

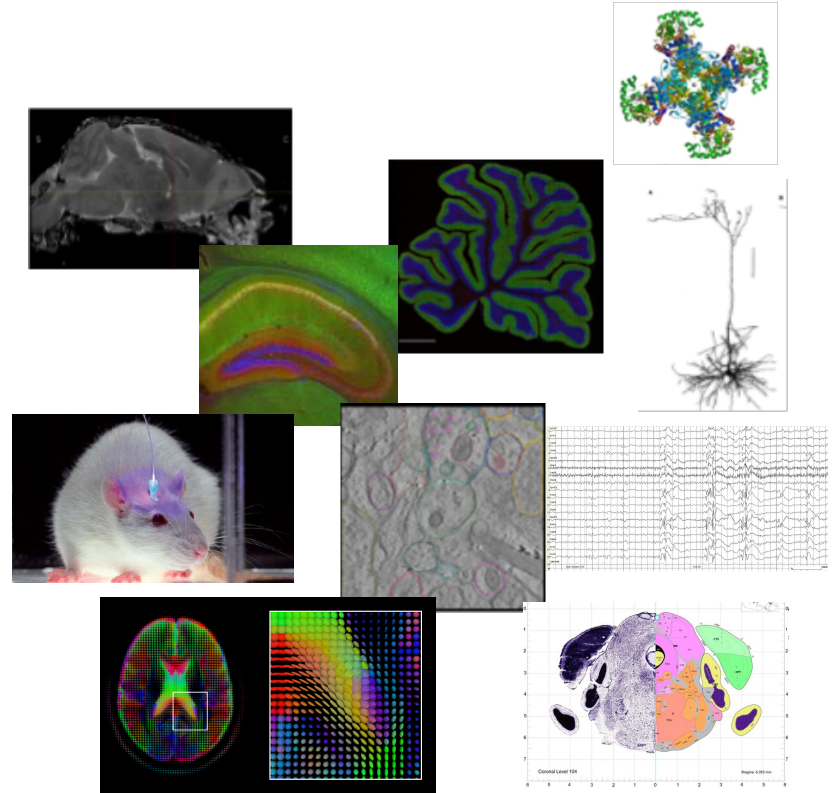
The State of FAIR in Neuroscience

Maryann E Martone
University of California San Diego
[SciCrunch.com](https://www.sci crunch.com)*

*SciCrunch is a UCSD start up that provides services around the use of Research Resource Identifiers (RRID)

Neuroscience...

- Very diverse experimental domain with often complex experimental paradigms combining
 - Physiology
 - Molecular
 - Behavior
 - Anatomy
- A single data set can comprise 1000's of individual files
- Heavily reliant on spatial information
 - Imaging
 - Brain atlases
 - Common coordinate systems
- Extraordinary number of vocabularies



Why principles?

“we don't want to re-invent the wheel”



Principles provide aspirations and guidance while respecting local needs and constraints and allowing infrastructures to scale with new technologies

Findable

- F1. (meta)data are assigned a *globally unique and persistent* identifier
- **F2. data are described with rich metadata**
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

Interoperable

- 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

Accessible

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- *A2. metadata are accessible, even when the data are no longer available*

Re-usable

- **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**

Notice of Data Sharing Policy for the BRAIN Initiative

“The BRAIN Initiative has made significant investments in a program to build the infrastructure that is needed to effectively share and interpret data. The goal of the program is to:

- build data science or informatics infrastructure that is useful to the research community;
- make data and tools openly available to the research community;
- help to enhance [FAIR principles](#) of data sharing and improve the [rigor and reproducibility](#) of BRAIN Initiative research; and enable or facilitate secondary analysis or data mining of BRAIN Initiative datasets.”

Data archives that have been established include:

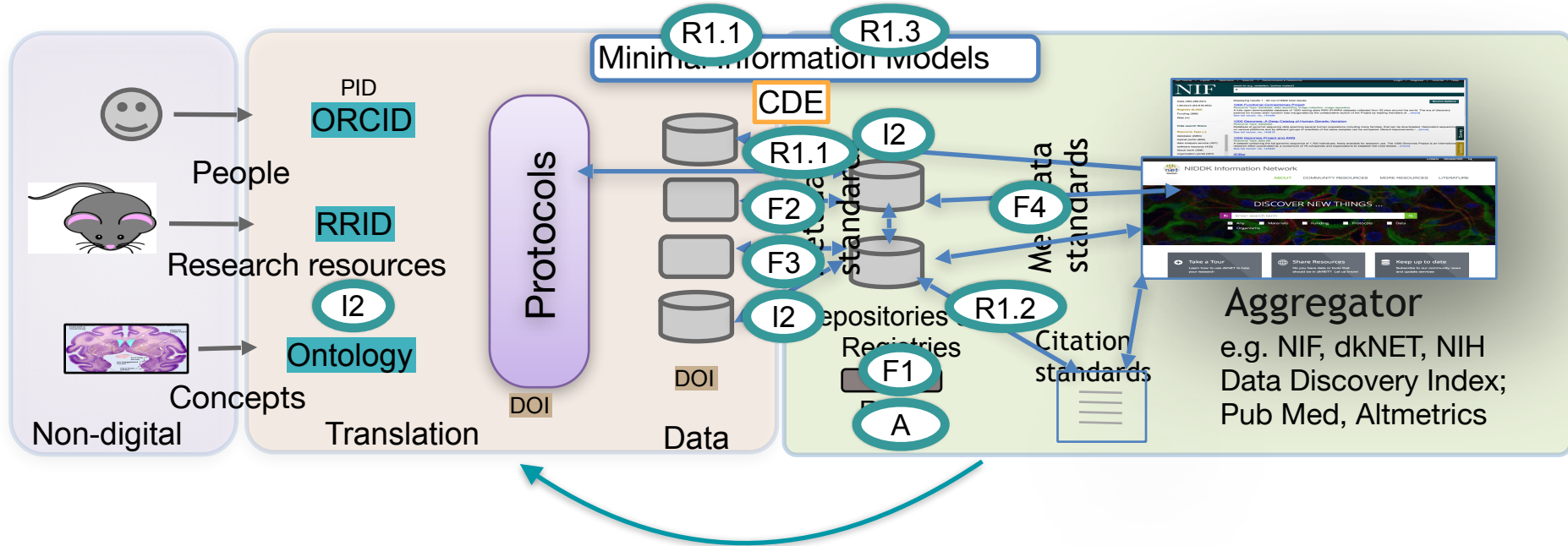
- 1) The Neuroscience Multi-omic Data Archive (<https://nemoarchive.org/about.php>, R24MH114788) to hold data from -omics experiments.
- 2) The Brain Image Library (<http://www.brainimagelibrary.org/index.html>, R24MH114793) to hold microscopy data.
- 3) Data Archive for the BRAIN Initiative (<https://dabi.loni.usc.edu>, R24MH114796) to hold data related to human electrophysiology experiments.
- 4) OpenNeuro (<https://openneuro.org/>, R24MH117179) to hold magnetic resonance imaging data.
- 5) Block and Object Storage Service (<https://bosddb.org/>, R24MH114785) to hold electron microscopy data.

You will put your data in a repository

You will use standards

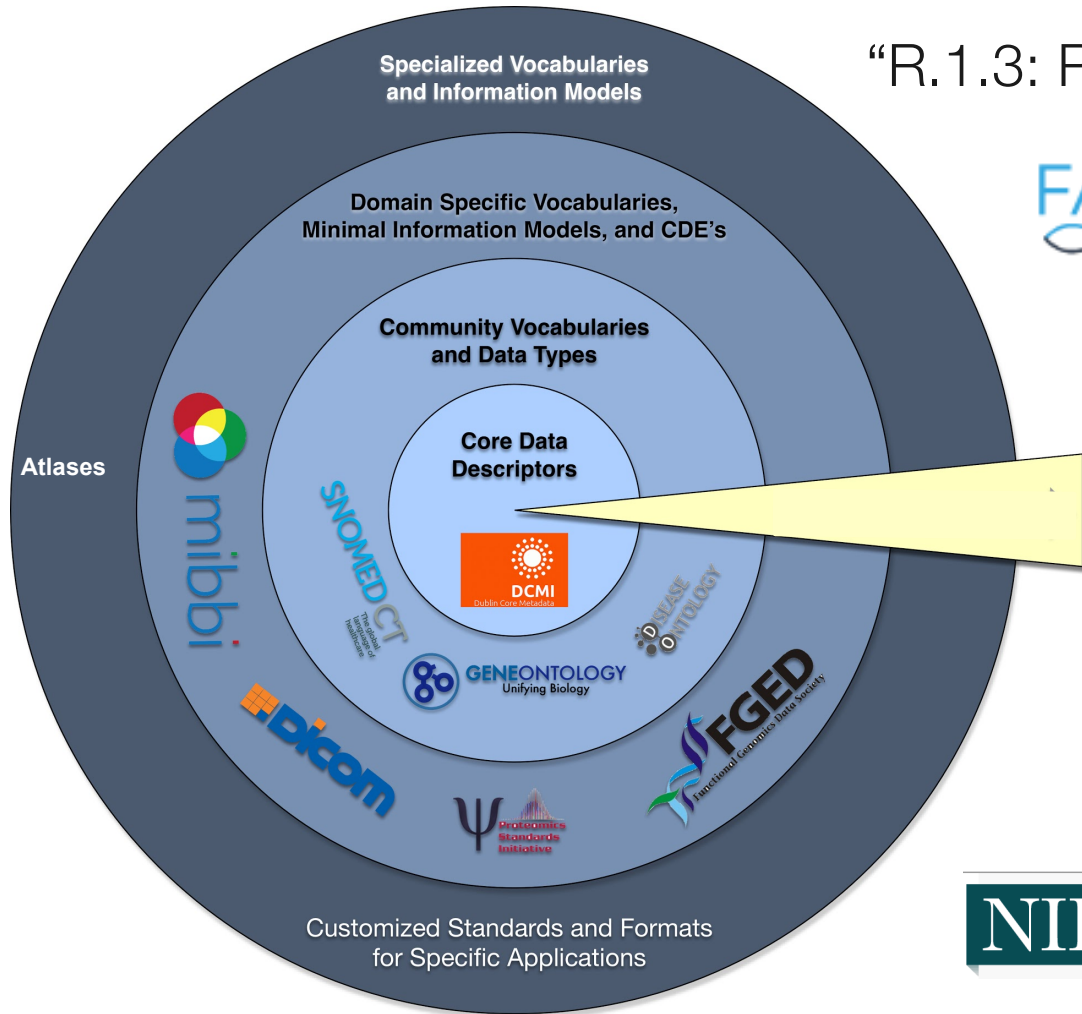
You will share your data

What is needed for FAIR neuroscience?



But what about the human infrastructure required?

“R.1.3: Relevant community standards”

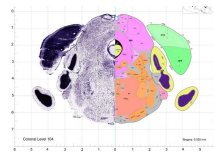


FAIRsharing.org
standards, databases, policies

FORCE11
The Future of Research Communications and e-Scholarship

RDA
RESEARCH DATA ALLIANCE

incf

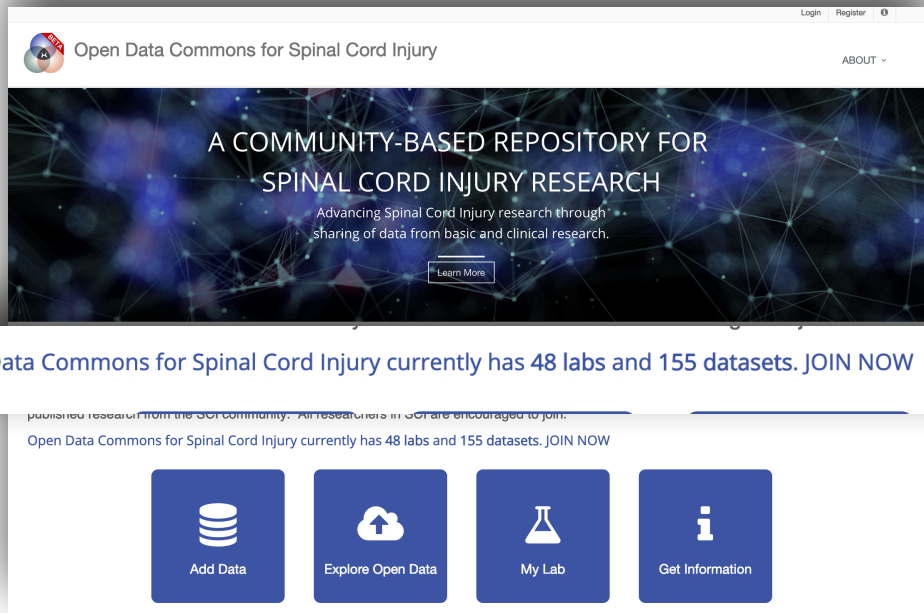


NIF

NEURODATA
WITHOUT BORDERS

Open Data Commons for Spinal Cord Injury

- Stakeholder governance: Researchers are taking charge of their field
- Majority of laboratories involved in preclinical spinal cord injury
- Sharing of data with each other; publishing data for the public
- Making SCI data: FAIR, open and citable



<https://scicrunch.org/odc-sci>

Jeff Grethe, Adam Ferguson, Karim Fouad and
ODC-SCI Steering Committee

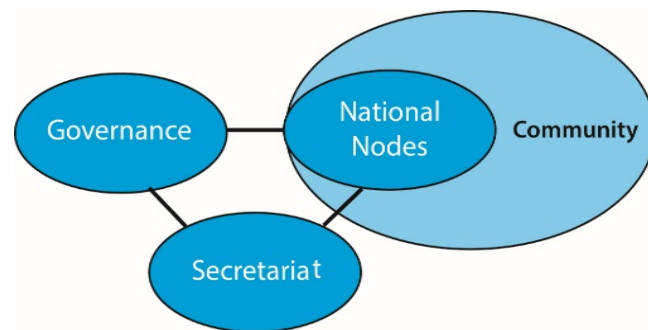
The International Neuroinformatics Coordinating Facility

INCF, established in 2006, is a network of researchers in 18 countries across 4 continents, working together with funders, publishers, industry, and organizations to promote and facilitate data reuse and reproducibility through the promulgation and development of open standards and best practices.

“The place for open and FAIR neuroscience”

Current Nodes*

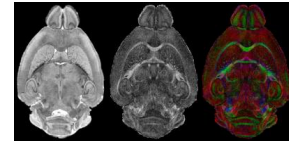
Australia	Belgium	Italy
Canada	Czech Republic	Netherlands
Japan	Finland	Poland
Malaysia	France	Republic of Korea
Norway	Germany	UK
Sweden	India	USA



*Membership model for INCF is changing

INCF: A standards organization to support global neuroscience

- International Neuroinformatics Facility (INCF): taking a leading role in coordinating standards and best practices for neuroscience data
- Adopted practices from W3C, NIST and other standards organizations for reviewing and endorsing standards and best practices
- Established the Standards and Best Practices Committee and a formal review and endorsement process
- **Standards need not have been developed by INCF working groups to be considered**

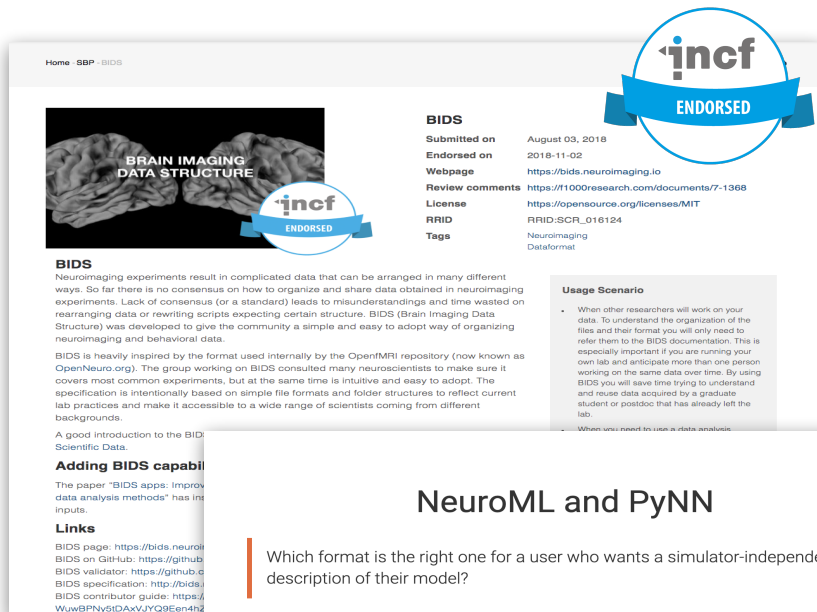


Focus on Standards and Best Practices Process

- Shepherd and support neuroscience during this new phase of open data and tools that will require a change in practice
- Ensure that neuroscience is supported by a robust set of interoperable **standards and best practices (plural)** that are embedded across the data lifecycle
- ***Provide guidance and a forum for neuroscience infrastructure providers on best practices and standards for their domain***
- Ensure that neuroscience interfaces with the broader life sciences community
- Help consortia, projects, individual users identify appropriate standards for their use case
- Provide a forum and process for evolution of existing standards and development of new ones
- Develop a community with expertise in the area of open and FAIR neuroscience
- Work towards long term sustainability of global neuroscience infrastructure

INCF standards and best practices review and endorsement

- Developed a set of consistent criteria supporting open and FAIR neuroscience
- Nomination and review process are community driven, e.g., 60 days of open comments
- Grievance procedure
- Developing standards portal
- Incorporate into training materials, workshops and Congress

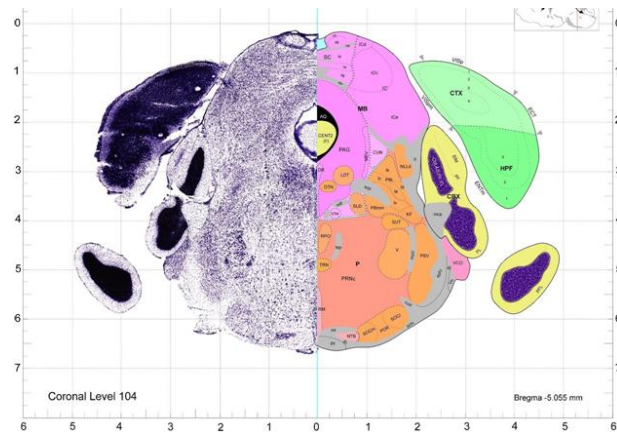


The screenshot shows the BIDS website with a prominent blue circular badge in the top right corner that reads "incf ENDORSED". The website content includes a header with "Home · BDP · BIDS", a main image of a brain with "BRAIN IMAGING DATA STRUCTURE" text, and a table of metadata. The metadata table lists: BIDS, Submitted on (August 03, 2018), Endorsed on (2018-11-02), Webpage (<https://bids.neuroimaging.io>), Review comments (<https://1000research.com/documents/7-1368>), License (<https://opensource.org/licenses/MIT>), RRID (RRID:SCR_016124), and Tags (Neuroimaging, Dataformat). Below the table, there is a "Usage Scenario" section with a bullet point: "When other researchers will work on your data. To understand the organization of the files and their format you will only need to refer them to the BIDS documentation. This is especially important if you are running your own lab and anticipate more than one person working on the same data over time. By using BIDS you will save time trying to understand and reuse data acquired by a graduate student or postdoc that has already left the lab." Below this, there is a section titled "NeuroML and PyNN" with a sub-header "Which format is the right one for a user who wants a simulator-independent description of their model?". The text below this section states: "The NeuroML model description language (neuroml.org) and the PyNN package for cross simulator neural network specification (neuralensemble.org/PyNN) were created independently as solutions to the issue of the multiple incompatible formats used by neuronal simulators for specifying cell and network models. They took differing approaches to this problem, with NeuroML creating a declarative format in XML for describing the elements of networks (cells, ion channel, populations, etc.) and PyNN developing a procedural description in Python which could be used to instantiate the same network across multiple simulators." At the bottom, a note says: "Nevertheless, there has been close collaboration between the developers of these formats since the start, with the aim of ensuring interoperability between the formats. There is currently an

[Abrams et al., OSF, 2018](#)

Define principles and best practices for neuroanatomy (infrastructure + practice)

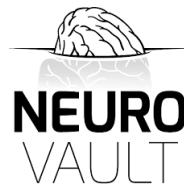
- Neuroscience is an experimental discipline; what we don't know is more than we know
- Have to manage neuroscience atlases, terminologies, etc for anatomical delineations and cell types as computational artefacts (FAIR)
- Community needs guidance on how to do that; do we have enough experience to lay that out?
- Can we provide guidance for researchers producing anatomical data?



Allen Reference Atlas

The central role of distributed repositories

- Neuroscience is characterized by many distributed data repositories, some specialized for data type (community), some for individual projects, e.g. Human Connectome
- Not large government-run databases; more of a cottage industry
- **Repositories are the publishers for data**
- **Repositories are key participants in FAIR**
- **FAIR is necessary for long term sustainability**
- Awareness of FAIR is very minimal
- Workforce development: INCF will provide a forum for those running neuroscience infrastructures **but is NOT assessing FAIRness at this time**



Assessment of Biomedical Repositories

- Working through dkNET: NIDDK Information Network ([dkNET.org](https://dknet.org)) to assess the state of biomedical repositories in digestive, diabetes and kidney disease: Open, FAIR, Citable and Trustworthy
- Look for obvious things like use of PID's (including ORCIDs), clear licenses, Core Trust Seal, Open licenses and consistent licenses across data sets, persistence policies, future friendly formats
- But also look for FAIR potential:
 - Landing page (even if it does not have a resolvable PID)
 - Structured metadata even if it isn't machine readable metadata
 - Help for researchers in applying any standard implemented
- Also look at principles of open infrastructures (Bilder et al., 2015) and Core Trust Seal
 - Stakeholder driven
 - Code maintained in GitHub
 - Governance is clear

FAIR Partnership



- Good data management
 - Rich metadata
 - Prepare to share
 - Open formats
 - Adopt/align to standards
 - Submit to repository
- Persistent identifier
 - Machine based access
 - Clear license
 - Support for open, domain specific standards
 - Machine readable metadata
 - Future friendly formats
 - Persistent metadata
 - Bidirectional links
 - Data citation
- Index
 - Effective Search
 - Persistent metadata

Recommendations

- Recognize that it isn't just researchers who need training, but those that build and maintain infrastructure for biomedical research
 - Research infrastructure is different than commercial or library-centered approaches but we can all learn from and help each other
 - Organizations like INCF, RDA, FORCE11 provide the opportunity for SUSTAINED interactions, not just one-off workshops
- Find funding mechanisms to support these organizations
- Assessments and metrics can backfire big time-don't rush into them before a community has come together around what is FAIR for their domain
- Concentrate on understanding FAIR potential that works **within resource constraints**: how easy is it for data and infrastructures to be made FAIR once we understand what that is. What are the best practices that allow this to happen?