



HAL
open science

Sharing research products FAIRly

Camille Maumet

► **To cite this version:**

Camille Maumet. Sharing research products FAIRly. Doctoral. OHBM educational course 2022: Cultivating open science practices in academic research and culture, Glasgow, United Kingdom. 2022. inserm-03698730

HAL Id: inserm-03698730

<https://inserm.hal.science/inserm-03698730v1>

Submitted on 18 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Course Cultivating open science practices
in academic research and culture



Image credits: "use all the colors" by Laura Bundesen,
[OHBM Brain Art SIG 2020](#)

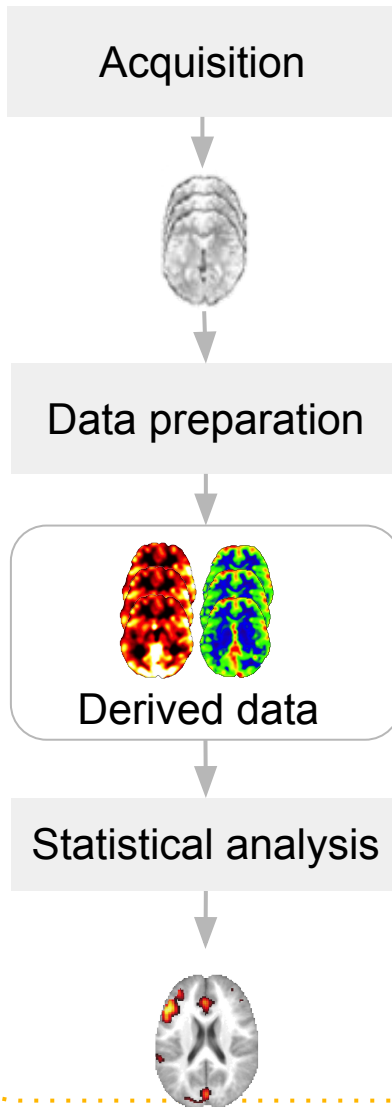
Sharing research products FAIRly

Camille Maumet

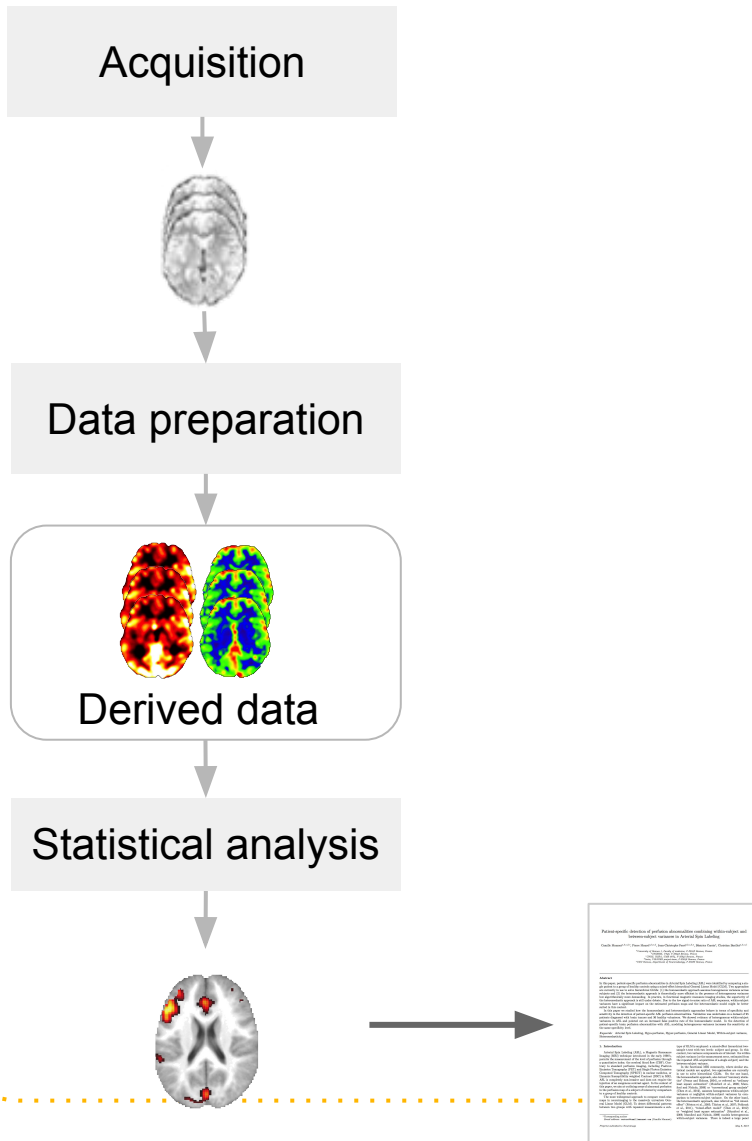
Univ Rennes, Inria, Inserm, CNRS, IRISA

Why sharing different
research products?

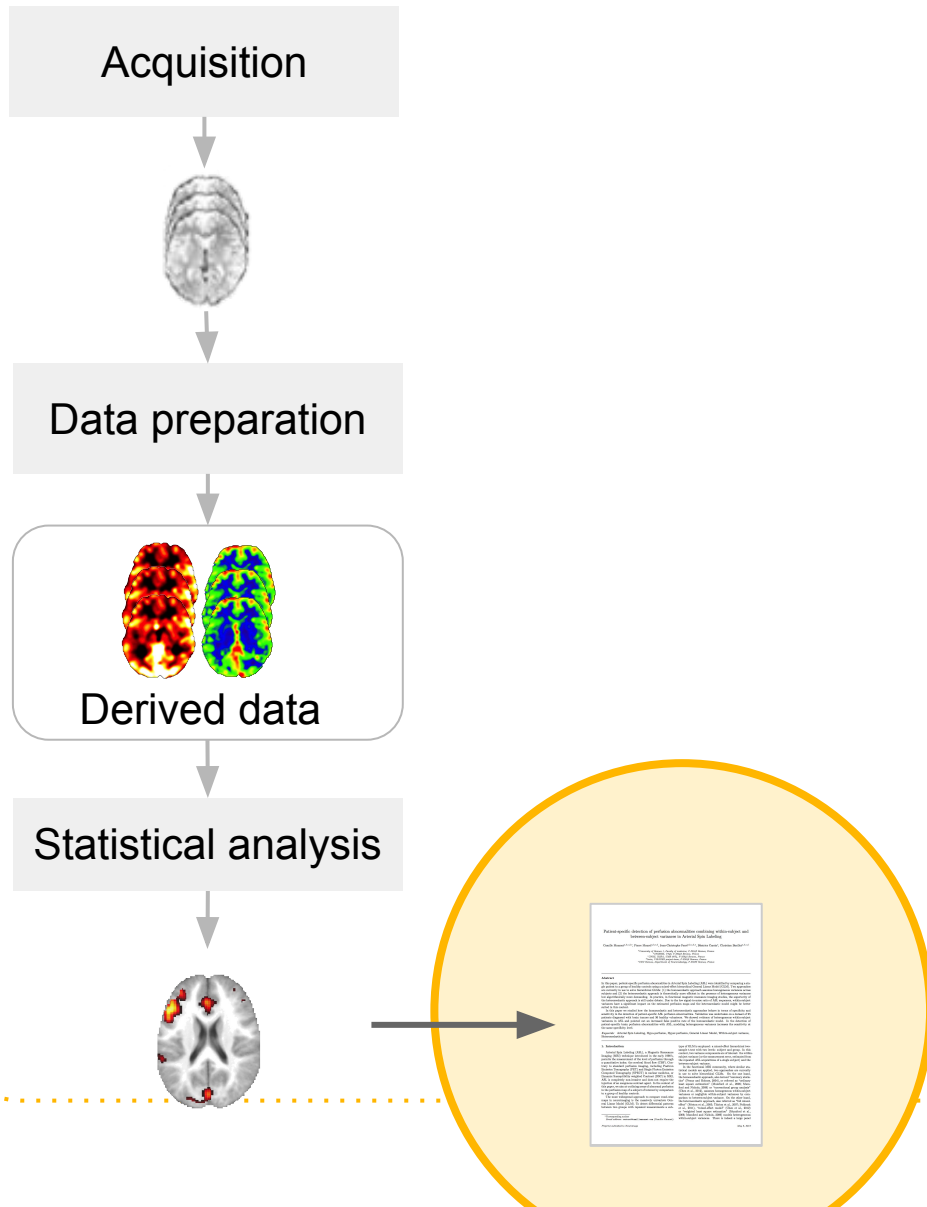
Why sharing research products?



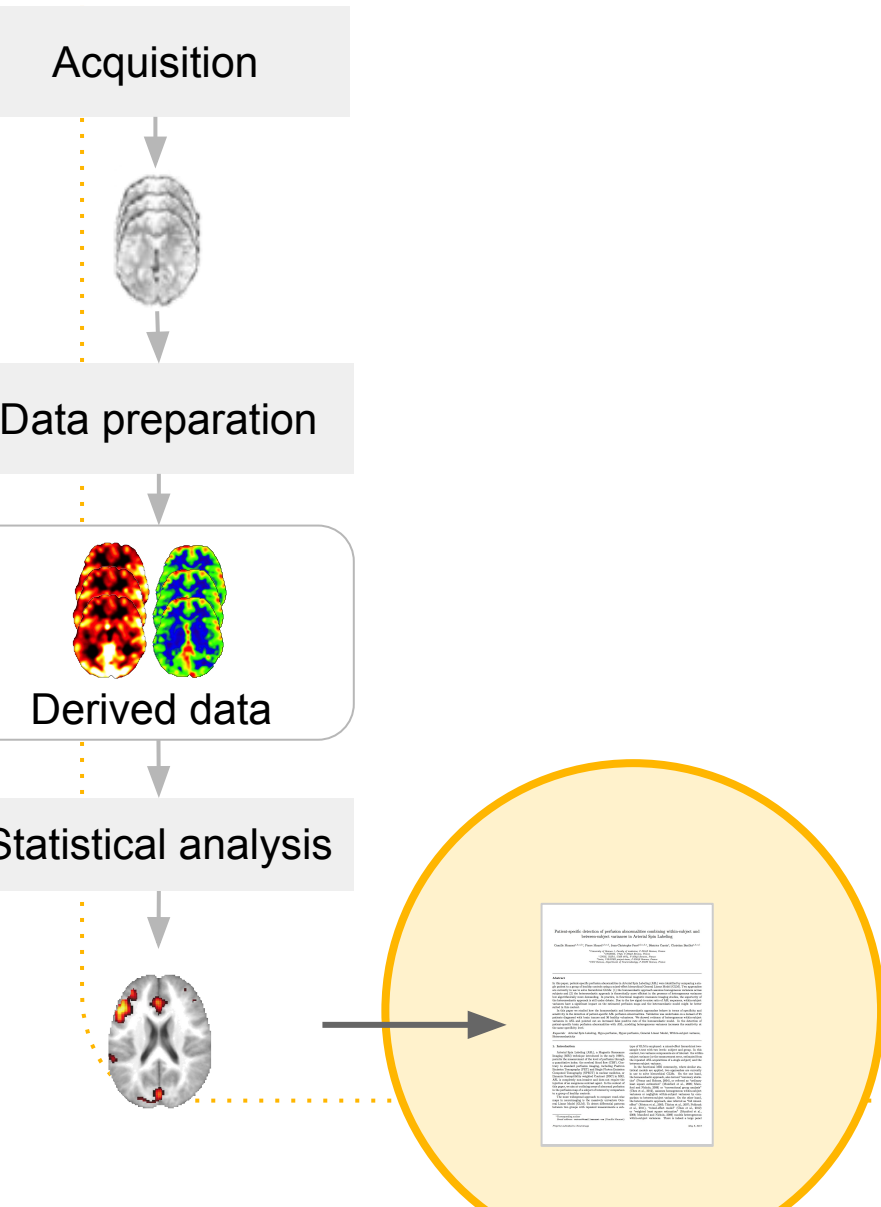
Why sharing research products?



Why sharing research products?



Why sharing research products?



Why sharing research products?

tion

paration

data

analysis



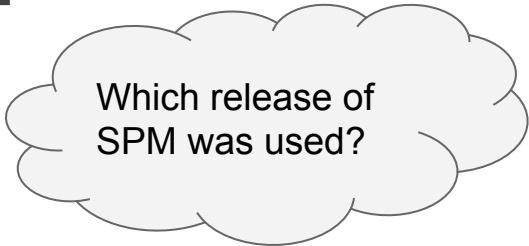
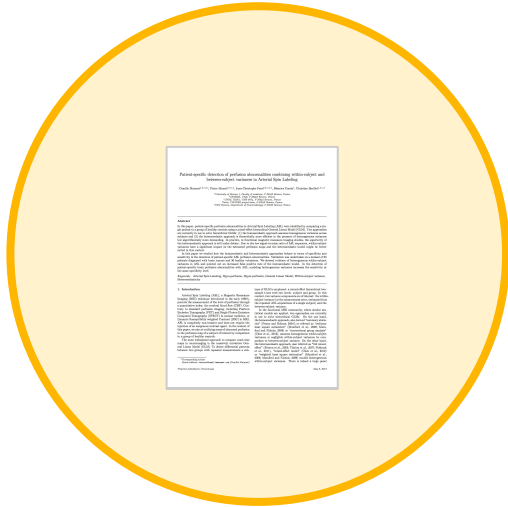
Why sharing research products?



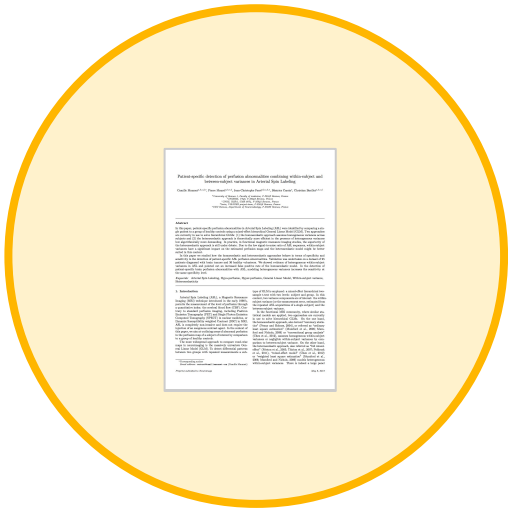
Why sharing research products?



Why sharing research products?



Why sharing research products?



Which release of SPM was used?

Which registration algorithm was selected?

Why sharing research products?



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

Why sharing research products?



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Why sharing research products?



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... I can't reproduce the original results...

Why sharing research products?



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Did I understand the method correctly?

Why sharing research products?



1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Did I understand the method correctly?

... or is this due to my use of a different dataset?

Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

Why sharing research products?



1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Did I understand the method correctly?

... a different computing platform?

... or is this due to my use of a different dataset?

How was the brain mask computed?

Which registration algorithm was selected?

Which release of SPM was used?

Why sharing research products?



1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... **I can't reproduce** the original results...

Did I understand the method correctly?

Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

... or is this due to my use of a different dataset?

... a different computing platform?

... something else!?

Why sharing research products?



Which release of SPM was used?

Which registration algorithm was selected?

How was the brain mask computed?

1. It is a lot of work to restart from zero when trying to build on the results of a scientific paper

Hum... I can't reproduce the original results...

Did I understand the method correctly?

... or is this due to my use of a different dataset?

... a different computing platform?

... something else!?

2. When results cannot be reproduced, it is often very difficult to understand why this is happening

How to share research
products FAIRly

FAIR neuroimaging

Findable **A**ccessible **I**nteroperable **R**eusable



FAIR neuroimaging

Findable **A**ccessible **I**nteroperable **R**eusable



Technical solutions: Standards, Databases,
Good practices (version control, etc.)

FAIR neuroimaging

Findable **A**ccessible **I**nteroperable **R**eusable

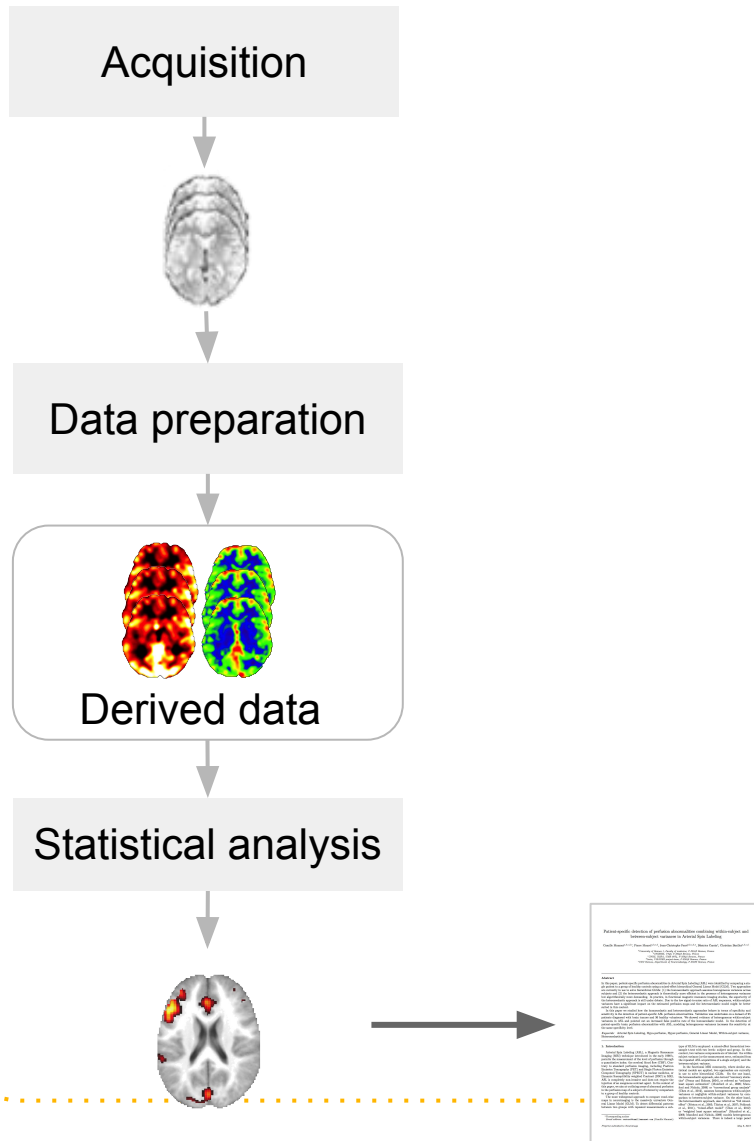


Technical solutions: Standards, Databases,
Good practices (version control, etc.)

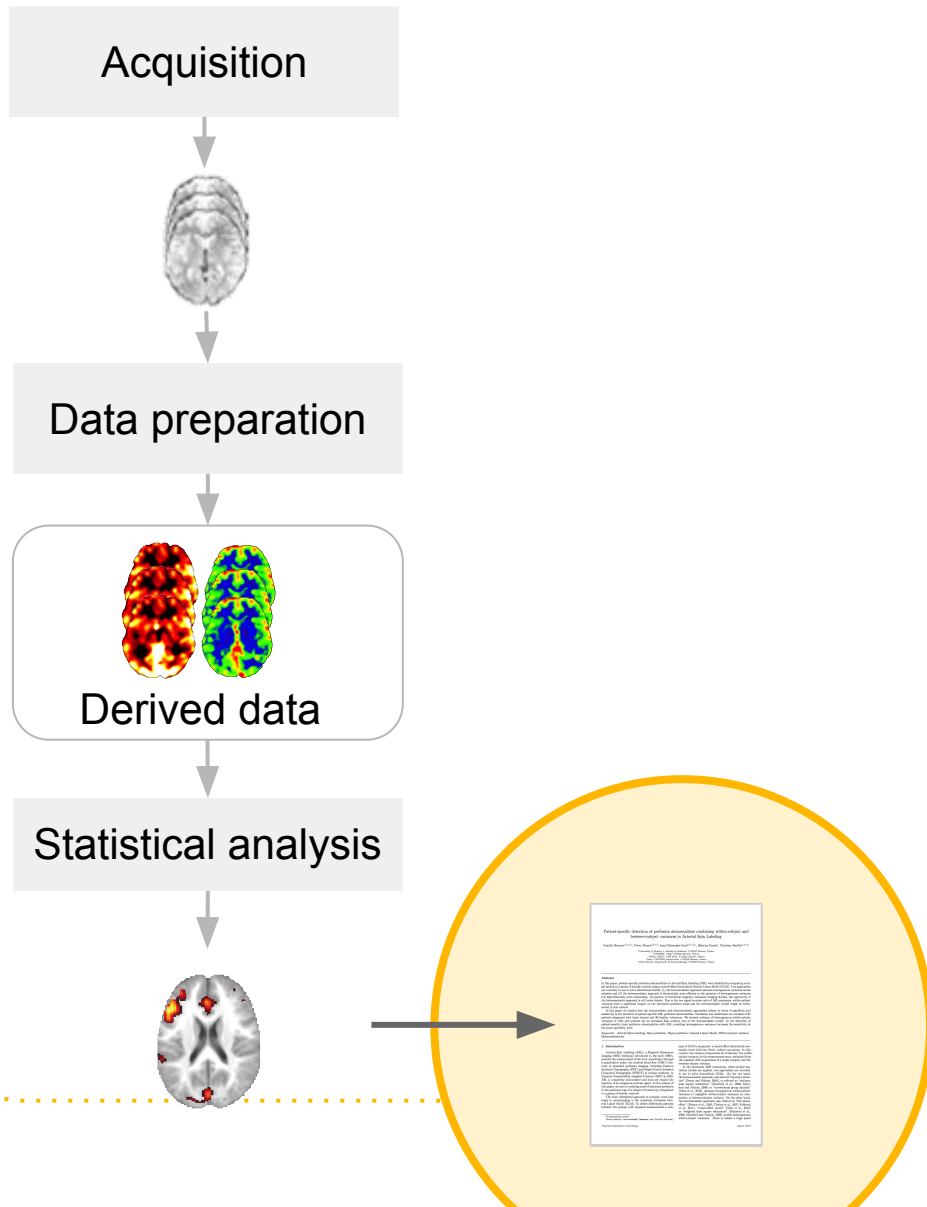
A culture change: towards community-driven research

- Making resources **available to the scientific community**
- **Reusing** existing resources (datasets, tools) that were created by others

Sharing FAIR research products



Sharing FAIR research products

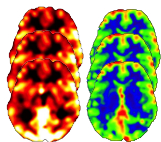


Sharing FAIR research products

Acquisition

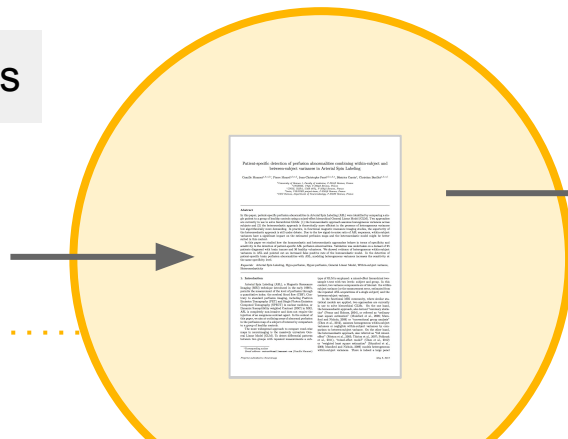
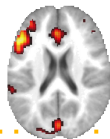


Data preparation

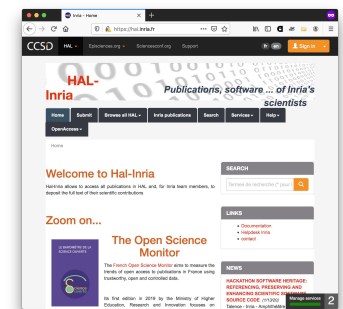


Derived data

Statistical analysis



Biorxiv, arxiv,
HAL,
Pubmed



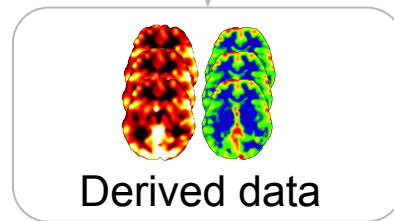
Credits: [Parcels 1 2 & 4 \(CCo\)](#), [Parcel 3 \(CCo\)](#), [Parcel 5 \(CCo\)](#).

Sharing FAIR research products

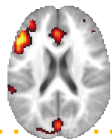
Acquisition



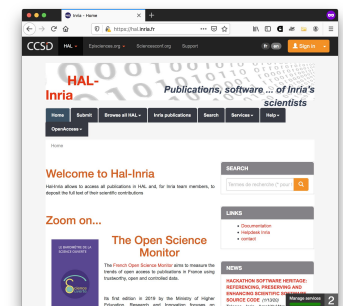
Data preparation



Statistical analysis



Biorxiv, arxiv,
HAL,
Pubmed



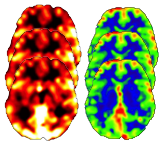
Credits: [Parcels 1 2 & 4 \(CC0\)](#), [Parcel 3 \(CC0\)](#), [Parcel 5 \(CC0\)](#).

Sharing FAIR research products

Acquisition

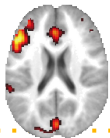


Data preparation

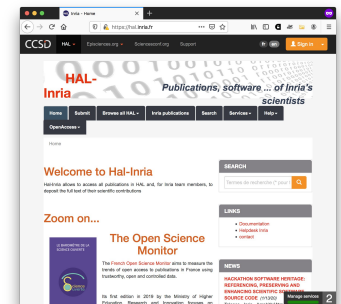


Derived data

Statistical analysis



Biorxiv, arxiv,
HAL,
Pubmed



Credits: [Parcels 1 2 & 4 \(CC0\)](#), [Parcel 3 \(CC0\)](#), [Parcel 5 \(CC0\)](#).

Sharing FAIR datasets



Sharing FAIR datasets



Donders Repository



Sharing FAIR datasets



Donders Repository



Findable we can search for open datasets in each repository

Sharing FAIR datasets



Donders Repository



Findable we can search for open datasets in each repository

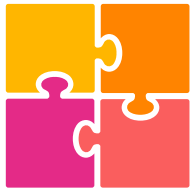


Accessible Direct download (e.g. *Openneuro*) or with Data User Agreement (e.g. *Donders*)

Sharing FAIR datasets



Donders Repository



Interoperable?



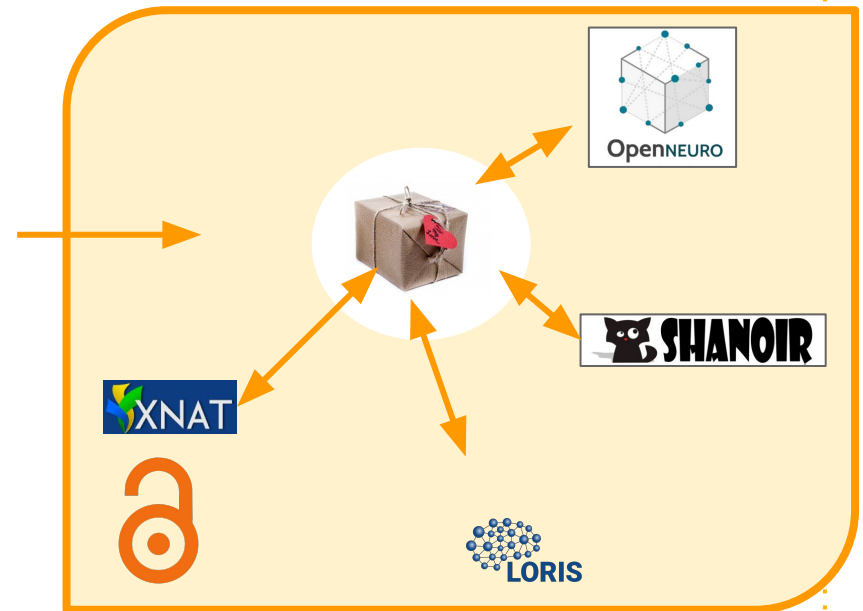
Findable we can search for open datasets in each repository



Accessible Direct download (e.g. *Openneuro*) or with Data User Agreement (e.g. *Donders*)

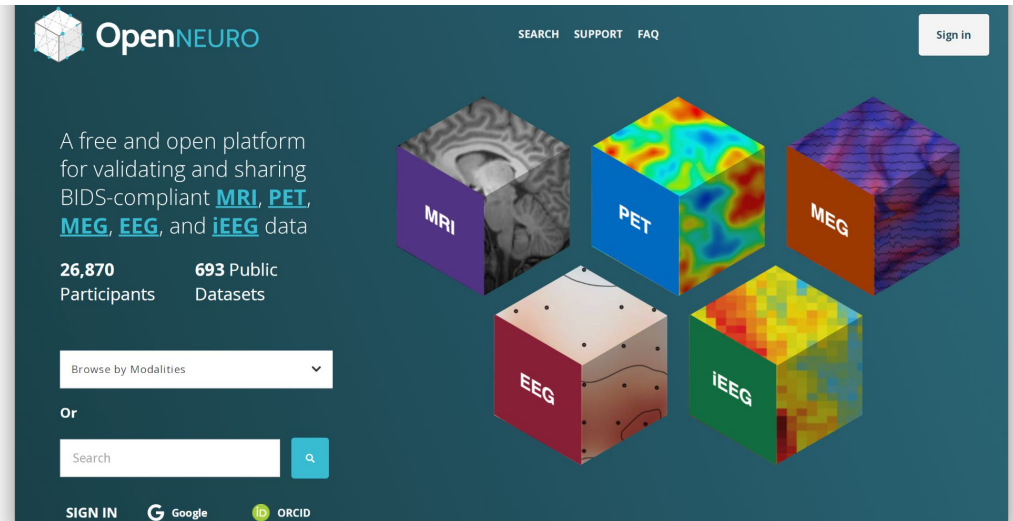
Standards for neuroimaging data sharing

- Data and **meta-data**
- A common “language”



Brain Imaging Data Structure

```
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/
```



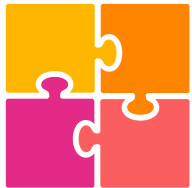
<https://openneuro.org/>

[Gorgolewski et. al, Scientific Data 2016]

Sharing FAIR datasets



Donders Repository



Interoperable?



Findable we can search for open datasets in each repository

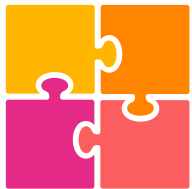


Accessible Direct download (e.g. *Openneuro*) or with Data User Agreement (e.g. *Donders*)

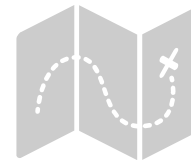
Sharing FAIR datasets



Donders Repository



Interoperable BIDS acts as a common languages across tools



Findable we can search for open datasets in each repository

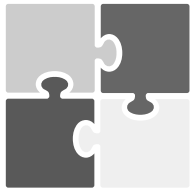


Accessible Direct download (e.g. *Openneuro*) or with Data User Agreement (e.g. *Donders*)

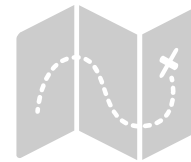
Sharing FAIR datasets



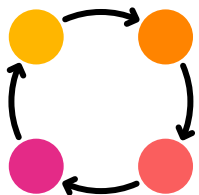
Donders Repository



Interoperable BIDS acts as a common languages across tools



Findable we can search for open datasets in each repository



Reusable Metadata in BIDS were selected to enable reuse. *e.g. onsets of the task for task-based fMRI.*

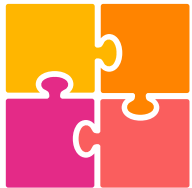


Accessible Direct download (*e.g. Openneuro*) or with Data User Agreement (*e.g. Donders*)

Sharing FAIR datasets



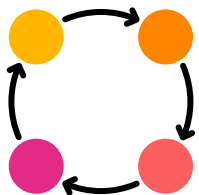
Donders Repository



Interoperable BIDS acts as a common language across tools



Findable we can search for open datasets in each repository



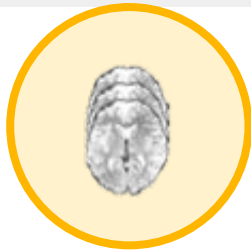
Reusable Metadata in BIDS were selected to enable reuse. *e.g. onsets of the task for task-based fMRI.*



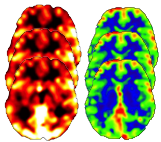
Accessible Direct download (*e.g. Openneuro*) or with Data User Agreement (*e.g. Donders*)

Sharing FAIR research products

Acquisition

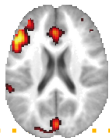


Data preparation

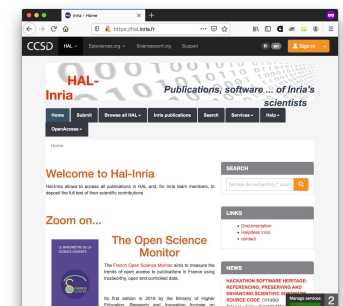


Derived data

Statistical analysis



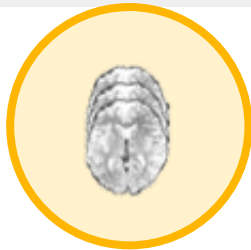
Biorxiv, arxiv,
HAL,
Pubmed



Credits: [Parcels 1 2 & 4 \(CC0\)](#), [Parcel 3 \(CC0\)](#), [Parcel 5 \(CC0\)](#).

Sharing FAIR research products

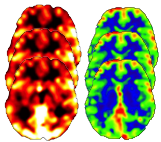
Acquisition



Donders Repository



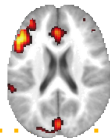
Data preparation



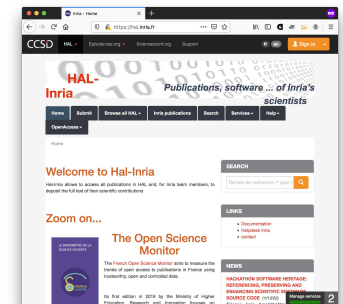
Derived data



Statistical analysis

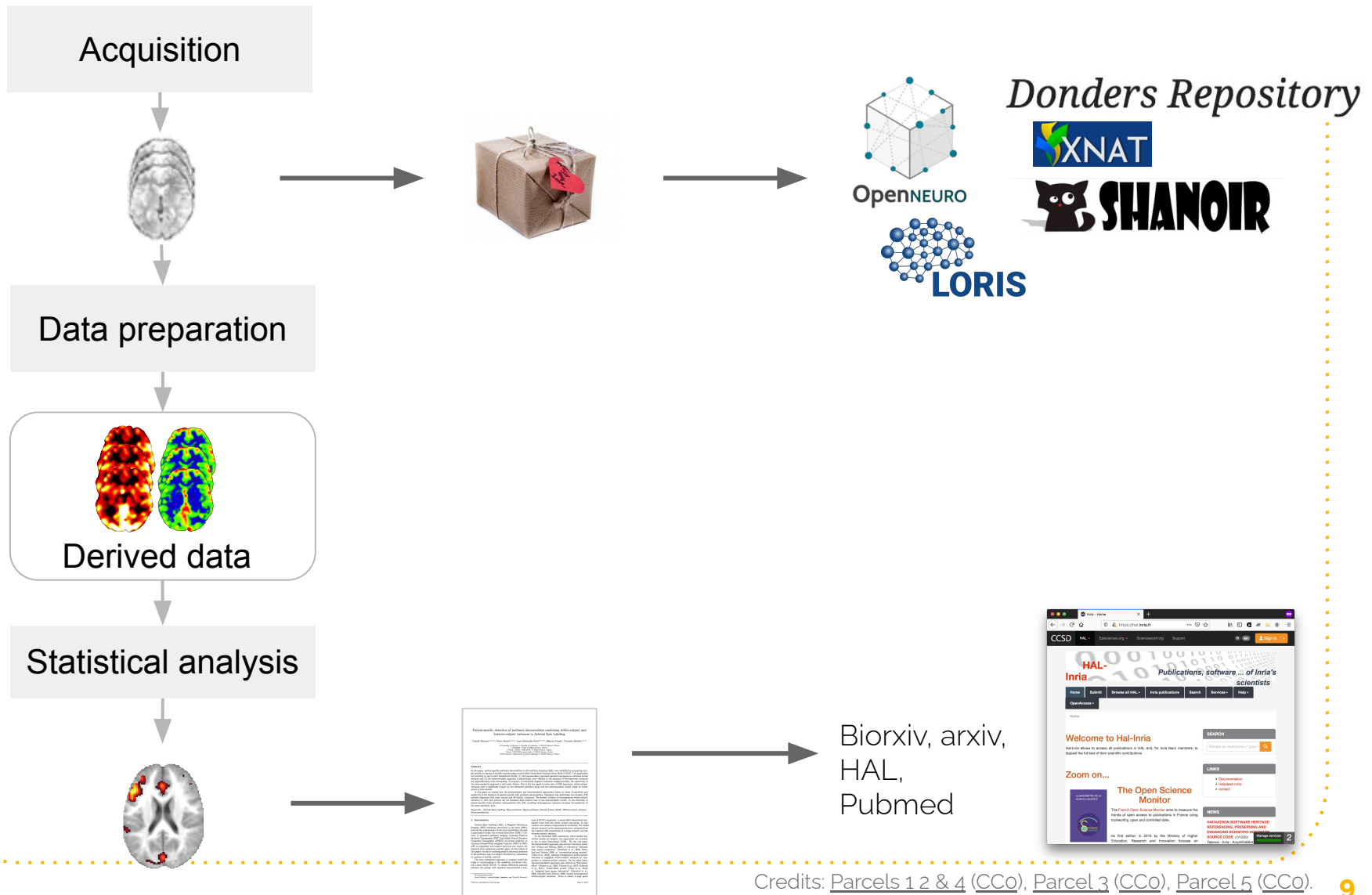


Biorxiv, arxiv,
HAL,
Pubmed

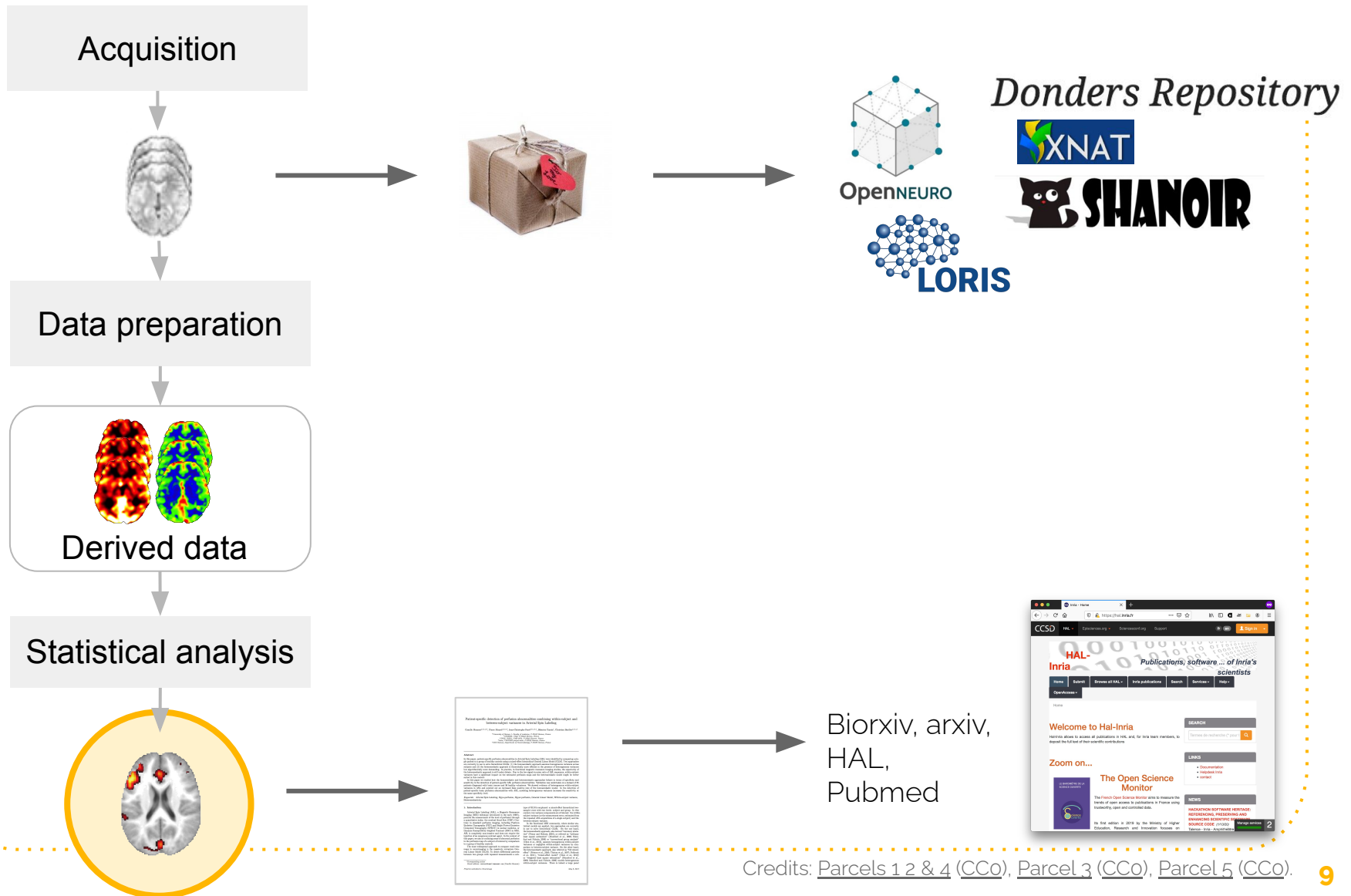


Credits: [Parcels 1 2 & 4 \(CC0\)](#), [Parcel 3 \(CC0\)](#), [Parcel 5 \(CC0\)](#).

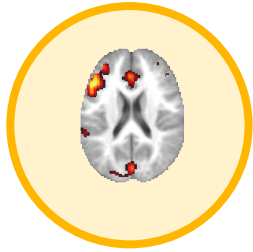
Sharing FAIR research products



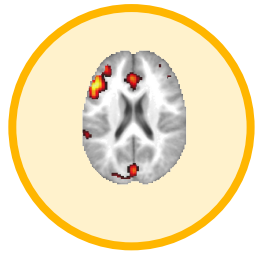
Sharing FAIR research products



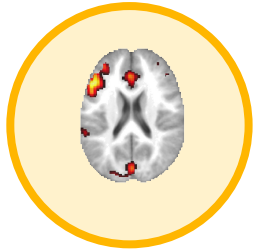
Sharing FAIR results



Sharing FAIR results (task fMRI)



Sharing FAIR results (task fMRI)

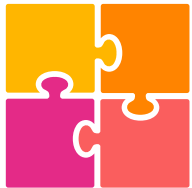
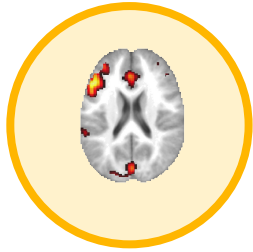


Findable we can search for datasets on NeuroVault

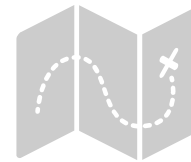


Accessible Direct download

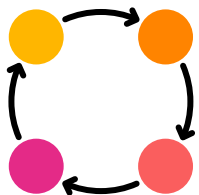
Sharing FAIR results (task fMRI)



Interoperable?



Findable we can search for datasets on NeuroVault

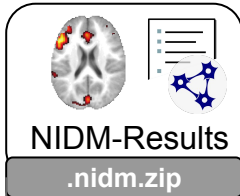


Reusable?



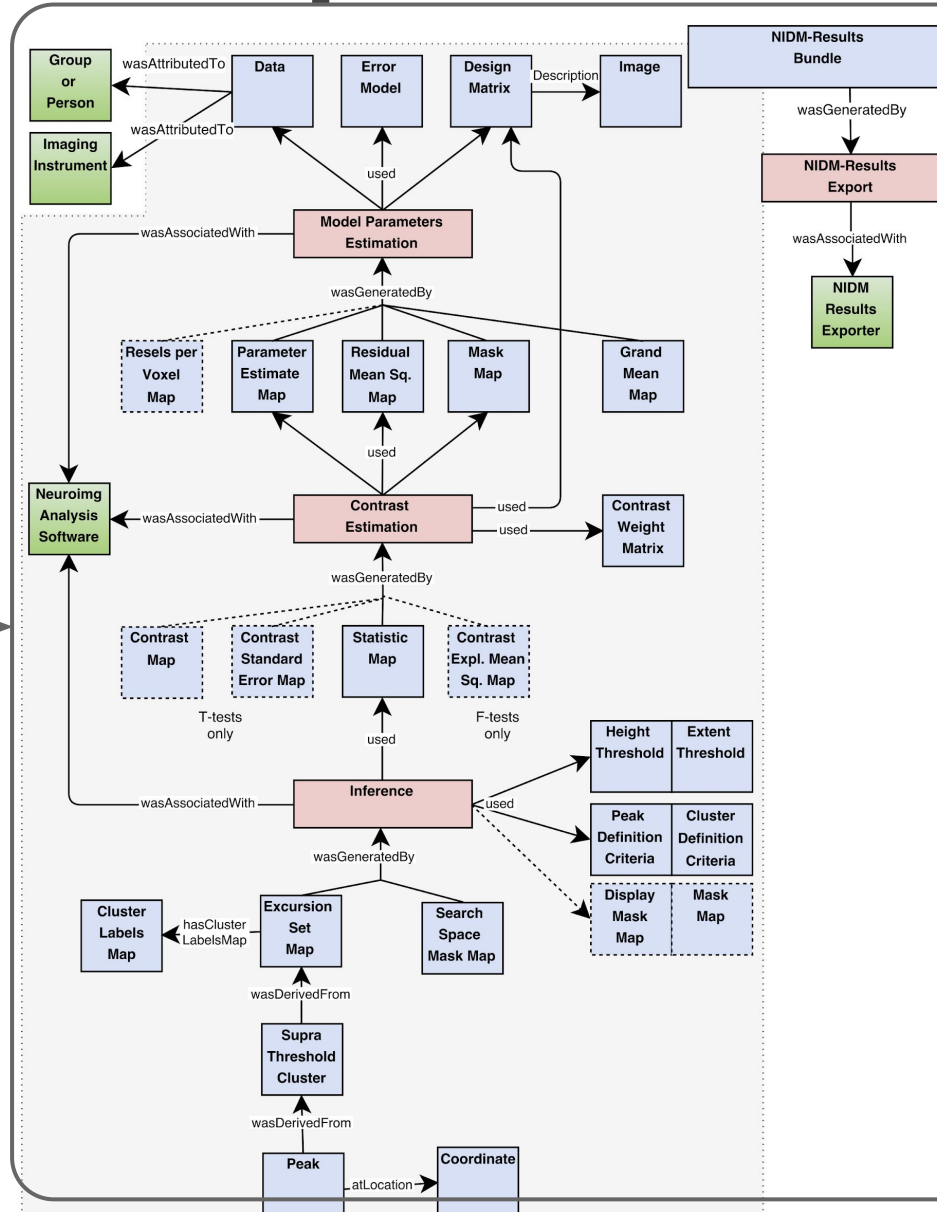
Accessible Direct download

NIDM-Results pack



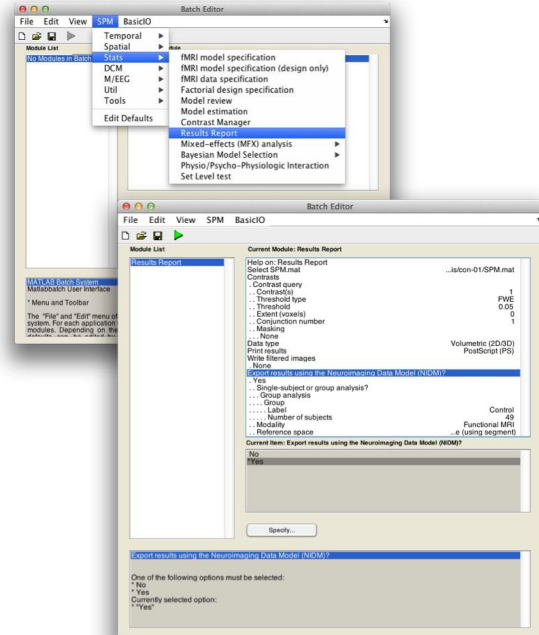
- ClusterLabels.nii.gz
- Contrast.nii.gz
- ContrastStandardError.nii.gz
- DesignMatrix.csv
- DesignMatrix.png
- ExcursionSet.nii.gz
- ExcursionSet.png
- GrandMean.nii.gz
- Mask.nii.gz
- nidm.json
- nidm.ttl
- ParameterEstimate_001.nii.gz
- ParameterEstimate_002.nii.gz
- ResidualMeanSquares.nii.gz
- SearchSpaceMask.nii.gz
- TStatistic.nii.gz
- ZStatistic.nii.gz

[Maumet et al.,
Sci. Data 2016]

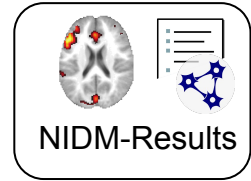


NIDM in SPM, FSL & NeuroVault

SPM



2. Publication on NeuroVault



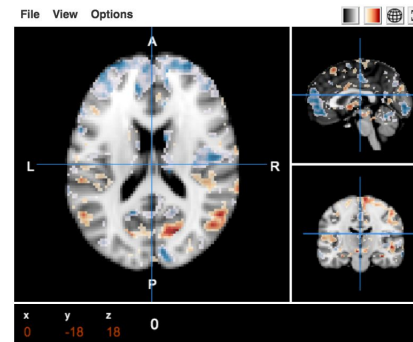
A Correspondence between Individual Differences in the Brain's Intrinsic Functional Architecture and the Content and Form of Self-Generated Thoughts

Contributed by ChrisFilioGorgolewski

Krzysztof J. Gorgolewski, Dan Lurie, Sebastian Urchs, Judy A. Kipping, R. Cameron Craddock, Michael P. Milham, Daniel S. Margulies, Jonathan Smallwood

[Link to the paper](#)

3D View



Images Details

Show 7 entries Search:

View	ID	Name	Type
	25	fALFF: Past	Z map
	26	fALFF: Future	Z map
	27	fALFF: Past > Future	Z map
	28	fALFF: Positive	Z map
	29	fALFF: Negative	T map
	30	fALFF: Positive > Negative	Z map
	31	fALFF: Social Cognition	Z map

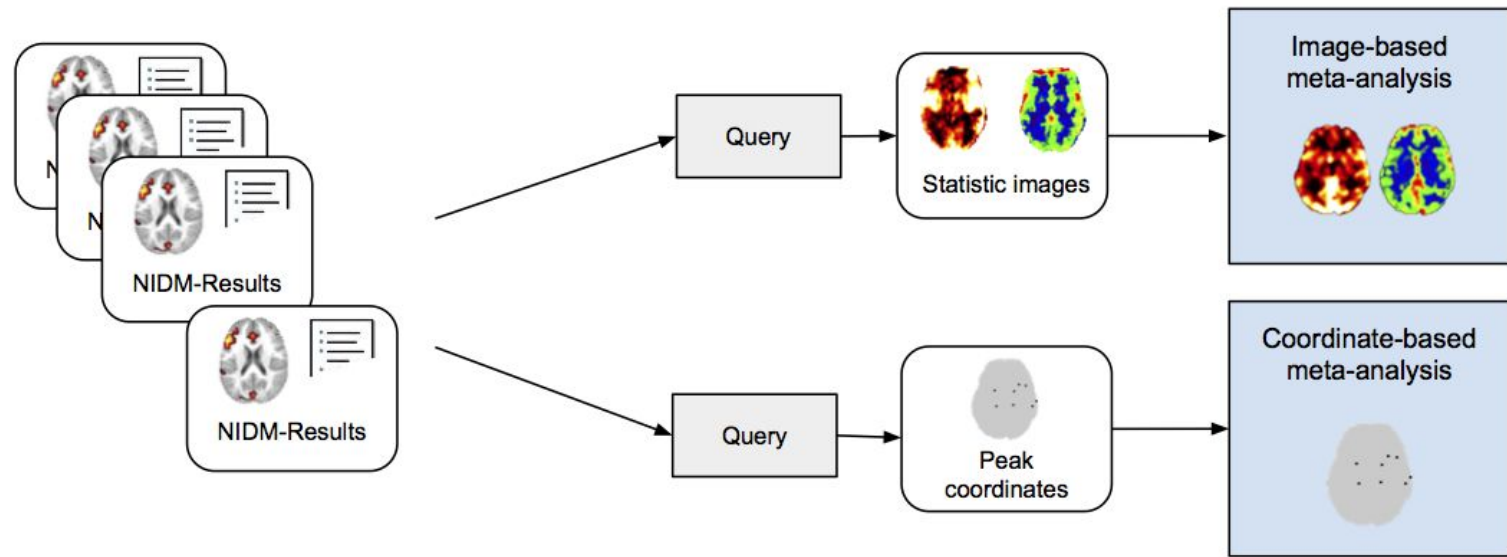
Showing 1 to 7 of 30 entries First Previous Next Last

1. Export



```
$ nidmfsfsl fsl_ds107_group 49 -g Control
```

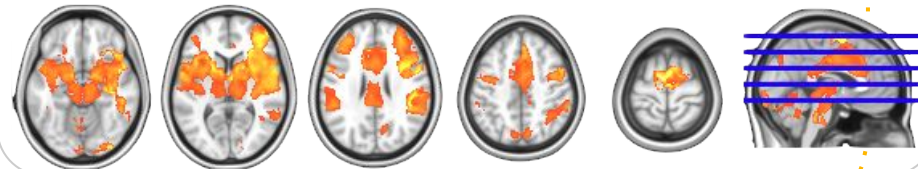
Meta-analysis w/ NIDM-Results



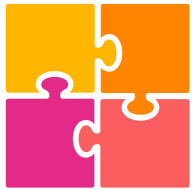
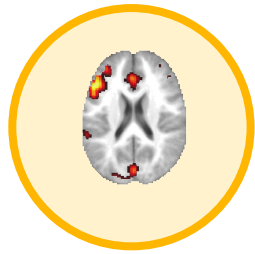
Coordinate-based meta-analysis



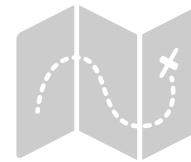
Image-based meta-analysis



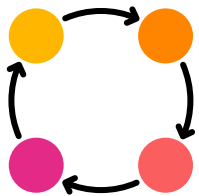
Sharing FAIR results (task fMRI)



Interoperable NIDM-Results acts as a common languages across tools



Findable we can search for datasets on NeuroVault

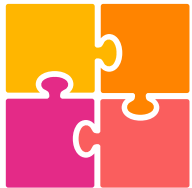


Reusable Metadata in NIDM-Results selected to enable meta-analyses



Accessible Direct download

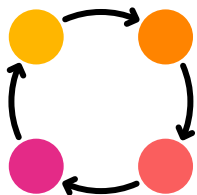
Sharing FAIR results (task fMRI)



Interoperable NIDM-Results acts as a common languages across tools



Findable we can search for datasets on NeuroVault

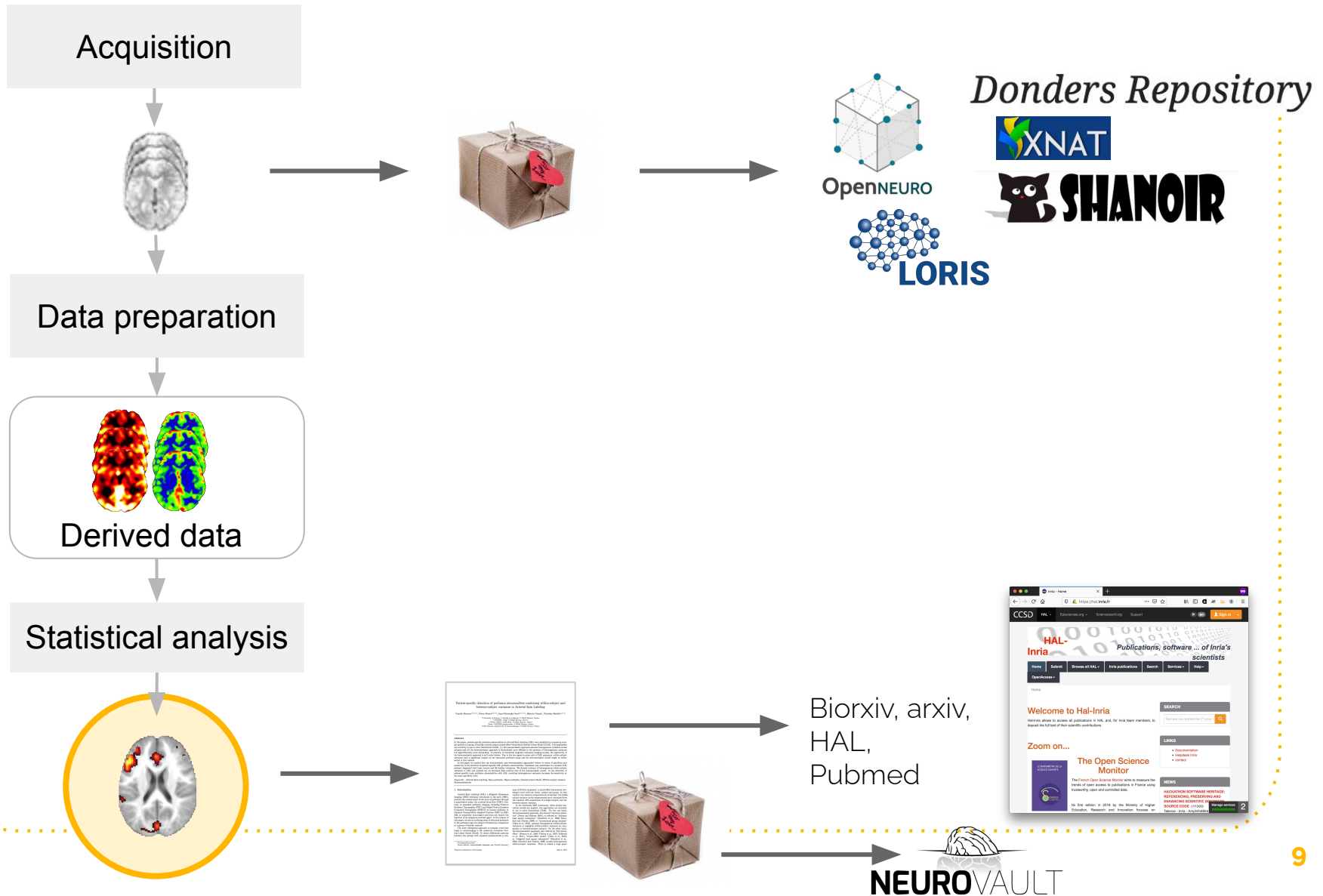


Reusable Metadata in NIDM-Results selected to enable meta-analyses

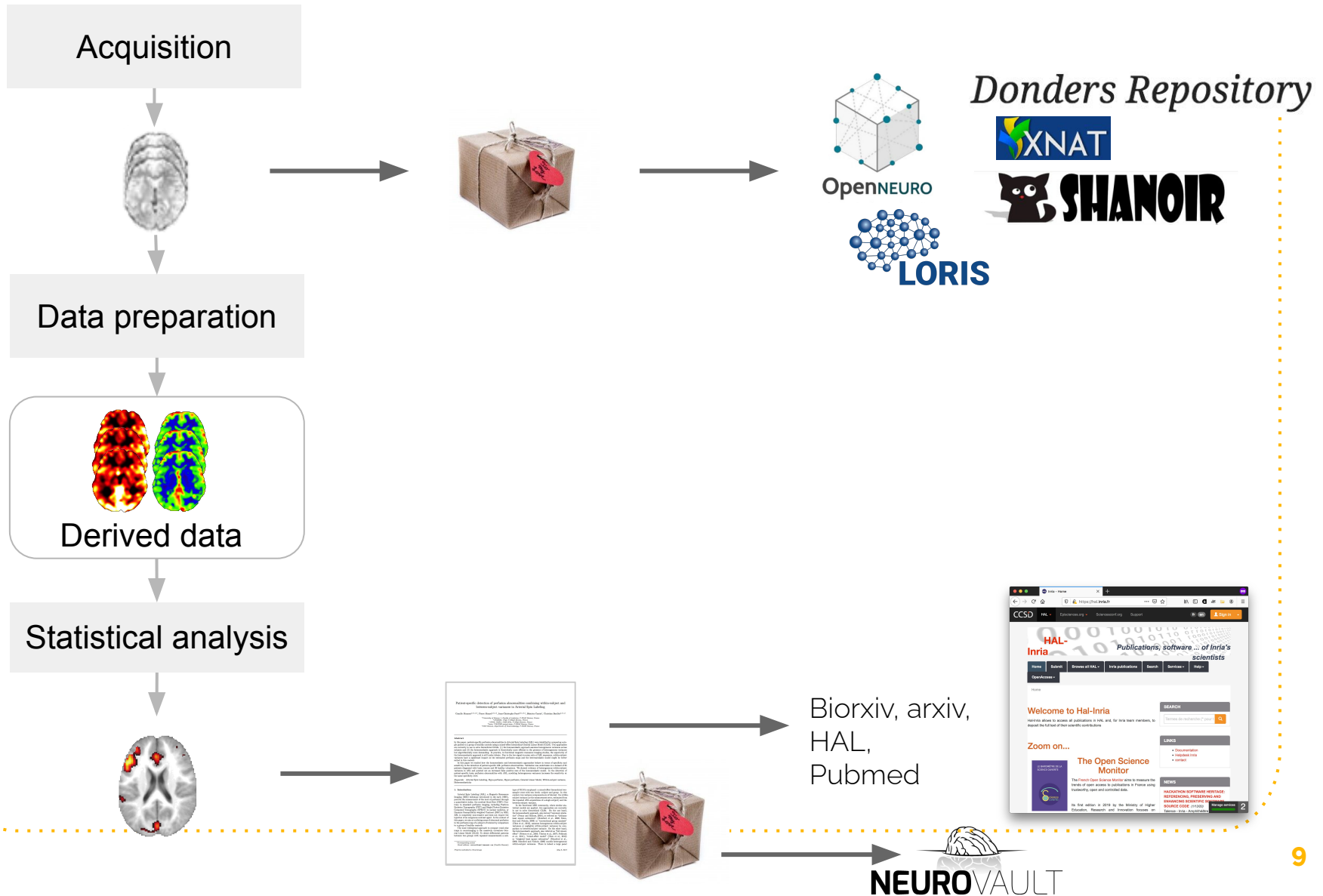


Accessible Direct download

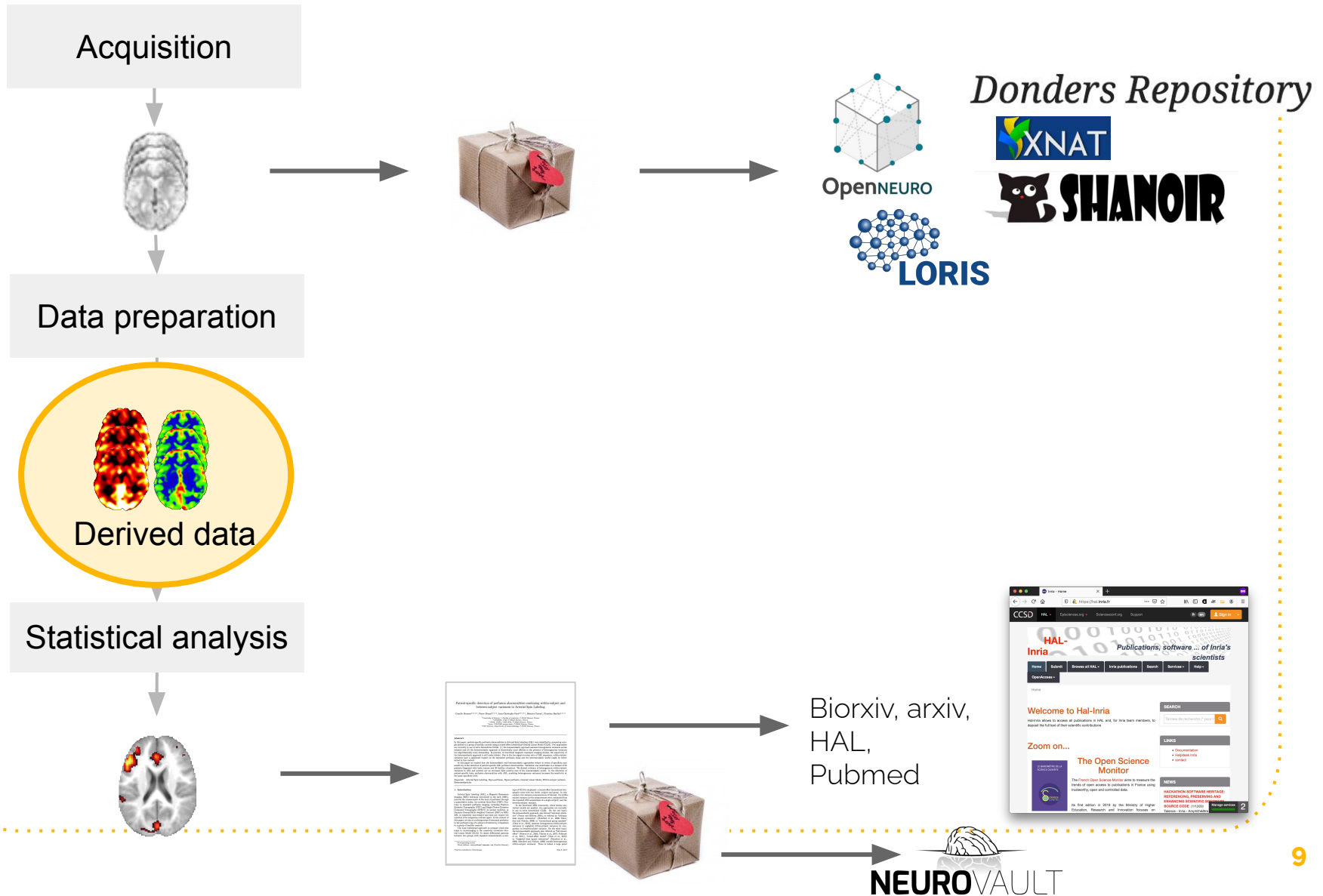
Sharing FAIR research products



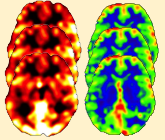
Sharing FAIR research products



Sharing FAIR research products

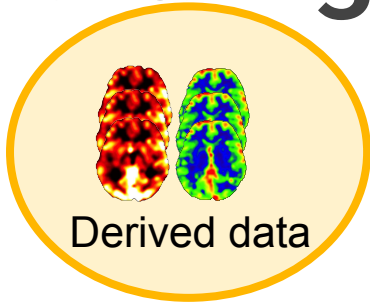


Sharing FAIR intermediary datasets



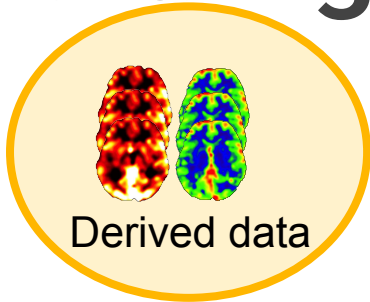
Derived data

Sharing FAIR intermediary datasets



General purpose repositories

Sharing FAIR intermediary datasets



General purpose repositories

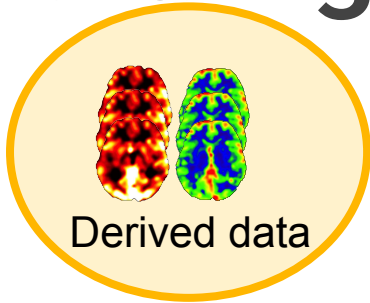


Findable can search for open datasets (less specific queries)



Accessible Direct download or Data User Agreement

Sharing FAIR intermediary datasets



General purpose repositories

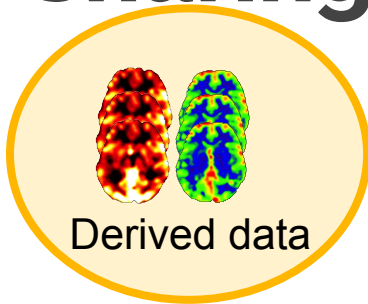


Findable can search for open datasets (less specific queries)

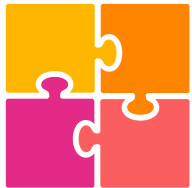


Accessible Direct download or Data User Agreement

Sharing FAIR intermediary datasets



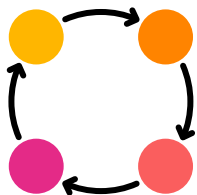
General purpose repositories



Interoperable?



Findable can search for open datasets (less specific queries)



Reusable?



Accessible Direct download or Data User Agreement

Sharing other FAIR datasets

1- Look for existing standards and databases:



2- You can help develop the next standards

BIDS Derivatives

<https://bids.neuroimaging.io/>

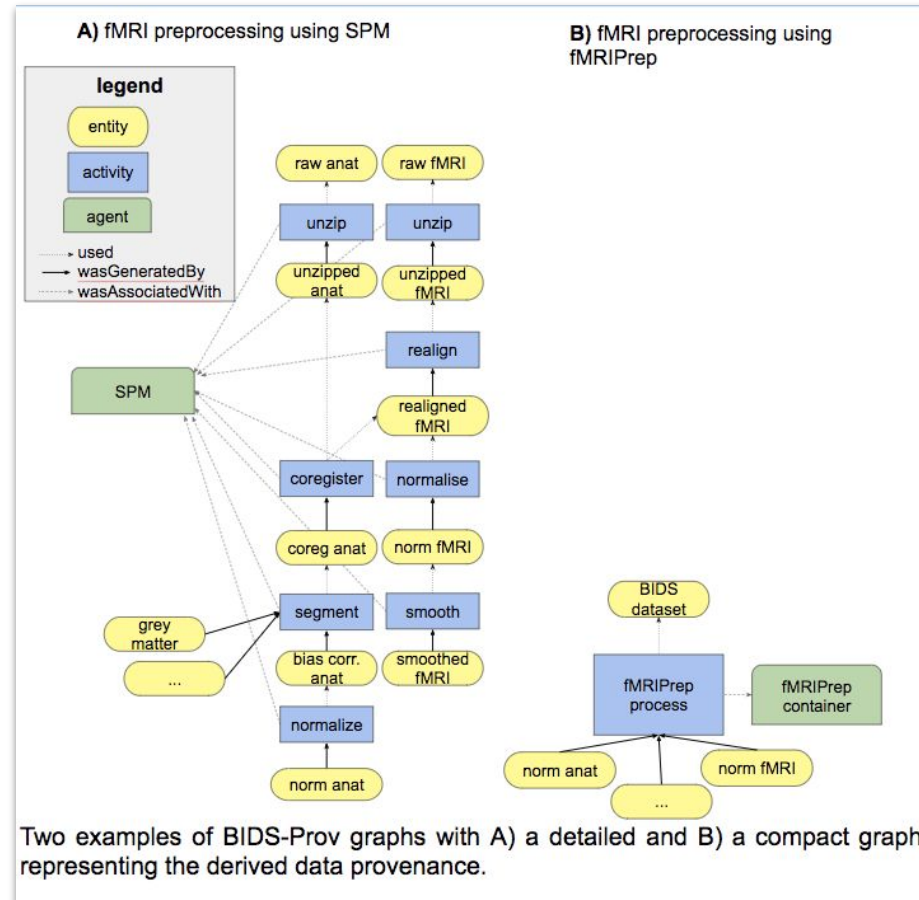
Derivatives are outputs of common processing pipelines, capturing data and meta-data sufficient for a researcher to understand and (critically) reuse those outputs in subsequent processing. Standardizing derivatives is motivated by use cases where formalized machine-readable access to processed data enables higher level processing.

- BEP011 The structural preprocessing derivatives
 - BEP012 The functional preprocessing derivatives
 - BEP013 The resting state fMRI derivatives
 - BEP016 The diffusion weighted imaging derivatives
 - BEP017 Generic BIDS connectivity data schema
 - BEP021 Common Electrophysiological Derivatives
- And more!

BIDS Provenance

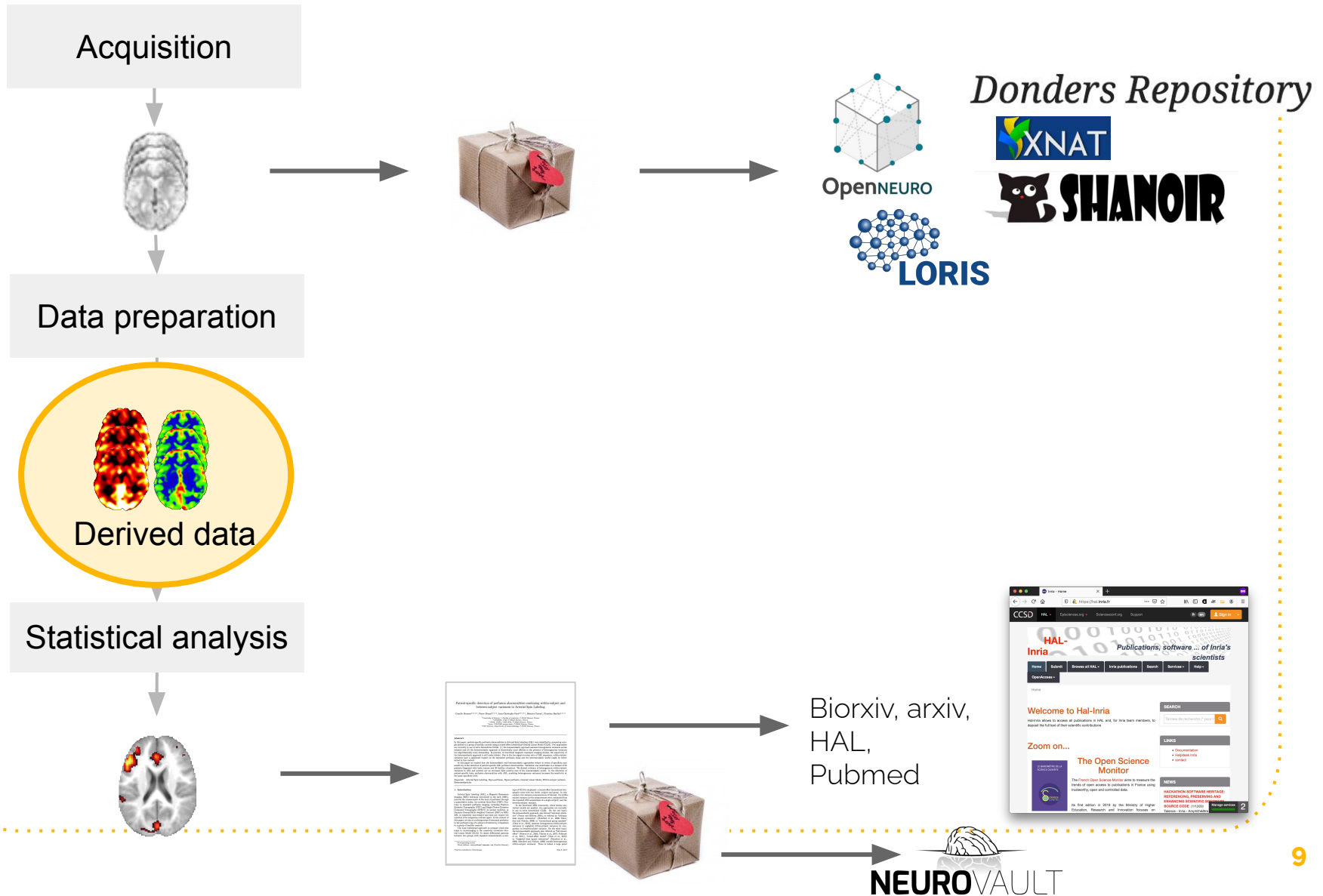


Satra Ghosh

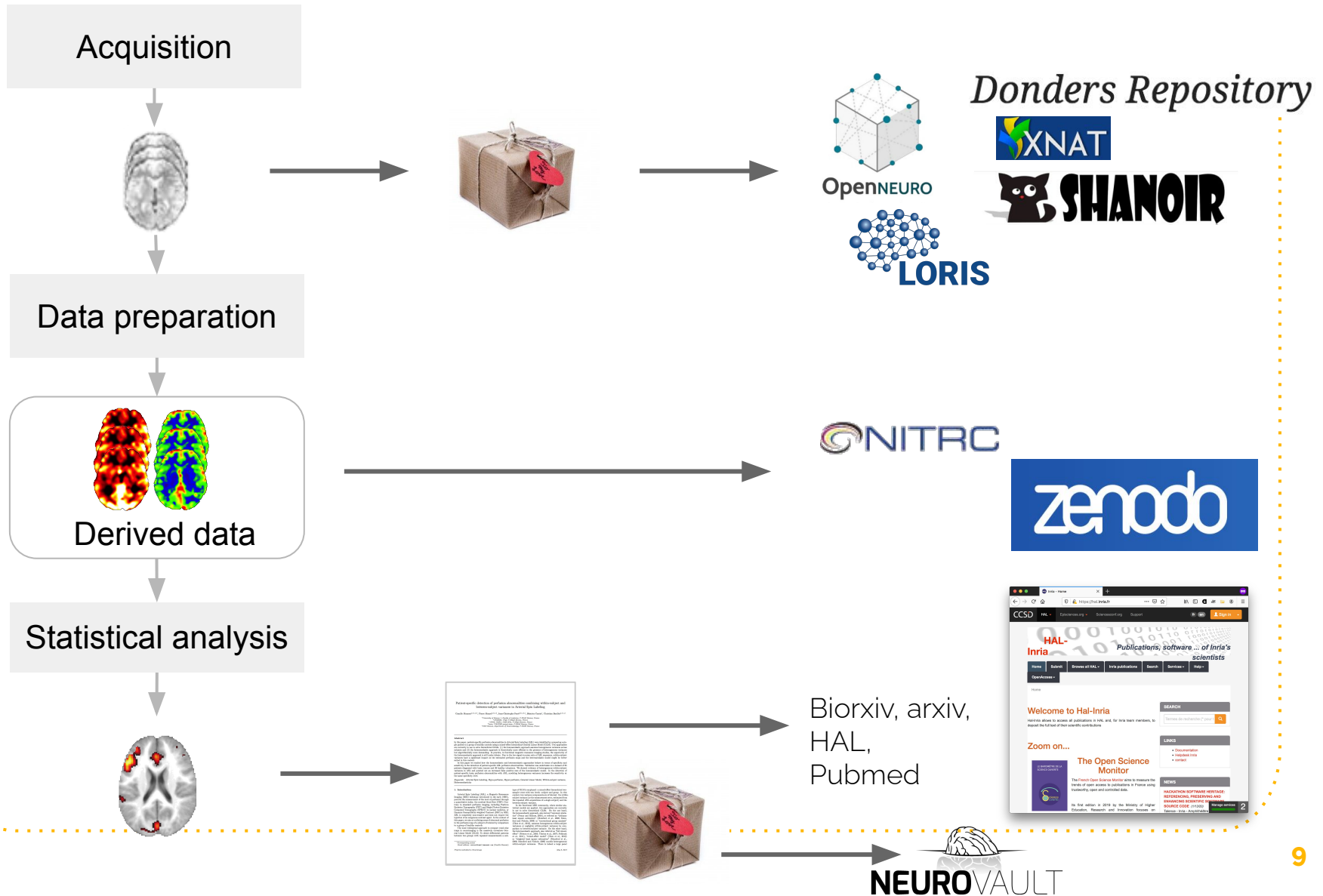


https://github.com/bids-standard/BEP028_BIDSprov

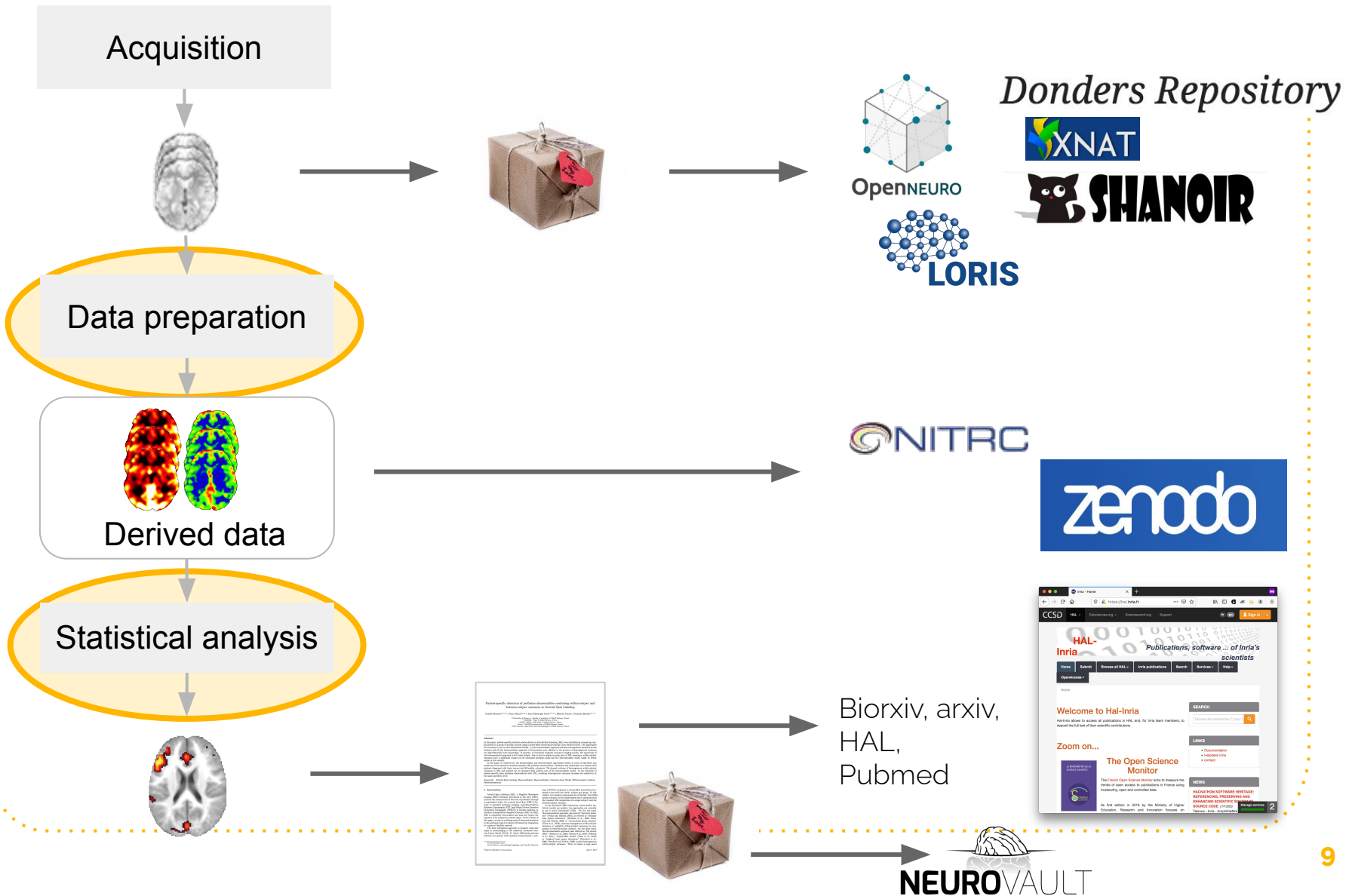
Sharing FAIR research products



Sharing FAIR research products



Sharing FAIR research products



Sharing analysis code

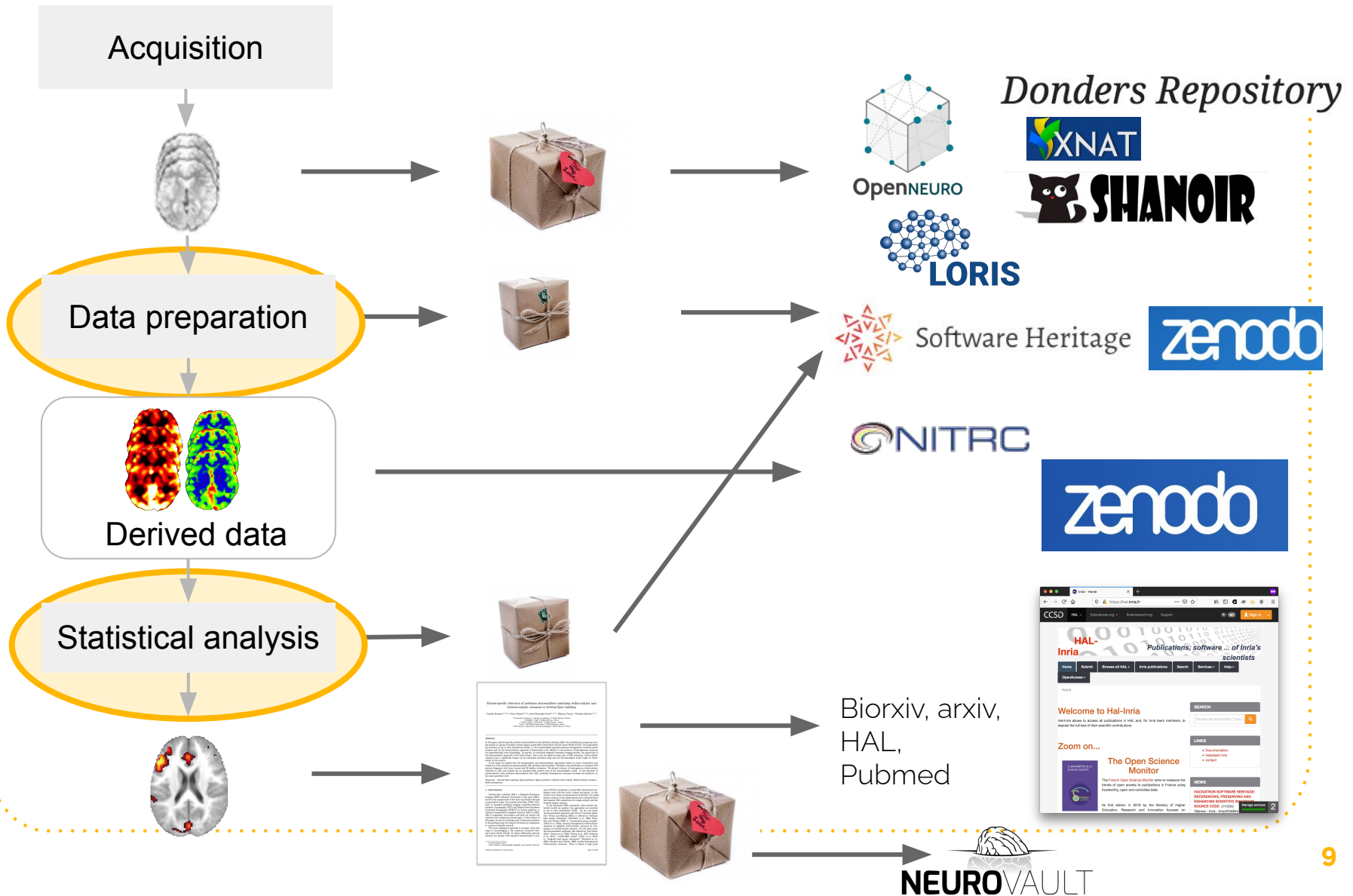


Different requirements for sharing code compared to datasets

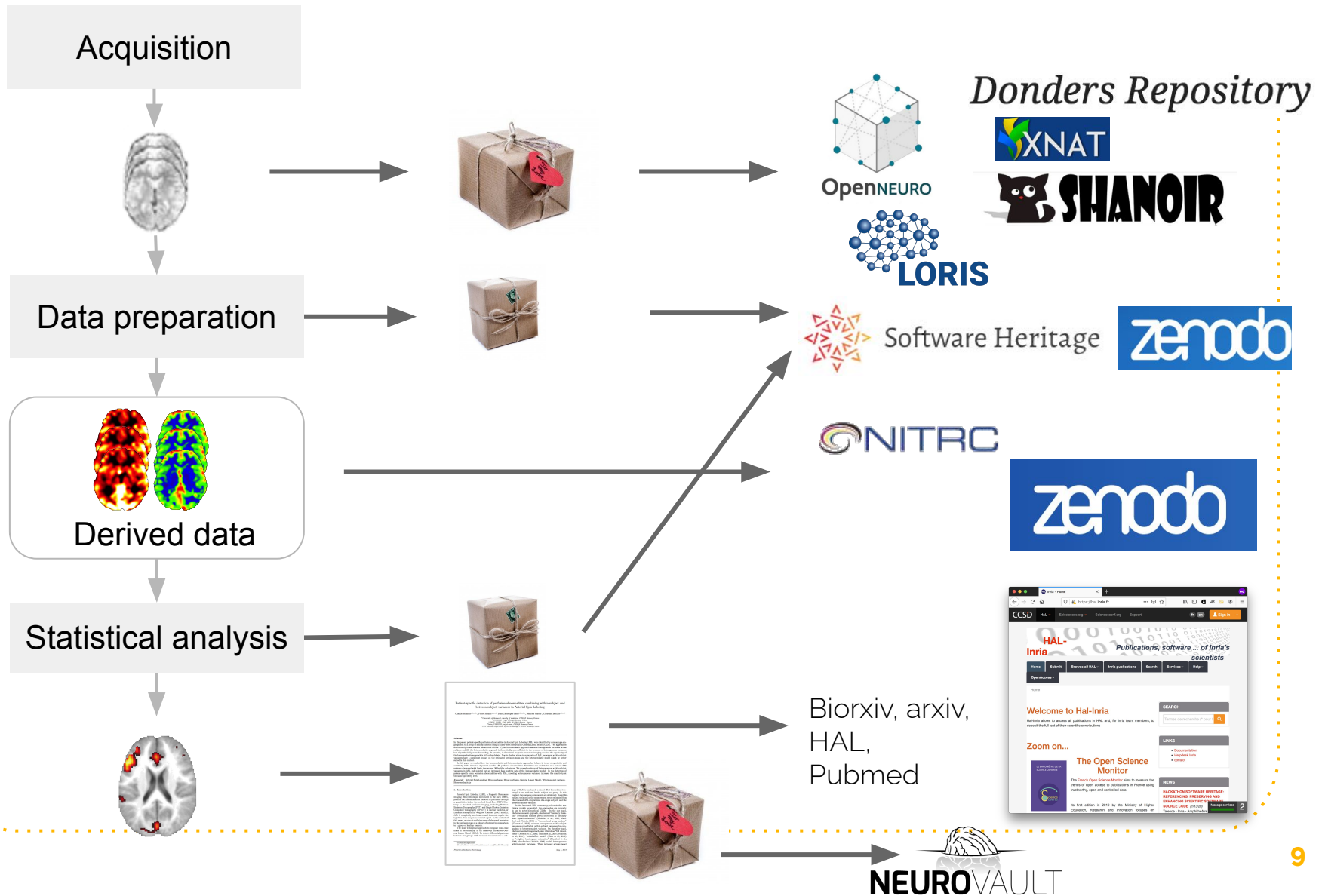
- Often already stored in **online repositories** (e.g. Github, Gitlab)
- **Many** versions (commits)
- Long term preservation is also important!
- **Code structure**, README

```
.
|-- CITATION
|-- README
|-- LICENSE
|-- requirements.txt
|-- data
|   -- birds_count_table.csv
|-- doc
|   -- notebook.md
|   -- manuscript.md
|   -- changelog.txt
|-- results
|   -- summarized_results.csv
|-- src
|   -- sightings_analysis.py
|   -- runall.py
```

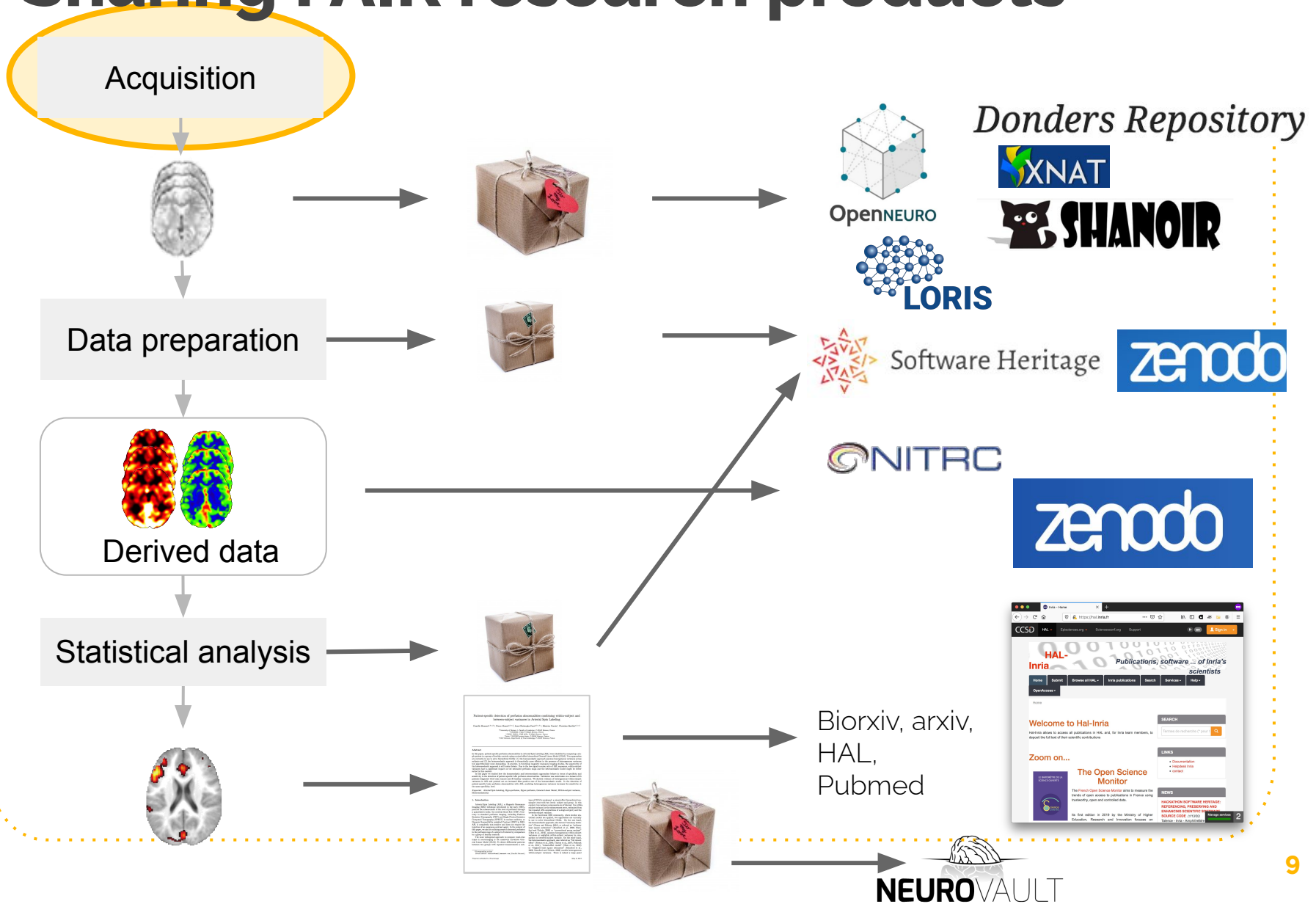
Sharing FAIR research products



Sharing FAIR research products



Sharing FAIR research products



Ready?

Let's get started!

- **You can choose.** Pick one or two research outputs for sharing (e.g. code, results).
- **You can share subparts.** Sharing only the code to build your figures or part of your results (e.g. manually delineated lesion masks).
- When you feel more comfortable, consider **'working open' all the way**, it will make sharing easier later on.
- Connect with the community **and help advance FAIR research outputs on the topics you love.**

Pick what you like the best!

Course Cultivating open science practices
in academic research and culture



Sharing research products FAIRly



Camille Maumet

Thank you!



We are hiring! Join us to
investigate pipeline-induced
variability in an
#openscience-friendly environment

<https://team.inria.fr/empenn/job-offers/>