Clinical Data Warehouse (CDW)
AKA Caboodle

Dana Ludwig
Enterprise Information and Analytics
### UCSF Clinical Data Warehouse (CDW)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Self-serve (Free)</th>
<th>Consult Required (May have recharge)</th>
<th>IRB needed</th>
<th>Requires MyResearch account or other secure environment</th>
<th>Includes clinical notes</th>
<th>UC Health data available in addition to UCSF data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counts</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>De-ided data</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes *</td>
</tr>
<tr>
<td>Limited data</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes *</td>
</tr>
<tr>
<td>Id’ed data</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes *</td>
</tr>
<tr>
<td>Recruit</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes *</td>
</tr>
</tbody>
</table>

* Process to obtain data from UC Health sites can be very lengthy
UCSF Clinical Data Warehouses (CDW)
Clinical Data Warehouse (CDW)
High-Value APeX Data for Research

• Created by Epic about 2 years ago
• Sourced from High-Value Clarity Tables
  • 170 Tables today (UCSF uses only 104)
  • 2698 Columns (UCSF uses only 1525)
  • Small enough that you can understand all of it!
    • Solves the Forest and Trees problem of Clarity
• Includes a web-based Data Dictionary including where the data came from
• Supports integration with Custom Non-Epic tables
• Supports merging Custom Non-Epic data into Epic tables
  • As long as you don't add new columns to Epic tables
  • Ideal for UCSF STOR Legacy data integration
• Includes a daily, incremental ETL supported by Epic
  • CDW is only 1 day behind APeX.
Clinical Data Warehouse (CDW)  
What it Contains

- Includes migrated legacy data before 6/1/2012
- All Data feeding RDB:
  - Demographics, Diagnoses, Med orders, procedure orders, billing procedures, Billing accounts, Insurance Coverage data, visits, hospitalizations, labs, providers, departments, allergies, immunizations, vital signs
- **UCSF Custom data**: STOR data warehouse, others in the future
- flowsheets feeding the current flat files
- Medication Administration Record (IP Only)
- Medication Dispensing (IP Only)
- Surgical Cases and Procedures
- Emergency Department Visits
- Care Teams per Patient per Time Interval
- Chief Complaints
- SmartData
- Lab Tests (not just Lab Results)
- ADT - Patient Location (IP Only)
Clinical Data Warehouse (CDW)
Limitations of CDW

• No clinical notes today
• HIPAA PHI - Like Clarity, it has identifiable patient information, so governance for research use is complicated
• Not the complete EHR - sometimes you need to go back to Clarity.
CDW Web-based Data Documentation - by Epic

FlowsheetValueFact

The flowsheet value fact contains information about measurements from flowsheets. Each row represents a flowsheet value.

Properties

- **Granularity**: A flowsheet value
- **Profile tables**: FlowsheetValueProfileDim
- **Links to**: DateDim, EmployeeDim, EncounterFact, FlowsheetFlowDim, FlowsheetTemplateDim, FlowsheetValueProfileDim, PatientDim, TimeOfDayDim
- **Links from**: none
- **Change type**: Type 1

Sources

- **Non-Epic**: Unspecified custom extracts

Columns

- **FlowsheetValueKey**: integer - Surrogate key used to uniquely identify the record
- **IdType**: string - This column should no longer be used and will be removed in a future version. Source ID type for th
## De-Identified CDW

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Self-serve (Free)</th>
<th>Consult Required (May have recharge)</th>
<th>IRB needed</th>
<th>Requires MyResearch account or other secure environment</th>
<th>Includes clinical notes</th>
<th>UC Health data available in addition to UCSF data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counts</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>De-ided data</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes *</td>
</tr>
<tr>
<td>Limited data</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes *</td>
</tr>
<tr>
<td>Id’ed data</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes *</td>
</tr>
<tr>
<td>Recruit</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes *</td>
</tr>
</tbody>
</table>

* Process to obtain data from UC Health sites can be very lengthy
Near Future: De-identified Clinical Data Warehouse

- Clinical Data for All UCSF Researchers:
  - As with CDW, it's 104 tables are easy to learn by everyone
  - Deidentified --> all UCSF investigators can see all the data on all the patients without IRB approval
  - Deidentified --> Self-service access to the data.
    - No "Honest Broker" required
    - no recharge
    - no delay to get your data
- **User-friendly views**
  - Internal codes are defined within the same view
    - e.g., Diagnosis had diagnosisKey = 12345
    - Now has DiagnosisName = 'Diabetes Mellitus'
  - A single, unique patient key is provided in every patient-related view
- User-friendly Flat Files - in case your software can't read relational tables.
- Available 4th Quarter, 2017