

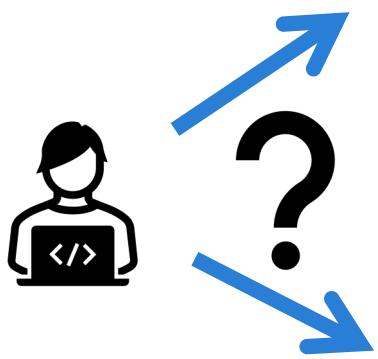
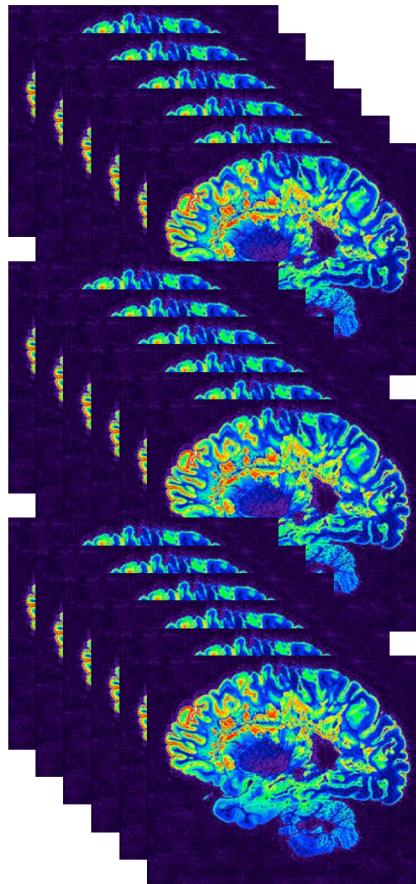
BigBrain Data Processing with CBRAIN and DataLad



7th BigBrain Workshop
October 4, 2023
Reykjavik, Iceland

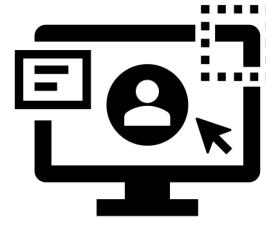
Bryan Caron, PhD
McGill University

Big Data Processing



```
user@debut:~/Desktop$ ls -l
total 30388
drwxr-xr-x  2 user  user  4096 Aug 18 21:02 2_Fetching_Sessions_and_Files.ipynb
-rw-r--r--  1 user  user   29575 Aug 18 21:02 2_Measure_Crops_and_Data_Procedures.ipynb
-rw-r--r--  1 user  user   98765 Aug 18 21:02 3_Measure_Sects_and_Tests.ipynb
-rw-r--r--  1 user  user   98765 Aug 18 21:02 4_Measure_Sects_and_Tests.ipynb
drwxr-xr-x  2 user  user  4096 Aug 18 19:38 alltests_cache-obj
drwxr-xr-x  2 user  user  4096 Aug 18 19:38 alltests_cache-obj
-rw-r--r--  1 user  user  144141 Aug 18 20:18 auto-104936_aec_PyMCdmc_run-01.tif.nii.gz
-rw-r--r--  1 user  user  144141 Aug 18 20:18 auto-104936_aec_PyMCdmc_run-01.tif.nii.gz
$ python3.6 main.py
[...]
Python 3.6.5 |Anaconda, Inc.| (default, Jul 11 2017, 13:36:48)
Type "help", "copyright", "credits" or "license" for more information.
IPython 7.0.0 -- An enhanced Interactive Python. Type "%?" for help.

In [1]: import nibabel
In [2]: import nibabel as nib # load the brain scan file
In [3]: brain = nib.load('auto-104936_aec_PyMCdmc_run-01.tif.nii.gz')
/home/jayyer/.local/lib/python3.7/site-packages/nibabel/datasets/_unit_.py:98: FutureWarning: Patchers from the nifti_npy module will be updated in version 3.8 to return NumPy strings instead of Bytes and Pandas DataFrames instead of Numpy arrays. Please update from the nibabel.datasets module via 'nipyapi'.
```



Web UI



Easy access to Big Data
and Big Compute



Reproducible
Science



Scale-up
Science



Data
Lad

cbrain



Data
Lad



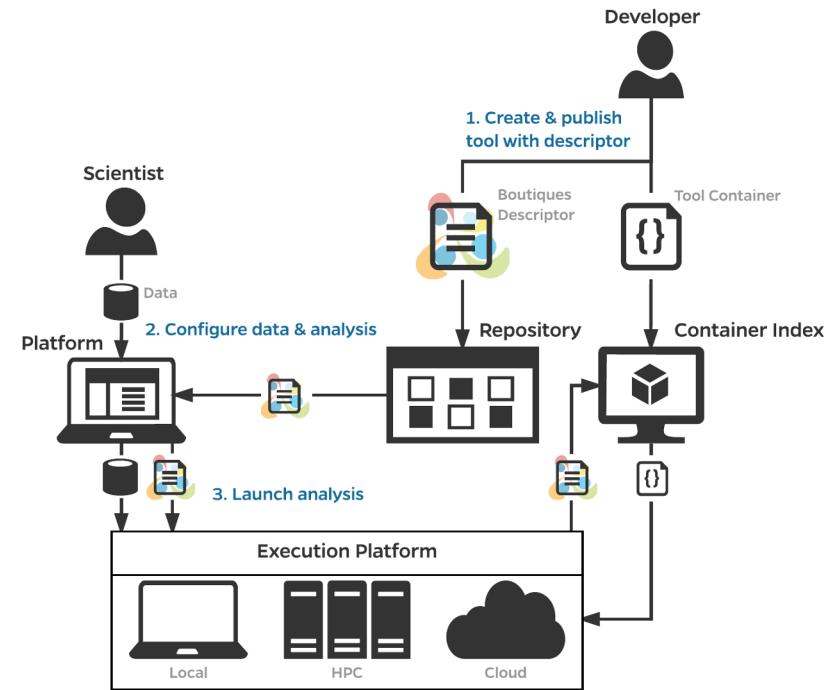
Collaborate
and Share

Big Data Processing



<https://github.com/boutiques>

- Fully-automated integration of applications
- Runs containers on heterogeneous computing resources
- Comprehensive input validation through a strict JSON schema
- Flexible application description through a rich JSON schema
- Pipeline publishing
 - The ability to uniquely reference the software execution
 - data and provenance for open science publication (ie. [Zenodo](#))

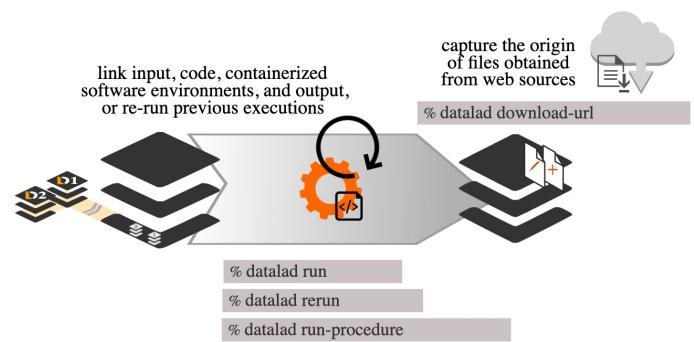
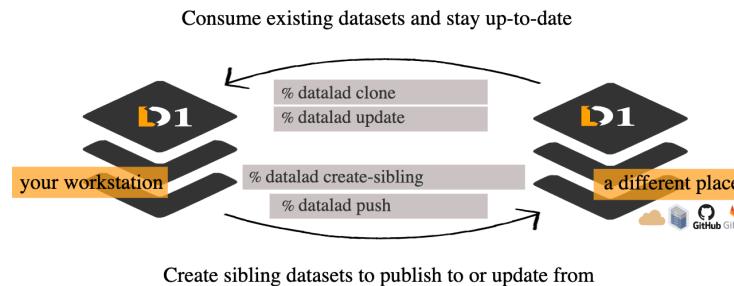
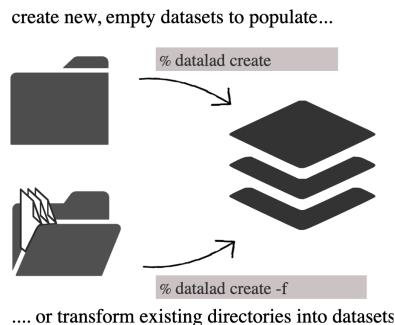


Big Data Processing

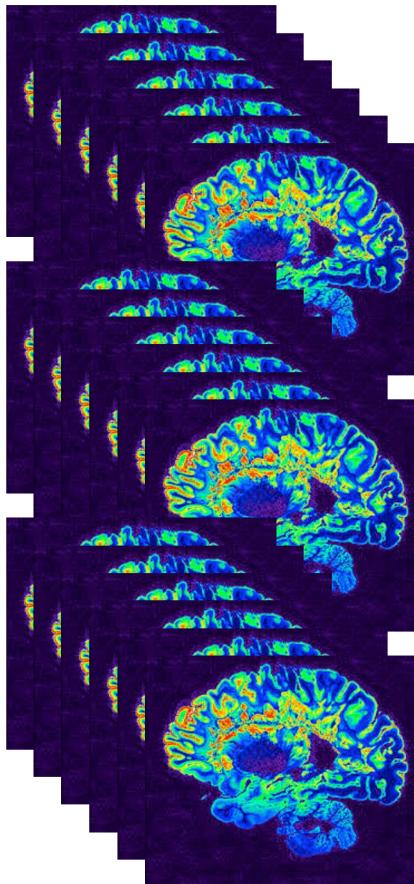


<https://www.datalad.org/>

- Free, open source command-line tool and Python API
- Built on-top of Git and Git-Annex
- Allows version controlling of arbitrarily large content
- Provides transport mechanisms for sharing and obtaining data
- Enables reproducible data analysis



Big Data Processing



```
jupyter-notebook:3fb2cf:-8 ls -l
total 53384
drwxr-x---  2 Jupyter users 12514 Aug 19 22:02 1_Hierarchical_Designs_and_Projex.ipynb
drwxr-x---  2 Jupyter users 12515 Aug 19 22:02 2_Hierarchical_Groups_and_Data_Providers.ipynb
drwxr-x---  2 Jupyter users 57820 Aug 19 22:02 3_Hierarchical_Applications.ipynb
drwxr-x---  2 Jupyter users 57821 Aug 19 22:02 4_Hierarchical_Projex_and_Providers.ipynb
drwxr-x---  2 Jupyter users 57822 Aug 19 20:54 5_Libraries_and_Cache.ipynb
drwxr-x---  2 Jupyter users 57823 Aug 19 20:54 6_Libraries_and_Cache.ipynb
drwxr-x---  2 Jupyter users 1444410 Aug 29 2013 sub-1004359_ece-PREBOLD_task-encoding_run-001_T1w.nii.gz
drwxr-x---  2 Jupyter users 34131890 Aug 19 20:56 sub-1004359_ece-PREBOLD_task-encoding_run-001_b01d.nii.gz
Python 3.8.6 packaged by conda-forge | | default; Jul 11 2021, 03:39:48
Tcl/Tk 8.6.4 -- Tk 8.6.4 -- An enhanced interactive Python. Type "t" for help.

[1]: import nilearn
[2]: import nibabel as nib # loads the brain scan files
[3]: from nilearn import plotting
[4]: from nilearn.datasets import fetch_nfm
[5]: mat_file = nib.load(file_to_load) #loading an anatomical (T1 weighted) scan
[6]: mat_file.shape
(176, 240, 232)
[7]: plotting.plot_imgmask_imgmat_file,output_file='test.png' #General purpose plotting, no specific preset
[8]: test.png
->writer-crc- 3 Jupyter users 253810 Aug 25 21:22 test.png
[9]:
```



cbrain

A screenshot of the Cbrain web interface, specifically the 'Files' section. It shows a list of four files: 'wb-01_ses-test_T1w.nii', 'wb-02_ses-test_T1w.nii', 'wb-03_ses-test_T1w.nii', and 'wb-04_ses-test_T1w.nii'. Each file is listed with its file type (Nifti File), owner (brian), creation date (2022-04-11), size (0.3 MB, 0.3 MB, 0.3 MB, 0.5 MB), tags (None), project access (Read Only, Read Only, Read Only, Read Only), provider (NeurodataOne, MaxCompute, NeurodataOne, NeurodataOne), and synchronization symbols (Sync, Sync, Sync, Sync).



CBRAIN orchestrates the data and computations on remote HPC and Cloud resources



Easy access to Big Data
and Big Compute



Reproducible
Science



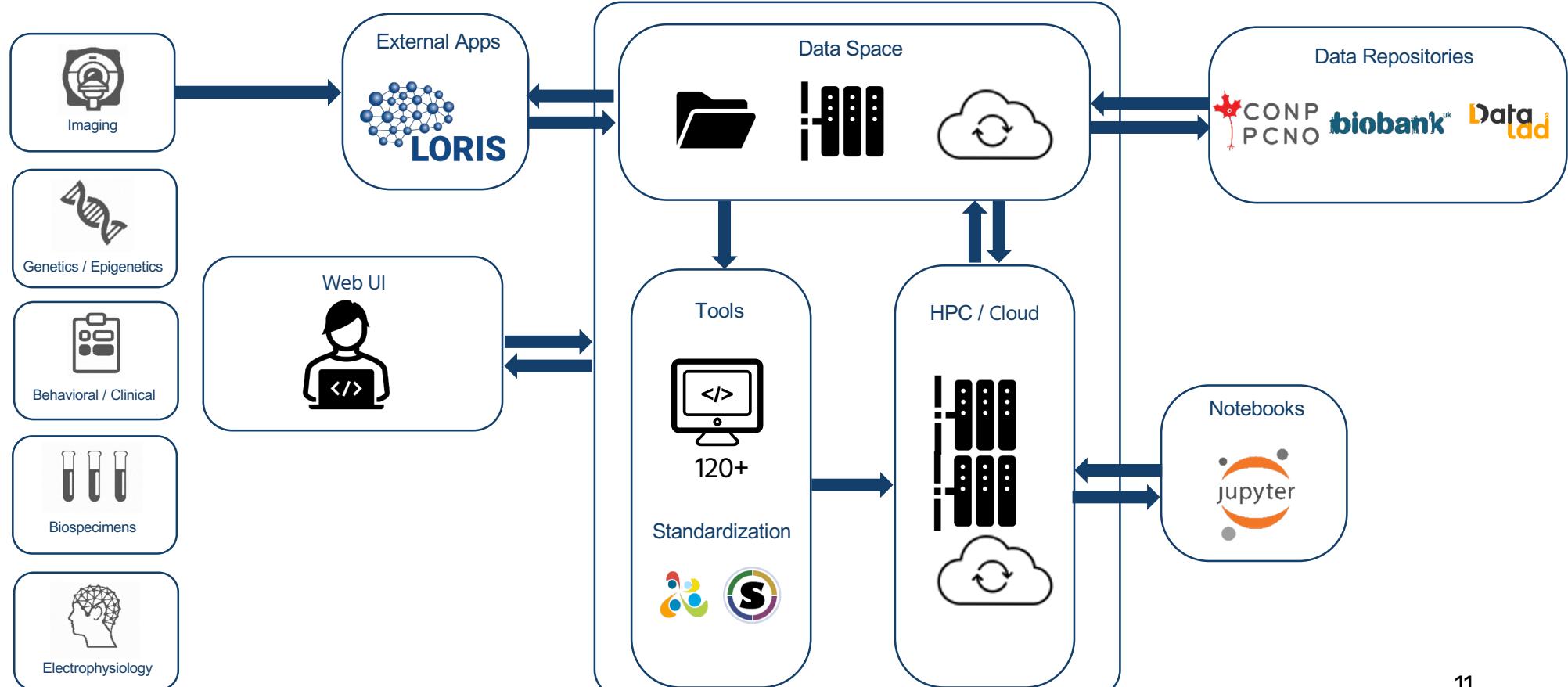
Scale-up
Science



Collaborate
and Share

Big Data Processing

cbrain



Features of CBRAIN



Enables distributed execution of software pipelines

NeuroHub Dashboard My Account Projects Messages (11) Resources Support Forum Rev: 6.2.0-197 (last updated 1m ago) Logged in as Bryan Caron Sign out

cbrain - bcaron - Tasks

Files Tasks

Update Attributes For Failed Tasks For Completed Tasks Terminating And Cleaning Up Archiving Filters

Switch to List View (128 tasks) Total task space used: 46.2 MB 25 per page.

Task Type	Version	Description	Owner	Execution Server	Current Status	Run Number	Workdir Size	Time Submitted	Last Updated	...
+ Civet	2.1.1	sub_01_ses-test_T1w.mnc	bcaron	Bridges2	3 x Failed		0 bytes	2021-05-20 12:45:28 EDT	2021-05-20 12:46:46 EDT	...
+ Civet	2.1.1	sub_01_ses-test_T1w.mnc	bcaron	Bridges2	3 x Failed		0 bytes	2021-05-20 12:55:37 EDT	2021-05-20 12:56:38 EDT	...
+ Civet	2.1.1	sub_01_ses-test_T1w.mnc	bcaron	Graham	3 x Completed		0 bytes	2021-05-20 13:28:30 EDT	2021-05-20 20:56:50 EDT	...
FailFirst	5.0.9	My test tasks!	bcaron	Converter-2	Completed	1	V (25.7 MB)	2021-08-10 17:06:47 EDT	2021-08-10 17:16:15 EDT	...
+ FMRiprepSingleSubject	20.2.0		bcaron	Beluga	25 x Failed		0 bytes	2021-12-15 13:51:54 EST	2021-12-16 21:25:26 EST	...
FreeSurfer Recon_all	v7.1.1		bcaron	GrahamPlatform	Completed	1	V (299.7 MB)	2022-03-16 13:17:34 EDT	2022-03-16 23:26:10 EDT	...
+ FMRiprepBidsSubject	21.0.0		bcaron	Narval	46 x Terminated		0 bytes	2022-05-11 14:51:49 EDT	2022-05-11 14:52:34 EDT	...
+ FMRiprepBidsSubject	21.0.0		bcaron	Narval	46 x Failed		46.2 MB	2022-05-11 15:50:07 EDT	2022-05-11 15:54:36 EDT	...

Workdir archiving status symbols: : On Cluster V: As File

Powered by CBRAIN Credits

Aggregates multiple distributed file systems into a uniform view

My Account Projects Messages Resources User Manager Help Support

cbrain - Tasks

File Tasks

- Update Attributes + For Failed Tasks + For Completed Tasks

Active Filters (User) + User groups + Filter in Bulk + User groups +

Type	Description
CARD	user_pierre_20090925_200127_2_A_Card
High-Res Reconstruction	User: 12 Proj: 8 Tasks: 6
CBRAIN-Hipp	User: 4 Proj: 12 Tasks: 6
CBRAIN-Hipp	User: 1 Proj: 68 Tasks: 302
All	Proj: 207 Task: 115

Next Proj > Next Nefile >

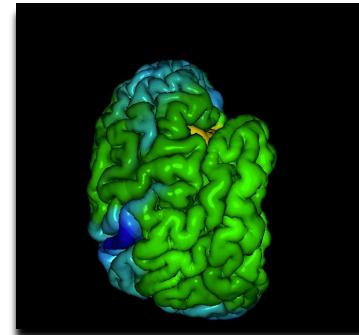
Atlas view

Screenshot

Volume Coordinates

Lobes surface parcellation (Left-Interside Left=40, Right=24)

BrainBrowser web-based 3D visualization for real-time manipulation and analysis of volumes and surfaces



Software Tools

cbrain

cbrain 120+ Tools

fMRI Analysis

fMRI Preprocessing

Structural MRI

Genetics

Tractography

PET

Arterial Spin Labeling

EEG

Data Conversion Tools

Statistics



Interactive Data Exploration



Projects and users can interact programmatically with data and tools



Session Management

Authentication
Authorization
Access Control

Tools

Tool selection
Libraries and Packages Loading
and import

Tasks

Task submission
Task status querying
Task management

Data Handling

File upload / download
File search / selection
File read / write

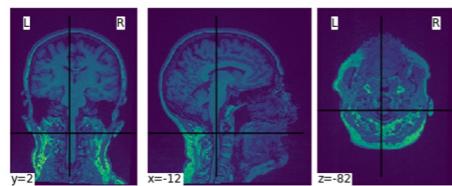
```
In [24]: file_to_load = 'sub-1004359_ses-PREBL00_run-001_T1w.nii.gz' #specify filename as string
anat_file = nib.load(file_to_load) #loading an anatomical (T1 weighted) scan
print(anat_file.shape) #3D shape

file_to_load2 = 'sub-1004359_ses-PREBL00_task-encoding_run-001_bold.nii.gz'
bold_file = nib.load(file_to_load2) #loading an fMRI (BOLD) scan
print(bold_file.shape) #4D shape

plotting.plot_img(anat_file) #General purpose plotting, no specific preset
plotting.plot_anat(anat_file) #Anatomical image plotting

(176, 240, 256)
(64, 64, 32, 183)

Out[24]: <nilearn.plotting.displays.OrthoSlicer at 0x7f352eae9b70>
```

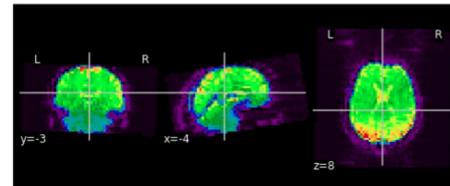


5. fMRI visualization

```
In [26]: from nilearn.image.image import mean_img
mean_fmri = mean_img(bold_file) #Compute the mean of the images over time or the 4th dimension

from nilearn.plotting import plot_epi, show
plot_epi(mean_fmri) #plot cuts of the image (by default 3 cuts: Frontal, Axial, and Lateral)

Out[26]: <nilearn.plotting.displays.OrthoSlicer at 0x7f352ecce908>
```



Access to Large Datasets



Enabling scientific discoveries that improve human health



CONP Portal | Data Search

Filter By: Modality: File Format: Third-party account required: Search

Available in CBRAIN

Results 1 - 10 displayed of 45. (Maximum results per page 10 -)

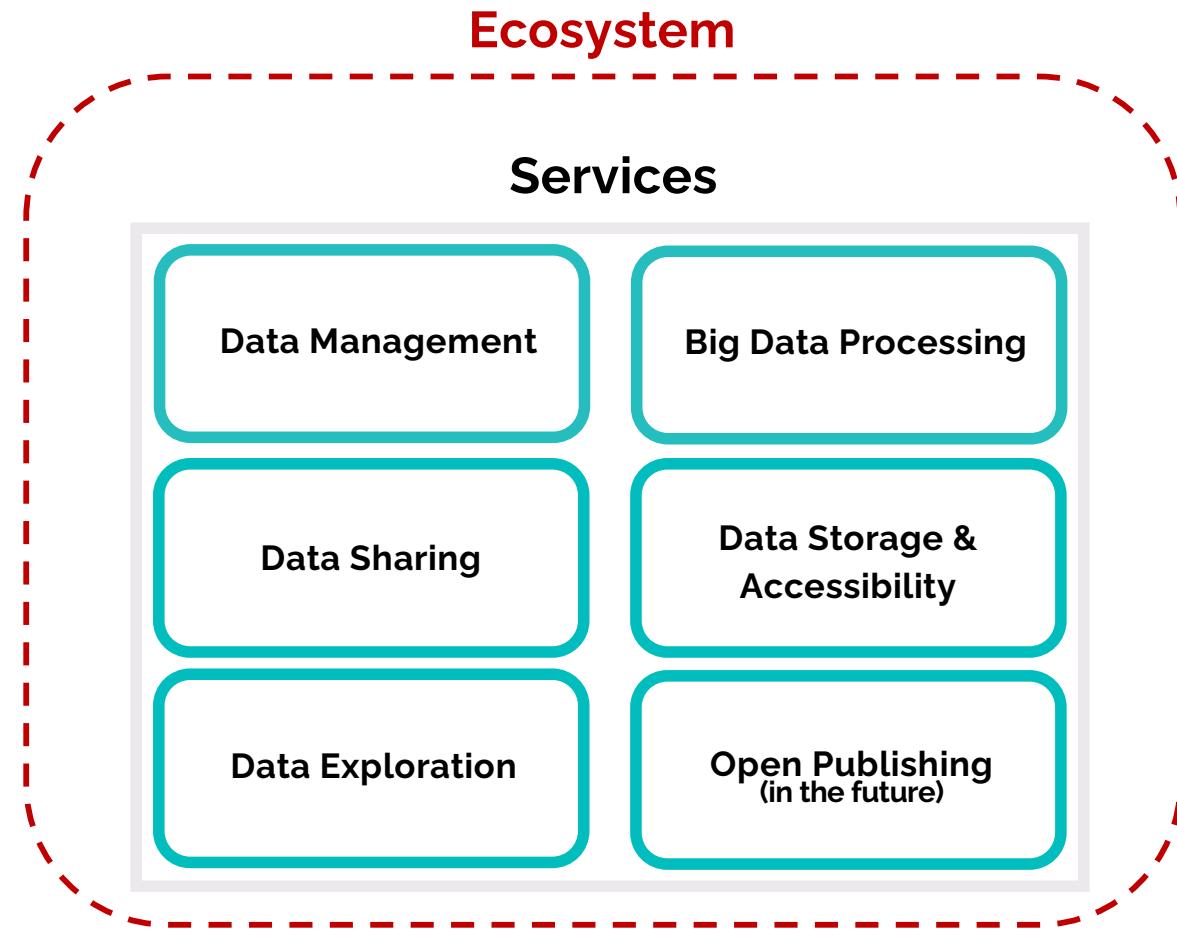
Sort By: Origin

SIMON	Direct Download (Not Available)	?
Creators: Laboratoire MEDICS	Download With DataLad	?
Version: 1.0 Date Added: 2019-04-26 Date Updated: 2021-11-12	Process On	cbrain
Modalities: MRI, fMRI License: CC-BY-SA		?
Files: 1829 Size: 84.0 GB Subjects: 1 Formats: NIfTI		?
ARK ID ⓘ: https://n2t.net/ark:/99999/d7zmp70228p8c8hx0		?
BigBrain dataset	Direct Download (Not Available)	?
Creators: BigBrain project	Download With DataLad	?
Version: 1.0 Date Added: 2020-01-12 Date Updated: 2022-05-02	Process On	cbrain
Modalities: histology License: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License		?
Files: 316 Size: 112.5 GB Subjects: 1 Formats: MINC, NIfTI		?
ARK ID ⓘ: https://n2t.net/ark:/99999/d7r3hp47g3419234m3		?
Learning Naturalistic Structure: Processed fMRI dataset	Direct Download (13.9 GB)	?
Creators: DuPre, Elizabeth	Download With DataLad	?
Version: 0.1.0 Date Added: 2020-02-07 Date Updated: 2021-06-02	Process On	cbrain
License: other-pd		?
Files: 240 Size: 13.9 GB Formats: TSV, NIfTI, TXT		?
ARK ID ⓘ: https://n2t.net/ark:/99999/d7pkf815v7mm6p17r		?



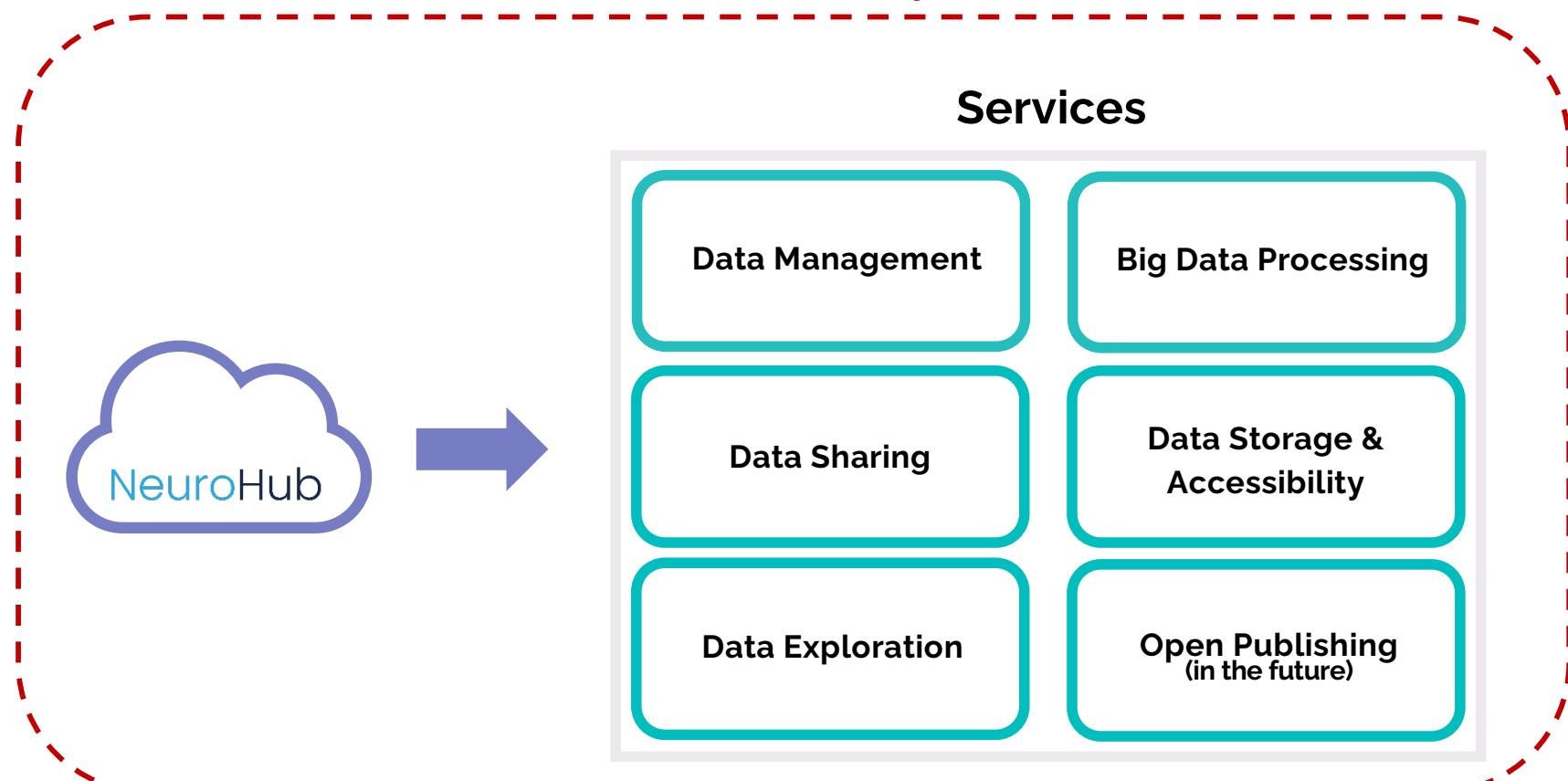
CONP-1000-Genomes	CONP-AlzheimerMouse	CONP-BigBrain	CONP-BigBrain-3D-ROIs	CONP-BigBrain-3DClass...
Files: 37 Tasks: 0 1000 Genomes Project	Files: 18 Tasks: 0 Access control project for CONP ...	Files: 317 Tasks: 0 Access to files in the BigBrain ...	Files: 11 Tasks: 0 BigBrain dataset - 3D ROIs (dev...)	Files: 25 Tasks: 0 Access to files in the BigBrain ...
View	View	View	View	View
CONP-BigBrain-3DSurfaces	CONP-BigBrain-A3D	CONP-BigBrain-Hippoc...	CONP-BigBrain-Layered	CONP-BigBrain-MBISIM
Files: 49 Tasks: 0 Access to files in the BigBrain ...	Files: 5 Tasks: 0 BigBrain dataset - A3D (derived ...)	Files: 3 Tasks: 0 BigBrain dataset - Hippocampus ...	Files: 7 Tasks: 0 BigBrain dataset - Layer Segment...	Files: 6 Tasks: 0 BigBrain dataset - MBISIM (deriv...
View	View	View	View	View
CONP-BigBrain-RawData	CONP-BigBrain-SurfParcels	CONP-BigBrainMRICoreg	CONP-BrainSpan	CONP-Calgary-Campinas
Files: 4 Tasks: 0 BigBrain dataset - Raw Data	Files: 6 Tasks: 0 BigBrain dataset - Surface Parce...	Files: 13 Tasks: 0 Accurate registration of the Big...	Files: 24 Tasks: 0 BrainSpan: Atlas of the Develop...	Files: 3 Tasks: 0 Access control project for CONP ...
View	View	View	View	View
CONP-CalgaryPreschool	CONP-CPMM-7T	CONP-Chik-Def-Mouse	CONP-Chik-Striatal	CONP-HCPUR109
Files: 410 Tasks: 0 Calgary Preschool fMRI Dataset	Files: 36 Tasks: 0 Files in CONP public dataset CP...	Files: 4 Tasks: 0 Brainstem differences between ...	Files: 4 Tasks: 0 Mouse models that differ in neur...	Files: 8 Tasks: 0 HCPUR109: Human Adult Human
View	View	View	View	View
CONP-Hippociefold	CONP-HumanRef-Genome	CONP-LearnNat	CONP-Mice-mice	CONP-MiceImmune
Files: 14 Tasks: 0 Hippocampal morphology and cyto...	Files: 3 Tasks: 0 Human sapio reference genome se...	Files: 422 Tasks: 0 CONP Learning Naturalistic Struct...	Files: 65 Tasks: 0 CONP Mice-mice	Files: 388 Tasks: 0 Longitudinal structural MRI and ...
View	View	View	View	View
CONP-Multimodal-iGCaMP	CONP-MurineNeocortical...	CONP-Muskate-Squirrels	CONP-Neural-Cog	CONP-Neurocom
Files: 12 Tasks: 0 Multimodal data with wide-field ...	Files: 96 Tasks: 0 Intracellular Recordings of Mur...	Files: 4 Tasks: 0 MEF and unbiased averages of ME...	Files: 95 Tasks: 0 Quantifying Neural-Cognitive Rel...	Files: 46 Tasks: 0 Parkinson's Disease Datasets - N...
View	View	View	View	View
CONP-NeuroRehab-MS	CONP-NIH-NHP-Atlas	CONP-OpenPreventAD	CONP-SIMON	CONP-SPINS-Phantoms
Files: 8 Tasks: 0 The Effect of Neurorehabilitatio...	Files: 381 Tasks: 0 NIH Blueprint Non-Human Primate ...	Files: 627 Tasks: 2 This project restricts access to ...	Files: 5 Tasks: 0 Access control project for CONP...	Files: 11 Tasks: 0 SPINS: Treating Human Phantoms
View	View	View	View	View
CONP-Tsne	CONP-VFA-T1-Mapping	CONP-VisualWorkingMemory	CONP-Integration_Testing	All
Files: 43 Tasks: 0 Parkinson's Disease Datasets - T...	Files: 7 Tasks: 0 VFA T1 mapping R7nowk (open) ...	Files: 35 Tasks: 0 CONP Dataset	Files: 0 Tasks: 0 This is a project to test the in...	File: 578569 Tasks: 256805
View	View	View	View	View

Interoperability of services is key to building researcher communities and their success

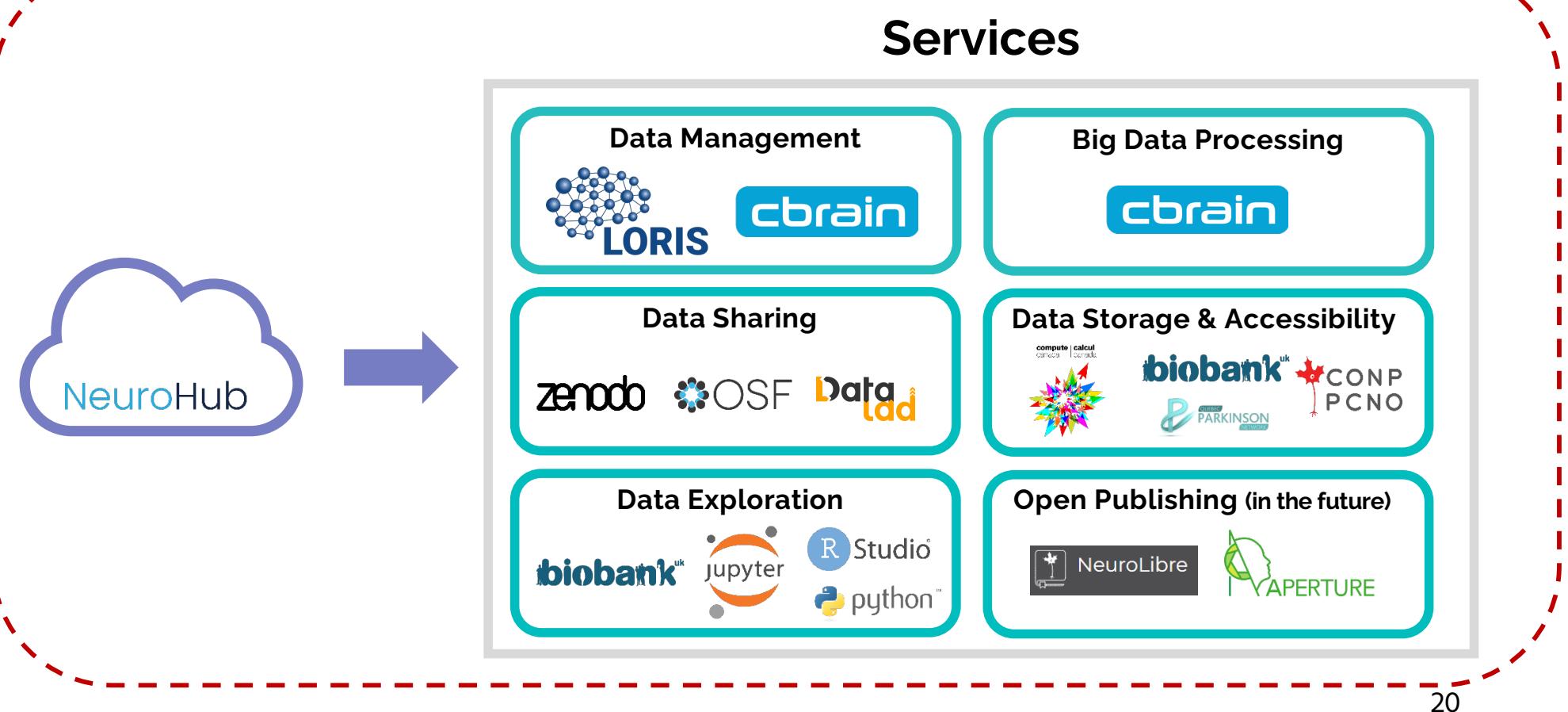


Interoperability of services is key to building researcher communities and their success

NeuroHub Ecosystem



NeuroHub Ecosystem



Conclusion



Big Data



Big Computation



International
Collaboration



Interdisciplinary
Health Research

Infrastructure and services for accomplishing science



CBRAIN and NeuroHub are core components for reproducibility and interoperability

- CBRAIN is a completely general platform that can be used beyond brain science
- Interoperate or integrate – not reimplement



cbrain.ca

github.com/aces/cbrain



neurohub.ca

github.com/neurohub

Acknowledgements



Alan Evans, Bryan Caron, Pierre Rioux, Natacha Beck, Serge Boroday, Darcy Quesnel, Xavier Lecours-Boucher, Najmeh Kalili-Manhani, Darius Valevicius, Safa Sanami, Reza Adalat, Tristan Glatard



J-B Poline, Bryan Caron, Alan Evans, Pierre Rioux, Natacha Beck, James Mehta, Samir Das, Xavier Lecours-Boucher, Rida Abou-Haidar, Darcy Quesnel, Emmet O'Brien, Sebastian Urchs, Derek Lo, Diana Le

Past Contributors (alphabetical order)

Adam Trefonides, Andrew Doyle, Angela McCloskey, Anton Zoubarev, Armin Taheri, Candice Czech, Ehsan Afkhami, Greg Kiar, Ksenia Zaytseva, Liam O'Callaghan, Marc Rousseau, Mathieu Desrosiers, Melanie Legault, Nicolas Kassis, Rémi Bernard, Shawn Brown, Sonam Tamang, Tarek Sherif, Tristan Aumentado-Armstrong, Verena Schuster



HEALTHY BRAINS
HEALTHY LIVES



Fondation
Brain Canada
Foundation



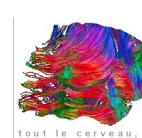
CANADA
FIRST
RESEARCH
EXCELLENCE
FUND



McGILL

MCIN MCGILL CENTRE
for INTEGRATIVE
NEUROSCIENCE

canarie



Centre de neuroimagerie
McConnell
Brain Imaging Center
tout le cerveau, en images... everything brain imaging



CENTRE
LUDMER
CENTER
NEUROINFORMATIQUE & SANTÉ MENTALE | NEUROINFORMATICS & MENTAL HEALTH