INCF: global collaboration on FAIR standards in neuroscience

Malin Sandström (PhD)
Community Engagement Officer
INCF

malin@incf.org
Twitter: @msandstr
ORCID ID: 0000-0002-8464-2494

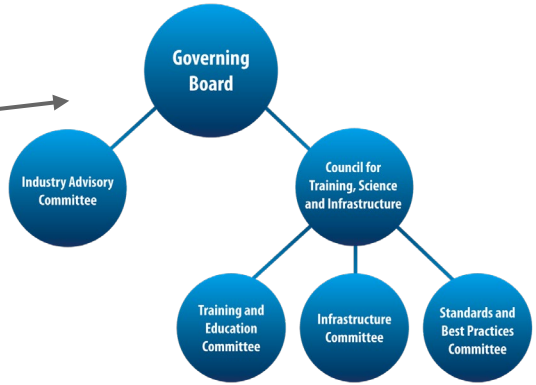
incf enabling open and FAIR neuroscience

INCF

The mission of INCF is to develop, evaluate, and endorse standards and best practices that embrace the principles of Open, FAIR, and Citable neuroscience. INCF also provides training on how standards and best practices facilitate reproducibility and enables the publishing of the entirety of research output, including data and code.

- Founded on recommendation by OECD
- Secretariat at KI since 2006
- Governed by member representatives

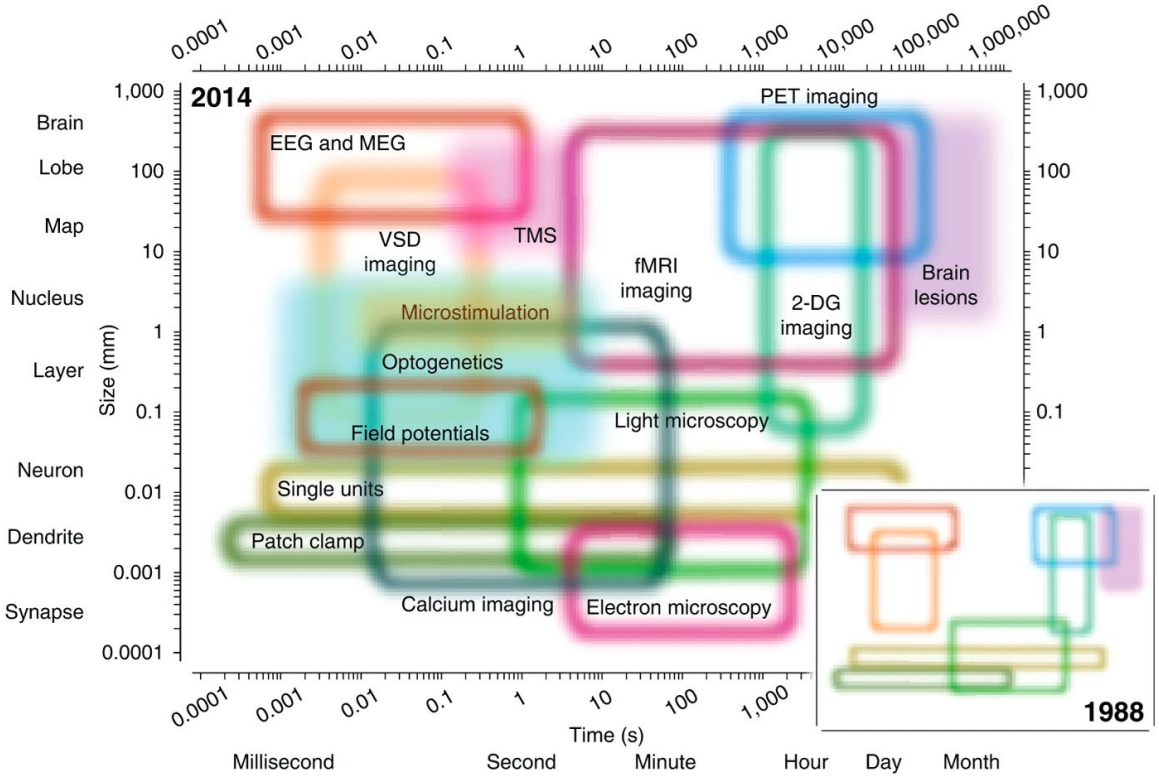
  Our governing councils are responsible for the strategic vision for the organization, and for setting INCF's priorities in terms of science, training, and infrastructure.
- Global network on 5 continents
- 13 Working Groups
  - 4 WGs are communities of practice
  - 9 WGs are developing or implementing community standards
- Advocacy & outreach for furthering implementation and adoption of **FAIR**, openness, reproducibility and scientific rigor in neuroscience



Neuroscience has always had problems with big & complex data

- 10^8 difference in length scales studied
- 10^{10} difference in time scales studied
- MANY different modalities
- Imaging tech progress gives "data explosion"

"This problem can only be solved with global collaboration on infrastructure, tools and methods."



Sejnowski et al. *Putting big data to good use in neuroscience* **Nature** (2014)
Huang & Luo *It takes the world to understand the brain* **Science** (2015)

INCF's focus

***outreach to** & **coordination** of the international neuroscience community on*

Focus areas 2005 - 2015

- Infrastructure
- Openness, reproducibility & scientific rigor
- Training & education in data science and computational skills
- International coordination to harmonize/avoid duplicate efforts

Added focus from 2016 →

- **FAIR**
  - Community standards (CS)
    - Facilitate collaborative CS conception & development
    - Harmonize CS efforts across brain initiatives
    - Community-driven CS endorsement

Abrams et al A Standards Organization for Open and FAIR Neuroscience: the International Neuroinformatics Coordinating Facility Neuroinformatics (2022) https://doi.org/10.1007/s12021-020-09509-0
M.Martone The importance of community organizations for open and FAIR efforts in neuroinformatics *GigaScience* (2022) https://doi.org/10.1093/gigascience/giac060

Our community members

Mainly **neuroscientists** interested in informatics, data science, **data sharing** and **collaboration**

- PIs who were early pioneers in open and reproducible neuroscience (now they form the core of our steering committees)
- Infrastructure providers
- Neuroscience researchers interested in openness & reproducibility
- Researchers developing tools part-time for their own research & research collaborations - that became widely used community tools
- PhD/PD students and ECRs with a strong interest in collaboration and open research; interested in FAIR because it increases the utility of open research outputs

What we (and our network) do

- serve as a collective arena for neuroscience researchers, tool developers and infrastructure providers to meet, identify joint problems and collaborate on solutions
- support the development, implementation, and adoption of open and FAIR *community standards* and *best practices* (SBPs)
    - Working Groups that develop or implement SBPs
    - **Endorsement process for neuroscience SBPs**
- develop, curate & promote training and educational resources in neuroinformatics; providing open training materials on FAIR data management, tools, standards & best practices via training.incf.org
- work with journals to implement standards and best practices
- facilitate harmonization & collaboration between international brain projects
- partner with stakeholders to promote and prioritize FAIR in neuroscience at global, national and local levels

incf

Why endorse community standards?

Many well-known (industry) standards are top-down standards, or standards *de jure* - designed and promoted by an authority (ISO, NISO, IEEE) whose authority and integrity are already guaranteed.

In contrast, most community standards are *de facto* standards, that gain their status by becoming widely used. They usually develop at a grass-roots level, in communities with close ties between users and developers. Adoption and promotion is slow and organic, driven by word-of-mouth and trusted personal contacts.

Our endorsement process includes review by an expert committee, but the most important step is community review - feedback and expressions of support from actual and intended users. By endorsement, we guarantee that a standard is deemed to be trustworthy and useful to the wider neuroscience community. This will help drive adoption and give standards developers credit for their work.

Why *metadata* standards? They are the backbone of FAIR:

Metadata are mentioned in every FAIR sub-principle except one, but especially

F2. Data are described with **rich metadata** (as described in R1)

I2. (Meta)data use vocabularies that follow FAIR principles

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

Why do communities need to standardize?

To be FAIR, but also:

- Standards increase *reproducibility* - both for yourself and others
- Standards make it easier to *collaborate*
- *Reuse* becomes much easier
- *Sharing* becomes more worthwhile
- Possible to *compare or merge* datasets
- Data standards (for open data) can catalyze growth of *a tool ecosystem* of compatible and interoperable tools

Some example standard efforts in neuroscience

Brain Imaging Data Structure (BIDS)

A directory structure + naming convention + controlled field-specific vocabularies for brain imaging data

Neuroscience Electrophysiology Objects (Neo)

A data model that can represent and write/convert a large amount of electrophysiology file formats

ARTEM-IS and eCOBIDAS

Minimal reporting standards for EEG and MRI data

Neuroscience Without Borders: Neurophysiology (NWB:N)

A file format for multimodal (~ many different kinds of combined) neurophysiology data

Research Resource IDentifier (RRID)

A versatile permanent identifier for *research objects* (antibodies, instruments, tools, projects)

BIDS: “The Marie Kondo for neuroimaging data”

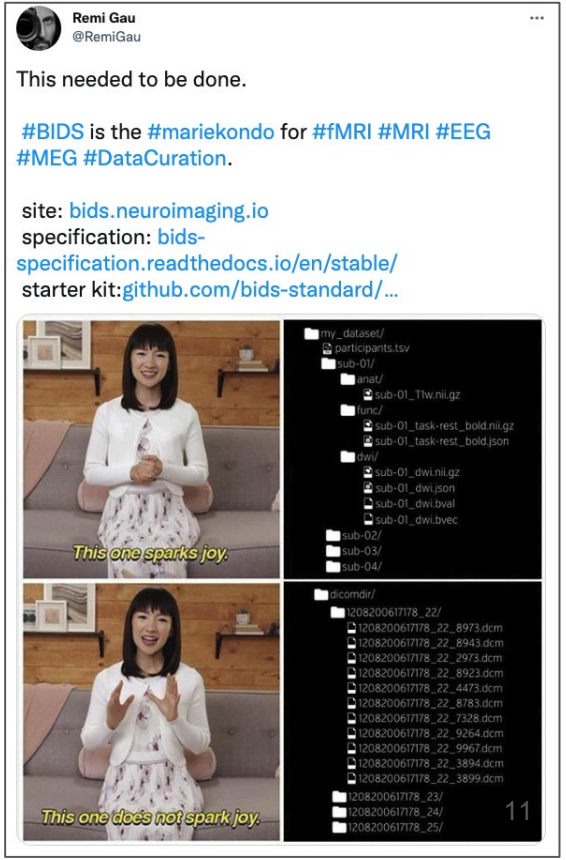
Original DICOM data

- dicomdir/
  - 1208200617178_22/
    - 1208200617178_22_8973.dcm
    - 1208200617178_22_8943.dcm
    - 1208200617178_22_2973.dcm
    - 1208200617178_22_8923.dcm
    - 1208200617178_22_4473.dcm
    - 1208200617178_22_8783.dcm
    - 1208200617178_22_7328.dcm
    - 1208200617178_22_9264.dcm
    - 1208200617178_22_9967.dcm
    - 1208200617178_22_3894.dcm
    - 1208200617178_22_3899.dcm
  - 1208200617178_23/
  - 1208200617178_24/
  - 1208200617178_25/

Data structured as BIDS

- my_dataset/
  - participants.tsv
  - sub-01/
    - anat/
      - sub-01_T1w.nii.gz
    - func/
      - sub-01_task-rest_bold.nii.gz
      - sub-01_task-rest_bold.json
    - dwi/
      - sub-01_dwi.nii.gz
      - sub-01_dwi.json
      - sub-01_dwi.bval
      - sub-01_dwi.bvec
  - sub-02/
  - sub-03/
  - sub-04/

BIDS is a format for standardizing and describing outputs of neuroimaging experiments (left) in a way that is intuitive to understand and easy to use with existing analysis tools (right).



What makes a good (endorse-able) standard?

A standard or best practice can be endorsed by INCF if it enables or facilitates FAIR research practice. But it must also be **developer-friendly**:

- A clear **specification** that can be implemented by others
- Clear and extensive **documentation**
- A **reference implementation** in a widely used programming language
- An **API** or **CLI** for programmatic access in at least one common programming language (e.g. Python)
- Transparent and sustainable **governance** open to **community input**
- Strong **user support**/engaged user community

https://www.incf.org/activities/standards-and-best-practices

Software tools & tool developers are essential to FAIR

- Cite all software your work depends on
  - github link - better than nothing
    Better: check for citation info in root repo (CITATION.cff file)
  - tool paper(s) - better
  - tool identifiers - best (ideally, do both)
    - DOI
    - RRID scicrunch.org
    - SWHID softwareheritage.org
      (all github code already has one)
- Make it easy to cite your own software
  - add a CITATION.cff file to your root repo
  - deposit software where you get a permanent identifier

scientific **data**

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › scientific data › articles › article

Article | Open Access | Published: 14 October 2022

Introducing the FAIR Principles for research software

Michelle Barker, Neil P. Chue Hong, Daniel S. Katz, Anna-Lena Lamprecht, Carlos Martinez-Ortiz, Fotis Psomopoulos, Jennifer Harrow, Leyla Jael Castro, Morane Gruenpeter, Paula Andrea Martinez & Tom Honeyman

Scientific Data **9**, Article number: 622 (2022) | Cite this article

1440 Accesses | **221** Altmetric | Metrics

Abstract

Research software is a fundamental and vital part of research, yet significant challenges to discoverability, productivity, quality, reproducibility, and sustainability exist. Improving the practice of scholarship is a common goal of the open science, open source, and FAIR (Findable, Accessible, Interoperable and Reusable) communities and research software is now being understood as a type of digital object to which FAIR should be applied. This emergence reflects a maturation of the research community to better understand the crucial role of FAIR research software in maximising research value. The FAIR for Research Software (FAIR4RS) Working Group has adapted the FAIR Guiding Principles to create the FAIR Principles for Research Software (FAIR4RS Principles). The contents and context of the

stay in touch!

- blog: incf.org/blogs-list
- newsletter: incf.org/incf-newsletters
- Twitter: twitter.com/INCForg
- Our Neurostars forum: neurostars.org
- Our yearly Assembly: neuroinformatics.incf.org

INCF community at our latest in-person conference, the INCF Assembly in Warsaw, Poland in 2019.

Current institution & industry members

incf.org/join-incf

Let’s collaborate!

How can we work together?

Is there someone you think we should interact with?