

Building a Perioperative Data Warehouse

Matthew A Levin MD
Assistant Professor
Department of Anesthesiology
Division of Cardiothoracic Anesthesia
Icahn School of Medicine at Mount Sinai



Disclosures

- None

Define: Data Warehouse

da·ta ware·house
noun COMPUTING
noun: **data warehouse**; plural noun: **data warehouses**.
1. a large store of data accumulated from a wide range of sources within a company and used to guide management decisions.

Define: Perioperative data warehouse

1. a large store of data accumulated from a wide range of sources within a health care organization and used to?

- support/enable research
- support/enable PQRS?
- support/enable departmental PI/QA?
- support/enable hospital PI/QA?
- archive data from legacy AIMS systems?

Difference between an DW and an AIMS

- AIMS is a tool for intraop charting
- AIMS usually only includes limited periop data
- AIMS reporting may be limited to pre-built reports
- AIMS cannot archive from other AIMS systems (e.g., legacy)

DW as an archive

- What to do with old data when migrating to new AIMS?
- Build a warehouse and populate with old data
- Build a feed that adds new data from new AIMS
- Many issues with schema mapping, patient identifier mapping, etc.

Tools for building a periop DW

- People
- Support
- Hardware
- Software

Challenges

- Getting data out of AIMS
- Integrating hospital feeds
- Validation/cleaning/phenotyping
- Ongoing HW/SW support
- Who will pay for all of this??

Can EPIC be a Periop DW?

- EPIC is definitely not a data warehouse
 - Its more like a rusty file cabinet that you can't open, with a slot on top for stuffing files in
 - The structure is hierarchical, there is literally no way to easily query across cases/patients/encounters



Getting data out of EPIC to a DW

- **Clarity - SQLServer (i.e., relational) extract from Cache**
 - “Standard” reports, which may or may not include all the data that you want
 - Batch
 - Limited access, no ad-hoc queries, have to go through reporting team

Getting data out of EPIC to a DW

- **Custom feed**
 - Web Services (Epic as provider, SOAP protocol)
 - External calls (Epic as consumer)
 - Message Passing (bi-directional SOA model, XML/XSD schemas)
- \$\$\$\$

EPIC data - Caveats

- **NO real-time intraoperative vitals***
 - Party line remains that Epic cannot provide realtime feed of intraop vitals, even via custom feed
- **NO MAR (Medication Administration Reconciliation) data***
 - MAR team remains adamant about not providing access to MAR data

*Disclaimer: hearsay, but believed to still be true

Cloud-based AIMS and Periop DW

- There are now several “cloud” based AIMS (e.g. Plexus, iPro Anesthesia, AnesthesiaOS, Talis)
- They all store case data on remote servers
- They all provide facilities for reporting and for sending PQRS data to AQI
- Unclear ability for ad-hoc querying
- Unclear how they integrate periop data, if at all

13

The Sinai Experience - ORDW

14

ORDW overview

- **Current design operational since 2011***
- **MySQL based**
- **Continuous load of intraop case data**
 - custom parser reads CompuRecord (CR) binary files
- **Batch load of periop data from enterprise DW (MSDW)**
- **Independent of CR SQLServer**

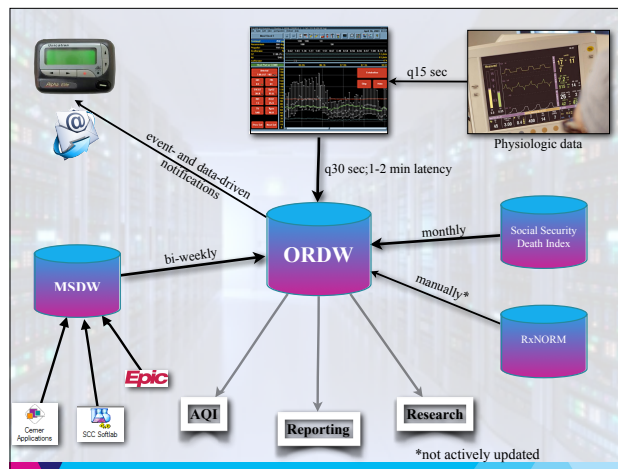
*Prior to 2011, no MSDW feed, CR data dumped into MySQL via CR research tool and spoofed ODBC link. Batch process, flaky.

15

ORDW overview

- **Social Security death index via automated loader**
 - custom loader
 - kept in separate db to limit access
- **Other data loaded manually**
 - CPT codes, HCUP CCS, personnel lists...
- **Continuously updated summary table**
- **Minimal other summary, validation or phenotyping**

18



ORDW - hardware & software

- **16 core machine (Quad core 2.53 GHz Xeon x4), 192 GB RAM, 3 TB disk**
- **Ubuntu 14.04, MySQL 5.6**
- **SQL Server 2008 added in 2013 for enhanced reporting**
 - 24 Core 2Ghz Intel Xeon, 32GB RAM, 543GB disk
 - SqlServer Reporting Services (SSRS) are great
- **Hardware and OS maintained by main IT**
 - fault monitoring, machine backup, OS patches and upgrades, some database support
- **DB mostly maintained by Anesthesiology Dept**
 - Not unwilling to give access, but understanding schema requires deep knowledge and dedicated personnel
 - Main IT DBA's are okay with MySQL but know Oracle and SQLServer better
 - MSDW team assists with periop schema

18

ORDW - schema

- **Main tables mimic CompuRecord schema**
 - Makes loading easier
 - Its a good schema - mix of relational and dimensional
 - Downside: likely not useable if/when we migrate to Epic
- **Periop data in separate db, quasi-relational**
 - Already abstracted so won't have to change
- **Separate schema for reporting tables**
- **“Scratch” schema**
- **Does not currently include all CR audit metadata**
 - Its in the binary files, also in CR SQLServer

19

ORDW - periop data

- **Custom feed, designed with and implemented by MSDW team**
- **Separate database, separate schema, quasi-normalized**
 - 2 week lag (matches lag in MSDW Epic feed)
- **Rich set of perioperative data**
 - Encounter data (encounter ID, admit/discharge date, discharge disposition)
 - Demographic and billing data (race, zip, payor, billing codes)
 - ICD-9-CM diagnosis and procedure codes, APR-DRG codes
 - ADT data (Admit, Discharge, Transfer)
 - Laboratory data 72h pre- and 7d post
 - Vitalsigns from floor 72h pre- and post
 - Preoperative medications (complex)
 - Medication administration data 72h pre- and post
 - ECG data (report only, no waveform data)

*Arbitrary windows

20

ORDW - summary table

- **Wide (300 column) table which aggregates data from many base CR tables**
- **Many calculated fields**
 - Age at surgery
 - BMI
 - Blood product and fluid totals
- **Various binary flags**
 - is_peds
 - is_aline
 - is_regional
- **Table definition stored in spreadsheet**
 - programmatic update and table regeneration (perl script) when adding new columns
- **Primary initial table for all research queries**
 - Replicated to billing server and used for billing extract
 - Replicated to reporting server and used for reporting and compliance notification

21

ORDW - size

- ~550,000 cases
- ~300-400 new cases added daily
 - now from 4 sites
 - CR data only from 3 new sites
- Total DB size: ~ 770 GB
- Total tables: ~18 core CR tables, 10 core periop tables, dozens of ancillary tables
- vitalsigns table: ~600 GB, 5+ billion rows
 - partitioned

ORDW - performance

- The key to performance for any RDBMS is RAM
 - We now have lots of RAM
- Performance not an issue for most queries
- The one exception is large aggregate queries of vitalsigns
 - Plan to add 2 replication machines (128GB RAM, 6 TB disk each)
 - Will deploy ShardQuery* to enable distributed/parallelized queries

*<http://www.percona.com/blog/2014/05/01/parallel-query-mysql-shard-query/>

ORDW and Epic

- MSDW acts as intermediary
 - Daily export from Cache to Clarity
 - From Clarity to MSDW every 2 weeks
 - feed designed and maintained by MSDW team in collaboration with Epic reporting team
- From MSDW to us every 2 weeks
- Result: we get what we want and don't really have to deal with Epic

ORDW - what we don't have

- **Full text of notes**
 - 90% of note content is redundant and useless anyway
- **Imaging reports, or image data**
- **Allergies (Epic reported)**
- **All Epic flowsheet data**
 - For example, vent data for ICU patients
 - not contained in any existing Clarity reports
 - would have to have Epic reporting team build a new report
 - low priority for reporting team
 - MSDW then has to modify their Clarity feed (and test/validate)
 - MSDW then has to modify our anesthesia feed (and test/validate)

25

The Vanderbilt* experience

- **SQL Server based**
- **Batch loads of both AIMS and external periop data**
 - AIMS data loaded from backup files 2x daily
 - Periop data loaded from Enterprise Data Warehouse
- **Cleaning/modeling using T-SQL MERGE statements**
- **Phenotyping i.e., risk score calculation, etc.**
- **Reporting via Tableau software**
 - Custom one-off queries for research requests
- **680,000+ cases**
- **5 billion+ vitalsigns**

*Thanks to Jon Wanderer MD for this information

26

ORDW - future directions

- **Enable distributed queries (Replication servers and ShardQuery)**
- **Summary tables for vitals**
 - q 1 min, q 5 min, etc
- **Automated cleanup on import, or routinely**
 - Most data cleaning currently done in R, after export
 - Same thing done over and over
 - "case_fixups" table
- **Automated calculation of comorbidity scores**
 - Charlson, Elixhauser, RCRI, etc
- **Automatic classification of preoperative medications**
- **Visualization**
- **Periop data from other Mount Sinai health system sites**

27

Summary

- Building a periop DW is a major effort
- Reward is a rich dataset that can be used to answer deep research questions
- Most groups will not need a periop DW or want to build one. Built in reporting facilities will be good enough.
- Epic is not a data warehouse

28

Thank you

