The MNI data-sharing and processing ecosystem

Samir Das [a,*], Tristan Glatard [a,c], Leigh C. MacIntyre [a], Cecile Madjar [b], Christine Rogers [a],
Marc-Etienne Rousseau [a], Pierre Rioux [a], Dave MacFarlane [a], Zia Mohades [a], Rathi Gnanasekaran [a],
Carolina Makowski [a], Penelope Kostopoulos [a], Reza Adalat [a], Najmeh Khalili-Mahani [a], Guiomar Niso [d],
Jeremy T. Moreau [d], Alan C. Evans [a]

[a] McGill Centre for Integrative Neuroscience (MCIN), Ludmer Centre for Neuroinformatics and Mental Health, Montreal Neurological Institute (MNI), McGill University, Montreal, Canada
[b] Douglas Mental Health University Institute, Montreal, Canada
[c] Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon ; Université Claude Bernard Lyon 1, France.
[d] McConnell Brain Imaging Centre (BIC), Montreal Neurological Institute (MNI), McGill University, Montreal, Canada

ARTICLE INFO

ABSTRACT

Neuroimaging has been facing a data deluge characterized by the exponential growth of both raw and processed data. As a result, mining the massive quantities of digital data collected in these studies offers unprecedented opportunities and has become paramount for today's research. As the neuroimaging community enters the world of "Big Data", there has been a concerted push for enhanced sharing initiatives, whether within a multisite study, across studies, or federated and shared publicly. This article will focus on the database and processing ecosystem developed at the Montreal Neurological Institute (MNI) to support multicenter data acquisition both nationally and internationally, create database repositories, facilitate data-sharing initiatives, and leverage existing software toolkits for large-scale data processing.

Introduction

Neuroimaging is facing a deluge of both raw and processed data. The amount of data reported by studies published in the journal *NeuroImage* has doubled every 26 months since 1995 and is expected to reach 20 GB per study within a few years (Van Horn et al., 2014). As a result, mining the massive quantities of digital data collected in these studies offers unprecedented opportunities and has become paramount for today's research. As the neuroimaging community enters the world of "Big Data," there has been a concerted push for enhanced sharing initiatives, whether within a multisite study, across studies, or federated and publicly shared (Mennes et al., 2013; Poline et al., 2012; Van Horn et al., 2014).

A single study rarely achieves a sample size comparable to those found in larger-scale data repositories. Smaller initiatives are limited in the range of possible analyses that can be applied on their population of interest (Xuan et al., 2010). Sharing data within a larger community is a practical way to address these issues with greater statistical power and propose scientific questions beyond the scope of a single research group. Furthermore, the robustness of biological findings across different methods or processing architectures encourages confidence in the reproducibility of results, a fundamental requirement of good scientific practice (Glatard et al., 2015; Poline et al., 2012).

At the same time, cross-site data sharing brings with it a broad range of issues in terms of site/scanner compatibility (Jovicich et al. 2009, 2013, 2014) and the logistical challenges of IT interoperability. The acquisition, storage, and curation of neuroimaging and related data files has therefore become a major Big Data challenge in the context of online sharing initiatives (Dinov et al., 2014; Wood et al., 2014). Several database systems have been constructed to address these issues, such as COINS (Wood et al., 2014), LONI IDA (Van Horn & Toga, 2009), LORIS (Das et al., 2011), XNAT (Marcus et al., 2007; Gutman et al., 2014), and others. However, many are under-equipped to deal with the wide array of data-sharing challenges (Dinov et al., 2014).

This article focuses on the database management and processing ecosystem developed at the Montreal Neurological Institute (MNI) to support multisite data acquisition, both nationally and internationally. Specifically, this manuscript addresses the creation of database repositories, facilitation of sharing initiatives, and leveraging of existing software toolkits for large-scale processing.

Materials and methods

The MNI ecosystem consists of two principal – and independently-functioning – platforms, LORIS and CBRAIN. The multi-purpose longitudinal database, LORIS, was created for data management, quality control

* Corresponding author at: McGill Centre for Integrative Neuroscience (MCIN), McConnell Brain Imaging Centre (BIC), Montreal Neurological Institute (MNI), McGill University, 3801 University, WB-208, Montreal, QC H3A 2B4, Canada.
E-mail address: samir.das@mcgill.ca (S. Das).

(QC), and dissemination (Das et al., 2011), while CBRAIN was implemented as a collaborative high-performance computing (HPC) platform for processing (Sherif et al., 2014). Together, they provide a comprehensive environment for large-scale processing, combining remote visualization tools (Sherif et al., 2015) and fully automated pipelines such as CIVET (www.bic.mni.mcgill.ca/ServicesSoftware/CIVET; Kim et al., 2014; Macdonald et al., 2000), NIAK (Bellec et al., 2012) and other software pipelines (SPM-batch: www.fil.ion.ucl.ac.uk/spm; FSL: http://fsl.fmrib. ox.ac.uk/fsl/fslwiki; Freesurfer http://freesurfer-software.org) (Fig. 1).

A key feature of LORIS is joint management of heterogenous data (i.e., behavioral, clinical, imaging, genomics), often collected within a longitudinal design. Each stage, from clinical and imaging acquisition to processing and statistical analysis, was designed to create a logical flow for data handling (Fig. 2).

(i) Core

Built on LAMP technologies (Linux, Apache, MySQL, PHP), LORIS is an open-source platform (https://github.com/aces/Loris) designed as an extensible and modular system, organizing heterogeneous data for multisite longitudinal studies (Fig. 3).

Behavioral and Clinical data are typically entered or uploaded directly via a web browser, with real-time summary statistics to track each stage of acquisition. A customizable behavioral battery of psychometric measures provides metrics of ongoing data collection.

Imaging data are stored in both raw and processed forms. The original DICOM files are archived and images are then converted to MINC format (Vincent et al., 2004), where extensive provenance capabilities track processing using MINC tools. Other file formats (e.g., NIfTI, NRRD) are available for use with separate processing environments.

Genomic data are handled through a genomics module that enables hosting, visualization, and dissemination of summary genetic and epigenetic data. Typically, imported directly into LORIS from genetic analysis pipelines or other sharing platforms (e.g., CBRAIN), input files

are cross-linked with imaging and behavioral data. Integrated links to resources (e.g., UCSC Genome Browser) facilitate rapid visualization of results (Kent et al., 2002).

(ii) Quality control (QC)

Visualization and QC of image files are tailored to each modality and customized per type of review (on-site, centralized, manual, or automated). Supported imaging modalities include structural MRI (T1, T2, PD), functional MRI (fMRI, PET, SPECT), diffusion, spectroscopy, and electrophysiology (MEG, EEG). Independent tools, such as DTIPrep, have been integrated as part of the QC pipeline for raw and processed diffusion imaging data (Liu et al., 2010). In addition, a feedback module with configurable permissions allows users (such as radiologists) to assess scan quality, selecting from over 30 intensity, movement, coverage, QC annotations, and diffusion-specific artifacts. Finally, LORIS provides protocol checking, where violations are flagged in a real-time scan protocol violations module. After images have passed automated protocol checking, multiple reviewers can then examine an image for diagnostic quality and pathology within the radiological review module.

LORIS also includes a suite of behavioral QC modules, such as (i) the conflict resolver module for double data entry mismatches; (ii) the feedback module, enabling users to input and query QC flags and annotations for every field; (iii) data integrity flagging, which verifies every field for consistency and completeness; (iv) a certification tool for assessment and maintenance of examiner reliability across sites, an essential feature for longitudinal studies with significant staff turnover; and (v) real-time statistics.

(iii) Data query tool (DQT)

LORIS facilitates web-based data-sharing, complete with viewing, querying, and downloading capabilities, requiring just a web browser. The DQT (MacFarlane et al., 2014; https://github.com/aces/Data-Query-
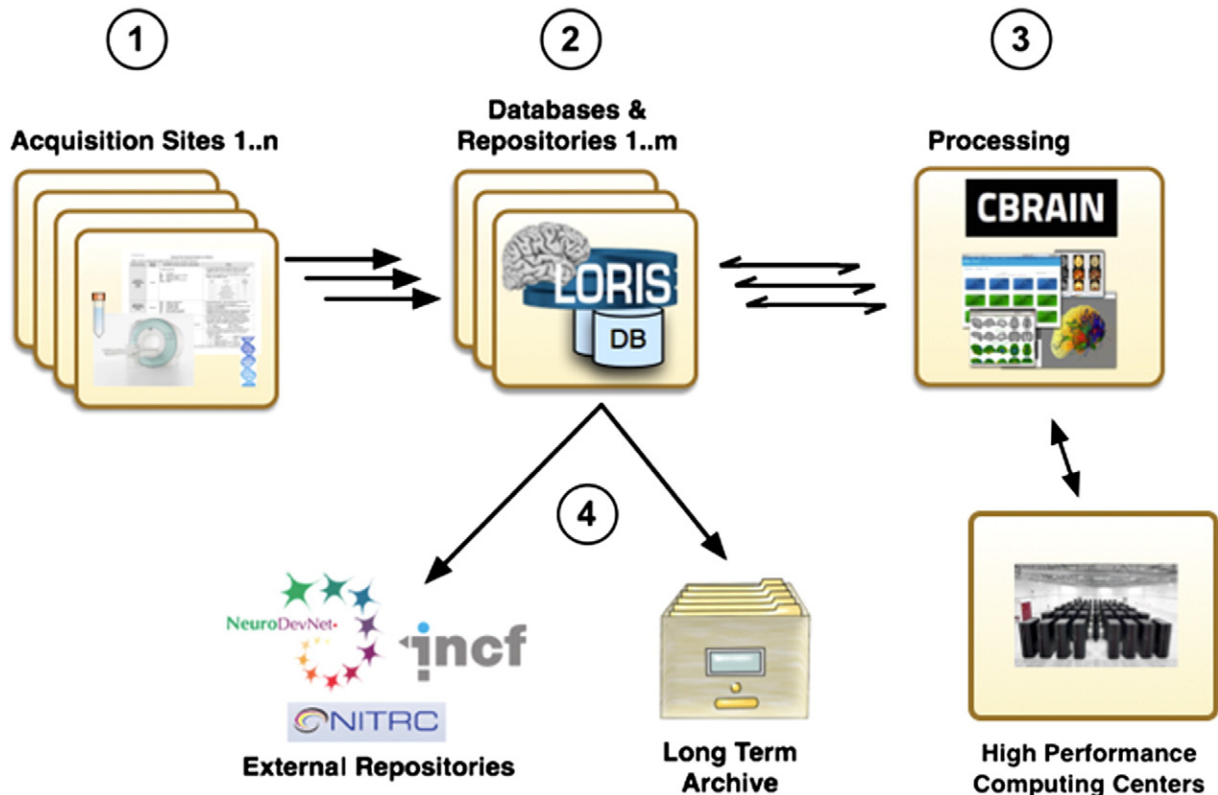
Fig. 1. LORIS and CBRAIN integration. Datasets hosted in LORIS' data-sharing platform can be processed via CBRAIN's processing utilities, and can be seamlessly disseminated to external pipelines and repositories.
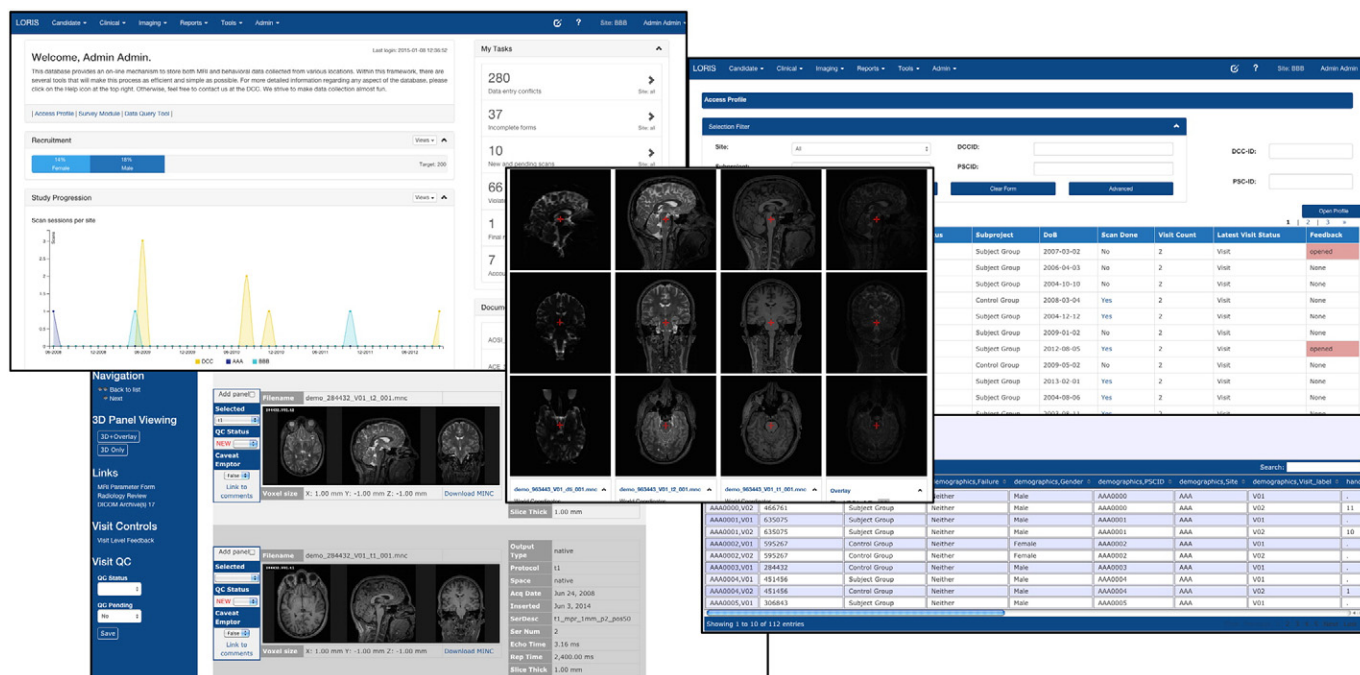
Fig. 2. LORIS Web Interface. LORIS provides an intuitive web-based interface for data entry, management, visualization and querying.
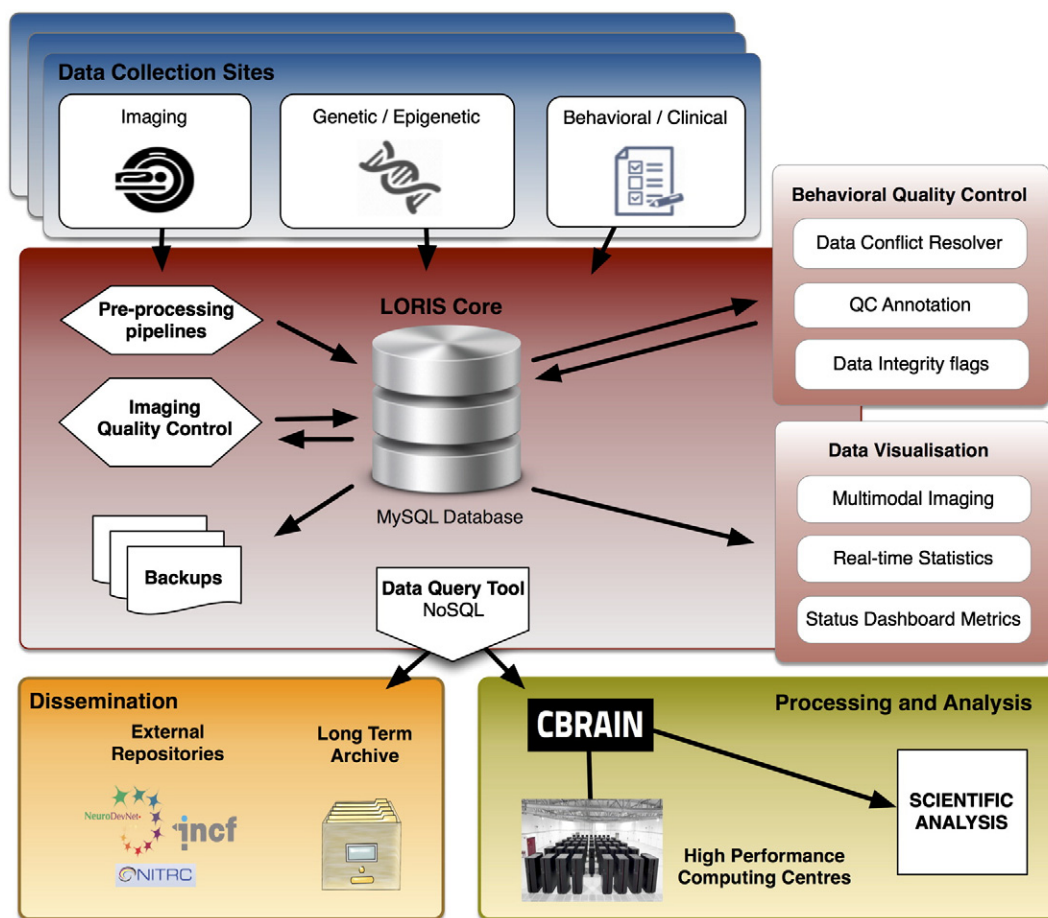

Fig. 3. LORIS architecture. LORIS acquires and stores multiple modalities of data (i.e., behavioral/clinical, imaging, genetic). Imaging data are stored independently from any image processing pipeline, enabling direct processing by external image analysis software. Dissemination tools allow users to aggregate, query and distribute subject data in a powerful and flexible manner to neuroimaging processing pipelines and analysis suites.

Tool) enables download of scalar data (e.g., clinical, behavioral) that can be simultaneously linked to multi-dimensional data (images) through a simple querying interface. This tool uses simple boolean algebra to customize searches and filter results, either longitudinally or cross- sectionally. Typically, this allows simultaneous querying from multiple researchers within a LORIS-based project but can also be configured for querying across multiple LORIS repositories. For future cross-platform interoperability, the extensible NoSQL architecture (i.e., CouchDB) facilitates federated sharing with other platforms.

(iv) Data processing via CBRAIN

CBRAIN (Sherif et al., 2014) is a web platform that enables distributed execution of software pipelines on computing networks and aggregates multiple distributed file systems into a uniform view (Fig. 4). The system was designed to abstract the complexities typically associated with the use of HPC resources for image processing.

CBRAIN supports multiple file servers called Data Providers, using various file transfer protocols, such as SFTP, SSH, and Amazon S3. Once files are registered on a data provider, they can be processed by any pipeline available in CBRAIN regardless of their physical location. To meet privacy, security and sharing requirements, access permissions can be defined on files and Data Providers—for example, CBRAIN users may keep their resources completely private, share them with specific groups, or make them public to the entire CBRAIN community. CBRAIN maintains a complete data provenance graph where an extensive log of all actions, parameters, software versions and locations are kept, and output files produced by a pipeline are linked to the inputs.

CBRAIN is linked with LORIS through a RESTful web API (Application Programming Interface) to facilitate data sharing. A specific LORIS data provider automates file transfer and registration operations between the two platforms, such that data can be processed using CBRAIN, and outputs re-inserted into LORIS.

Figure 5 exemplifies the interaction of three external platforms with CBRAIN: LONI (integration described in (Frisoni et al., 2011)), Virtual Imaging Platform (Glatard et al., 2013), and LORIS. All three projects use the CBRAIN API and Data Providers to share data or virtual machines (VMs), and to access computing resources.

A CBRAIN service is operated by the MNI, providing access to a national network of computing resources, which supports numerous projects. For security and support reasons, access to this service is restricted to MNI collaborators, as large resource allocations are limited to selected Canadian scientific projects. However, for testing purposes, trial accounts can be requested by emailing cbrain-support.mni@mcgill.ca.

The CBRAIN source code is freely available (GitHub: https://github.com/aces/cbrain, NITRC: www.nitrc.org/projects/cbrain). Detailed documentation allows anyone to install and configure a CBRAIN service on their own computing and storage resources (e.g., single machines, large

Fig. 4. CBRAIN web interface. CBRAIN provides an intuitive web-based interface for launching neuroimaging processing workloads on a distributed high-performance computing infrastructure and for built-in visualization utilities.
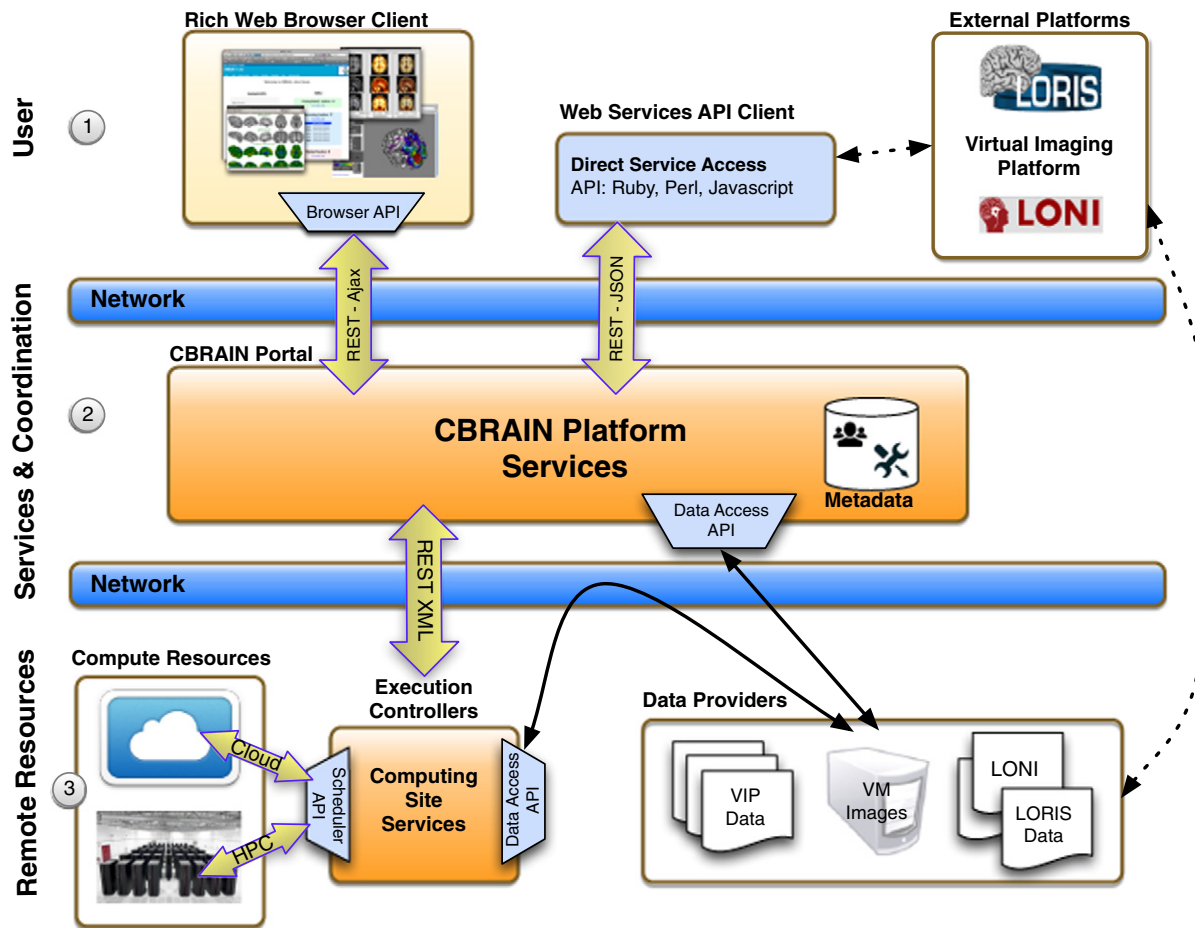
Fig. 5. CBRAIN architecture. The top User layer (1) exposes services to web browser clients or external platforms via an API. The central services and coordination layer (2) hosts the CBRAIN portal, which provides services and business logic for requests from the top User layer, and orchestration for the lower Resource layer. Metadata for users, tools, files, privileges, computing resources, etc. is stored in a central database. The lower remote resources layer (3) holds remote data providers (servers, databases or virtual machine images, and tool repositories) and execution controllers, which coordinate the computing sites (research HPCs, commercial cloud, or small servers). External platforms are loosely coupled to CBRAIN; they can interact with data providers (layer 3) and trigger any event through the Web API (layer 1).

national clusters, grids or clouds, etc.). Any scientific software installed on these resources (including licensed software) can technically be available through CBRAIN.

(v) Additional features

Beyond the foundation described previously, LORIS has supplementary functionality:

1. *Imaging Uploader* via a web browser
2. *Enhanced Visualization and QC tool*—BrainBrowser (Sherif et al., 2015)
3. *Instrument Builder* to create data forms in a web browser
4. Personalized dashboard
5. *Candidate Information Module* for status, demographics, and consent
6. *Data Dictionary Builder* for ontological consistency and codebook creation
7. Document repository
8. Mobile-friendly interface
9. *Survey Module* to email measures to participants
10. Configuration module
11. *REST API* for automated retrieval or modification of data

(vi) Requirements

To use CBRAIN or LORIS, a user simply requires a computer with a modern web browser. LORIS installation requires a Linux platform with LAMP stack, while CBRAIN is a Ruby on Rails application. Installation and customization of either infrastructure requires some technical expertise depending on the scale of deployment. An administrator with good knowledge of the dataset, ethics requirements, processing pipelines, and scientific questions would be an asset for assurance of high-quality data.

Results

LORIS (www.loris.ca) was first implemented at the MNI in 1999 to organize data acquisition and analysis for the multisite MRI Study of Normal Brain Development (Evans and Brain Development Cooperative, 2006). It is currently active in 130 sites worldwide with over 30,000 data collection time points, more than 500 behavioral instruments, and 75,000 variables hosted on its platform. Over 30 TBs of imaging datasets are stored in LORIS projects, comprising more than 200,000 acquisitions over 20,000 separate sessions.

CBRAIN (www.cbrain.ca) has been continuously operating at the MNI since 2009, hosting numerous raw and processed datasets. CBRAIN serves more than 300 users from 60 cities and 20 countries, including diverse projects such as PreventAD (Tremblay-Mercier et al., 2014), K-ADNI (Kim et al., 2014), the Lothian Birth Cohort (Karama et al., 2013), MAVAN, and NeuroDevNet (Goldowitz and McArthur, 2011). Users from 130 virtual sites access a data storage grid of more than 600 TBs across 25 storage servers. Between 2010 and 2014, CBRAIN users launched 246,000 jobs on an allocation of 24 million

CPU hours from Compute Canada (www.computecanada.ca) and private servers.

The MNI ecosystem serves as repositories for the following publicly available studies:

The NIH MRI Study of Normal Brain Development (Evans and Brain Development Cooperative, 2006): This was the first LORIS repository, completing the full lifecycle of a longitudinal project in 2007. Data were collected for more than 60 clinical instruments with over 8,000 variables, including raw and processed data from different MNI scanners. Data were shared with the research network through the DQT and were eventually federated to the NIH National Database for Autism Research (NDAR; https://ndar.nih.gov/) for public access, in accordance with the NIH Data Use Certification policy. Access procedures are detailed within the study's official website (http://pediatricmri.nih.gov/nihpd/info/index.html). The NIH has included additional processed data since 2007 and may consider the addition of supplementary data on a case-by-case basis.

Infant Brain Imaging Study (IBIS, Wolff et al., 2012): IBIS is an ongoing multisite U.S. Autism Spectrum Disorder study (www.ibisnetwork.org), collecting data from high-risk siblings of autistic probands. Data are incrementally federated to NDAR for public access (NDAR; https://ndar.nih.gov/access.html). The study has collected close to 10,000 phenotypic and imaging variables, with 4,000 subject time points. Approximately 3 TBs of imaging data are presently housed in IBIS, including 2,300 acquisitions that encompass a wide range of modalities, such as structural MRI, resting-state fMRI, diffusion tensor imaging (DTI), and MR spectroscopy. Selected genetic data (SNPs, CNVs) are also included. Collaborators of the IBIS project (e.g., EARLI, http://nationalautismnetwork.com/research/research-programs/earli-project.html) are able to contribute to this dataset, which can be organized as subprojects within the IBIS database.

BigBrain (Amunts et al., 2013): The BigBrain dataset (https://bigbrain.loris.ca) has been accessed by over 20,000 different visitors. It comprises 7,404 stained sections of a postmortem human brain, digitally reconstructed in 3D. The raw data, totaling approximately 1 TB at full resolution, are made freely available for download in slice format as MINC and NIfTI formats. Lower-resolution versions can also be downloaded in 3D for ease of visualization. Crowd sourcing efforts have recently been initiated enabling external research groups to contribute to processing and parcellation of the BigBrain.

The Autism Brain Imaging Data Exchange (ABIDE) (Di Martino et al., 2014): ABIDE (https://abide.loris.ca) employs LORIS' data-sharing platform in hosting 1,112 datasets, comprised of both resting-state and structural MRI, and federated phenotypic data acquired on 539 individuals with autism across 16 international sites. Public access to the ABIDE dataset is granted in accordance with the 1,000 Functional Connectomes Project data usage policies, as outlined at http://fcon_1000.projects.nitrc.org/indi/abide/. No new data are currently being collected for ABIDE; thus, all available data reflect the complete dataset for this neuroimaging initiative.

The Open MEG Archive (OMEGA) (Niso et al., 2015): Initiated in 2014, the Open MEG Archive (https://omega.bic.mni.mcgill.ca/main.php) is among the first diversified MEG open-science projects. It houses growing amounts of resting-state and task-related data, presently totaling 138 sessions from 79 study subjects. These datasets are publicly available for download via the OMEGA LORIS repository, including demographic and behavioral data, MEG, structural MRI for optimal source imaging, and processed data in Brainstorm format (Tadel et al., 2011). Access is freely granted to any individual who requests an account via the login page. The OMEGA repository is expected to house additional datasets, including electrophysiology (e.g., EEG, single cell and field recordings) and imaging (e.g., fMRI, DWI, PET) modalities.

Discussion

There is clearly an increased global effort to make data available across projects, borders, and technological platforms (Keator et al., 2009). As a result, data sharing has become a popular term of late. However, the disarmingly simple term obscures a multitude of challenges that go far beyond utilizing a database system to host data through a public website. The questions are almost endless: What type of data…? Sharing with whom…? For how long…? What transfer formats…? Who "owns" the data…? What privacy safeguards exist…? Data need to be managed and processed in a clear and auditable way, with unique consideration devoted to the curation and sharing of clinical data. For example, subject information cannot be released without consent (Brakewood and Poldrack, 2013; Choudhury et al., 2014; Mennes et al., 2013). LORIS has several mechanisms to address de-identification, both through removal of identifying data fields and more esoteric anonymization techniques, such as defacing of anatomical images, masking of DICOM headers, and addressing privacy issues, such as HIPAA compliance (Matlock et al., 2012).

Often lacking in data dissemination initiatives, documentation is a multi-faceted task that must include detailed provenance tracking of data processing (Keator et al., 2009). Comprehensive guidelines on how to navigate the database are a *sine qua non* for a repository to be effective. The careful development of a logical ontology for study variables is critical for the seamless integration, organization, and analysis of the study. This need for common data elements increases exponentially with the introduction of multi-institutional, federated or publicly shared repositories. Properly defining and documenting a common lexicon allows for precise querying and analysis to be performed on "Big Data" (Temal et al., 2008; Turner et al., 2010; Larson, 2013), as a small and nuanced change in semantics between sites and researchers can critically undermine the dissemination process. Thus, LORIS has integrated several tools such as a Data Dictionary builder to enable granular definition of data fields, and facilitate project-specific modifications for sharing purposes.

Terms such as "interoperability" and "data harmonization" have become popular buzzwords surrounding data-sharing across platforms. Despite this trend, many existing platforms do not fully address the range of dissemination issues, which critically impact the integrity of publicly shared datasets (Mendelson et al., 2014). For instance, the type of imaging format researchers use to share and process data can vary (e.g., DICOM, MINC, NIfTI, etc.), either requiring converters and essential libraries (that can produce lossy results) or employing other platforms such as ITK (www.itk.org) or LONI Babel (Van Horn et al., 2014). As sharing initiatives proliferate, the "order of complexity" of data organization increases disproportionately (Mennes et al., 2013). As a result, an API becomes essential for automated retrieval and dissemination of data, particularly when dealing with database federation. The importance of providing collaborators with easy access to all relevant information cannot be overstated (Keator et al., 2013).

When dealing with large multisite studies, reproducibility across scanners, sites, and time is imperative (Liu et al., 2010; Jovicich et al., 2009, 2013, 2014). To identify and track scanner variability and system errors over time and across sites, both living phantom and geometric phantom data are used to assess scanner consistency, and to develop compensatory transformations for distortion corrections. Geometric phantoms include the ADNI phantom (Gunter et al., 2009), the ACR phantom used by NIH MRI Study of Normal Brain Development (Fu et al., 2006), and the LEGO phantom used by the IBIS study (Caramanos et al., 2010; Fonov et al., 2010), which are used to assess temporal and geometric corrections in gradient field nonlinearity. Additionally, "Living phantom" acquisitions, performed on project staff who are regularly scanned at all sites, provide reproducibility information in a more realistic biological setting than can be obtained from a relatively simple geometric phantom (Jovicich et al., 2009, 2013).

Finally, the capacity to share information beyond a project's funding cycle is paramount for the long-term utility of data-sharing initiatives. A PubMed literature review conducted by Poline et al. (2012) estimated that of 12,000 datasets with more than 12 subjects, "only a small percentage" of data is available in public repositories, reflecting the notably

ephemeral state of most repositories. Concrete planning for long-term stewardship and sustainability is an important element in public data-sharing, one that has also begun to garner consideration from funding agencies (Berman & Cerf, 2013; Toga & Dinov, 2015). To this effect, LORIS and CBRAIN are both scalable initiatives, with stable long-term funding and an open-source community that is actively contributing to their codebases.

Conclusion

The MNI data-sharing ecosystem, based on the LORIS and CBRAIN platforms, represents a powerful combination of digital infrastructure for international data-sharing initiatives. These platforms leverage 20 years of established MNI tools (e.g., MINC toolkit, CIVET image processing pipeline) and numerous external software pipelines (e.g., SPM-batch, FSL, Freesurfer). Future work will involve an increased focus on federating and sharing of existing datasets, with particular emphasis in the areas of authentication, security, and pipeline reproducibility. We will also enhance our use of cloud technology and international platform interoperability for greater flexibility in storage, sharing and remote processing of multi-domain data.

References

Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M.E., et al., 2013. BigBrain: an ultrahigh-resolution 3D human brain model. Science 340, 1472–1475. http://dx.doi.org/10.1126/science.1235381.

Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J.P., Zijdenbos, A.P., Evans, A.C., 2012. The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. Front. Neuroinform. 6, 7. http://dx.doi.org/10.3389/fninf.2012.00007.

Berman, F., Cerf, V., 2013. Who will pay for public access to research data? Science 341, 616–617. http://dx.doi.org/10.1126/science.1241625.

Brakewood, B., Poldrack, R.A., 2013. The ethics of secondary data analysis: considering the application of Belmont principles to the sharing of neuroimaging data. Neuroimage 82, 671–676. http://dx.doi.org/10.1016/j.neuroimage.2013.02.040.

Caramanos, Z., Fonov, V.S., Francis, S.J., Narayanan, S., Pike, B., Collins, D.L., Arnold, D.L., 2010. Gradient distortions in MRI: characterizing and correcting for their effects on SIENA-generated measures of brain volume change. NeuroImage 49 (2), 1601. http://dx.doi.org/10.1016/j.neuroimage.2009.08.008.

Choudhury, S., Fishman, J.R., McGowan, M.L., Juengst, E.T., 2014. Big data, open science and the brain: lessons learned from genomics. Front. Hum. Neurosci. 8, 239. http://dx.doi.org/10.3389/fnhum.2014.00239.

Das, S., Zijdenbos, A.P., Harlap, J., Vins, D., Evans, A.C., 2011. LORIS: a web-based data management system for multi-center studies. Front. Neuroinform. 5, 37. http://dx.doi.org/10.3389/fninf.2011.00037.

Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., ... Milham, M.P., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry 19 (6), 659–667. http://dx.doi.org/10.1038/mp.2013.78.

Dinov, I.D., Petrosyan, P., Liu, Z., Eggert, P., Hobel, S., Vespa, P., ... Toga, A.W., 2014. High-throughput neuroimaging-genetics computational infrastructure. Front. Neuroinform. 8, 41. http://dx.doi.org/10.3389/fninf.2014.00041.

Evans, A.C., Brain Development Cooperative, G., 2006. The NIH MRI study of normal brain development. Neuroimage 30 (1), 184–202. http://dx.doi.org/10.1016/j.neuroimage.2005.09.068.

Fonov, V., Janke, A.L., Caramanos, Z., Arnold, D.L., Narayanan, S., ... Collins, D.L., 2010. Improved precision in the measurement of longitudinal global and regional volumetric changes via a novel MRI gradient distortion characterization and correction technique. Proceedings MIAR 324–333.

Frisoni, G.B., et al., 2011. Virtual imaging laboratories for marker discovery in neurodegenerative diseases. Nat. Rev. Neurol. 7, 429–438. http://dx.doi.org/10.1038/nrneurol.2011.99.

Fu, L., Fonov, V., Pike, B., Evans, A.C., Collins, D.L., 2006. Automated analysis of multi site MRI phantom data for the NIHPD project. Med Image Comput Comput Assist Interv 9 (Pt 2), 144–151. http://dx.doi.org/10.1007/11866763_18.

Glatard, T., Lartizien, C., Gibaud, B., da Silva, R.F., Forestier, G., ... Friboulet, D., 2013. A virtual imaging platform for multi-modality medical image simulation. IEEE Trans. Med. Imaging 32 (1), 110–118. http://dx.doi.org/10.1109/TMI.2012.2220154.

Glatard, T., Lewis, L.B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., ... Evans, A.C., 2015. Reproducibility of neuroimaging analyses across operating systems. Front. Neuroinform. 9, 12. http://dx.doi.org/10.3389/fninf.2015.00012.

Goldowitz, D., McArthur, D., 2011. The NeuroDevNet Vision. Semin. Pediatr. Neurol. 18 (1), 2–4. http://dx.doi.org/10.1016/j.spen.2011.02.007.

Gunter, J.L., Bernstein, M.A., Borowski, B.J., Ward, C.P., Britson, P.J., Felmlee, J.P., Jack, C.R., 2009. Measurement of MRI scanner performance with the ADNI phantom. Med. Phys. 36 (6), 2193–2205. http://dx.doi.org/10.1118/1.3116776.

Gutman, D.A., Dunn Jr., W.D., Cobb, J., Stoner, R.M., Kalpathy-Cramer, J., Erickson, B., 2014. Web based tools for visualizing imaging data and development of XNATView, a zero footprint image viewer. Front. Neuroinform. 8, 53. http://dx.doi.org/10.3389/fninf.2014.00053.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., ... Fischl, B., 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 46 (1), 177–192. http://dx.doi.org/10.1016/j.neuroimage.2009.02.010.

Jovicich, J., Marizzoni, M., Sala-Llonch, R., Bosch, B., Bartres-Faz, D., Arnold, J., PharmaCog Consortium, 2013. Brain morphometry reproducibility in multi-center 3 T MRI studies: a comparison of cross-sectional and longitudinal segmentations. Neuroimage 83, 472–484. http://dx.doi.org/10.1016/j.neuroimage.2013.05.007.

Jovicich, J., Marizzoni, M., Bosch, B., Bartres-Faz, D., Arnold, J., Benninghoff, J., PharmaCog Consortium, 2014. Multisite longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging of healthy elderly subjects. Neuroimage 101, 390–403. http://dx.doi.org/10.1016/j.neuroimage.2014.06.075.

Karama, S., Bastin, M.E., Murray, C., Royle, N.A., Penke, L., Munoz Maniega, S., Gow, A.J., Corley, J., Valdes Hernandez, M., Lewis, J.D., Rousseau, M.E., Lepage, C., Fonov, V., Collins, D.L., Booth, T., Rioux, P., Sherif, T., Adalat, R., Starr, J.M., Evans, A.C., Wardlaw, J.M., Deary, I.J., 2013. Childhood cognitive ability accounts for associations between cognitive ability and brain cortical thickness in old age. Mol. Psychiatry 19, 555–559. http://dx.doi.org/10.1038/mp.2013.64.

Keator, D.B., Wei, D., Gadde, S., Bockholt, J., Grethe, J.S., Marcus, D., ... Ozyurt, I.B., 2009. Derived data storage and exchange workflow for large-scale neuroimaging analyses on the BIRN grid. Front. Neuroinform. 3, 30. http://dx.doi.org/10.3389/neuro.11.030.2009.

Keator, D.B., Helmer, K., Steffener, J., Turner, J.A., Van Erp, T.G., Gadde, S., ... Nichols, B.N., 2013. Towards structured sharing of raw and derived neuroimaging data across existing resources. Neuroimage 82, 647–661. http://dx.doi.org/10.1016/j.neuroimage.2013.05.094.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., 2002. The human genome browser at UCSC. Genome Res. 12 (6), 996–1006.

Kim, J.H., Song, P., Lim, H., Lee, J.H., Lee, J.H., Park, S.A., Alzheimer's Disease Neuroimaging Initiative, 2014. Gene-based rare allele analysis identified a risk gene of Alzheimer's disease. PLoS One 9 (10), e107983. http://dx.doi.org/10.1371/journal.pone.0107983.

Larson, E.B., 2013. Building Trust in the Power of "Big Data" Research to Serve the Public Good. JAMA. 309 (23), 2443–2444. http://dx.doi.org/10.1001/jama.2013.5914.

Liu, X., Wang, Y., Liu, H., Liu, Z., Zhou, W., 2010. Diffusion tensor imaging and resting state functional magnetic resonance imaging on young patients with major depressive disorder. Zhong Nan Da Xue Xue Bao Yi Xue Ban 35 (1), 25–31. http://dx.doi.org/10.3969/j.issn.1672-7347.2010.01.004.

MacDonald, D., Kabani, N., Avis, D., Evans, A.C., 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. Neuroimage 12 (3), 340–356.

MacFarlane, D., et al., 2014. Enhanced Data Querying for Neuroinformatics Databases. Proceeding of Organization of Human Brain Mapping (OHBM). Hamburg, Germany.

Marcus, D.S., Olsen, T.R., Ramaratnam, M., Buckner, R.L., 2007. The extensible neuroimaging archive toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. Neuroinformatics 5, 11–33.

Matlock, M., Schimke, N., Kong, L., Macke, S., Hale, J., 2012. Systematic redaction for neuroimage data. Int J Comput Models Algorithms Med 3 (2). http://dx.doi.org/10.4018/jcmam.2012040104.

Mendelson, D.S., Erickson, B.J., Choy, G., 2014. Image sharing: evolving solutions in the age of interoperability. J. Am. Coll. Radiol. 11 (12 Pt B), 1260–1269. http://dx.doi.org/10.1016/j.jacr.2014.09.013.

Mennes, M., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2013. Making data sharing work: the FCP/INDI experience. Neuroimage 82, 683–691. http://dx.doi.org/10.1016/j.neuroimage.2012.10.064.

Niso, G., Rogers, C., Moreau, J.T., Chen, L.Y., Madjar, C., Das, S., ... Baillet, S., 2015. OMEGA: The Open MEG Archive. NeuroImage S1053–8119 (15), 00318-3. http://dx.doi.org/10.1016/j.neuroimage.2015.04.028.

Poline, J.B., Breeze, J.L., Ghosh, S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., ... Kennedy, D.N., 2012. Data sharing in neuroimaging research. Front. Neuroinform. 6, 9. http://dx.doi.org/10.3389/fninf.2012.00009.

Sherif, T., Rioux, P., Rousseau, M.E., Kassis, N., Beck, N., Adalat, R., ... Evans, A.C., 2014. CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. Front. Neuroinform. 8, 54. http://dx.doi.org/10.3389/fninf.2014.00054.

Sherif, T., Kassis, N., Rousseau, M.-É., Adalat, R., Evans, A.C., 2015. BrainBrowser: distributed, web-based neurological data visualization. Front. Neuroinform. 8, 89. http://dx.doi.org/10.3389/fninf.2014.00089.

Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 8. http://dx.doi.org/10.1155/2011/879716.

Temal, L., Dojat, M., Kassel, G., Gibaud, B., 2008. Towards an ontology for sharing medical images and regions of interest in neuroimaging. J. Biomed. Inform. 41 (5), 766–778. http://dx.doi.org/10.1016/j.jbi.2008.03.002.

Toga, A.W., Dinov, I.D., 2015. Sharing big biomedical data. J. Big Data 2 (7), 1–12. http://dx.doi.org/10.1186/s40537-015-0016-1.

Tremblay-Mercier, J., Madjar, C., Étienne, P., Poirier, J., Breitner, J.C.S., 2014. A program of Presymptomatic Evaluation of Experimental or Novel Treatments for Alzheimer's Disease (PREVENT-AD): Design, Methods and Perspective. Alzheimer Association International Conference 2014. July16th 2014. Copenhagen, Denmark.

Turner, J.A., Mejino, J.L., Brinkley, J.F., Detwiler, L.T., Lee, H.J., Martone, M.E., Rubin, D.L., 2010. Application of neuroanatomical ontologies for neuroimaging data annotation. Front. Neuroinform. 4. http://dx.doi.org/10.3389/fninf.2010.00010.

Van Horn, J.D., Toga, A.W., 2009. Is it Time to Re-Prioritize Neuroimaging Databases and Digital Repositories? Neuroimage 47 (4), 1720–1734. http://dx.doi.org/10.1016/j.neuroimage.2009.03.086.

Van Horn, J.D., et al., 2014. Neuroimaging and genetics in aging and age-related disease. Brain Imaging Behav. 8 (2), 141–142. http://dx.doi.org/10.1007/s11682-014-9299-7.

Vincent, R.D., Janke, A., Sled, J.G., Baghdadi, L., Neelin, P., Evans, A.C., 2004. MINC 2.0: a modality independent format for multidimensional medical images. In 10th Annual Meeting of the Organization for Human Brain Mapping. Budapest, Hungary.

Wolff, J.J., Gu, H., Gerig, G., Elison, J.T., Styner, M., Gouttard, S., ... Network, I.B.I.S., 2012. Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. Am. J. Psychiatry 169 (6), 589–600. http://dx.doi.org/10.1176/appi.ajp.2011.11091447.

Wood, D., King, M., Landis, D., Courtney, W., Wang, R., Kelly, R., ... Calhoun, V.D., 2014. Harnessing modern web application technology to create intuitive and efficient data visualization and sharing tools. Front. Neuroinform. 8, 71. http://dx.doi.org/10.3389/fninf.2014.00071.

Xuan, W., Dai, M., Buckner, J., Mirel, B., Song, J., Athey, B., Meng, F., 2010. Cross-domain neurobiology data integration and exploration. BMC Genomics 11 (Suppl. 3), S6. http://dx.doi.org/10.1186/1471-2164-11-S3-S6.