

Enabling efficient federated neuroimaging studies through FAIR infrastructure

Michelle Wang
McGill University, Canada

15 June 2026

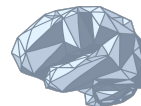


McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab



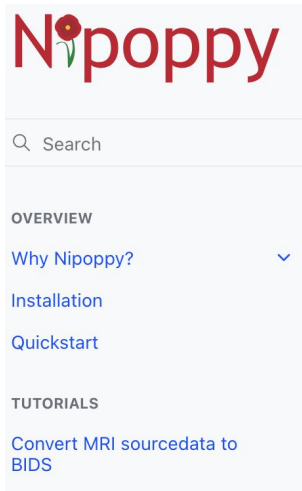
OHBM 2026

Disclosures

The NeuroDataScience-ORIGAMI lab created the Nipoppy and Neurobagel projects.

I am a core maintainer of Nipoppy.

I am a user of Fed-BioMed and collaborate with their development team.



Team

Core team



Core maintainer



Principal Investigator



Michelle Wang



Mathieu Dugré



Nikhil Bhagwat



Julia-Katharina Pfarr



Jean-Baptiste Poline



McGill



neuro

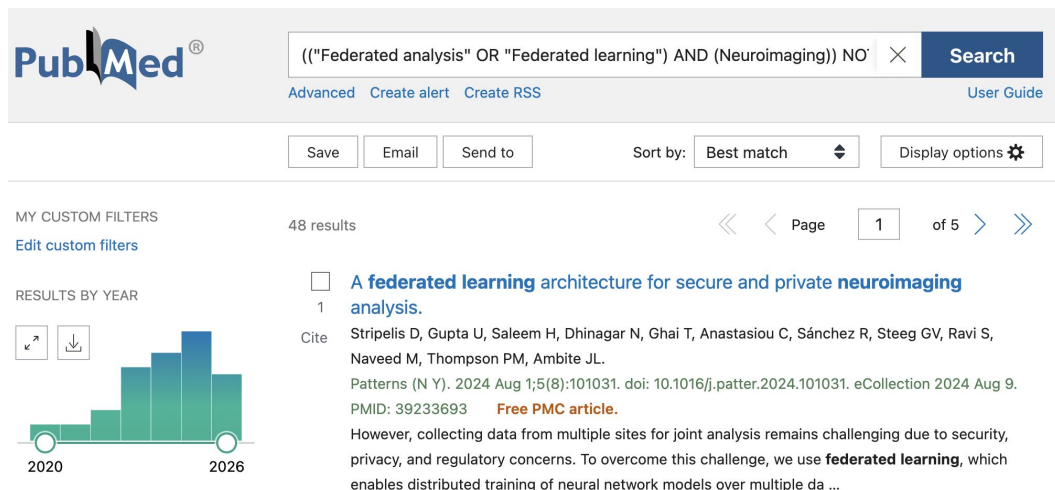
Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Reality check: Federated learning (FL) is not commonly used in neuroimaging

`((("Federated analysis" OR "Federated learning") AND (Neuroimaging))) NOT (Review[Publication Type])`



< 48 results / ~ 28,818

Primarily methods development
(simulated federated setups)



McGill



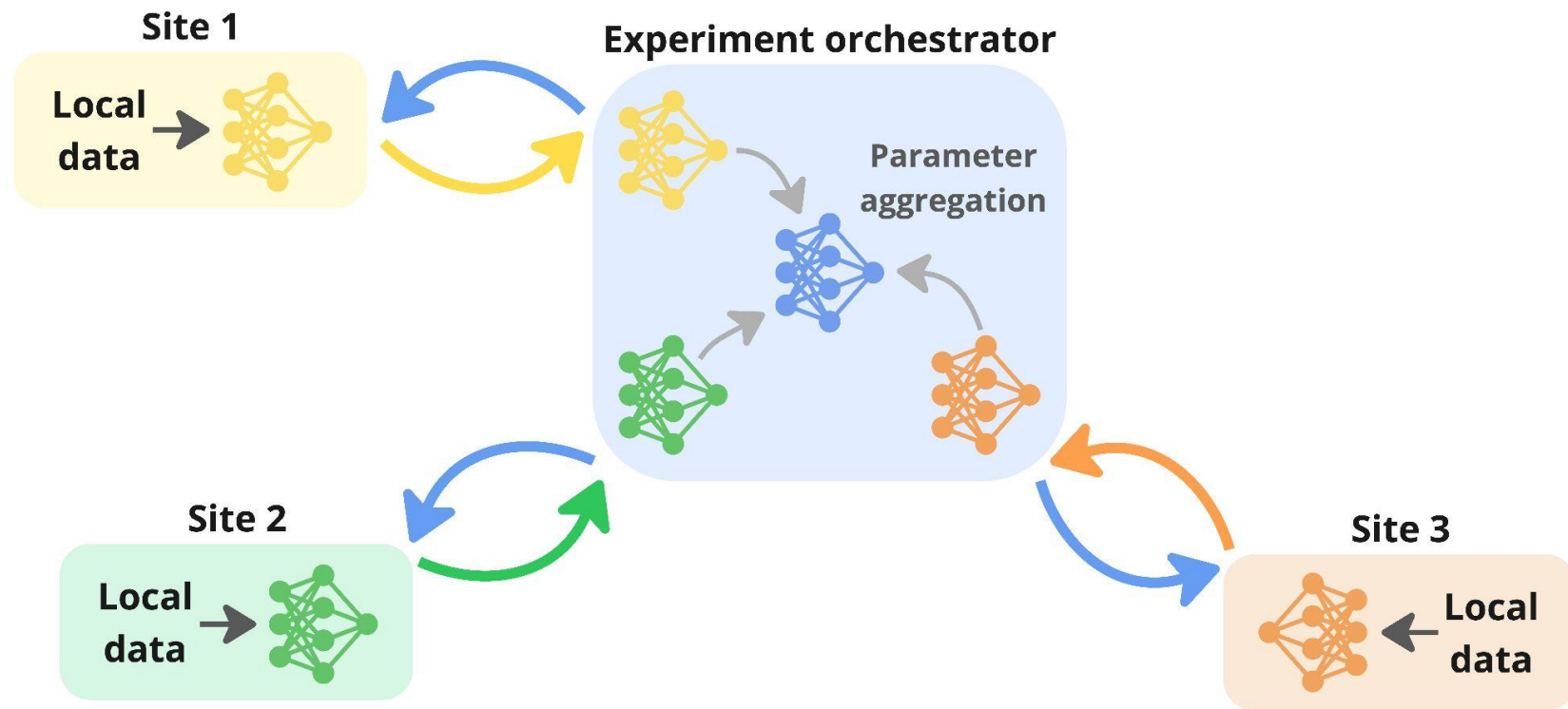
neuro

Montreal Neurological
Institute-Hospital

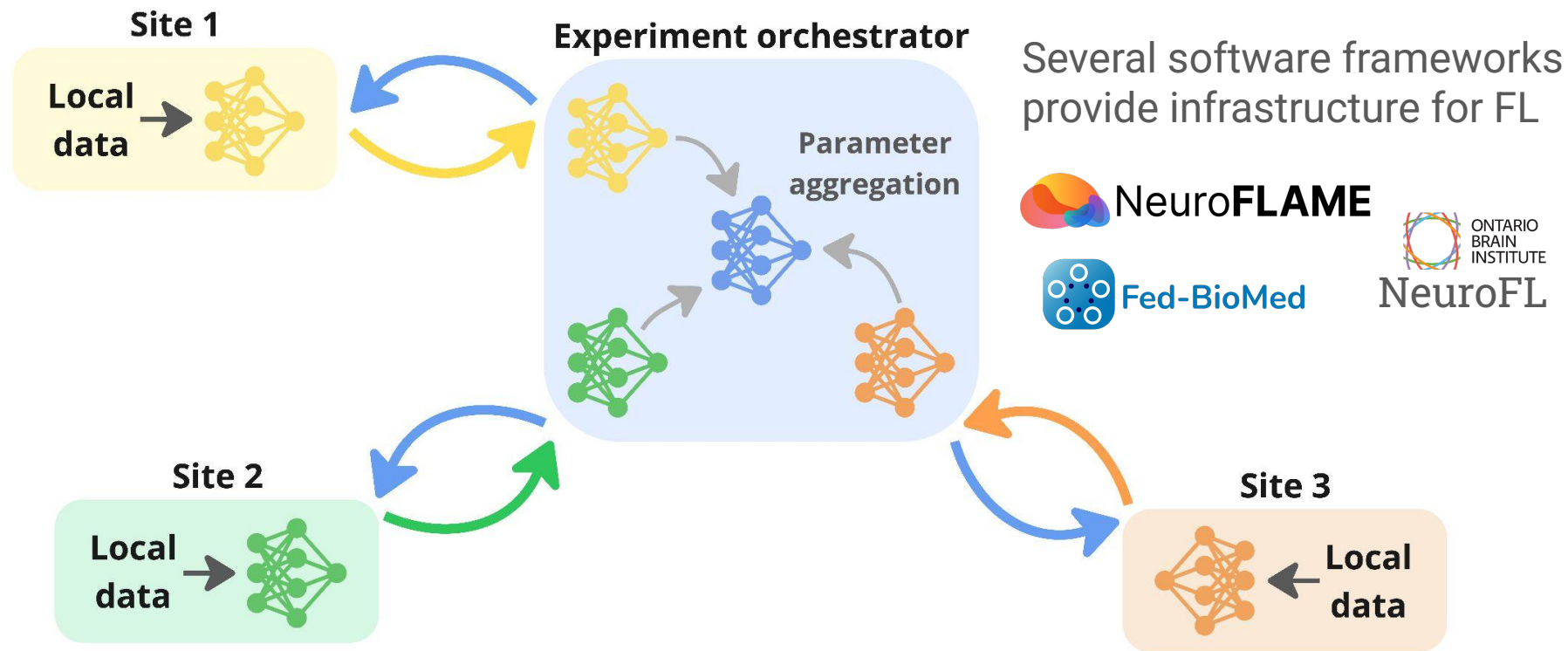


ORIGAMI
Lab

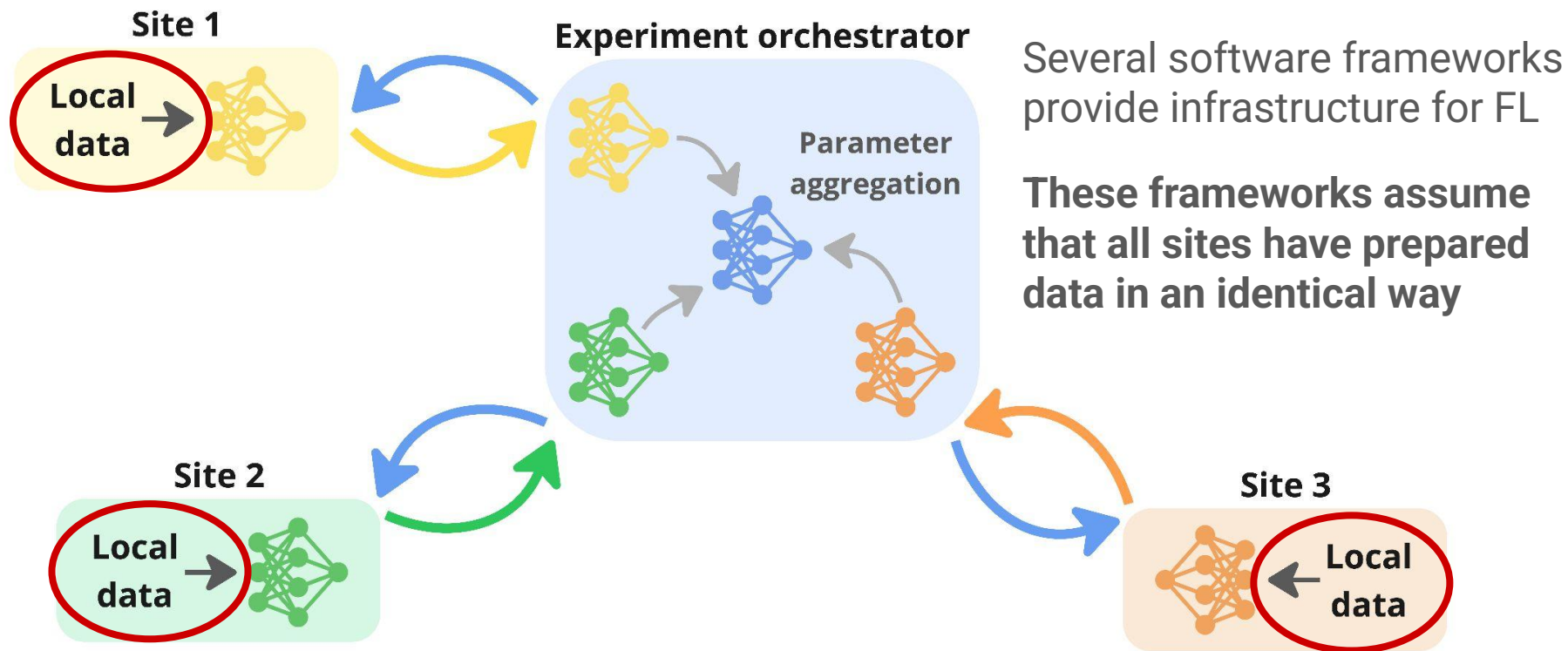
Technical hurdles for FL adoption in neuroimaging



Technical hurdles for FL adoption in neuroimaging



Technical hurdles for FL adoption in neuroimaging



Neuroimaging data preparation challenges

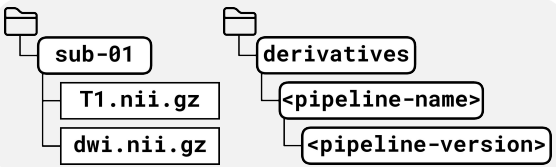
Filesystem organization

- Where are raw and derived data stored?
- How are files named?

Site 1



Site 2



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Neuroimaging data preparation challenges

Filesystem organization

- Where are raw and derived data stored?
- How are files named?

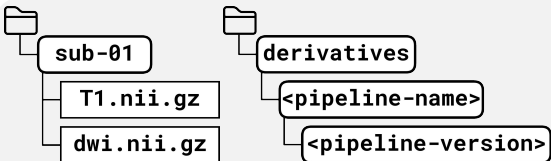
Data processing

- How to ensure the same processing at each site?
- How to launch pipelines in different environments?

Site 1 



Site 2 



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Neuroimaging data preparation challenges

Filesystem organization

- Where are raw and derived data stored?
- How are files named?

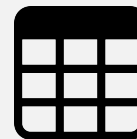
Data processing

- How to ensure the same processing at each site?
- How to launch pipelines in different environments?

Variable semantics

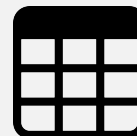
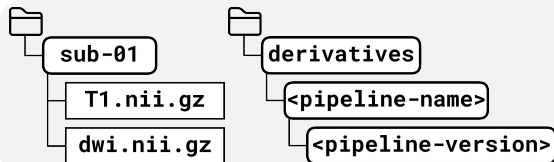
- What measures do the columns contain?
- How are the values coded?

Site 1 



- Age: 50.0; 62.3; ...
- Sex: F; M
- Diag: HC; PD
- MoCA

Site 2 



- age: 78,1; 58,3; ...
- sex: 0; 1
- diag: 0; 1
- moca_score



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Challenge 1: Filesystem organization

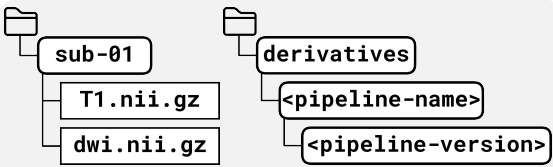
Filesystem organization

- Where are raw and derived data stored?
- How are files named?

Site 1 



Site 2 



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Challenge 1: Filesystem organization

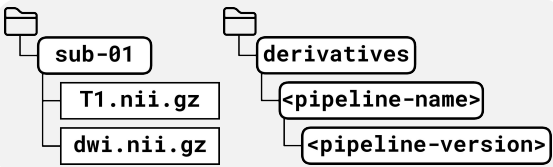
Filesystem organization

- Where are raw and derived data stored?
- How are files named?

Site 1



Site 2



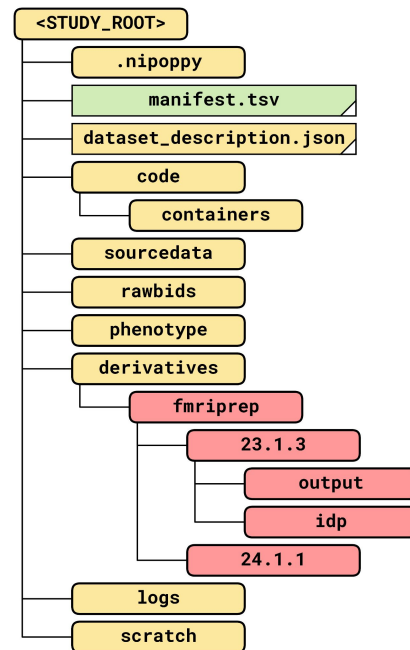
Standardized layout for phenotypic and imaging study files



Study dataset

BIDS allows one to organize the source, raw, and derived data for the entire study as a valid BIDS dataset (see [study DatasetType](#)) in the following way

```
study-1/  
├── sourcedata/  
│   ├── dicoms/  
│   └── ...  
├── rawbids/  
│   ├── sub-01/  
│   ├── sub-02/  
│   └── ...  
├── dataset_description.json  
├── derivatives/  
│   ├── pipeline1-v1/  
│   ├── pipeline2/  
│   └── ...  
└── dataset_description.json  
...
```



Challenge 2: Data processing

Data processing

- How to ensure the same processing at each site?
- How to launch pipelines in different environments?

Site 1 



APTAINER

Site 2 



docker®



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Challenge 2: Data processing

Data processing

- How to ensure the same processing at each site?
- How to launch pipelines in different environments?

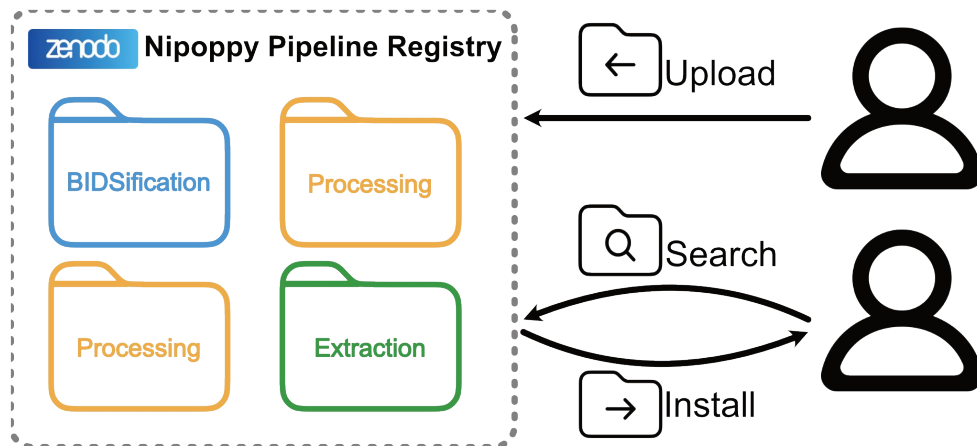
Site 1 



Site 2 



Consistent derivatives processed with identical pipelines with parameters shared (e.g. via Zenodo)



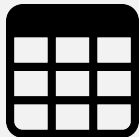
Nipoppy

Challenge 3: Variable semantics

Variable semantics

- What measures do the columns contain?
- How are the values coded?

Site 1 



- **Age:** 50.0; 62.3; ...
- **Sex:** F; M
- **Diag:** HC; PD
- **MoCA**

Site 2 



- **age:** 78,1; 58,3; ...
- **sex:** 0; 1
- **diag:** 0; 1
- **moca_score**



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Challenge 3: Variable semantics

Variable semantics

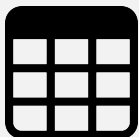
- What measures do the columns contain?
- How are the values coded?

Site 1 



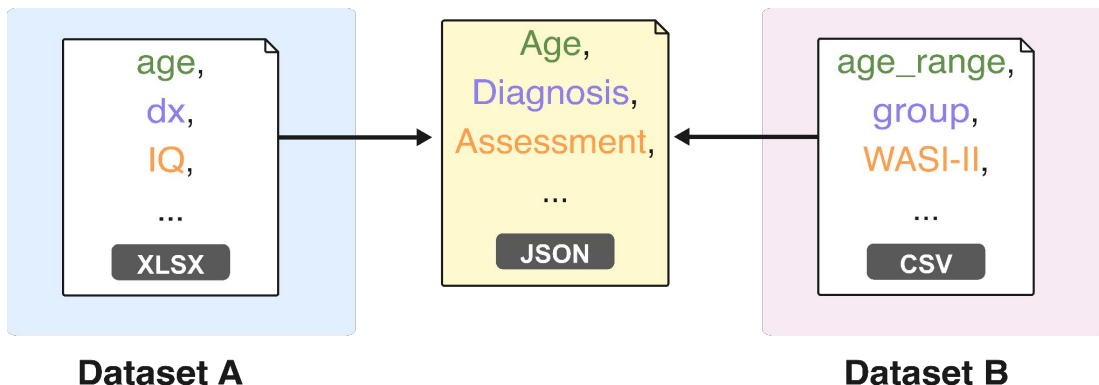
- Age: 50.0; 62.3; ...
- Sex: F; M
- Diag: HC; PD
- MoCA

Site 2 



- age: 78,1; 58,3; ...
- sex: 0; 1
- diag: 0; 1
- moca_score

Semantically harmonized data annotated with a common vocabulary



neurobagel

annotate.neurobagel.org



McGill



neuro

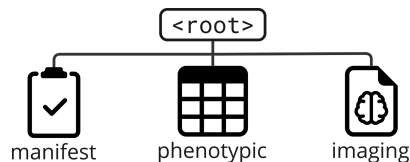
Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Together: FAIR ecosystem for FL-ready data

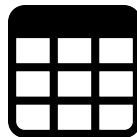
Standardized layout



Consistent derivatives



Semantically harmonized data



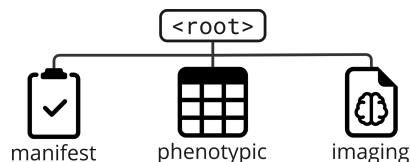
- **nb:Age:** 50.0; 62.3; 78.1; 58.3...
- **nb:Sex:** snomed:248152002; snomed:248153007
- **nb:Diagnosis:** ncit:C94342; snomed:49049000
- **snomed:859351000000102**

Npoppy

neurobagel

Together: FAIR ecosystem for FL-ready data

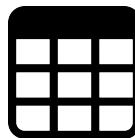
Standardized layout



Consistent derivatives



Semantically harmonized data



- **nb:Age:** 50.0; 62.3; 78.1; 58.3...
- **nb:Sex:** snomed:248152002; snomed:248153007
- **nb:Diagnosis:** ncit:C94342; snomed:49049000
- **snomed:859351000000102**

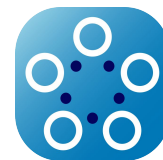
Npoppy

neurobagel

Machine-understandable data
ready for decentralized analysis



Federated learning engine



e.g., Fed-BioMed

Proof-of-concept use-case



neurobagel



Fed-BioMed

Setup	Train data	Model
Siloed (PPMI)	PPMI	$M_{\text{siloed, ppmi}}$
Siloed (Calgary)	Calgary	$M_{\text{siloed, calgary}}$
Siloed (QPN)	QPN	$M_{\text{siloed, qpn}}$
Siloed (ADNI)	ADNI	$M_{\text{siloed, adni}}$
Siloed (PREVENT-AD)	PREVENT-AD	$M_{\text{siloed, prevent-ad}}$
Siloed (COGTIPS)	COGTIPS	$M_{\text{siloed, cogtips}}$
Federated	PPMI	M_{ppmi}
	Calgary	M_{calgary}
	QPN	M_{qpn}
	ADNI	M_{adni}
	PREVENT-AD	$M_{\text{prevent-ad}}$
	COGTIPS	M_{cogtips}
		$M_{\text{federated}}$ (weighted avg. of params.)
Mega-analysis	PPMI	$M_{\text{mega-analysis}}$
	Calgary	
	QPN	
	ADNI	
	PREVENT-AD	

- Alzheimer's and Parkinson's disease datasets
- From FreeSurfer measures, predict
 - Diagnosis (case vs control)
 - Future cognitive decline
 - Age
- Siloed (no sharing) vs federated vs mega-analysis
- **Hypothesis:** Siloed < Federated ≤ Mega-analysis



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Proof-of-concept use-case



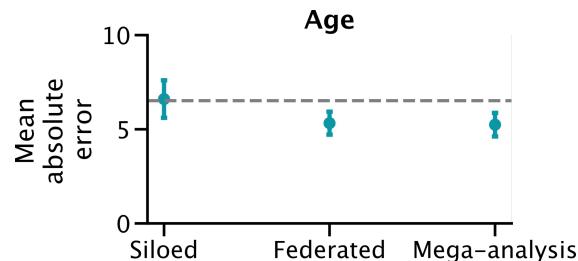
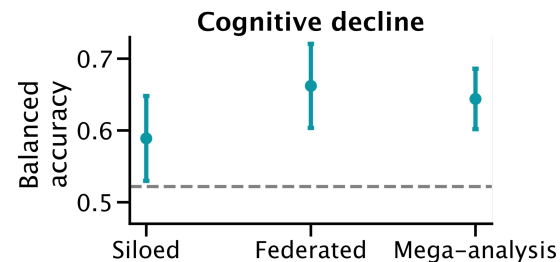
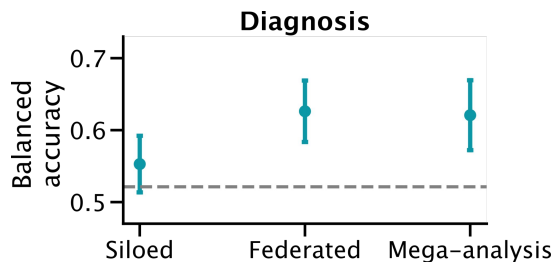
neurobagel



Fed-BioMed

Setup	Train data	Model
Siloed (PPMI)	PPMI	$M_{\text{siloed, ppmi}}$
Siloed (Calgary)	Calgary	$M_{\text{siloed, calgary}}$
Siloed (QPN)	QPN	$M_{\text{siloed, qpn}}$
Siloed (ADNI)	ADNI	$M_{\text{siloed, adni}}$
Siloed (PREVENT-AD)	PREVENT-AD	$M_{\text{siloed, prevent-ad}}$
Siloed (COGTIPS)	COGTIPS	$M_{\text{siloed, cogtips}}$
Federated	PPMI	M_{ppmi}
	Calgary	M_{calgary}
	QPN	M_{qpn}
	ADNI	M_{adni}
	PREVENT-AD	$M_{\text{prevent-ad}}$
	COGTIPS	M_{cogtips}
		$M_{\text{federated}}$ (weighted avg. of params.)
Mega-analysis	PPMI	$M_{\text{mega-analysis}}$
	Calgary	
	QPN	
	ADNI	
	PREVENT-AD	

- Siloed (no sharing) vs federated vs mega-analysis
- **Hypothesis:** Siloed < Federated $\leq$ Mega-analysis



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab

Thank you!



NeuroDataScience-ORIGAMI Lab

Nikhil Bhagwat

Mathieu Dugré

Julia-Katharina Pfarr

Jean-Baptiste Poline

Alyssa Dai

Arman Jahanpour

Sebastian Urchs

Fed-BioMed team

Francesco Cremonesi

Sergen Cansiz

Lucie Chambon

Ali Tolga Dinçer

Jhonatan Torres

Marc Vesin

Marco Lorenzi

Preprint



Data collaborators

Emile d'Angremont

Chris Vriend

Ysbrand D. van der Werf

Gabriel Pinilla-Monsalve

Yuan Song

Oury Monchi



McGill



neuro

Montreal Neurological
Institute-Hospital



ORIGAMI
Lab