Enabling Collaboration Using the Biomedical Informatics Research Network (BIRN)

Bruce Rosen M.D., Ph.D.
Data Lifecycle and BIRN

1. Create metadata (tag the data)
2. (query metadata)
3. Find data of interest
4. (move data)
5. (workflow)
6. Process and analyze data
7. retrieve data
The Challenge: Create Metadata for Data Identification

Metadata creation

Tagfiler
Collaborative data catalog web service

Formal Data Model

Data Provenance

Domain-specific terms in public lexicons
Data integration architecture approaches

**Data Warehouse(s)**

- Extract Transform Load
- Relational DB Ex: HID
  - Schema1
- Web Service Ex: XNAT
  - Schema2
- XML DB Ex: eXist-db
  - Schema3

Transform data into common schema and load into one or more warehouses

**Mediation between Data Sources**

- Mediator
  - Schema Translation
  - Distributed Query Evaluation
  - Live Queries

- User / Application
- Common Schema, Data Model
- Different Schemas, Data Models

- Relational DB Ex: HID
  - Schema1
- Web Service Ex: XNAT
  - Schema2
- XML DB Ex: eXist-db
  - Schema3

Mediator translates queries for each individual source, while data and local schemas stay the same
BIRN Mediator provides deep integration across complex data sources, with access control and encryption while supporting complex queries.

User query: patients with rectal bleeding grade 2+ treated with dose of 70 Gy

Integrated results

Security: Encryption, Authentication, Authorization,

Query Optimization/Execution

Reconcile Semantics/Query Rewriting

Source queries/results

Prostate Radiation

Prostate Clinical

Head/Neck Radiation

Head/Neck Clinical
The Challenge:
Moving and Sharing Data Easily – Globus Online

• Globus Online
  – Hosted service with “fire-and-forget” transfers
  – Cloud service can be used as endpoint
  – Web, command line, and representational state transfer (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, works behind a firewall

• Globus Connect Multiuser
  – Create transfer endpoints in minutes
  – Enable multi-user access for a resource
  – GridFTP + MyProxy server + Certificate Authority
The Challenge: Customization of Processing and Analysis

- Galaxy and LONI-based Pipeline Tools
• Executive Committee
  – Carl Kesselman, PI
  – Bruce Rosen, co-PI
  – Karl Helmer, CTSN
  – Arthur Toga, co-PI
  – Ian Foster, co-PI
  – Steven Potkin, co-PI
  – Amy Swain, NIGMS
  – Peter Lyster, NIGMS
• Data Management
  – Ann Chervenak, Rob Schuler
• Information Integration
  – Jose Luis Ambite
• Security
  – Rachana Ananthakrishnan
  – Laura Pearlman
• Operations
  – Lee Liming, JP Navarro

• Knowledge Engineering
  – Gully Burns, Jessica Turner
• Communications
  – Cristina Williams
• Workflows
  – Ivo Dinov, Ravi Madduri
• Genomics
  – Fabio Macciardi
• Executive & Steering Committees
  – Joe Ames, Karl Helmer

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