MCIN MCGILL CENTRE for INTEGRATIVE NEUROSCIENCE

Evolution of Neurosciences:
Cyberinfrastructure Needs for a National Neuroinformatics Platform.

Dr. Alan C. Evans
Director - McGill Center for Integrative Neuroscience
Montreal Neurological Institute, McGill University
Montreal, QC
Contact: alan.evans@mcgill.ca

February 25, 2016

MCIN MCGILL CENTRE for INTEGRATIVE NEUROSCIENCE

The rapid transformation of neurosciences

We are at the beginning of a new era in brain research, an era that combines the burgeoning capabilities of the IT revolution with powerful emerging data acquisition technologies. On all sides, we are surrounded by new techniques for capturing neuroimaging, genetic, epigenetic, various 'omics, and behavioural data non-invasively during development, learning, aging and plasticity. These techniques allow us to investigate the pathological mechanisms that underlie brain disorders as well as their environmental and genetic origins. We apply these techniques in animal models, where environmental and genetic manipulations provide a unique window that is not possible with human studies. Let alone data growth due to DNA sequencing and neuro-genetics, recent post-mortem ultra high-resolution imaging techniques, used to create the first microscopic-scale full brain atlases, can generate tens of petabytes for a single rodent brain; a human brain would produce up an estimated 200 exabytes[1]. Neuroscience is overwhelmed by a veritable tsunami of multi-modal data at all spatial scales. The datasets are getting larger as we increase spatial and temporal resolution. Moreover, the complexity of brain analysis is growing rapidly as we consider (i) the N X N problem of pairwise interaction among all 3D brain locations, i.e. "connectomics", where N may be 10^4-10^6, at multiple time points, or (ii) imaging-genetic analysis where we seek to solve a complex genetic analysis across perhaps 10^6 genome locations (0.1% of the entire genome), in 1000's of subjects…..at every one of ~10^6 3D image locations. The analytic strategies being developed for extracting useful information from this wealth of raw data, sometimes termed 'big data analytics', involve a vast range of univariate and multivariate statistical approaches, such as the various flavours of machine learning. The challenge of the coming decade in brain research will be to adapt our analysis toolsets and data management processes to integrate massively multi-dimensional data into a holistic model of brain organization in order to gain insights on its normal trajectory through the lifespan and the perturbations that represent disease or dysfunction. Only with a profound understanding of the brain will we be able to design effective treatments for brain illnesses that shatter lives and inflict a massive clinical and socio-economic burden on society.

Our group, the McGill Center for Integrative Neuroinscience (MCIN), has been at the forefront of this evolution for the past two decades, providing the first comprehensive platforms for large multi-center studies and portals for collaborative distributed computing. Dr. Evans' MCIN has been selected to lead large cyber-platform efforts for several national and international projects. Recent awards from CIHR and Brain Canada establishes our group as the platform technology leader to support large Canadian multi-center consortia such as; NeuroDevNet NCE, Canadian Consortium on Neurodegeneration in Aging (CCNA), Quebec Alzheimer's Neuroimaging Network (CIMA-Q) [for more information, please see project links at end of document]. Through these applied research initiatives, MCIN represents tens of Canadian neuroscience research sites and hundreds of researchers. In addition, MCIN is a core partner to major international projects, as per our central work in U.S. Autism Centres of Excellence - Infant Brain Imaging Study (IBIS), the European Human Brain Project (HBP), the Indian Dementia Imaging Network (IBRAIN) and the International Consortium for Brain Mapping (ICBM). A representative sample of direct partners for these projects is shown in Table 1.

1 http://www.nature.com/news/neuroscience-solving-the-brain-1.13382

Name	Position	Institution	Country	Expertise
Alan Evans	Professor	McGill U.	Canada	Neuroinformatics
Louis Collins	Professor	McGill U.	Canada	Image processing
Bruce Pike	Professor	U. of Calgary	Canada	MRI physics
Oury Monchi	Professor	U. of Calgary	Canada	Cognitive neuroscience
Stephen Strother	Professor	U. of Toronto	Canada	Neuroinformatics
Stephen Scherer	Professor	U. of Toronto	Canada	Neurogenetics
Peter Szatmari	Professor	U. of Toronto	Canada	Autism, Neurogenetics
Julien Doyon	Professor	U. de Montreal	Canada	Cognitive Neuroscience
Pierre Bellec	Professor	U. de Montreal	Canada	fMRI connectivity analysis
Patricia Conrod	Professor	U. de Montreal	Canada	Neurodevelopment
Sebastien Jacquemont	Assoc Prof	U. de Montreal	Canada	Neurogenetics
Alain Dagher	Professor	McGill U.	Canada	Neurology
Mayada Elsabbagh	Asst Prof	McGill U.	Canada	Autism, EEG, imaging
Jon Stoessl	Professor	U. British Columbia	Canada	Neurology
Philip Awadalla	Professor	U. de Montreal	Canada	Genetics
Michael Meaney	Professor	McGIll University	Canada	Epigenetics
Celia Greenwood	Assoc Prof	McGill University	Canada	Genetics
Faisal Beg	Professor	Simon Fraser U.	Canada	Image Processing
Vesna Sossi	Professor	U. British Columbia	Canada	PET physics
Maxime Descoteaux	Assoc Prof	U. de Sherbrooke	Canada	Diffusion imaging
Simon Duchesne	Assoc Prof	U. de Laval	Canada	Neuroinformatics
Christophe Grova	Professor	Concordia U.	Canada	Multi-modal analysis
Louis Bherer	Professor	Concordia U.	Canada	Cognitive Neuroscience
Jason Steffener	Asst Prof	Concordia U.	Canada	Multi-modal analyis
Habib Benali	Professor	Concordia U.	Canada	Statistics, modelling
Jason Lerch	Assoc Prof	U. of Toronto	Canada	Mouse imaging-genetics
Christian Beaulieu	Assoc Prof	U. of Alberta	Canada	Diffusion Imaging
Barry Bedell	Assoc Prof	McGill U.	Canada	Animal imaging,pathology
Sultan Darvesh	Professor	Dalhousie U.	Canada	Neuropathology
Howard Chertkow	Professor	McGill U.	Canada	Neurology
Guy Rouleau	Professor	McGill U.	Canada	Neurology, genetics
Tristan Glatard	Asst Prof	U. de Lyon	France	Computer Science
Roberto Toro	Researcher	Institut Pasteur	France	Neuroinformatics, genetics
Hasse Karlsson	Professor	U. of Turku	Finland	Neurodevelopment
Jean-Francois Mangin	Group leader	Neurospin	France	Databasing
Katrin Amunts	Professor	Julich Research C.	Germany	Neuroanatomy
Jong-Min Lee	Professor	Hanyang U.	S. Korea	Image processing
Yong He	Assoc Prof	Beijing Normal U.	China	Neuroinformatics
David Glahn	Professor	Yale U.	US	Imaging-Genetics
Jim Hudziak	Professor	U. of Vermont	US	Genetics, Psychiatry
Pedro Valdes-Sosa	Professor	Neurosci Centre	Cuba	EEG, statistics
Dezhong Yao	Professor	UESTC	China	Neuroinformatics

Table 1: CBRAIN participants and international collaborators, using Brain Canada-funded national networks.

Established platforms

Through our LORIS and CBRAIN platforms, the MCIN team coordinate and facilitate large neuroscience studies and data processing for hundreds of collaborators around the globe (Fig. 1). Closer to high-performance computing, the CBRAIN web-platform currently interacts transparently with 11 computing clusters and 40 remote data sources across the country and the world. **The seven largest of these computing clusters are located in major Compute Canada sites.** Since 2009, the CBRAIN platform has been used by **roughly 420 scientists/labs spread over 50 cities in 22 countries** (Fig. 2) to launch millions of individual operations on brain scans (**70% of users are our Canadian**, others are international).

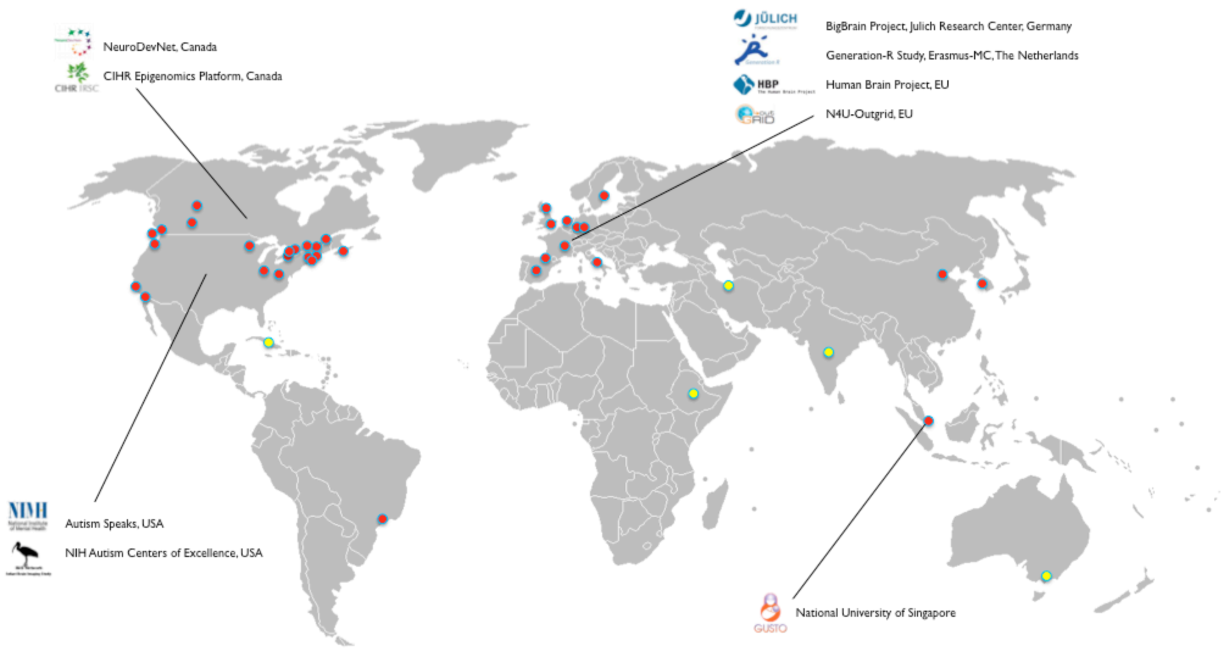

Figure 1: **Geographical spread of LORIS and CBRAIN collaborators**. Some high profile projects are outlined.

Most of these portal users do not have any of the experience normally required to interact with high-performance supercomputers. Our community now benefits from an unmatched ability to share analysis and visualization tools, computing resources and data online, entirely within the ease and comfort of a web-browser. This has the immediate benefit of standardizing tools, methods and traceability, thus ensuring better experimental result reproducibility between dispersed teams; a currently problematic area where we are actively promoting unified platforms. Importantly, the CBRAIN platform code itself is generic in nature, it can accommodate any data type and any scriptable tools, whether in life sciences, astronomy, physics or humanities. Not only major published research advances have been directly supported by these technologies, our group has maintained a respected Canadian leadership presence at an international level.

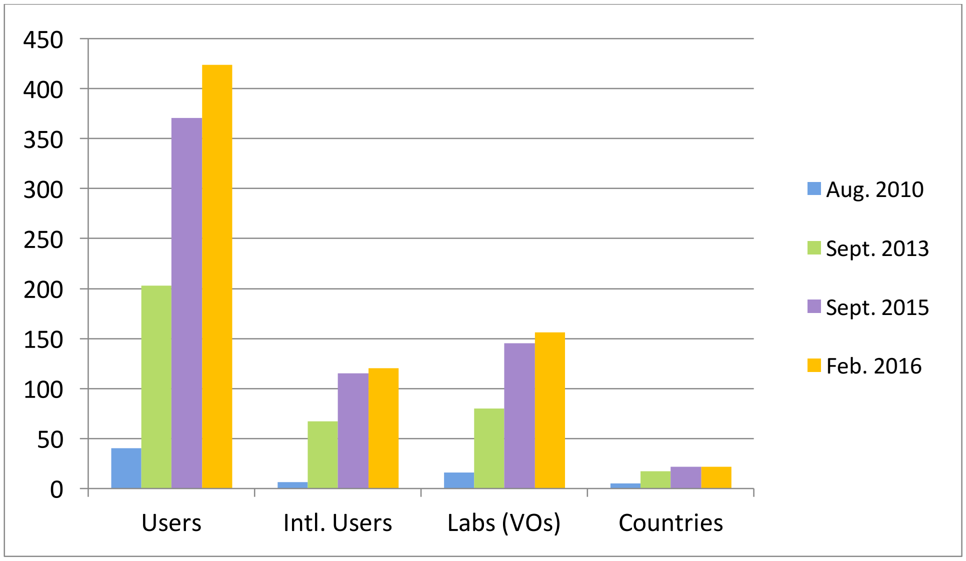

Figure 2: Growth of CBRAIN user community between 2010 and 2016. This histogram shows the number of total registered accounts, international users, virtual laboratories and countries within CBRAIN for 4 time points between early production in 2010 and February 2016. We maintained a constant proportion of approximately 70% Canadian core user base across this timeline.

Importance of Compute Canada for research platforms

In the early phases of design and deployment, tight collaboration with Compute Canada sites from coast-to-coast has played a paramount role in the success of CBRAIN. Their service model, at the time, was not oriented towards automated platform support; mutual trust and dialogue had to be established in order to coordinate technical aspects. After using tens of millions of computing hours on their systems, we feel it is important to state that Compute Canada's staff open-mindedness, efficiency and expertise allowed us to progress at a speed unthinkable in most other international cyberinfrastructure projects we have been involved with. In multiple international projects, we have been able with the support of Compute Canada sites to perform major research breakthroughs that better equipped institutions in other countries simply could not deliver on, despite firm agreements and sustained coordination efforts.

This type of flexible service model and specialized support is extremely difficult to find in comparable research infrastructure or commercial clouds providers. Due to the nature of their activities, research platforms need reasonable fluidity to remain relevant and drive innovation at a global level. For this to be a reality, research cyberinfrastructure services cannot always be reduced to the commercial model of a uniform access to resources and a uniform support offering mapped to a simple scaling fees structure. We are in strong support of maintaining and, more importantly, further developing Compute Canada cyberinfrastructure capabilities for the benefit of all Canadian research platforms and research and general.

Anticipated cyberinfrastructure needs

To support these increasing needs for our neuroscientific community alone, multiple important trends have to be considered: the past average growth of our in-house systems (non-Compute Canada), our Compute Canada usage, and upcoming "game changing" data-related technologies. With the clear indication that future national platform infrastructure investments will be handled by Compute Canada on behalf of the researchers, our in-house infrastructure needs will to be addressed within Compute Canada. A summary of anticipated needs is presented in Table 2.

Current in-house resources: The MCIN currently develops, operates and monitors on a 24/7 basis; platforms, visualization services, databases, web sites and portals, storage, high performance networking and computing servers serving dozens of research projects for hundreds of scientists. **In total, MCIN hosts over 60 virtualized production services, 200+ CPU core years, and close to 1PB of total disk storage** (as well as an insufficient 400TB of tape storage). Past growth indicates that we should **forecast these needs to triple every 2 years**.

Current Compute Canada resources: Our platform allocation obtained for 2015-2018 anticipate 1200 CPU core years and 0.6PB disk storage by 2017. These represented as close as realistic estimates as possible for regular operations, **prior to the knowledge of the shift against in-house infrastructure investments**. In addition, our workloads are also evolving towards higher memory requirements, making traditional high performance computing nodes less suitable and requiring a realignment of our needs.

Game changing data acquisition: As discussed in the introduction, several new acquisition methods generate unprecedented amounts of data. While we cannot ignore them, and are starting to experiment and plan new tools and methodologies for multi-petabyte level datasets, we could not in the right mind justify a demand for exabytes of disk storage in this document. However, there is little doubt this is where neuroscience research is heading; our data administrators have to handle new datasets which are growing at an alarming speed, old paradigms and expectations are shattered on a regular basis.

	2016	**2018**	**2020**
Hosting (dedicated cores)	150	450	1,350
Compute (cores years)	1,200	3,600	15,000
Disk Storage (PB)	1.6	6	25
Tape Storage (PB)	0.5	3	12

Table 2: Anticipated cyberinfrastructure needs for MCIN related platforms. Conservative assumptions based on current growth and trajectory: we used our minimal projection showing a tripling of resources every 2 years, with a slight increase of the factor for 2020 to account for new acquisition methods.

MCIN MCGILL CENTRE for INTEGRATIVE NEUROSCIENCE

Platform related publications and references

McGill Center for Integrative Neuroscience

http://mcin-cnim.ca/

Technology papers

CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research, http://dx.doi.org/10.3389/fninf.2014.00054

LORIS: a web-based data management system for multi-center studies, http://dx.doi.org/10.3389/fninf.2011.00037

The MNI data-sharing and processing ecosystem, http://dx.doi.org/10.1016/j.neuroimage.2015.08.076

Controlling the Deployment of Virtual Machines on Clusters and Clouds for Scientific Computing in CBRAIN, http://dx.doi.org/10.1109/CCGrid.2014.42

Virtual imaging laboratories for marker discovery in neurodegenerative diseases, http://dx.doi.org/10.1038/nrneurol.2011.99

Dr. Evans' full publication history

https://scholar.google.ca/citations?user=FxPzh9kAAAAJ&hl=en

Partner Projects (sample)

http://ccna-ccnv.ca/
http://www.ibis-network.org/
www.cima-q.ca
http://www.neurodevnet.ca/
https://www.humanbrainproject.eu/

MCIN Open source projects

CBRAIN (platform): https://github.com/aces/cbrain
LORIS (platform): https://github.com/aces/Loris
BrainBrowser (3D web visualization): https://github.com/aces/brainbrowser
Boutiques (cross-platform tools exchange): http://boutiques.github.io