MCGILL CENTRE FOR INTEGRATIVE NEUROSCIENCE

LORIS GOVERNANCE FRAMEWORK

April 2019

neuro
Institut et hôpital neurologiques de Montréa
Montreal Neurological Institute and Hospita

MCIN MCGILL CENTRE for INTEGRATIVE NEUROSCIENCE

McGill

TABLE OF CONTENTS

What is LORIS?

LORIS is an open-source web-based research data management platform. It is freely available for download on GitHub (http://github.com/aces/loris) and can be installed for use under GPL version 3 license. Technical system specifications are detailed in the GitHub repository Readme. A demonstration LORIS website has been created to try at: https://demo.loris.ca

LORIS is designed to store, cross-link, curate, visualize and share many types of research data over the web, including neuroimaging, behavioural, clinical, electrophysiological, biospecimen and genetic data. LORIS is designed to make it easy for research teams to manage large multi-modal datasets acquired over time in a longitudinal study, or at different locations in a large multi-site study. Features of its secure, modular and extensible tools for automating the flow of research data are described at: LORIS.ca. As part of expanding Open Science initiatives, such as the Montreal Neurological Institute's Open Science policy (Poupon 2017), the LORIS design and infrastructure supports open data sharing with adherence to the FAIR principles (Wilkinson 2016).

Researchers may install LORIS on any web-accessible Unix server in order to run their own self-supported data platform. As a service, LORIS may be installed, developed and actively maintained by our staff serving as a Data Coordinating Centre based at the Montreal Neurological Institute (MNI). Paired with the CBRAIN data processing portal (Das 2018), LORIS can automate multi-modal data analysis and storage at scale, allowing investigators to focus on research instead of data handling.

The LORIS Group

LORIS has been developed since 1999 at the Montreal Neurological Institute, by Alan Evans' research group at the McGill Centre for Integrative Neuroscience (MCIN.ca). It is actively developed and maintained by a core team of developers based at MCIN and led by Samir Das, Director of Technology.

A supporting group of over 20 developers, based both within MCIN and at affiliated institutions, actively contributes to its shared public codebase on a weekly basis, with most committing at least 25% of their weekly time to core LORIS development and activities. Project Officers, Data Managers and affiliated clinical, research and technical staff support each research data platform.

LORIS' global contributor and user community spans 4 continents and over 400+ research sites. As projects around the world customize their LORIS platforms, their technical contributions to the open-source codebase are welcomed.

Working with LORIS

For researchers who wish to install LORIS on their own, the LORIS group at MCIN may serve in a consultative capacity for self-supported projects. MCIN also provides LORIS as a data management service, serving as the LORIS Data Coordinating Centre (DCC) for several major initiatives.

Two common models for data sharing via LORIS are: Collaborative data access within a research project or consortium, and Open Science data sharing.

1. LORIS as a supported service

The LORIS group based at MCIN and the MNI can also serve as the Data Coordinating Centre (DCC) for a LORIS instance, providing services to install, customize, maintain, upgrade, and develop features for the lifecycle of the study or consortium. Several major installations of LORIS have been actively supporting research consortia for over a decade, such as the IBIS research network.

Collaborative data access model:

- IBIS : Infant Brain Imaging Study : Long-term imaging study of high-risk and low-risk infant siblings, to detect early biomarkers of Autism Spectrum Disorder in neurodevelopment (http://www.ibis-network.org/)
- NIHPD : NIH database of Normal Pediatric Development
- Brain Canada Autism Platform : Development of integrative tools and technologies to harmonize and cross-link data from diverse types and platforms across national cohorts
- NeuroDevNet imaging subprojects

Open Science model:

- ABIDE : Autism Brain Imaging Data Exchange, hosted in LORIS since 2014
- BigBrain : open-access histological dataset and visualization of 20-micron brain slices, in collaboration with the Juelich Institute, Germany
- CCNA : Canadian Consortium for Neurodegeneration in Aging : National multi-cohort network studying long-term aging and the brain (Mohaddes 2018; http://ccna-ccnv.ca/)
- Prevent-AD : Prevent Alzheimer's Disease : longitudinal study of multiple treatment cohorts
- Open MNI iEEG Atlas : Open-access Atlas of healthy intracranial electrophysiological activity (https://mni-open-ieegatlas.research.mcgill.ca)
- C-BIGR : Biobank at the Montreal Neurological Institute
- QPN : Quebec Parkinson's Network

2. Self-supported projects

Many LORIS projects are run by local technical personnel, with community support or consultation provided by the LORIS group. Code contributions from external projects are welcomed.

Collaborative data access model:

- PERFORM centre, Concordia University: institutional LORIS data platform housing multiple studies, labs, and data modalities
- Canadian Neonatal Brain Platform: Based at Ste-Justine children's hospital, CNBP federates data across multiple platforms
- CALSNIC: Canadian ALS Neuroimaging consortium: Canada-wide ALS imaging study
- MAVAN: Maternal Adversity study based at the Douglas Hospital University Institute and the Ludmer Centre for Mental Health and Neuroinformatics
- University of Edinburgh, Centre for Clinical Brain Sciences: Stroke database collection of 3000 scans with clinical data
- Baby Connectome Project: infant imaging subproject of the Human Connectome Project
- Douglas Mental Health University Institute: Brain Imaging Centre
- 1000 Brains (1000-Gehirne) project, Germany
- Leukodystrophy: MNI-based harmonized data collection for rare progressive genetic disease
- Cuban Neuroscience Centre: national cohorts of imaging, clinical/behavioural, and EEG data
- KISTI, Korea: neurodegeneration research group
- Child-Mind Institute, New York: Milham research group
- GUSTO, Singapore: population data from national birth cohorts

Data Governance & Ethics

This section includes best practices followed by the LORIS group in its development and when serving as DCC, and recommended for all implementations, as well as features of LORIS as an open-source software.

1. Data Quality

As a data platform, LORIS can contribute to projects adopting the FAIR data principles to make data Findable, Accessible, Interoperable and Reusable (Wilkinson 2016).

Our aim is to store and process data in a way that is accurate, verifiable, unbiased, proportionate, and current, so as to enhance their interoperability and replicability and preserve their long-term searchability and integrity.

LORIS provides workflows for data collection, quality review, curation and data sharing across different modalities, including neuroimaging, clinical/behavioural, genetic and biobanking data. Each workflow is designed to promote consistency and accuracy in research data management. The transparency of these procedures provides critical confidence in the quality and accuracy of data resources, and facilitates verification of the reliability and reproducibility of research results.

The LORIS group gathers community feedback on the utility, quality, security, and accuracy of its systems with a view to improving quality and interoperability and appropriate re-use of research resources by others.

2. Data Protection & Access

2.1 Privacy

The LORIS group strives to respect privacy in its sharing of neuroscience resources and LORIS was designed to facilitate compliance with applicable privacy, data protection and security regulations.

When serving as DCC, privacy and data protection safeguards (see sections below) will be proportionate to the nature and use of the data, taking into consideration both the risk of re-identification of data and its potential for misuse.

LORIS will publicly disclose lawful requests for data granted based on law enforcement, public health, or national security concerns.

2.2 De-identification

The LORIS team has developed and adopts a suite of de-identification tools and procedures to protect the privacy of individual research participants whose data are shared through its systems. These include the “defacing” of imaging scans, removal of potentially identifying dates (e.g., hospital visits, dates of birth), and further anonymization of demographic information (e.g., profession data).

Research participant identities are replaced by LORIS participant codes that enable participant re-contact by the partner providing the resource (e.g., to participate in further research or for the return of results), but LORIS does not hold information about individual identities beyond that required to generate the Open Science codes (described below) and the participant codes themselves.

For Open Science resources, a hash encryption tool is implemented to generate Open Science participant codes based on personal information that is not stored in LORIS. Through this one-way de-identification, participants can be matched to their datasets from other Open Science studies that are using this system if their personal details are known, without enabling any access to those personal details (Das 2017). Particular Open Science studies can opt out of this matching system if required.

Researchers accessing protected Open Science resources (i.e., resources that are not open access; see section 2.5) will agree to forego any attempt to re-identify data and samples unless they have written permission from the sponsoring institution of the partner providing the shared resource.

2.3 Confidentiality

Researchers accessing protected Open Science resources (i.e., resources that are not open access; see section 2.5) will agree to protect confidential data against unauthorized access.

This will not preclude the publication of research results from approved uses of the data following standard academic practices.

2.4 Security

The LORIS group has established proportionate data security measures that mitigate the risk of unauthorized access, data loss, and misuse.

Members of the LORIS group are dedicated to maintaining consistent practices in testing and security review of the codebase, including actively researching and documenting potential security concerns.

User access permissions embedded in LORIS allow granular control of access by administrators on the level of sites, modules, workflows and data modalities.

IT planning and practices prioritize backup and archiving of data and protection of data assets against security risks. These measures include regular audits of data backups in addition to automated server monitoring using nagios, verifying user credentials for each LORIS and protecting against attacks using fail2ban, restricting port availability and applying SSL encryption.

2.5 Access

This section describes how data can be shared using LORIS depending on the Open Science or Collaborative model governing sharing for each resource. Some data resources will be available without restrictions while others will be protected to maintain participant privacy and consent.

LORIS enables the sharing of data through three distinct data access policy models: Open Access, Registered Access and Controlled Access (see sections below). Both Registered Access and

Controlled Access are protected access levels, which limit access to users such as approved researchers and also limit how the data can be used.

To determine the level of access for a particular dataset, the type of data to be shared and the potential risk of its re-identification (how detailed it is, and its potential for misuse) should be considered.

2.5.1 Open Access

Open Access data will be publicly and freely available without any conditions of use.

However, a statement requesting appropriate acknowledgment may be attached to Open Access data.

2.5.2 Registered Access

LORIS is currently developing a new data access level for Registered Access data, which will be available to all bona fide researchers and clinical care professionals who have registered and agreed to basic good data use practices (e.g., do not attempt to re-identify participants, meet ethics requirements, keep the data secure) (Dyke 2018).

2.5.3 Controlled Access

Controlled Access data will be available on a case-by-case basis upon request to the original data provider.

The Controlled Access level will include data and samples that are only available to members of a research consortium under the Collaborative model.

3. Consent

Data (and biospecimen sample) sharing plans should be clearly explained in consent information and discussions with individuals from whom data may be shared.

In particular, for Open Science projects, individuals should be made aware of the possibility of international sharing and eventual commercial use of data and biospecimens.

3.1 Consent Codes

As data and other resources are shared and become widely available for secondary uses in Open Science, the original consent of research participants and patients must be respected.

Data resources available through LORIS will have consent-based permissions and restrictions on their use indicated by Consent Codes (Dyke 2016). Consent Codes are a structured way of recording and communicating conditions such as permitted area of research use (e.g., health/medical/biomedical). The use of Consent Codes will facilitate the discovery of suitable resources for a given project as well as their ethical re-use.

The Consent Codes will be extended as required with any further data use categories identified with the neuroscience research community.

3.2 Re-contact of participants

Re-contact of research participants and patients may be required to seek consent to further research or to inform them of medical information that could be useful to them.

The LORIS participant coding system will enable such re-contact (see section 2.2), but LORIS will not hold information about individual identities beyond that involved in generating the Open Science participant codes and the participant codes themselves.

Partners remain responsible for re-contact of individual research participants according to participant consent and local policies and procedures.

3.3 Withdrawal

Research participation is voluntary. LORIS allows for the removal of an individual's data from distribution tools if consent is withdrawn.

As DCC, the LORIS group will facilitate withdrawal when notified by partners that consent has been withdrawn.

However, for LORIS Open Science resources, data (and biospecimen samples) that have already been used or published by researchers around the world cannot be withdrawn.

4. Publications

4.1 Authorship

LORIS Open Science resources shared through Open, Registered or Controlled Access (see section 2.5) are available for analysis and use without collaboration (and co-authorship) with the original team providing the resource.

4.2 Acknowledging LORIS and Open Science resources

The use of shared LORIS Open Science resources should be acknowledged in any related publications and presentations.

The appropriate form of acknowledgment may be specified on a case-by-case basis, but should comprise, at a minimum:

- acknowledgment of the team and/or institution sharing the resource;
- citation of relevant research publications; and
- a clear description of the resource used and where it can be accessed - including the doi or URL of the LORIS platform (to be included in the methods section of research publications, where possible).

Researchers should also acknowledge the use of LORIS software and support by citation of Das 2016.

5. Research oversight

Projects using LORIS resources are responsible for obtaining all ethics approvals required by their institutions.

6. Return of individual results (including incidental findings)

LORIS allows for re-contact of participants by partners through its participant coding system (see section 2.2).

LORIS Infrastructure, Specifications & Technical Procedures

LORIS recommends international best-practice standards in the implementation and maintenance of research data platforms. Its architecture has been specifically designed to overcome the challenges of large research consortia and data federation platforms, enabling investigators to seamlessly validate, process, query and disseminate diverse datasets at scale.

LORIS is fully customizable for any data type and any activity workflow. It is mobile-friendly for tablet data entry, and includes features such as surveys for at-home secure data entry by participants. A typical LORIS setup will also include a NoSQL Data Querying Tool to enable quick and granular cross-modal longitudinal querying and dissemination of data. LORIS actively supports integration with the BIDS neuroimaging standard, and provides projects with granular user access control features.

As a storage platform, LORIS is designed to interface with any data platform or processing system via its RESTful API. Together with CBRAIN, MCIN's open-source processing portal mountable on HPC clusters, the combined platforms serve as the MNI ecosystem (Das 2016) for large-scale data storage and analysis.

1. Roles and resourcing

A LORIS project requires support by one or more individuals. Skill sets typically include:

- A technical expert, administrator/developer: maintain the UNIX server, data storage, network connectivity, codebase installation and dependencies, and customize technical code.
- A Data Manager or Project Officer: track, curate and manage the data, and document procedures and define workflows for user groups (e.g. clinical staff, imaging technicians) and design feature roadmaps in coordination with investigators involved in the research study.

Additionally, Project Officers may coordinate ethics, data access and data use agreements, with a view to governing access, storage and dissemination controls.

For projects using LORIS as a supported service, members of the LORIS group serve in these capacities as DCC.

2. IT practices, Support and Community engagement

A global community listserv (loris-dev@bic.mni.mcgill.ca) provides an open forum for support discussions. Bugs and software issues can be reported to the GitHub Issues public page (github.com/aces/loris/issues). Online documentation (via GitHub) details installation and troubleshooting procedures, and includes guides to system components. Each module is documented with web-accessible specifications.

For projects where the LORIS group serves as DCC, the IT infrastructure, support and maintenance is overseen by a highly skilled team of system administrators within MCIN who provide an extensive range of services in maintaining and scaling each project. Virtual machine environments are used to separate development, staging (testing) and production activities, ensuring a minimum of downtime and optimizing service stability. Security is actively maintained and proportionately optimized throughout the lifecycle of the project (see above).

3. Development procedures

The LORIS group centrally manages code upkeep, improvements, and documentation. Bugs reported on GitHub are addressed and resolved by developers in the LORIS group. The LORIS feature roadmap is available publicly on GitHub and is updated by senior staff several times per year, primarily with release updates. Special-interest working groups across the local LORIS development community drive forward domain-specialized features or dimensions of the software, such as working groups dedicated to Imaging features or User Experience.

Almost all LORIS development procedures and practices are documented through its GitHub repository (and wiki), in order to promote accessibility to contributors from the global open-source development community. These practices enable remote administrators to access and customize every aspect of their LORIS.

LORIS development practices and guidelines evolve through group discussion including at open meetings held weekly at the Montreal Neurological Institute. Visitors are welcome to present issues or feature requirements arising in their research context, and consult on troubleshooting questions with the technical team. Open cross-collaboration is also fostered through outreach activities including through engagement with the International Neuroinformatics Coordinating Facility (INCF.org).

3.1 Releases

The LORIS codebase is released open-source on GitHub (https://github.com/aces/LORIS), averaging two major releases per year, with numerous minor releases interspersed. The main repository is

accompanied by neuroimaging insertion and preprocessing pipeline code released in matching increments (https://github.com/aces/LORIS-MRI) which are closely coordinated with each major/minor LORIS release.

Release notes provide public information on updates and improvements for administrators. Upgrade instructions are provided for existing projects, including database patches compiled for each release, to ensure backwards compatibility, minimal downtime, and optimal data integrity.

Quick Links

- LORIS.ca
- Try LORIS at https://demo.loris.ca
- MCIN.ca

References

Academic publications describing LORIS infrastructure, key project deployments, and ethical and governance frameworks include:

Das S., Lecours-Boucher X., Rogers, C., et al., 2018. Integration of “omics” Data and Phenotypic Data Within a Unified Extensible Multimodal Framework. Frontiers in Neuroinformatics 12:91. doi:10.3389/fninf.2018.00091

Das S., Glatard T., Rogers C., Saigle J., Paiva S., MacIntyre L., et al. (2017). Cyberinfrastructure for open science at the montreal neurological institute. Front. Neuroinform 10:53. doi:10.3389/fninf.2016.00053

Das S., Glatard T., MacIntyre L. C., Madjar C., Rogers C., Rousseau M. E., et al. (2016). The MNI data-sharing and processing ecosystem. Neuroimage 124, 1188–1195. doi:10.1016/j.neuroimage.2015.08.076

Dyke SOM, Linden M, Lappalainen I, Rambla De Argila J, Carey K, et al. (2018). Authorizing Data Access. European Journal of Human Genetics. doi:10.1038/s41431-018-0219-y

Dyke SOM, Philippakis AA, Rambla De Argila J, Paltoo DN, Luetkemeier ES, et al. (2016). Consent Codes: Upholding Standard Data Use Conditions. PLoS Genetics 12(1): e1005772. https://doi.org/10.1371/journal.pgen.1005772

Mohaddes Z., Das S., Abou-Haidar A., Safi-Harab M., Blader D., Callegaro J., et al. . (2018). National neuroinformatics framework for canadian consortium on neurodegeneration in aging (CCNA). Front. Neuroinf. doi:10.3389/fninf.2018.00085

Poupon, V., Seyller, A., Rouleau, Rouleau, G.A., (2017). The Tanenbaum Open Science Institute: Leading a Paradigm Shift at the Montreal Neurological Institute. Neuron. 95:5, 1002-1006. https://doi.org/10.1016/j.neuron.2017.07.026

Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3:16 doi:10.1038/sdata.2016.18