

i2b2 – what it is, and what it is not

Keith Marsolo, PhD
Associate Professor
Division of Biomedical Informatics
March 9, 2014

i2b2 – what it is

- Acronym: Informatics for Integrating Biology & the Bedside
- Data model
 - Way to structure observations (facts) about patients
- Integration platform
 - Combine disparate observations into a common data structure
 - Typical examples
 - Diagnoses
 - Demographics
 - Lab results
 - Procedures
 - Genetic information

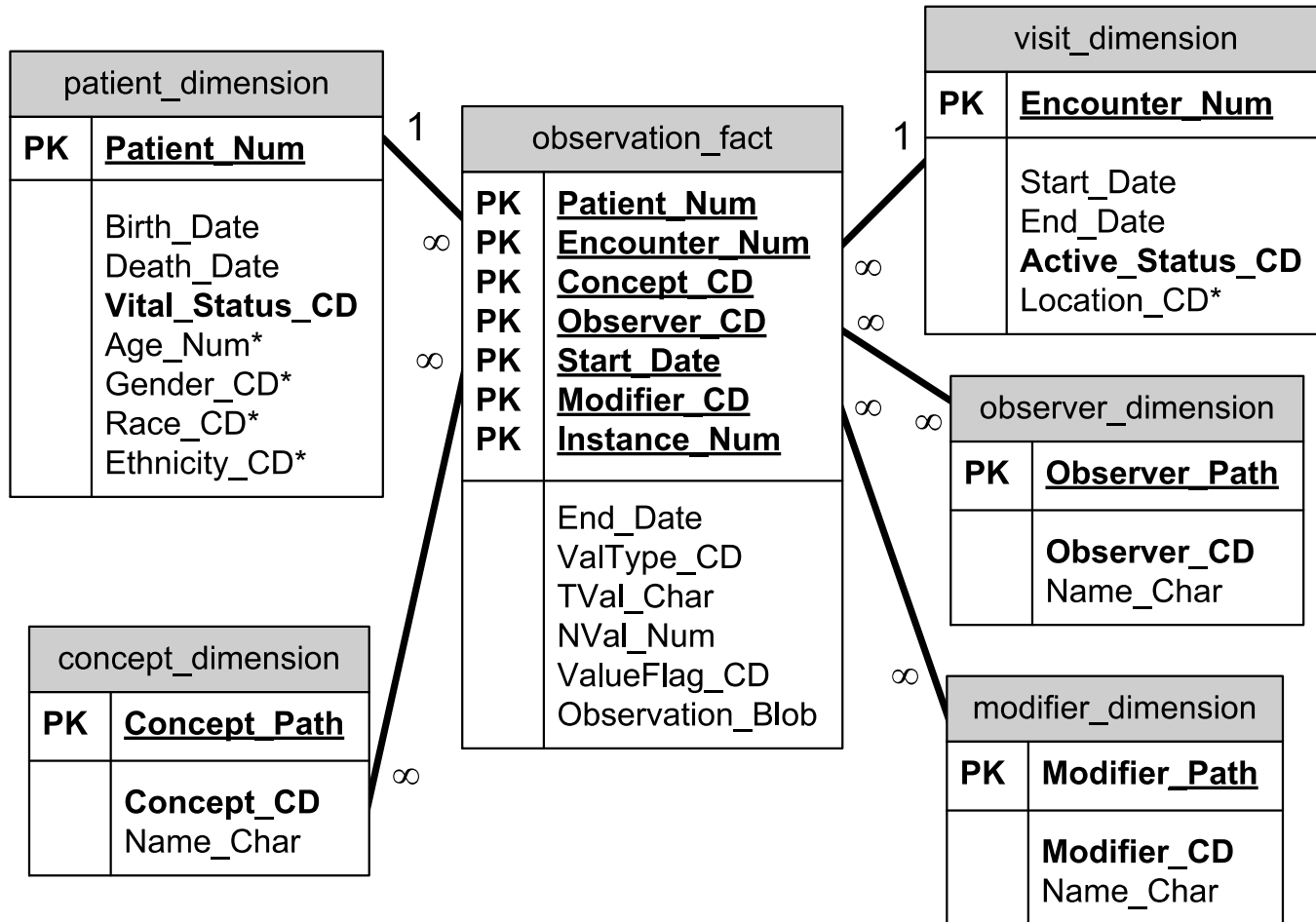
i2b2 – what it is (2)

- Query tool
 - Cohort identification and study feasibility
 - Structured using Boolean algebra
 - Counts of based based on inclusion/exclusion criteria
 - YOU MUST KNOW WHAT YOU'RE DOING!!!!
- Based on Research Patient Data Registry developed at Massachusetts General Hospital (MGH)

Representing observations in i2b2

```
P_num|E_num|Instance_num|C_CD|S_d|Modifier_CD|ValType_CD|TVal|NVal
-----
123|107|1|cpt:59622|20060304|@|<null>|<null>|<null>
123|107|1|cpt:59622|20060304|cptmod:62|<null>|<null>|<null>
123|107|1|cpt:59622|20060304|cptmod:AA|<null>|<null>|<null>
123|107|1|cpt:59622|20060304|cptmod:TH|<null>|<null>|<null>
123|567|1|med:aspirin|20100404|@|<null>|<null>|<null>
123|567|1|med:aspirin|20100404|MED:DOSE|N|E|325
123|567|1|med:aspirin|20100404|MED:FREQ|T|QD|<null>
123|567|1|med:aspirin|20100404|MED:ROUTE|T|PO|<null>
123|567|2|med:aspirin|20100404|@|<null>|<null>|<null>
123|567|2|med:aspirin|20100404|MED:DOSE|N|E|83
123|567|2|med:aspirin|20100404|MED:FREQ|T|BID|<null>
123|567|2|med:aspirin|20100404|MED:ROUTE|T|PO|<null>
```

i2b2 data model



i2b2 query tool

The screenshot displays the i2b2 Query Tool interface, which is used for constructing and executing queries on clinical data. The interface is divided into several main sections:

- Ontology:** A tree view on the left showing the CCHMC ontology structure, including categories like Allergy, Demographics, Diagnoses, Flowsheet, GWADB, Laboratory Tests, Medications, Procedure, and Samples.
- Query Tool:** The central workspace for building queries. It features three groups (Group 1, Group 2, and Group 3) for defining criteria. Each group has options for 'Dates', 'Occurs > [value]', and 'Exclude'. Group 1 is set to 'Occurs > 2', Group 2 to 'Occurs > 0', and Group 3 to 'Occurs > 0'. The criteria are: Group 1: 'Fragile X Syndrome(759.83)'; Group 2: 'Antipsychotics'; Group 3: (empty). The groups are connected by 'AND' operators, and each group contains an 'OR' operator.
- Run Query:** A button to execute the query.
- Getting Started Guide:** A section with tabs for 'Query Verbalizer', 'Query Results', and 'Patient/Visits'. The 'Query Verbalizer' tab is active, showing the generated SQL query.
- Previous Queries:** A section at the bottom left showing a list of saved queries, including one from 'Mar 09, 2014' titled 'Frag X-Anti-psy'.

The generated SQL query for the current configuration is as follows:

```
Patients matching ANY of the following criterion:  
HAVING Diagnoses \ Congenital Anomalies \ Other And Unspecified Congenital Anomalies \ Fragile X Syndrome(759.83)  
WHERE 'Date of Event' after 01/01/2011 AND before 12/31/2013  
WHERE criteria Occurs greater than 2  
  
AND  
Patients matching ANY of the following criterion:  
HAVING Medications \ Central Nervous System Drugs \ Antipsychotics
```

i2b2 – what it is not

- Magic
- Why not?
 - Not sure, but often criticized for it

i2b2 – what it is not (2)

- A giant spreadsheet
- Why not?
 - Data dimensionality in the age of the EHR
 - 200K+ potential diagnoses
 - 3K+ different lab tests
 - 60K allergens
 - 60K flowsheets
 -
 - Example – extract of 4K patients
 - 280K diagnoses
 - 1.5 million lab results
 - 130K vitals

i2b2 – what it is not (3)

- An end-to-end analysis tool
- A data discovery tool
- Why not?
 - i2b2 was designed to count
 - Data structure can be suboptimal for some analyses
 - “Long & skinny” structure can be confusing
 - i2b2 can be used to represent varying data
 - Designing generalized tools is difficult
 - Often easier to create extracts for use in other programs

i2b2 – what it is not (4)

- A replacement for understanding clinical workflows
- Why not?
 - Specialty clinics often have customized EHR builds
 - Allows for collection of condition-specific data elements
 - However: workflow used to collect data dictates where those data are stored
 - Can be almost impossible for non-clinicians to identify the most relevant data elements

Epic data model (reporting – partial)

- <image removed>

i2b2 – what it is not (5)

- Smart
- Why not?
 - The query tool is dumb
 - It does exactly what you ask (which is not necessarily what you think)
 - Example: Find all asthma patients
 - Usual query: Number of patients with ICD-9 code 493.X
 - Other alternatives:
 - Patients with visit to the asthma clinic
 - Patients on clinical asthma registry
 - Patients with order for albuterol

i2b2 is one tool in a suite of offerings

- <image removed>

How is i2b2 being used currently?

- De-identified cohort queries
 - Generate initial counts, move on to mediated query
 - <https://i2b2.cchmc.org> (Access tab)
- Mediated data requests
 - De-identified or identified
- Federated network of warehouses / distributed data sharing
 - Broadcast query to several sites, receive aggregated counts
- Multi-center quality improvement / research registries
 - Centers can query own data (patient-level), or data of entire network (aggregate counts)
- Quality assurance of genetic lab results (not shown)

De-identified i2b2 – Data

- Current (or pending)
 - Demographics
 - Diagnoses
 - Medication orders
 - Procedures
 - Lab results
 - Visit location
 - Surgical History
 - Flowsheets
 - Samples / metadata
 - Existence of notes
- Coming soon
 - De-identified notes
 - Outcomes
 - Registry populations
 - Derived phenotypes

Federated networks

The screenshot displays a web-based query tool interface for a federated network. It is divided into several sections:

- Navigate Terms / Find Terms:** A hierarchical tree view of medical terms. The 'Endocrine, nutritional, and metabolic diseases and immunity disorders' category is expanded, showing sub-terms like 'Diabetes mellitus with complications' and 'Diabetes mellitus without complication'.
- Query Tool:** A central area for building queries. It shows a 'Query Name: Test' and three groups:
 - Group 1:** Contains the term 'Male'.
 - Group 2:** Contains the term 'Diabetes mellitus with com'.
 - Group 3:** Is currently empty.
 The groups are connected by 'AND' operators. A yellow box in Group 3 says 'drag a term to here'.
- Run Query / New Query / Print Query:** Buttons at the bottom of the query tool. The status '2 Groups' is displayed.
- Previous Queries:** A list of recent queries, such as 'type I -Female@9/28/12 [9-28-2012] [shrine]' and 'Diabetes - Female [8-31-2012] [shrine]'.
- Query Status:** A section titled 'Finished Query: "Test"' showing results from various sites:

Site	Patients	Status	Time
Cincinnati	713 ±3	FINISHED	[171.0 secs]
Partners HealthCare	19081 ±3	FINISHED	[53.5 secs]
Wake Forest	6833 ±3	FINISHED	[31.0 secs]
UTHealth	8856 ±3	FINISHED	[41.4 secs]
UCSF	883 ±3	FINISHED	[90.0 secs]
ITHS	14327 ±3	FINISHED	[118.7 secs]

Data and participating sites contingent on project specifics

Future plans

- Participation in national data sharing networks
 - PCORI National patient-centered clinical research network (PCORnet)
- Summary statistics on warehouse/extracts
- Upgrade to version 1.7
 - Keyword search
 - Temporal queries
- “Chart review” workflow / revised UI

Questions?

- Funding:
 - R01HS020024
 - R01DK076893
 - UL1RR025758
 - UL1RR026314
 - RC2AR058934
 - U01HG006828
 - R24HL105333