Clinical Data Architecture for Business Intelligence and Quality Reporting

Aaron Abend
Managing Director, Recombinant Data Corp.
• The role and value of an effective data architecture
• The unique challenges of healthcare data
• How to define and achieve effective clinical data integration
• How to define and achieve Data Quality
• How to create effective analytics for clinical quality and reporting for mandates such as Meaningful Use require using data from multiple systems
• An approach to creating a clinical data warehouse for reporting and analytics
Why do you need Data Architecture?
Data Architecture Defined

• Data Architecture is the design of data structures and data semantics to create data resources that can be trusted by the institution’s stakeholders, including:
  • Leadership & staff
  • Affiliated clinics, institutions & suppliers
  • Patients
• The result is that the institution can make decisions that benefit all of its stakeholders
For Selling analytics…

• Business Intelligence benefits the institution, not the individual
• Good BI shows the individual how their individual success and institutional success are connected
• Data Architecture is necessary for good data quality
• It is an important first step, institutionally, if everyone can agree on What, Why, & How
• But you cannot implement change until you can agree on WHO!
Data Architecture in Healthcare
A Unique Challenge
What makes healthcare different?

• Other industries talk about individual treatment for customers, but healthcare has no other model

• Data capture in healthcare is weak because adoption of technology is 20 years behind

• In the absence of centrally implemented systems, departmental systems have emerged

• Workflows are unlike any other industry – the business side of the healthcare enterprise is weakly connected to the care side

• Data is scattered in many places…

• And data quality is extremely poor (especially compared with other industries)
Today’s pressures in Healthcare

• Healthcare costs have escalated to levels considered unsustainable *

• Government regulation will soon penalize institutions that cannot prove “meaningful” use of their systems

• The new healthcare reform bill calls for changes in the business model that will put exceptional focus on understanding risk and reward – an understanding that is difficult in such a complex system even if data were available

Pressure for better analytics

Need for Analytics

- HITECH & HIPAA requirements
- EHR Adoption / Meaningful Use
- Personalized Medicine / Genomics tipping point
- Escalating Costs
- Accountable Care Payment Reform
- Public confusion
Sources of data in healthcare

- Billing and administrative systems
  - Financial data – weakly connected to medical practice
  - Coded for collection
  - Not trusted by clinicians
- Legacy Ordering Systems
  - Labs
  - Medications
- Electronic Medical Records
  - Mostly unstructured notes
  - Inpatient and outpatient systems often separate
- Departmental registries
  - “Dark matter” of healthcare data
  - Governance challenge
- External systems
  - Labs, Pharmacies, Affiliated Institutions, Referring providers & clinics …
Data States in Healthcare

Service  
Bill  
Paid Claim  

Prescription  
Pharmacy  
Correctly taken  

Lab Order  
Results  
Interpretation
**Typical healthcare data environment**

### Delivery Network
- Hospitals
- Independent Physicians
- Community Clinics
- Specialty Care
- Rehabilitation
- Hospice

### Payers
- Private Insurers
- Medicare
- States
- Accountable Care Networks
- HMOs
- Third Party Administrators

### Research affiliates
- Medical/Nursing Schools
- Private Research Institutes
- Clinical Research partners

### External entities
- Private labs
- Suppliers
Data Architecture Landscape

Technology
- Systems
- Databases
- Information Exchange
- Security & Privacy

Data
- Structured data
- Unstructured data
- Semantics
- Data Integration

Organization
- Communication
- Governance
- Trust
1. Period covered, etc., Report request
2. Patient data – EMPI
3. Last visit – Registration
4. LDL values – Lab System
5. Vital Signs – EMR
6. Who is diabetic? Definitions from HEDIS
7. Comparison with peers – Computed summary data sets
8. Patient procedures - Claims
Current state of analytics

- Redundant investments in tools/data cleansing
- One off projects repeat development
- Multiple versions of truth
- Not enough resources for key projects
• In any environment:
  • Offloads analytic queries from transactional systems
  • Allows data to be organized for analytical purposes
  • Integrates data from different sources – “Single Source of Truth”

• In Healthcare
  • Collection is complicated by legacy data systems
  • Aggregation is often viewed as the goal
  • Integration is harder, but essential for answering important questions that must take financial, operational, and medical data into consideration
Data Warehouse Architecture
What we need is a data warehouse!
Why not a data warehouse?

• We tried a data warehouse – it failed
• We built a data warehouse – it works great, but it only has financial data in a cube
• A data warehouse is too expensive and will take too much time
• We already have several data warehouses!
Data Warehouse Architecture

- Sources
  - EMR
  - EMPI
  - Orders
  - Billing
  - Labs
  - Schedule

Data Warehouse

- Staging Database
  - Load programs
  - Correct
  - Map
  - Merge

- Data Warehouse
  - Process Control & Monitoring
    - Rules
    - Exceptions
    - Data Quality Metrics

- Interfaces
  - Analytic Data Mart
  - Disease-Specific Registry
  - Quality Reporting Data Mart

Recombinant
### Preventive Care Dashboard > Provider Performance [Institution: Progressive Health]

**Measure: Preventive Care and Screening: Colorectal Cancer Screening**

Note: Click on provider name to see patient list.

#### Reporting Period: 2H 2010 [Measure Year: 2010]

<table>
<thead>
<tr>
<th>Provider</th>
<th>Specialty</th>
<th>NPI</th>
<th>N</th>
<th>Score</th>
<th>* Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donnellson, Jessica</td>
<td>Family Medicine</td>
<td>124587455</td>
<td>2175</td>
<td>99.8%</td>
<td>18%</td>
</tr>
<tr>
<td>Marigold, Sally</td>
<td>Cardiology</td>
<td>124587452</td>
<td>360</td>
<td>92.8%</td>
<td>23%</td>
</tr>
<tr>
<td>Provider, James</td>
<td>Cardiology</td>
<td>876589898</td>
<td>133</td>
<td>91.3%</td>
<td>-6%</td>
</tr>
<tr>
<td>Jefferson, Lisa</td>
<td>Cardiology</td>
<td>876589892</td>
<td>1220</td>
<td>88.6%</td>
<td>38%</td>
</tr>
<tr>
<td>Bennett, William</td>
<td>Family Medicine</td>
<td>124587454</td>
<td>2528</td>
<td>79.3%</td>
<td>44%</td>
</tr>
<tr>
<td>Goodie, Margie</td>
<td>Cardiology</td>
<td>876589893</td>
<td>1758</td>
<td>76.0%</td>
<td>6%</td>
</tr>
<tr>
<td>Stephenson, Alexandra</td>
<td>Family Medicine</td>
<td>124587453</td>
<td>2119</td>
<td>75.3%</td>
<td>3%</td>
</tr>
<tr>
<td>Barrie, Johnathon</td>
<td>Family Medicine</td>
<td>876589894</td>
<td>2413</td>
<td>62.3%</td>
<td>17%</td>
</tr>
<tr>
<td>Bonds, Barry</td>
<td>Family Medicine</td>
<td>124587456</td>
<td>3659</td>
<td>60.6%</td>
<td>-24%</td>
</tr>
<tr>
<td>McCallister, Joseph</td>
<td>Cardiology</td>
<td>876589891</td>
<td>1694</td>
<td>59.1%</td>
<td>-37%</td>
</tr>
</tbody>
</table>

* Indicates % change from prior reporting period: 1H 2010.

#### Provider Score Metrics

**Frequency Distribution**

(Excludes Providers with N < 5)

![Bar Chart](image)

**Score Percent Range**

**Distribution Profile:**

- Mean: 78.50
- Median: 77.63
- Std Dev: 14.52
- Skewness: 0.07
- Provider Min: 59.15
- Max: 99.82
# Incentive Scorecard > Measure Patients Summary

## Measure [5]
Heart Failure: Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)

### Definition

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>Age</th>
<th>Service Dt</th>
<th>Numerator Codes</th>
<th>Denom Codes</th>
<th>Performance Criteria</th>
<th>Participation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1, Male</td>
<td>M</td>
<td>44</td>
<td>06/01/2010</td>
<td>4009F</td>
<td>4280, 99214</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 15, Male</td>
<td>M</td>
<td>44</td>
<td>06/01/2010</td>
<td>4009F with 2P</td>
<td>4254, 99239</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 16, Female</td>
<td>M</td>
<td>44</td>
<td>06/01/2010</td>
<td>4009F with 2P</td>
<td>4254, 99239</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 17, Male</td>
<td>M</td>
<td>44</td>
<td>06/01/2010</td>
<td>4009F</td>
<td>99238, 42822, 4259, 99203</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 18, Female</td>
<td>M</td>
<td>44</td>
<td>06/01/2010</td>
<td>4009F</td>
<td>99238, 42822, 4259, 99203</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 19, Male</td>
<td>M</td>
<td>38</td>
<td>06/01/2010</td>
<td>4009F</td>
<td>4254, 99213</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 2, Female</td>
<td>F</td>
<td>38</td>
<td>06/01/2010</td>
<td>3021F</td>
<td>4254, 99214</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 28, Male</td>
<td>M</td>
<td>44</td>
<td>06/01/2010</td>
<td>4009F</td>
<td>4254, 4280, 99214</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 29, Female</td>
<td>M</td>
<td>44</td>
<td>06/01/2010</td>
<td>4009F</td>
<td>4254, 4280, 99214</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 3, Male</td>
<td>M</td>
<td>38</td>
<td>06/01/2010</td>
<td>3021F</td>
<td>4254, 99214</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient 34, Male</td>
<td>M</td>
<td>44</td>
<td>06/01/2010</td>
<td>3022F</td>
<td>99214, 4254</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
## PQRI Summary - Current Period (01/01/2010 thru 05/03/2010)

<table>
<thead>
<tr>
<th>Code</th>
<th>Measure</th>
<th>% Participation</th>
<th>% Performance</th>
<th>Quality Information Available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Diabetes Mellitus: Hemoglobin A1c Poor Control in Diabetes Mellitus</td>
<td>78% (n=200)</td>
<td>71% (n=200)</td>
<td>80% (n=126)</td>
</tr>
<tr>
<td>2</td>
<td>Diabetes Mellitus: Low Density Lipoprotein (LDL-C) Control in Diabetes Mellitus</td>
<td>51% (n=200)</td>
<td>53% (n=200)</td>
<td>54% (n=160)</td>
</tr>
<tr>
<td>4</td>
<td>Heart Failure: Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)</td>
<td>87% (n=296)</td>
<td>62% (n=297)</td>
<td>62% (n=92)</td>
</tr>
</tbody>
</table>

## Good Citizenship Measures

<table>
<thead>
<tr>
<th>Metric</th>
<th>Current</th>
<th>Target</th>
<th>Earned Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIM Related Emails</td>
<td>9</td>
<td>N/A</td>
<td>$500</td>
</tr>
<tr>
<td>Clinical Training Completed</td>
<td>0</td>
<td>N/A</td>
<td>$0</td>
</tr>
<tr>
<td>Comments Submitted</td>
<td>0</td>
<td>N/A</td>
<td>$0</td>
</tr>
<tr>
<td>Suspensions</td>
<td>0</td>
<td>N/A</td>
<td>$500</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>--</td>
<td>--</td>
<td><strong>$1,000</strong></td>
</tr>
</tbody>
</table>
Data Warehouse Architecture

- Data warehouse is normalized for integration
  - Master data cleaned and integrated
  - Coding systems and other semantics used to validate data
  - Exception management for managing errors
- Data Marts created for defined needs
  - Report-specific models for production reports
  - Dimensional models for ad hoc reporting and analytics
  - Research marts with IRB approval
- Additional Interfaces
  - Health Information Exchange
Limits to dimensional modeling

• Dimensional modeling has its place in healthcare – but it is insufficient:
  • No single source system will answer most institutional level questions
  • Integration is required before dimensional modeling can even begin

• A dimensional modeling approach does not work well for clinical data:
  • Data matrix is sparsely populated
  • Many different “roll-ups” are needed – by provider, by payer, by disease
Data Integration

Collection

Operational Data Store

Source Data
- EMR
- EMAR
- Billing
- Claims
- Labs
- Scheduling

Collection

Patient Data Aggregation
- Operational Reporting systems
- HIE

Aggregation

Mapping/Pointers

Data Warehouse

Patient Problems (part of the patient subject)

Integration

Ontologies, Coding, Data Quality

Diagnosis - ICD9

Meds - Insulin

Lab - A1C>9
Data Quality

• “Everything is exactly as entered in the source system”
  • Easy to measure
  • Easy to achieve
  • Not very valuable

• “Data represents the truth as best it can”
  • Easy to measure – but you need help from clinicians and other users
  • Hard to achieve – a constant process of improvement for users and data analysts
  • Extremely valuable – a source of data that represents the truth as best as it can
Data Quality

- Good Data Quality: Institutionally accepted
- Better Data Quality: Individually accepted
- Best Data Quality: Globally accepted

**Goal:** Data that can support difficult decisions
The Subject Area Model
Patient Model

Patient
- Patient Identifiers
  - Master Patient Index

Attributes
- Problems
- Medications
- Allergies
- Biomarkers / 'Omics
- Demographics
- Socioeconomics

Medical Interactions
- Appointments
- Registration
- Patient Flow
- Biomaterial Collection
- Orders

Financial Interactions
- Bills
- Revenues
- Insurance Membership
- Costs

Services
- Observations
  - Vital Signs
  - Narratives
  - Images
  - Test Results
  - Incidents
  - Patient Reported Outcomes

Interventions
- Medications
- Procedures
- Vaccinations
- Radiology
Additional Models

• Financial
  • Revenue focus
  • Cost focus
• Chronic Diseases
  • Diabetes
  • Heart Disease
• Research registries
  • Cancer
Case: UMass Memorial Health Care

• 800 Bed Tertiary Care
• Academic Medical Center
• Pay-for-performance Challenges:
  • What can we measure?
  • What can we improve?
  • What can we attribute?
• Data Environment
  • Legacy Meditech ordering systems
  • Implementing Allscripts
  • Payer data from paid claims
• Centralized data in a single MS SQL Server Database
• Highly secure environment (2 factor security, auditing, and limited users)
• Integration process using open source (Pentaho Kettle) and SQL processes
• Data validated against reference tables (ICD-9, etc.)
• Data checked with doctors before rollout
• Internal process challenges (changes to lab codes)
Governance – a longer discussion

**Capture**
- Accurate
- Comprehensive
- Standards Based
- Timely

**Usage**
- Business Intelligence
- Performance Management
- Research

**Management**
- Integrity
- Security
- Integration
- Performance

**Definitions/Models**

**Security/Privacy**

**Policies**

**Manage**

**Capture**
Lessons Learned

• Engage top level executives – data governance cannot be resolved at the departmental level
• There are “no magic beans” – integration is still a time consuming challenge
• Perfection is the enemy of the good – Best data possible, not perfection
• Constant monitoring and improvement
• Incremental approach was key to keeping costs acceptable
Incremental Approach

Clinical Data
Conclusion: Analytics & Data Architecture

• A good data architecture ensures data that can be a foundation for institutional reporting

• A better data architecture provides attribution so that the links from cause to effect can be established and process can be improved

• The best data architecture connects institutional data to environmental data and partner institutions to link providers, payers, and other parties, so that risk can be identified and mitigated through improved care to patients

• Don’t let the Best architecture be a reason for not having a Good architecture!
Questions and Discussion
Thank You

Aaron Abend
aabend@recomdata.com