Enabling Collaboration Using the Biomedical Informatics Research Network (BIRN):

Karl Helmer Ph.D.

Athinoula A. Martinos Center for Biomedical Imaging,

Massachusetts General Hospital

June 4, 2010

BIRN Research Networks

- The main goal of the BIRN is to enable collaborative bioscience/biomedicine
- Use Cases:
  - share data/tools
  - provide security
  - develop, add to, merge ontologies
  - query across disparate data sources
  - search for outside resources

BIRN Biomedical Informatics Research Network

BIRN Background

Funded by National Center for Research Resources in 2001.

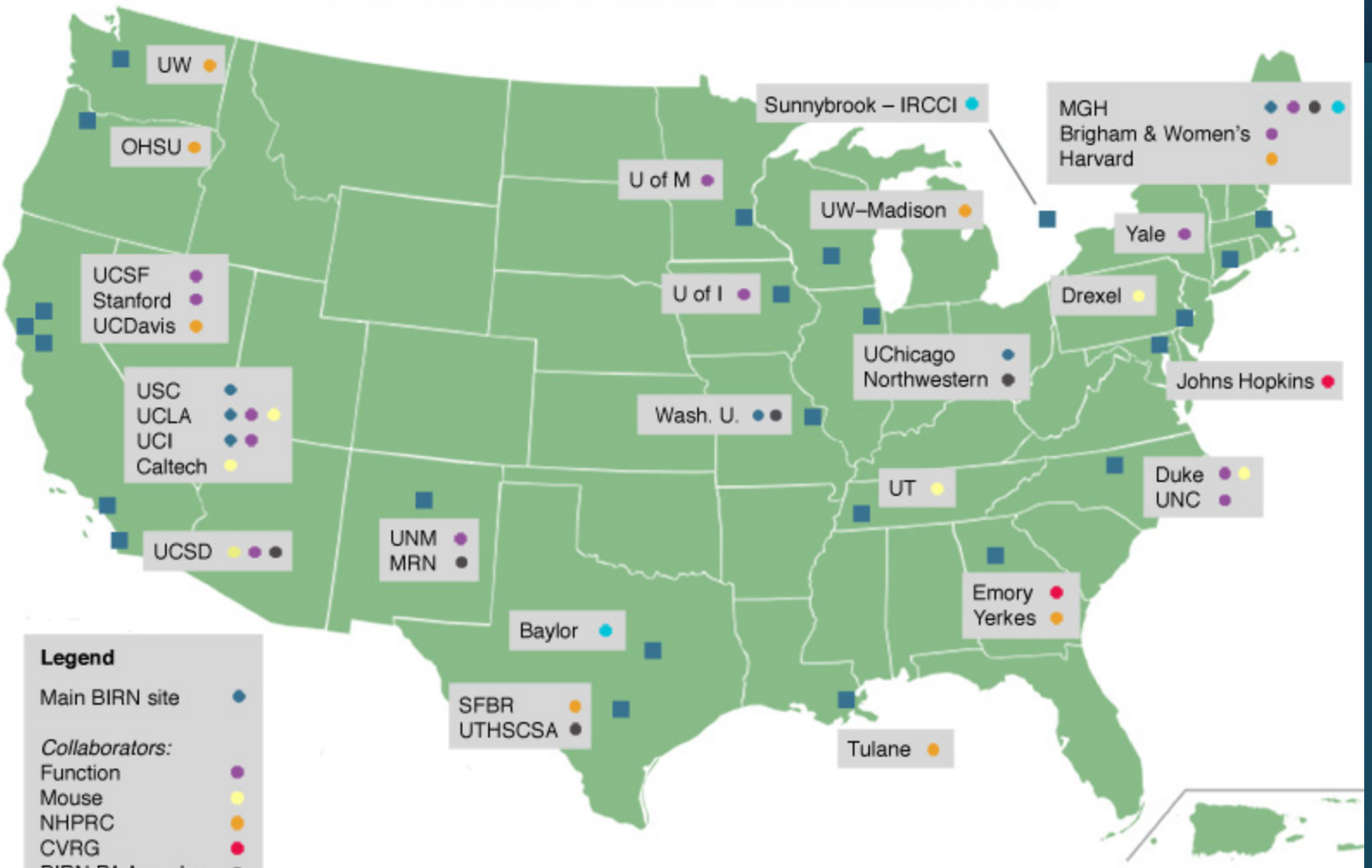
Three neuroimaging testbeds initially provided use cases for developed technology.

Reorganized in 2008 to provide software-based solutions for data sharing in the biosciences. No longer focused on neuroimaging (or even imaging).

One can no longer 'log into BIRN'. New model of providing modular capabilities that users can fit into existing toolset.

BIRN Biomedical Informatics Research Network

BIRN Sites and Collaborators



matics
Research Network

Technical Approach

- Bottom up, not top down
- Focus on user requirements - what they want to do
- Create solutions - factor out common requirements
  - Capability model includes software and process
  - Avoid “Big Design up Front” (BDUF)

Capability Model

- Software and services are only useful in bioinformatics if they do things that scientists need.
  - What is the problem?
  - How do the tools address the problem?
- BIRN's capabilities are defined in terms of problems and solutions, not in terms of software and services.
  - This is true from beginning (definition and documentation) to end (quality assurance and assessment).

Dissemination Models

- Different problems require different kinds of solutions.
  - BIRN *operates services* for all users; e.g., user registration service
  - BIRN *provides kits* for project teams to deploy services for their members; e.g., data sharing
  - BIRN *provides downloadable tools* for individuals to use on their own; e.g., image manipulation tools
- Understanding deployment needs is part of defining the problem to be solved.

Working With BIRN

- New capabilities are defined, designed, and disseminated by BIRN *Working Groups*

Data Management

Information Integration

Knowledge Engineering

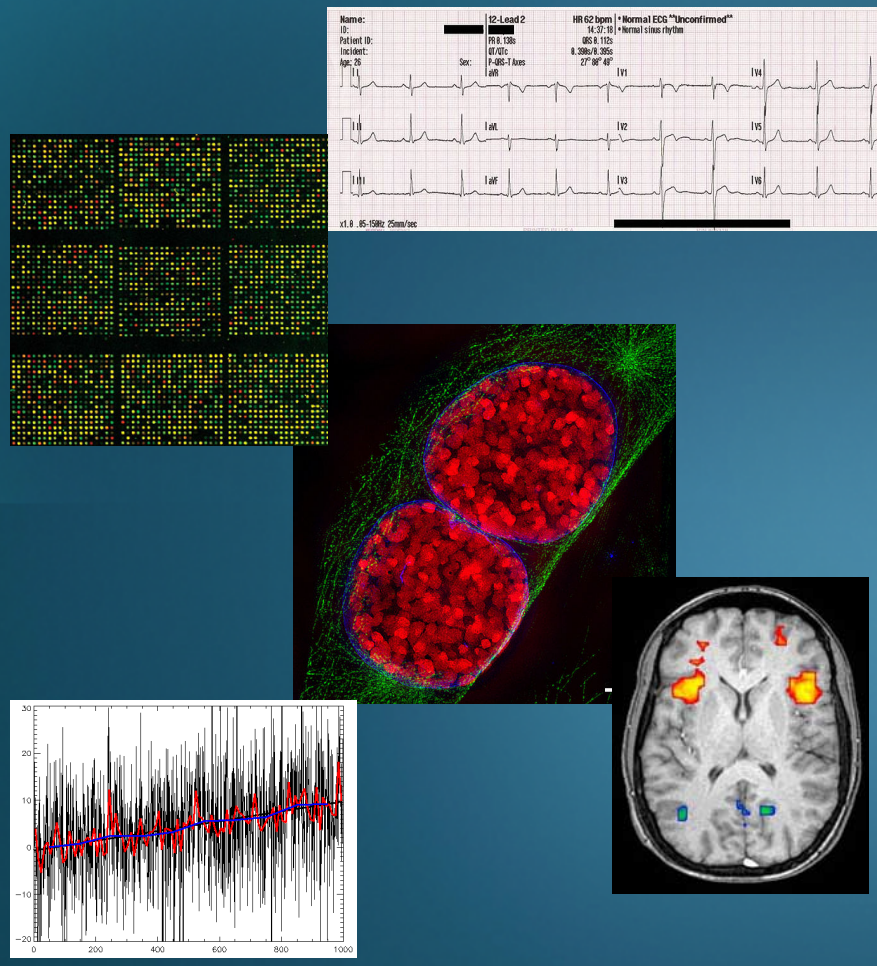
Workflow Tools

Security

Genomics

BIRN Biomedical Informatics Research Network

Data Management



- BIRN seeks to support data intensive activities including:
  - Imaging, Microscopy, Genomics, Time Series, Analytics and more…
- BIRN utilities scale:
  - TB data sets
  - 100 MB – 2 GB Files
  - Millions of Files

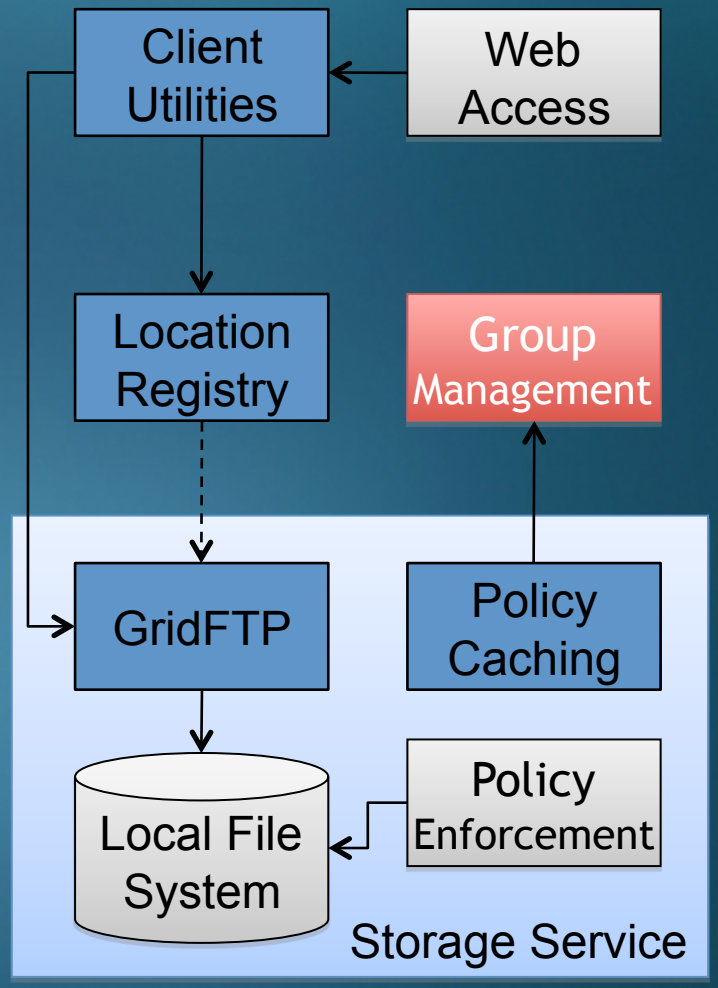
BIRN Biomedical Informatics Research Network

Data Management

- *Security*
  - Protect sensitive data: PHI, Researcher identity, IRB restrictions
- *Scalability*
  - Support growth in projected data usage: data volumes, file sizes, users, sites, ops/sec
- *Infrastructure*
  - Work with commodity networking, storage, etc.; and strict firewalls, etc.

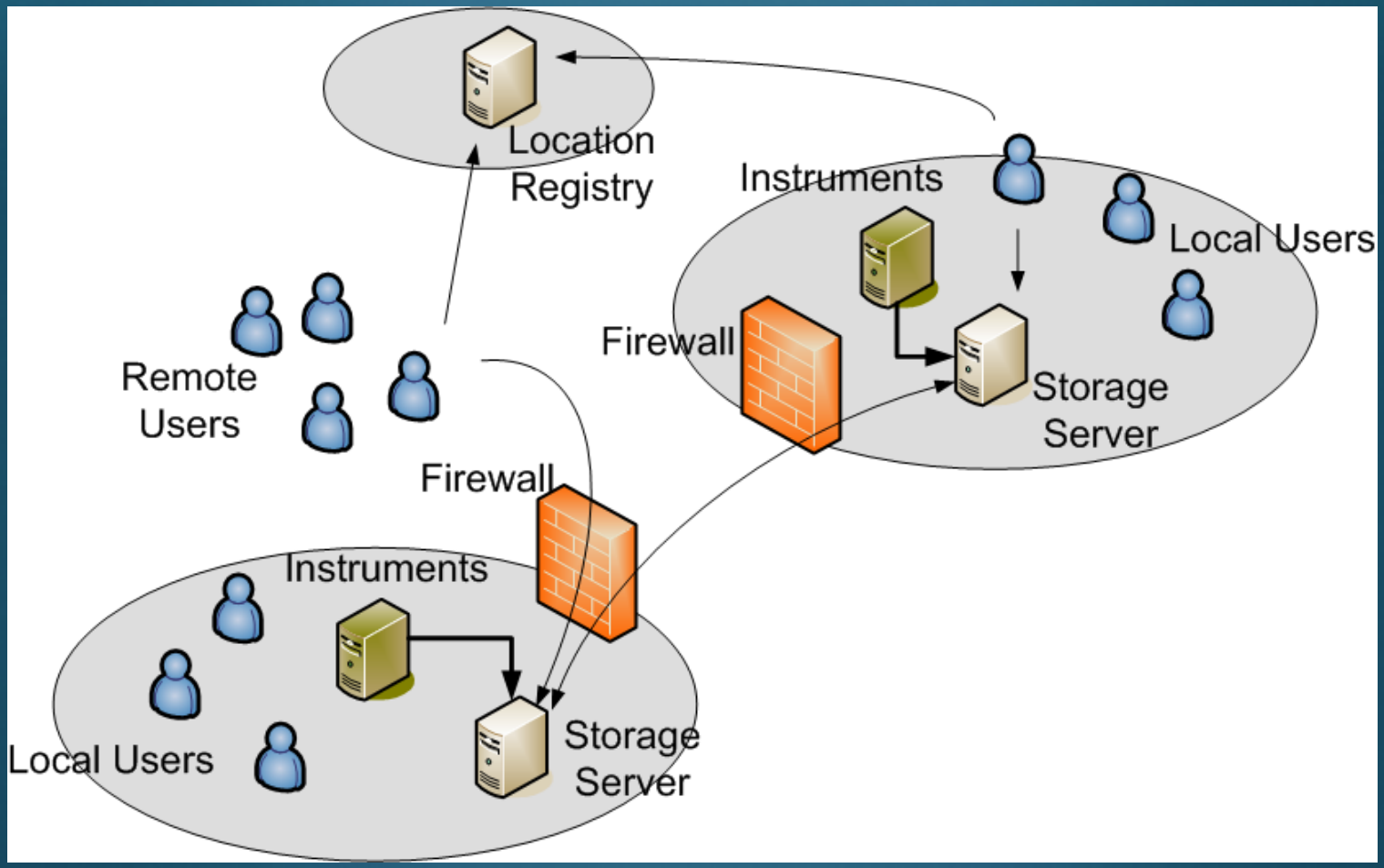
Data Service Architecture

- Client Utilities
  - UNIX utilities
  - Java & C APIs
- Shared Services
  - Manage File Locations
  - Manage User Groups
- Local Services
  - Integrate with conventional file systems



BIRN Biomedical Informatics Research Network

Data Service Deployment



Data Management Summary

- High Performance Transfer Protocol
  - Strong security
  - High performance, proven technology
- Location Registry
  - Distributed registry to track file locations
  - Low overhead, supports 100s ops/s
- Group Management
  - Centralized user and group management
  - Caching at local servers
- Policy Enforcement
  - Flexible access control policies
  - User and group ownership rights

BIRN Biomedical Informatics Research Network

Information Integration

Mediator: uniform structured query access to heterogeneous sources

Challenges:

Syntactic (Access/Format) heterogeneity: → *Wrappers*

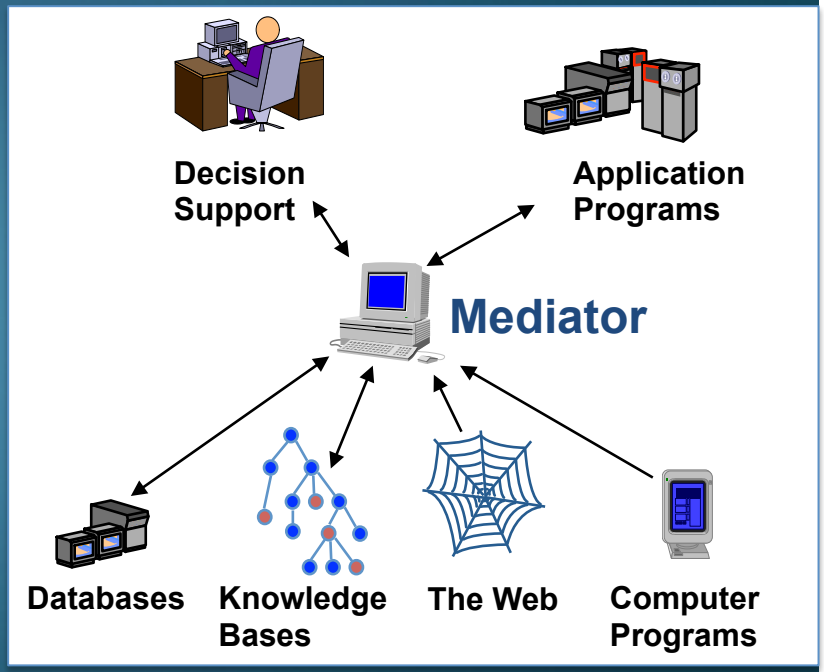
- Structured Sources: DBMS, XML/XQuery DBs
- Semi-structured Sources: HTML, text, pdf
- Web services → XML, SOAP, WSDL

Semantic heterogeneity → ***Mediator***

- Schema → *Source modeling*
- Data → *Record Linkage*

Scalability:

- Mediation
- Security
- Source Addition
- Record Linkage
- Efficient Query Execution



BIRN Biomedical Informatics Research Network

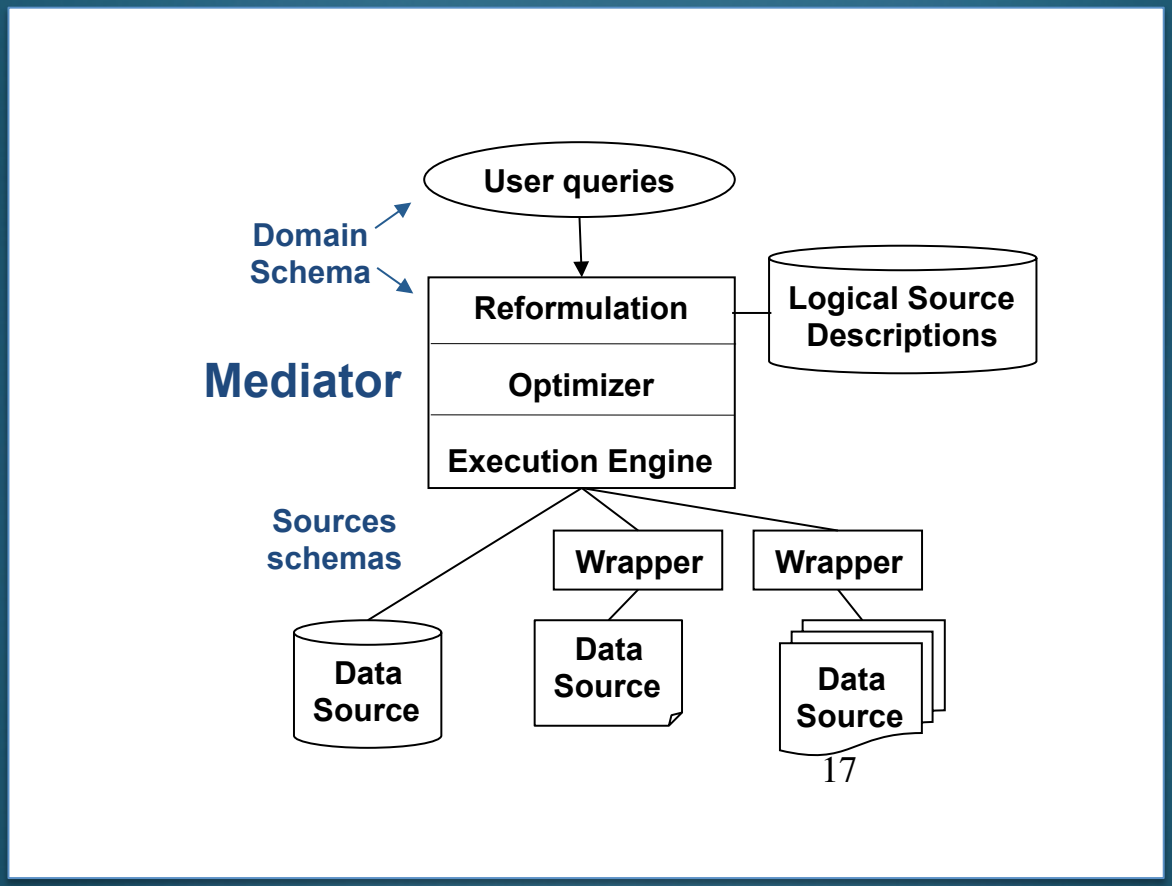
Information Mediator

- Virtual Integration Architecture:
  - *Virtual organization*: community of data providers and consumers that want to share data for specific purpose
  - *Autonomous sources*: data, control remains at sources; no change to access methods, schemas; data accessed real-time in response to user queries
  - *Mediator*: integrator defines domain schema and describes source contents
    - Domain schema: agreed upon view of the domain preferred by the virtual organization
    - Source descriptions: logical formulas relating source and domain schemas

Information Mediator

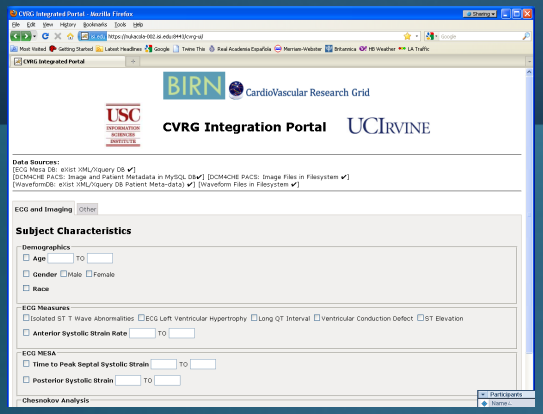
- Query Answering
  - User writes query in domain schema
  - Mediator:
    - Determines sources relevant to user query
    - Rewrites query in sources schemas
    - Breaks query into sub-queries for sources
    - Optimizes query evaluation plan
    - Combines answers from sources
  - Efficient query evaluation
    - Streaming dataflow

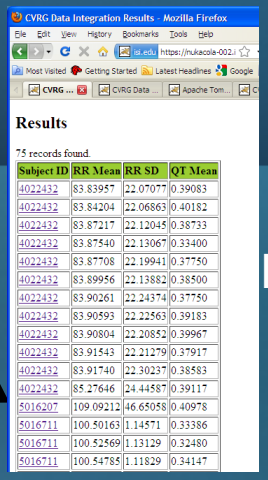
Information Mediator



Mediator Use Case - CVRG

Domain query





Subject ID	RR Mean	RR SD	QT Mean
4022432	83.83957	22.07077	0.39083
4022432	83.84204	22.06863	0.40182
4022432	83.87217	22.12045	0.38733
4022432	83.87540	22.13067	0.33400
4022432	83.87708	22.19941	0.37750
4022432	83.89956	22.13882	0.38500
4022432	83.90261	22.24374	0.37750
4022432	83.90593	22.22563	0.39183
4022432	83.90804	22.20852	0.39967
4022432	83.91543	22.21279	0.37917
4022432	83.91740	22.30237	0.38583
4022432	85.27646	24.44587	0.39117
5016207	109.09212	46.65058	0.40978
5016711	100.50163	1.14571	0.33386
5016711	100.52569	1.13129	0.32480
5016711	100.54785	1.11829	0.34147

Integrated results

Just plug in CVRG source descriptions

Logical Source descriptions

BIRN Mediator

Same BIRN mediator

ECG_Mesa (MySQL DB)

and additional wrapper for eXistDB (XML/XQuery database)

WaveformDB (eXistDB XML DBMS)

DICOM Image Files (file system)

Image Metadata dcm4che PACS (MySQL DB)

Chesnokov Analysis (eXistDB XML DBMS)

Waveform Files (file system)

BIRN Biomedical Informatics Research Network

Knowledge Engineering

Ontology development challenges:

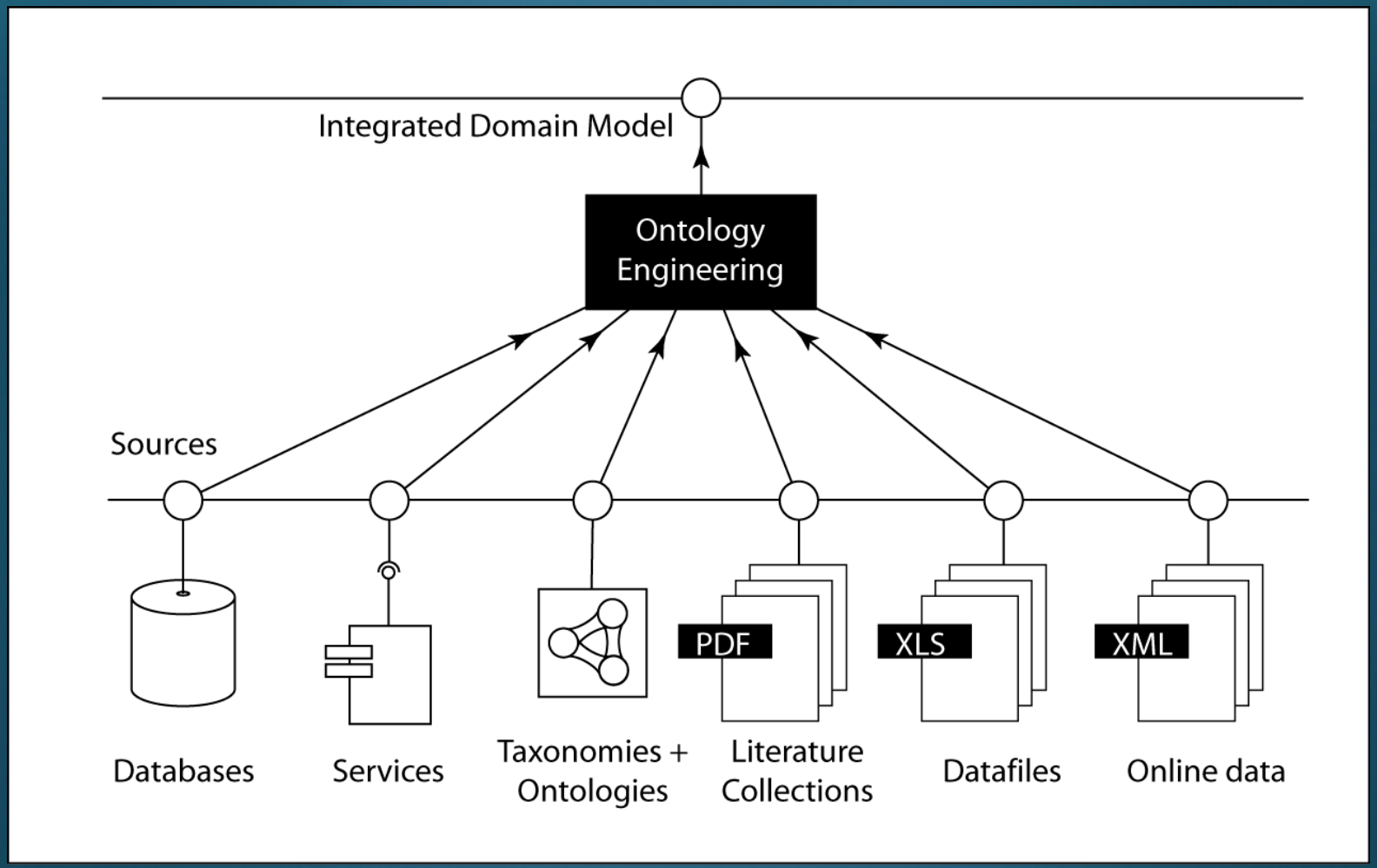
- Modeling complex domains is challenging and requires specialized expertise
- The community of ontology development efforts is large and somewhat daunting to navigate

Our goal: to provide an ontology development process that

- *leverages existing ontology development*
- *creates effective dialog between domain users and biomedical ontologists*
- *informs and documents the design of domain models for integration*
- *publishes curated domain ontologies to the community*

BIRN Biomedical Informatics Research Network

Domain/Ontology Engineering



BIRN Biomedical Informatics Research Network

BIRN Security

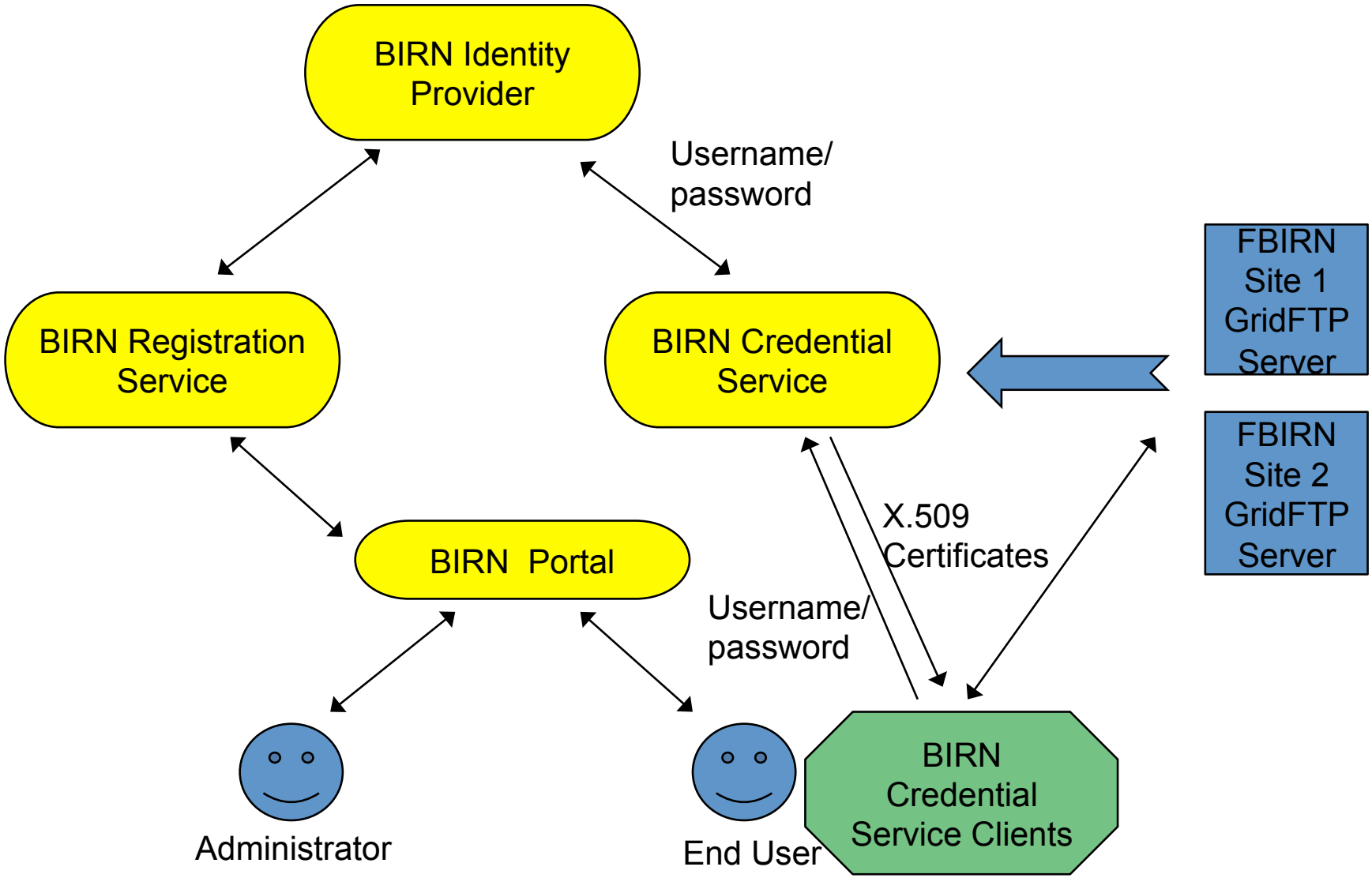
- Vet new user request for community membership
- Provide new user identity
- Provide single sign-on for users
- Manage access control policy
- Enforce access policy on resources
- Manage user groups for projects within community
- Address potential security issues

BIRN Biomedical Informatics Research Network

BIRN Security

- Hosted services for common requirements
- Tools and clients for integration with community resources
- Security Vulnerability Handling System
- Expertise and consultation for community application integration

Function BIRN Security



BIRN Best Practices Consulting

- "We need 8, 40 GB uploads and 4, 10 GB downloads simultaneously" is achievable.
- "We need to transfer data…" is not.

Data provenance, data curation, best practices, federated versus repository…

- What does it take to get things done?

BIRN Biomedical Informatics Research Network

Working with BIRN

- Projects are vetted by BIRN Steering Committee for appropriateness and resource planning
- Collaborators are expected to be actively involved in development and deployment.
- Collaborators are not expected to use complete BIRN capability kit.

Summary

- BIRN provides a capability kit and consulting to facilitate internal and external data sharing.
- BIRN takes a bottom-up, modular, capability-based approach that is driven by end user needs.
- End users are viewed as collaborators and are actively involved in development process.

BIRN Program Announcements

- Provides funds to work with BIRN on projects relating to data sharing (PAR-07-426) and the associated ontology (PAR-07-425) for the data.
- Mechanism to fund the leveraging of BIRN capabilities
- BIRN provides assistance in framing proposals.

BIRN Biomedical Informatics Research Network

Contact Information

- Email address: info@birncommunity.org

  BIRN Representatives

  Joe Ames: jdames@uci.edu

  Karl Helmer: helmer@nmr.mgh.harvard.edu

  Seth Ruffins: sruffins@loni.ucla.edu

- Web site: www.birncommunity.org

BIRN Biomedical Informatics Research Network

Acknowledgements

- Jose Luis Ambite (ISI, USC)
- Rachana Ananthakrishnan (ANL, UC)
- Gully Burns (ISI, USC)
- Carl Kesselman (ISI, USC)
- Lee Liming (ANL, UC)
- J-P Navarro (ANL, UC)
- Robert Schuler (ISI)

BIRN Biomedical Informatics Research Network