Introduction to the EDW

Daniel Schneider
Manager of Research Analysts

IRB Brown Bag
3/16/2016
Agenda

• What is the EDW?
• What types of research can it help with?
• FSM Policy
• Q/A
What is the EDW?
The EDW Provides a Campus Data Platform for Your Approved Projects

Participating Entities

Campus Data

Contribute

Report

Analyze

EDW

Load

Cleanse

Enhance

Provide Campus Data Platform

Analyze Data and Create Reports
What is the EDW?

The EDW is a SQL database that contains data from many of the transactional systems used by the campus.
What is the EDW?

Clinical data from NMG (EPIC)

Clinical data from NMH (Cerner)

Billing data from NMH (PRIMES)

Billing data from NMG (IDX/Epic)

More clinical systems (ancillary)

Research systems from FSM
What is the EDW?

The end goal of the EDW is to provide a more comprehensive dataset and options for interacting with and understanding it.
What is the EDW?

EDW (Enterprise Data Warehouse) is a tool that provides access to a comprehensive collection of data across Northwestern Medicine. It allows users to access and analyze data for various purposes, such as research, operational oversight, and patient care improvement. The EDW includes features like reporting, custom data requests, and notifications.

For example, the EDW provides reports such as the Average Hours by Month, Dollars YTD, and EDW Realtime. These reports can help users understand trends and make informed decisions.

The EDW also includes features like MY REPORTS, which allows users to view their favorite, frequently run, and recently granted access reports. It also includes NOTIFICATIONS, which inform users of any new data or updates.

The EDW Portal Admin is a feature that allows administrators to manage the EDW portal and its permissions.
Common Workflows
Intro to the EDW

Common Workflows

Research, Operations or Quality

**Research** IRB approval is required if EDW is used for Research purposes, especially when results may be published in a journal.

**Operational** Reports that display metrics used by administrators on a day-to-day basis.

**Quality** Reports must support an NMH DMAIC, be sponsored by a physician department chair at NMG or otherwise officially sponsored.

*Research reports should not be used for operational or quality purposes; the reverse typically holds true, as well.*
Research Uses of the EDW

- Study Feasibility
- Clinical Trial Subject Recruitment
- Retrospective Chart Review
- Prospective, Longitudinal Studies
Study Feasibility

- Aggregate counts of patients meeting defined criteria
- No approval needed for prep-to-research data containing no PHI
- i2b2 is currently available and is being actively developed

Example:

“How many patients we've had over the years with ovarian cancer and a recurrence in the brain...”
Clinical Trial Subject Recruitment

- Real-time scheduling data from NMG
- Primary care physician contact information
- Daily updates to the subject population

Example:

“Patients with the diagnoses of peripheral arterial disease and/or coronary artery disease...”
Retrospective Chart Review

• Data in the EDW reaches back about to 1998…
• … but becomes robust after 2000
• Discrete fields can be used to efficiently define and capture data elements for a population
• Complex NLP can be leveraged for free-text, though not as reliably as discrete fields

Example:

“The study is evaluating patients at NMG Ophthalmology with the following ICD9 codes between Jan 1, 2004 and Jan 1, 2011…”
Prospective, Longitudinal Studies

• Disease, other study registries can be built to keep track of a population, events and outcomes
• Data can be integrated with custom applications

Example:

“For consented patients provide the following data points for the first available Cardiac Cath Report: 1. Mean RA 2. Mean PA 3. PA Saturation % 4. Mean PCW 5. Mean LV 6. Fick Output...”
Getting Started
Getting Started…

• Submit Data Request using the EDW Portal
• Work with an EDW Analyst to refine requirements
• Service Agreement
  • Data Steward approvals
• Reports retrieved from the EDW Portal

It’s an iterative process!
Requesting an EDW Report

- IRB and NMHC approval is needed
- Discrete data elements are best
- Be able to send screen shots of the data elements you’re looking for
- Free text is difficult to work with
  - NLP can be used but is time consuming (expensive) and rarely as accurate as a discrete field
- Recharge rate of $75/hour
What is a Data Steward?

- The data in the EDW is still owned by the contributing organizations.
- Data Stewards are representatives from these organizations and approve data and access requests.

More information on the EDW Wiki

http://edw.bioinformatics.northwestern.edu/edwwiki/index.php/Data_Steward_Policy
Research Use of EMR Data Policy

• Data recorded in Northwestern Medicine electronic medical records systems for clinical care and desired to be used for research must be obtained from the Northwestern Medicine Enterprise Data Warehouse (NMEDW)
  – Chart abstraction for research purposes or chart abstraction for clinical care purposes then later repurposed for research purposes is prohibited
  – Data requests should be limited to data that is required to conduct study
• Exceptions are granted per study for:
  – Adverse Event Forms
  – Case Report Forms for FDA-Regulated Clinical Trials
  – Images/Scanned Attachments
  – Real-time Data
  – Interpretive Chart Review (under waiver of consent)
Important Issues in Clinical Data Warehousing
Important Issues in Clinical Data Warehousing

GIGO: Garbage In, Garbage Out

- “We don’t make the news, we just report it.”
- If a system has a 43yo man with ovarian cancer, so will the EDW

When to clean data? Inbound vs outbound

- If data are cleaned as they arrive in the EDW, faith is sacrificed. One needs to be able to see the 43yo man with ovarian cancer in the EMR
- Few researchers want men with ovarian cancer in their reports; so we exclude the outliers when data are sent to the requester

Data provenance

- One should be able to trace any datum in the EDW all the way back to the source system
Future Directions
Future Directions

• Integrated Data Structures (IDS)
  – Incorporate new source data (Cadence, CDH)

• Predictive Analytics/Biostatistics
  – Integrate R into our analytics infrastructure
  – Provide preliminary statistical analysis
  – Utilize advanced data visualizations

• Self-Service Query and Annotation
  – REDCap
  – Abstractor

• Predictive Analytics
  – RAPID

• Advanced Natural Language Processing
  – Siddhartha Jonnalagadda (sid@northwestern.edu)
Quick Reference

NMEDW@northwestern.org
http://edw.northwestern.edu
https://grants.nubic.northwestern.edu/
Questions?