



Vers des données FAIR en neurosciences

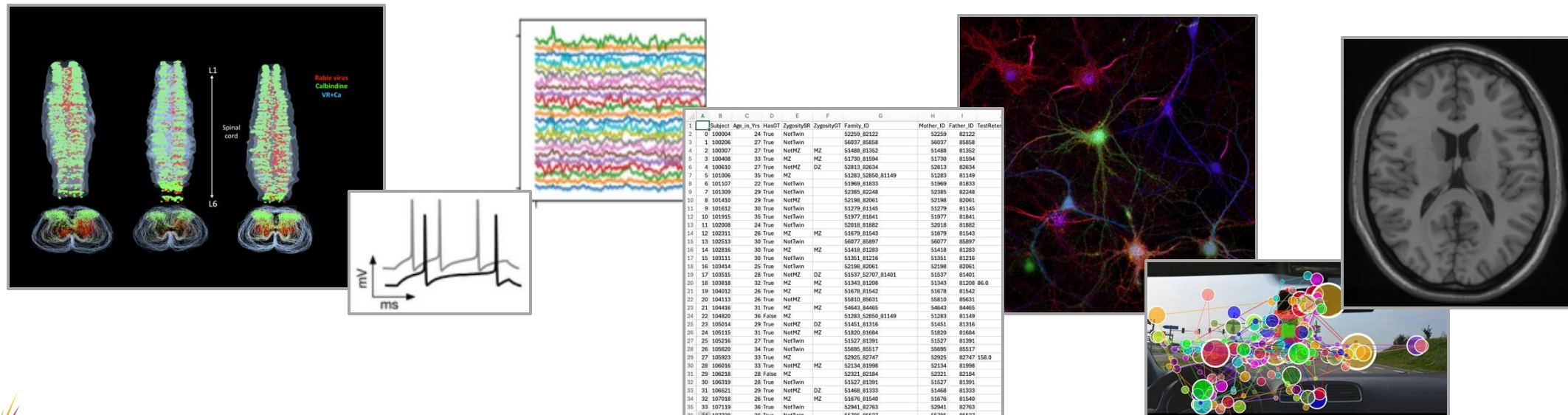
Olivier Coulon,
Institut de Neurosciences de la Timone,
Aix-Marseille Université, CNRS



L'Institut de Neurosciences de la Timone

L'INT est un institut de recherche en neurosciences fondamentales et cliniques, du niveau moléculaire au niveau cognitif.

~200 personnes, 13 équipes, 3 grosse plateformes d'acquisition de données



Beaucoup de données

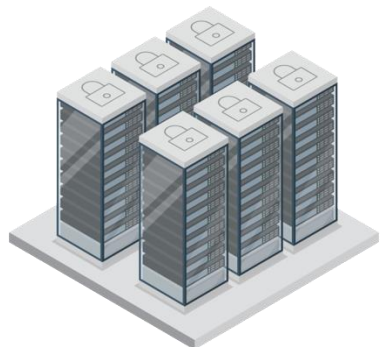
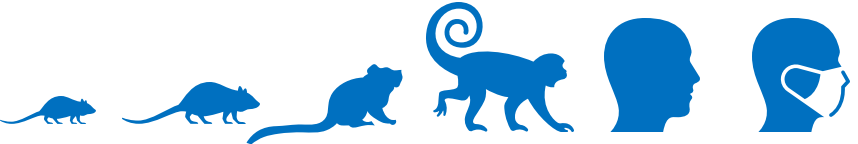
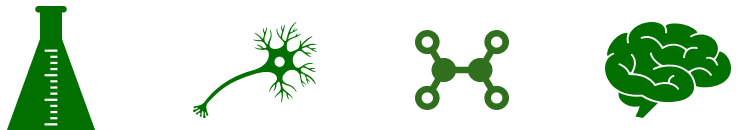
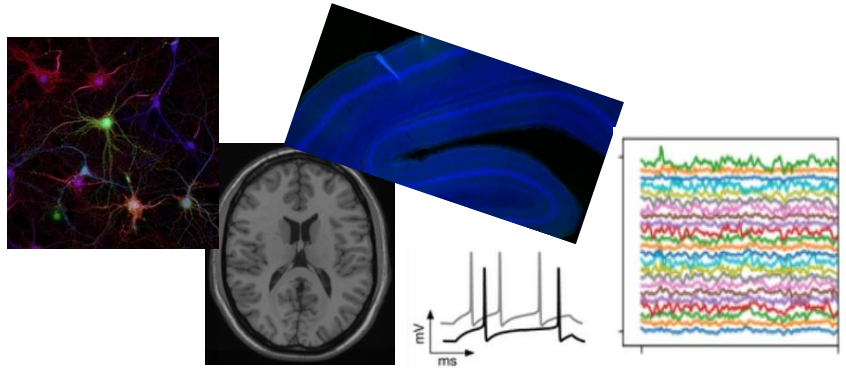
Des données...

... hétérogènes

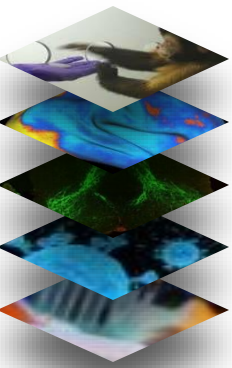
... à de multiples échelles

... sur des modèles variés

... en grandes quantités



Plateformes



Clinique



Grandes bases internationales



MRI neuroimaging

- Anatomical
- Diffusion
- Functional
- Quantitative
- ...

Optical Imaging

- Bi-photon microscopy
- Confocal microscopy
- Mesoscopic optical imaging
- Spectroscopy
- Laser doppler flowmetry
- Optical coherence tomography
- Histology / tracing
- Light sheet imaging

Electrophysiology

- EEG/MEG
- Multi-electrodes array
- Single cell recordings
- Deep brain stimulation recordings

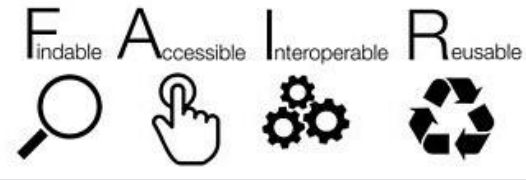
Molecular

- Genomics
- Transcriptomics

Behavioral

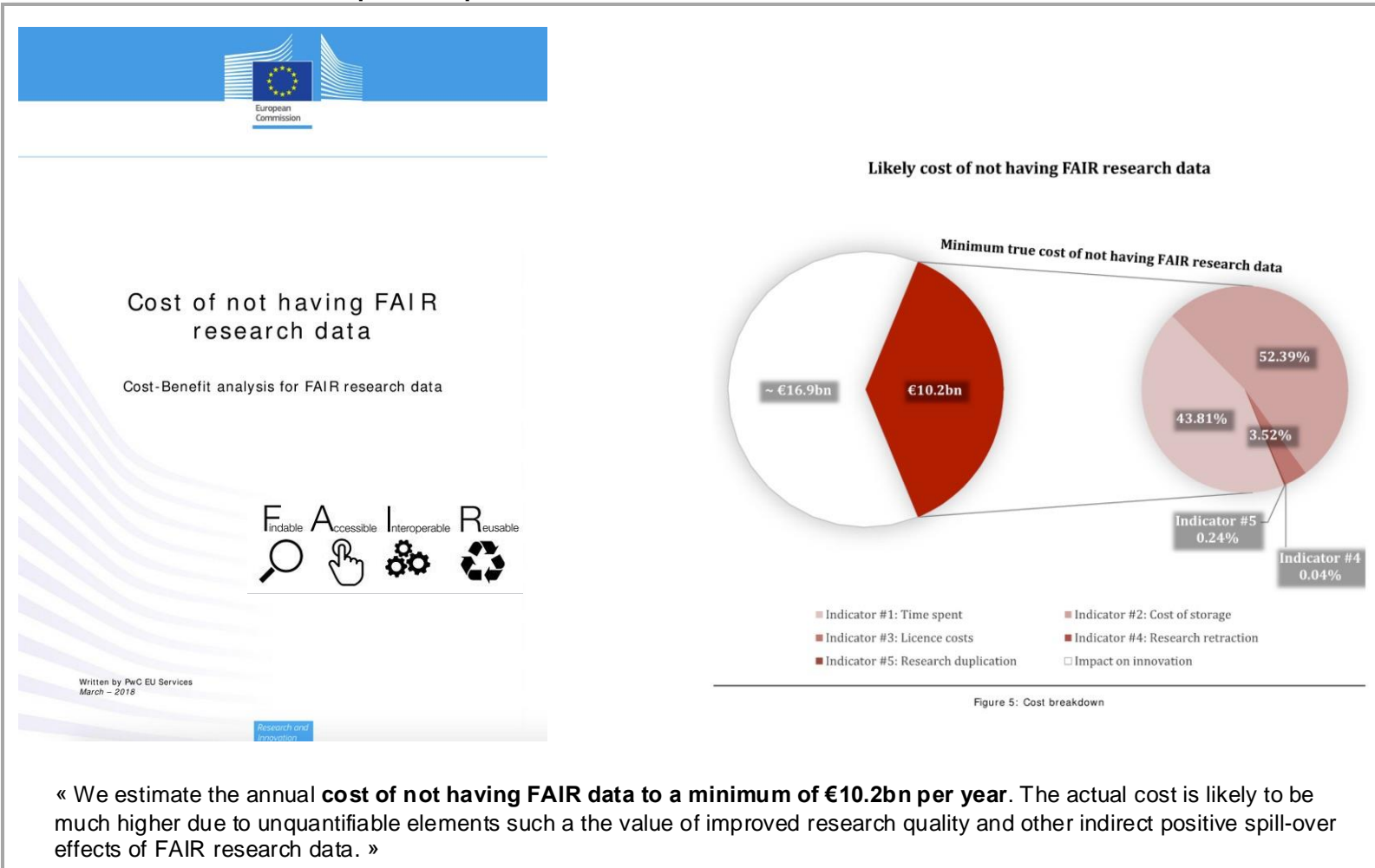
- Eye tracking
- ...

Le formalisme FAIR



Facile à trouver

- Identifiant unique et pérenne



identifiant unique.

accès, conditionnelle, ou payante. utilisation, en particulier

Stockage: où sont mes données ?

« Sur un disque dur portable. Mon doctorant l'avait, je lui envoie un mail »

Stockage non sûr, non sécurisé, non sauvegardé.

Risque majeur: perte complète des données
Autre risque: perte de méta données, et impossibilité de réutilisation.



« Sur une station de travail dans la salle expérimentale. De temps en temps j'en fais une copie sur un disque dur »

Stockage non sûr, non sécurisé, sauvegardes aléatoires.

Risque majeur: perte complète des données
Autre risque: perte de méta données, et impossibilité de réutilisation.



« Sur un serveur de stockage (professionnel) »

Stockage sûr et sécurisé, sauvegardes garanties.

Peut-on trouver les données et les utiliser ?



Structuration : pourquoi faire ?

Pouvoir retrouver les données



Métadonnées descriptives

Pouvoir réutiliser les données



Une organisation standardisée

Des métadonnées contenant l'ensemble des détails expérimentaux nécessaires.

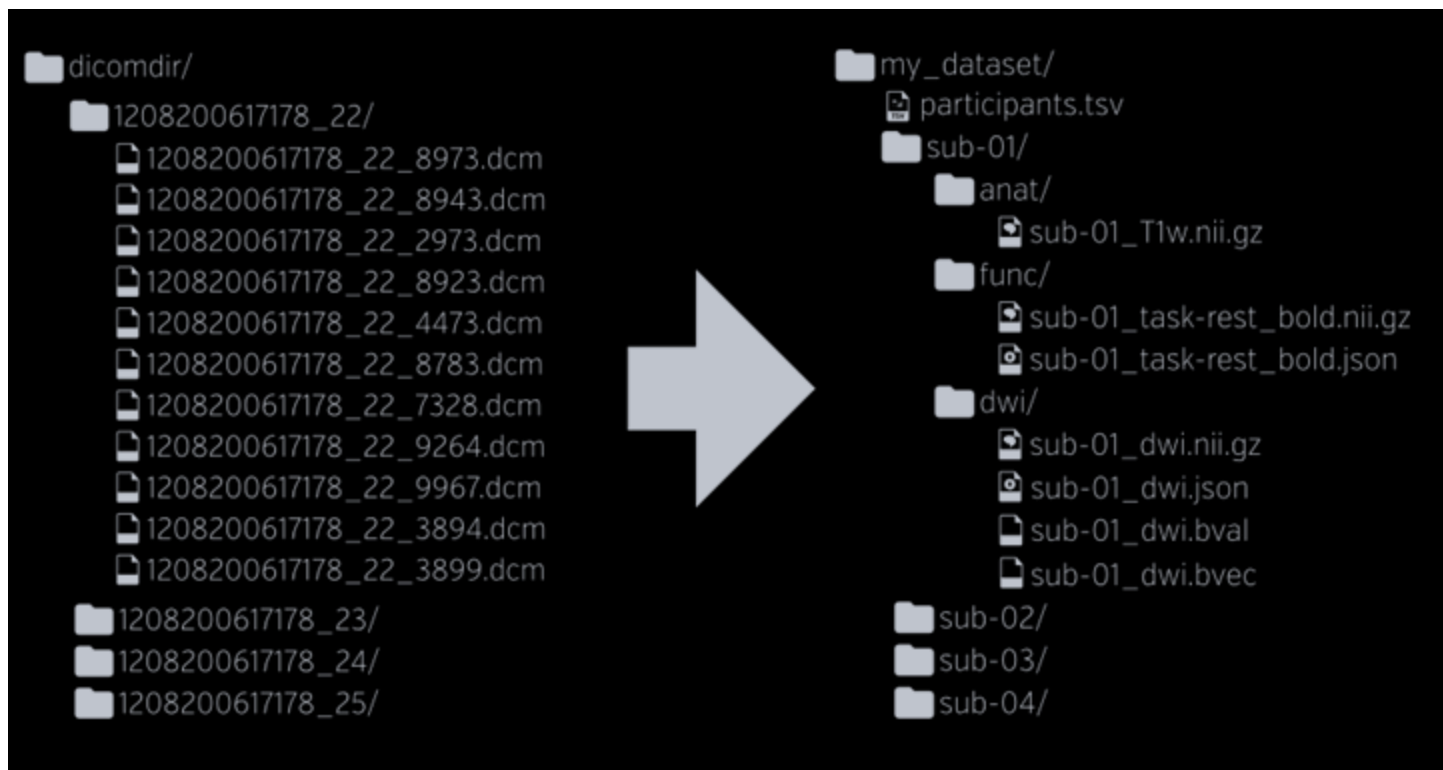
Des fichiers lisibles dans des formats standards

Une structure de données dédiée



- Une structure de donnée standard pour les données cérébrales
- Une initiative libre et communautaire
- Un écosystème croissant: 'validator', intégration dans des bases de données, BIDS Apps
- Un choix de modalités croissant: IRM, MEG, EEG, IEEG, PET, données physiologiques (respiration, activité cardiaque,...), données comportementales, microscopie, NIRS
- BIDS :
 - Organise les données dans une hiérarchie de répertoires
 - Nomme les fichiers de manière standard
 - Documente les metadonnées
 - Facilite la réutilisation pour vous et pour les autres

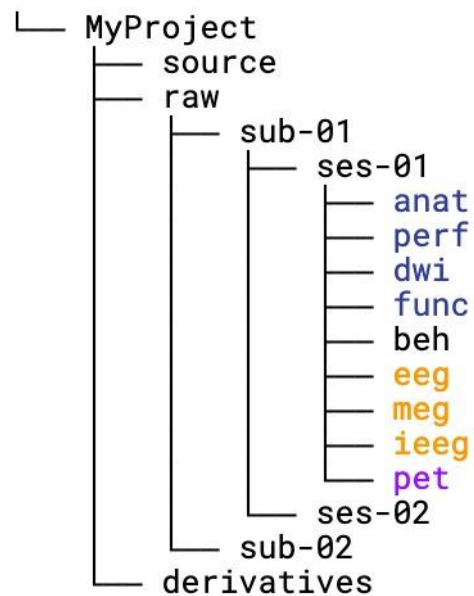
Une structure de données dédiée



Une structure de données dédiée



- BIDS définit une structure de données



- BIDS définit une nomenclature de noms de fichiers



BIDS filenames:

- **suffix** preceded by an **underscore**
- **entity-label** pairs separated by **underscores**
- **Entities**, **labels**, **suffixes** can only contain letters and / or numbers.
- For a given **suffix**, some **entities** are required and some others are [optional].
- **Entity-label** pairs have a specific order in which they must appear in filename.

Une structure de données dédiée

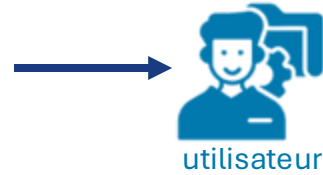
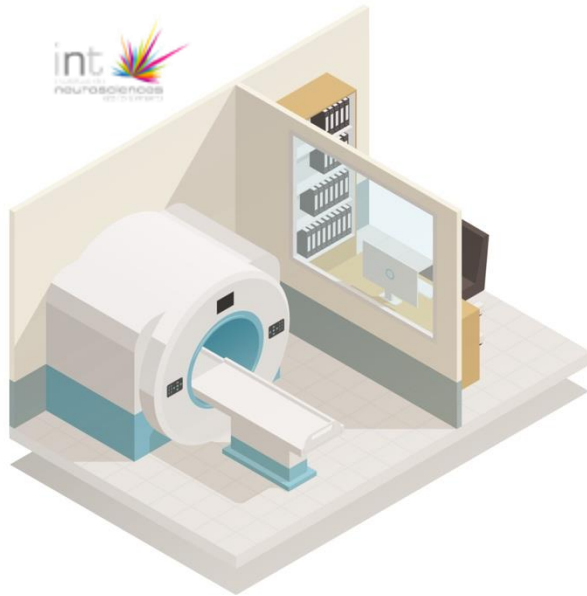


- BIDS standardise et gère les métadonnées

```
sub-01
├── ses-01
│   ├── anat
│   │   ├── sub-01_ses-01_acq-tse_T2w.nii
│   │   └── sub-01_ses-01_T1w.nii.gz
│   └── func
│       ├── sub-01_ses-01_task-olfloc_run-01_bold.nii
│       ├── sub-01_ses-01_task-olfloc_run-01_bold.json
│       ├── sub-01_ses-01_task-olfloc_run-01_events.tsv
│       └── sub-01_ses-01_task-olfloc_run-01_physio.tsv.gz
```

- **JSON files:** JavaScript Object Notation
 - for attribute-values pairs
- **TSV files:** Tabulation Separated Values
 - for spreadsheet data

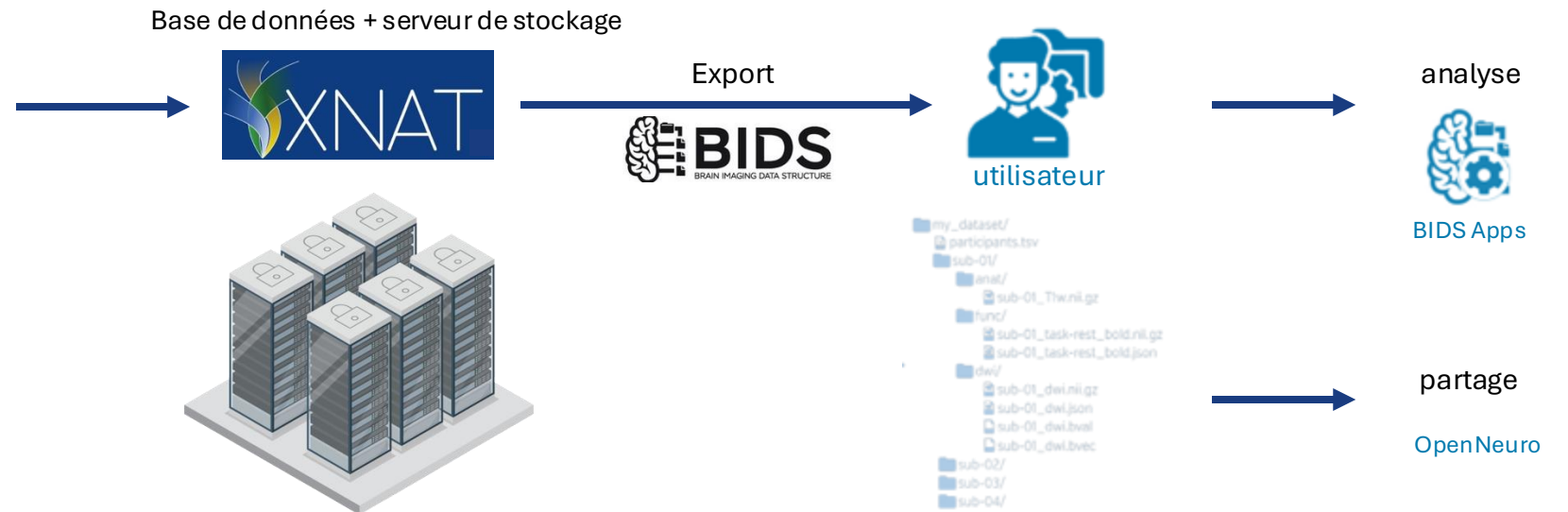
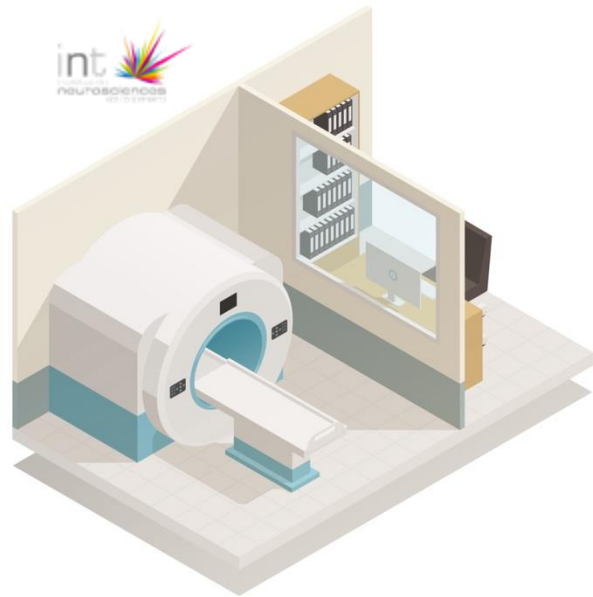
Un cas facile: l'imagerie par résonance magnétique



utilisateur

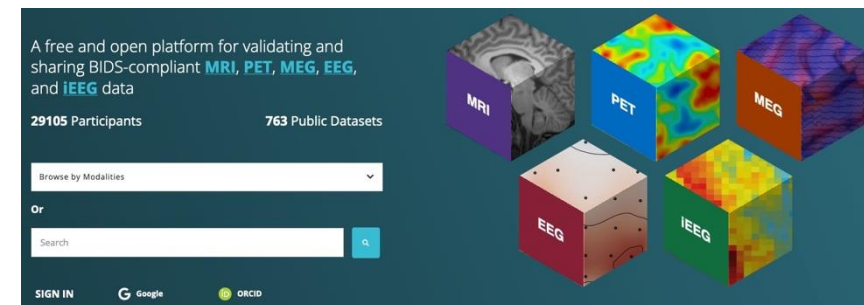
- dicomdi/
 - 1208200617178_22/
 - 1208200617178_22_8973.dcm
 - 1208200617178_22_8943.dcm
 - 1208200617178_22_2973.dcm
 - 1208200617178_22_8923.dcm
 - 1208200617178_22_4473.dcm
 - 1208200617178_22_8783.dcm
 - 1208200617178_22_7328.dcm
 - 1208200617178_22_9264.dcm
 - 1208200617178_22_9967.dcm
 - 1208200617178_22_3894.dcm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

Un cas facile: l'imagerie par résonance magnétique



Le stockage ne passe pas par l'utilisateur

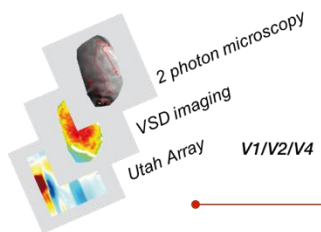
La récupération des données se fait directement en BIDS.



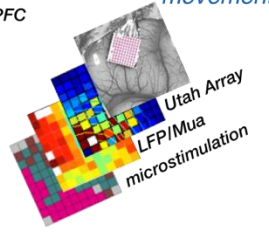
Un cas difficile: l'électrophysiologie du primate

Dynamics of cortical maps for decision, action and perception
 (Teams: CoMCo, NeOpto, InViBe, BanCO)

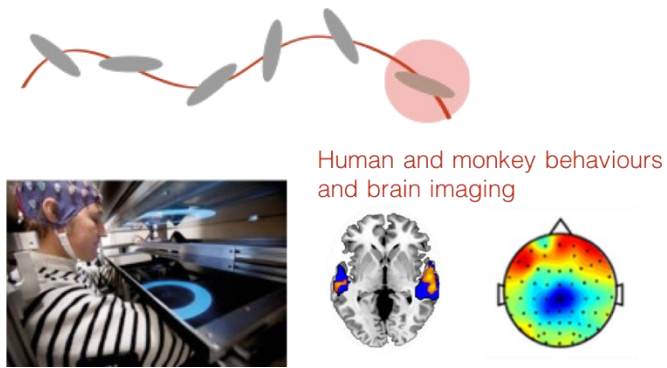
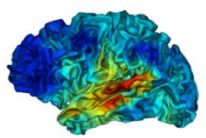
Visual maps for motion computation



Motor maps for hand/eye movements



Auditory maps for voice processing



Neural data in:
 · 2 files
 · 2 formats

Metadata in:
 · ~ 14 files
 · ~ 8 formats

General Specifications

Electrode-Array
 .pdf .xls .txt

Monkey
 .pdf

Hardware
 .pdf

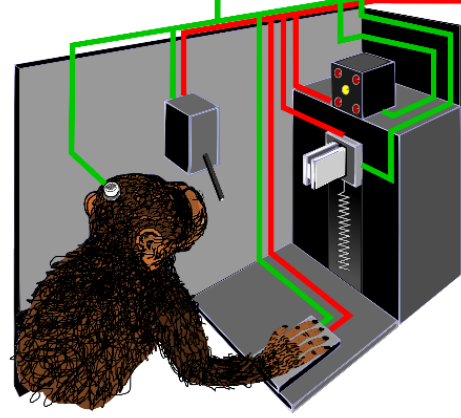
Recordings

Cerebus
 .nev .nsX .txt

Setup & Monkey
 .xls

Task control

LabVIEW
 .html .xls



Preprocessing

Plexon
 .txt .xls .nev

Event Detection
 .nev

Quality Check
 .hdf5

Analysis

(courtesy Thomas Brochier, INT)

Un format BIDS pour l'électrophysiologie animale



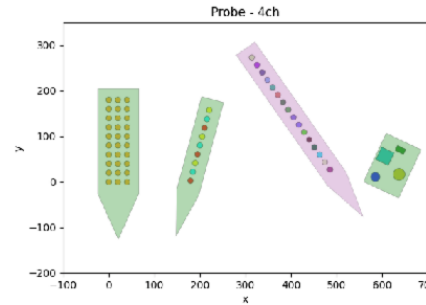
BIDS Extension Proposal 032

<http://bids.neuroimaging.io/bep032>



Contenu:

- Open file format (nix/nwb)
- Description des électrodes et du câblage
- Métadonnées



<https://probeinterface.readthedocs.io>

an example...

```
my_dataset/
├── dataset_description.json
├── participants.json
├── participants.tsv
├── tasks.json
├── tasks.tsv
├── sub-i/
│   ├── sub-i_sessions.json
│   ├── sub-i_sessions.tsv
│   └── ses-140703/
│       └── ephys/
│           ├── sub-i_ses-140703_task-r2g_run-001_channels.tsv
│           ├── sub-i_ses-140703_task-r2g_run-001_contacts.tsv
│           ├── sub-i_ses-140703_task-r2g_run-001_ephys.json
│           ├── sub-i_ses-140703_task-r2g_run-001_ephys.nix
│           └── sub-i_ses-140703_task-r2g_run-001_probes.tsv
├── sub-l/
│   ├── sub-l_sessions.json
│   ├── sub-l_sessions.tsv
│   └── ses-101210/
│       └── ephys/
│           ├── sub-l_ses-101210_task-r2g_run-001_channels.tsv
│           ├── sub-l_ses-101210_task-r2g_run-001_contacts.tsv
│           ├── sub-l_ses-101210_task-r2g_run-001_ephys.json
│           ├── sub-l_ses-101210_task-r2g_run-001_ephys.nix
│           └── sub-l_ses-101210_task-r2g_run-001_probes.tsv
```

Naming of files and directories :

- follows the generic rules of BIDS
- intuitive hierarchy (*project/animal/session/modality*)
- redundancy of information in file and directory names
- added specific infos for electrophysiology

Supported data file format (INCF standards) :

- NIX
- NWB

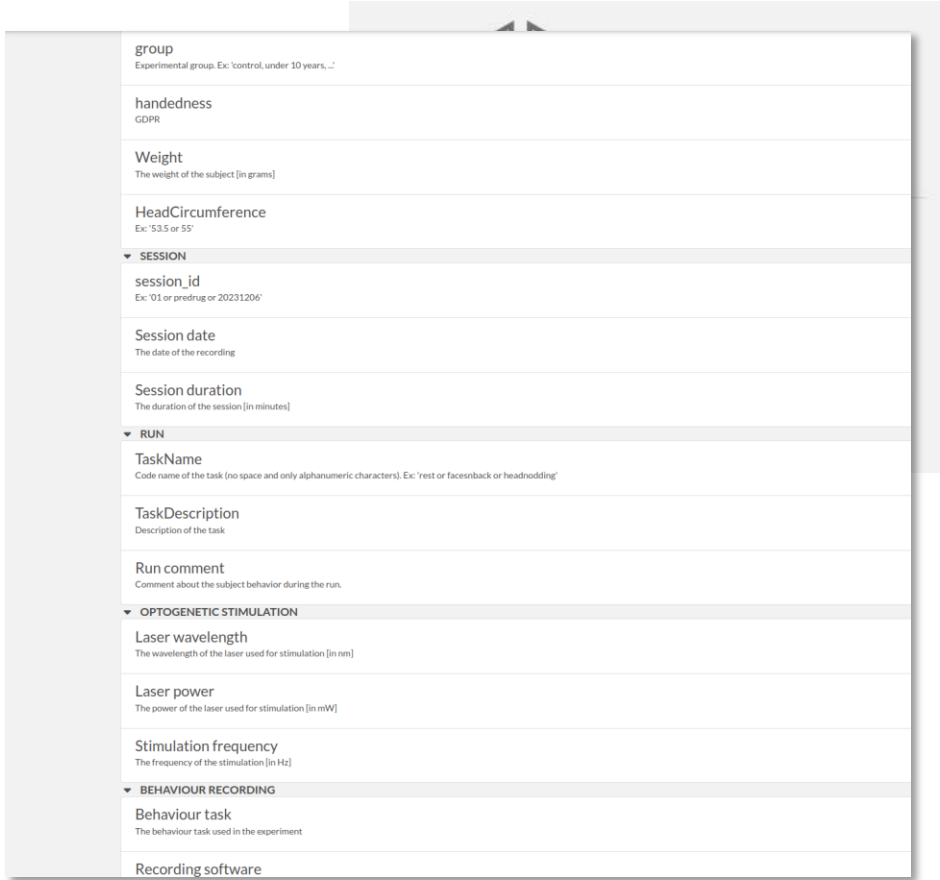
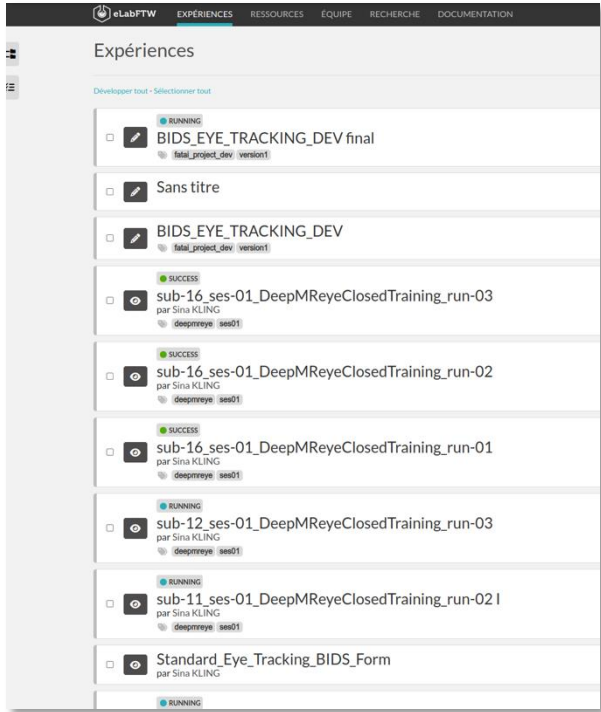
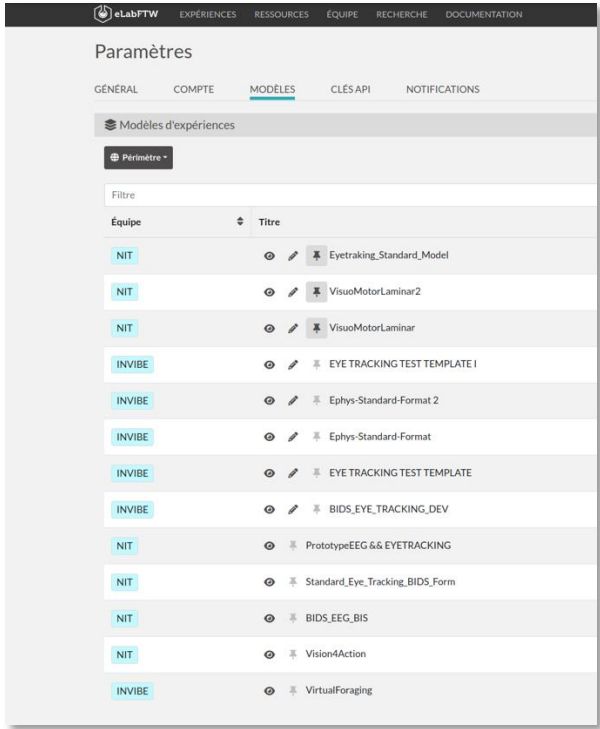


Supported metadata file formats (as in generic BIDS) :

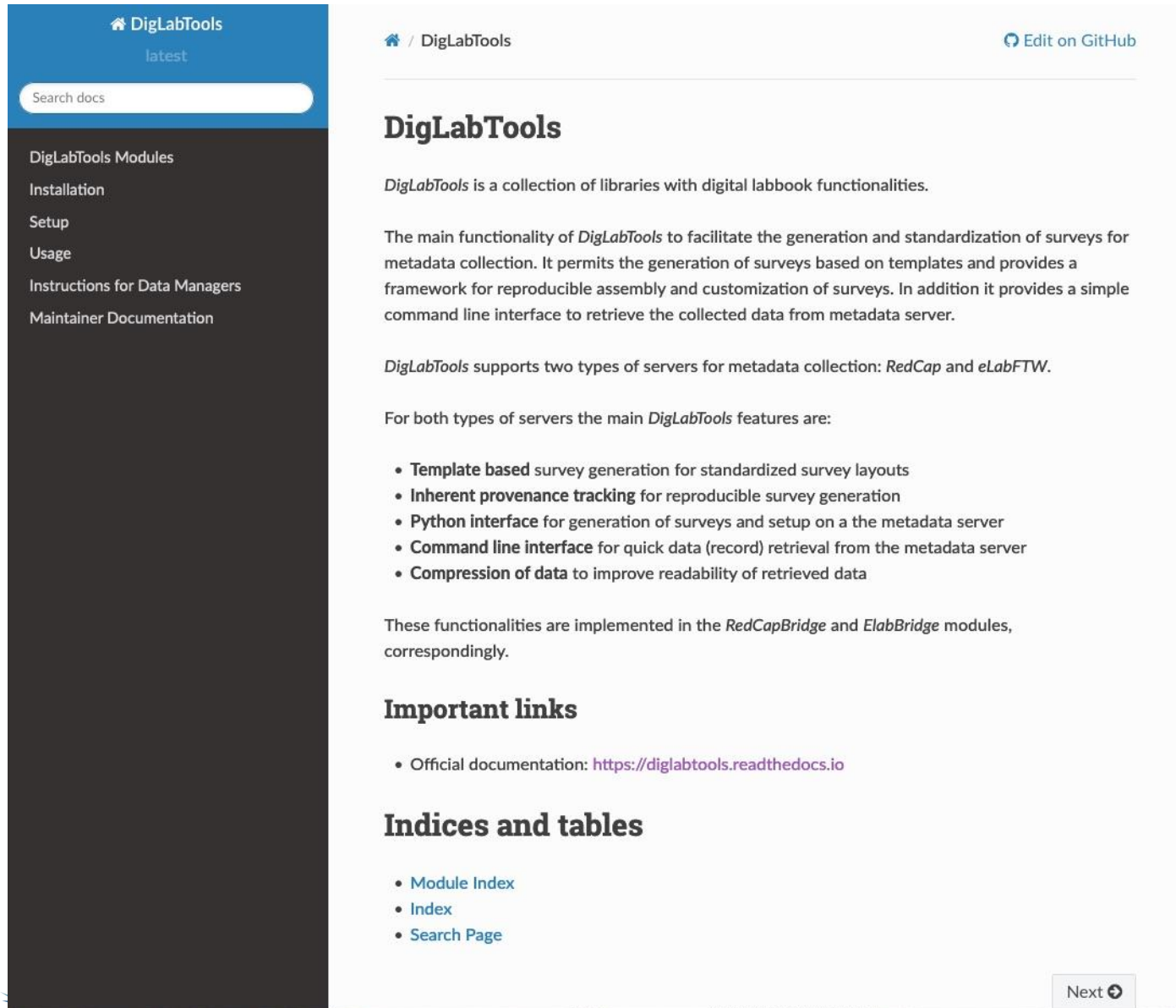
- tsv
- json

La gestion des métadonnées

Choix : utiliser l'arrivée des cahiers de laboratoires électroniques (eLabFTW) pour métadonnées.



Une API pour gérer les formulaires eLab



The screenshot shows the documentation page for DigLabTools. The page has a blue header with the DigLabTools logo and 'latest' version indicator. A search bar is present. The left sidebar lists navigation items: DigLabTools Modules, Installation, Setup, Usage, Instructions for Data Managers, and Maintainer Documentation. The main content area includes the title 'DigLabTools', a description of its purpose as a collection of libraries for digital labbook functionalities, and details about its main functionality for survey generation and metadata collection. It also lists supported servers (RedCap and eLabFTW) and features like template-based survey generation, provenance tracking, Python interface, command line interface, and data compression. Important links and indices are provided at the bottom.

🏠 / DigLabTools [Edit on GitHub](#)

DigLabTools

DigLabTools is a collection of libraries with digital labbook functionalities.

The main functionality of *DigLabTools* to facilitate the generation and standardization of surveys for metadata collection. It permits the generation of surveys based on templates and provides a framework for reproducible assembly and customization of surveys. In addition it provides a simple command line interface to retrieve the collected data from metadata server.

DigLabTools supports two types of servers for metadata collection: *RedCap* and *eLabFTW*.

For both types of servers the main *DigLabTools* features are:

- **Template based** survey generation for standardized survey layouts
- **Inherent provenance tracking** for reproducible survey generation
- **Python interface** for generation of surveys and setup on a the metadata server
- **Command line interface** for quick data (record) retrieval from the metadata server
- **Compression of data** to improve readability of retrieved data

These functionalities are implemented in the *RedCapBridge* and *ElabBridge* modules, correspondingly.

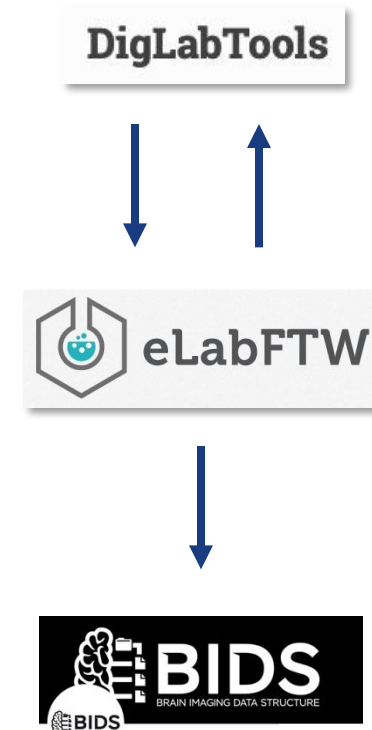
Important links

- Official documentation: <https://diglabtools.readthedocs.io>

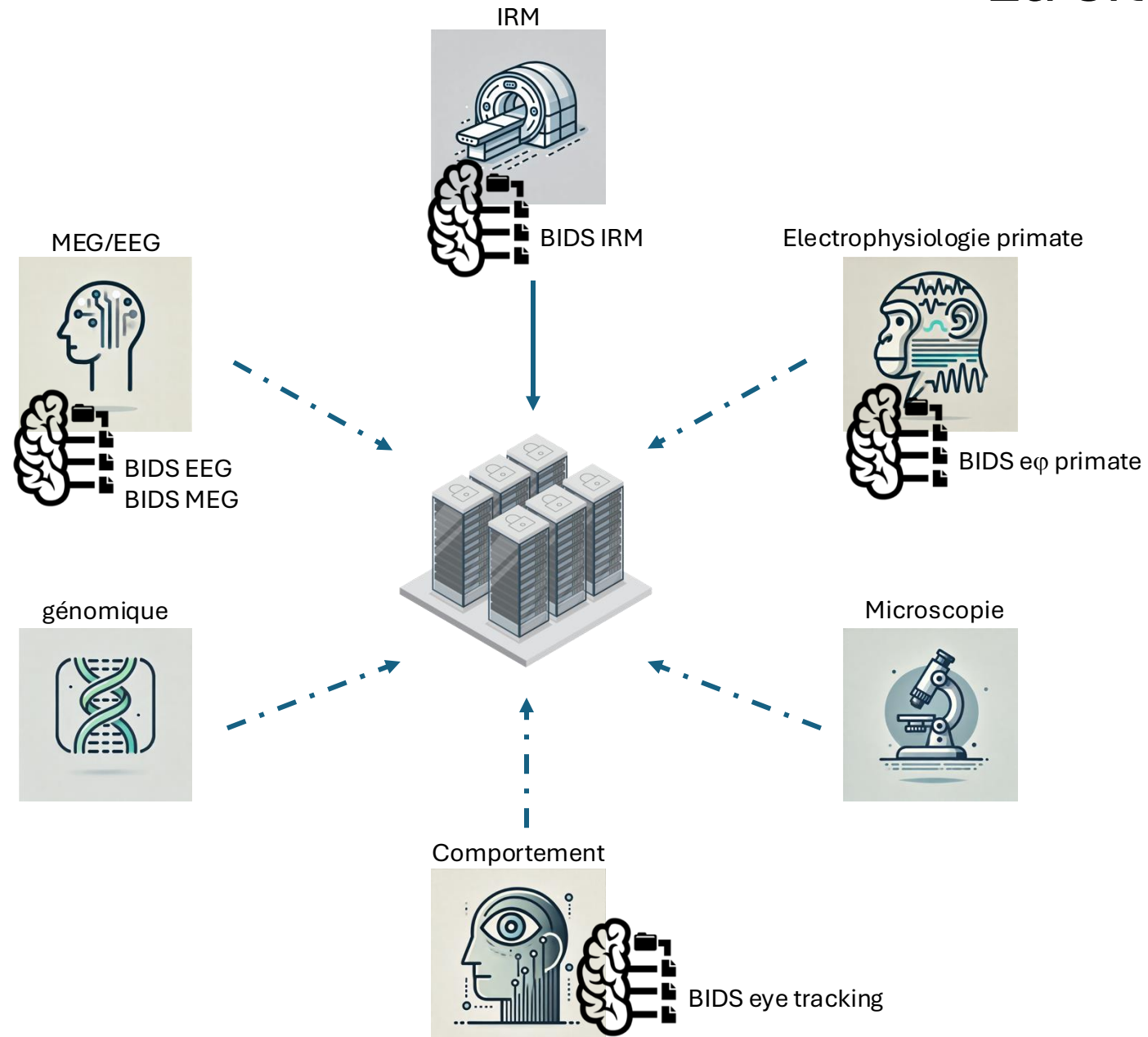
Indices and tables

- [Module Index](#)
- [Index](#)
- [Search Page](#)

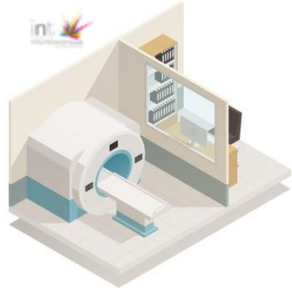
Next ➔



La situation à l'INT



Quelles leçons en tirer ?



Solutions existantes

Un ingénieur + centre IRM

Du cahier des charges à la mise en production: 2 ans.

Il faut du temps

Il faut des ressources humaines et financières



Solutions à créer

ANR ShareElec: 2019-2021

Interactions avec BIDS Microscopy et BIDS eye-tracking

Support de l'INCF

BIDS électrophysiologie animale en cours de validation, opérationnel en 2025

2 ingénieurs, deux alternants + 2 équipes de recherche

Du cahier des charges à la mise en production: 6 ans

Quelles leçons en tirer ?

La définition de formats de données+métadonnées « universels » :

- Nécessite une connaissance et une pratique du domaine de recherche

-> Les unités de recherche doivent être impliquées

- Est par définition une initiative qui concerne une communauté scientifique

-> Les unités de recherche ne doivent pas agir seules



Alors même que cela tend à devenir une obligation réglementaire, le coût et la charge des projets de gestion de données revient souvent (au moins en partie) aux unités de recherche, y compris lorsqu'il s'agit d'une initiative institutionnelle (e.g. eLabFTW).

-> Les unités de recherche doivent être aidées financièrement, humainement, et parfois techniquement (e.g. le CEDRE, centre de formation et de soutien aux données de la recherche)

Sylvain Takerkart

Julia Sprenger

Killian Rochet

Fataï Idrissou

David Meunier

Dipankar bachar

Frédéric Barthélémy

Bjorg Kilavik

Thomas Brochier

Julien Sein

Bruno Nazarian

Jean-Luc Anton

Martin Szinte

Merci !

**Institute of Neuroscience and Medecine
Jülic, germany**

Sonja Grün

Michael Denker

Junji Ito

CEDRE

[Centre de formation et de soutien
aux données de la recherche]

Nadine Babaguela Matsono

William Domingues Vinhas

