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

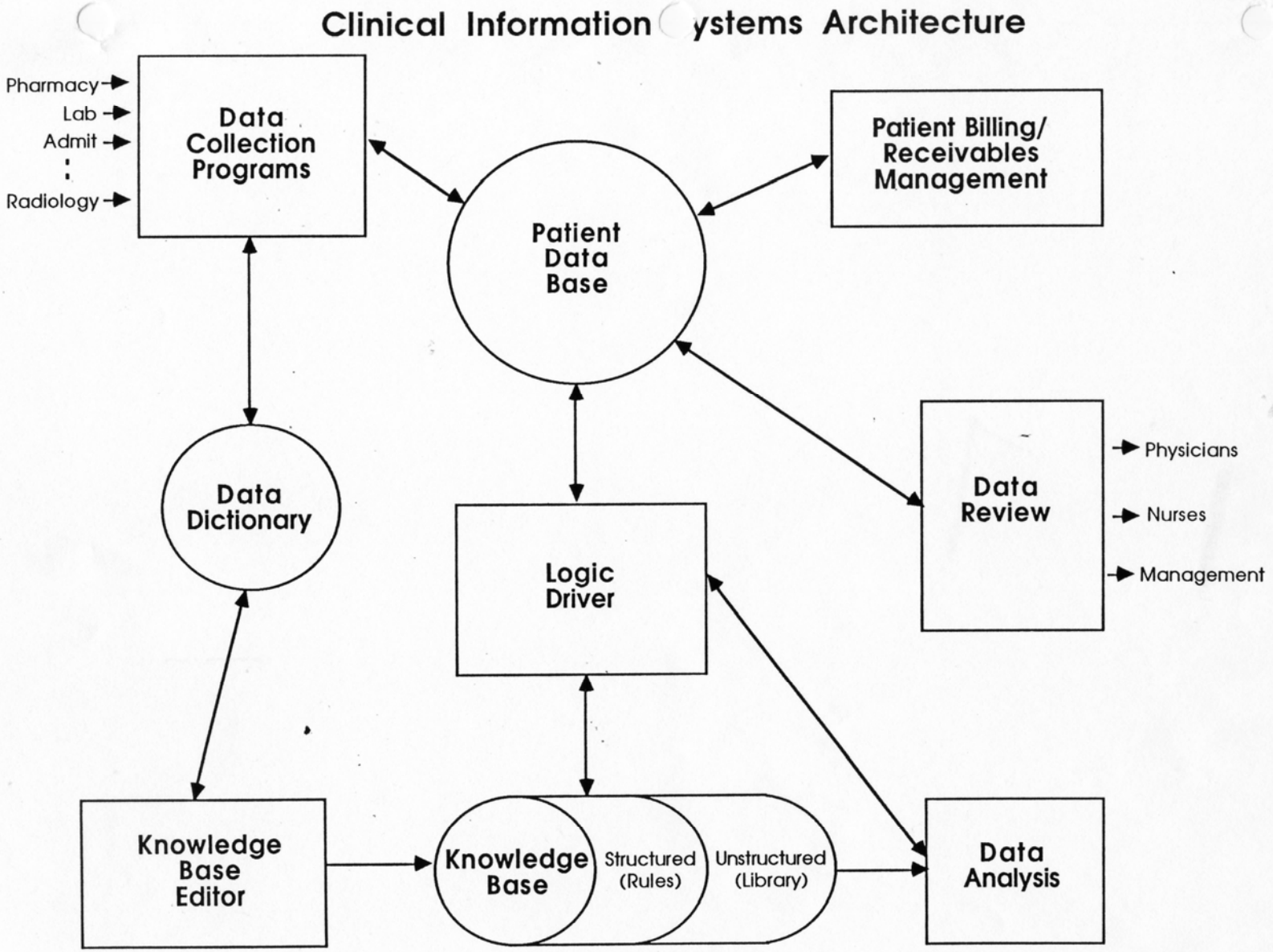
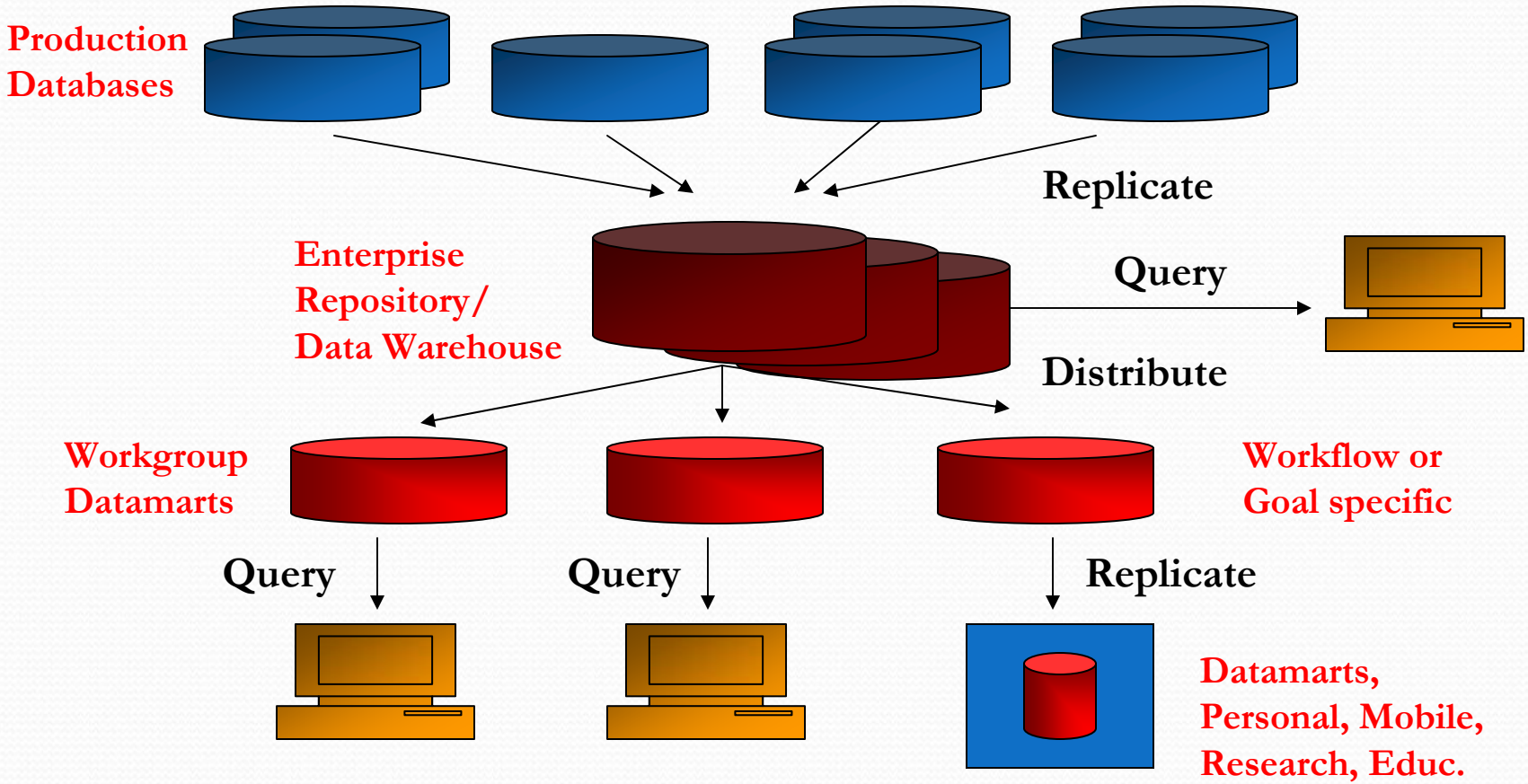
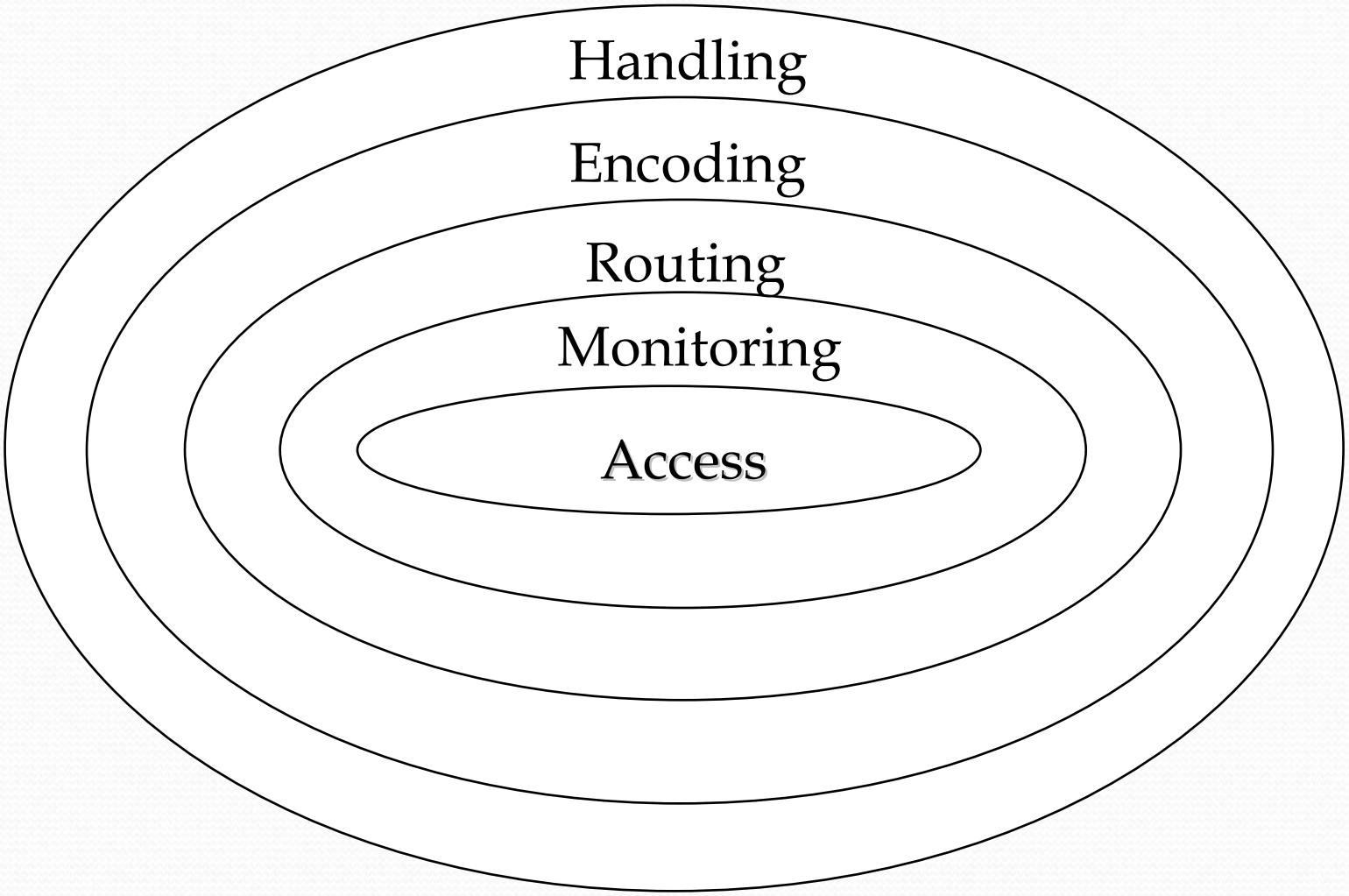

Clinical Information ystems Architecture
Pharmacy
Lab
Admit
Radiology
Data Collection Programs
Patient Billing/ Receivables Management
Patient Data Base
Data Dictionary
Logic Driver
Data Review
Physicians
Nurses
Management
Knowledge Base Editor
Knowledge Base
Structured (Rules)
Unstructured (Library)
Data Analysis

FIGURE B.1

EMR environment



Architecture



Other Level 2 Systems

- Intermountain
- VA
- Partners
- Regenstrief
- Vanderbilt

Level 3 Systems

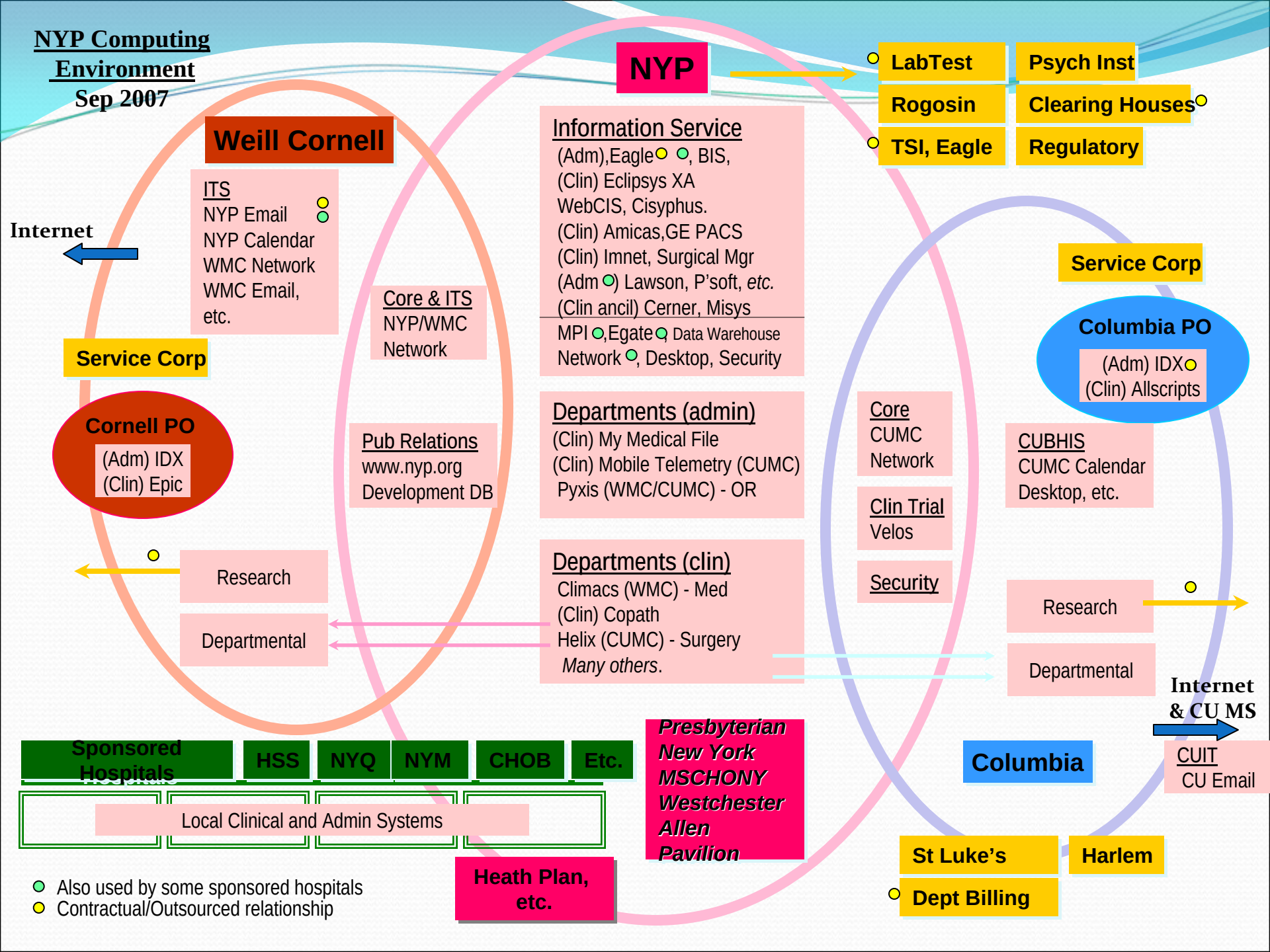
- Cerner
- Epic
- Eclipsys
- GE
- McKesson

Challenges at Columbia

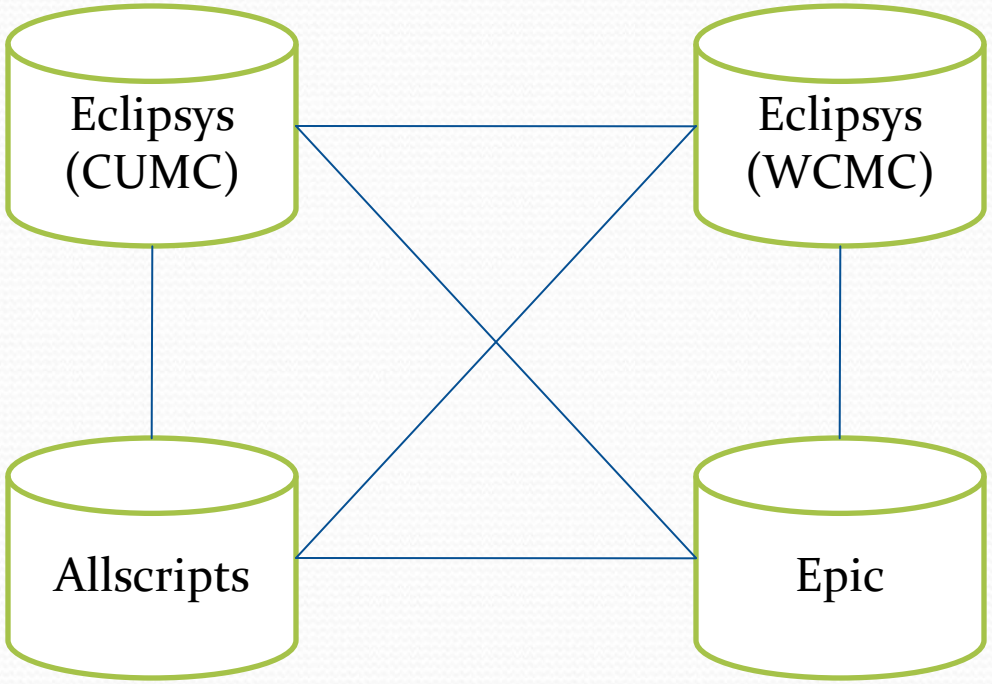
- Moved from Stage 3 through Stage 4 to Stage 5
- Purchased a vendor system (Level 3)
- How to get to Stage 5 and Level 4?

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)


NYP Computing Environment
Sep 2007
NYP
LabTest
Psych Inst
Rogosin
Clearing Houses
TSI, Eagle
Regulatory
Weill Cornell
ITS
NYP Email
NYP Calendar
WMC Network
WMC Email, etc.
Internet
Information Service
(Adm),Eagle, BIS,
(Clin) Eclipsys XA
WebCIS, Cisyphus.
(Clin) Amicas,GE PACS
(Clin) Imnet, Surgical Mgr
(Adm) Lawson, P'soft, etc.
(Clin ancil) Cerner, Misys
MPI,Egate Data Warehouse
Network, Desktop, Security
Service Corp
Columbia PO
(Adm) IDX
(Clin) Allscripts
Core & ITS
NYP/WMC
Network
Service Corp
Cornell PO
(Adm) IDX
(Clin) Epic
Pub Relations
www.nyp.org
Development DB
Departments (admin)
(Clin) My Medical File
(Clin) Mobile Telemetry (CUMC)
Pyxis (WMC/CUMC) - OR
Core
CUMC
Network
CUBHIS
CUMC Calendar
Desktop, etc.
Clin Trial
Velos
Security
Research
Departmental
Departments (clin)
Climacs (WMC) - Med
(Clin) Copath
Helix (CUMC) - Surgery
Many others.
Research
Departmental
Internet & CU MS
Sponsored Hospitals
HSS
NYQ
NYM
CHOB
Etc.
Local Clinical and Admin Systems
Presbyterian
New York
MSCHONY
Westchester
Allen
Pavilion
Columbia
CUIT
CU Email
St Luke's
Harlem
Dept Billing
Heath Plan, etc.
Also used by some sponsored hospitals
Contractual/Outsourced relationship

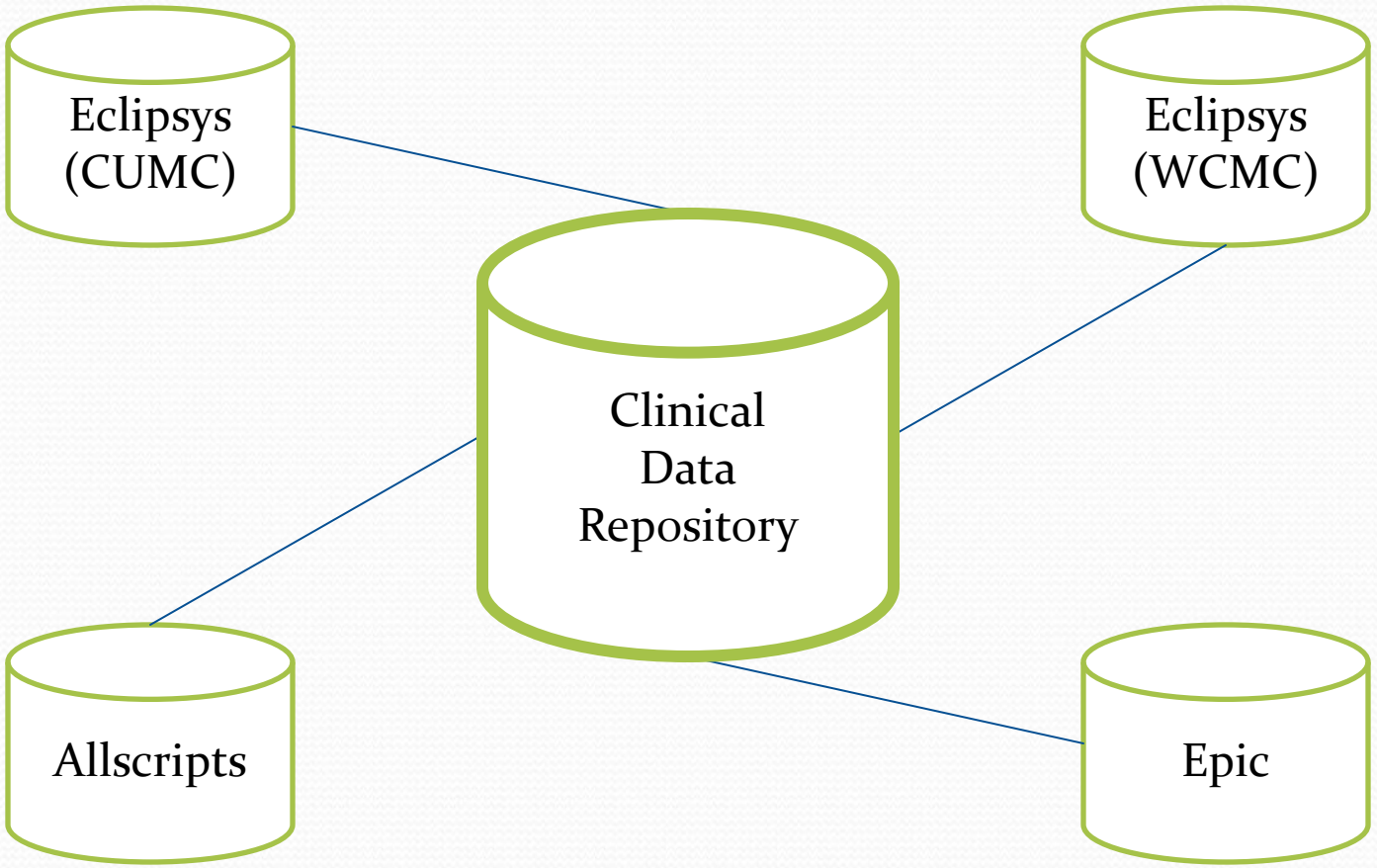
Integrating Among Multiple EHRs



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



Benefits of CDR

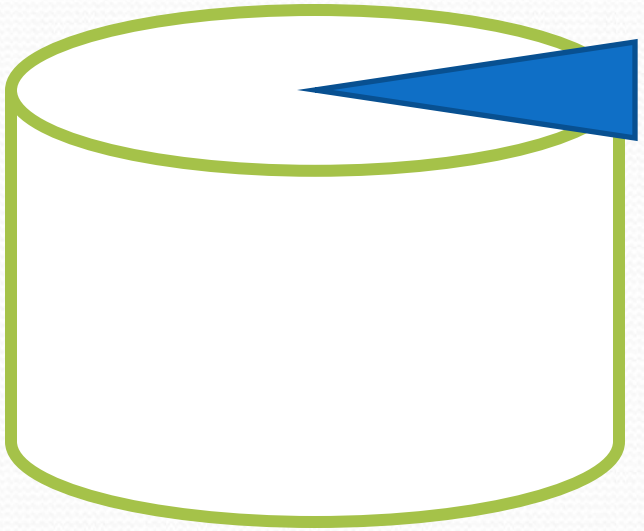
- Only model data from source systems once
- Common data store
- Data are read only
  - Optimized for read
- Historical data included
- Web-based viewer adaptable to multiple applications
- Adaptable to future health information exchange efforts
- Platform of innovation

Optimized for Retrieval

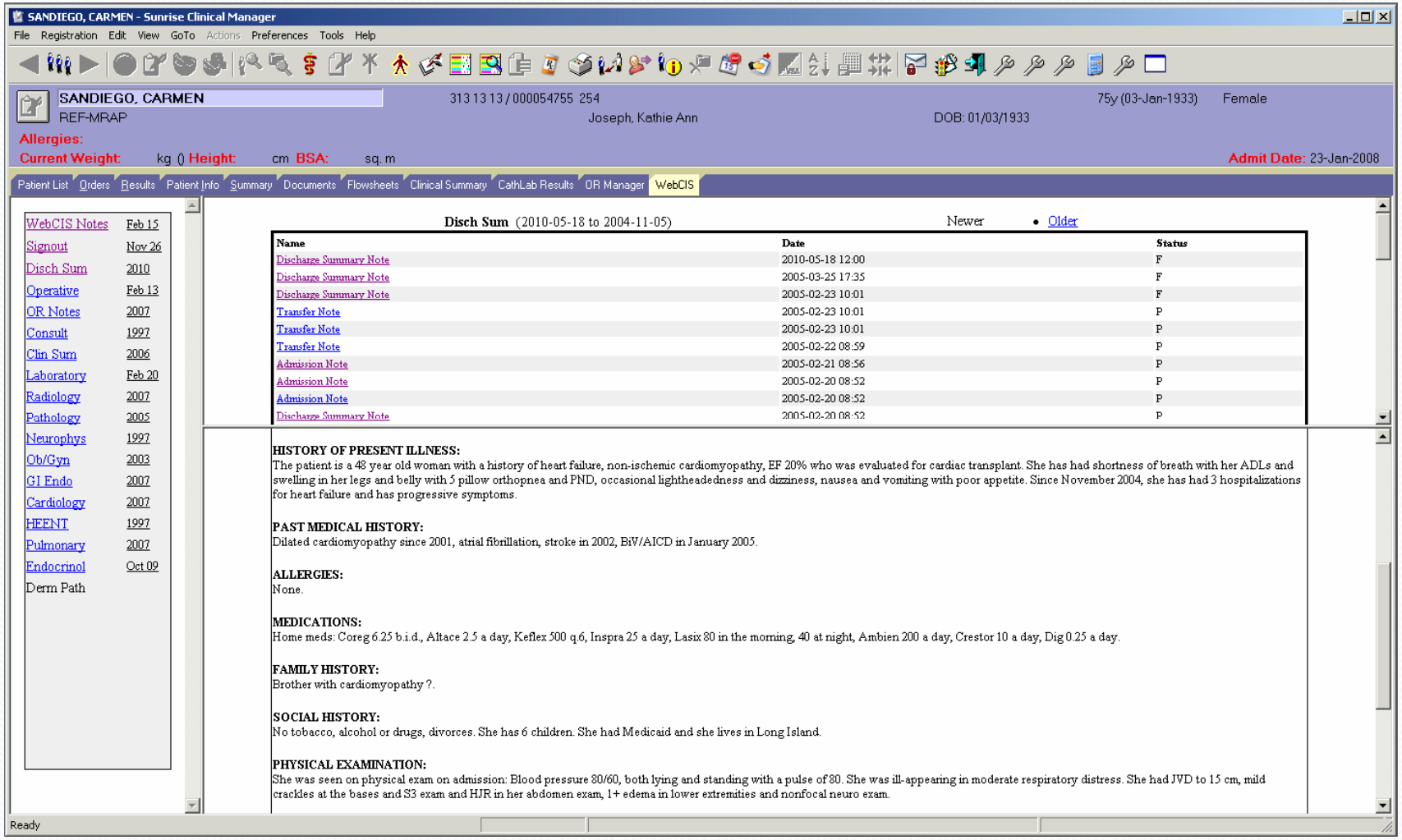
- Relational structure can be difficult to query for both data and context
  - Gathering multiple elements requires multiple table joins
  - Good for data storage
  - Good for aggregating across multiple patients
- Event-based model good for querying across data types
  - Data organized according to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

Retrieval optimization

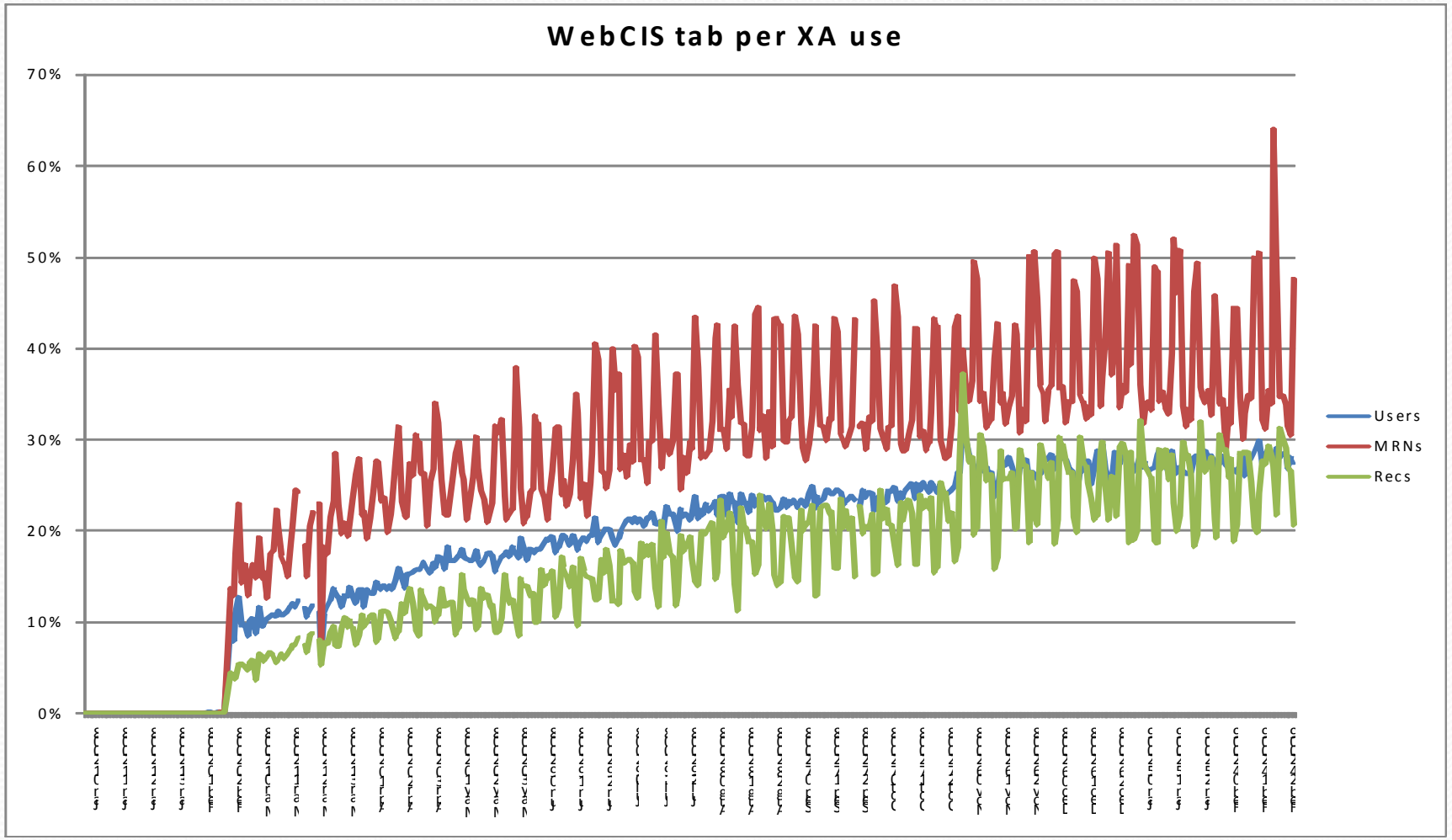
- Paradigm shift in how data are used
  - Paper records mainly for primary use
  - Electronic allows secondary use
  - Secondary use can be multiple times



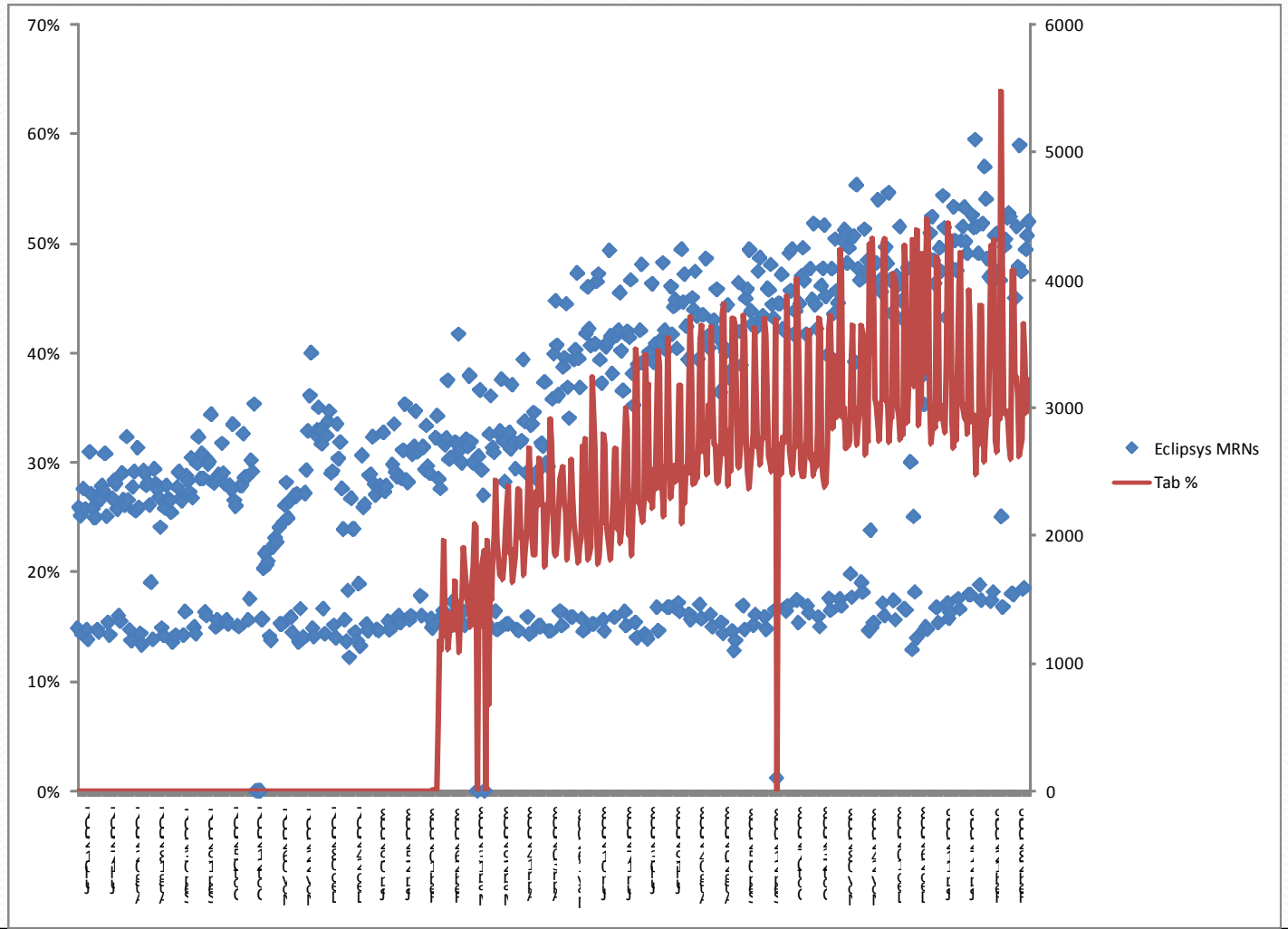
CDR View in Eclipsys



Proportion of CDR Viewer Access



Increase in CDR View Access



CUMC/NYP Clinical Data Warehouse History

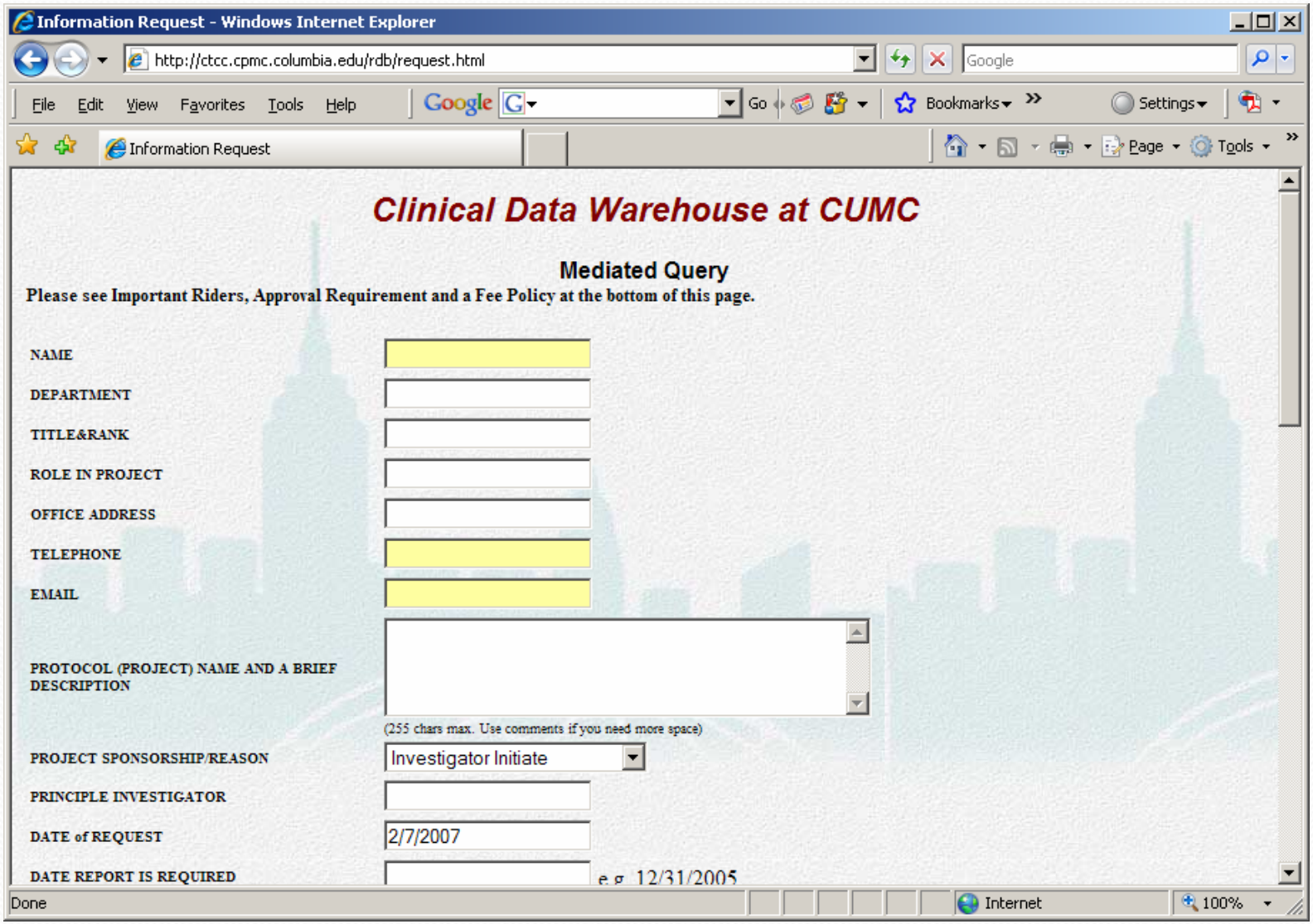
- 1994: Created, sponsored by Columbia University Department of Medical Informatics and Office of Clinical Trials
  - Populated with data from existing clinical data repository
  - Supporting clinical research
- 1998: Columbia + Cornell = NewYork Presbyterian Hospital
  - Warehouse funded by NYPH
  - 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

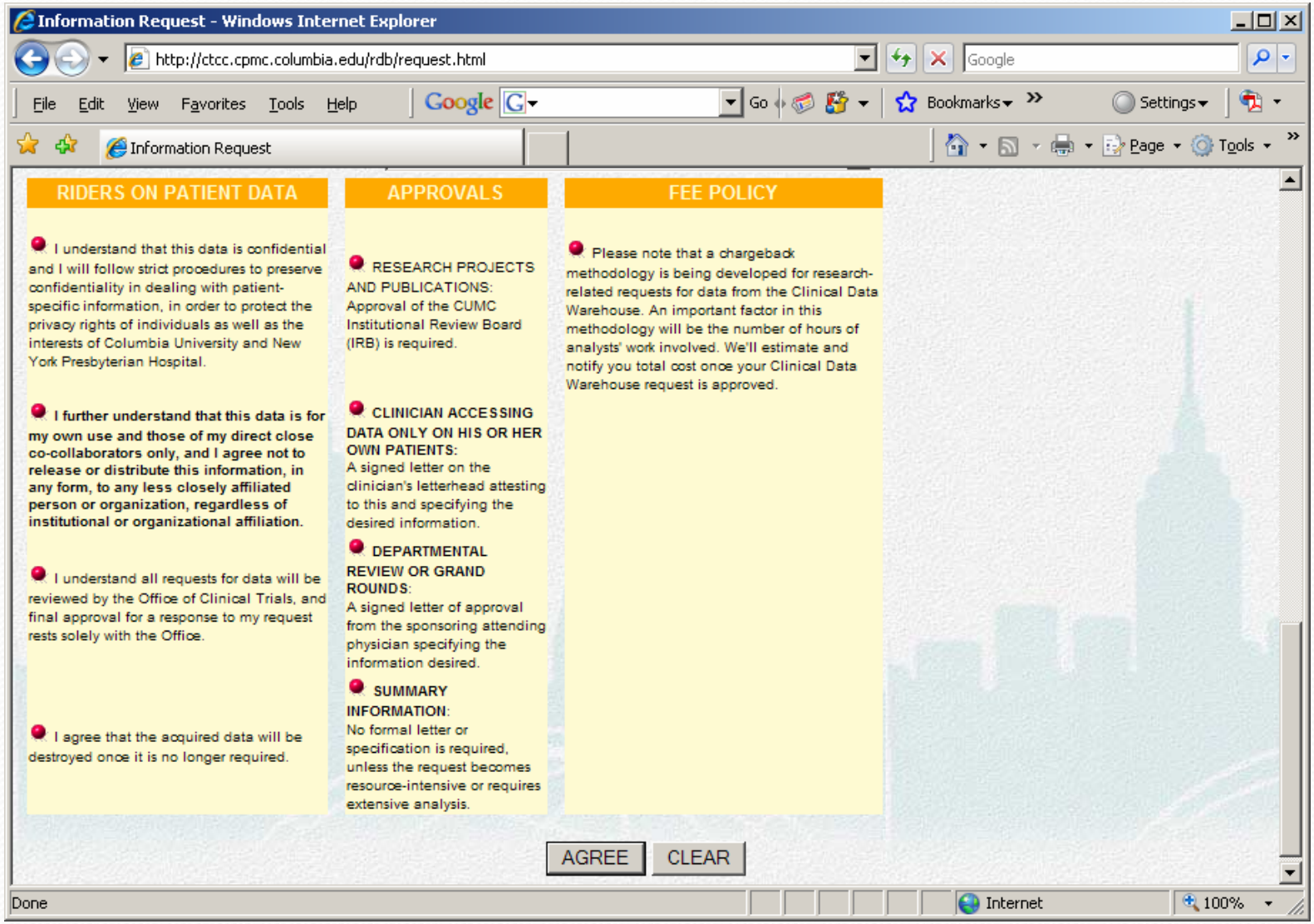
Uses of the Warehouse

- 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


Information Request - Windows Internet Explorer
http://ctcc.cpmc.columbia.edu/rdb/request.html
Google
File Edit View Favorites Tools Help
Bookmarks
Settings
Information Request
Page
Tools
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
(255 chars max. Use comments if you need more space)
PROJECT SPONSORSHIP/REASON
Investigator Initiate
PRINCIPLE INVESTIGATOR
DATE of REQUEST
2/7/2007
DATE REPORT IS REQUIRED
e.g. 12/31/2005
Done
Internet
100%


Information Request - Windows Internet Explorer
http://ctcc.cpmc.columbia.edu/rdb/request.html
Google
File Edit View Favorites Tools Help
Bookmarks
Settings
Information Request
Page
Tools
RIDERS ON PATIENT DATA
I understand that this data is confidential and I will follow strict procedures to preserve confidentiality in dealing with patient-specific information, in order to protect the privacy rights of individuals as well as the interests of Columbia University and New York Presbyterian Hospital.
I further understand that this data is for my own use and those of my direct close co-collaborators only, and I agree not to release or distribute this information, in any form, to any less closely affiliated person or organization, regardless of institutional or organizational affiliation.
I understand all requests for data will be reviewed by the Office of Clinical Trials, and final approval for a response to my request rests solely with the Office.
I agree that the acquired data will be destroyed once it is no longer required.
APPROVALS
RESEARCH PROJECTS AND PUBLICATIONS: Approval of the CUMC Institutional Review Board (IRB) is required.
CLINICIAN ACCESSING DATA ONLY ON HIS OR HER OWN PATIENTS: A signed letter on the clinician's letterhead attesting to this and specifying the desired information.
DEPARTMENTAL REVIEW OR GRAND ROUNDS: A signed letter of approval from the sponsoring attending physician specifying the information desired.
SUMMARY INFORMATION: No formal letter or specification is required, unless the request becomes resource-intensive or requires extensive analysis.
FEE POLICY
Please note that a chargeback methodology is being developed for research-related requests for data from the Clinical Data Warehouse. An important factor in this methodology will be the number of hours of analysts' work involved. We'll estimate and notify you total cost once your Clinical Data Warehouse request is approved.
AGREE
CLEAR
Done
Internet
100%

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

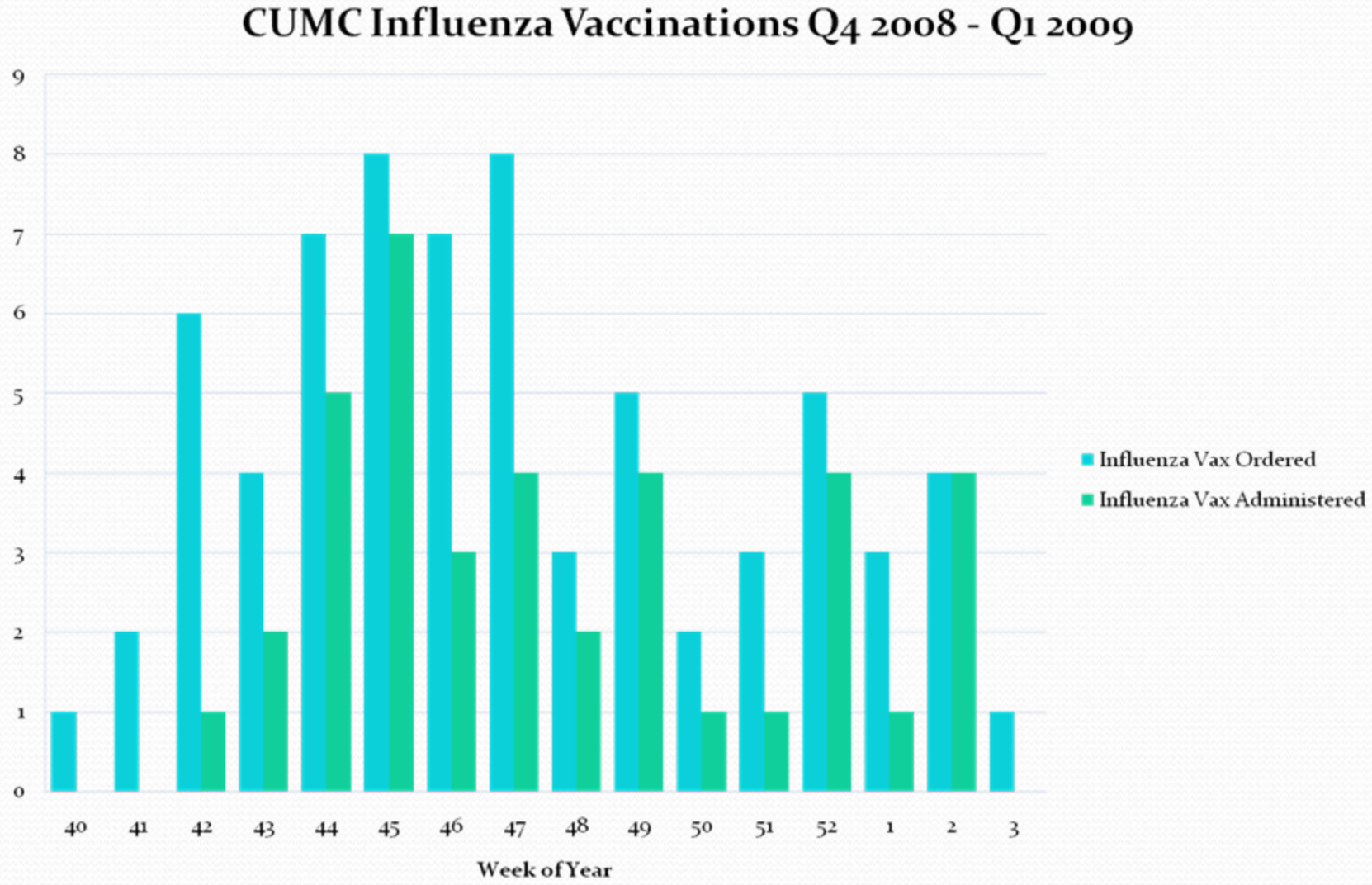
CUMC/NYP Clinical Data Warehouse History

- 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**

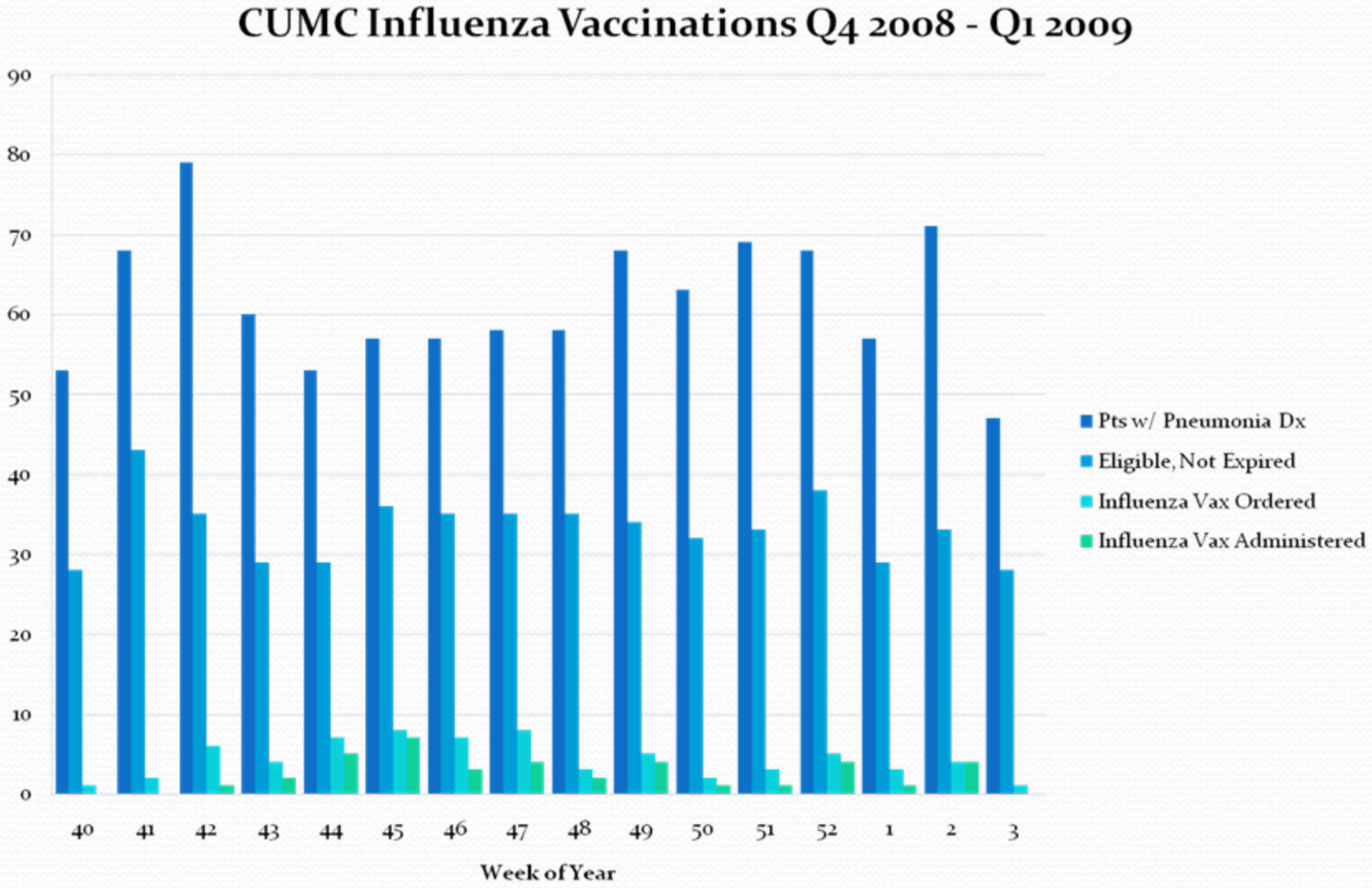
Analysis of Challenges

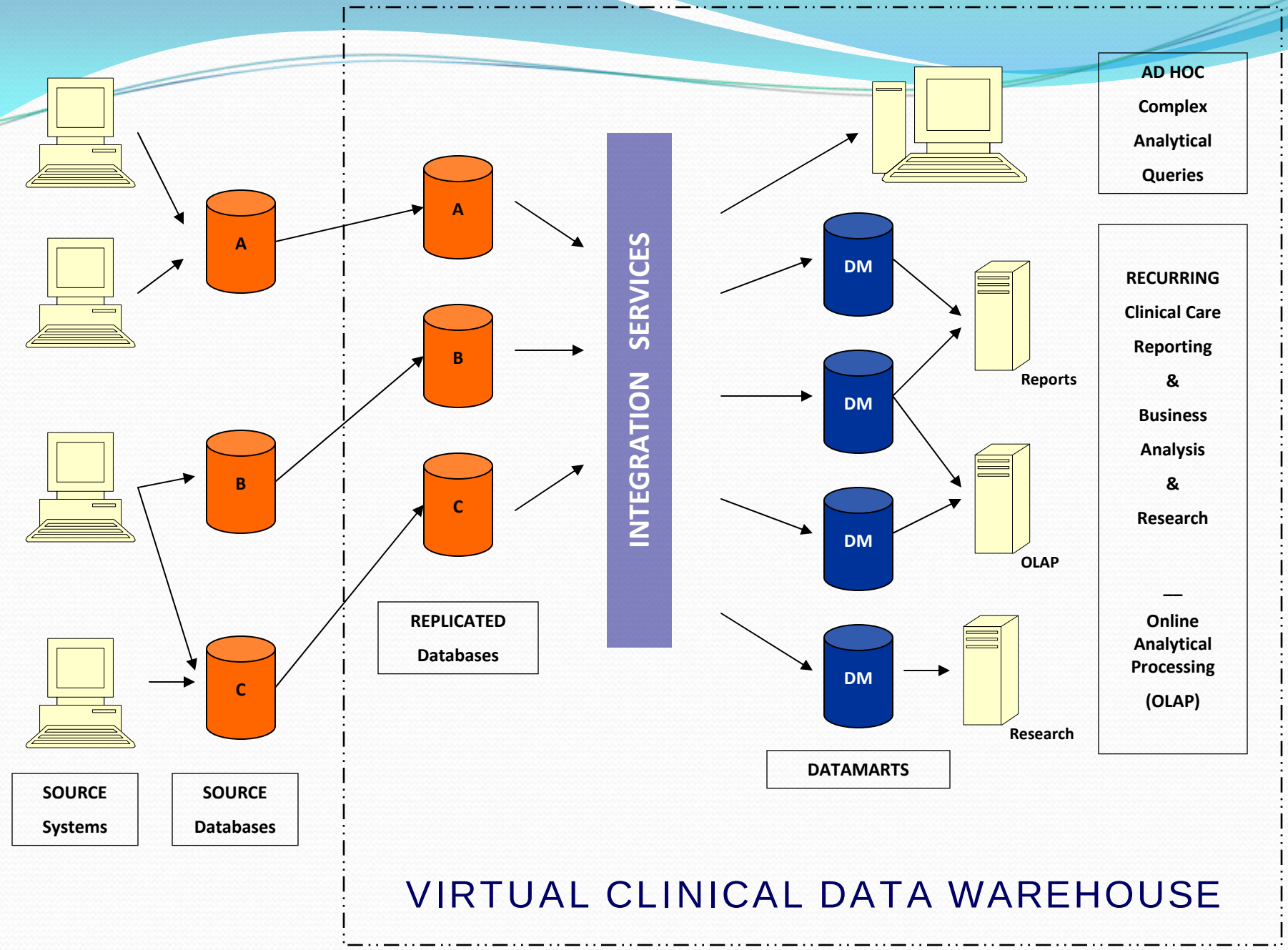
- Data in vendor-based transactional systems
- Could not query across transactional systems
- Users needed help in defining their needs
- Mature initiatives required more robust data solutions

Pneumonia Core Measures

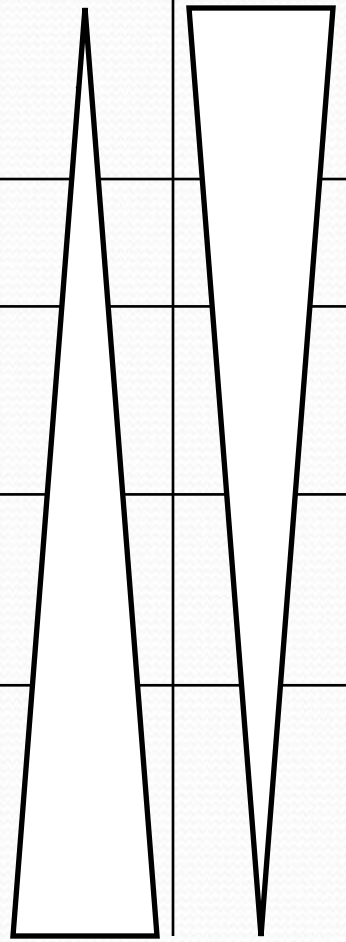


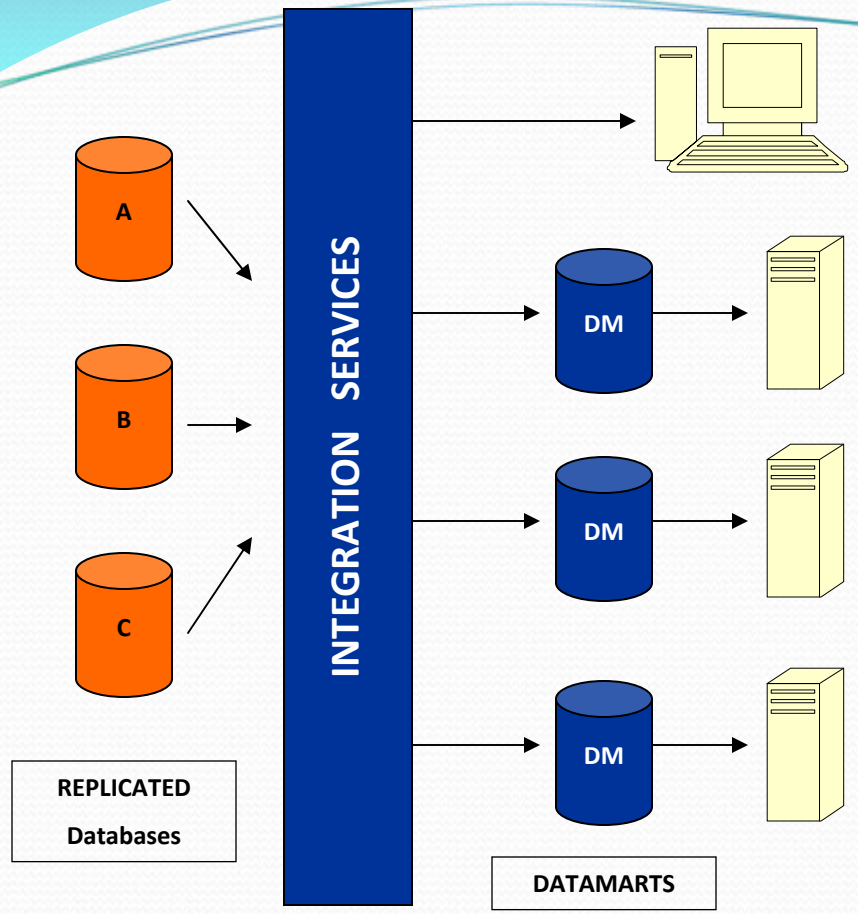
Pneumonia Core Measures




A
B
C
A
B
C
INTEGRATION SERVICES
DM
DM
DM
DM
Reports
OLAP
Research
AD HOC
Complex
Analytical
Queries
RECURRING
Clinical Care
Reporting
&
Business
Analysis
&
Research
—
Online
Analytical
Processing
(OLAP)
SOURCE
Systems
SOURCE
Databases
REPLICATED
Databases
DATAMARTS
VIRTUAL CLINICAL DATA WAREHOUSE

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





VIRTUAL DATA WAREHOUSE

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

DATA WAREHOUSE TOOLS

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