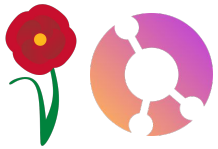
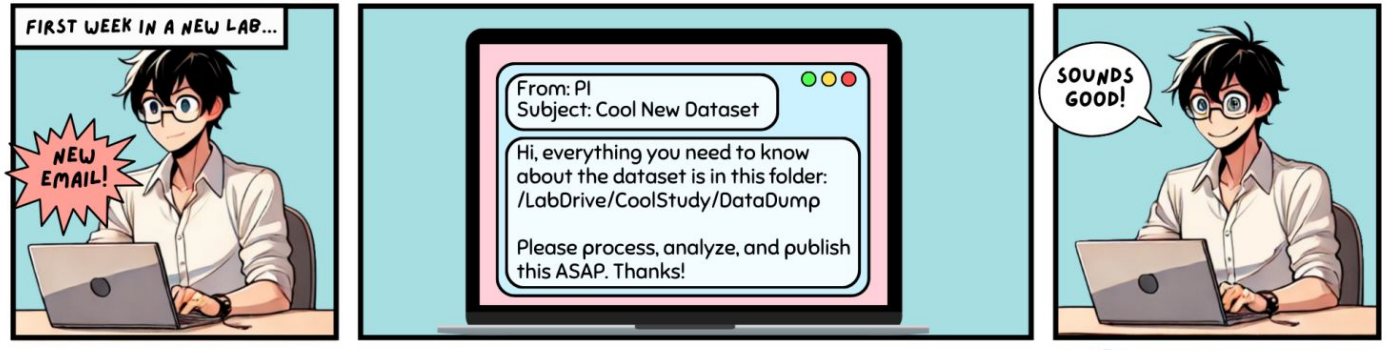


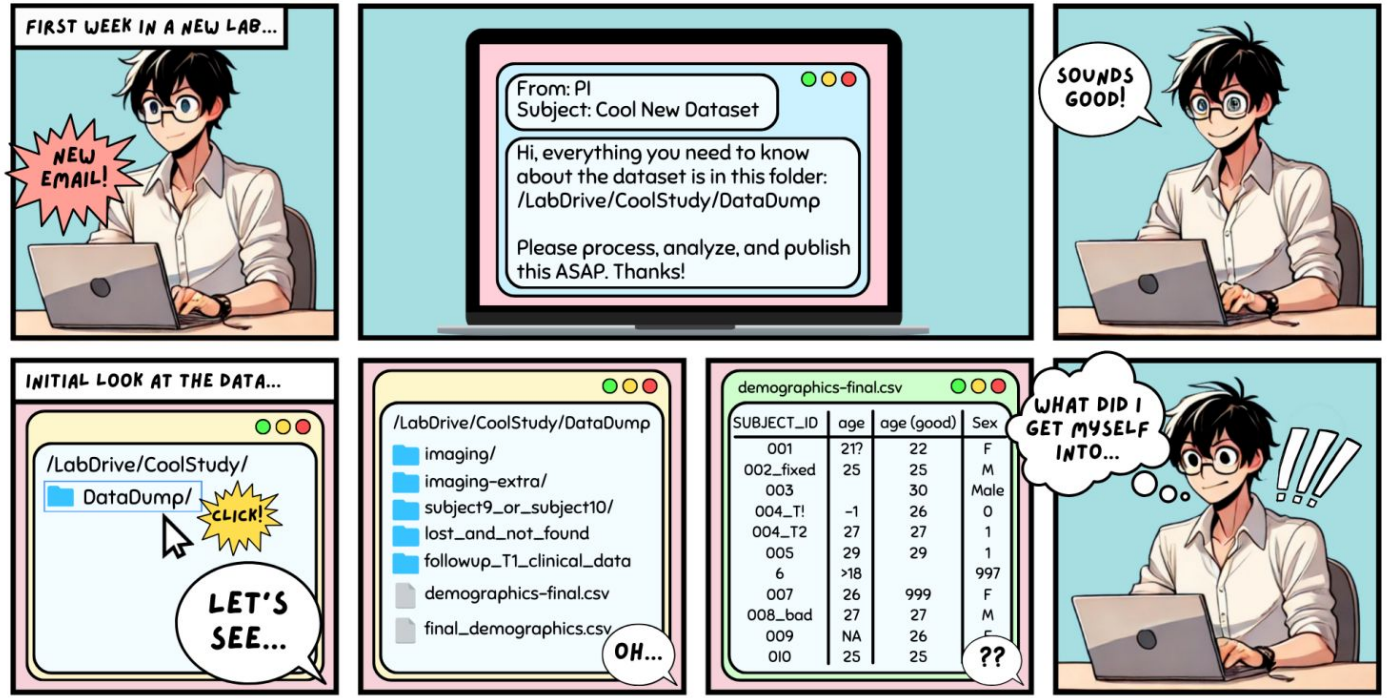
Nipoppy and **Neurobagel**: Practical tools to navigate barriers for FAIR research

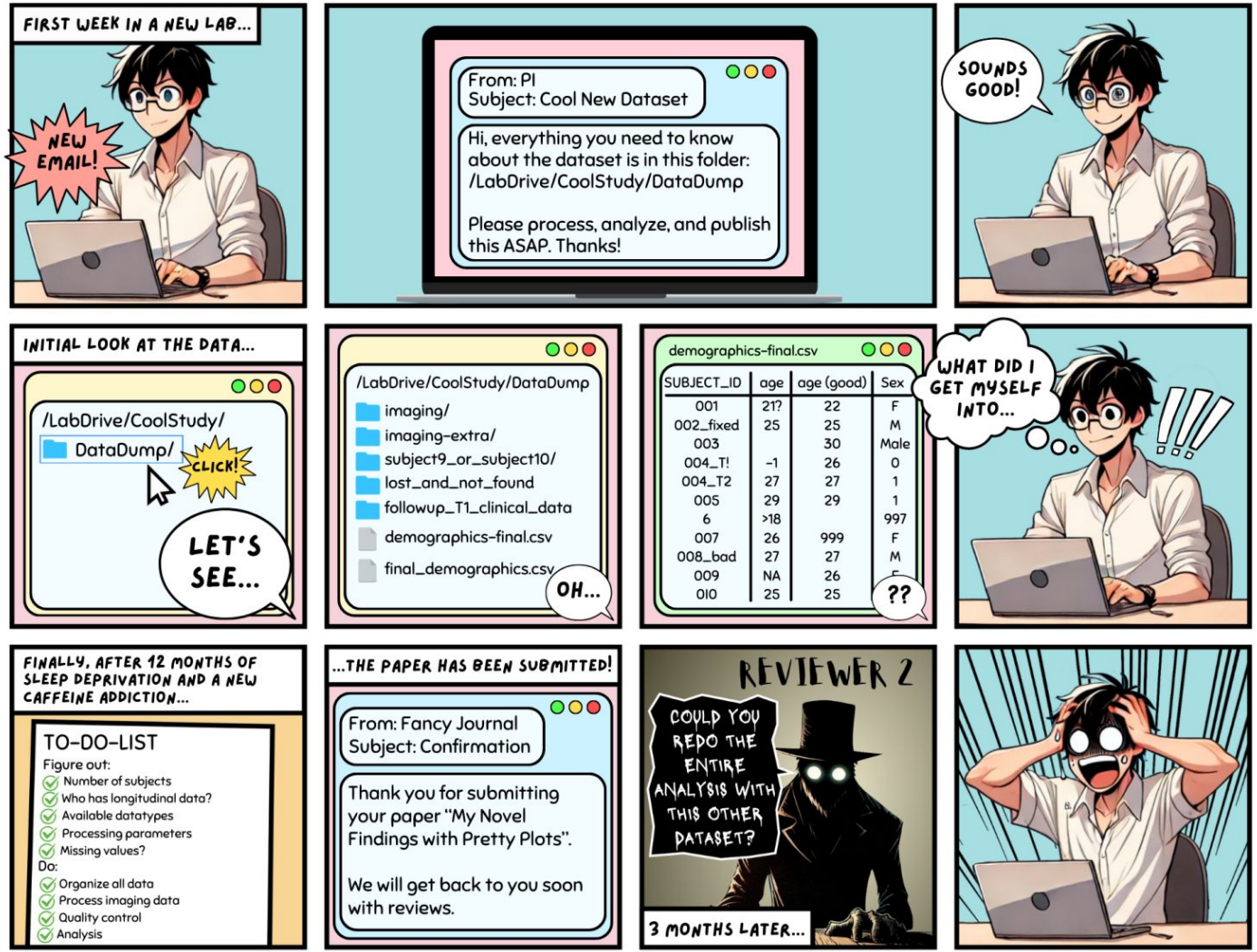
Michelle Wang
PhD candidate, NeuroDataScience-ORIGAMI Lab (JB Poline)

2025 June 18

ORIGAMI
Lab


FIRST WEEK IN A NEW LAB...
NEW EMAIL!
From: PI
Subject: Cool New Dataset
Hi, everything you need to know about the dataset is in this folder: /LabDrive/CoolStudy/DataDump
Please process, analyze, and publish this ASAP. Thanks!
SOUNDS GOOD!


FIRST WEEK IN A NEW LAB...
NEW EMAIL!
From: PI
Subject: Cool New Dataset
Hi, everything you need to know about the dataset is in this folder: /LabDrive/CoolStudy/DataDump
Please process, analyze, and publish this ASAP. Thanks!
SOUNDS GOOD!
INITIAL LOOK AT THE DATA...
/LabDrive/CoolStudy/
DataDump/
CLICK!
LET'S SEE...
/LabDrive/CoolStudy/DataDump
imaging/
imaging-extra/
subject9_or_subject10/
lost_and_not_found
followup_T1_clinical_data
demographics-final.csv
final_demographics.csv
OH...
demographics-final.csv
SUBJECT_ID | age | age (good) | Sex
001 | 21? | 22 | F
002_fixed | 25 | 25 | M
003 | | 30 | Male
004_T! | -1 | 26 | 0
004_T2 | 27 | 27 | 1
005 | 29 | 29 | 1
6 | >18 | | 997
007 | 26 | 999 | F
008_bad | 27 | 27 | M
009 | NA | 26 |
010 | 25 | 25 |
??
WHAT DID I GET MYSELF INTO...
!!!


FIRST WEEK IN A NEW LAB...
NEW EMAIL!
From: PI
Subject: Cool New Dataset
Hi, everything you need to know about the dataset is in this folder: /LabDrive/CoolStudy/DataDump
Please process, analyze, and publish this ASAP. Thanks!
SOUNDS GOOD!
INITIAL LOOK AT THE DATA...
/LabDrive/CoolStudy/
DataDump/
CLICK!
LET'S SEE...
/LabDrive/CoolStudy/DataDump
imaging/
imaging-extra/
subject9_or_subject10/
lost_and_not_found
followup_T1_clinical_data
demographics-final.csv
final_demographics.csv
OH...
demographics-final.csv
SUBJECT_ID | age | age (good) | Sex
001 | 21? | 22 | F
002_fixed | 25 | 25 | M
003 | | 30 | Male
004_T! | -1 | 26 | 0
004_T2 | 27 | 27 | 1
005 | 29 | 29 | 1
6 | >18 | | 997
007 | 26 | 999 | F
008_bad | 27 | 27 | M
009 | NA | 26 |
010 | 25 | 25 |
??
WHAT DID I GET MYSELF INTO...
!!!
FINALLY, AFTER 12 MONTHS OF SLEEP DEPRIVATION AND A NEW CAFFEINE ADDICTION...
TO-DO-LIST
Figure out:
Number of subjects
Who has longitudinal data?
Available datatypes
Processing parameters
Missing values?
Do:
Organize all data
Process imaging data
Quality control
Analysis
...THE PAPER HAS BEEN SUBMITTED!
From: Fancy Journal
Subject: Confirmation
Thank you for submitting your paper "My Novel Findings with Pretty Plots".
We will get back to you soon with reviews.
REVIEWER 2
COULD YOU REDO THE ENTIRE ANALYSIS WITH THIS OTHER DATASET?
3 MONTHS LATER...

The ORIGAMI Lab

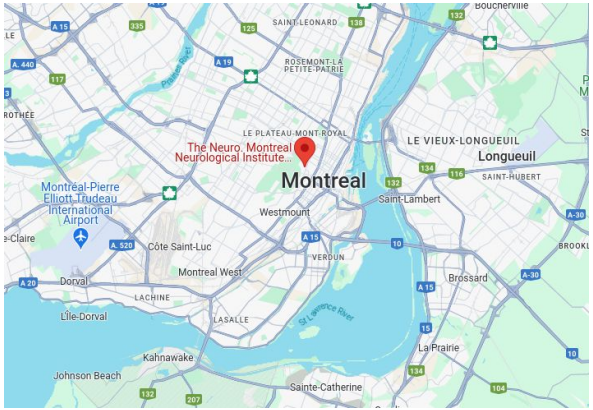
- **Promote** FAIR neuro-data-science
- **Build** neuroinformatic workflows
- **Improve** clinical research practices

JB Poline

ORIGAMI Lab

neuro
Montreal Neurological Institute-Hospital

McGill UNIVERSITY



REPRONIM

nilearn

OpenNEURO

BIDS
BRAIN IMAGING DATA STRUCTURE

CONP
PCNO

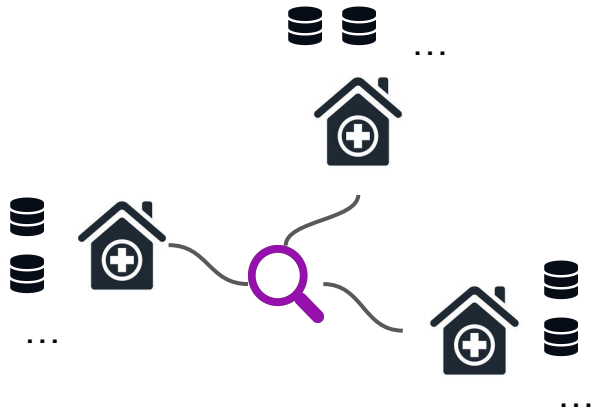
THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH

Douglas
CENTRE DE RECHERCHE RESEARCH CENTRE

Centre de recherche iugm
Institut universitaire de gériatrie de Montréal

Data challenges in **multicentric** studies

- **Curation**: inevitability of heuristics
- **Processing**: reproducibility and versioning
- **Harmonization**: clinical knowledge
- **Sharing**: privacy, access control



A lot of time-consuming manual work that is difficult to coordinate!

ORIGAMI solutions

Nipoppy (curate + process)

A protocol for organizing and processing neuroimaging data

Neurobagel (annotate + search)

Tools for decentralized data harmonization and search

ORIGAMI solutions

Nipoppy (curate + process)

A protocol for organizing and processing neuroimaging data

Neurobagel (annotate + search)

Tools for decentralized data harmonization and search

Bagel

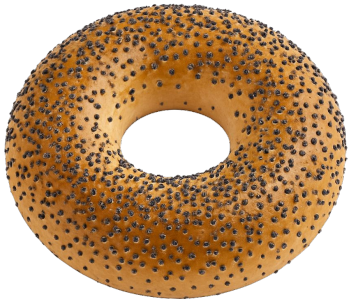
ORIGAMI solutions

Nipoppy (curate + process)

A protocol for organizing and processing neuroimaging data

Neurobagel (annotate + search)

Tools for decentralized data harmonization and search



Poppy seed **bagel**

Michelle Wang

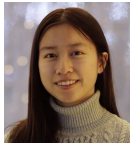
Nikhil Bhagwat

Nipoppy: A framework for the **organization** and **decentralized processing** of neuroimaging-clinical studies

Mathieu Dugré

Julia Fan

Brent McPherson



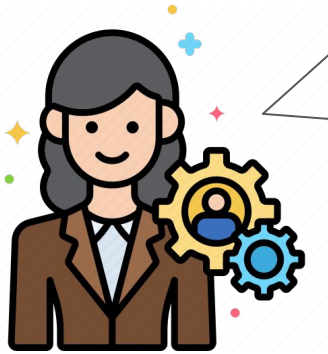
Jean-Baptiste Poline

https://nipoppy.readthedocs.io/

https://github.com/nipoppy/nipoppy

Neuroimaging-clinical datasets are **messy**

- **Curating/processing** the data is time-consuming and error-prone
  - Reproducibility concerns
- **Linking** data is difficult



How many participants with
- Parkinson's disease diagnosis
- age under 65
- two imaging sessions
- processed with FreeSurfer

are in my dataset?

How can we make neuroimaging data more FAIR?

- **FAIR**: findable, accessible, interoperable, reusable (Wilkinson *et al.*, 2016)

Existing open standards/tools developed by the community

(Gorgolewski *et al.*, 2016)

(Kurtzer *et al.*, 2017)

(Glatard *et al.*, 2018)

We leveraged existing open science tools to build a flexible framework for data organization and processing of neuroimaging-clinical data

Introducing the **Nipoppy** framework

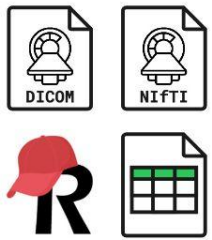
1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets

Introducing the **Nipoppy** framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets

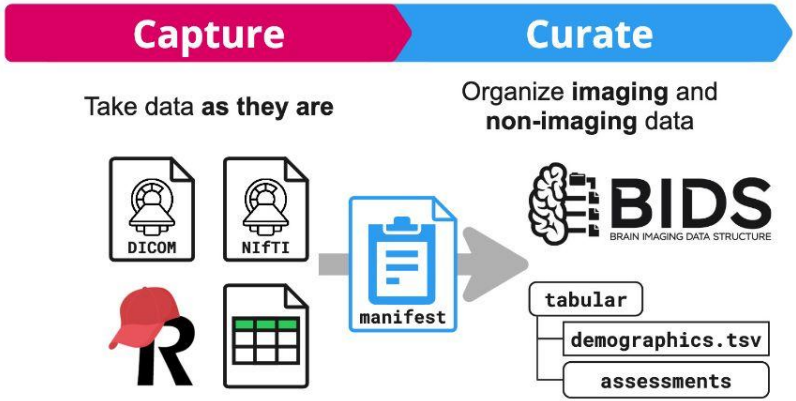
Capture

Take data **as they are**



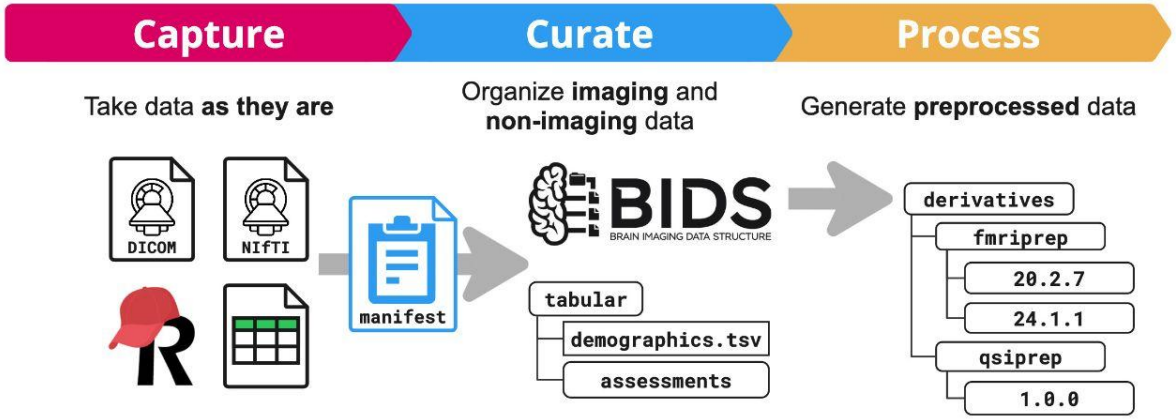
Introducing the **Nipoppy** framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets



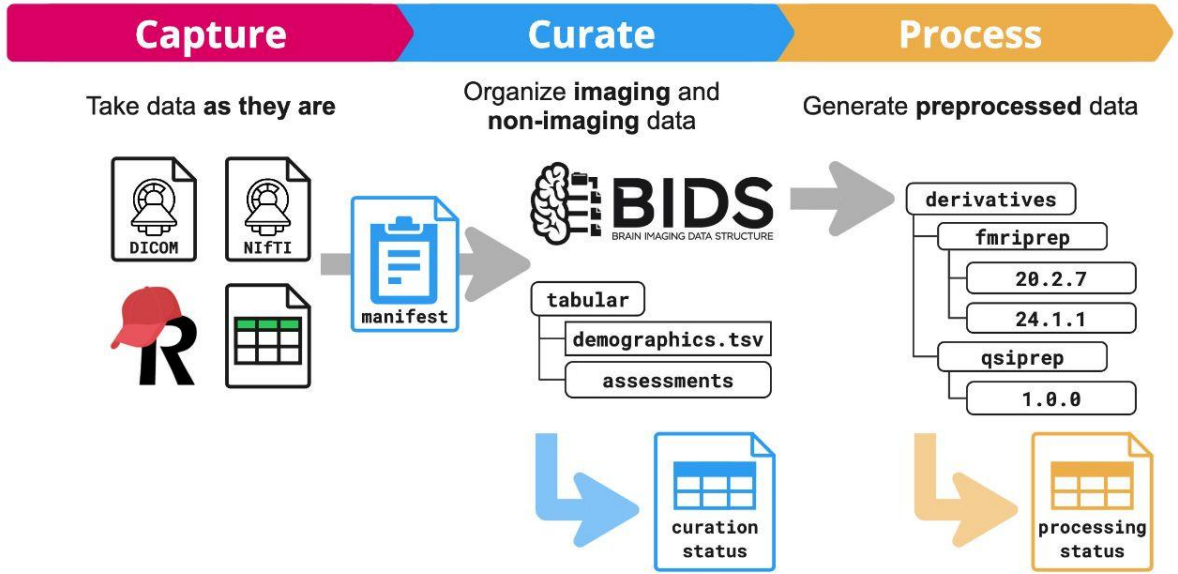
Introducing the Nipoppy framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets



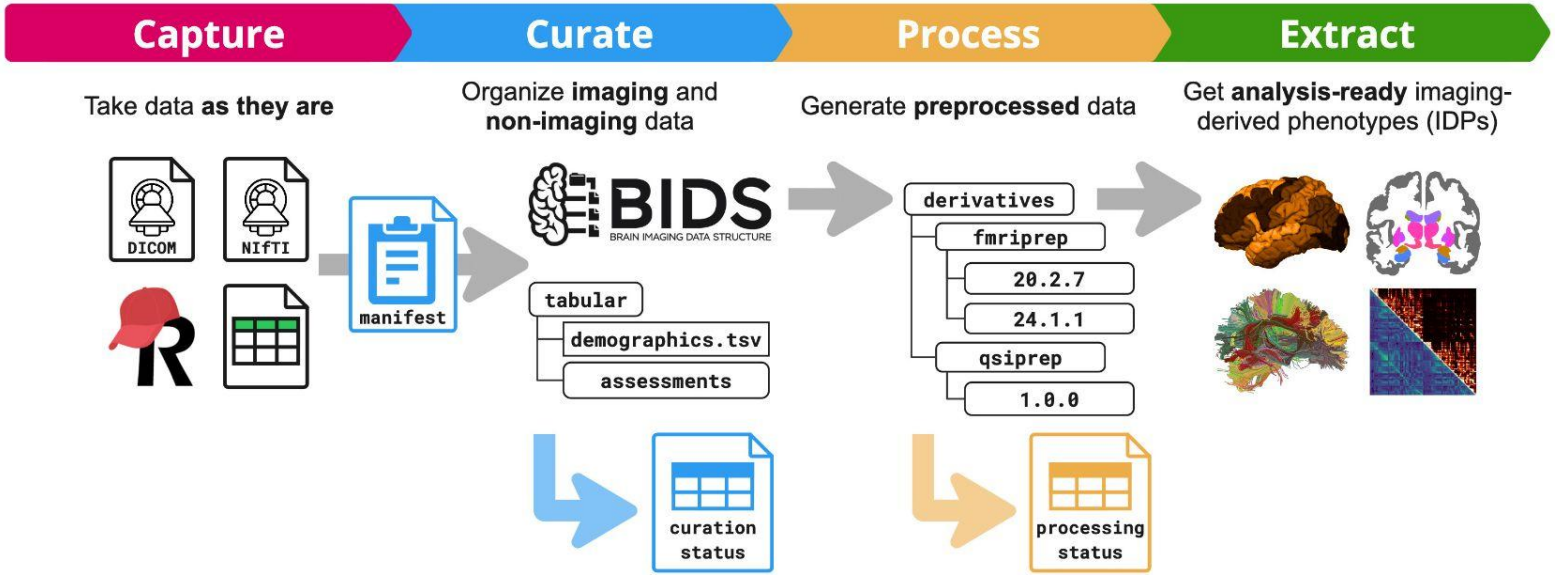
Introducing the Nipoppy framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets



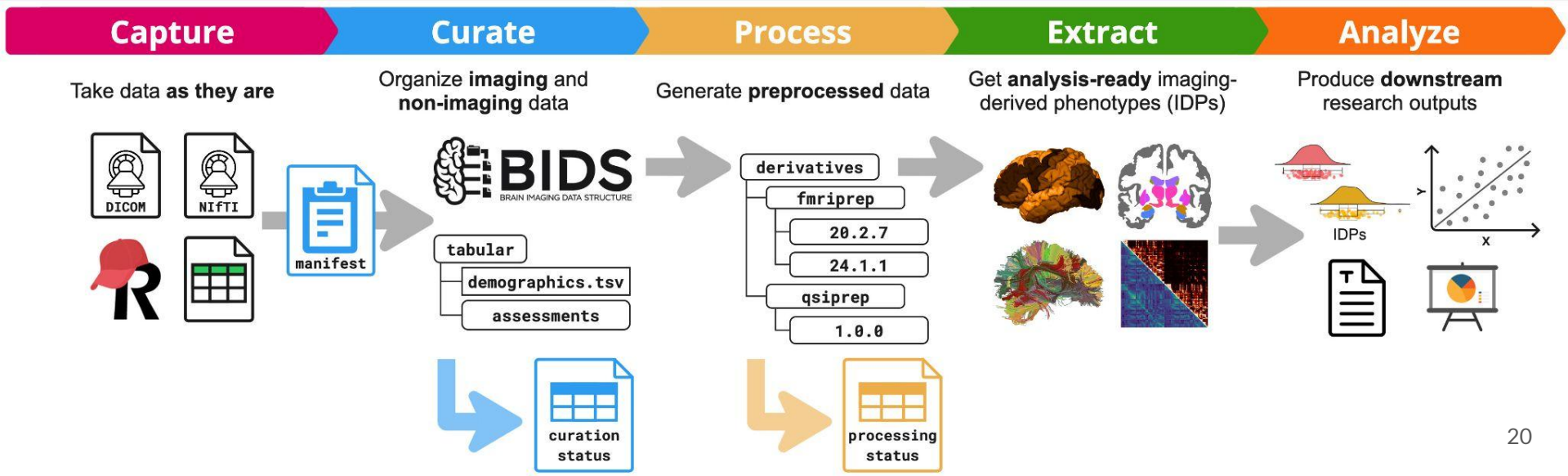
Introducing the **Nipoppy** framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets



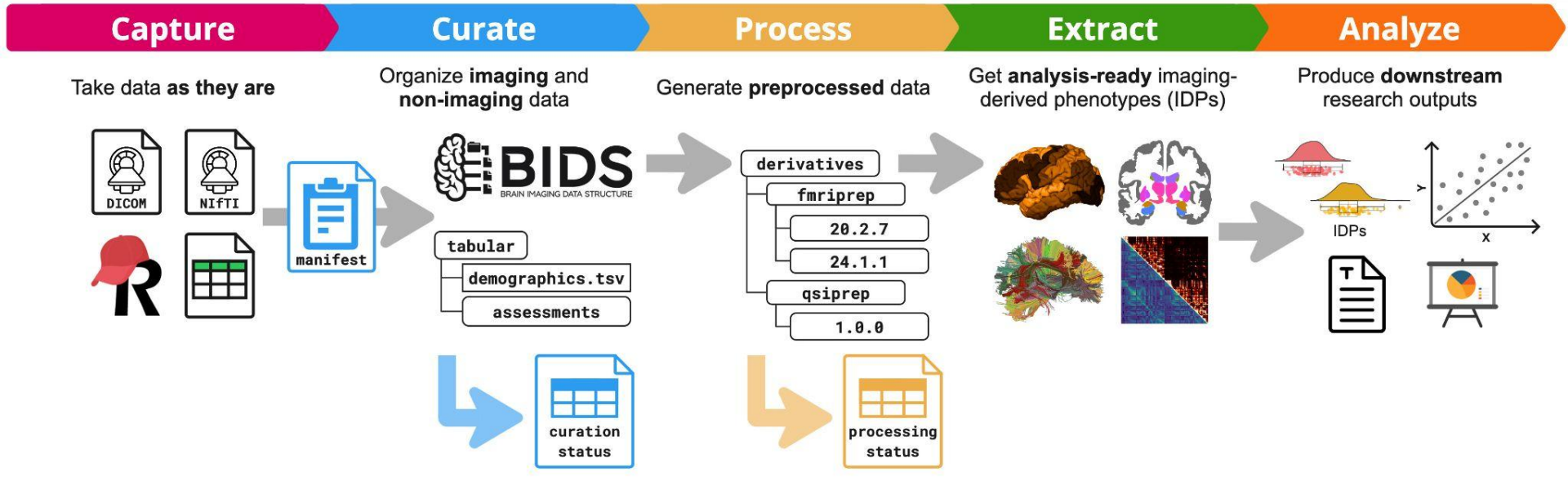
Introducing the **Nipoppy** framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets

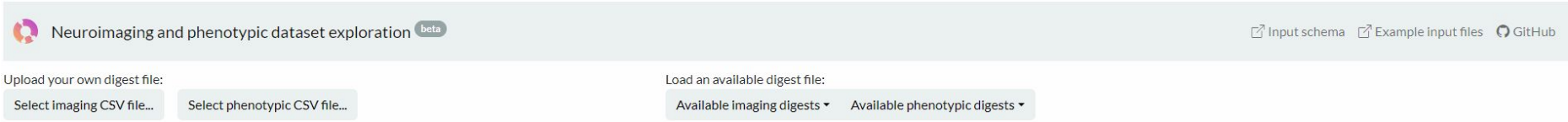


Introducing the **Nipoppy** framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets
   a. From raw scanner data to extracted imaging-derived phenotypes (IDPs)
   b. Visualizing tabular data/metadata on a dashboard (https://digest.neurobagel.org/)



https://digest.neurobagel.org/



Introducing the **Nipoppy** framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets
   a. From raw scanner data to extracted imaging-derived phenotypes (IDPs)
   b. Visualizing tabular data/metadata on a dashboard
      (https://digest.neurobagel.org/)

Introducing the **Nipoppy** framework

1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets
   a. From raw scanner data to extracted imaging-derived phenotypes (IDPs)
   b. Visualizing tabular data/metadata on a dashboard
   (https://digest.neurobagel.org/)

2. Data organization **specification**
   a. For imaging and non-imaging data
   b. At the whole study level

Introducing the **Nipoppy** framework

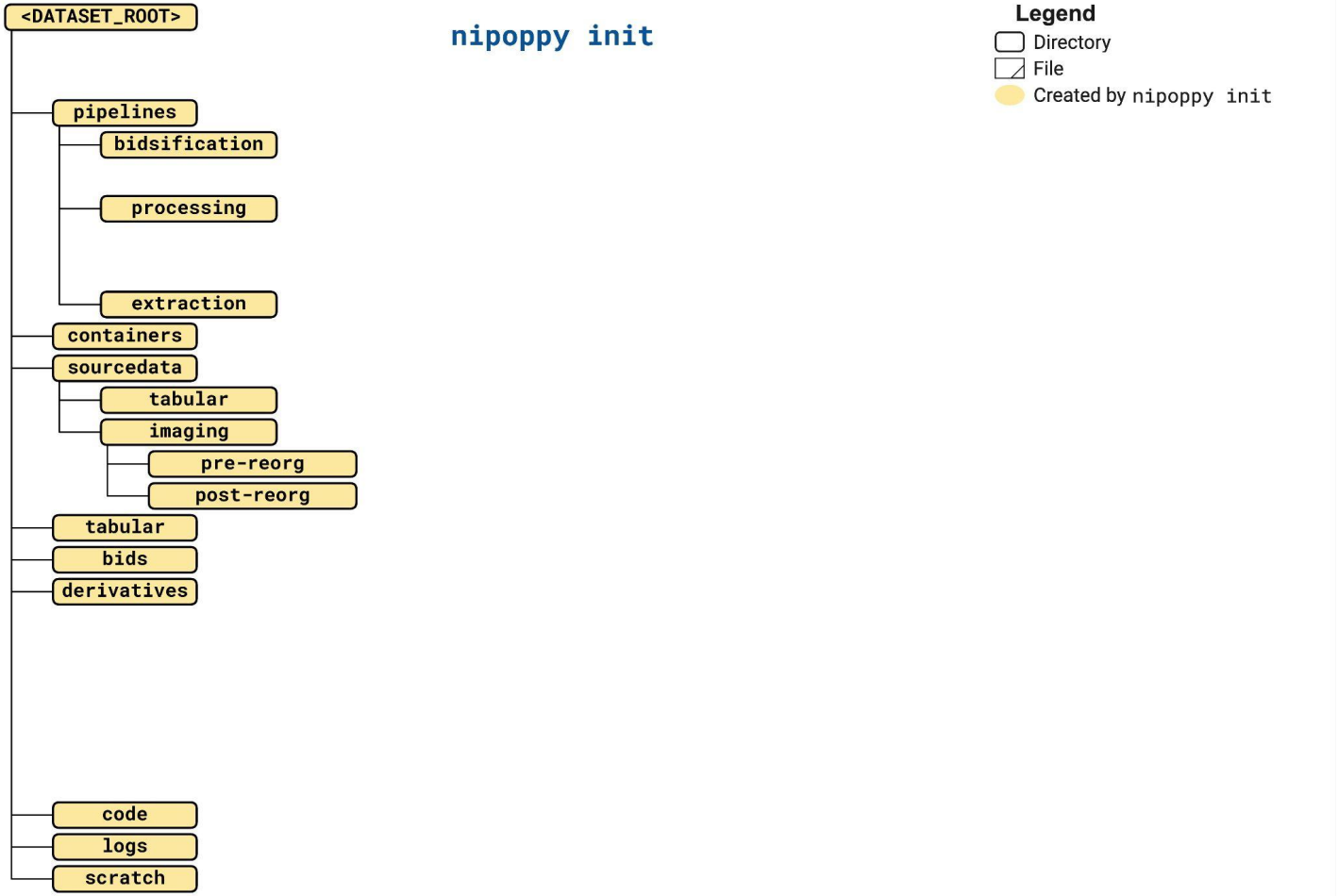
1. **Protocol** for best practices (FAIR principles) in organization/processing of neuroimaging-clinical datasets
   a. From raw scanner data to extracted imaging-derived phenotypes (IDPs)
   b. Visualizing tabular data/metadata on a dashboard
   (https://digest.neurobagel.org/)

2. Data organization **specification**
   a. For imaging and non-imaging data
   b. At the whole study level

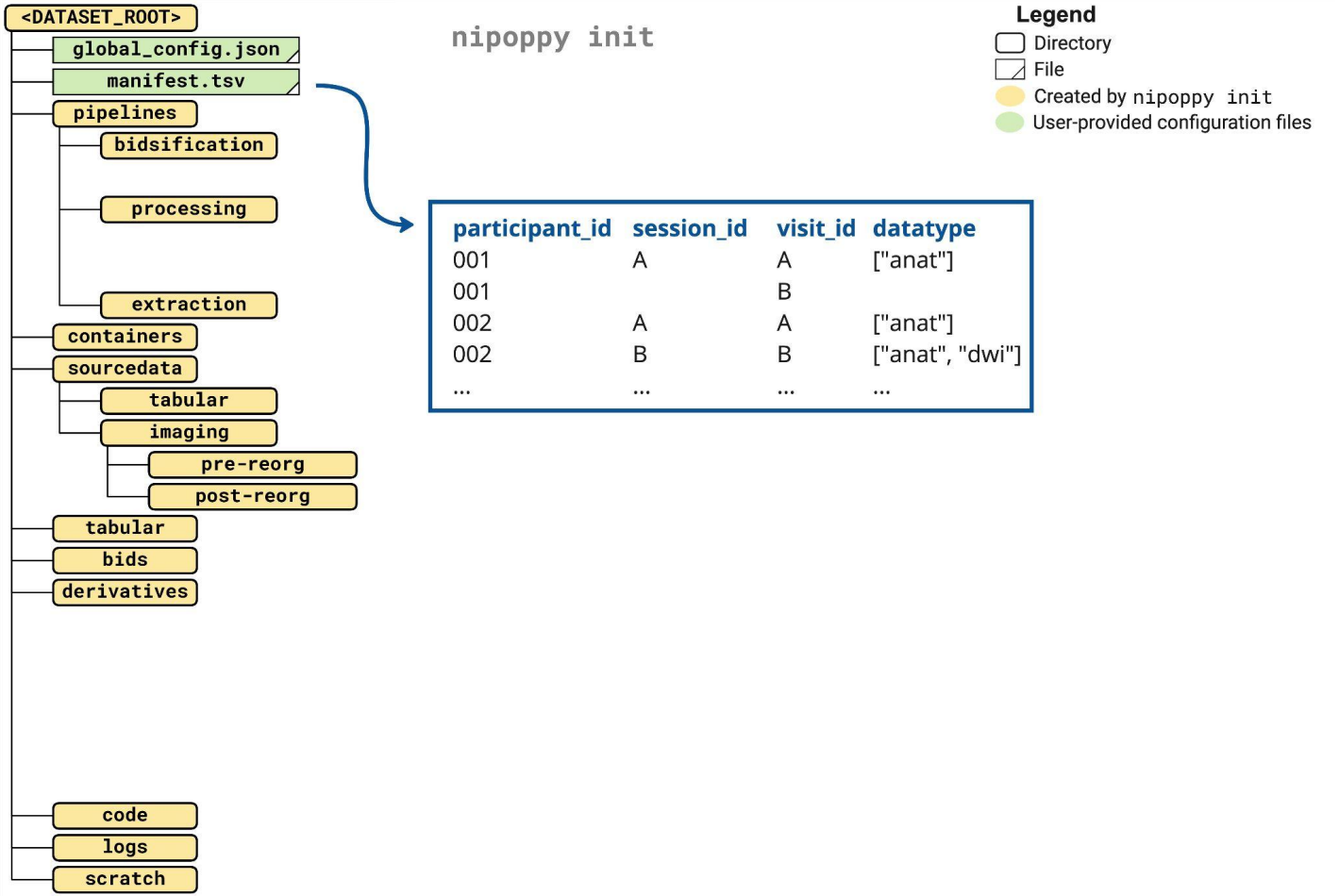
3. Software package with **tools** to help work with this framework **locally**

A typical **Nipoppy** workflow

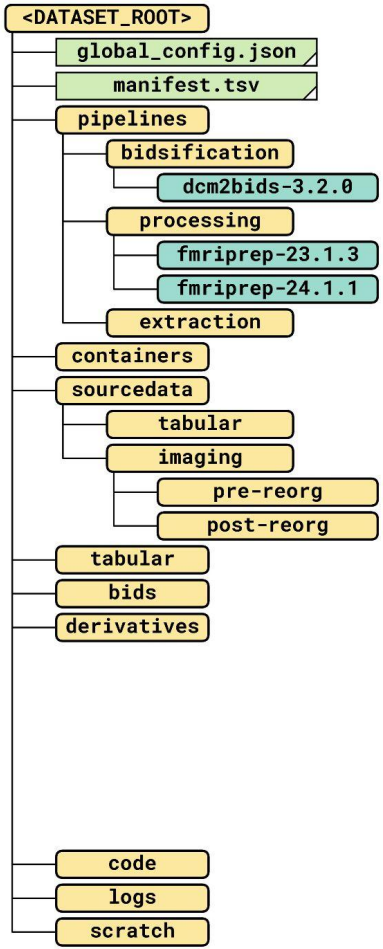
A typical Nipoppy workflow



A typical Nipoppy workflow



A typical **Nipoppy** workflow



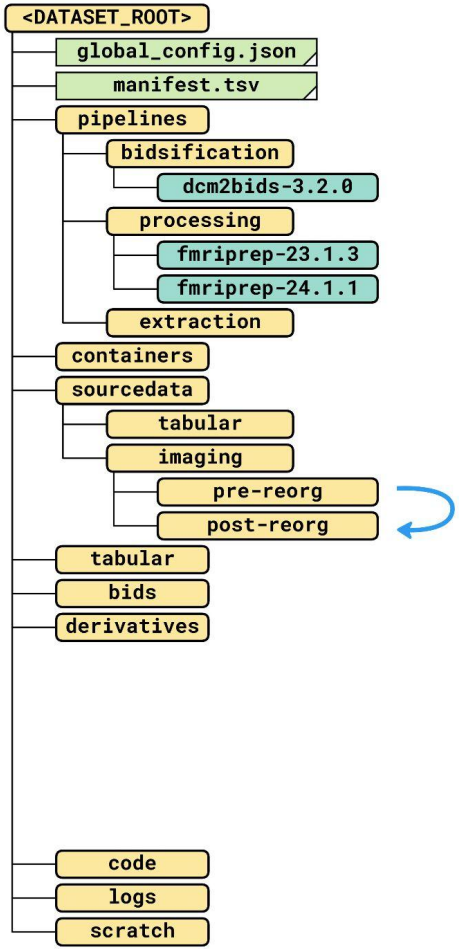
`nipoppy init`

`nipoppy pipeline install`

Legend

- Directory
- File
- Created by `nipoppy init`
- User-provided configuration files
- Pipeline configuration files

A typical Nipoppy workflow



`nipoppy init`

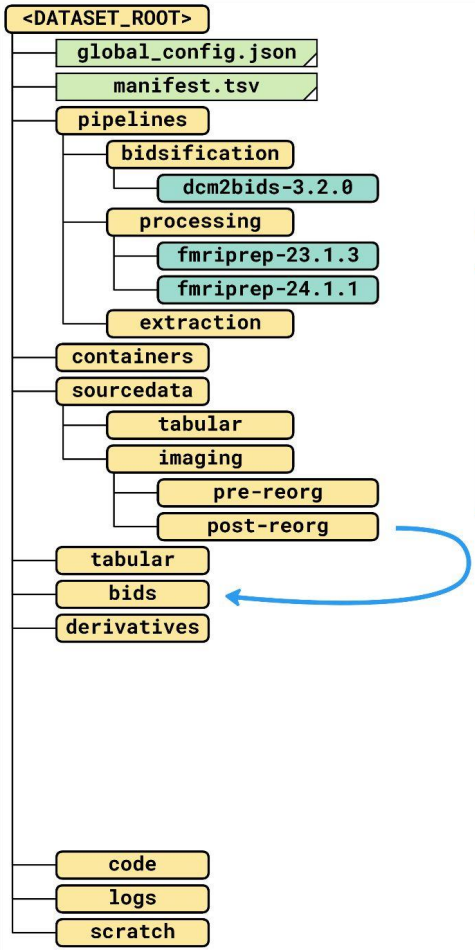
`nipoppy pipeline install`

`nipoppy reorg`

Legend

- Directory
- File
- Created by `nipoppy init`
- User-provided configuration files
- Pipeline configuration files

A typical **Nipoppy** workflow



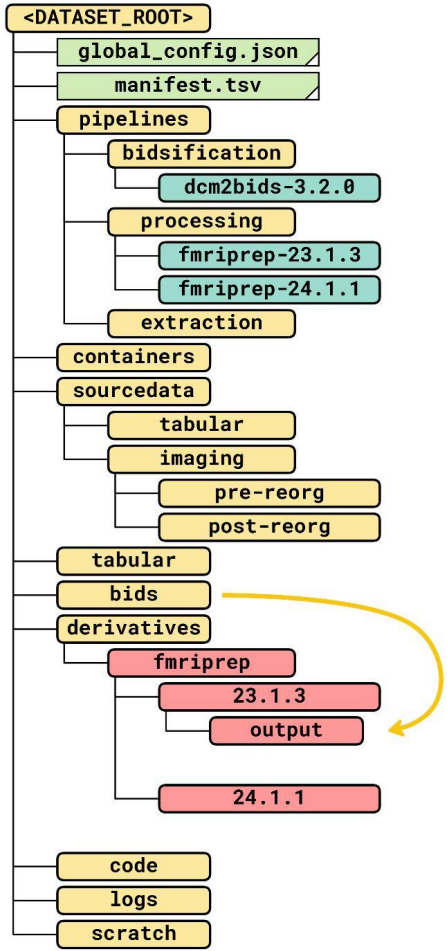

```
nipoppy init

nipoppy pipeline install

nipoppy reorg

nipoppy bidsify
```

```
{
  "sourcefolder": "[[NIPOPPY_DPATH_SOURCEDATA]]",
  "bidsfolder": "[[NIPOPPY_DPATH_BIDS]]",
  "participant_label": [
      "[[NIPOPPY_PARTICIPANT_ID]]"
  ]
}
```

Legend

- Directory
- File
- Created by `nipoppy init`
- User-provided configuration files
- Pipeline configuration files

A typical Nipoppy workflow



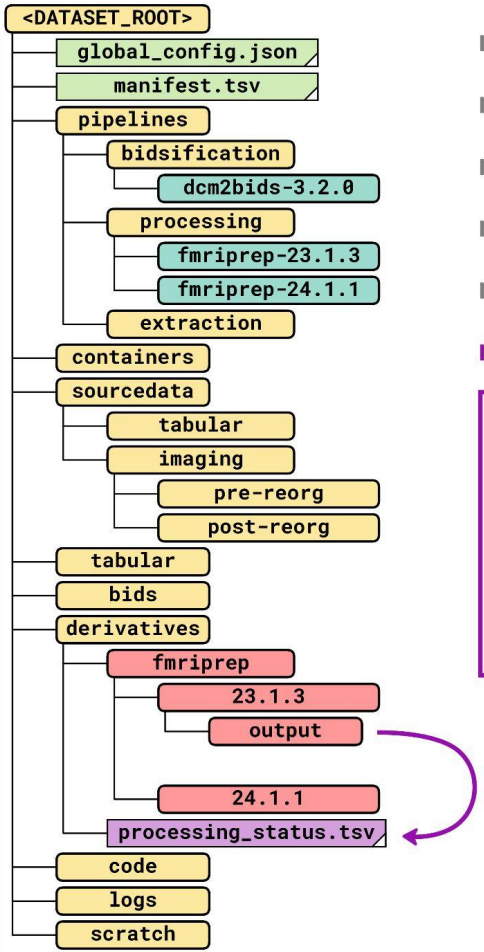

```
nipoppy init

nipoppy pipeline install

nipoppy reorg

nipoppy bidsify

nipoppy process
```

```
{
  "bids_dir": "[[NIPOPPY_DATASET_ROOT]]",
  "output_dir": "[[NIPOPPY_DPATH_PIPELINE_OUTPUT]]",
  "analysis_level": "participant",
  ...
}
```

Legend

- Directory
- File
- Created by `nipoppy init`
- User-provided configuration files
- Pipeline configuration files
- Imaging derivatives data

A typical **Nipoppy** workflow

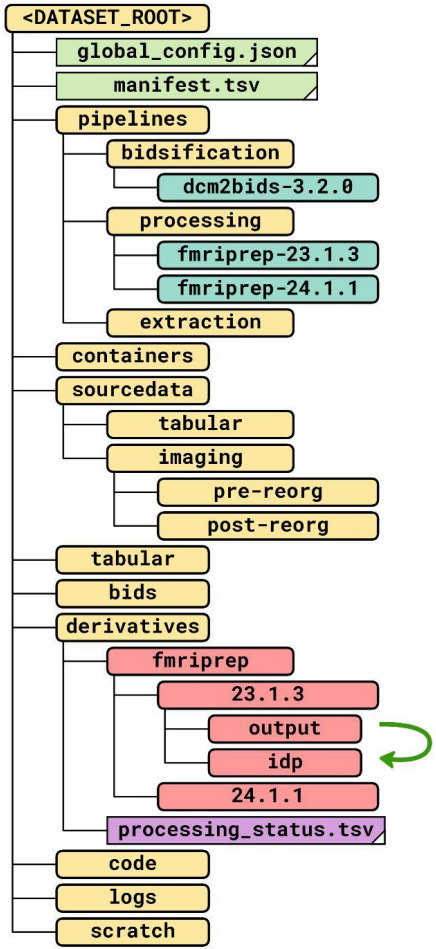



```
nipoppy init

nipoppy pipeline install

nipoppy reorg

nipoppy bidsify

nipoppy process

nipoppy track-processing
```

participant_id	session_id	pipeline	version	status
01	A	freesurfer	6.0.0	SUCCESS
01	A	freesurfer	7.4.1	SUCCESS
02	A	freesurfer	6.0.0	FAIL
02	A	freesurfer	7.4.1	FAIL
02	B	freesurfer	6.0.0	INCOMPLETE
02	B	freesurfer	7.4.1	INCOMPLETE
...	...	...	...	...

Legend
- Directory
- File
- Created by `nipoppy init`
- User-provided configuration files
- Pipeline configuration files
- Imaging derivatives data
- Tracking files

A typical Nipoppy workflow




```
nipoppy init

nipoppy pipeline install

nipoppy reorg

nipoppy bidsify

nipoppy process

nipoppy track-processing

nipoppy extract
```

participant_id	session_id	gm_vol	wm_vol	...
001	A	399445	463167	...
002	A	432054	422803	...
002	B	432913	422634	...
...	...	...	...	...

Legend

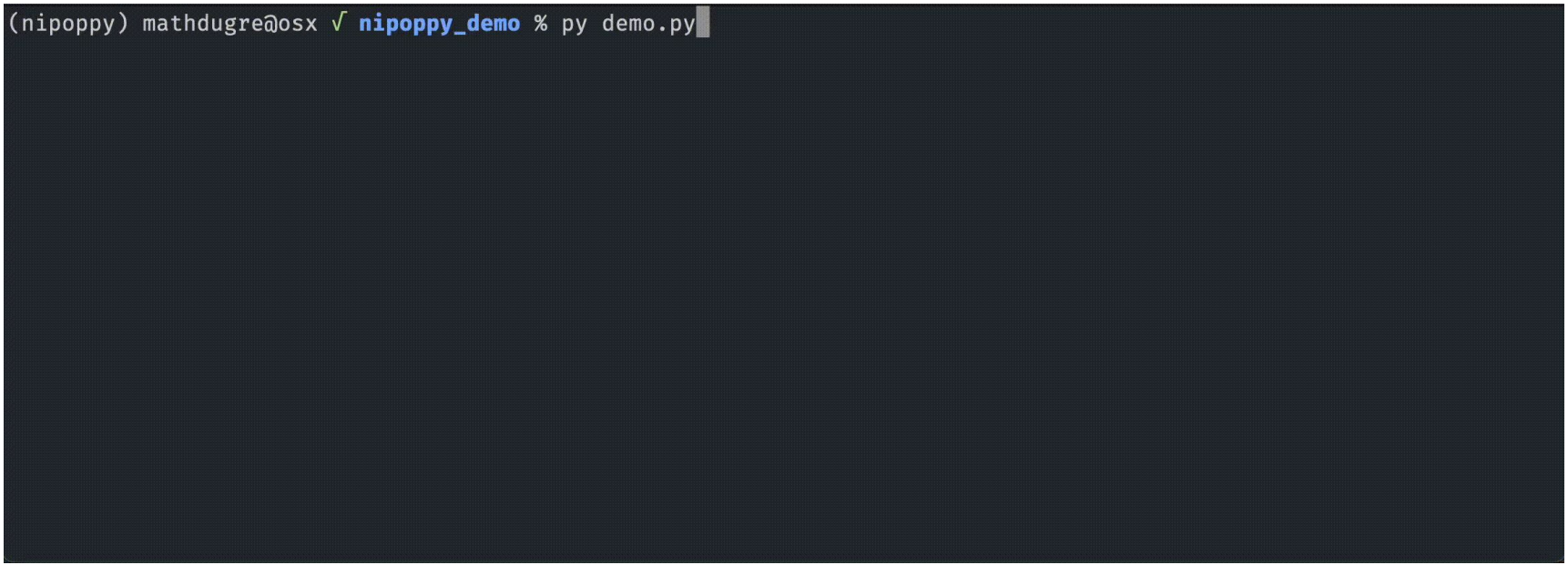
- Directory
- File
- Created by `nipoppy init`
- User-provided configuration files
- Pipeline configuration files
- Imaging derivatives data
- Tracking files

Demo

The Nipoppy pipeline catalog

Available on the Zenodo data repository

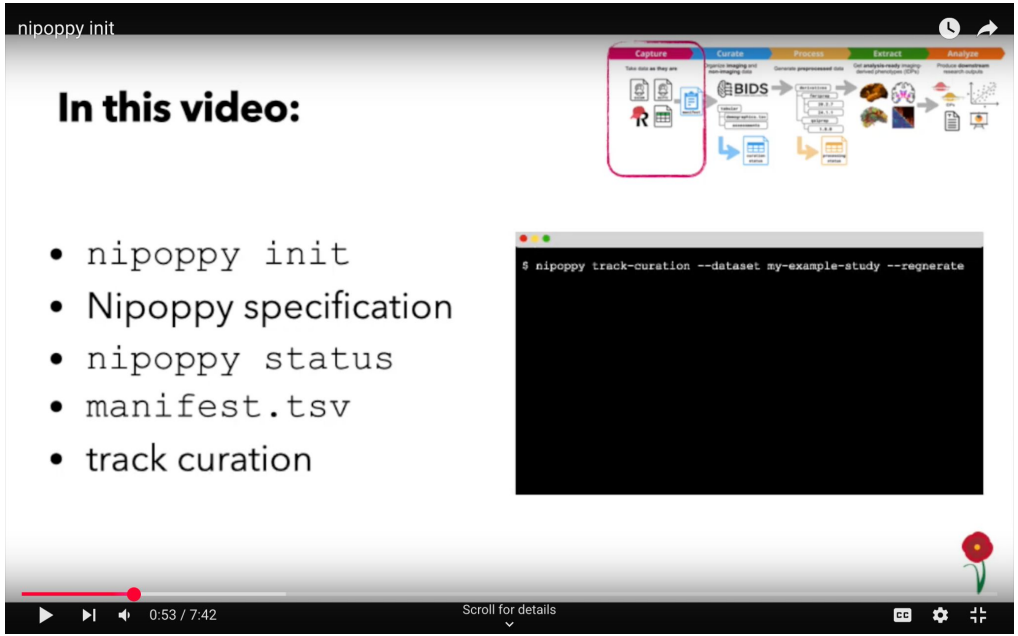
- Community-driven
- Seamless pipeline configuration sharing (container, runtime parameters)



Documentation and video tutorials

Written documentation available at https://nipoppy.readthedocs.io/en/latest/

Coming soon: video tutorials on YouTube



Nipoppy takeaways

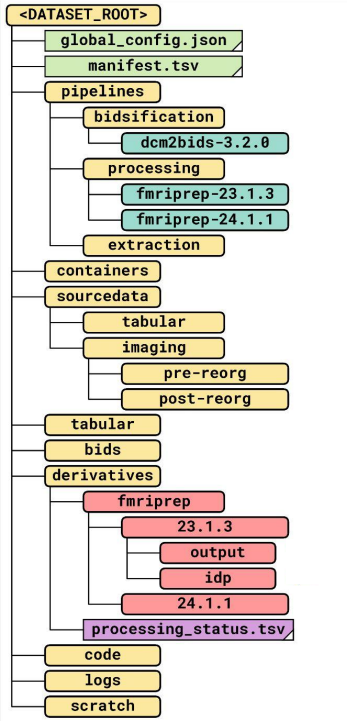
Framework combining **existing tools**

BIDS BRAIN IMAGING DATA STRUCTURE

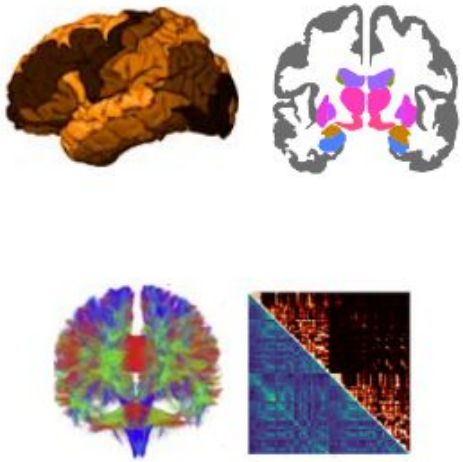
APPTAINER

Boutiques

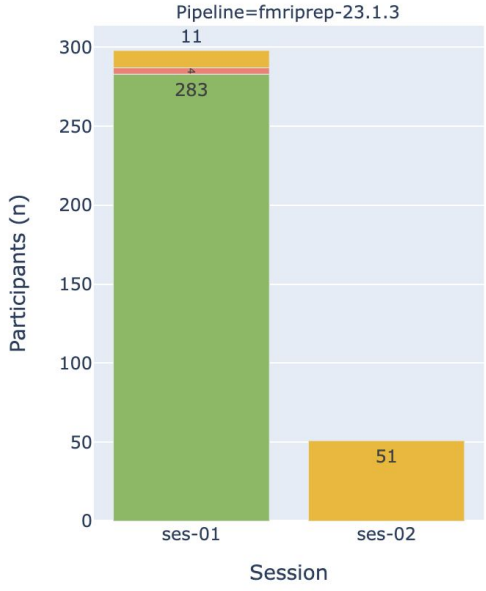
Local data management



Reproducible MRI processing and IDP extraction



Progress tracking

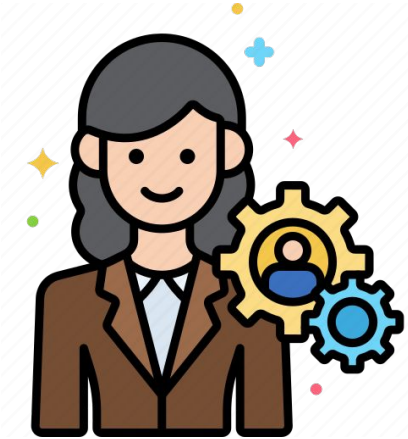


Nipoppy takeaways

For trainees

Best/FAIR practices

For labs



Efficiency and consistency

For institutions

Data-sharing

Metadata discovery with **Neurobagel**

Neurobagel © 2025 by Neurobagel Team, Origami Lab, McGill
slides are licensed under CC BY 4.0

Alyssa Dai

Rémi Gau

Arman Jahanpour

Brent McPherson

ikhil Bhagwat
Jean-Baptiste Poline

Sebastian Urchs

Michelle Wang

neurobagel

neurobagel.org

neurobagel

A neuroinformatics framework for decentralized data harmonization and search

ORIGAMI Lab

HEALTHY BRAINS HEALTHY LIVES

neuro Montreal Neurological Institute-Hospital

CONP PCNO

NIH National Institutes of Health

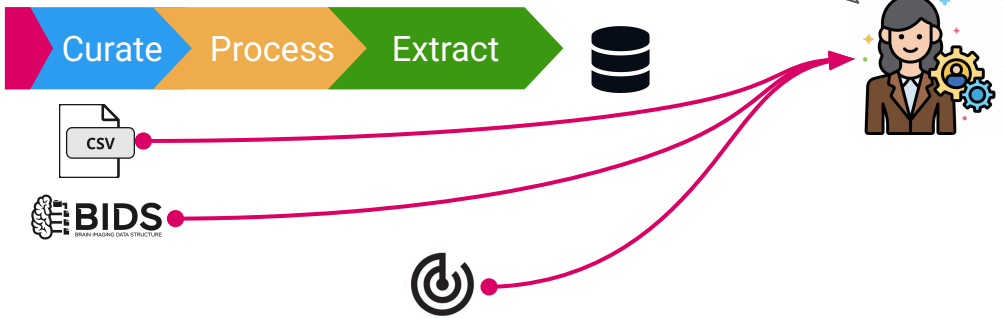
Fondation Brain Canada Foundation

NIBIB

Nipoppy individual study

How many participants
- with PD diagnosis
- under 65
- 2 imaging sessions
- run with freesurfer v6

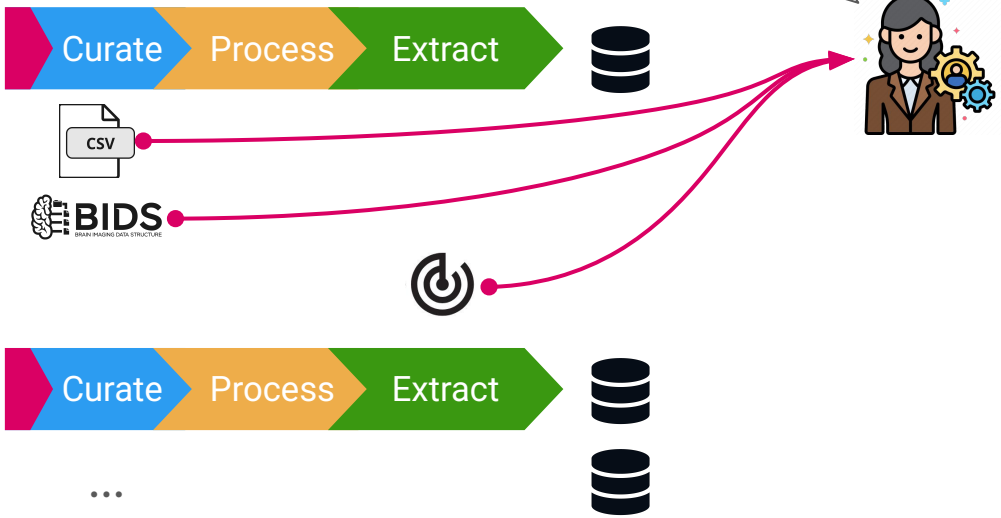
are in **this one dataset?**



Nipoppy individual study

How many participants
- with PD diagnosis
- under 65
- 2 imaging sessions
- run with freesurfer v6

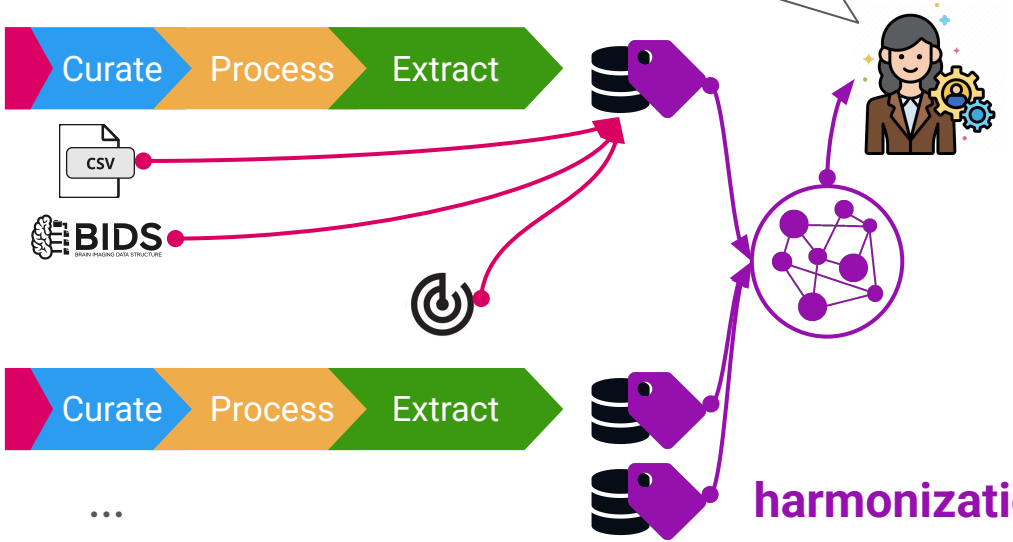
do our PIs have?



Neurobagel individual site

How many participants
- with PD diagnosis
- under 65
- 2 imaging sessions
- run with freesurfer v6

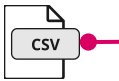
do our PIs have?



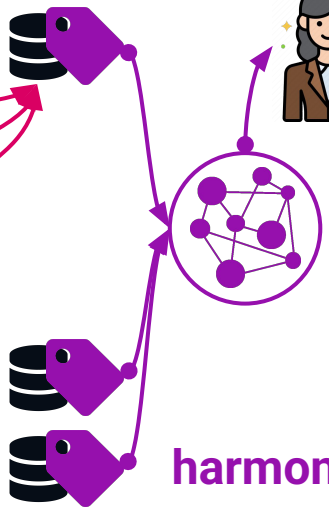
Neurobagel multiple sites

How many participants
- with PD diagnosis
- under 65
- 2 imaging sessions
- run with freesurfer v6

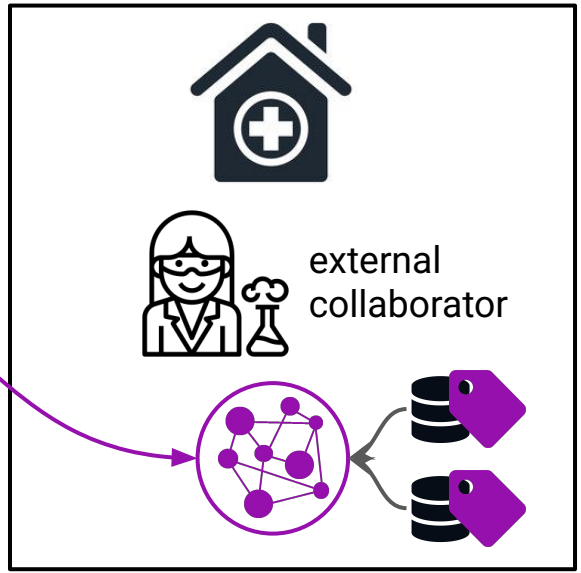
exist within our network?



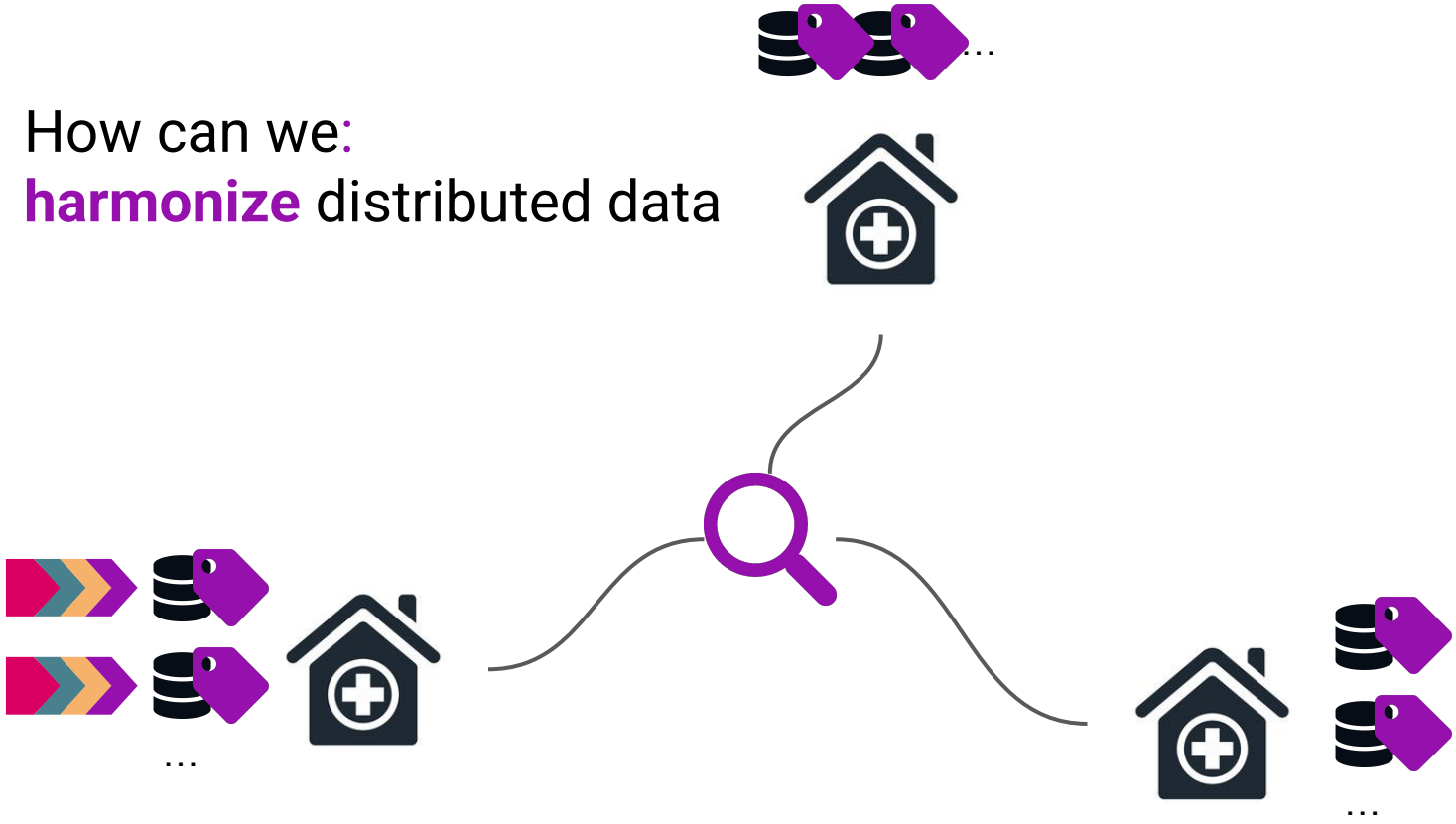
...



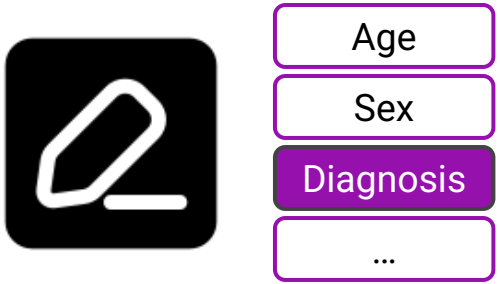
harmonization



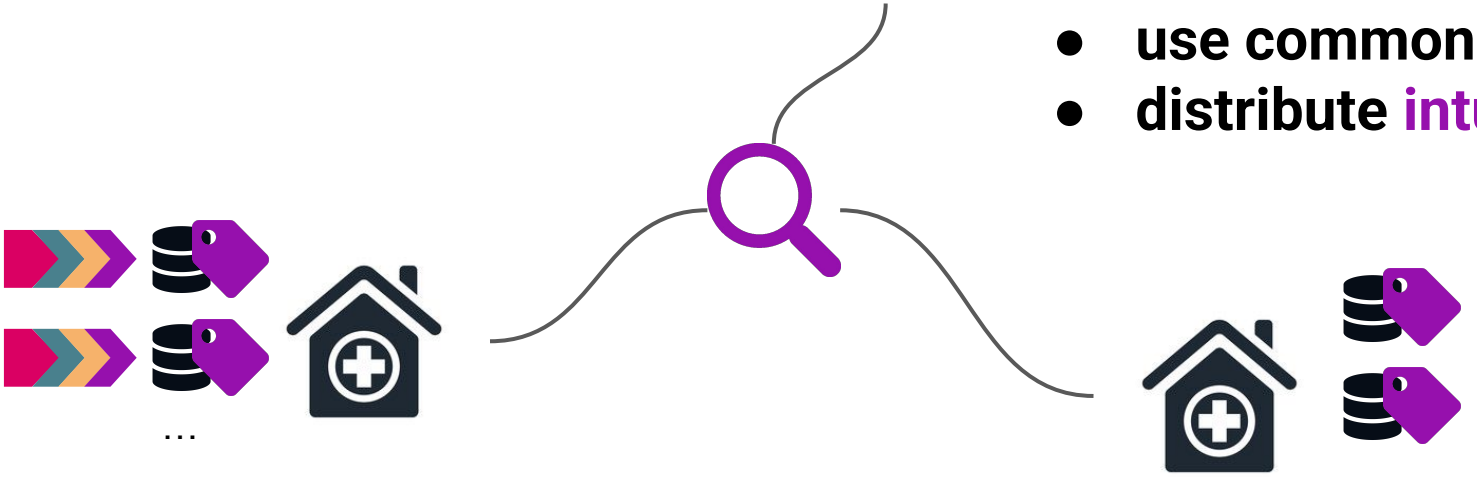
How can we:
harmonize distributed data



How can we:
harmonize distributed data



- **use common data model**
- **distribute intuitive GUI**



SNOMED CT
The global language of healthcare

RDF

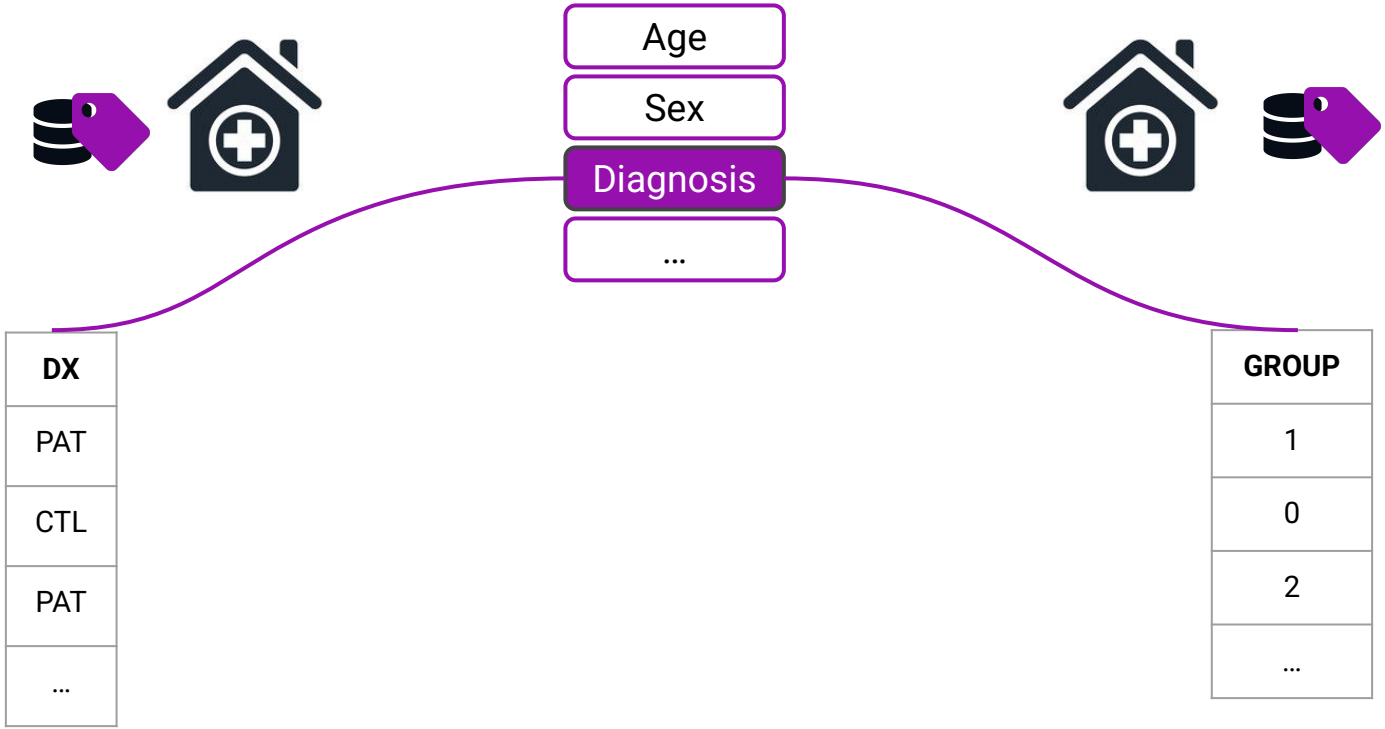
W3C
Word Wide Web Consortium

SPARQL

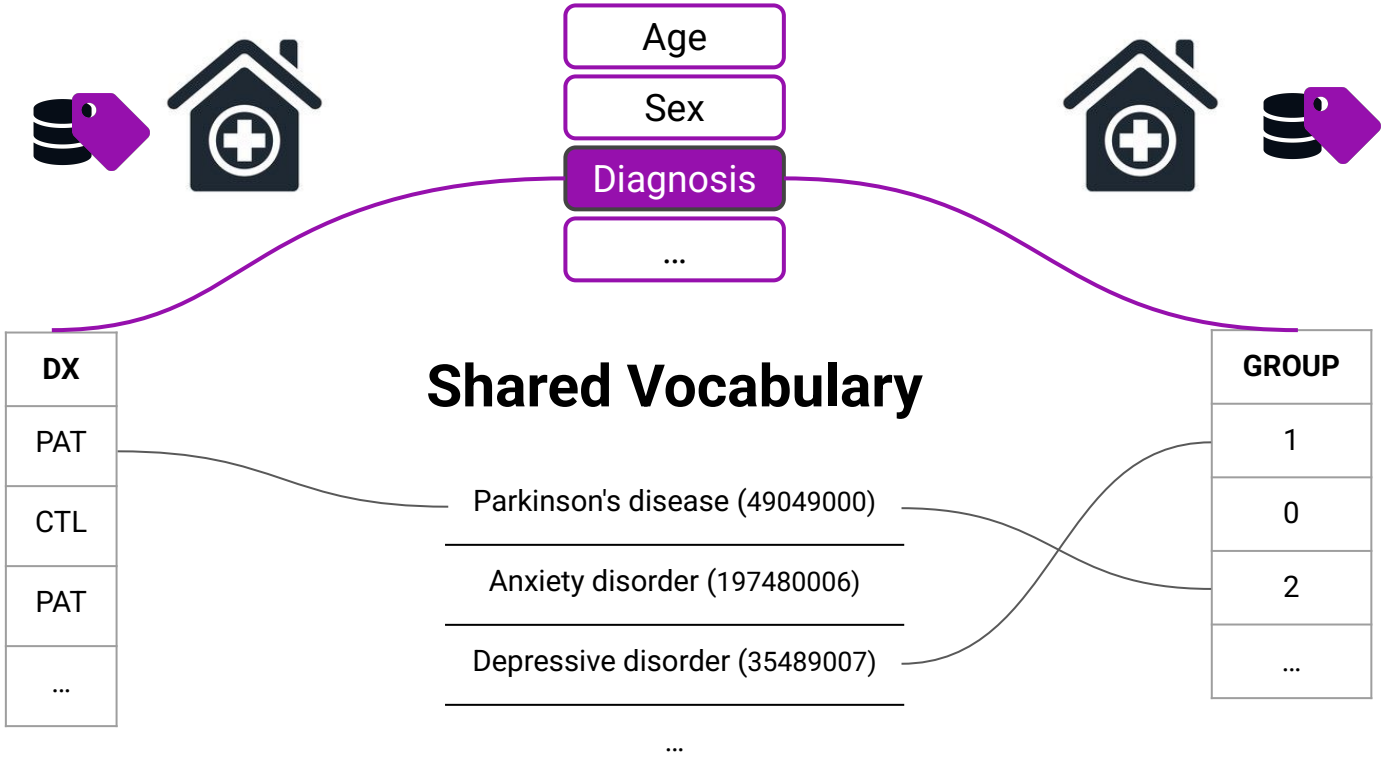
BIDS
BRAIN IMAGING DATA STRUCTURE

Findable Accessible Interoperable Reusable

Common participant level data model

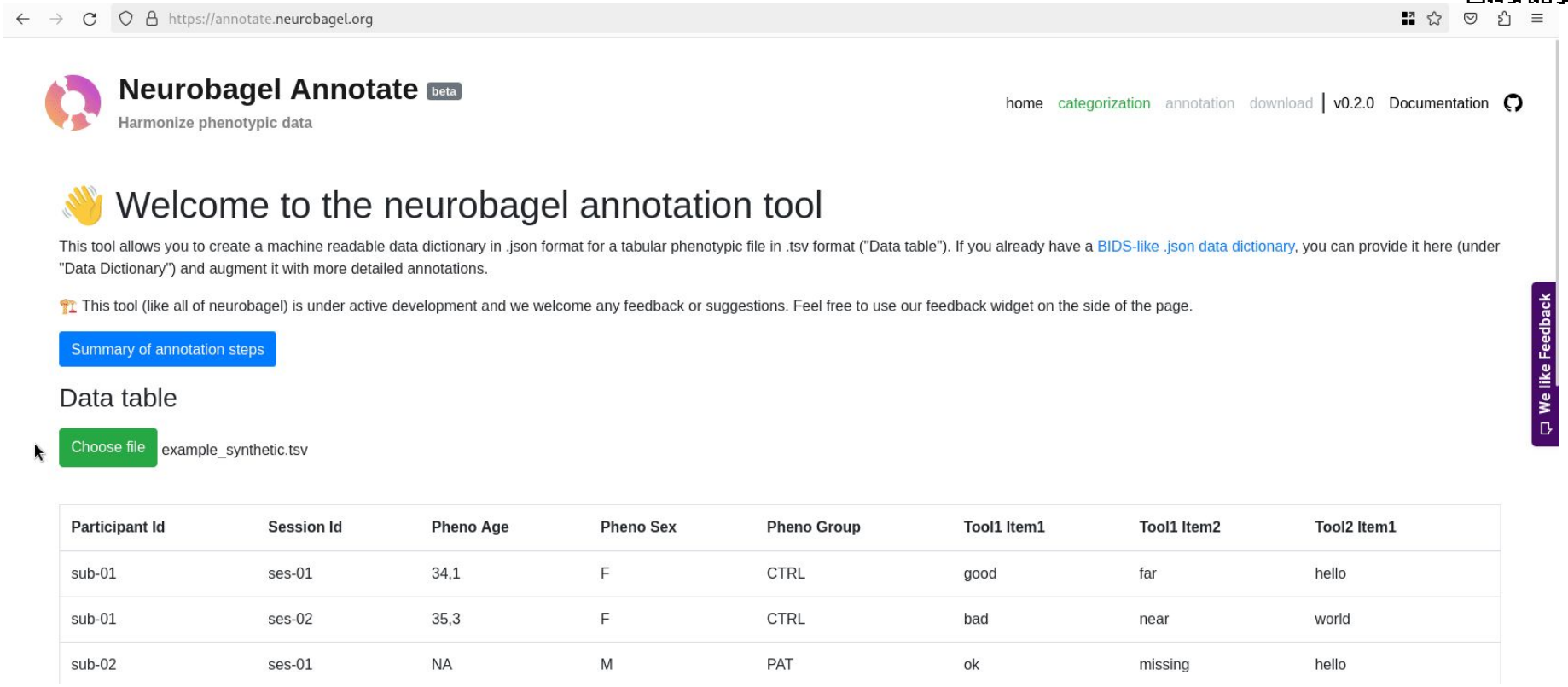


Common participant level data model



SNOMED CT
The global language of healthcare

Intuitive, web-based annotation tool

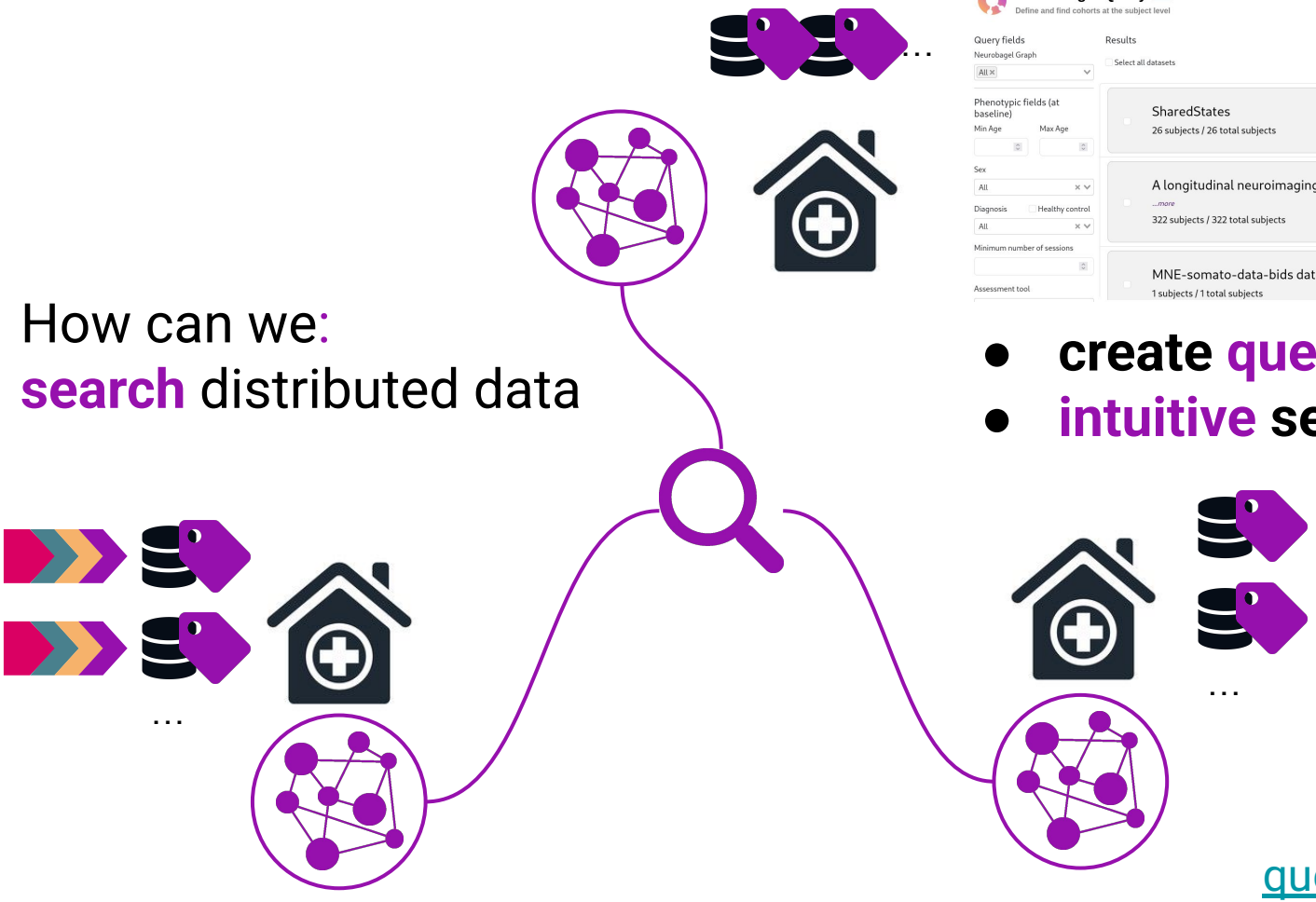


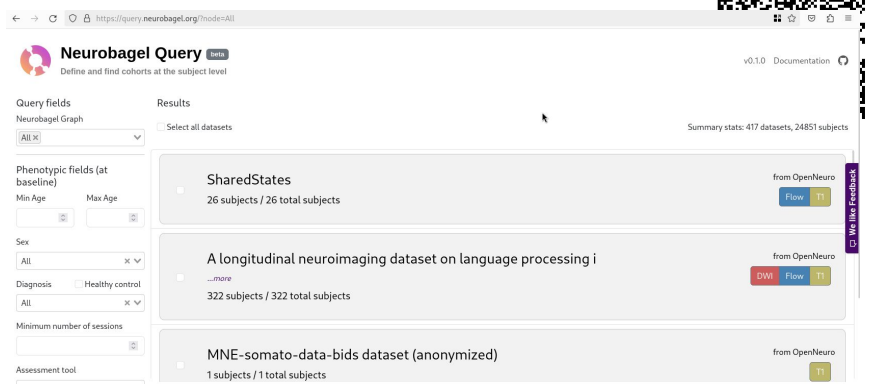
Participant Id	Session Id	Pheno Age	Pheno Sex	Pheno Group	Tool1 Item1	Tool1 Item2	Tool2 Item1
sub-01	ses-01	34,1	F	CTRL	good	far	hello
sub-01	ses-02	35,3	F	CTRL	bad	near	world
sub-02	ses-01	NA	M	PAT	ok	missing	hello

https://annotate.neurobagel.org/

How can we:
search distributed data

How can we:
search distributed data

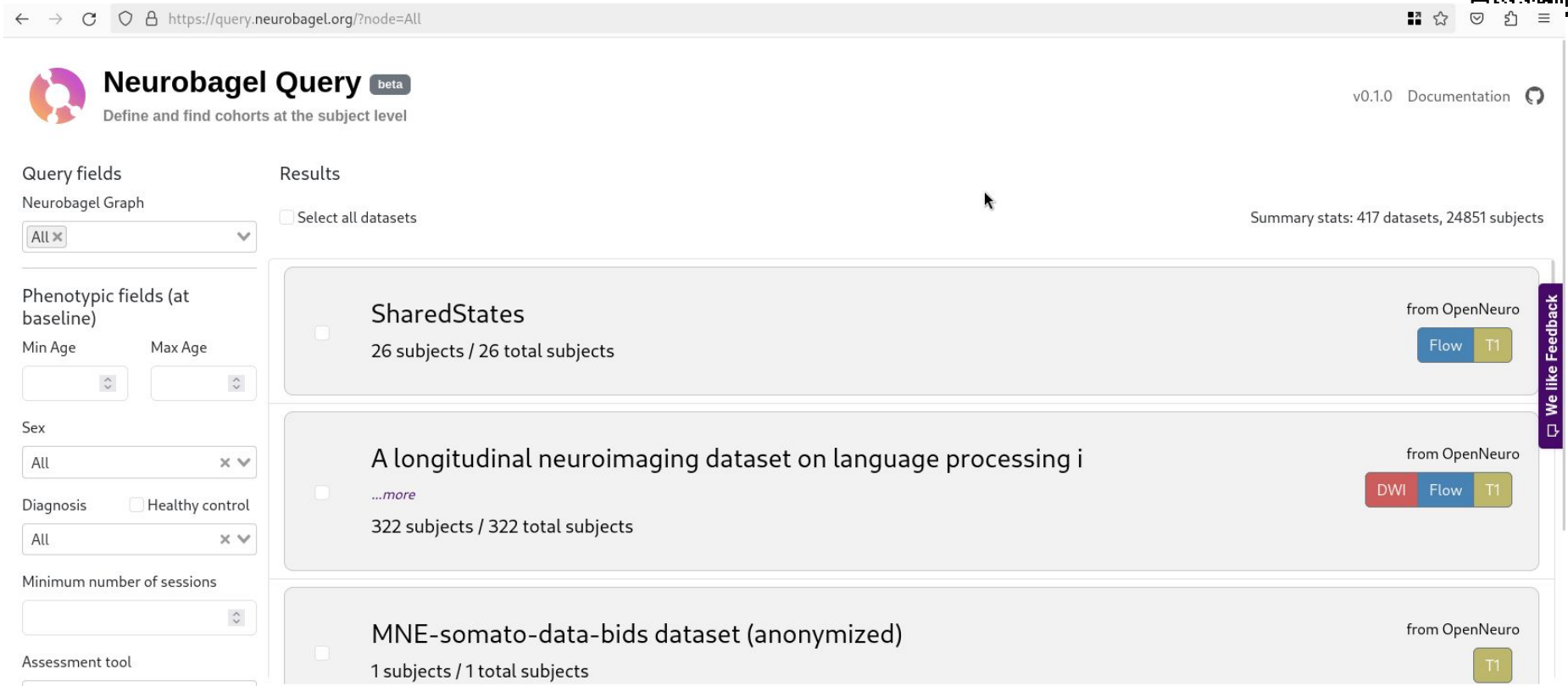




- **create query federation**
- **intuitive search interface**

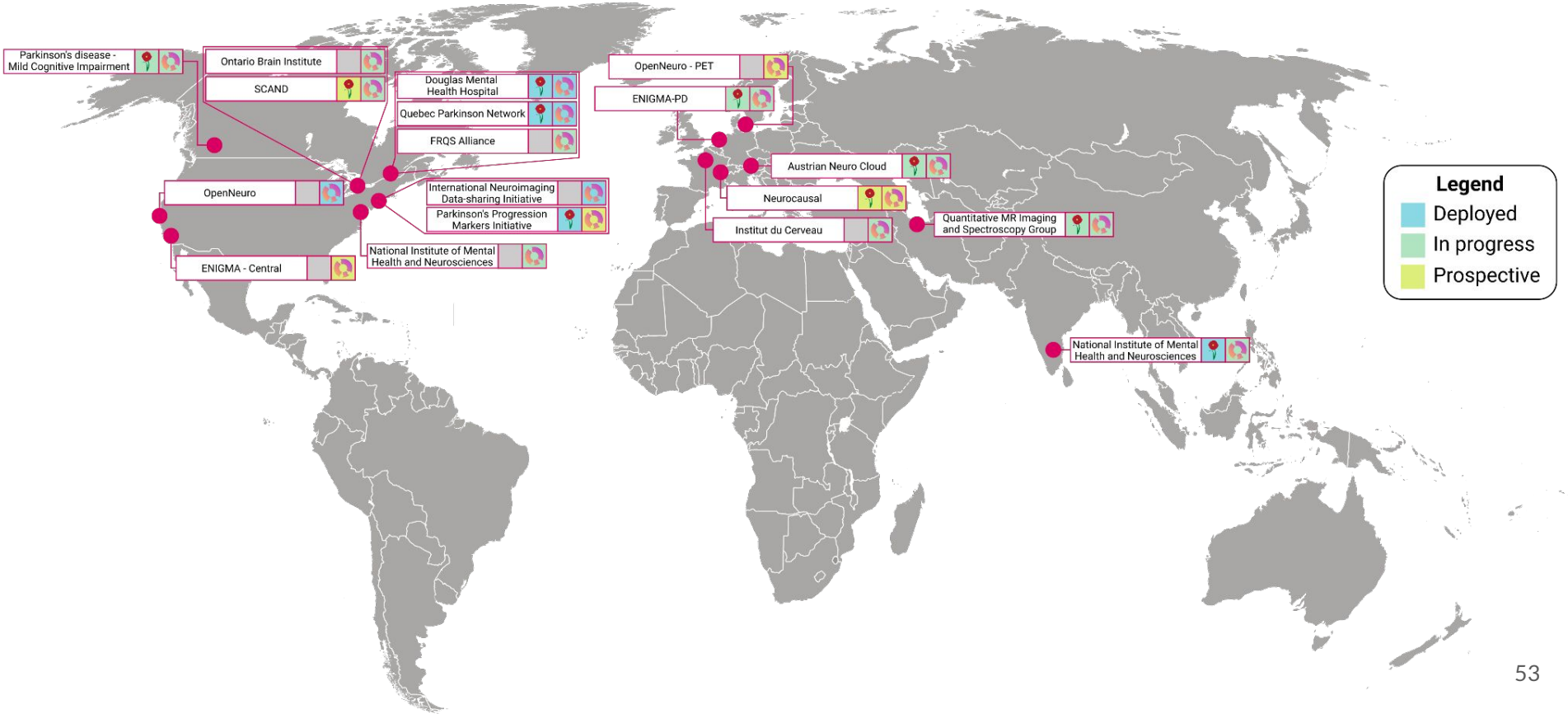
query.neurobagel.org/

Web-based federated query tool



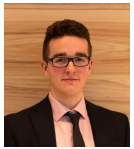
query.neurobagel.org/

Nipoppy and Neurobagel adoption so far



Thank you!

Michelle Wang


Mathieu Dugré

Brent McPherson

Nikhil Bhagwat

Arman Jahanpour

Sebastian Urchs

Alyssa Dai

Julia Pfarr

Jean-Baptiste Poline

ORIGAMI Lab

neuro Montreal Neurological Institute-Hospital

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

Québec Fonds de recherche – Nature et technologies; Fonds de recherche – Santé; Fonds de recherche – Société et culture

PARKINSON QUÉBEC

UNIQUE CENTRE

NIH National Institutes of Health

THE MICHAEL J. FOX FOUNDATION FOR PARKINSON'S RESEARCH

McGill UNIVERSITY

Chan Zuckerberg Initiative

HBHL NEUROHUB

QUEBEC PARKINSON NETWORK

Fondation Brain Canada Foundation