Experience with a Clinical Data Repository and Warehouse

Adam Wilcox, PhD
Columbia University
March 24, 2009
Outline

- History
- Clinical Data Repository
- Clinical Data Warehouse
Clinical Information Systems

- Stage 1: Early computers calculated data in context
- Stage 2: Client applications provided access to ancillary data
- Stage 3: Systems began aggregating data from multiple sources
- Stage 4: Data storage provided historical view
  - And analysis
- Stage 5: Workflow applications formalize processes between clinical roles
Clinical Information System Technology Levels

- Level 1: Departmental applications
- Level 2: Internally-developed integrated systems
- Level 3: Functional vendor-based systems
- Level 4: Comprehensive clinical information systems
Clinical Information Systems at Columbia University

- Began at Stage 3
- Pushing a Level 1 system to Level 2
- Issues
  - Vocabulary
  - Data modeling
  - Interfaces
  - Decision support
  - Data processing
- Recipient of first Nicholas Davies Award
EMR environment

Production Databases

Enterprise Repository/Data Warehouse

Workgroup Datamarts

Query

Replicate

Distribute

Query

Workflow or Goal specific

Datamarts, Personal, Mobile, Research, Educ.
Architecture

Information Service Layers

- Handling
- Encoding
- Routing
- Monitoring
- Access
Other Level 2 Systems

- Intermountain
- VA
- Partners
- Regenstrief
- Vanderbilt
Level 3 Systems

- Cerner
- Epic
- Eclipsys
- GE
- McKesson
Challenges at Columbia

- Moved from Stage 3 through Stage 4 to Stage 5
- Purchased a vendor system (Level 3)
- How to get to Stage 5 and Level 4?
Challenges at CPMC/CUMC/NYPH/WCMC

- In 1998, merged two academic medical centers into NewYork Presbyterian Hospital
  - Columbia Presbyterian campus became Columbia University Medical Center
  - New York Hospital became Weill Cornell Medical Center
- Currently 4 different electronic health records
  - Eclipsys (WCMC)
  - Eclipsys (CUMC)
  - Epic (WCMC)
  - Allscripts (CUMC)
Integrating Among Multiple EHRs

- Eclipsys (CUMC)
- Eclipsys (WCMC)
- Allscripts
- Epic
Problems with Integrating to Application Databases

- Must model each system multiple times
  - Increased effort and complexity
- Overloading workflow databases
- Protecting external data consistency (no updates)
- Increased complexity of data protection
- Bringing in data for a new patient
  - When to pull data in
  - Interfaces don’t naturally pull in historical data
- Increases complexity as move toward RHIOs
Repository Model

- Eclipsys (CUMC)
- Eclipsys (WCMC)
- Allscripts
- Epic

Clinical Data Repository
Benefits of CDR

- Only model data from source systems once
- Common data store
- Data are read only
  - Optimized for read
- Historical data included
- Web-based viewer adaptable to multiple applications
- Adaptable to future health information exchange efforts
- Platform of innovation
Optimized for Retrieval

- Relational structure can be difficult to query for both data and context
  - Gathering multiple elements requires multiple table joins
  - Good for data storage
  - Good for aggregating across multiple patients
- Event-based model good for querying across data types
  - Data organized according to patient
  - Not good for querying across patients
Retrieval optimization

- Paradigm shift in how data are used
  - Paper records mainly for primary use
  - Electronic allows secondary use
  - Secondary use can be multiple times
CDR View in Eclipsys

**Patient Information**
- Name: Josephina, Kathy Ann
- DOB: 01/01/1933
- Admit Date: 22-Jan-2009
- Gender: Female
- Current Weight: kg
- Height: cm
- US: sq m

**Disch Sum (2010-03-18 to 2004-11-05)**

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2010-03-18 12:00</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>2005-10-31 09:55</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>2005-10-31 10:05</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>2005-10-31 10:05</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>2005-10-31 08:59</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>2005-10-31 08:56</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>2005-10-31 08:53</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>2005-10-31 08:52</td>
<td>F</td>
</tr>
</tbody>
</table>

**History of Present Illness:**
The patient is a 45-year-old woman with a history of heart failure and end-stage heart disease. She was evaluated for cardiac transplantation. She has had shortness of breath with her ACE-I and sulfa antibiotics, which have had little benefit. She has been treated with digoxin, furosemide, and nitroglycerin. She has been on a low-sodium diet and has been taught to LIMIT fluids.

**Past Medical History:**

**Allergies:**
None.

**Medications:**
- Norepinephrine 1:1000
- Dopamine 1:1000
- Doxapram 1:1000
- Nitroglycerin 1:1000
- Furosemide 1:1000
- Digoxin 1:1000

**Family History:**
Onset of heart failure at age 50. Her mother had heart failure at age 60. Her father had diabetes and heart disease.

**Social History:**
Non-smoker, non-drinker, does not use alcohol.

**Physical Examination:**
- Blood pressure: 120/80
- Respiratory rate: 18
- Temperature: 98.6°F
- Pulse: 80
- Oxygen saturation: 98%
- GCS: 15
- Cardiac: S1, S2, S3
- Chest: Clear to auscultation
- Abnormalities: None noted.

**Diagnosis:**
- Heart failure with reduced ejection fraction

**Plan:**
- Continue medical therapy
- Refer to cardiologist for further management

**Follow-up:**
- Cardiology appointment scheduled for 1 week.
Proportion of CDR Viewer Access

WebCIS tab per XA use
Increase in CDR View Access
CUMC/NYP Clinical Data Warehouse History

- **1994**: Created, sponsored by Columbia University Department of Medical Informatics and Office of Clinical Trials
  - Populated with data from existing clinical data repository
  - Supporting clinical research
- **1998**: Columbia + Cornell = NewYork Presbyterian Hospital
  - Warehouse funded by NYPH
  - Goal to incorporate and provide data across whole system
- **2004**: Formal analysis of CDW user needs by Clinical Quality and Information Technology Committee (CQIT)
  - Creation of Data Warehousing Subgroup
  - Need to bring together disparate clinical data sources
  - Need to manage user requests for data
Uses of the Warehouse

- Clinical research queries
- Management reports
- Clinical trial recruitment
CDW Content Issues

- Began as a copy of the repository
  - Data already gathered
- Mainly for research queries
  - Some data marts built for common queries
- Ability to query rapidly across patients increases security risk
Clinical Data Warehouse at CUMC

Mediated Query

Please see Important Riders, Approval Requirement and a Fee Policy at the bottom of this page.

NAME
DEPARTMENT
TITLE/RANK
ROLE IN PROJECT
OFFICE ADDRESS
TELEPHONE
EMAIL

PROTOCOL (PROJECT) NAME AND A BRIEF DESCRIPTION

PROJECT SPONSORSHIP/REASON
Investigator Initiate

PRINCIPLE INVESTIGATOR

DATE OF REQUEST
2/7/2007

DATE REPORT IS REQUIRED
12/31/2005
<table>
<thead>
<tr>
<th>RIDERS ON PATIENT DATA</th>
<th>APPROVALS</th>
<th>FEE POLICY</th>
</tr>
</thead>
<tbody>
<tr>
<td>• I understand that this data is confidential and I will follow strict procedures to</td>
<td>• RESEARCH PROJECTS AND PUBLICATIONS Approval of the CUMC Institutional</td>
<td>• Please note that a chargeback methodology is being developed for research-</td>
</tr>
<tr>
<td>preserve confidentiality in dealing with patient-specific information. In order to</td>
<td>Review Board (IRB) is required.</td>
<td>related requests for data from the Clinical Data Warehouse. An important</td>
</tr>
<tr>
<td>protect the privacy rights of individuals as well as the interests of Columbia</td>
<td>• CLINICIAN ACCESSING DATA ONLY ON HIS OR HER OWN PATIENT’S</td>
<td>factor in this methodology will be the number of hours of analysis work</td>
</tr>
<tr>
<td>University and New York Presbyterian Hospital.</td>
<td>A signed letter on the clinician’s letterhead attesting to this and</td>
<td>involved. We’ll estimate and notify you total cost once your Clinical Data</td>
</tr>
<tr>
<td>• I further understand that this data is for my own use and that of my direct</td>
<td>specifying the clinician’s information.</td>
<td>Warehouse request is approved.</td>
</tr>
<tr>
<td>close collaborators only, and I agree not to release or distribute this information,</td>
<td>• DEPARTMENTAL REVIEW OR GRAND ROUNDS: A signed letter of approval</td>
<td></td>
</tr>
<tr>
<td>in any form, to any less closely affiliated person or organization, regardless of</td>
<td>from the sponsoring attending physician specifying the information desired.</td>
<td></td>
</tr>
<tr>
<td>institutional or organizational affiliation.</td>
<td>• SUMMARY INFORMATION: No formal letter or specification is required,</td>
<td></td>
</tr>
<tr>
<td>• I understand all requests for data will be reviewed by the Office of Clinical Trials</td>
<td>unless the request becomes resource-intensive or requires extensive analysis.</td>
<td></td>
</tr>
<tr>
<td>and final approval for a response to my request rests solely with the Office.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• I agree that the acquired data will be destroyed once it is no longer required.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Goal of Access Policy

- Provide broader access to data
  - Central control is resource limited
- Allow collection of more data sources
  - Reassure data stewards
    - Three separate institutions
    - Data ownership not completely defined for all data
CDW Structure

- Identifying data
  - Patient identifying information
- Main data
  - Event tables for clinical repository
- Lookup tables
  - Vocabulary translation
  - Contains no patient data
- Specialty data marts
Access Policy

- Identifying data
  - Most restricted
  - Create a research identifier to replace the patient ID
  - Allow access to only ResearchID, sex, birth date (month and year only), marital status, race, death status

- Specialty data
  - Access policy defined by data steward

- Patient clinical data
  - No access to text data
  - Modified dates

- Lookup tables
  - Full access (contain no patient data)
Access Policy

- Specific patient information
  - Sometimes needed to create initial queries
  - Analysts get access only to a randomly selected subset
  - Access request through supervisor
- De-identified patient data
  - Test patients
  - Full access given
CUMC/NYP Clinical Data Warehouse History

- 1994: Created, sponsored by Columbia University Department of Medical Informatics and Office of Clinical Trials
  - Populated with data from existing clinical data repository
  - Supporting clinical research
- 1998: Columbia + Cornell = NewYork Presbyterian Hospital
  - Warehouse funded by NYPH
  - Goal to incorporate and provide data across whole system
- 2004: Formal analysis of CDW user needs by Clinical Quality and Information Technology Committee (CQIT)
  - Creation of Data Warehousing Subgroup
  - Need to bring together disparate clinical data sources
  - Need to manage user requests for data
Analysis of Challenges

- Data in vendor-based transactional systems
- Could not query across transactional systems
- Users needed help in defining their needs
- Mature initiatives required more robust data solutions
Pneumonia Core Measures

CUMC Influenza Vaccinations Q4 2008 - Q1 2009

Week of Year

Influenza Vax Ordered
Influenza Vax Administered
INTEGRATION SERVICES

SOURCE Systems

SOURCE Databases

REPLICATED Databases

AD HOC Complex Analytical Queries

RECURRING Clinical Care Reporting & Business Analysis & Research

Online Analytical Processing (OLAP)

VIRTUAL CLINICAL DATA WAREHOUSE

A

B

C

DM

DM

DM

OLAP

Reports

A

B

C

DATAMARTS

Research
<table>
<thead>
<tr>
<th>Goal</th>
<th>Task</th>
<th>Use</th>
<th>User</th>
<th>Tool</th>
<th>Six Sigma</th>
<th>Cost/Instance</th>
<th>Instances</th>
<th>Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer a specific question</td>
<td>Ad hoc query</td>
<td>Research</td>
<td>Researcher</td>
<td>SQL</td>
<td>Define</td>
<td></td>
<td>Defined request</td>
<td></td>
</tr>
<tr>
<td>Observe trends</td>
<td>Recurring query</td>
<td>Management reports</td>
<td>Manager</td>
<td>Reporting application</td>
<td>Measure</td>
<td></td>
<td>Available owner</td>
<td></td>
</tr>
<tr>
<td>Identify dependencies</td>
<td>OLAP</td>
<td>Operational analysis</td>
<td>Analyst</td>
<td>Analytics / Data cubes</td>
<td>Analyze</td>
<td></td>
<td>Content expert/ analyst</td>
<td></td>
</tr>
<tr>
<td>Assist decision making</td>
<td>Dashboard display</td>
<td>Point of care</td>
<td>Clinical team</td>
<td>Registries</td>
<td>Improve</td>
<td></td>
<td>Pilot site</td>
<td></td>
</tr>
<tr>
<td>Automate processes</td>
<td>Application</td>
<td>Decision support</td>
<td>Clinician/ Role</td>
<td>EMR application</td>
<td>Control</td>
<td></td>
<td>Institutional sponsor</td>
<td></td>
</tr>
</tbody>
</table>
INTEGRATION SERVICES

VIRTUAL DATA WAREHOUSE

REPLICATED Databases

INTEGRATION SERVICES

DATAMARTS

DATA WAREHOUSE TOOLS

Ad-Hoc Queries – Questions

Research

Define

Recurring – Automated Queries

Management Reports

Measure

OLAP – Analytics

Operational Reports

Analyze

Dashboards

Point of Care Reporting

Improve

Applications

Decision Support

Control

Black Belt Six Sigma Approach
Conclusion

- Integrating clinical data repository view into workflow applications can improve use
- Access policies need to isolate data to reassure data use from different stakeholders
- Data access tools need to account for users’ evolving data needs along the quality improvement life cycle