Developing a Self-Service Query Interface for Re-Using De-Identified Electronic Health Record

James J. Cimino, Elaine J. Ayres, Andrea Beri, Robert Freedman, Ellen Oberholtzer, Sachi Rath

Laboratory for Informatics Development
NIH Clinical Center
Bethesda, Maryland, USA
Motivation for a Self-Service Data Repository

- Human mediator is expensive
- Using a mediator slows the process
- User will learn to better understand the data
- User empowerment will improve spontaneity and creativity
The US National Institutes of Health (NIH)

- Premiere US government health research institution
- 27 institutes and centers
- Intramural and extramural research
- Main campus just north of Washington, DC
- 1800 active intramural clinical protocols
- 8100 terminated intramural clinical protocols
- 400,000 research subjects
- The Biomedical Translational Research Information System (BTRIS)
Where are the Clinical Research Data at NIH?

- Electronic Health Record
- Institute System
- Personal “System”
- Lab System

![Diagram showing the flow of clinical research data at NIH](image)
Where BTRIS Fits In

- BTRIS
- Electronic Health Record
- Lab System
- Institute System
- Personal “System”
Querying BTRIS for De-identified Data

- Version 0.0: home-grown system
- Policy-based access to de-identified data
- Ontology-based and text-based queries
- Summary reports available immediately
- Single-domain queries
- Detailed reports:
  - Require completion of OHSRP form
  - Download old data only (no permissions required)
### De-identified Queries, Version 0.0

Please use EITHER the "Search and Select" prompt OR the "tree" prompt to find or select Vital Sign Related Observation Name(s).

(Optional) Please search and select Vital Sign and Related Observation Names

- Arterial Blood Temperature Measurement CCN, (CC, SOFTLAB, TEMPA)
- Arterial Blood Temperature Measurement Point of Care (POCT), (CC, SOFTLAB, TEMPA)
- Arterial Blood Temperature Measurement, (CC, SOFTLAB, TEMPA)
- Arterial Blood Temperature Measurement Point of Care, (CC, SOFTLAB, TEMPA)

#### Results:

<table>
<thead>
<tr>
<th>Subject Count: <strong>2,978</strong></th>
<th>Record Count: <strong>25,873</strong></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Row Number</th>
<th>Observation_Name</th>
<th>Subject Count</th>
<th>Mean (Average)</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DTM Temperature (post)</td>
<td>40</td>
<td>36.85</td>
<td>0.41972635</td>
<td>36</td>
<td>37.7</td>
</tr>
<tr>
<td>2</td>
<td>DTM Temperature (pre)</td>
<td>88</td>
<td>36.519791</td>
<td>0.61842241</td>
<td>34.8</td>
<td>37.8</td>
</tr>
<tr>
<td>3</td>
<td>DTM Temperature2</td>
<td>77</td>
<td>36.317021</td>
<td>0.63747457</td>
<td>35</td>
<td>37.8</td>
</tr>
<tr>
<td>4</td>
<td>MRD_PNN_S Temperature</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>MRD_SCT Temperature</td>
<td>4</td>
<td>36.775</td>
<td>0.20615528</td>
<td>36.6</td>
<td>37</td>
</tr>
<tr>
<td>6</td>
<td>NURS CP Temperature</td>
<td>1</td>
<td>37.566666</td>
<td>0.05773503</td>
<td>37.5</td>
<td>37.6</td>
</tr>
<tr>
<td>7</td>
<td>NURS Dialysis Temperature</td>
<td>4</td>
<td>37.032203</td>
<td>0.9322647</td>
<td>35.3</td>
<td>39.6</td>
</tr>
<tr>
<td>8</td>
<td>NURS Post Procedure Temperature</td>
<td>79</td>
<td>36.930894</td>
<td>0.49574089</td>
<td>35.8</td>
<td>38.6</td>
</tr>
<tr>
<td>9</td>
<td>NURS Temperature</td>
<td>2,876</td>
<td>36.92438</td>
<td>1.9212647</td>
<td>0</td>
<td>99.9</td>
</tr>
<tr>
<td>10</td>
<td>NURS Temperature (NF)</td>
<td>2</td>
<td>37.966666</td>
<td>0.35118846</td>
<td>37.6</td>
<td>38.3</td>
</tr>
<tr>
<td>11</td>
<td>RESP Temperature</td>
<td>31</td>
<td>37.8</td>
<td>6.26569717</td>
<td>35</td>
<td>92</td>
</tr>
<tr>
<td>12</td>
<td>Temperature</td>
<td>23</td>
<td>37.154385</td>
<td>0.60417304</td>
<td>35.4</td>
<td>38.8</td>
</tr>
<tr>
<td>13</td>
<td>Temperature, Patient, Arterial, POCT</td>
<td>5</td>
<td>36.328571</td>
<td>0.73123799</td>
<td>35</td>
<td>37.6</td>
</tr>
<tr>
<td>14</td>
<td>Temperature, Patient, Venous, POCT</td>
<td>1</td>
<td>36.944444</td>
<td>0.05270463</td>
<td>36.9</td>
<td>37</td>
</tr>
<tr>
<td>15</td>
<td>Temperature, CCNMD</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Querying BTRIS for De-identified Data

- Version 1.0: Home-grown system
- Policy-based access to de-identified data
- Ontology-based and text-based queries
- Summary reports available immediately
- Select one or more data domains
- Complete the filter “widgets”
- Detailed reports:
  - Require completion of OHSRP form
  - Download instructions sent via e-mail
  - Original investigators notified of download
  - User must get permission for some data prior to use
Methods

• Development of Version 1.0
  – Based on user reactions to version 0.0
  – Chose an i2b2-like user interface design
  – Domains prioritized by ease, volume and popularity
  – Query features limited to allotted development time

• Collection of requirements
  – Based on user requests that were mediated by staff
  – Did not alter development of 1.0
To Start Querying for De-Identified Data,

Select Data Types on the Left Menu
Select Data Type

Diagnosis
Demographics
Medications
Lab Tests

Training & User Guides
Data Source
BTRIS Policies
Searching in BTRIS

BTRIS Customer Support
Submit Comment
M-F 8:30am to 5pm
301-827-8270
BTRISsupport@nih.gov

System Hours
M-F 8:30am to 5 pm
Maintenance and upgrades are scheduled for after hours and may impact system access and performance

Updated: Feb 21, 2013
Diagnosis

Training & User Guides
Data Source
BTRIS Policies
Searching in BTRIS

BTRIS Customer Support
Submit Comment
M-F 8:30am to 5pm
301-827-6270
BTRISsupport@nih.gov

System Hours
M-F 8:30am to 5 pm
Maintenance and upgrades are scheduled for after hours and may impact system access and performance

Updated: Feb 21, 2013
Coverage of Requirements by Version 1.0

- 30 user queries (see Table 2 in manuscript)

<table>
<thead>
<tr>
<th>Query #</th>
<th>Domains</th>
<th>Relationship</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$B, C, D, L, M$</td>
<td>AND</td>
<td>c, d, t, v</td>
</tr>
<tr>
<td>2</td>
<td>$C, L$</td>
<td>AND</td>
<td>d, n, u</td>
</tr>
<tr>
<td>3</td>
<td>$D, P$</td>
<td>AND</td>
<td>d, t, v</td>
</tr>
<tr>
<td>4</td>
<td>$C, L$</td>
<td>AND</td>
<td>d, t, v</td>
</tr>
<tr>
<td>5</td>
<td>$A, C, D, Dx, E, L, M, Mi, V$</td>
<td>AND</td>
<td>d, e, t, v</td>
</tr>
<tr>
<td>6</td>
<td>$D, Mi$</td>
<td>AND</td>
<td>t, v</td>
</tr>
<tr>
<td>7</td>
<td>$AE, D, L$</td>
<td>AND</td>
<td>v</td>
</tr>
</tbody>
</table>

**Bold items** are Version 1; **italic items** planned for future versions.

(A=Admission/Discharge/Transfer, AE=Adverse events, B=Blood Bank, C=Clinical Documents, D=Demographics, Dx=Diagnosis, E=Echo, L=Laboratory Tests, M=Medications, Mi=Microbiology, P=Pathology, R=Radiology, S=(Research) Study, V=Vital Signs)

Attributes: $a=$age, $c=$controlled terms, $d=$date range, $e=$expiration date, $g=$gender, $l=$location, $n=$normal ranges, $t=$text search, $u=$units of measure, $v=$discrete values
## Coverage of Requirements by Version 1.0

- 4 of 30 queries handled by Version 1
- 21 of 30 queries handled by later planned versions

<table>
<thead>
<tr>
<th>Function</th>
<th>Version 1</th>
<th>Later Versions</th>
<th>Not Planned</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domains</strong></td>
<td>Demographics (18), Diagnoses and Procedures (13), Medications (14), Laboratory Results (23)</td>
<td>Adverse Events (1), Alerts (0), Allergies (0), Blood Bank (1), Clinical Documents (14), ECG (0), Echo (1), Microbiology (3), Pathology (8), Pulmonary Functions (0), Radiology (2), Vital Signs (6)</td>
<td>Admission/Discharge/Transfer (1), Research Study (3)</td>
</tr>
<tr>
<td><strong>Features</strong></td>
<td>Age (2), Gender (1), Dates (11), Controlled Terminology (3), Value Range (28), Expiration Date (1)</td>
<td>Cardinality (0), Location (1), Normal Range (1), Text Search (21), Units of Measure (1), Discrete Values (28)</td>
<td></td>
</tr>
<tr>
<td><strong>Relationships</strong></td>
<td>AND (29), OR (3)</td>
<td>NOT (1), Before/After (4)</td>
<td></td>
</tr>
</tbody>
</table>
Next Steps for Versions 1.5 and 2.0

- Add data domains (vital signs, pathology)
- Add filters (text search, value range, cardinality)
- Add temporal relationships
- Text de-identification
- Warehouse appliance
- Streamline permission policies
- Explore access by outside investigators
Conclusions

✓ Current tool provides modest functionality

✓ Improved terminology user interface

✓ No technical obstacles to continued development

✗ Resources challenges remain

✗ Policy challenges remain

✓ Sharing (ontology, technology, data) is possible
Welcome to BTRIS

The Biomedical Translational Research Information System (BTRIS) is a resource available to the NIH intramural community that brings together clinical research data from the Clinical Center and other NIH Institutes and Centers. BTRIS provides clinical investigators with access to identifiable data for the subjects on their own active protocols, while providing all NIH investigators with access to de-identified data across all protocols. BTRIS provides users with advanced search, filtering, and aggregation methods to create data sets to support ongoing studies and stimulate ideas for new research. BTRIS Version 1.0 contains subject data from CRIS/MIS, NIAID (Crimson) and NIAAA. Future versions of BTRIS will contain additional historical and clinical data back to 1976, images, and subject data from other IC Systems such as NCI.

BTRIS is two applications:

BTRIS Data Access and BTRIS Preferences. Data Access allows users to query their active protocols with filters and criteria. Preferences is a Web based application that allows principal investigators or their designees to verify subject enrollment in their protocol(s). This ensures that reports created in BTRIS Data Access include all subjects. It also allows the principal investigator to designate an alternate investigator from the protocol to manage subject enrollment and create reports in BTRIS Data Access.

For questions or comments about BTRIS contact Dr. Jim Cimino, Chief, Laboratory for Informatics Development, NIH Clinical Center, National Institutes of Health, Bethesda, MD.