Importing UNC-CRDM Data Mart into an I2B2 Instance

Presenter: Xialan Dong
Advisor: Dr. Javed Mostafa
Mentor: Ashraf Farrag
Project Objectives

- Set up the I2B2 instance in a virtual machine
- Convert UNC CRDM data model to I2B2 star schema
- Explore/develop approaches to query UNC CRDM using I2B2 tools
- Provide insights into improving UNC patient cohort discovery portals
Project Design

- Explore the current literature on i2b2
- Install I2B2 instance and familiarize with the user interfaces
- Analyze the I2B2 database schema
- Analyze UNC CRDM data model
- Convert UNC CRDM data model to I2B2 schema
- Write Oracle sql scripts to import UNC CRDM data Mart into I2B2 Database
- Test /validation
Why I2B2?

- I2B2 = Informatics for Integrating Biology and the Bedside
- A open source database system funded by NIH developed by Harvard Medical School (National Center for Biomedical Computing based at Partners HealthCare System)
- An ontology based object-oriented database system with highly simple and flexible database schema
- Enables the user to build complex dynamic queries using an understandable user interface and a set of well-defined patient characteristics.
The i2b2 Hive: Take care of security, access rights and managing the underlying data repository

star schema: one fact table (a patient object) surrounded by four dimension tables (its attributes)

Dimension tables describe who (patient information), when (dates), what (ontology), and where (hospital and doctors)
Query and data mining tools
Allowing users to ask questions about the data
- Demographics->Gender-> Female (drag to panel 1)
- Diagnosis -> Respiratory System -> Chronic Obstructive Diseases -> Asthma
## UNC CRDM Dictionary

### CRDM_PROCEDURE [UNC] / HOA_PROCEDURE_STAGE_TABLE [SAS]
#### Data Domain:
Merged hospital-side and P&A-side billed procedures

<table>
<thead>
<tr>
<th>UNC Column</th>
<th>SAS Column</th>
<th>PK?</th>
<th>Definition</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENCOUNTER_ID</td>
<td>ENCOUNTER_ID</td>
<td>Y</td>
<td>Account number associated with an encounter</td>
<td>Linking field to all encounter-based tables</td>
</tr>
<tr>
<td>SERVICE_DATE</td>
<td>PROCEDURE_START_DTTM</td>
<td>Y</td>
<td>Date transaction occurred</td>
<td></td>
</tr>
<tr>
<td>PROCEDURE_CD</td>
<td>PROCEDURE_CD</td>
<td>Y</td>
<td>Code (either ICD-9 or CPT) for procedure</td>
<td></td>
</tr>
<tr>
<td>PROCEDURE_SCHEME</td>
<td>TERMINOLOGY_TYPE_CD</td>
<td>Y</td>
<td>Indicates whether procedure is coded ICD-9 or CPT</td>
<td></td>
</tr>
<tr>
<td>PROCEDURE_DESC</td>
<td>PROCEDURE_LONG_DESC</td>
<td></td>
<td>Description of procedure</td>
<td></td>
</tr>
<tr>
<td>PROVIDER_ID</td>
<td>PROCEDURE_BLG_PROVIDER_ID</td>
<td>Y</td>
<td>ID number for billing provider</td>
<td>Links to provider dictionary table</td>
</tr>
<tr>
<td>RESIDENT_ID</td>
<td>PROCEDURE_RESIDENT_ID</td>
<td>Y</td>
<td>ID number for resident, if a resident performed the transaction</td>
<td>Links to provider dictionary table</td>
</tr>
<tr>
<td>SOURCE_LOCATION</td>
<td>PROCEDURE_LOCATION</td>
<td></td>
<td>Location procedure was performed</td>
<td></td>
</tr>
<tr>
<td>SOURCE_SYSTEM_CD</td>
<td>PROCEDURE_SOURCE_SYSTEM_CD</td>
<td>Y</td>
<td>Indicates whether data came from PDS or ADS</td>
<td></td>
</tr>
</tbody>
</table>

### CRDM_ECHEART_ONC [UNC] / HOA_CHEMO_STAGE_TABLE [SAS]
#### Data Domain:
Chemo data from eChart

<table>
<thead>
<tr>
<th>UNC Column</th>
<th>SAS Column</th>
<th>PK?</th>
<th>Definition</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENCOUNTER_ID</td>
<td>ENCOUNTER_ID</td>
<td>Y</td>
<td>Account number associated with an encounter</td>
<td>Linking field to all encounter-based tables</td>
</tr>
<tr>
<td>FORM_NAME</td>
<td>FORM_NAME</td>
<td>Y</td>
<td>Name of the eChart form where the data were entered</td>
<td></td>
</tr>
<tr>
<td>OBS_NAME</td>
<td>CHEMO_OBSV_NAME</td>
<td>Y</td>
<td>Name of the field on the eChart form</td>
<td></td>
</tr>
<tr>
<td>OBS_NAME_TEXT</td>
<td>CHEMO_OBSV_RAW_CHAR_VAL</td>
<td>Y</td>
<td>Text value entered in the field on the eChart form</td>
<td></td>
</tr>
<tr>
<td>OBS_NAME_NUM</td>
<td>CHEMO_OBSV_RAW_NUM_VAL</td>
<td></td>
<td>Numeric value entered in the field on the eChart form (often null)</td>
<td></td>
</tr>
<tr>
<td>CREATE_TS</td>
<td>CHEMO_OBSV_CREATE_DTTM</td>
<td>Y</td>
<td>Datetime when value was first entered</td>
<td></td>
</tr>
<tr>
<td>UPDATE_TS</td>
<td>CHEMO_OBSV_UPDATE_DTTM</td>
<td></td>
<td>Datetime when value was last updated</td>
<td></td>
</tr>
</tbody>
</table>

- 19 data tables
- Assume that the ontology is the same as that in I2B2 (so ontology mapping is not dealt in this project)
The rest of the 19 UNC tables also be mapped to Observation_fact, although no significance in query for now.
Some unfitted fields are mapped to the blob field, so no field in UNC tables will be stripped off.

This mapping sheet are the basis of data converting algorithm.
How to Take Advantage of I2B2?

Use I2B2

Import Database into I2B2

- Use I2B2 Import Data View

I2B2 Front End Access to UNC Database

- Build a New Import Tool
- Use UNC Current Query Engine
- Use FURTHEr Query Engine

- Challenge 1: too many new tools, new concepts
- Challenge 2: No clinic experience, no data sample from UNC
- All the possibilities are explored in this project
Flow of Importing Data from UNC CRDM to I2B2

- Create User Account: UNCCRDM
- Create Tables (for i2b2 and CRDM)
- Load Data to CRDM Tables
- Move Data from CRDM Tables to I2B2's
- Replace Tables in CRC with the new i2b2 tables

- Sql script files used to create tables are from Ashraf and I2B2 documentation.
- 8 SQL script files are built and used to facilitate the importing process.
Against some made up data, the presented importing approach we designed is proved to be reliable and feasible, although it may need some modifications and ontology mapping when Ashraf test against the real data.

The presented approach will also pave a way for us to explore integrating I2B2 front end, FURTHeR’s query engine, and UNC data warehouse.

We may need to build an I2B2 instance from scratch if the i2b2 demo suit will not provide a way for us to replace the i2b2 demodata with our own.
1. http://www.i2b2.org/software/(2012 Apr 16)
I thank Dr. Mostafa for his insightful guidance and constant encouragement, so I can finish this project during my grueling schedule with heavy coursework and a full time job.

I also take this opportunity to express my deep gratitude to Ashraf, for his cordial support and valuable guidance, which helped me to embrace the mind stretching challenges during the process.

I am obliged to Heather for her unconditional supports both professionally and personally.