**Slide 1**

**Experience with a Clinical Data Repository and Warehouse**

Adam Wilcox, PhD
Columbia University
March 24, 2009

**Slide 2**

**Outline**

History
Clinical Data Repository
Clinical Data Warehouse

**Slide 3**

**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
ation Systems

**Slide 4**

**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

**Slide 5**

**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

**Slide 6**

**Image:** Clinical Information Systems Architecture.

**Slide 7**

**Image:** EMR environment

**Slide 8**
**Image:** Architecture. Handling, Encoding, Routing, Monitoring, Access

**Slide 9**
**Other Level 2 Systems**
Intermountain
VA
Partners
Regenstrief
Vanderbilt

**Slide 10**
**Level 3 Systems**
Cerner
Epic
Eclipsys
GE
McKesson

**Slide 11**
**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?

**Slide 12**
**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)

**Slide 13**
**Image:** NYP Computing, Environment Sep 2007

**Slide 14**
**Image:** Integrating Among Multiple EHRs
Eclipsys (CUMC)
Eclipsys (WCMC)
Allscripts
Epic

**Slide 15**

**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

**Slide 16**

**Image:** Repository Model
Eclipsys (CUMC)
Eclipsys (WCMC)
Clinical Data Repository
Allscripts
Epic

**Slide 17**

**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

**Slide 18**

**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

**Slide 19**

**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

**Slide 20**
**Image:** CDR View in Eclipsys

**Slide 21**
**Graph:** Proportion of CDR Viewer Access

**Slide 22**
**Graph:** Increase in CDR View Access

**Slide 23**
**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

**Slide 24**
**Uses of the Warehouse**
Clinical research queries
Management reports
Clinical trial recruitment

**Slide 25**
**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

**Slide 26**
**Screen Grab**

**Slide 27**
**Screen Grab**

**Slide 28**

**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

**Slide 29**

**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

**Slide 30**

**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)

**Slide 31**

**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

**Slide 32**

**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

**Slide 33**

**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

**Slide 34**

**Graph:** Pneumonia Core Measures

**Slide 35**

**Graph:** Pneumonia Core Measures

**Slide 36**

**Image:** VIRTUAL CLINICAL DATA WAREHOUSE

**Slide 37**

**Graph**

**Slide 38**

**Image:** VIRTUAL DATA WAREHOUSE

**Slide 39**

**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