i2b2 Roadmap
June 18, 2013

Shawn Murphy MD, Ph.D.
Michael Mendis
Nich Wattanasin MS
Lori Phillips MS
Wensong Pan MS
Janice Donahoe
Susanne Churchill Ph.D.
Isaac Kohane MD, Ph.D.
I2b2 Roadmap

1 – Supporting cohort discovery and recruitment “out of the box” for clinical and observational trials.

2 – Supporting future query systems either outside of the Data Repository or within NoSQL systems.
   - Imaging
   - Genomics
   - Text / Unstructured data

3 – Supporting add-on plug-ins, web services, and ETL processes
Supporting cohort discovery and recruitment

- First - Allow Manager to obtain Identified Patient Sets

  - First step in creating a new data mart
  - Patient MRNs can be obtained though Workbench or though Identity Management Cell for Web Client
  - Assumes Patient Mapping table contains the list of encrypted MRNs mapped to i2b2 patient numbers (could be plain text MRNs depending on site policy)
Working with patient sets – Patient Set Viewer
Drag single or sets of patients into View
Patient Sets are first-class objects (like queries)

- Patients are first-class objects and can be used in queries and grouped together in Workplace Cell

- Patient Mapping table can manage groups of patients for projects in one database
Patient is a first-class object
Creating computation Patient Sets from patients

Add Patients to the Query Tool

Run the Query . . .
Creating Patient Sets

New Patient Set appears in Previous Queries & Patient Sets Views
Patient Sets in mapping table are strategy to creating several projects in one database

- Patient Mapping table can manage groups of patients for projects in one database
Workflow to support Clinical Trials

- Person does query as obfuscated user in large data mart
- Optimal query results can be used to create request for approval so that patients can be viewed as a limited data set
- Approval is obtained and a new project is created where those patients in the Optimal patient set can be viewed in plug-ins such as timeline, charts, and de-identified SMART views.
- Patients are carefully screened in limited-data displays to sort into good candidate patients for the Clinical trial.
- PHI is viewed on the truly Optimal patients in a specially Audited view that resembles EMR
- PHI enables patients to be contacted though mechanisms that abide by hospital policy
Management of Projects

“demo” project has the following i2b2 patients
  1000000001
  1000000005

“demo2” project has the following i2b2 patients
  1000000001
  1000000126

* Patient number 1000000001 exists in both projects
Mix and match ontology tables and patient databases for various projects.
Patients in PATIENT_MAPPING table for the two projects called demo=X and demo2=Z

<table>
<thead>
<tr>
<th>PATIENT_IDE</th>
<th>PATIENT_IDE_SOURCE</th>
<th>PATIENT_NUM</th>
<th>PATIENT_IDE_STATUS</th>
<th>PROJECT_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2000001961</td>
<td>1000000001</td>
<td>A</td>
<td>demo</td>
</tr>
<tr>
<td>2</td>
<td>1000000001</td>
<td>1000000001</td>
<td>A</td>
<td>demo</td>
</tr>
<tr>
<td>3</td>
<td>3000001821</td>
<td>1000000001</td>
<td>A</td>
<td>demo</td>
</tr>
<tr>
<td>4</td>
<td>2000001965</td>
<td>1000000005</td>
<td>A</td>
<td>demo</td>
</tr>
<tr>
<td>5</td>
<td>1000000005</td>
<td>1000000005</td>
<td>A</td>
<td>demo</td>
</tr>
<tr>
<td>6</td>
<td>3000001825</td>
<td>1000000005</td>
<td>A</td>
<td>demo</td>
</tr>
<tr>
<td>7</td>
<td>2000001961</td>
<td>1000000001</td>
<td>A</td>
<td>demo2</td>
</tr>
<tr>
<td>8</td>
<td>1000000001</td>
<td>1000000001</td>
<td>A</td>
<td>demo2</td>
</tr>
<tr>
<td>9</td>
<td>3000001821</td>
<td>1000000001</td>
<td>A</td>
<td>demo2</td>
</tr>
<tr>
<td>10</td>
<td>17028580</td>
<td>1000000026</td>
<td>A</td>
<td>demo2</td>
</tr>
<tr>
<td>11</td>
<td>1000000026</td>
<td>1000000026</td>
<td>A</td>
<td>demo2</td>
</tr>
<tr>
<td>12</td>
<td>4000002026</td>
<td>1000000026</td>
<td>A</td>
<td>demo2</td>
</tr>
</tbody>
</table>
CT Web Client being developed by CBMI
Supporting Future Query Systems

- Supporting future query systems either outside of Data Repository or NoSQL
  - Imaging
  - Genomics
  - Text / Unstructured data
Supporting Future Query Systems

- Genomic Subsystem
- Imaging Subsystem
- I2b2 Project
- Registries
- SNV:PPARq
- Male
- Diabetes mellitus

Other concepts found in CRC

Genomic concept found in Genomic Subsystem

CRC
Query Command Result

Male/Diab/PS Run Query processing

Depends upon Patient Set Infrastructure

<table>
<thead>
<tr>
<th>Query</th>
<th>Male</th>
<th>Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEL:Chr</td>
<td>Male</td>
<td>Diabetes</td>
</tr>
<tr>
<td>DEL:Chr</td>
<td>Male</td>
<td>Diabetes</td>
</tr>
<tr>
<td>Male/Diab/PS</td>
<td>Run Query</td>
<td>processing</td>
</tr>
</tbody>
</table>

CRC

Subsystem Plug-in

Subsystem Plug-in

DEL:Chr3

HIVE:153
HIVE:149
HIVE:132
Depends on Ontology Management Tools
Depends on Management of Patient Identity

- Identity Management (IM) cell is one of the primary core cells within the i2b2 hive
- Capable of converting i2b2 patient numbers back into identifying MRNs
- Allows multiple identifiers to be resolved by an Enterprise Master Patient Indexes
- Allows auditing in to be centralized among multiple cells
Enables Data flow of next-gen sequencing

- base calls from the sequencer
- FASTQ files with base calls
- SAM with standard alignment
- VCF digests variants
- GVF maps to ontologies

Map-Reduce Queries
- Sequence patterns

De-identified Data Warehouse

Sequence patterns
Supporting community add-on plug-ins, web services, and ETL processes

- Streamlined management of download process for supporting plug-ins (will discuss tomorrow)

- Web service architecture to feed data to i2b2

- ETL Bulk loading and Library (will continue tomorrow)
EHR systems will generate C-CDA compliant Continuity of Care Documents (CCDs) as part of Meaningful Use.

The Services ETL cell (SETL) will support their use to:

- Retrieve the latest information on a patient
- ETL CCDs into i2b2
- View CCDs in SMART apps
Data flow with Services ETL

- **Data Repository (CRC)**
  - PDO to **SMART**
  - PDO to **SETL**
  - PDO to **Mapper**
- **PDO** to **CCDA Documents**
- **PDO** to **Query Health CEDD-based Information Model (Ontology)**

**SMART App**

Red = Local Codes
Green = Standard Terminologies
Project Status

- Currently imports demographics from C-CDA documents and the SMART cell can consume them.
  - Running at Partners.

- Design is underway for a full version that will import all document sections required by Meaningful Use.
  - It will use Open Health Tools and the C-CDA ontology developed as part of Query Health.

- Major CCD sections:
  - Problems – SNOMED
  - Demographics – HL7 Codes
  - Meds – RxNorm
  - Labs – LOINC
  - Procedures – SNOMED
  - Allergies – RxNorm
  - Vitals – LOINC
  - Immunizations - CVX codes
  - Smoking, cognitive, and functional status – SNOMED
Existing Workflow: A SMART app makes a Medications API call to the SMART cell. The SMART cell retrieves Medications data from the i2b2 data warehouse.

New Workflow: A SMART app makes a Demographics API call to the SMART cell. The SMART cell sends the request to the new SETL cell to receive real-time contact data using a CCDA from Partners clinical web services.
Services ETL: Input PDO Request

```xml
<ns3:request xsi:type="ns3:GetPDOFromInputList_requestType"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <input_list>
        <patient_list>
            <patient_id>XXXXXXXX</patient_id>
            <patient_id>MGH:0000004</patient_id>
        </patient_list>
    </input_list>
    <filter_list>
        <panel name="DEM">
            <panel_number>0</panel_number>
            <panel_accuracy_scale>0</panel_accuracy_scale>
            <invert>0</invert>
            <item>
                <hlevel>1</hlevel>
                <item_key>\i2b2_DEMO\i2b2\Demographics</item_key>
                <dim_tablename>concept_dimension</dim_tablename>
                <dim_dimcode>\i2b2_DEMO\i2b2\Demographics</dim_dimcode>
                <item_is_synonym>N</item_is_synonym>
            </item>
        </panel>
    </filter_list>
    <output_option>
        <patient_set select="using_input_list" onlykeys="false"/>
        <pid_set select="using_input_list" onlykeys="true"/>
        <observation_set blob="true" onlykeys="false" select="using_input_list"/>
    </output_option>
</ns3:request>
```
Services ETL: Output PDO Response

<ns2:patient_set>
    <patient>
        <patient_id source="hive">X0000000</patient_id>
        <param column="vital_status_cd" name="vital_status_cd">U</param>
        <param column="birth_date" name="birth_date">19490101</param>
        <param column="sex_cd" name="sex_cd">M</param>
        <param column="language_cd" name="language_cd">SPANISH</param>
        <param column="religion_cd" name="religion_cd">PROTESTANT</param>
        <param column="race_cd" name="race_cd">OTHER</param>
        <param column="ethnicity_cd" name="ethnicity_cd">AFRICAN</param>
        <param column="marital_status_cd" name="marital_status_cd">UNKNOWN</param>
        <param column="legal_first_name" name="legal_first_name">SANTA</param>
        <param column="legal_middle_initial" name="legal_middle_initial">J</param>
        <param column="legal_last_name" name="legal_last_name">CLAUS</param>
        <param column="legal_suffix" name="legal_suffix">JR</param>
        <param column="permanent_line1" name="permanent_line1">1010 TENTH STREET</param>
        <param column="permanent_line2" name="permanent_line2">APT 110</param>
        <param column="permanent_city" name="permanent_city">BOSTON</param>
        <param column="permanent_city" name="permanent_city">MA</param>
        <param column="permanent_city" name="permanent_city">02114</param>
        <param column="permanent_city" name="permanent_city">US</param>
        <param column="local_line1" name="local_line1">55 FRUIT ST</param>
        <param column="local_line2" name="local_line2">APT 2</param>
        <param column="local_city" name="local_city">BOSTON</param>
        <param column="local_city" name="local_city">MA</param>
        <param column="local_city" name="local_city">02114</param>
        <param column="local_city" name="local_city">US</param>
        <param column="primary_phone" name="primary_phone">9781231231217</param>
        <param column="work_phone" name="work_phone">7814445553333</param>
        <param column="mobile_phone" name="mobile_phone">6039275569</param>
        <param column="other_phone" name="other_phone">(987)111-1111</param>
    </patient>
</ns2:patient_set>
SMART-i2b2 patient-centric views

Demographics data is retrieved in real-time from the SETL cell
1. Send the i2b2 file to the FR

2. Tell the CRC the file is ready to load

3. SSIS package loads the i2b2 file to observation_fact table
Integration of CRC and FR for file processing

1. Send Bulk i2b2 message (as attachment?)
2. Tell the CRC the file is ready to load
3. Trigger Native Bulk Loaders to Import files

CRC

FR

Native App

I2B2
Continued Development

- Simplification of Temporal Query Interface
- Streamline i2b2 Administration
- Support of PostgreSQL
Temporal Query Terminology

- **Patient First Visit**: 1/1/2005
- **Patient Last Visit**: 12/31/2010

**Event 1**
- **First Occurrence of Event 1**

**Event 2**
- **Last Occurrence of Event 2**

**Event 3**
- **Span between Last Occurrence of Event 1 and First Occurrence of Event 3**
- **Any Occurrence of Event 3**
Simplification to the Temporal Query Interface

- Background, non-temporal, part of query tool
  - Changed panel orientation to accommodate longer names, modifiers, and value settings
  - Consolidated controls for occurrence and exclusion
  - Panel timing: All occurs in same patient or in same visit

- Interface for temporal queries
  - Everything in an event occurs in one observation
Defining Underlying Patient or Visit Set
Defining Temporal Relationships

The image shows a software interface for defining temporal relationships among events. The interface includes a query editor where users can add events, specify date constraints, and define relationships between events. The example includes an event named "Asthma" and another event named "Albuterol," with the relationship defined as "Start of Assay before Event 1."
Viewing Temporal Relationships in Time Align
Viewing Temporal Relationships in Time Align
Updates

- JBoss 7.1
- All POJO
- Axis 1.6.1
- Oracle 11g
JBoss 7

✓ Improved Performance
✓ Better Security
✓ J2EE 6
✓ Administration Improvements
✓ Integrate into Eclipse
DATASOURCE CAN BE CONFIGURED VIA WEB
Debugging in eclipse
### JBoss Application Server 7.1

#### Console Handlers
Defines a handler which writes to the console.

<table>
<thead>
<tr>
<th>Name</th>
<th>Log Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSOLE</td>
<td>DEBUG</td>
</tr>
</tbody>
</table>

#### Details
- **Name**: CONSOLE
- **Log Level**: DEBUG
- **Target**: System.out
- **Encoding**:
- **Formatter**:
- **Auto Flush**: true
Summary

- Supporting cohort discovery and recruitment “out of the box” for clinical and observational trials.
- Supporting future query systems either outside of the Data Repository or within NoSQL systems
- Supporting add-on plug-ins, web services, and ETL processes
- Simplification of Temporal Query Interface
- Streamline i2b2 Administration
- Support of PostgreSQL
Special thanks to...

- **SMART i2b2 Team**
  - Nich Wattanasin - Project Manager
  - Alyssa Porter - Analyst
  - Stella Ubaha – Developer
  - Jeff Klann - Informatician

- **i2b2 – SHRINE Team**
  - Nich Wattanasin
  - Michael Mendis
  - William Simons
  - Douglas MacFadden

- **SMART Core Team**
  - Isaac Kohane – Co-PI
  - Kenneth Mandl – Co-PI
  - Joshua Mandel
  - Rachel Ramoni
  - David Kreda

- **i2b2 Core Team**
  - Isaac Kohane
  - Susanne Churchill
  - Shawn Murphy
  - Michael Mendis
  - Lori Phillips
  - Wensong Pan
  - Janice Donahoe
  - Nich Wattanasin
  - David Wang
  - Christopher Herrick
  - Bill Wang
  - Vivian Gainer
  - Andrew Cagan
Web Resources

- SMART i2b2 Homepage: www.smarti2b2.org
- SMART Platforms Homepage: www.smartplatforms.org
- i2b2 Community Site: community.i2b2.org
- i2b2 Software: www.i2b2.org/software
- i2b2 Homepage: www.i2b2.org