

Enabling Biomedical Collaboration with the Biomedical Informatics Research Network (BIRN)

BIRN is supported by NIH grants 1U24-RR025736, U24-RR021992, U24-RR021760 and the Collaborative Tools Support Network Award 1U24-RR026057-01.



NIH National Institute of
General Medical Sciences
Basic Discoveries for Better Health

What is BIRN?

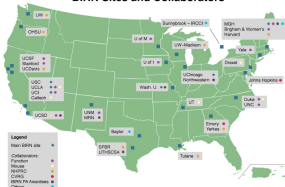
BIRN is a national initiative to advance biomedical research through data sharing and online collaboration. Funded by NIGMS, BIRN provides data-sharing infrastructure, software tools, strategies and advisory services – all from a single source.

BIRN focuses on the biomedical research community's intensive data sharing and analysis needs evident in fields such as biomedical imaging and genetics.

BIRN provides a software-based framework based on the needs of research teams for sharing significant quantities of data – rapidly, securely and privately – across incompatible computing systems. Groups may choose whether to share data internally or with external audiences. In either scenario, hardware and data remain under the control of individual user groups.

We also offer data-sharing software tools specific to biomedical research, best practices references, expert advice, and other resources. BIRN actively is developing new data-sharing tools, authorization capability, and knowledge engineering tools to help researchers make sense of new information in unique ways.

BIRN Sites and Collaborators



BIRN is a nationwide consortium that:

- Consists of technology development sites and specialized domain-knowledge sites.
- Facilitates data sharing in the biomedical community.

<http://www.birncommunity.org>

BIRN Fast Facts:

- Begins with user needs in a 'bottom-up' approach.
- Does not require extensive retooling or strict standards compliance by users.
- Employs leading-edge, grid software-based approach.
- Involves users as domain experts and technology development partners.

Data Movement and Sharing using BIRN Tools and Services

- Facilitate collaborative research through sharing of research data
- Find data by querying across heterogeneous data sources
- Manage users and access control of data
- Use BIRN capabilities to enhance your own infrastructure

1) Easy-to-use "fire and forget" data movement using Globus

• Globus Online

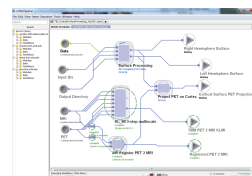
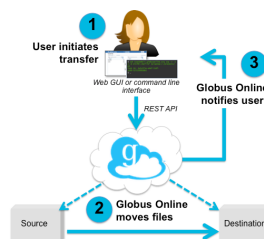
- Hosted service that automates the tasks associated with moving files between "endpoints."
- Easy "fire-and-forget" transfers
- Cloud-based server (e.g., AWS) can be used as endpoint
- Web, command line, and REST interfaces
- User/Group management with BIRN Access

• Globus Connect

- Transfer files to and from your local machine
- Installed and run as a user w/o admin privileges
- Outbound connections only so works behind a firewall

• Globus Connect Multiuser

- Create transfer endpoints in minutes
- Enable multi-user access for a resource
- No GridFTP install necessary



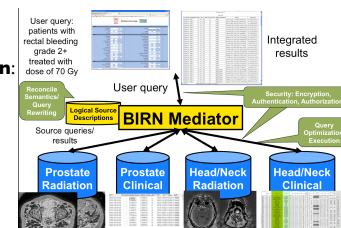
3) High-performance workflows - Galaxy and the LONI Pipeline

- **Galaxy** is an extensible web-based platform; almost any software tool can easily be integrated
- Our **Galaxy** system includes integrated mechanisms to transfer big datasets from geographically distributed centers into our cloud computing infrastructure using Globus Online
- Our managed platform includes flexible on-demand storage management using Amazon EBS and Amazon S3. We have also developed a novel on-demand computational infrastructure using Amazon EC2 and the Condor framework that is capable of elastically scaling the compute cluster
- **LONI Pipeline** is a cross-domain workflow application that allows users to quickly and easily create workflows that utilize leading tools in biomedical fields such as neuroimaging, genomics, and bioinformatics. Genomic data can be accommodated using our SNPLim database.
- The **Distributed Pipeline Server** allows access to local hardware resources, local software tools, as well as the LONI Pipeline library of tools (including FSL, Freesurfer, AFNI, LONI and other tools)

2) Mediation between data sources and source modeling

Advantages of Data Mediation:

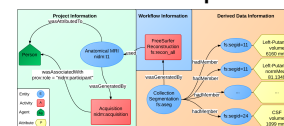
- Source of data controls curation
- Existing schemas allowed
- Use of a domain model rather than full ontology
- Mediator can exist on non-local server



BIRN Mediator Features:

- Works with relational databases, XML databases, Web Services, HTML, ...
- Complex structured queries (SQL)
- Queries rewritten and optimized for each source
- Secure system with access control and encryption (Grid Security Infrastructure)
- Scalable Integration – query rewriting time over 10,000 sources in under 1 sec

4) Derived data capabilities



Tagfiler is a collaborative data catalog rendered as an easy-to-use web service with rich query support. Tagfiler allows data to be annotated with extensible, domain-specific metadata.

Provenance Libraries: a C prototype library that enables easier construction of PROV-DM compatible XML formatted provenance representations for neuroimaging applications.

NI-DM is an extensible model for neuroimaging derived data with support for rich, query-able provenance and metadata.

Adding to existing lexicons:

Currently BIRN is working with Neurolex to add terms relating to the DICOM standard and neuroimaging processing and analysis tools.