



Global Forecast: Cloudy yet FAIR

Maryann E. Martone, Ph. D.

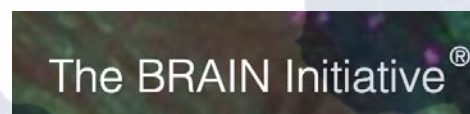
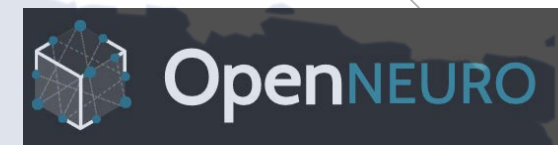
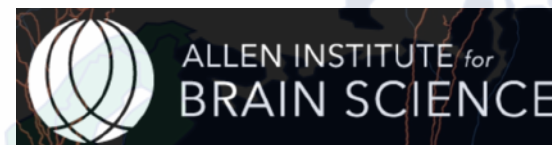
International Neuroinformatics Coordinating Facility

University of California San Diego

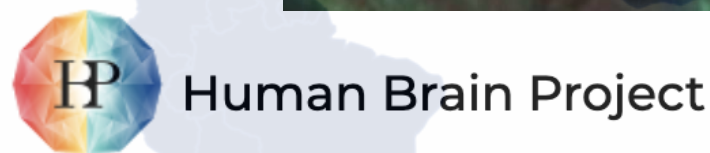
CSO, SciCrunch, Inc*

*SciCrunch Inc is a tech start up developing tools and services around Research Resource Identifiers (RRIDs)

Exciting time for global neuroscience



China Brain Project



About

This registry exists to help people discover and share datasets that are available via AWS resources. [Learn more about sharing data on AWS.](#)

See [all usage examples for datasets listed in this registry.](#)

See datasets from [Facebook Data for Good](#), [NOAA Big Data Project](#), and [Space Telescope Science Institute](#).

OpenNeuro

biology imaging neuro imaging neurobiology

OpenNeuro is a database of openly-available brain imaging data. The data are shared according to a Creative Commons CC0 license, providing a broad range of brain imaging data to researchers and citizen scientists alike. The database primarily focuses on functional magnetic resonance imaging (fMRI) data, but also includes other imaging modalities including structural and diffusion MRI, electroencephalography (EEG), and magnetoencephalography (MEG). OpenfMRI is a project of the [Center for Reproducible Neuroscience at Stanford University](#). Development of the OpenNeuro resource has been funded by the National Science Foundation, National Institute of Mental Health, National Institute on Drug Abuse, and the Laura and John Arnold Foundation.

[Details →](#)

The Human Connectome Project

life sciences neuro imaging

The Human Connectome Project aims to provide an unparalleled compilation of neural data, an interface to graphically navigate this data and the opportunity to achieve never before realized conclusions about the living human brain.

[Details →](#)

Allen Cell Imaging Collections

biology cell imaging image processing machine learning microscopy

This bucket contains multiple datasets (as Quilt packages) created by the Allen Institute for Cell Science (AICS). The imaging data in this bucket contains either of the following: 1) field of view images from glass plates 2) cell membrane, DNA, and structure segmentations 3) cell membrane, DNA and structure contours 4) machine learning imaging predictions of the previously listed modalities. In addition, many of the datasets include CSVs that contain feature sets related to that data.

[Details →](#)

Usage examples

- [Visual Guide to Human Cells](#) by [Allen Institute for Cell Science](#)
- [AICS ImageIO](#) by [Matthew Bowden](#), [Jackson Brown](#), [Jamie Sherman](#), [Dan Toloudis](#)
- [Allen Cell Structure Segmenter](#) by [Jianxu Chen](#), [Liya Ding](#), [Matheus P. Viana](#), [Melissa C. Hendershott](#), [Ruian Yang](#), [Irina A. Mueller](#), [Susanne M. Rafelski](#)
- [AICS Volume Viewer](#) by [Dan Toloudis](#)
- [Download and train label-free models](#) by [Greg Johnson](#)

[See 10 usage examples →](#)

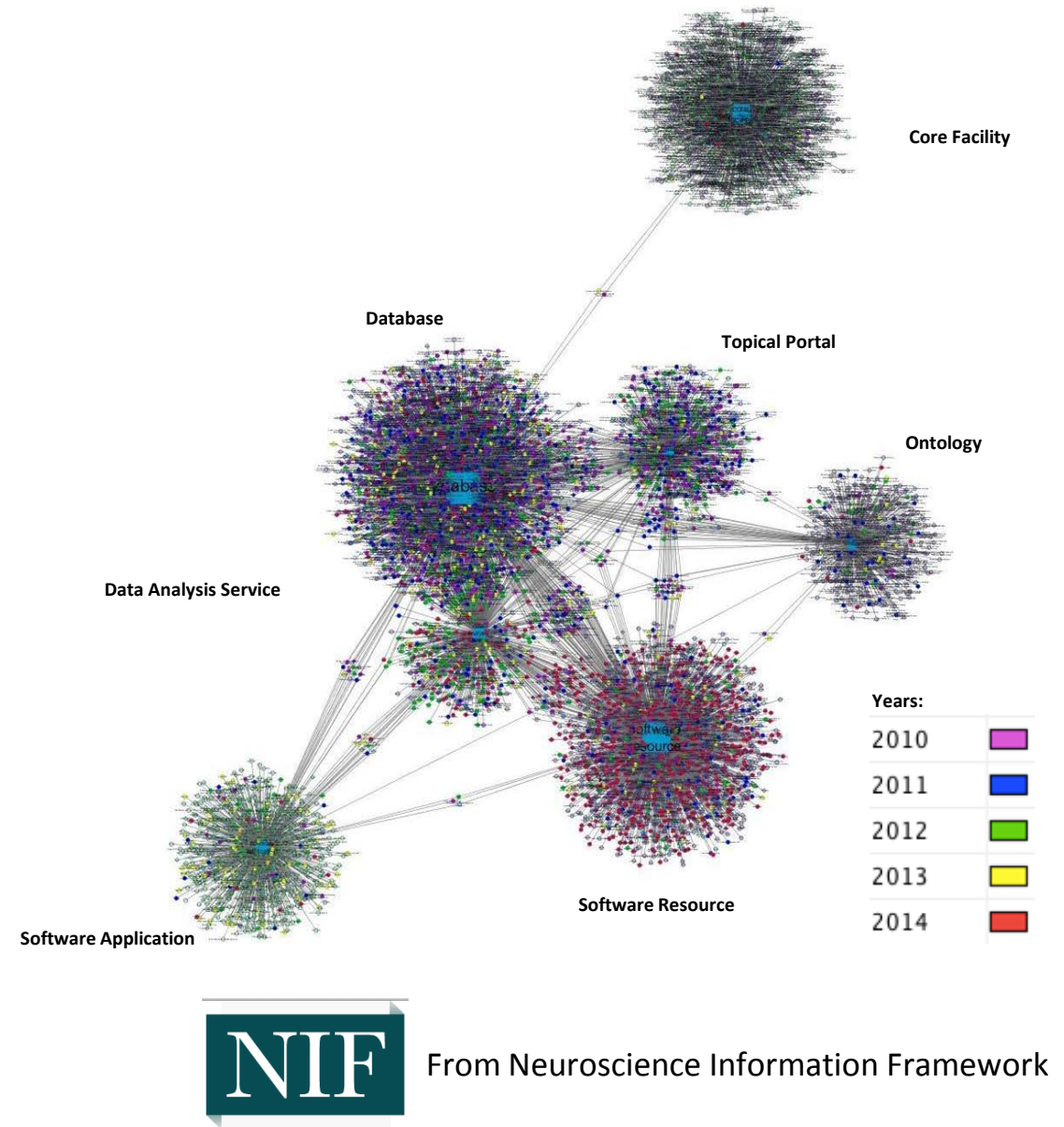
Allen Brain Observatory - Visual Coding AWS Public Data Set

image processing life sciences machine learning neuro imaging neurobiology

The Allen Brain Observatory – Visual Coding is the first standardized in vivo survey of physiological activity in the mouse visual cortex, featuring representations of visually evoked calcium responses from GCaMP6-expressing neurons in selected cortical layers,

In other words...

- Different platforms, same situation
- Neuroscience is served by multiple repositories (archives) representing data across scales modalities, species and conditions
- No single central repository or platform
- How do we share data so that they are maximally useful?



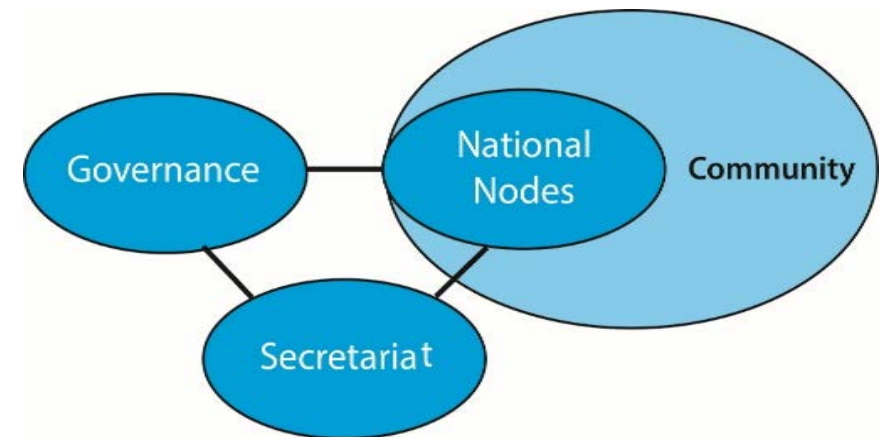
The International Neuroinformatics Coordinating Facility

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”

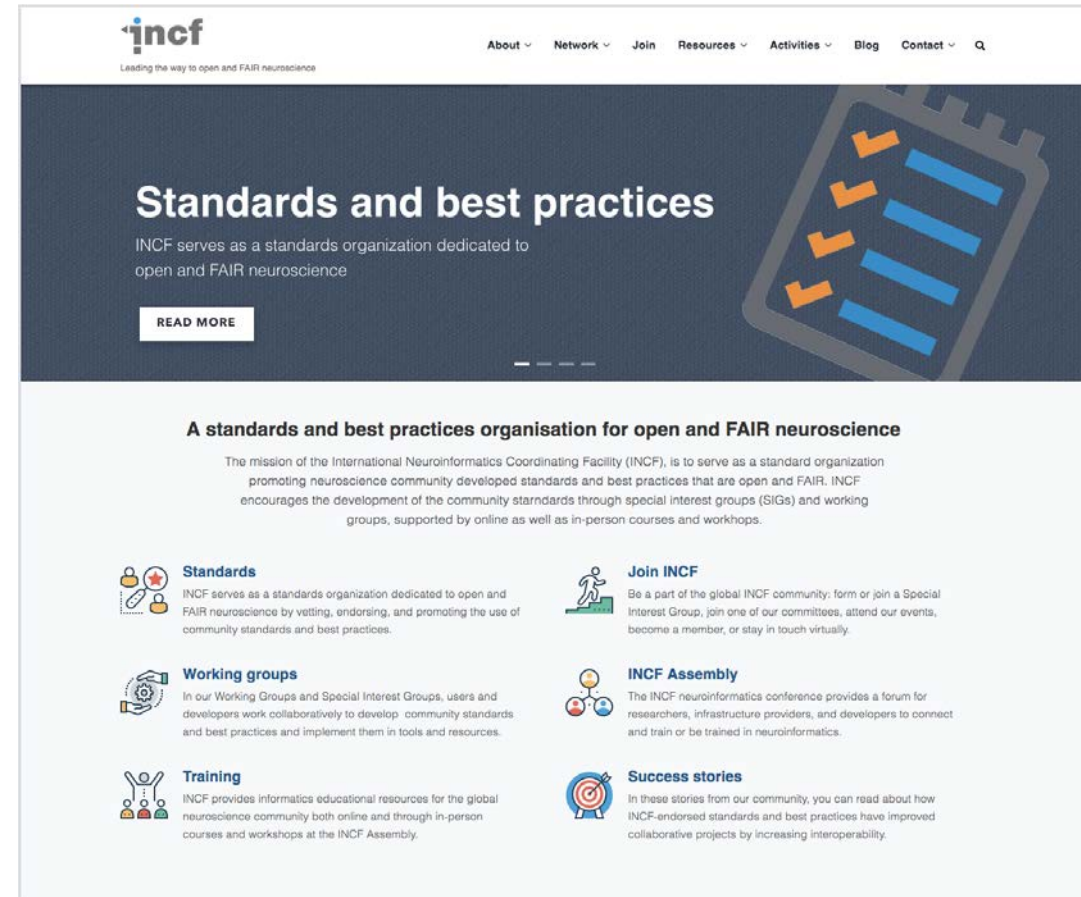
Current Nodes*

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



The INCF network

- supports the development and endorsement of open and FAIR community **standards and best practices (SBPs)** for neuroscience
- leads the development and provision of **training and educational resources** in neuroinformatics
- serves as an interface between the international large-scale brain projects (e.g. BrainMINDS, HBP, CONP, Allen Inst for Brain Science)
- partners with **international stakeholders** to promote and prioritize neuroinformatics at global, national and local levels
- engages scientific, clinical, technical, industry, and funding partners in **collaborative, community-driven projects**



Who is INCF for?

- Neuroinformaticians
- Neuroscience researchers who need to manage, share and analyze their data
- Those building and maintaining infrastructures
- Funders who are trying to implement and encourage open and FAIR neuroscience



INCF is expanding the network with new membership models:
Corporate, Institutions and Organizations and individual

The FAIR Guiding Principles for scientific data management and stewardship

High level principles to make data:



- Findable
- Accessible
- Interoperable
- Re-usable



...for humans *and* machines

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

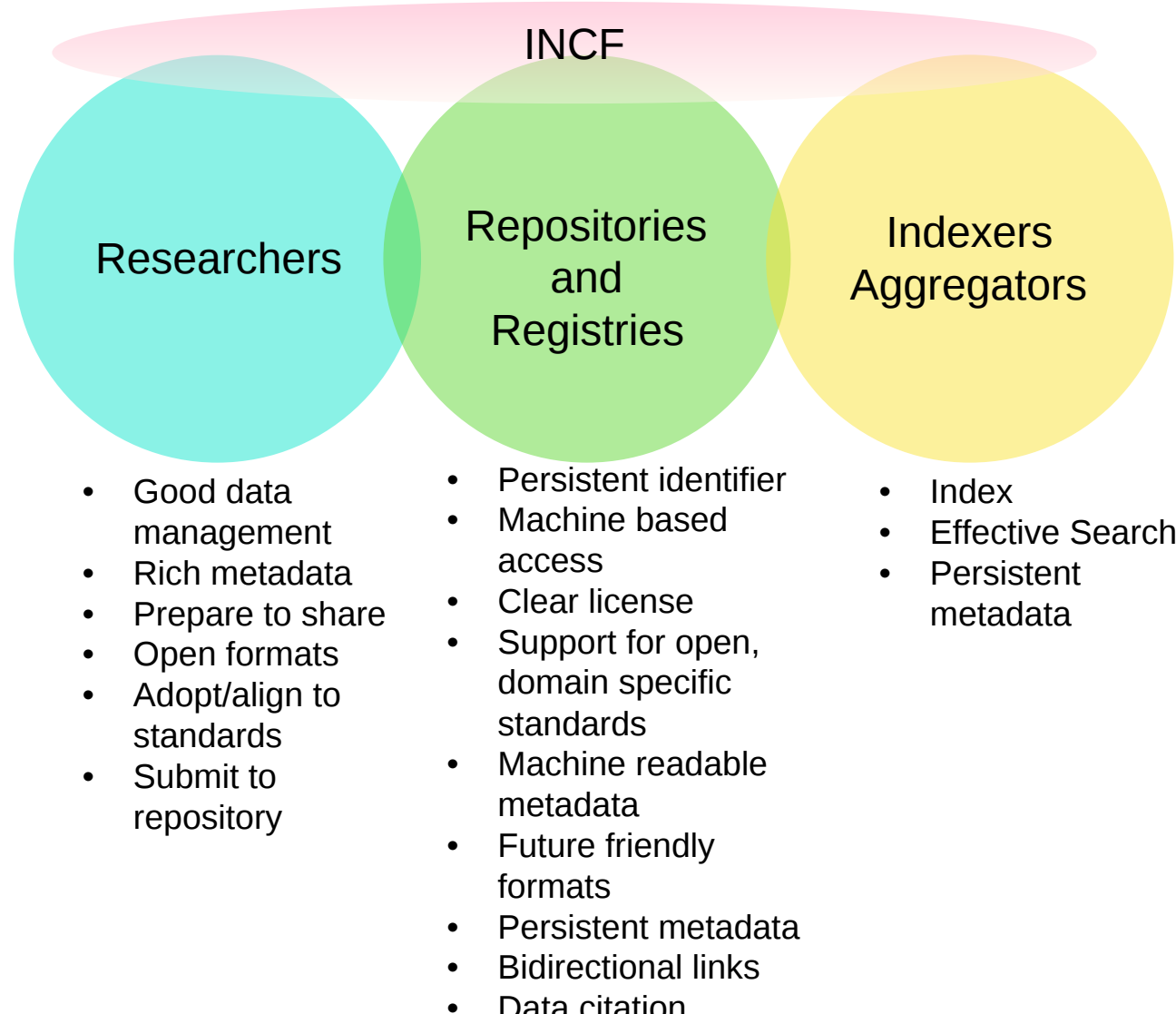
Accessible

- 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*

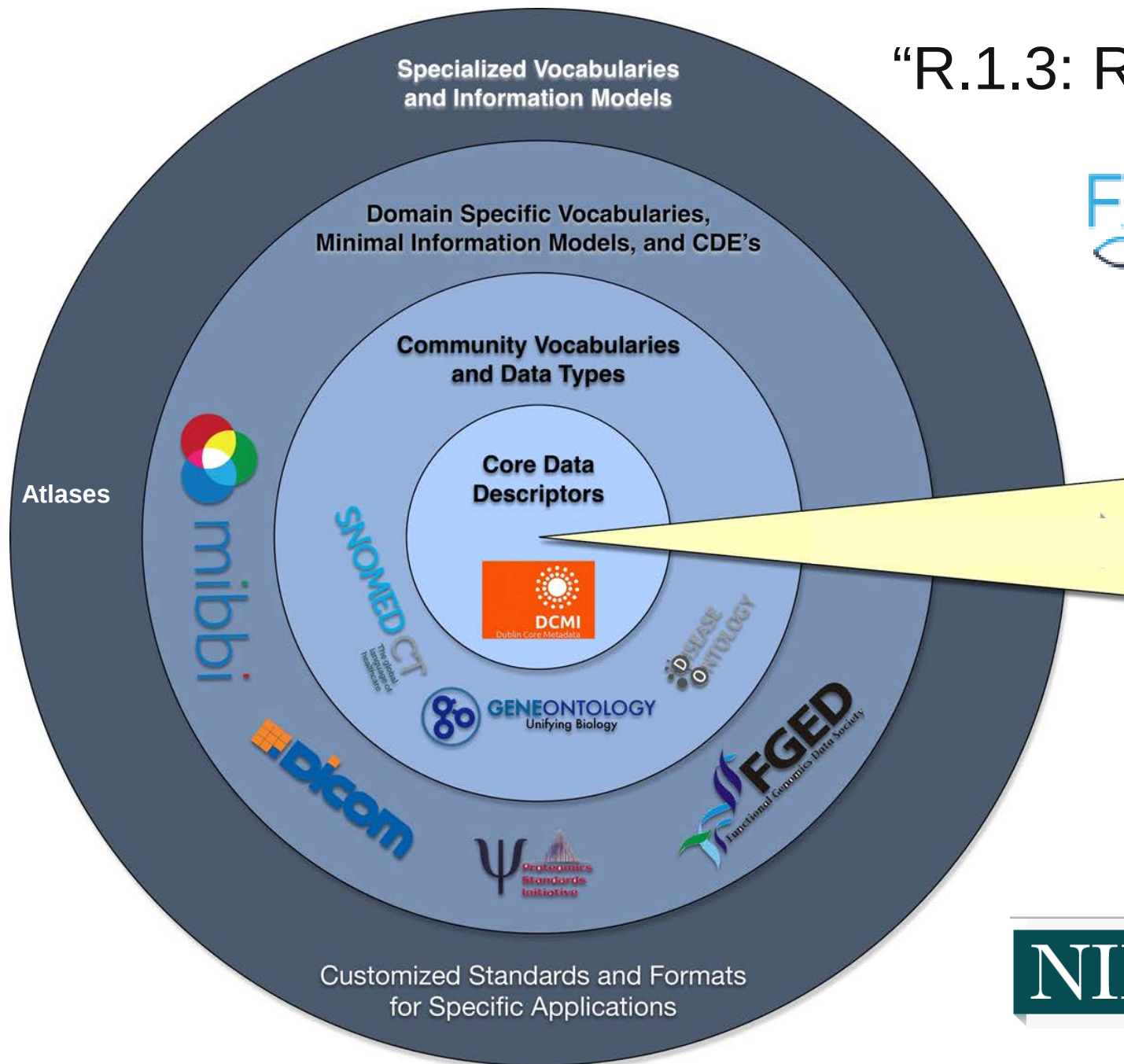
Re-usable

- **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**

FAIR Partnership



“R.1.3: Relevant community standards”

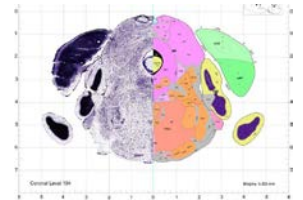


FAIRsharing.org
standards, databases, policies

FORCE11
The Future of Research Communications and e-Scholarship

RDA
RESEARCH DATA ALLIANCE

incf

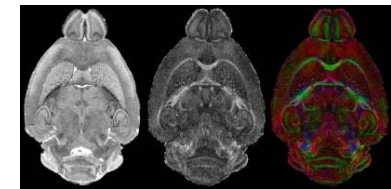


NIF

NEURODATA
WITHOUT BORDERS

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



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
- Ensure neuroscience is supported by robust, harmonious standards

The screenshot displays the BIDS website, which is endorsed by the INCF. The endorsement is highlighted by a blue circular badge with the INCF logo and the word "ENDORSED". The website content includes a header with "Home · SBP · BIDS", a main section titled "BIDS" with a brain image and the text "BRAIN IMAGING DATA STRUCTURE", and a table of metadata. The metadata table lists the following information:

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
Tags	Neuroimaging Dataformat

Below the metadata 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."

The bottom section of the screenshot is titled "NeuroML and PyNN" and contains the following text:

Which format is the right one for a user who wants a simulator-independent description of their model?

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.

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

TrainingSpace: facilitating the use and uptake of standards



Where To Start

4 study tracks that provide a general overview of neuroinformatics, brain medicine, computational neuroscience, and neurobiology for those new to the fields.

Full Courses

Watch lectures from the INCF short course that provides an introduction to neuroinformatics by leading experts in the field.

Tools And Resources

Here you will find resources to help you incorporate neuroinformatics approaches in your classroom as well as links to repositories offering open neuroscience data and models.

Featured courses

Computational neuroscience: the basics
INCF

Introduction to modeling the brain.

- Neuron anatomy and signaling, and different types of models
- Dynamic basis of the action potential
- Modelling of chemical computation in the brain
- Abstract models of neurons.
- Forms of plasticity on many levels
- Synaptic plasticity on the molecular level

[VIEW THE COURSE](#)

Browse Our Top Courses

COMPUTATIONAL NEUROSCIENCE | NEUROINFORMATICS | BRAIN MEDICINE | NEUROBIOLOGY | DATA SCIENCE

Biophysical models
INCF
[VIEW THE COURSE](#)

Biochemical models
INCF
[VIEW THE COURSE](#)

Statistical models
COSYNE
[VIEW THE COURSE](#)

Computational neuroscience: the basics
INCF
[VIEW THE COURSE](#)

Dynamical neural systems
ICTS
[VIEW THE COURSE](#)

Cajal Course in Computational Neuroscience
HBP Education Programme
[VIEW THE COURSE](#)

Introduction to computational neuroscience
INCF
[VIEW THE COURSE](#)

Building the Brain
INCF
[VIEW THE COURSE](#)

[View all the courses >>](#)

Browse Courses on Tools and Resources

TOOLS

OpenNeuro.org Tutorials
OpenNeuro.org
[VIEW THE COURSE](#)

Jupyter Notebook
EuroPython Conference
[VIEW THE COURSE](#)

Open Science Framework (OSF)
Center for Open Science
[VIEW THE COURSE](#)

Hypothes.is
Hypothes.is: a web-based annotation tool
[VIEW THE COURSE](#)

TECHNICAL TOUR
How to use Allen Institute for Brain Science resources
Allen Institute for Brain Science
[VIEW THE COURSE](#)

TrainingSpace provides informatics educational resources for the global neuroscience community, and is continuously updated with materials from INCF and partner organizations.

TrainingSpace is supported by KnowledgeSpace, a community encyclopedia linking brain research concepts to data, models, and literature

KnowledgeSpace

A community encyclopedia linking brain research concepts to data, models, and literature.

Search in KnowledgeSpace

Have any questions regarding TrainingSpace?

[DISCUSS IN NEUROSTARS](#)

Brain Summits

Purpose: Bring together infrastructure providers and developers working on the large international brain initiatives to learn from each other and identify opportunities for harmonization and coordination

Brain Summit 2020: Planning underway to possibly hold in conjunction with the US BRAIN Initiative 2020 annual meeting, June, Washington DC

1. Introduce the INCF SBP process and the FAIR data principles to US BRAIN Initiative
2. Highlight US infrastructure supporting BRAIN and other large consortia
3. Practical issues (lectures and hands on): what every infrastructure provider needs to know: trends in biomedical infrastructure (e.g., machine readable landing pages to support data citation, designing for Google data set search, choosing appropriate licenses, how to conduct agile user testing)
4. User forums to hear from end users: researchers (data management and use of standards) and computational scientists (likely end users)
5. Engage industry representatives to address issues around SBP implementation
6. Identify and support opportunities for harmonization

So is the cloud the answer to everything?

- Resources (data, tools, etc.) in the cloud are not guaranteed to be available indefinitely and require continued funding and support. ***Therefore, the cloud should not be seen as a replacement for contributing data to a proper data repository or archive.***
- Data download costs (“egress fees”) from the cloud can become very expensive for large data volumes...
- Cost and capacity of commercial provider services and tools are very different from local, on-premise services and tools.
- The Academy should proceed cautiously with ceding too much of our data to commercial providers without protections
- ***Data uploaded into the cloud are not automatically findable, accessible, interoperable, and reusable (FAIR). Effort is required to make them FAIR prior to upload, and doing so will allow other researchers to more easily use the data.***

