Data Integrity and Governance

What should we do and how do we get there?

Andy Tauber MCSE, MCDBA
Data Architect
Snapshot of The Vancouver Clinic

- Physician-owned and governed, professionally managed
- Founded 1936
- 265 providers
- 990 employees
- 5 sites
Data Growth

• Byte = 1 Character
• Kilobyte = 1,000 Characters
• Megabyte = 1,000,000 Characters
• Gigabyte = 1,000,000,000 Characters
• Terabyte = 1000 Gigabytes (1 Trillion bytes)
• Petabyte = 1000 Terabytes ($10^{15}$)
• Exabyte = 1000 Petabytes ($10^{18}$)
• Zetabyte = 1,000,000,000 Terabytes ($10^{21}$)
Data Growth

THE SPHERES OF BIG DATA ARE CONVERGING

30 billion pieces of data are added to Facebook each minute

72 hours of video are added to YouTube each minute

$600 buys you a disk that stores all music on Earth

92% of the world’s data was created in the past 2 years

DATA AGGREGATED BY GONICUS RESEARCH
Data Growth

DATA GROWTH RATE

Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020.

One Zetta Byte (ZB) = 1000 Exa Bytes = 1 Billion Terra Byte (TB)

Source: Oracle, 2010
Data Growth in Healthcare

Total Data Healthcare Providers (PB)

Medical Imaging Archive Projection
Case from just 1 healthcare system

Data Explosion projected to reach 35 Zetabytes by 2020, with a 44-fold increase from 2009^5
Cost of Bad Data

Business cost of bad data may be as high as 10-25% of an organization’s revenue.

REVENUE

A BUSINESS PROBLEM

$314 billion:
The cost of bad data in healthcare.

A HEALTHCARE PROBLEM

Bad data costs the U.S. economy over $3 trillion a year - over twice the amount of the 2011 Federal Deficit:

COST OF BAD DATA

AN ECONOMIC PROBLEM

HOW CAN BAD DATA COST MONEY?

A lack of visibility into the right data caused one major retailer to lose more than $3 million annually.

As much as 50% of a typical IT budget may be spent in "information scrap and rework."

The average company wastes $180,000 per year simply on direct mail that does not reach the intended recipient because of inaccurate data.
The Legal Climate

EMRs can hurt physicians during lawsuits

Lawyers smell blood in electronic medical records

Lousy Electronic Medical Records Fuel Successful Lawsuits
Testimony

• Plaintiff attorney: Doctor, if the emergency renal consult was called in at 11:30, why did you wait until 6 p.m. to see the patient, during which time his kidneys became severely damaged?

• Doctor: I did see the patient within 30 minutes.

• Plaintiff attorney: Where does it show that in the chart?

• Doctor: Uh … it doesn’t, I guess. I saw the patient but wrote the note later.

• Plaintiff attorney: So you claim you saw this critically ill patient for 30 minutes, spent one hour evaluating him, but did not document your findings for another six hours?
Horror Stories

This SSN is already in the system.

Just add 1 to it.

Change chart name to file claim

Twins on same chart
Our Process

• Data Flow
• Data Inventory
  • Servers
  • Databases
  • PHI
• Report Inventory
• Current State Assessment
  • Maturity Model Assessment
  • Maturity Model
• Define Governance Structure
Our Process

• Training
  • Management
  • DG Council Members
  • Data Stewards
• Define Subject Area Priorities
• Begin Implementation
• Monitor
# Maturity Model Assessment - Foundational

## Data Governance Maturity Model

### Guiding Questions for each Component-Dimension

<table>
<thead>
<tr>
<th>Component</th>
<th>People</th>
<th>Policies</th>
<th>Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness</td>
<td>What awareness do people have about their role within the data governance program?</td>
<td>What awareness is there of data governance policies, standards and best practices?</td>
<td>What awareness is there of data governance enabling capabilities that have been purchased or developed?</td>
</tr>
<tr>
<td>Formalization</td>
<td>How developed is the data governance organization and which roles are filled to support data governance activities?</td>
<td>To what degree are data governance policies formally defined, implemented and enforced?</td>
<td>How developed is the toolset that supports data governance activities and how consistently is that toolset utilized?</td>
</tr>
<tr>
<td>Metadata</td>
<td>What level of cross functional participation is there in the development and maintenance of metadata?</td>
<td>To what degree are metadata creation and maintenance policies formally defined, implemented and enforced?</td>
<td>What capabilities are in place to actively manage metadata at various levels of maturity?</td>
</tr>
</tbody>
</table>
# Maturity Model Assessment - Project

<table>
<thead>
<tr>
<th>Project</th>
<th>People</th>
<th>Policies</th>
<th>Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stewardship</td>
<td>To what degree have stewardship roles been defined and filled?</td>
<td>To what degree are stewardship policies defined, implemented and enforced?</td>
<td>What capabilities are implemented to support the effective stewardship?</td>
</tr>
<tr>
<td>Data Quality</td>
<td>To what degrees have data quality competencies developed?</td>
<td>To what degree are data quality policies defined, implemented and enforced?</td>
<td>What capabilities are implemented to support the production and maintenance of high quality data?</td>
</tr>
<tr>
<td>Master Data</td>
<td>To what degree has a formal master data management organization been developed and assigned consistent responsibilities across data domains?</td>
<td>To what degree are master data policies defined, implemented and enforced?</td>
<td>What capabilities are available and implemented to actively master and provision master data?</td>
</tr>
</tbody>
</table>
# Maturity Model Assessment

<table>
<thead>
<tr>
<th>Maturity Level</th>
<th>Awareness</th>
<th>Formalization</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Limited awareness of purpose or value of Data Governance program.</td>
<td>There are no defined roles related to data governance.</td>
<td>Limited understanding of types and value of metadata.</td>
</tr>
<tr>
<td>2</td>
<td>Executives are aware of existence of program. Little knowledge of program outside upper management.</td>
<td>Data governance roles and responsibilities have been defined and vetted with program sponsors.</td>
<td>All roles responsible for production of technical metadata on structured data are defined during system design.</td>
</tr>
<tr>
<td>3</td>
<td>Executives understand how program benefits/impacts their portion of the organization, knowledge workers are aware of program. Executives actively promote program within their groups.</td>
<td>Common institutional data policies are documented and available through a common portal.</td>
<td>Metadata best practices are produced and made available. Most best practices are focused on the metadata associated with structured data.</td>
</tr>
<tr>
<td></td>
<td>Most existing data policies are undocumented and there may be inconsistent understanding of data policies within a department.</td>
<td>A small subset of the organization understands the general classes of data governance capabilities and technologies.</td>
<td>Metadata templates are adopted to provide some consistency in content and format of captured metadata. Metadata is consolidated and available from a single portal.</td>
</tr>
<tr>
<td></td>
<td>There is little awareness of data governance capabilities and technologies.</td>
<td>High-level data governance meta-policies are defined and distributed.</td>
<td>Capabilities focus on capture of metadata of structured content.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>People</th>
<th>Policies</th>
<th>Capabilities</th>
<th>People</th>
<th>Policies</th>
<th>Capabilities</th>
<th>People</th>
<th>Policies</th>
<th>Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Data Governance Foundational Component Maturity**

- **Limited understanding of types and value of metadata.**

- **Metadata templates are adopted to provide some consistency in content and format of captured metadata.**

- **Capabilities focus on capture of metadata of structured content.**
Maturity Model

Source: Gartner (December 2008)
Structure

Board of Directors

Data Owners

Executive Leadership

Decision Making Input

Data Governors

Policy Decisions
Requirements Definition

Data Stewards

User Acceptance Testing

End Users, Customers, etc.

Each governor represents an interest group and line of business within the organization and makes policy decisions on behalf of the interests and the enterprise.

This ensures clear accountability for all aspects of data governance within each line of business as well as across the entire organization.
Structural Evolution
Documents

• Charter
  • Vision
  • Mission Statement
  • Capabilities
  • Roles and Responsibilities
• Data Governance Checklist
  • Major Items Needed
Documents

• Job Descriptions
  • Data Steward
  • Data Owner
  • Data Custodian
  • Data Governance Council
Framework

**Discover**
- Data Discovery
- Data Profiling
- Data Inventories
- Process Inventories
- CRUD Analysis
- Capabilities Assessment

**Define**
- Business Glossary Creation
- Data Classifications
- Data Relationships
- Reference Data
- Business Rules
- Data Governance Policies
- Other Dependent Policies
- Key Performance Indicators

**Measure**
- Proactive Monitoring
- Operational Dashboards
  - Reactive Operational DQ Audits
  - Dashboard Monitoring/Audits
- Data Lineage Analysis
- Program Performance
- Business Value/ROI

**Apply**
- Automated Rules
- Manual Rules
- End to End Workflows
- Business IT Collaboration
Training

• Executive Team
  • What is it?
  • Governance as part of the System (manual or computer) Development Lifecycle
  • Setting Data Governance Priorities

• Data Governance Council
  • What is it?
  • Roles and Responsibilities
Training

• Data Stewards
  • Responsibilities
  • Tools
  • Empowerment
Goals

• Standardize Data Definitions
• Improve Data Quality
• Enable better decision-making
• Reduce operational friction
• Protect the needs of data stakeholders
• Train management and staff to adopt common approaches to data issues
• Build standard, repeatable processes
• Reduce costs and increase effectiveness through coordination of efforts
• Ensure transparency of processes
Initial Steps

• Data Infrastructure Map
• Data Inventory
• Report Inventory
• Data Flow
Database Infrastructure
Flow of Information
What We Did

• Don’t Boil the Ocean (Bite sized chunks)
• Prioritize Subject Areas
  • Patient Demographics
  • Provider Data
  • Insurance and Patient Accounts
• Identify Data Owners
• Identify Data Stewards
Organizational Changes

• Initially
  • PSRs reported to department they worked in.
  • Loyalty was to department.
  • Operated as independent islands.

• Now
  • At least one PSR Supervisor in each location
  • PSR Supervisor report to Data Owner
  • PSR Supervisors held accountable for data quality
Empowerment
Process

• Weekly Meeting
  • Attendees
    • Data Stewards
    • Data Architect
    • Trainer
    • EPIC Analysts as needed
    • Security and Compliance as needed
  • Training

• Now we meet monthly
Process

• Identify data issues
  • Inconsistent procedures
  • Exceptions handling
  • EPIC issues
    • Hard Stops
    • Screen Design
    • Reports
  • Time Consuming Issues
Meeting Process

- Prioritize
- Discussion
- Research
- Negotiation
- Outcome
  - Policy
  - Procedure
  - Software Change
Accomplishments

• No longer collect Social Security Numbers
• Patient Identity Policy and Procedure
• Don’t make up data policy
• Employee update to medical record policy
Next Steps

• Define measurements to monitor success
• Automate manual processes
• Identify additional subject areas
• Continue transition to a data centric culture
  • Initial and ongoing training
  • Communication
  • Measure
Tools
Tools
<table>
<thead>
<tr>
<th><strong>Age</strong></th>
<th>Business Terms</th>
<th>In this context age refers to a patient's demographic information. Time elapsed from date of...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Confidential Address</strong></td>
<td>Business Terms</td>
<td>The address held in strict privacy or secrecy. An example of when to use a confidential...</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td>Business Terms</td>
<td>The facts that make up the characteristics of a person or certain group of people. At...</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td>Business Terms</td>
<td>Identity with or membership in a particular racial, national, or cultural group and observance...</td>
</tr>
<tr>
<td><strong>Guarantor Address</strong></td>
<td>Business Terms</td>
<td>The address where financial correspondence is sent.</td>
</tr>
<tr>
<td><strong>Hyperlink</strong></td>
<td>Business Terms</td>
<td>A link, phrase, picture, or icon in a document on which a user may click to move to another...</td>
</tr>
<tr>
<td><strong>Marital Status</strong></td>
<td>Patient Demographics</td>
<td>The state of being married, unmarried, or separated as defined by applicable state law. At...</td>
</tr>
<tr>
<td><strong>PAT_NAME (EpicCombinedModel.dm1, Logical, PATIENT)</strong></td>
<td>Attribute (Linked Term)</td>
<td></td>
</tr>
</tbody>
</table>
Tools

**PAT_NAME**

Attribute
EpicCombinedModel.dm1 > Logical > PATIENT > PAT_NAME
Linked to Term PAT_NAME (EpicCombinedModel.dm1, Logical, PATIENT) [Delete Linked Term]

⚠️ This information is highly confidential and is intended for internal use only. Authorization must be obtained to release.

<table>
<thead>
<tr>
<th>General Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute Name</td>
</tr>
<tr>
<td>Column Name</td>
</tr>
<tr>
<td>Logical Only</td>
</tr>
<tr>
<td>Role Name</td>
</tr>
<tr>
<td>Logical Role Name</td>
</tr>
<tr>
<td>Physical Role Name</td>
</tr>
<tr>
<td>Composite Datatype</td>
</tr>
<tr>
<td>Primary Key</td>
</tr>
<tr>
<td>Foreign Key</td>
</tr>
<tr>
<td>Check Constraint</td>
</tr>
<tr>
<td>Identity Ind</td>
</tr>
<tr>
<td>Identity Increment</td>
</tr>
</tbody>
</table>
# Tools

**Definition**

The patient's name in the format Lastname, Firstname MI.

**Security Properties**

<table>
<thead>
<tr>
<th>Name</th>
<th>Data Type</th>
<th>Value</th>
<th>Dictionary</th>
<th>Security Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy Level</td>
<td>Test List</td>
<td>Internal Use Only</td>
<td>Epic Data Dictionary</td>
<td>Data Security Classification</td>
</tr>
<tr>
<td>Compliance Mapping</td>
<td>Test List</td>
<td>HIPAA</td>
<td>Epic Data Dictionary</td>
<td>Compliance</td>
</tr>
<tr>
<td>Security Impact</td>
<td>Test List</td>
<td>Critical to Business</td>
<td>Epic Data Dictionary</td>
<td>Data Security Classification</td>
</tr>
</tbody>
</table>

**Where Used**

<table>
<thead>
<tr>
<th>Bound Object</th>
<th>Object Type</th>
<th>User Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>EpicCombinedModel.cdm &gt; Epic Combined Physical Model &gt; PATIENT &gt; PAT_NAME</td>
<td>Column</td>
<td>False</td>
</tr>
</tbody>
</table>

**Submodels**

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Events</td>
<td>View Image</td>
</tr>
<tr>
<td>Access Statistics</td>
<td>View Image</td>
</tr>
<tr>
<td>Account</td>
<td>View Image</td>
</tr>
<tr>
<td>Accounts Receivable</td>
<td>View Image</td>
</tr>
<tr>
<td>Allergies</td>
<td>View Image</td>
</tr>
<tr>
<td>Alternative Alerts</td>
<td>View Image</td>
</tr>
<tr>
<td>Alternatives</td>
<td>View Image</td>
</tr>
<tr>
<td>Appointments</td>
<td>View Image</td>
</tr>
<tr>
<td>Availability</td>
<td>View Image</td>
</tr>
<tr>
<td>Best Practice Alerts</td>
<td>View Image</td>
</tr>
<tr>
<td>Capitalization</td>
<td>View Image</td>
</tr>
<tr>
<td>Case Management</td>
<td>View Image</td>
</tr>
<tr>
<td>Chart Correction</td>
<td>View Image</td>
</tr>
<tr>
<td>Chart History</td>
<td>View Image</td>
</tr>
<tr>
<td>Chart Statistics</td>
<td>View Image</td>
</tr>
<tr>
<td>CLAIMS</td>
<td>View Image</td>
</tr>
<tr>
<td>Clinical Documentation Improvement</td>
<td>View Image</td>
</tr>
<tr>
<td>Coding and Abstracting - Accounts</td>
<td>View Image</td>
</tr>
<tr>
<td>Coding and Abstracting - Clinical</td>
<td>View Image</td>
</tr>
<tr>
<td>Concept Unique Identifier (CUJ)</td>
<td>View Image</td>
</tr>
<tr>
<td>CRM</td>
<td>View Image</td>
</tr>
<tr>
<td>Diagnosis/Reason for Visit</td>
<td>View Image</td>
</tr>
<tr>
<td>Drug Interaction Alerts</td>
<td>View Image</td>
</tr>
<tr>
<td>ENROLLMENT</td>
<td>View Image</td>
</tr>
<tr>
<td>Episode</td>
<td>View Image</td>
</tr>
</tbody>
</table>
Tools

• The tool is not important
• COMMUNICATION IS
What It Takes To Get This Done

COMMITMENT
COMMITMENT Defined

Think of the Success of this Breakfast

Involved

Committed
Questions