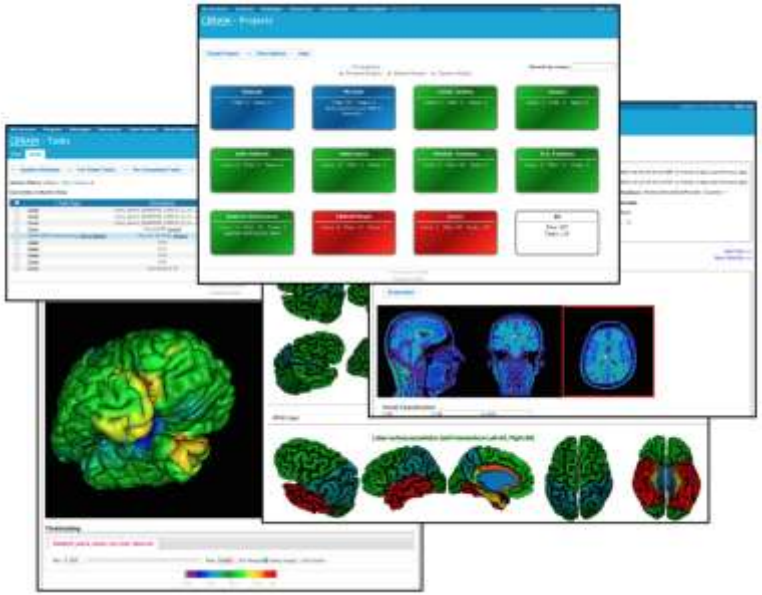
MNI ECOSYSTEM

neuro
Montreal Neurological Institute and Hospital
Institut et hôpital neurologiques de Montréal

LORIS
cbrain
LUDMER
MCIN

database management & data-sharing



CBRAIN and LORIS are both cloud infrastructures

CBRAIN

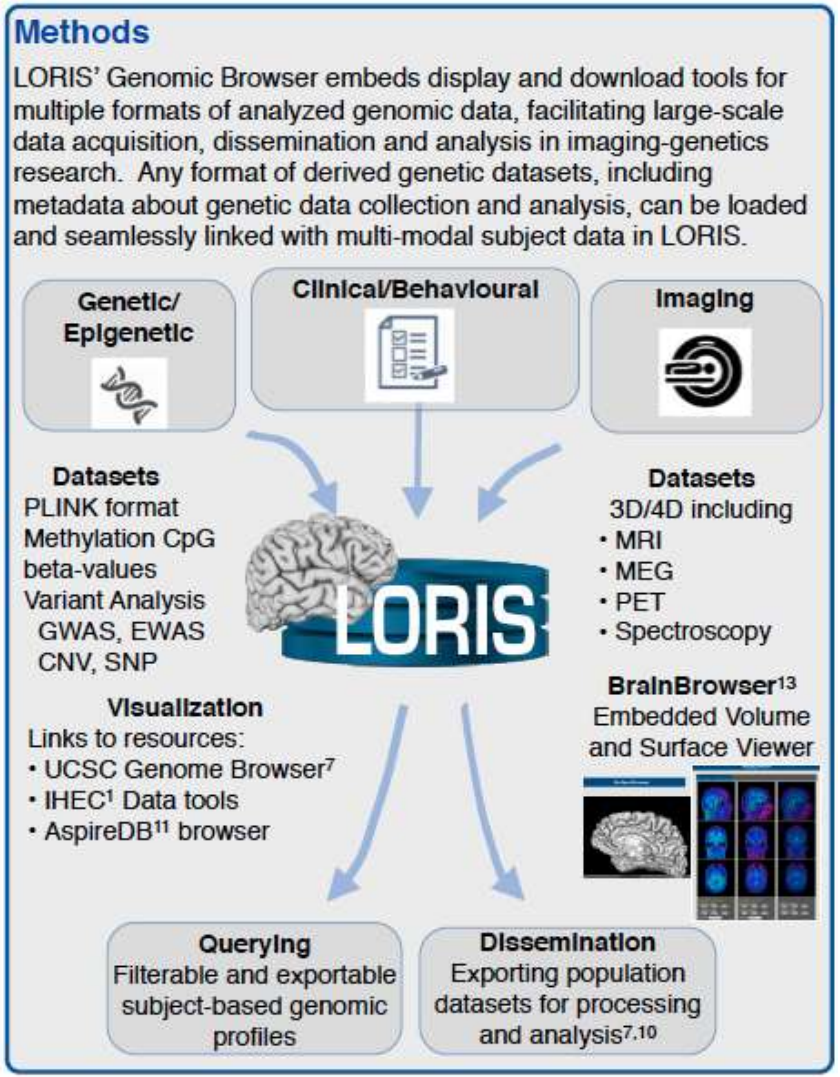
- cloud-based web portal
- uses Compute Canada as its cloud backbone
- interoperable with other cloud services
- uses Boutiques to describe pipelines

LORIS

- can be served from any cloud resource
- has built-in API functionality
- incorporates standardization to facilitate cloud usage

Data Types

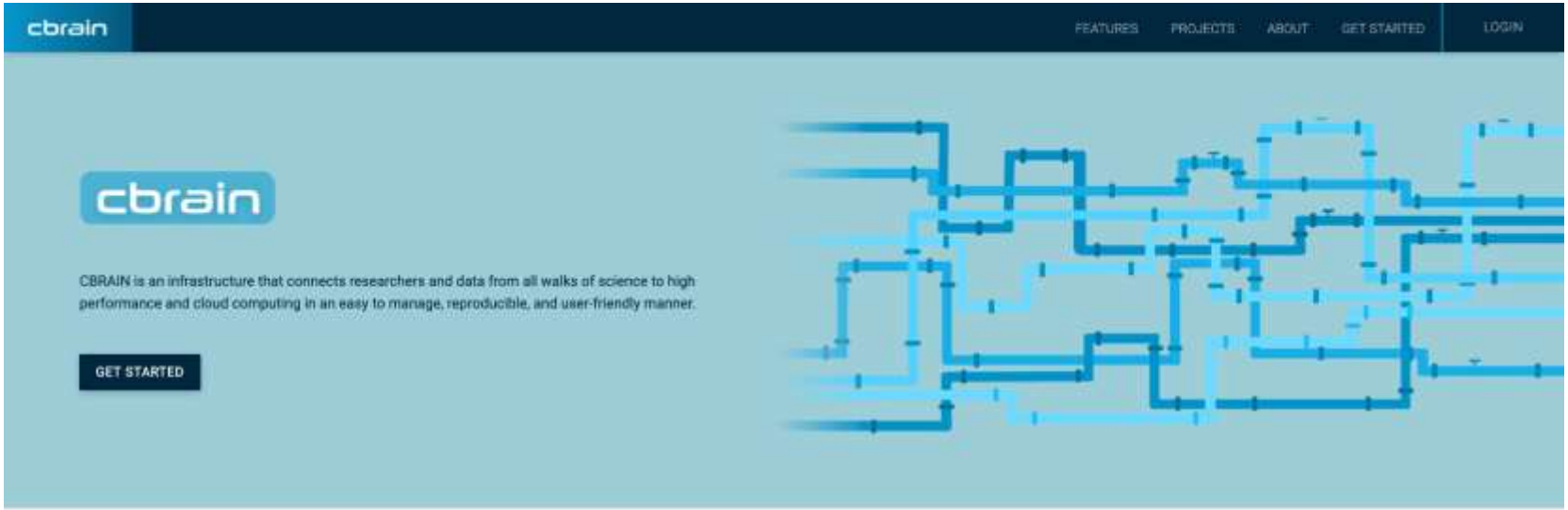
Imaging
Behavior
Genetics
Epigenetics
Tissue samples



International Partners

For list of 400+ measures: https://sites.google.com/site/lorisinstrumentlist/

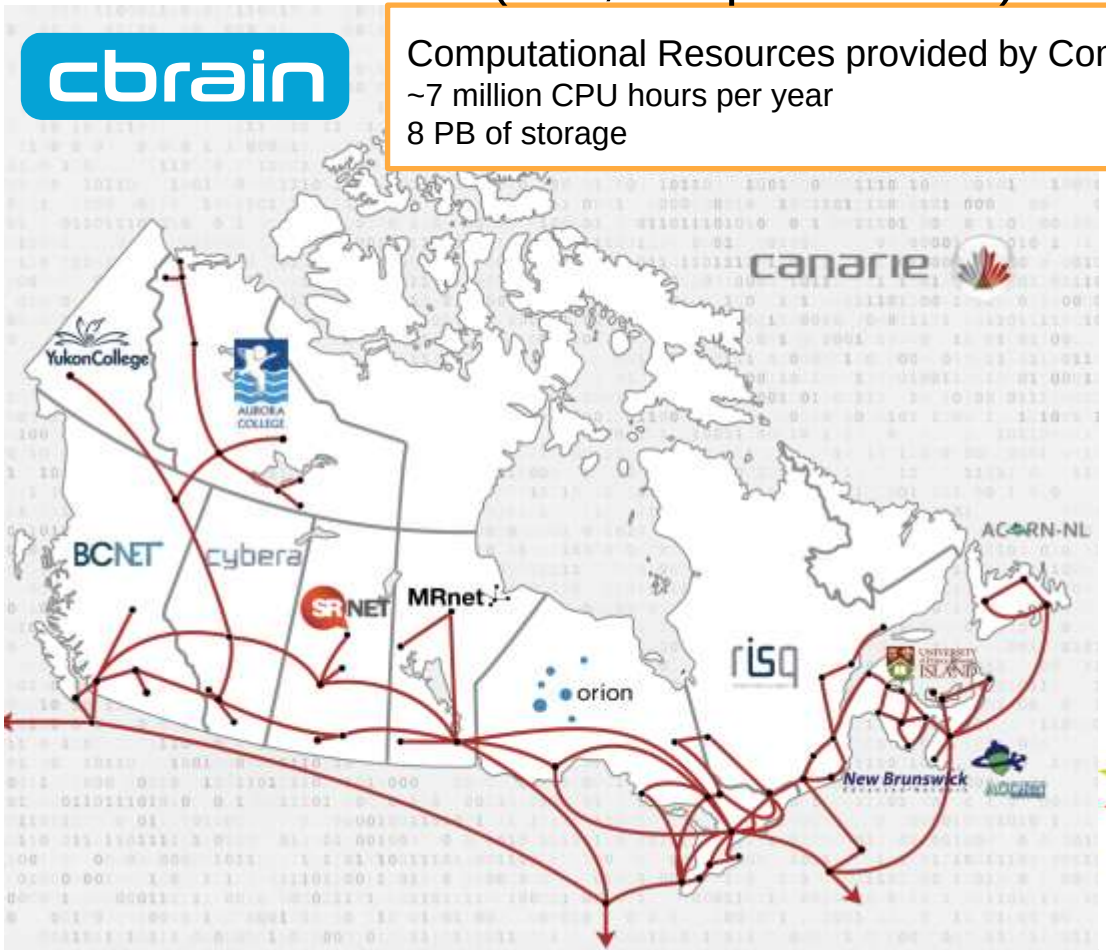
On the web at http://cbrain.ca



CBRAIN National HPC Integration
(300,000 processors)

cbrain

Computational Resources provided by Compute Canada
~7 million CPU hours per year
8 PB of storage



compute canada | calcul canada

CBRAIN / Texas Advanced Computing Center (TACC) University of Texas at Austin

MCIN MCGILL CENTRE for INTEGRATIVE NEUROSCIENCE

cbrain

High Performance Computing Portal

XSEDE Extreme Science and Engineering Discovery Environment

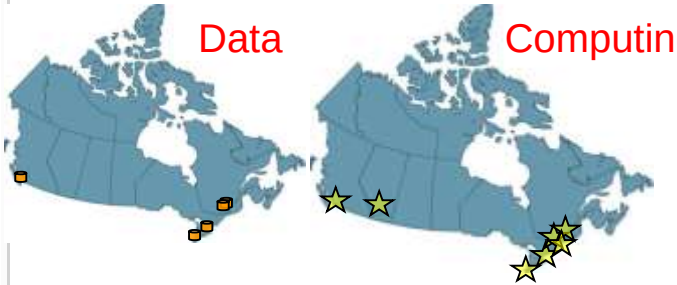
TACC

Stampede2

Largest supercomputer in the world for academic science

18 Petaflops (~370,000 cores)

CBRAIN Canadian Distributed Neuroinformatics Platform



Distributed Architecture

compute • calcul CANADA

canarie

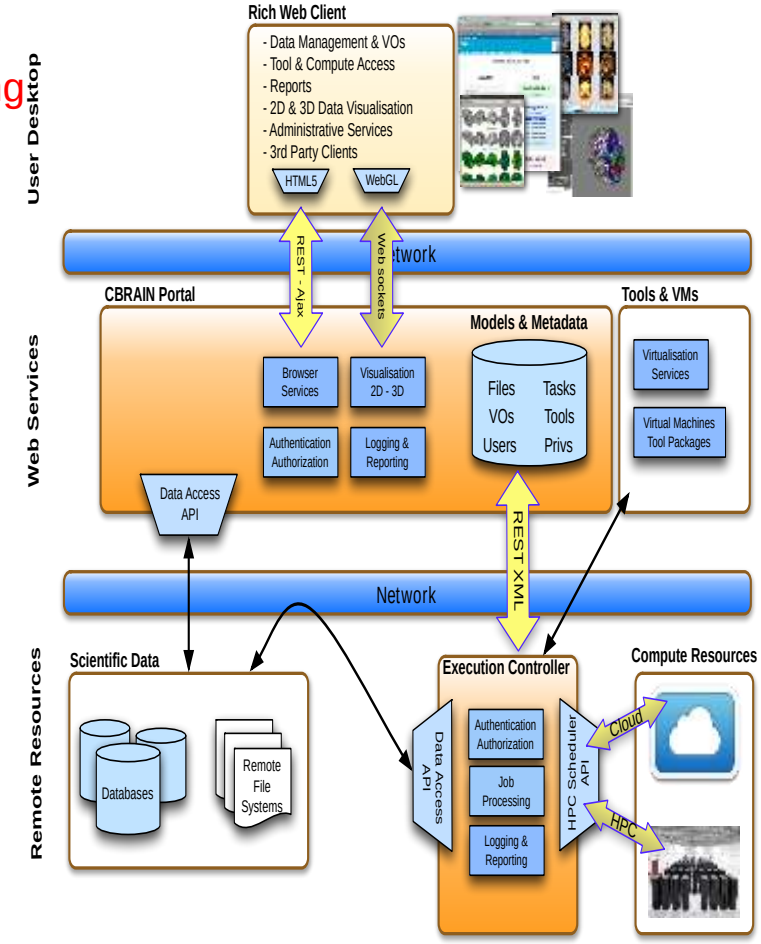
UBC

Western

ARC·CNRC

Université de Montréal

McGill



https://github.com/boutiques

“Boutiques”

Glatard T et al., 2018

- Fully-automated integration of applications
- Deployment on heterogeneous computing resources through containers
- Comprehensive input validation through a strict JSON schema
- Flexible application description through a rich JSON schema

OXFORD (GIGA)n SCIENCE

Giga Science, 7, 2018, 1–11

doi: 10.1093/gigascience/giy016
Advance Access Publication Date: 23 March 2018
Technical Note

TECHNICAL NOTE

Boutiques: a flexible framework to integrate command-line applications in computing platforms

Tristan Glatard[1,*], Gregory Kiar[2,3], Tristan Aumentado-Armstrong[2,3], Natacha Beck[2,3], Pierre Bellec[4], Rémi Bernard[2,3], Axel Bonnet[5], Shawn T Brown[2,3], Sorina Camarasu-Pop[5], Frédéric Cervenansky[5], Samir Das[2,3], Rafael Ferreira da Silva[6], Guillaume Flandin[7], Pascal Girard[5], Krzysztof J. Gorgolewski[8], Charles R.G. Guttmann[9], Valérie Hayot-Sasson[1], Pierre-Olivier Quirion[4], Pierre Rioux[2,3], Marc-Étienne Rousseau[10] and Alan C. Evans[2,3]

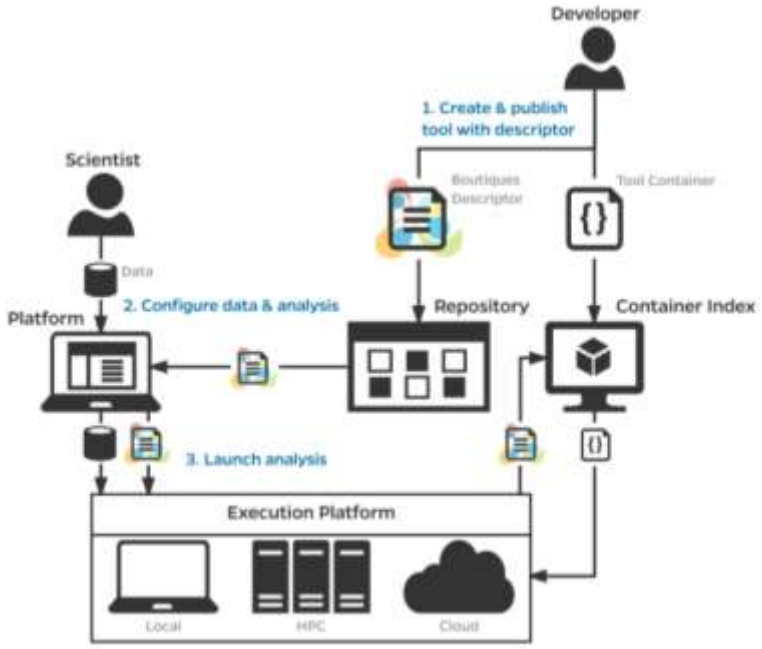


Figure 1. Publication, integration, and execution of applications with Boutiques.

Pipelines in CBRAIN

CBRAIN has over 50 execution pipelines available to users

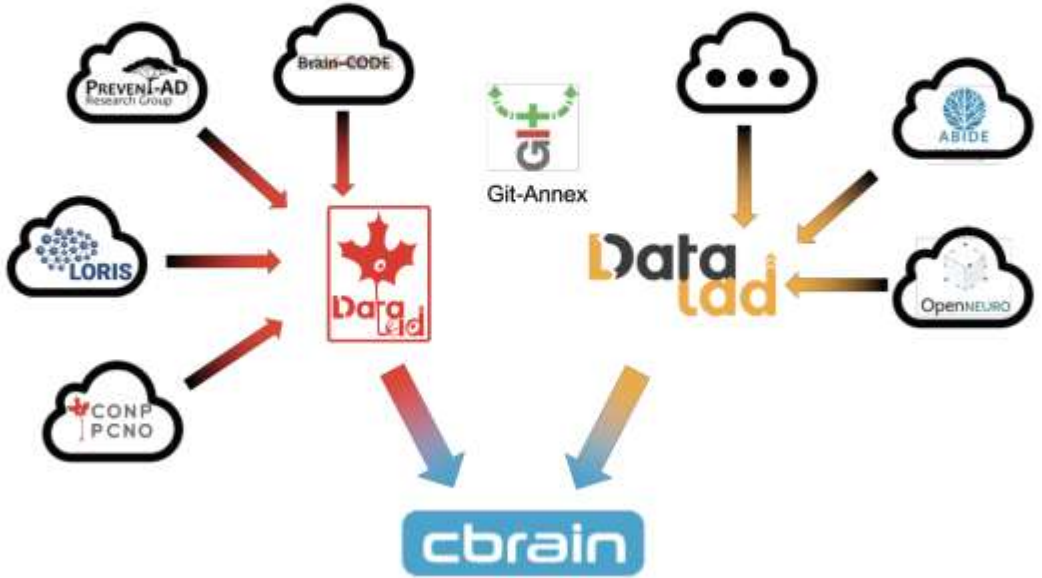
Software	Pipelines
Civet 2.1	Civet Quality Control Tool, Civet Structural Processing Pipelines, NuCorrect
FSL 5.0	FslBedpostx, FslBet, FslFast, FslFirst, FslFlirt, FslAnat, FslRandomise, FslMelodic, FslProbtrackX, FslProbtrackX2, FslSub, FslFeat
FreeSurfer 6.0.0	ReconAll, ReconAll.ongl
MINCTools	Mincaverage, Minmac, Mincpik, Mincresample, Mnc2nii, Nii2mnc, Dcm2mnc, MINCBet, MincConvert, Dcm2nii
Ants 2.1.0	AntsRegistration
Niak	FRMI/T1 Preprocessing
HERMES	Supply Chain Simulation Model
PSOM	PSOM Worker Launching
INCA-AROMA	Functional MRI Pipeline
NDMG	Connectome Estimation
QEEG	EEG / MEG Analysis
pCEV	Principal Component analysis

Datalad Integration

Datalad is a Python tool that builds on top of git-annex and extends it with an intuitive command-line interface to enable transparently operating and managing data.

A CBRAIN Datalad DataProvider allows data from any Datalad repository to be imported into CBRAIN for pipeline execution.

Pull data from the Canadian Open Neuroinformatics Platform

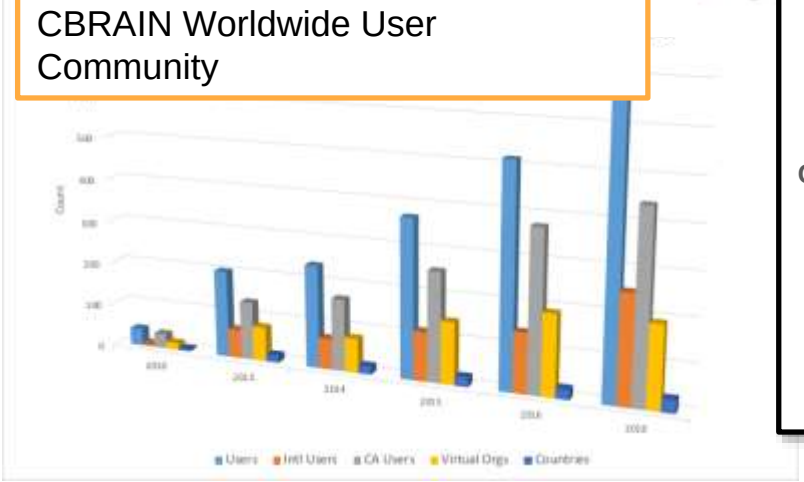


Pull data from any Datalad enabled resource

If data is in Datalad, it is now automatically available in CBRAIN!

cbrain

CBRAIN Worldwide User Community



Users

- > 900 Users (100s active at a given time)
- 193 Sites from 30 different countries
- > 30 Million CPU Hours served
- > 1 Million Files (100s TB of data)

Development Team

- 5 Full Time Developers

Support

- Canadian Foundation for Innovation
- CANARIE
- Healthy Brains for Healthy Lives (HBHL)
- Canadian Open Neuroscience Platform (CONP)
- Canada/Cuba/China Axis (CCCAxis)

Collaborations

- Big Data Infrastructures in Neuroinformatics (Glatard)
- CARMIN Project
- Compute Canada
- Texas Advanced Computing Center. (TACC)
- Pittsburgh Supercomputing Center (PSC)
- PERFORM Centre
- Human Brain Project
- OpenNeuro Platform

Become a CBRAIN User

- Sign Up for an account (completely free!)
  - http://portal.cbrain.mcgill.ca

MNI data-sharing and processing ecosystem

Open-source, documented code repositories

- CBRAIN https://github.com/aces/cbrain
- LORIS https://github.com/aces/Loris

Canadian Open Neuroscience Platform

https://conp.ca/

Organizational Design

Scalability

Interoperability

Analysis Packages

Training

International Partnerships

Ethics and Data Governance

Communications Platform

Public Release of Prospective AD cohort

PREVENT-AD

Global Brain Consortium

Focused on **EEG**, behaviour, interoperability, outreach to LMIC
GBC workshop in Q2, 2019 (~50 people)
World Health Organization (OHBM-WHO Geneva meeting, 2017)

GBC Steering committee

Gary Egan
Monash U.
Australian Brain Alliance

Maryann Martone
UCSD
INCF, NIF

Jean-Baptiste Poiine
McGill U.
CONP, INCF

Katrin Amunts
Julich
HBP

Alan Evans
McGill U.
CONP, INCF

Jane Roskams
UBC, CONP,
Cascadia Data Alliance

Pedro Valdez-Sosa
UESTC/CNEURO
China/Cuba/CCC

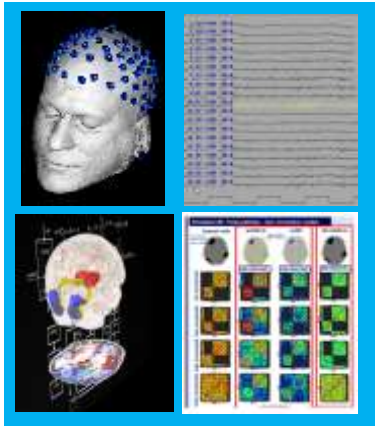
Bartha Knoppers
McGill U.
CONP, GA4GH

Paul Thompson
USC
ENIGMA

Greg Farber
NIH
US Brain Initiative

World Health Organization

4D spatiotemporal dynamics



LMIC applications

Interoperability with the cloud is key

- Our platforms have the capability to already interact with the cloud

Fully documented APIs are imperative !

- Our platforms have fully documented RESTful APIs
- Using our APIs can fully interact with the cloud for storage

Standardization is key to facilitating cloud based initiatives

- Independent of where data is stored, but needs to be done

Using cloud resources can be $$ (i.e. redundant data transfers)

- Depends on the use-case (cost of maintaining storage vs. cloud)
- Once we get to the PB level, cloud storage becomes attractive
- Neuroimaging datasets are made up many small files (ideal for object stores)

Containerization is a best practice (e.g. Docker, Singularity)

- Can take advantage of more scalable technologies (e.g. Kubernetes)

Neuroscience workflows are amenable to cloud computing

- Many small memory (4-16GB) naturally parallel processes
- Requires a large amount of data movement and orchestration

Having data storage close to processing can be efficient

- If cloud computing is used, data transfer will be cheaper
- If compute is outside the cloud, transfer much more expensive

Important cloud considerations

Potential cloud risks and limitations

Security is outsourced to provider to some degree

- Cloud providers provide differing levels of secured storage

Privacy might be in the fine print

- Must be careful the provider doesn’t suddenly own data (e.g. Dropbox)

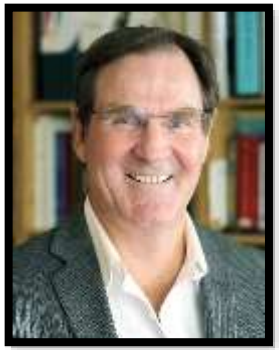
Prohibitively expensive for very large memory workflows

- Still more cost-effective to perform on traditional HPC resources

Cost structure is much more complex

- Users left to their own devices could rack up large bills
- Poorly implemented infrastructures could lead to large costs for users
- Funding agencies will now have every grant paying for computing
- Should explore large cloud purchase for the whole community

CBRAIN Team



Alan Evans

Shawn T. Brown

Pierre Rioux

Reza Adalat

Natasha Beck

Darcy Quesnel

Najma Mahani

Tristan Glatard

Candice Czech

Serge Boroday

Xavier Leours-Boucher

Gregory Kiar