The MUSC Clinical Data Warehouse (CDW) use for Research

Jihad Obeid, MD
Jean Craig, PhD
Outline

• Definition of terms:
  – EDW
  – CDW (MUSC/HSSC)
  – Biobank
  – Biobank data mart

• Focus on CDW for Research
  – About the project
  – Demo
Enterprise Data Warehouse (EDW)

http://musc.edu/projects/edw

The Enterprise Data Warehouse

MUSC has established an Enterprise Data Warehouse to support clinical, management, research and other operational initiatives. The EDW consolidates data primarily from Oacis, Clindoc and other applications currently in use at MUSC.

The vision for the EDW is to convert data into meaningful information that is secure, accurate and real-time so that staff, faculty and administrators throughout the campus will be able to use this data for decision making and research activities.

The EDW currently contains secure electronic clinical data from the OACIS Clinical Data Repository, including patient demographics, ICD-coded diagnoses and procedures, laboratory test results and, for inpatient admissions, nursing documentation and medications ordered and administered.
Enterprise Data Warehouse (EDW)
Clinical Data Warehouse (CDW)

• MUSC EDW
  – Used for quality measures and improvement
  – CDW is the *subset of EDW* used for research (CDW-R)

• Health Sciences South Carolina (HSSC) CDW
  – Includes MUSC EMR data only
  – Two options: MUSC data alone or HSSC hospitals (aggregated)
MUSC self-service (MicroStrategy)
HSSC self-service (i2b2) – coming soon

i2b2 (Informatics for Integrating Biology and the Bedside) www.i2b2.org
## CDW (local+state) vs. Biobank

<table>
<thead>
<tr>
<th></th>
<th>EDW</th>
<th>HSSC CDW</th>
<th>Bio Bank/ Data Mart</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source of data</strong></td>
<td>MUSC</td>
<td>MUSC EMR + HSSC partners</td>
<td>MUSC (+ HSSC partners?)</td>
</tr>
<tr>
<td><strong>What type of data is available?</strong></td>
<td>Clinical + other</td>
<td>Clinical</td>
<td>Clinical + Biospecimen</td>
</tr>
<tr>
<td><strong>Coming soon</strong></td>
<td>HR/Payroll, Surgical, and Patient Charges/Reimbursement</td>
<td>Rollout for MUSC then other HSSC partners</td>
<td>(Still in planning)</td>
</tr>
<tr>
<td>**Is there identifiable data? *</td>
<td>Yes (but not for self-service)</td>
<td>Yes (but not for self-service)</td>
<td>No</td>
</tr>
<tr>
<td><strong>Database platform</strong></td>
<td>Sybase IQ/Emergis</td>
<td>Oracle/Recombinant</td>
<td>Oracle /caTissue</td>
</tr>
<tr>
<td><strong>Self-service data (DUA required)</strong></td>
<td>Aggregate numbers (no identifiers)</td>
<td>Aggregate numbers (no identifiers)</td>
<td>De-identified records + biospecimens</td>
</tr>
<tr>
<td><strong>Self-service interface</strong></td>
<td>Microstrategy (BI)</td>
<td>i2b2</td>
<td>i2b2</td>
</tr>
<tr>
<td><strong>Status</strong></td>
<td>In production</td>
<td>Under development</td>
<td>In planning</td>
</tr>
</tbody>
</table>

*IRB approval required for brokered access to data – see schematic
The CDW-R Overview

Aim 1: Access to aggregate data (no identifiable information)

Self-service

Aim 2: Access for IRB approved protocols - Brokered access

IRB approval

Data Request Cmte

Honest Broker

Researchers

aggregate data (no identifiable information)

CDW

DUA

MicroStrategy Reporting tool (Restricted View)
More Information
https://sctr.musc.edu/index.php/cdw

MUSC Clinical Data Warehouse for Research (CDW)

WHAT IS THE CDW?
The MUSC CDW is a single, secure, integrated database extracted from the MUSC OACIS Clinical Data Repository, which includes patient demographics, ICD-coded diagnoses, ICD-coded procedures, medications, and laboratory test results.

WHAT IS THE PURPOSE OF THE CDW FOR RESEARCH?
The CDW will provide clear pathways for data requests with full regulatory compliance and accelerate our translational research programs that will lead to new diagnostic tools and therapies for a range of diseases.

ACCESS
Only MUSC faculty and faculty-sponsored staff may access the CDW. There are two methods for accessing the CDW for research:

METHOD 1: Access to:
- De-identified aggregate data through a restricted query interface and a signed data use assurance (DUA) in accordance with HIPAA guidelines.

A typical example would be the number of diabetics seen in the year 2009 stratified by gender. This will provide researchers with a self-service data request model for managing high demand queries for aggregate de-identified data.

*Method 2 could also be used for aggregate de-identified data queries if the query cannot be performed using self-service access or for more complex queries with multiple variables.

Self Service Access Click Here to Enter
Collaborative Effort

- Office of the Associate Provost for Research
  - Loretta Lynch-Reichert
  - Stephen M. Lanier, PhD
- SCTR (CTSA grant # UL1 RR029882)
  - Jean Craig, PhD
  - Jihad Obeid, MD
  - Amanda Zimmerman
- Enterprise Analytics Team & Hospital Administration
  - Nestor Esnaola, MD
  - Pat Wagstaff
  - Amy Wilson, PhD
  - Stuart Smith
  - Patrick Cawley, MD
  - Casey Liddy
- OCIO Information Services
  - Jeff Burdick
  - Crystal Smith
  - Larry Gale
  - Christopher Hafer
  - Mitchelle Morrison
  - Mike Coffman
  - John Imholz
  - Dan Furlong
  - Frank Clark, Ph.D.
- Data Request Committee
  - Mike Wheeler
  - Patrick Mauldin, PhD
  - Jihad Obeid, MD