OpenFurther: An Infrastructure for Clinical & Translational Research in a Distributed Environment

AMIA Systems Demonstration
November 19, 2013
Overview

• Introduction

• OpenFurther

• Technical Components of OpenFurther

• Deployments

• Demonstration of Federated Queries
## Distributed Computing Environments

<table>
<thead>
<tr>
<th>Industry</th>
<th>Semantic Complexity</th>
<th>Penetration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finance</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Travel</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Retail</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Engineering</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Defense</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Translational &amp; Clinical Sciences and Practice</td>
<td>Very High</td>
<td>Very Low</td>
</tr>
</tbody>
</table>
Federating Infrastructures

Static Federation (caGrid)

Static Translation Layer to Common Data Model

Static Aggregation (i2b2)

Dynamic Federation (OpenFurther)

Dynamic Translation Layer

Static Translation/Aggregation Layer
Heterogenous Federation

[Diagram showing a network of ADAPT nodes connected to a central 'F' node, with further connections to diverse ADAPT nodes with different colors representing various data or entity types.]
Design Goals

• Modular design to support component based implementation

• On-the-fly real-time federation of health information from heterogeneous data sources

• Data source partners do not need to extract data and/or build a new database
  
  • Data remains in its native format
  
  • Is as up-to-date as the data source

• Join data from multiple sources for research

• Framework to support granular security control to join targeted data across data sources
Typical Applications

• Cohort Finding for Prospective Research
• Comparative Effectiveness Research (CER) Infrastructure
• Public Health Surveillance
• Datasets for Observational Studies
OpenFurther: Demo Version

• Open Source Instance

• Demonstrative version that can be downloaded, tried, modified and deployed for testing and experimentation purposes.

• Public Datasets:
  • OMOP: Secondary data for Comparative Effectiveness Research
  • OpenMRS: Medical Record System

• Release: AMIA 2013

• http://openfurther.org/
OpenFurther Technical Architecture
OpenFurther

• Utilizes components available from standards organizations and open source initiatives
  • Service Oriented Architecture (SOA) and Enterprise Service Bus
  • Relevant to national projects
  • Architecture is open and sharable.

• Systematically support centralized and distributed governance models.
Component Overview

- Query Tool
- Federated Query Engine
- Data Source Adapters
- Admin & Security Components
- Virtual Identity Resolution on the GO (VIRGO)
- Quality & Analytics Framework
- Metadata Repository
- Terminology/Ontology Server

Quality Analysis

VIRGO

Query Tool

Terminology Server

Metadata Repository

Counts & Data

Security

Data Sources
i2b2/OpenFurther Query Tool Architecture

Web Server

i2b2 Web Client

XML Query

InterceptorFilter

web.xml configuration

FQE Web Service

QueryToolService

JBoss/Tomcat - Axis2 Webapp
Federated Query Engine

FQE Storage  In Memory

FQE

Query Topic

A ∩ B

A

B

Status Msgs

Result Msgs

In Memory

Data Source Façade Layer

Subscribe

Data Source 1 Adapter

ADAPT

Data Source 2 Adapter

ADAPT

XML

1

2

3

4

Subscribe

Subscribe

ADAPT
Data Source Adapters

Datasource Façade

1. Initialize
2. Query Translation
3. Execution
4. Result Translation

MDR

DTS

Data Source

FQE Layer

Subscribed

Status

Result
Security Components

Unauthenticated

Authentication System

LDAP, CAS, SAML

Authenticated to OpenFurther

Contextual:
User properties, roles, and authorization is different depending on the context (namespace)

Authorization:
Authorization happens at multiple levels: within the FQE and within the data source adapters.

Authorization within the OpenFurther namespace may be different than within the data source adapter (a different namespace).

Already Authorized:
Some data source adapters may not even require explicit authorization.

Authenticated to OpenFurther

Security Module: OpenFurther Namespace

Federated Query Engine

Request Topic

Subscribed

Security Module: DS1 Namespace

Data Source Adapter

Subscribed

Data Source Adapter
Virtual Identity Resolution on the GO (VIRGO)

**On the fly demographics analysis for record linkage**
- No permanent PHI storage
- Response composed of OpenFurther IDs

**Algorithms**
- Field-specific weights
- Adaptable to others

**Technologies used**
- Java, Groovy, Grails
- Web-services
- MySQL (temp storage)
- ElasticSearch (indexing)
Quality & Analytics Framework

A service oriented architecture that can assess the quality of heterogeneous electronic health data sources in a distributed environment.
Metadata Repository (MDR)

• Built in-house, standards-based

• Artifacts (Knowledge)
  • Logical Models, Local Models, model mappings
  • Administrative information
  • Descriptive information
  • XQuery Translation Programs

• Models supported - Open Source Initiative
  • Local: FURTHeR, UUEDW, UPDBL, Intermountain Healthcare Datamart
  • Public: OMOP, i2b2, OpenMRS
Translating Metadata

Translating Coded Values

<table>
<thead>
<tr>
<th>Data Element</th>
<th>Value Namespace</th>
<th>Value Code</th>
<th>Value Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenFurther.Person.administrativeGender</td>
<td>SNOMED</td>
<td>248153007</td>
<td>Female</td>
</tr>
<tr>
<td>OMOP.Person.genderConceptId</td>
<td>OMOP</td>
<td>2940</td>
<td>Female</td>
</tr>
</tbody>
</table>
Query: Females age 30 to 40 who have diabetes

OMOP Query:
- genderConceptId = 2940
- AND yearOfBirth between 1972 and 1982
- AND conditionOccurrence.conditionConceptId = 64547

OpenFurther Query:
- administrativeGender = SNOMED:248153007
- AND age between 30 and 40
- AND observation.observationCode = ICD9:250
Result Translation

1,000 records require > 12,000 translations (Person data class only)

OMOP.Patient

- personId
- yearOfBirth
- genderConceptId
- raceConceptId
- ethnicityConceptId
- providerId
- careSiteId

Get Master Person ID

Calculate Age

Translate Metadata

Translate Metadata

Translate Metadata

Translate Code

Translate Metadata

Translate Code

Translate Metadata

Translate Code

Translate Metadata

Translate Code

Translate Metadata

Translate Code

OpenFurther.Person

- masterPatientId
- age
- dateOfBirth
- administrativeGender
- race
- maritalStatus
- religion
- dateOfDeath
- pedigreeQuality
Terminology/Ontology Server

• Apelon’s Distributed Terminology Server – an integrated set of open source components that provides comprehensive terminology services in distributed application environments

• Provides tools for:
  • Management standard and local terminologies
  • Mapping local to standards
  • Browsing & Searching terminologies
  • Creation subsets
  • Extension of standards terminologies
  • Versioning of terminologies

• OpenFurther includes a layer of web-services that leverage DTS APIs
DTS Terminology Content

Within UU’s Installation

• ~25 Standard Namespaces
  (ICD-9, ICD-10, SNOMED, LOINC, RxNorm & More)

• ~33 Local Namespaces

• ~1 Million Concepts

• ~2 Million Total Mappings

• Subscribe for content updates

With Demo Installation

• Free content available from Apelom
  (subsets of ICD9, SNOMED, LOINC)

• Local content for two simulated data sources
  • Schultz Cancer Repository
    (OMOP Data Source)
  • Schultz Healthcare Systems
    (OpenMRS Data Source)

• Mappings between local data sources and standards

• Additional content could obtained from their respective sources or Apelom.
Deployments

- University of Utah FURTHeR: Cohort Identification & Datasets for Analysis
- PHIS+: Aggregated database for performing Comparative Effectiveness Research
- University of North Carolina: Cohort Identification
Vision for OpenFurther at the University of Utah
Comparative Effectiveness Research Infrastructure
Augment Children’s Hospital Association’s (CHA) existing electronic database of administrative data - Pediatric Health Information System (PHIS) with clinical data to conduct Comparative Effectiveness Research studies.

UU Biomedical Informatics - Informatics Partners

Agency for Healthcare Research and Quality (AHRQ) funded project.
PHIS+ Overview

3. Data Streams
   - Laboratory
   - Microbiology
   - Radiology

4. CER Studies
   - Pneumonia
   - Appendicitis
   - Osteomyelitis
   - Gastroesophageal Reflux Disease

5. Years Data
   - 2007 – 2011
   - 2009 – Development
   - 2012….
The PHIS+ Process

Pediatric Research in Inpatient Setting (PRIS) Sites
Developmental Process Overview

Narus et. al, Federating Clinical Data from Six Pediatric Hospitals: Process and Initial Results from the PHIS+ Consortium. AMIA 2011
**PHIS+ CER Database – 2007-11**

### Laboratory

<table>
<thead>
<tr>
<th>Site</th>
<th>Results</th>
<th>LOINC Lab Test Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15,011,312</td>
<td>538</td>
</tr>
<tr>
<td>B</td>
<td>33,214,540</td>
<td>1,214</td>
</tr>
<tr>
<td>C</td>
<td>16,868,383</td>
<td>860</td>
</tr>
<tr>
<td>D</td>
<td>25,706,608</td>
<td>1,089</td>
</tr>
<tr>
<td>E</td>
<td>38,422,668</td>
<td>1,016</td>
</tr>
<tr>
<td>F</td>
<td>14,507,629</td>
<td>2,131</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>143,731,140</strong></td>
<td><em>6,848 (2,992)</em></td>
</tr>
</tbody>
</table>

### Radiology

<table>
<thead>
<tr>
<th>Site</th>
<th>Reports</th>
<th>CPT Radiology Procedure Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>445,681</td>
<td>280</td>
</tr>
<tr>
<td>B</td>
<td>1,151,383</td>
<td>349</td>
</tr>
<tr>
<td>C</td>
<td>635,458</td>
<td>296</td>
</tr>
<tr>
<td>D</td>
<td>980,740</td>
<td>482</td>
</tr>
<tr>
<td>E</td>
<td>1,098,693</td>
<td>497</td>
</tr>
<tr>
<td>F</td>
<td>201,708</td>
<td>477</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,513,663</strong></td>
<td><em>2,381 (714)</em></td>
</tr>
</tbody>
</table>

### Microbiology

<table>
<thead>
<tr>
<th>Site</th>
<th>Culture Results</th>
<th>SNOMED Specimen Code</th>
<th>SNOMED Culture Procedure Code</th>
<th>SNOMED Organism Code</th>
<th>RxNorm Antimicrobial Code</th>
<th>Susceptibility Results</th>
<th>LOINC Susceptibility Test Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>247,933</td>
<td>114</td>
<td>70</td>
<td>113</td>
<td>57</td>
<td>487,813</td>
<td>97</td>
</tr>
<tr>
<td>B</td>
<td>359,780</td>
<td>58</td>
<td>42</td>
<td>56</td>
<td>58</td>
<td>393,594</td>
<td>85</td>
</tr>
<tr>
<td>C</td>
<td>231,071</td>
<td>179</td>
<td>46</td>
<td>162</td>
<td>59</td>
<td>340,100</td>
<td>99</td>
</tr>
<tr>
<td>D</td>
<td>335,606</td>
<td>110</td>
<td>34</td>
<td>145</td>
<td>57</td>
<td>376,844</td>
<td>75</td>
</tr>
<tr>
<td>E</td>
<td>486,315</td>
<td>130</td>
<td>56</td>
<td>160</td>
<td>59</td>
<td>605,000</td>
<td>76</td>
</tr>
<tr>
<td>F</td>
<td>176,848</td>
<td>264</td>
<td>71</td>
<td>121</td>
<td>51</td>
<td>283,865</td>
<td>89</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,837,553</strong></td>
<td><em>855 (451)</em></td>
<td><em>319 (95)</em></td>
<td><em>757 (203)</em></td>
<td><em>341 (74)</em></td>
<td>2,487,216</td>
<td><em>521 (136)</em></td>
</tr>
</tbody>
</table>

* The first number is the total number of standard codes, the second in parenthesis is the distinct number of standard codes across all sites.
Thank You
Contribute to OpenFurther

Learn more at: openfurther.org

To contribute, join and/or post on our mailing lists, submit a pull request on GitHub, or a bug on JIRA.

Community email Lists:
openfurther-user@googlegroups.com for user and implementation discussion.
openfurther-dev@googlegroups.com for development discussion.
openfurther-security@googlegroups.com for security sensitive issues and discussions.
ortherfurther-commits@googlegroups.com for developer commits to version control

Source code:
http://github.com/openfurther/

Reference manual:

Bugs & feature requests:
https://openfurther.atlassian.net
Acknowledgements

Current OpenFurther Team

Rick Bradshaw, MS
Dustin Schultz, MS
Ryan Butcher, MS
Ram Gouripeddi, MBBS, MS
Randy Madsen, BS
Phillip Warner, MS
Peter Mo, MS
Naresh Sundar Ranjan, MS
Bernie La Salle, BS
Julio Facelli, PhD

Collaborations

- UU EDW Team
- UU PPR/UPDB Team
- UU CHPC

- Collaboration Partners
  - Intermountain Healthcare
  - SLC VAMC
  - UDOH
  - PHIS+ Team members across 6 institutions
  - Children’s Hospitals Association

- Apelon
- Center for High Performance Computing, UU

OpenFurther supported by the NCRR/NCATS Grants UL1RR025764 and 3UL1RR025764-02S2, National Center for Clinical and Translational Science UL1TR001067, along with the University of Utah Research Foundation, and grant 1D1BRH20425 (DHHS). PHIS+ funded by grant R01 HS019862 from AHRQ, (DHHS). VIRGO (MPI) supported by NLM Grant 5RC2LM010798. REDCap is supported by Center for Clinical and Translational Sciences grant support 8UL1TR000105.
References

- Schultz ND, Bradshaw RL, Mitchell JA. Utilizing Previous Result Sets as Criteria for New Queries within FURTHEr. 2012 SHARPn Summit on Secondary Use, Rochester, MN.
- Website: http://openfurther.org