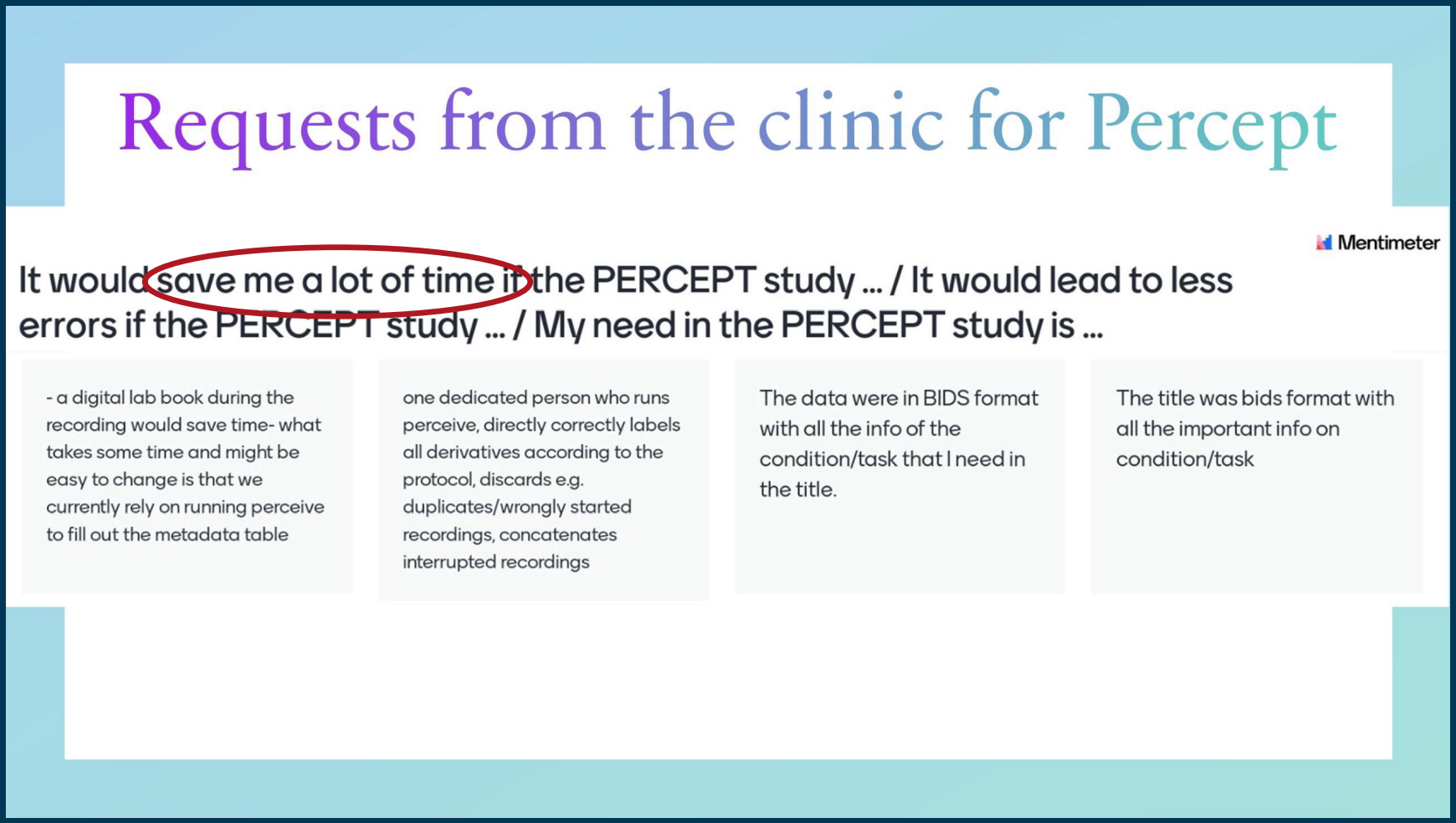
- Different priorities
- Mainly associated usefulness for others
- Personal benefits have to be highlighted

Perceived Usefulness



Slide from Daniel Parthier RETAIN project

From an internal survey by Jojo Vanhoecke:



Slide from Jojo Vanhoecke RETAIN project

Motivation for RDM – Altruism or personal gains ?

- **Both need to be balanced**
- **Consider „future personal gains“ →**
  - Data can be checked & reviewed faster with more confidence
  - Work continuity when responsbility transitions (your work survives your absense; you can quickly jump into abandoned projects)
  - Cognitive offloading & error avoidance
  - Faster restart after project interruption

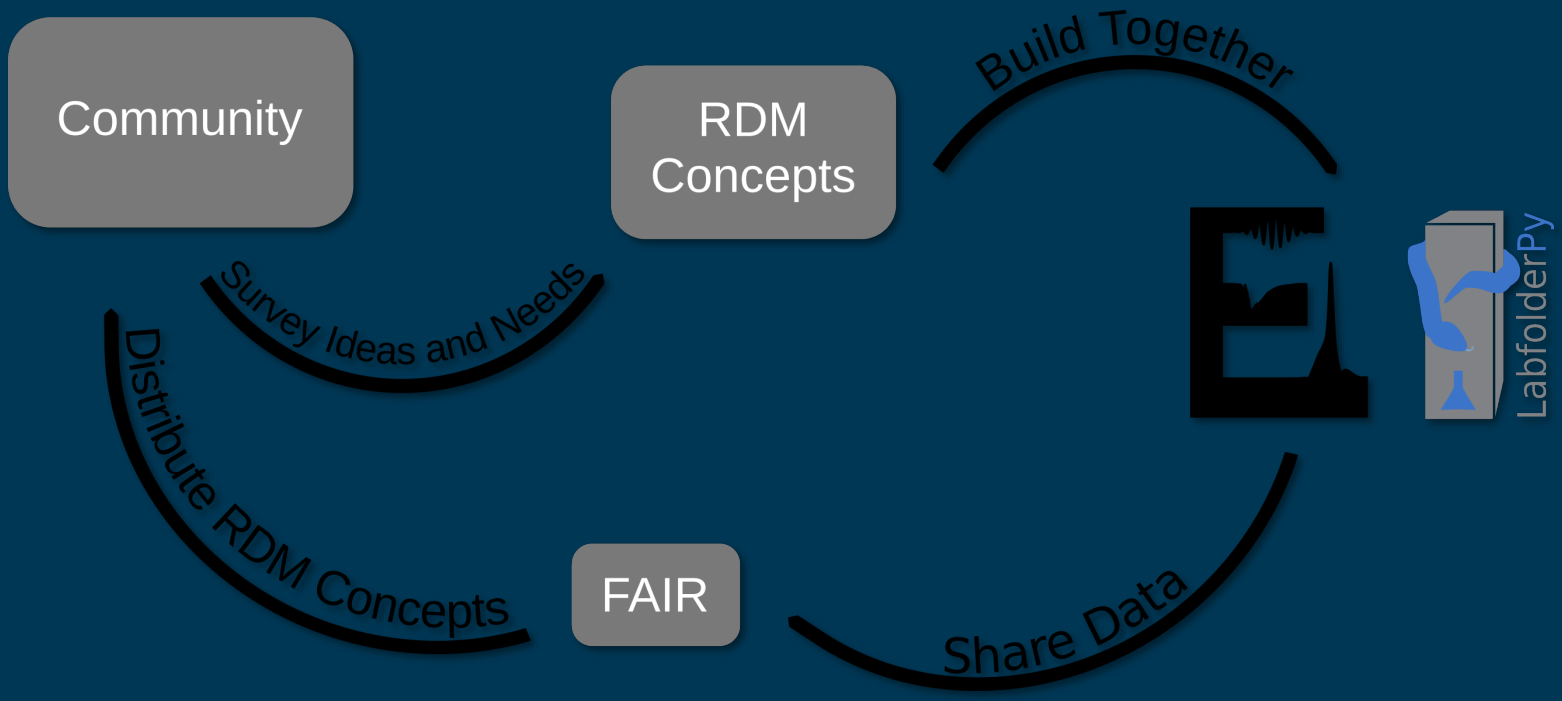
Preliminary learnings from the RETAIN fellowship

1. RDM workflows & solutions need to have benefits beyond future sharing

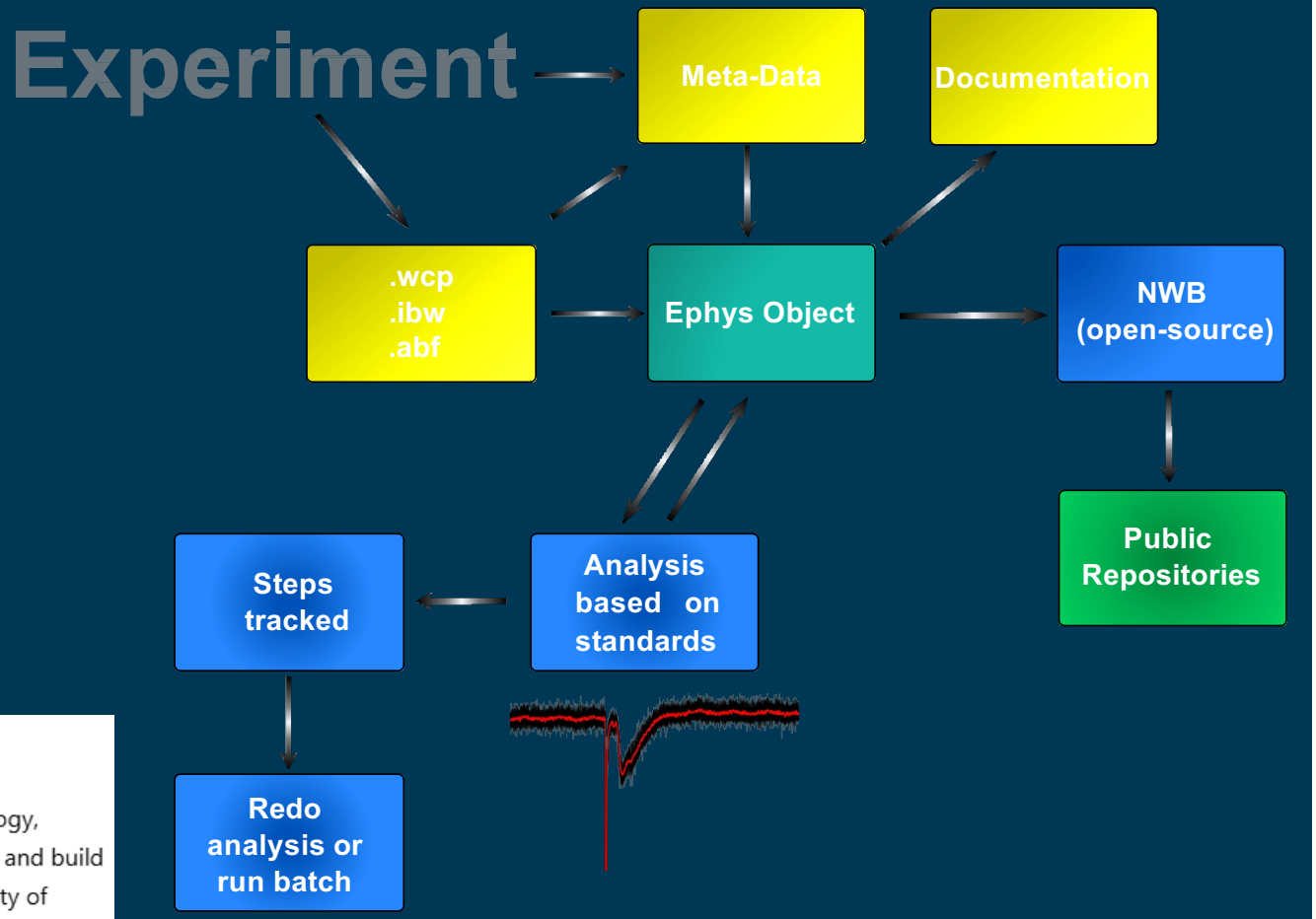
- Time saving & efficient
- Improve data quality
- Improve follow up worksteps (analyses)
- Anticipate changes

BIH Berlin Institute of Health at Charité

Project Building Blocks



by Daniel Parthier:



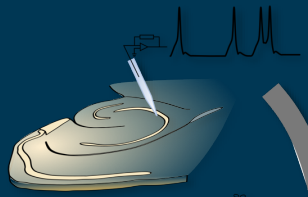
Neurodata Without Borders

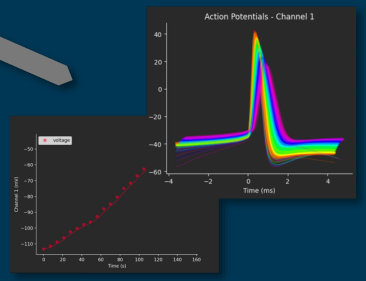
Neurodata Without Borders (NWB) is a data standard for neurophysiology, providing neuroscientists with a common standard to share, archive, use, and build analysis tools for neurophysiology data. NWB is designed to store a variety of neurophysiology data, including data from intracellular and extracellular electrophysiology experiments, data from optical physiology experiments, and tracking and stimulus data.

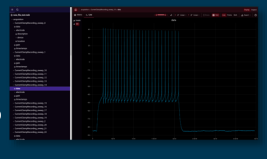
Slide from Daniel Parthier RETAIN project

BIH Berlin Institute of Health at Charité

Ephys







Feedback from group members

Open-Source and Python based

Integration of Ephys into workflow

Compatible with different file formats (extendable)

Guides you through steps

Output to NWB

NEURODATA WITHOUT BORDERS

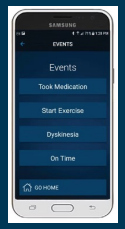
Slide from Daniel Parthier RETAIN project

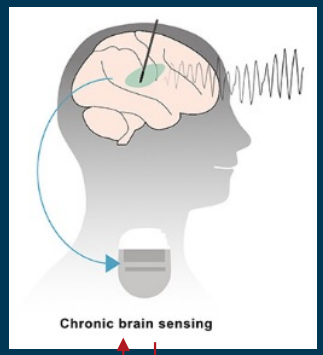
BIH Berlin Institute of Health at Charité

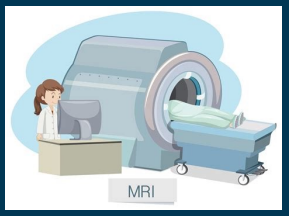
Data plan for closed-loop adaptive deep brain stimulation

Intro

- Medtronic Percept PC
- **Next-generation neurostimulator**
- continuous monitoring
- **Lab/ home environment**
- Long-term follow-up
- **150+ datasets** with 5300 recordings (ongoing)







PHÖNIX PACS GMBH

Challenges

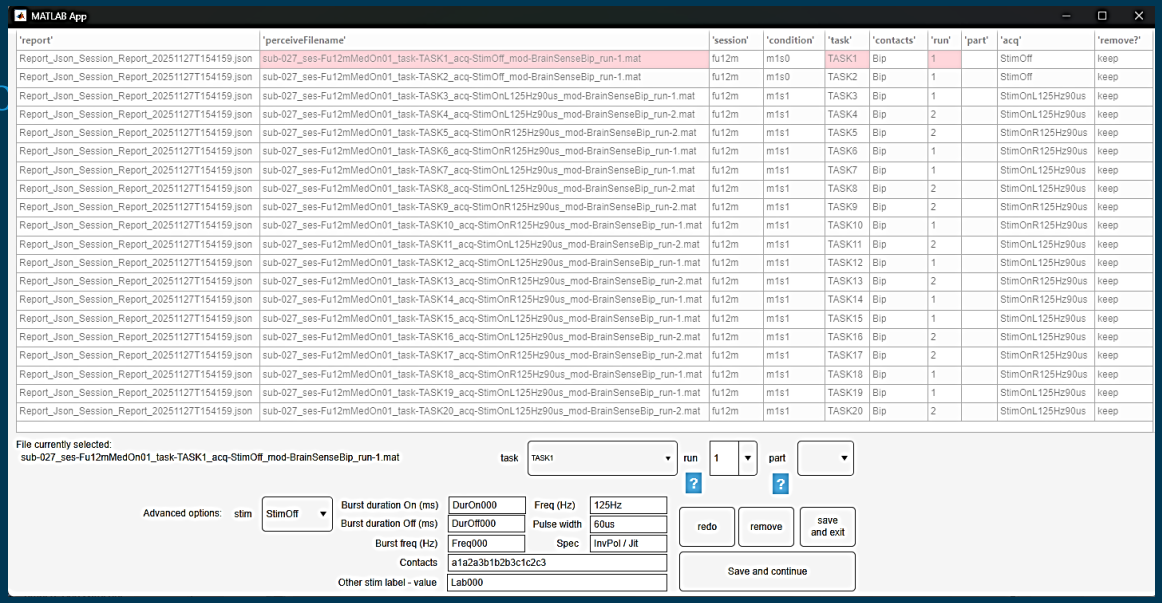
- new realm of continuous recording
  - **data integration**
- **time-consuming** for clinicians (!)
- **privacy concerns** – identification
- How to make big data usable in the clinic?
- How to do **Open Science**?

Slide from Jojo Vanhoecke RETAIN project

BIH Berlin Institute of Health at Charité

Perceive Toolbox: What it is all abo

- **Percept Data extraction**
- **Digitalized labbook (GUI)**
- Integrated GitHub page
- **Data standard using Brain Imaging Data** Structure
- Automatized data workflow
  - Labeling of data
  - Concatenation of data
  - Artefact cleaning
  - Automated test flows



MATLAB GUI

GitHub

BIDS
BRAIN IMAGING DATA STRUCTURE

```
Sub-001
Ses-MedOn
Task-Rest
Acq-StimOnR
Mod-BrainSense
Run-1
Part-2
```

Slide from Jojo Vanhoecke RETAIN project

BIH Berlin Institute of Health at Charité

by Jojo Vanhoecke:

Data plan for closed-loop adaptive deep brain stimulation

Deliverables

- DMP on dmptool.org
- electronic clinical lab book
- **Brain Imaging Data Structure (BIDS)**
- code to retrieve & standardize data

- automatic data analysis with figures that inform the clinician
  e.g. dotbase platform

- Open Science: FAIR principles

main deliverable: clean clinical data set and work flow

Slide from Jojo Vanhoecke RETAIN project

Preliminary learnings from the RETAIN fellowship

1. RDM workflows & solutions need to have benefits beyond future sharing

- Time saving & efficient
- Improve data quality
- Improve follow up worksteps (analyses)
- Anticipate changes

2. Investment in composite formats (containers) and toolboxes

- Efficient: data and metadata and traceable analyses together
- Scaleable
- Standardized

BIH Berlin Institute of Health at Charité

Preliminary learnings from the RETAIN fellowship

1. RDM workflows & solutions need to have benefits beyond future sharing

- Time saving & efficient
- Improve data quality
- Improve follow up worksteps (analyses)
- Anticipate changes

2. Investment in composite formats (data/metadata containers)

- Efficient: data and metadata together
- Scaleable
- Standardized

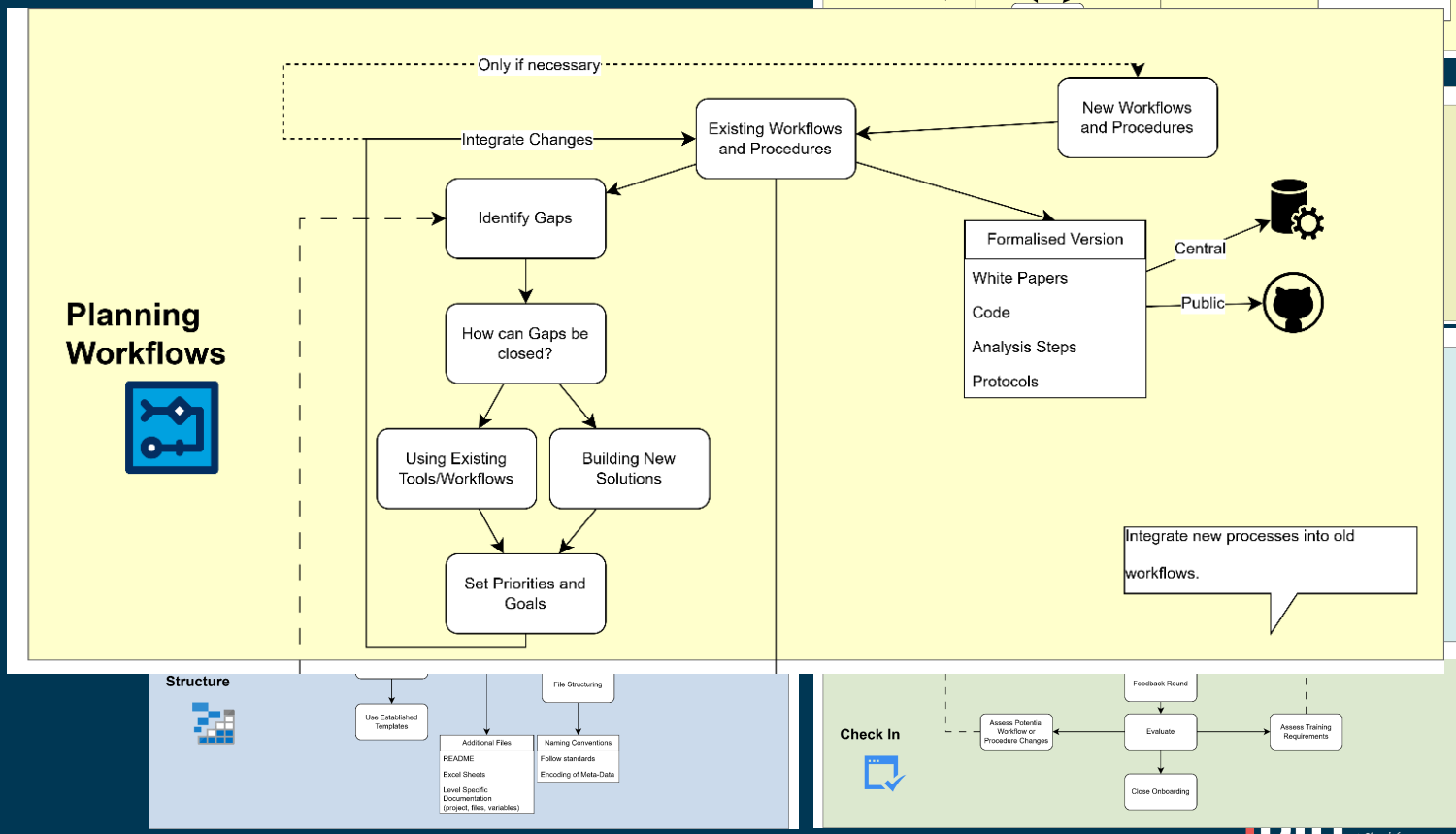
3. Internal communication and internal training are key

- Data onboarding for new member (maybe also data offboarding?)
- Mandatory training on new tools

BIH Berlin Institute of Health at Charité

Data Onboarding Framework

- Assess
- Access
- Structure
- Planning
- Needs
- Practice
- Check-in



Slide From Daniel Parthier RETAIN project

What is the role of VOS in RETAIN ?

- Based on knowledge interest - RDM courses and resources were suggested
- Ensure that projects advance and results align with the goals of the fellowship
- Make tailored suggestions (papers, meeting, ressources, courses)
- Problem solving - when tasks are stuck
- Available for questions or concerns
- Conduct regular project updates meetings - provide feedback
- Ensure progress towards deliverables
- Reporting to board of directors

BIH QUEST
Center for Responsible Research

BIH Berlin Institute of Health @ Charité

RETAIN fellowship – next steps

- Not all deliverables were met in time, e.g. community rollout and feedback from them
- Both solutions are in testing phase for their environment
- Both fellows are committed to finish up beyond the funding phase
- Publication of process, solutions and tools (most of them are available on GitHub) by the fellows
- Ultimate: FAIR data availability along side publications from both labs

SUMMARY

- Fellowships for RDM have the potential to generate true progress towards the successful implementation of RDM measures that produced reusable FAIR data
- Shared solutions can be to adopted by other Neuroscience community members
- Domain-data-expertise can help the fellow's professional career development
- Empower ECRs
- If overall successful, the RETAIN fellowship concept could be a funding mechanism for other communities that also struggle with their RDM efforts
- Another RETAIN fellowship will start in 2026

Thank you very much!

https://neurocure.de/career/NeuroCure-Research-Fellows/retain-fellowships.html
https://zenodo.org/records/13771491
https://zenodo.org/records/13772103

BIH QUEST
Center for Responsible Research

BIH Berlin Institute of Health @Charité