

# BIRN: Where We Have Been, Where We are Going.



Carl Kesselman  
BIRN Principal Investigator

Professor of Industrial and Systems Engineering  
Information Sciences Institute Fellow  
Viterbi School of Engineering  
University of Southern California

# Biomedical Informatics Research Network

- Infrastructure and services to support collaborative biomedical research
- Established by NCRR in 2001
- Restructured in 2008

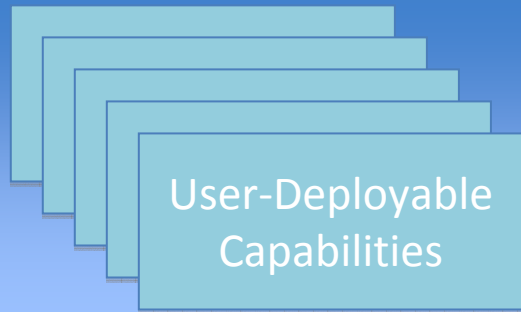
# BIRN Approach

- Use-case driven requirements, bottom up design
- Well defined roadmaps, features, functions
- Follow capabilities approach from TeraGrid
  - Promote reuse across communities
- Move to software and services model
  - BIRN not defined by the “rack”
- Outreach to new communities

# 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)

# BIRN Services



BIRN-wide Shared Services

Prototype  
Service  
Deployments

## User Teams

User-deployed  
BIRN Capabilities

User-deployed  
BIRN Capabilities

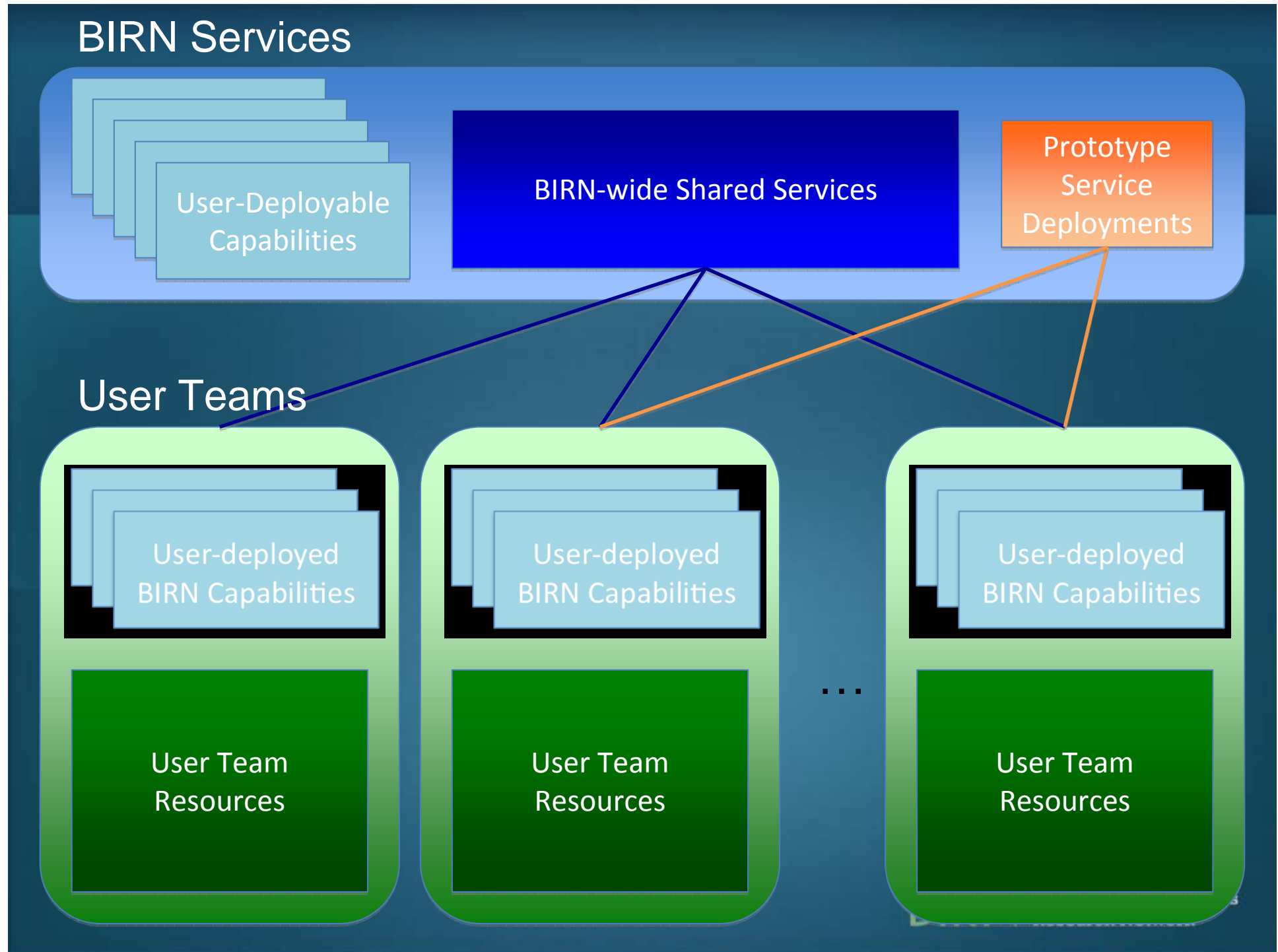
User-deployed  
BIRN Capabilities

User Team  
Resources

User Team  
Resources

User Team  
Resources

...



# Common User Needs

- Data sharing
  - Among well defined communities, e.g. NPRC
  - Facilitate sharing with the public
  - Different data types
- Infrastructure, e.g. grid FTP for sharing
- Data Integration
  - Mediation across data sources / query
- Security
  - Registration, credentials

# BIRN Services

- Website,
- Email lists
- WIKI (Confluence)
- Service Registry
- Source Repository
- User Registration (GAMA)
- Credential Access (MyProxy)

# BIRN Capabilities (Now)

Collaboration Tools	Security Services	Data Sharing
<p>Science teams have found BIRN's <a href="#">Wiki hosting</a> useful for building their communities and orienting new members.</p> <p>BIRN's <a href="#">email list manager</a> makes it easy to share information with team members, fostering a sense of community and keeping members "in the loop."</p>	<p>BIRN's <a href="#">user registration website</a> allows new users to sign up and obtain security credentials to access BIRN services.</p> <p>BIRN's <a href="#">online credential service</a> stores user credentials so that they can be retrieved and used from any computer system to access BIRN services.</p>	<p><a href="#">Data access software</a> gives BIRN teams a simple way to share data on their systems with other team members.</p> <p><a href="#">Data movement tools</a> give team members ability to upload/download or transfer data from site to site.</p> <p><a href="#">Metadata services</a> allow teams to describe their data in scientific terms and keep track of what data exists and where to get it.</p>



# BIRN Capabilities (Coming Soon)

## Collaboration Support

A group policy management service allows teams to define groups and assign BIRN service privileges (access control) to groups.

An [online capability registry](#) allows users to browse and search for BIRN capabilities and data collections and allows teams to “advertise” the capabilities they provide.

## Data Sharing Features

Data publication tools allow teams to upload data and metadata together, improving the accessibility of new data.

A [data mirroring service](#) allows BIRN sites to mirror (copy and keep up-to-date) collections of data on local servers for improved accessibility.

Our data access software supports the group access controls defined in the group policy management service.

A first release of [query tools](#) allows users to perform structured queries across multiple (heterogeneous) data sources.

# Operational Security Capabilities

- User registration
  - Portal interface for users
  - Administrator notification system
- Credential Management
  - X.509 Credential to all registered users
  - Manage long term credential for user
- Manual CA for host credentials

# Operational Data Capabilities

- Data location
- Bulk data movement of large data sets
- Bulk movement of many files

# Operational Collaboration Capabilities

- Capability registry
- Project membership management
- Project specific collaboration space
  - Wiki

# FBIRN Goals

- Develop the capability to analyze, as a single data set, data acquired from multiple sites using tools developed from multiple sites
- Develop multi-site functional neuroimaging tools.
- Develop a federated data management system to support these multi-site imaging and genetics studies

# Multi-site Data Collection Challenges

- Pre-data collection scanner QA
- Data collection
  - Effect of different stimuli and response devices
- Experimental design:
  - Cognitive task sensitivity, robustness, reliability
  - Non-neuronal measures of BOLD signal
- Post-data collection corrections
  - Use of the breathholding (BH), ASL, smoothness, etc.
- Data sharing across sites
  - Data format issues, orientations, provenance, etc.
  - Data analysis issues

# Automatic Federated Study Management Tracking

Site: UCI

## A. Clinical Data Collection

HID Database

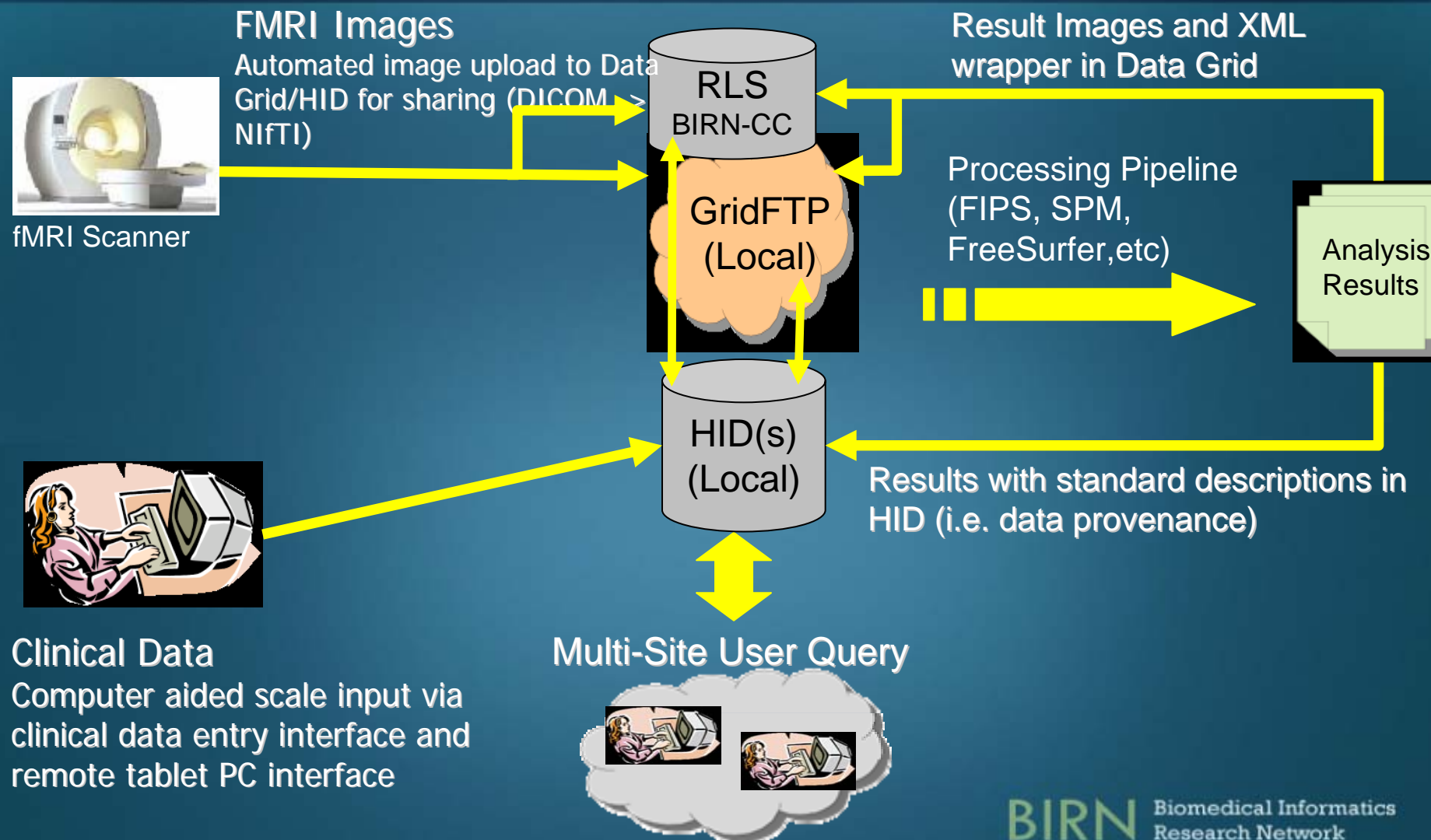
SUBJECTID	TYPE	GEN	AIM	BAR	CGI	CDI	DEM	EHS	FSS	ESR	NAA	NSA	PAN	PAS	QOL	SAN	SAP	SCI	DSS	SES
000932818211	C	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
00912345678	C	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000932764670	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000937714842	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000951069528	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000952864231	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000975340629	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000994325540	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000996678022	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000999765287	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●

## B. Imaging Data Collection

BIRN DataGrid

SUBJECTID	TYPE	HID	FTP	RLS	QAS	T1S	T2S	B01	B02	ASL	OT1	OT2	OT3	R77	OT4	OT5	R10	OT6	OT7	OD1	OD2	DTI
000932818211	C	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
00912345678	C	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000932764670	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000937714842	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000951069528	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000952864231	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000975340629	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000994325540	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000996678022	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
000999765287	Sz	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●

# Data Federation Workflow





# Clinical Assessment Form Design: Graphical Layout Editor

The screenshot displays the 'Clinical Assessment Layout Manager' software interface. The main window is divided into several panes:

- Left Pane (Tree View):** Lists form elements such as SCID\_P, Date, Time, InformantID, InformantRelation, ClinicalRefer, SCID\_P5, SCID\_P6, SCID\_P7, SCID\_P\_CHECKED\_BY, SCID\_P\_DATE\_CHECKED, SCID\_P8, SCID\_P9, SCID\_P10, SCID\_P11, SCID\_P12, SCID\_P13, SCID\_P14, SCID\_P15, SCID\_P16, SCID\_P17, SCID\_P18, SCID\_P19, SCID\_P20, SCID\_P21, SCID\_P22, SCID\_P23, SCID\_P24, SCID\_P25, SCID\_P26, SCID\_P27, SCID\_P28, SCID\_P29, SCID\_P30, SCID\_P31, SCID\_P32, SCID\_P33, SCID\_P34, and SCID\_P35.
- Central Workspace:** Shows a graphical layout of the form. It includes sections for 'MOOD DISORDERS' (Bipolar Disorder (D. 1), Bipolar Disorder (D. 2), Other Bipolar Disorder (D. 5)), 'Skip Lifetime', 'Skip pastMonth', and 'Skip2 Lifetime'. It also features a 'Summary Score Sheet' and a 'Logout' button.
- Right Pane (Preview):** Displays the final form design. It includes a header 'BDI Page 1 of 4' and a list of symptoms: 1.sadness, 2.pessimism, 3.Sense of failure, and 4.Dissatisfaction. Each symptom is followed by a list of statements with radio buttons for selection.

The preview window shows the following content:

BDI Page 1 of 4

1.sadness

- ☐ I do not feel sad.
- ☐ I do feel sad.
- ☒ I am blue or sad all the time and I can t snap out of it.
- ☐ I am so sad or unhappy that it is very painful.
- ☐ I am so sad or unhappy that I can t stand it.

2.pessimism

- ☐ I am not particularly pessimistic or discouraged about the future.
- ☐ I feel discouraged about the future.
- ☒ I feel I have nothing to look forward to.
- ☐ I feel that I won't ever get over my troubles.
- ☐ I feel that the future is hopeless and that things cannot improve.

3.Sense of failure

- ☐ I do not feel like a failure.
- ☐ I feel I have failed more than the average person.
- ☒ I feel that I have accomplished very little that is worthwhile or that means anything.
- ☐ As I look back on my life, all I can see is a lot of failures.
- ☐ I feel that I am complete failure as a person.

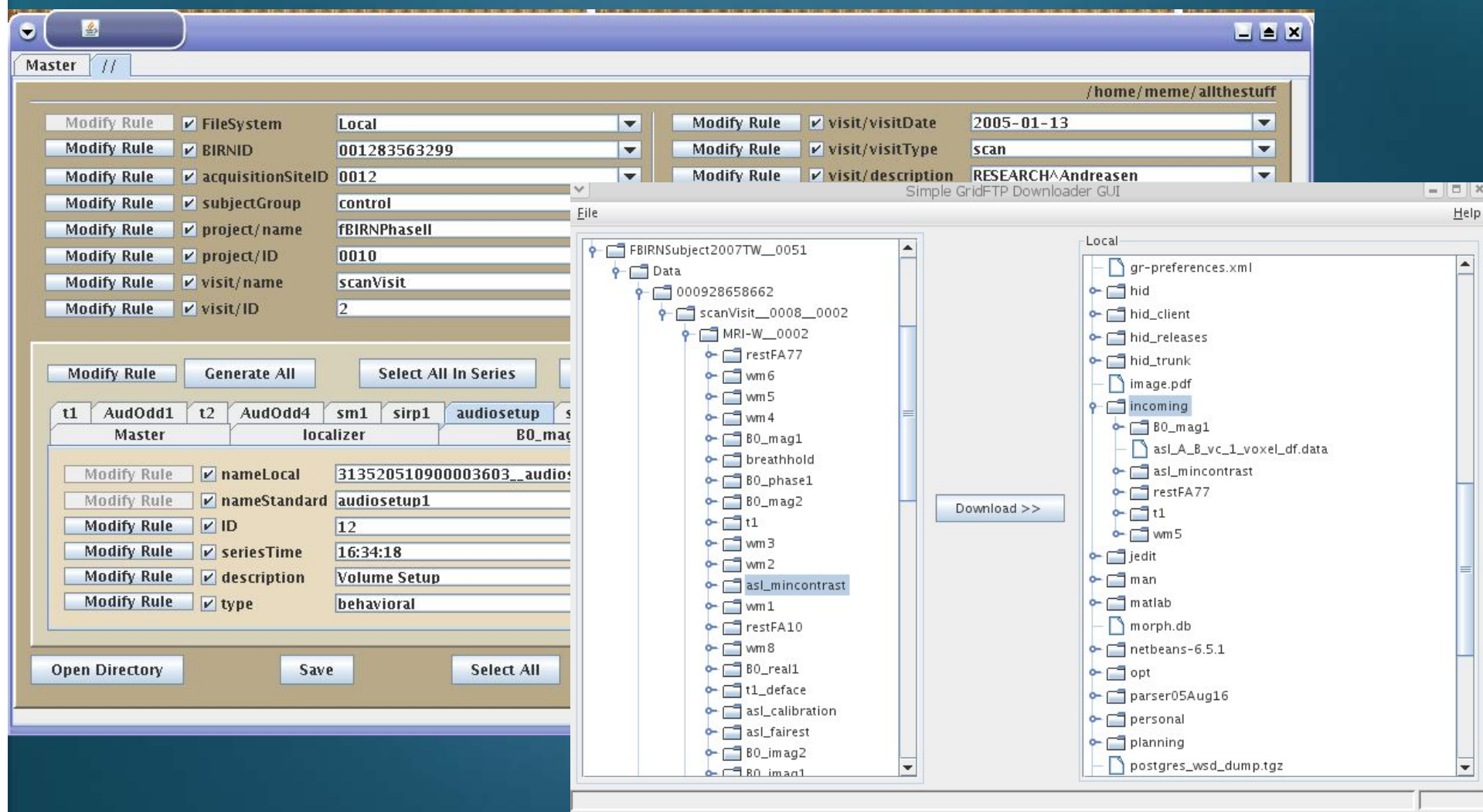
4.Dissatisfaction

- ☐ I am not particularly dissatisfied.
- ☐ I feel bored most of the time.
- ☒ I don't enjoy things the way I used to.
- ☐ I don't get satisfaction out of anything anymore.

# Data Storage Requirements

- Federated, sites control own data collections
  - Create project groups dynamically across federation
  - Assign permissions for groups dynamically
- Low cost, low maintenance overhead
- Data access quality of service requirements
- Access control
  - Site=RW, Project Group= R, User=Site
- Replication
- Performance
  - faster than scopy & sftp
- Human time
  - Improve efficiency over scopy/sftp alternatives and/or data warehousing

# Image/File GridFTP Upload/Download Tools



# HID: Federated Data Exploration

## Query Assessments

### Experiment

### Subject

2 new results

Manage Jobs

Home

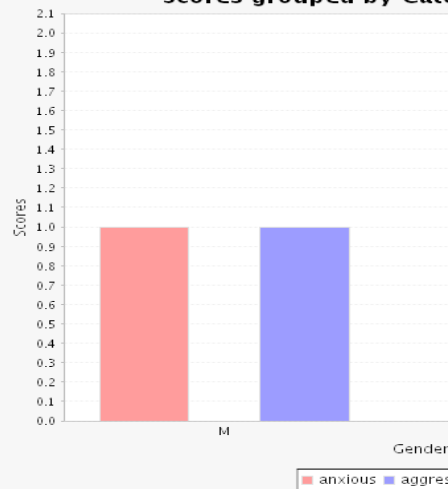
Bug Report

Logout

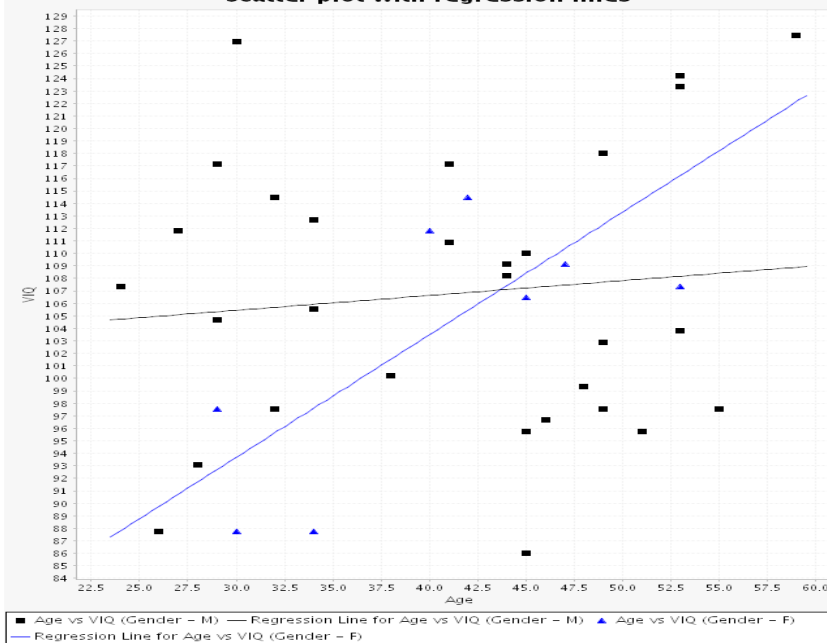
## Univariate Statistics Results

Variable	N	Mean	SD	Min	Max	Kurtosis	Skewness
anxious	3	1.67	0.58	1	2	?	0
aggressive	3	1.67	0.58	1	2	?	0

Scores grouped by Category 'Gender'



Scatter plot with regression lines



Age vs VIQ (Gender - M)

Intercept: 101.893

MSE: 127.7

Age vs VIQ (Gender - F)

Intercept: 64.245

MSE: 60.4

Slope: 0.119

N: 30

Slope: 0.981

N: 9

Pearson's R: 0.107

Coefficient of Determination  $R^2$ : 0.011

half-width of 95% confidence interval for the slope estimate: 0.427

Pearson's R: 0.753

Coefficient of Determination  $R^2$ : 0.568

half-width of 95% confidence interval for the slope estimate: 0.766

tics

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:*

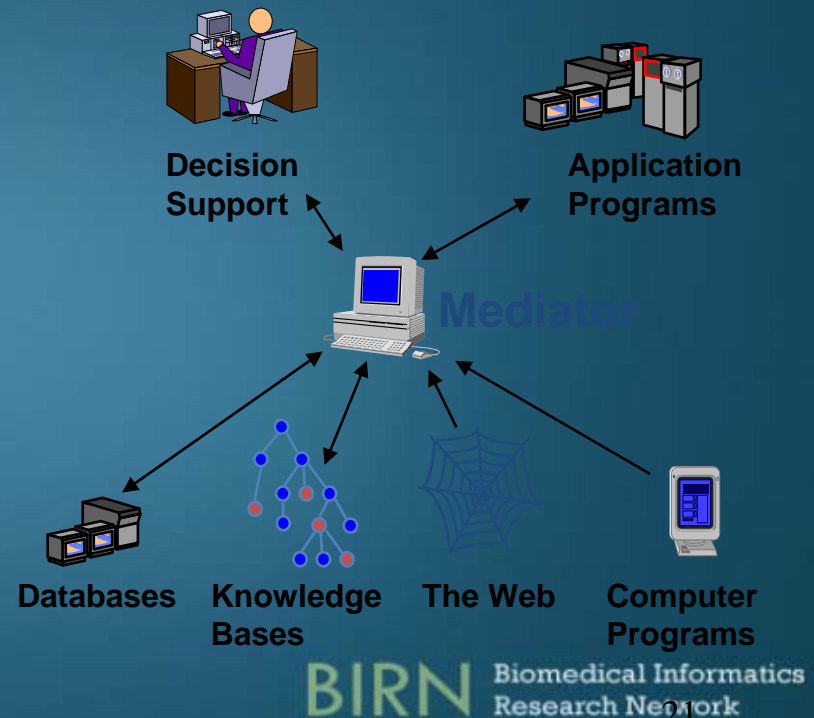
Mediation

Security

Source Addition

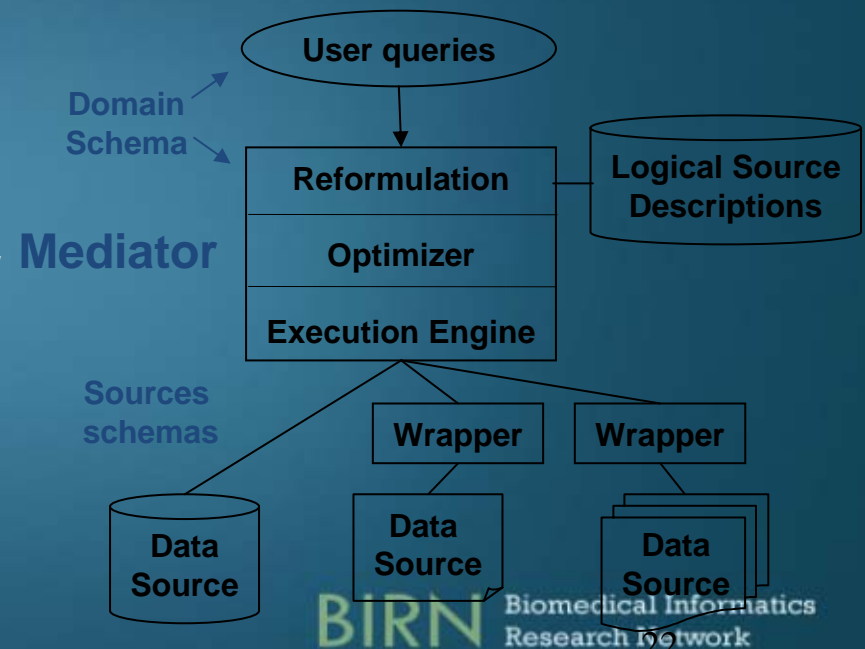
Record Linkage

Efficient Query Execution



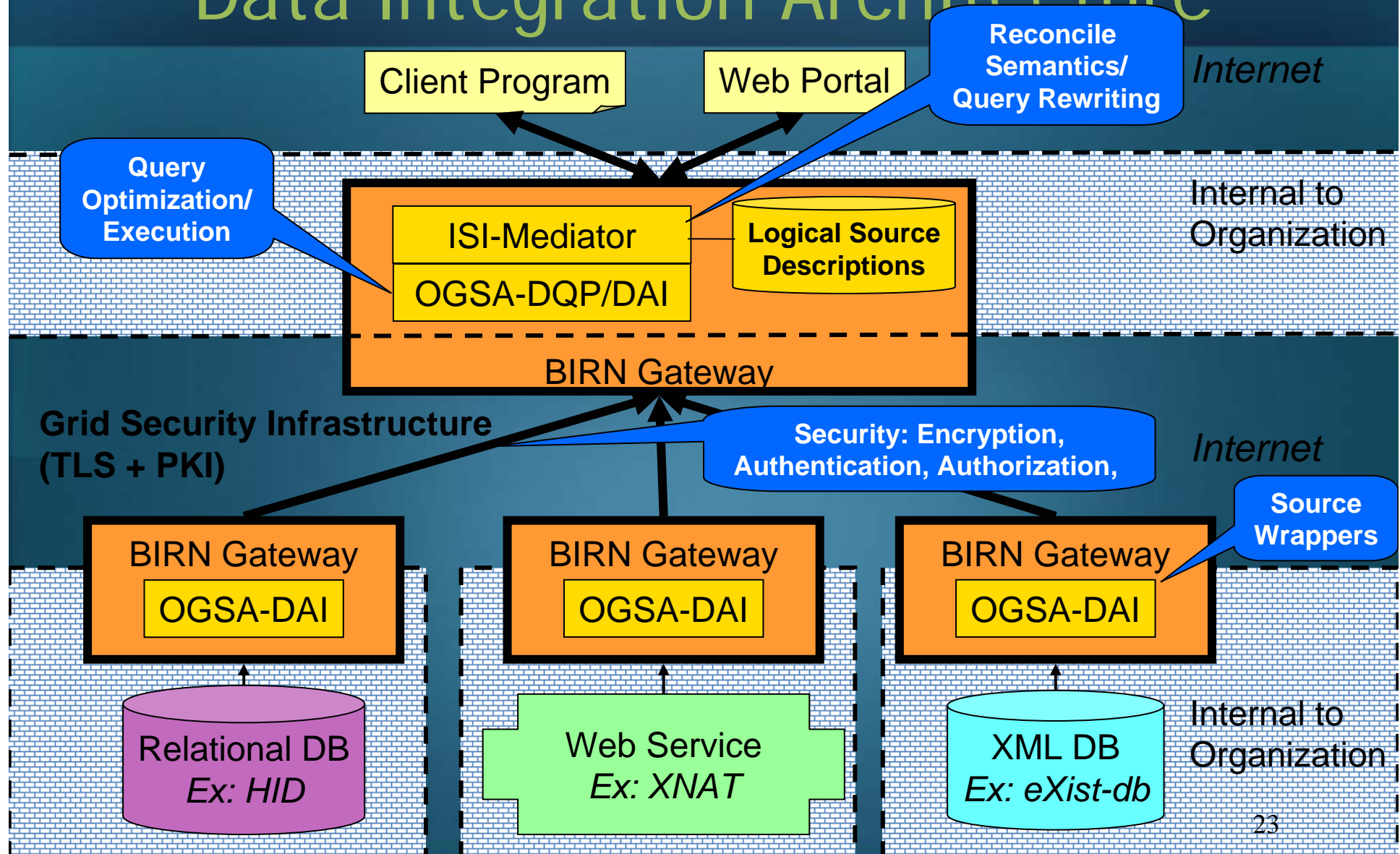
# Information Mediator

- 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
- Easy to add new sources
- 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





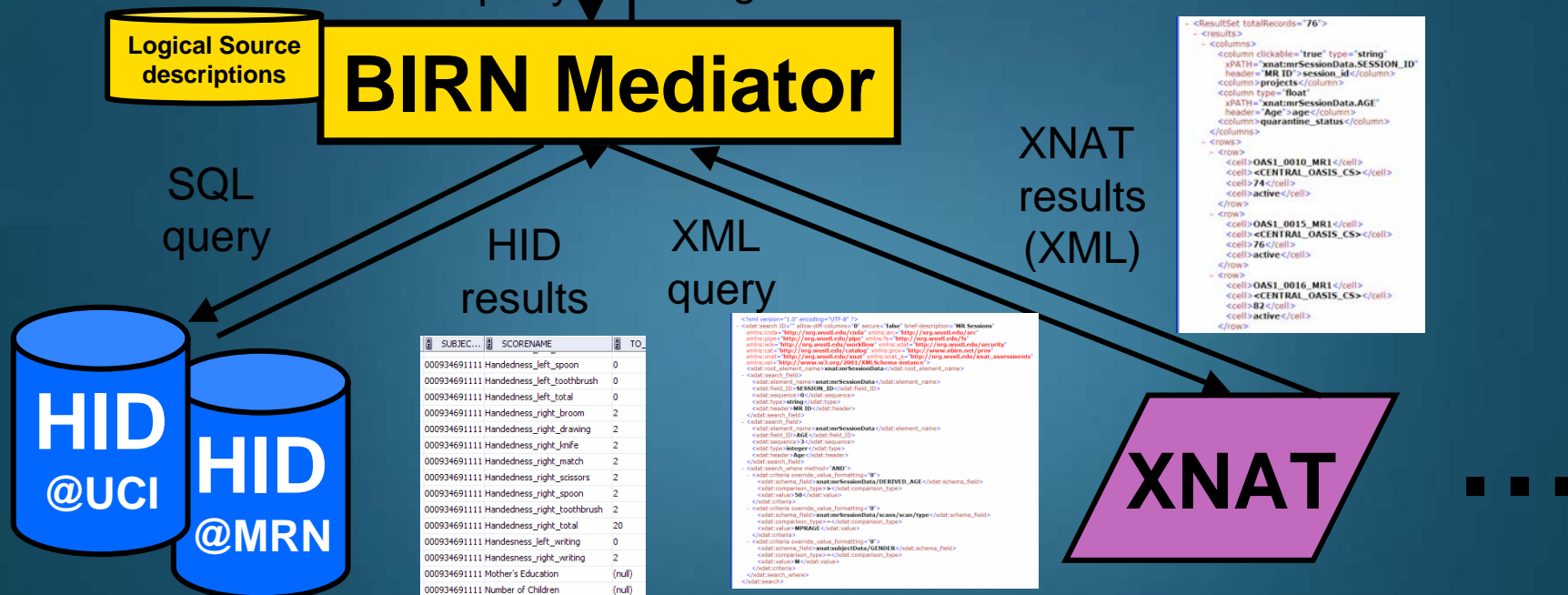
# BIRN Grid-based Virtual Data Integration Architecture



User query:  
find all male  
patients over 50  
with t1 scans

The screenshot shows the HXD NAT Integration Portal interface. At the top, there's a navigation bar with links like 'Home', 'About', 'Contact', 'Help', and 'Login'. Below this is a header with the USC logo and the text 'HDX NAT Integration Portal UCIRVINE'. The main content area is titled 'Subject Characteristics' and contains two sections: 'Scan Types' and 'Subject Characteristics'. The 'Scan Types' section has a 'Scan Type' dropdown set to 'Functional SFC (S)' and a 'Scan Interval' dropdown set to '100'. The 'Subject Characteristics' section has a 'Race' dropdown set to 'Caucasian' and a 'Complexion' dropdown set to 'Caucasian'. There is also a 'Page' number input field set to '101'.

Domain query  $\downarrow \uparrow$  Integrated results



# Human Imaging Database(s)

## Oracle DB

Extensible Neuroimaging Archive Toolkit  
Web service API



# KNAT



000934691111	Handesness_right_writing	2
000934691111	Mother's Education	(null)
000934691111	Number of Children	(null)

MR Sessions<sup>u</sup>  
du/arc<sup>u</sup>  
u/b<sup>u</sup>  
ustl.edu/security<sup>u</sup>  
rm.net/prov<sup>u</sup>  
ll.edu/xmat\_assessments<sup>u</sup>

## Results from X

Its $n >$ 

```

<cell>76</cell>
<cell>active</cell>
</row>
- <row>
  <cell>OAS1_0016_MR1</cell>
  <cell><CENTRAL_OASIS_CS></cell>
  <cell>82</cell>
  <cell>active</cell>
</row>

```

# XNAT

■ ■ ■

# Imaging Archive Toolkit

25

# Acknowledgements

- Steve Potkin, David Keater, Jose-Luis Ambite, Joe Ames