Big Data Governance

Sunil Soares
Founder & Managing Partner
Information Asset, LLC
+1-201-693-2216
@sunilsoares1
www.information-asset.com
sunilsoares@information-asset.com
1. Banking and Financial Markets
2. Insurance
3. Healthcare
4. Manufacturing
5. Retail
6. Travel and Transportation
7. Government
8. Oil and Gas
9. Telecommunications
10. Utilities
Convergence of Data Governance, Big Data, MDM, BPM

Part I: Big Data Governance
1. Introduction
2. Big Data Governance Framework
3. IBM Big Data Platform
4. Big Data Integration
5. Metadata
6. Big Data Security & Privacy
7. Big Data Quality
8. Master Data Integration
9. Managing the Lifecycle of Big Data

Part II: Process Data Governance
10. Introduction
11. Retail
12. Oil and Gas
13. Healthcare
Big Data

Volume

Data at Rest

Velocity

Data in Motion

Variety

Data in Many Formats
Big Data Governance

Big data governance is part of a broader data governance program that formulates policy relating to the optimization, privacy, and monetization of big data by aligning the objectives of multiple functions.
Big Data Governance Involves Leveraging Information by Aligning the Objectives of Multiple Functions

- Telecommunications carriers have access to vast troves of customer information in the form of Call Detail Records and Geolocation data
- Marketing might seek to monetize this information by selling to third parties or selling value added services such as location-specific coupons
- Privacy, Legal and the Wireline business will want to protect the core business from reputational risk
Big Data Governance Framework

Industries and Functions

- Healthcare
  - Sentiment Analysis
  - Patient Monitoring
  - Claims Analytics
  - Genetic Testing
  - Electronic Medical Records

- Utilities
  - Smart Meters

- Retail
  - Facebook Loyalty Programs
  - RFID Tags
  - Facial Recognition

- Telcos
  - Customer Churn Analytics
  - Location-Based Services
  - Customer Churn Analytics

- Insurance
  - Investigate Claims
  - Vehicle Telematics
  - Claims Fraud Analytics
  - Underwriting

- Customer Service
  - Call Quality Assurance

- IT
  - IT Log Analytics

Big Data Types

- Web & Social Media
- Machine-to-Machine
- Big Transactional Data
- Bio-metrics
- Human Generated

Governance Disciplines

- Organization
- Metadata
- Privacy
- Data Quality
- Business Process Integration
- Master Data Integration
- Information Lifecycle Management

© 2013 Information Asset
Big Data Governance and the Mars Climate Orbiter

- In 1999, a navigation error sent the satellite into an orbit 170 kilometers lower than the intended altitude above Mars.
- NASA engineers used English units (pounds) instead of metric units (newtons).
- Loss of $328 million for the orbiter and lander plus setback to space exploration.

Source: NASA Jet Propulsion Laboratory/Corby Waste
Use of Social Media Data in Insurance

- Claims investigations
  - Woman filed insurance claim declaring hit-and-run auto accident but Facebook comments indicated that her daughter was responsible for accident
  - She was convicted of filing a fraudulent claim
- Underwriting to set pricing on policies
  - Can a life insurer use an applicant’s Facebook profile indicating that she is a skydiver to increase premiums?
  - Most U.S. states used to restrict the use of credit scores for underwriting, but that has changed (California is one exception)
  - Most U.S. states have regulations that restrict the use of social media for underwriting purposes, but that may change
Big Data Types in Healthcare

Web and Social Media
- Clickstream Data
- Twitter Feeds

Biometrics
- Employee Credentialing
- Genetics

Machine-to-Machine
- Patient Monitoring

Human Generated
- Call Center Voice Recordings
- Electronic Medical Records

Big Transaction Data
- Claims Records
Use of Social Media Data in Banking

- **Credit**
  - Use of social media for credit decisions may be impacted by regulations such as the U.S. Fair Credit Reporting Act that limit data collection

- **Collections**
  - Social media may be used to get up-to-date contact information for skip tracing
  - Collectors must adhere to regulations such as the U.S. Fair Debt Collection Practices Act that are designed to protect the debtor’s privacy and to avoid harassment
Leveraging Twitter Data in Retail

@acme123: Best day ever, went mountain biking today, will get one for my husband too.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Inferred Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Married</td>
</tr>
<tr>
<td>Sports</td>
<td>Biking</td>
</tr>
<tr>
<td>Age</td>
<td>25-55 (since she is married and engaged in active sports)</td>
</tr>
</tbody>
</table>
# Twitter Data Quality

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Data Quality Concerns</th>
<th>Business Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweet Timestamp</td>
<td>Format of timestamps in Tweets is inconsistent with the standard format, which can cause when joining other datasets</td>
<td>• Reformat all timestamps to YYYY-MM-DD HH:MM:SS</td>
</tr>
<tr>
<td>References to Acme Corporation</td>
<td>Is the Tweet about Acme Corporation or is it noise that needs to be filtered out?</td>
<td>• If Tweet contains “@Acme” then confidence level = 99 percent</td>
</tr>
<tr>
<td>Location</td>
<td>Location data from Tweets is important and can be used to answer questions such as, “Are users in the southeastern United States more likely to be disgruntled than those in other parts of the country?”</td>
<td>• If Tweet contains “Acme” and Acme product names then confidence level = 75 percent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If Tweet is on the ignore list then confidence level = zero percent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Extract the Tweet.user.location string from the Tweet metadata and validate the city and state names</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If the location is not recognizable, then flag it as unknown</td>
</tr>
</tbody>
</table>
How Representative is Twitter Data?

- Customer demographics
  - Predominantly female
  - Age 30-44 years
  - Average household income in excess of $100,000

- CEO to Listening Department
  - “How representative is Twitter sentiment analysis relative to our customer base?”
  - “I expect Twitter users to be younger and less affluent than our target customers”

- Listening department ran marketing surveys and found that Twitter users were actually representative of retailer’s customers
Overview of Smart Meters

- Traditional water, gas and electricity meters measure consumption on a gross basis every month or quarter.
- Smart meters typically capture usage data every 15-60 minutes for residential and commercial customers.
- Smart meters can lower peak electricity demand by incenting consumers to move usage to off-peak hours through differentiated pricing.
- Utilities may be able to build less generation capacity.
- Data governance implications of smart meter “interval” data:
  - Explosion of data volumes
  - Personally Identifiable Information (PII) due to the ability to discern household activities.
Regulations Governing Smart Meter Data in Utilities

- **European Union Article 29 Data Protection Working Party**
  - Privacy is key because smart meter readings can be used to profile household patterns
  - Data is considered personal data since they include a unique identifier
  - Privacy should be designed at the start to minimize the amount of personal data processed
  - Privacy Impact Assessments
  - Data retained only for as long as necessary

- **United States**
  - State Public Utility Commissions such as California also adopting rules to protect privacy
Smart Meter Data Governance Policies

- Data Access – Establish policies around which departments can access interval data which constitutes Personally Identifiable Information (PII)
- Database Monitoring – Monitor access by privileged users to “interval data” that pinpoints customers’ activities
- Archiving and Retention – Establish archiving and retention policies to reduce storage costs associated with vast quantities of data
- Metadata – Need consistent definitions for terms such as “outage”
Utility Smart Meters

• Database Monitoring
  — Monitor access by privileged users to “interval data” that pinpoints customers’ activities

• Archiving and Retention
  — Establish archiving and retention policies to reduce storage costs associated with vast quantities of data

• Metadata
  — Need consistent definitions for terms such as “outage”
Human-Generated Data Reduce 30-Day Readmission Rates for Patients with Congestive Heart Failure

- Hospital system with 15 facilities and significant indigent population
- Big Data Analytics to reduce 30-day readmission rate for patients with congestive heart failure
- Reduction in costs not reimbursed by insurance
- Increase in quality of care based on proactive early intervention
- Predictive model based on 150 variables and 20,000 patient encounters over five years
- Data sourced from multiple applications
Leverage Unstructured Data to Improve the Quality of Structured Data for Smoking Status and Drug & Alcohol Abuse

- Only 25% of records initially populated with Yes/No answers for “Smoking Status”

- Content analytics revealed interesting results
  - “Patient is restless and asked for a smoking break”
  - Patient quit smoking yesterday”
  - “Quit”

- Based on content analytics, the team was able to increase the population rate for “Smoking Status” to 85%

- Drug & Alcohol Abuse also increased from 20% to 76% of records
Extract Additional Relevant Clinical Factors Not Found in Structured Data

- Assisted Living Facility
- Pharmacology Compliance Indicator
- Variables were not captured in structured data
- Analytics team used text analytics from discharge summaries, echocardiograms, patient histories, doctors’ notes and physicals to gather data

photo: Creative Commons / bitzi
Business Definitions for “Readmission”

- Clinical Perspective – 30 days, all causes
- Clinical Perspective – 30 days, same diagnosis
- Finance Perspective – Quarterly and Annually

Photo: KOMU Photos/Eric Staszczak via Flickr/Creative Commons
Inconsistency in Patient Master Data Across Facilities

- Lack of Enterprise Master Patient Index across hospital system made it difficult to manage allergies with potentially life-threatening consequences

- Admission #1
  - Name: Jim Smith
  - Patient Record: 0A7654
  - Date: August 12, 2010

- Admission #2
  - Name: Jim Smith
  - Patient Record: 7B9871
  - August 31, 2010
Enterprise Master Patient Index at Large U.S. Health System

- Hospital 1: Epic
- Hospital 2: Cerner
- Hospital 3: Eclipsys
- Hospital 4: Cerner
- Hospital 5: Eclipsys
- Hospital 6: Epic
- Hospital 7: Eclipsys

**EMPI**

- Name
- Date of Birth
- Social Security Number
- Driver’s License Number
- Race
- Religion
- VIP Indicator
- Home Phone
- Address
- Insurance Information
- Emergency Contacts
- Guarantor Information

- Admit
- Discharge
- Transfer

- Physician Portal

- Enterprise Data Warehouse

- Radiology

- Cath Lab
HIPAA Compliance

- Optimize the use of expensive paramedic services
- Patient address was Protected Health Information (PHI)
- Compliance with HIPAA by requesting patient approvals before using their address in research
Retailer uses innocuous data to identify an expectant teenager before her father knew about her pregnancy

- Advanced analytics team found that pregnant women bought larger quantities of unscented lotion, special dietary supplements, scent-free soap, and large bags of cotton balls.
- Marketing team used this pregnancy-prediction model to send offers to tens of thousands of female shoppers all over the United States.
- About a year later, a man walked into a store: “My daughter got this in the mail! She’s still in high school, and you’re sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?”
- A few days later, the father said: “I had a talk with my daughter. She’s due in August. I owe you an apology.”

Source: New York Times
Data Governance Policies

Enterprise Data Governance Policies
1. Data Ownership
2. Organization
3. Data Architecture & Modeling
4. Master Data Management
5. Reference Data Management
6. Metadata Management
7. Data Quality Management
8. Data Security Management

Domain & Organization Specific Data Governance Policies, e.g.,
9. Customer On-boarding in Corporate Banking
10. Email Address in Insurance
11. Role of Internal Audit
# EDM Data Policy Dashboard by Discipline

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Policies</th>
<th>Standards</th>
<th>Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Roles</td>
<td><img src="logo.png" alt="Developed" /></td>
<td><img src="logo.png" alt="Developed" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
<tr>
<td>2. Metadata</td>
<td><img src="logo.png" alt="Developed" /></td>
<td><img src="logo.png" alt="Defined" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
<tr>
<td>3. Data Quality Management</td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Defined" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
<tr>
<td>4. Data Integration</td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
<tr>
<td>5. Master Data Management</td>
<td><img src="logo.png" alt="Defined" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
<tr>
<td>6. Data Modeling</td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
<tr>
<td>7. Data Security Management</td>
<td><img src="logo.png" alt="Developed" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
<tr>
<td>8. Reference Data Management</td>
<td><img src="logo.png" alt="Developed" /></td>
<td><img src="logo.png" alt="Defined" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
<tr>
<td>9. Data Architecture</td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
<td><img src="logo.png" alt="Unidentified" /></td>
</tr>
</tbody>
</table>

**LEDGEN**
- **Developed**
- **Defined**
- **Unidentified**
# Top 50 Data Repositories Dashboard Example

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Med</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**LEGEND**
- **Developed**
- **Defined**
- **Unidentified**
Data Ownership Policies – Role of Data Owner

- Business Glossary
- Critical Data Elements
- Sensitive Data
- Data Access Requests
- Data Quality
- Data Policies, Rules & Standards
- Data Stewards
- Answerable to Internal Audit

<table>
<thead>
<tr>
<th>Data Repository</th>
<th>Data Domain</th>
<th>Owning Organization</th>
<th>Data Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>Customer</td>
<td>Marketing</td>
<td>John Doe</td>
</tr>
<tr>
<td>XYZ</td>
<td>Household</td>
<td>Marketing</td>
<td>Sally Smith</td>
</tr>
<tr>
<td>DEF</td>
<td>Product</td>
<td>Risk</td>
<td>Jane Duffy</td>
</tr>
</tbody>
</table>
RACI Matrix

- **Responsible**
  - Those who do the work to achieve the task

- **Accountable**
  - Approver
  - The one ultimately answerable for the correct and thorough completion of the deliverable or task, and the one from whom *Responsible* is delegated the work

- **Consulted**
  - Those whose opinions are sought, typically subject matter experts and with whom there is two-way communication

- **Informed**
  - Those who are kept up-to-date on progress, often only on completion of the task or deliverable
  - Those with whom there is just one-way communication
Retail Banking – RACI for Date of Birth

- **Accountable**
  - Customer Information Management

- **Responsible**
  - Operations

- **Consulted**
  - Compliance
  - Marketing
  - Internet Banking

- **Informed**
  - Finance

Source:
Creative Commons
## Retail Banking – RACI Matrix for Customer Data

<table>
<thead>
<tr>
<th>Category</th>
<th>Critical Attributes</th>
<th>CIM</th>
<th>Operations</th>
<th>Customer Service</th>
<th>Marketing</th>
<th>Credit Risk</th>
<th>Internet Banking</th>
<th>Channels</th>
<th>Compliance</th>
<th>Privacy</th>
<th>Collections</th>
<th>Finance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core Demographics</td>
<td>Date of Birth</td>
<td>A</td>
<td>R</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>Party Relationships</td>
<td>Spouse Of</td>
<td>AR</td>
<td>R</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>Other Demographics</td>
<td>Income</td>
<td>I</td>
<td>AR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Demographics</td>
<td>Net Worth</td>
<td>I</td>
<td>AR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contacts</td>
<td>Phone Number</td>
<td>A</td>
<td>R</td>
<td>C</td>
<td>C</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Master Data Management Rules:
Master Customer Information at Financial Services Company

- Master Customer Information File across Bank and Life Insurance
- 21 attributes for customer data including first name, last name, street address and phone number
- The proposed data policy was that a customer could make a change to his or her profile by calling either the Bank or the Life Insurance company and have those changes cascade across the entire enterprise
Date of Birth Became a Sticking Point

• Date of Birth became a key sticking point

• The Bank used to accept changes to Date of Birth with limited documentation

• The Life Insurance division would require substantial supporting documentation due to the impact on life insurance premiums

• The Bank believed that the Life process would slow them down

• The Life Insurance division felt that the bank’s processes were too risky

Data Policy around Date of Birth

- If the party was a Life customer, there would be a flag set up on the CSR screen and the Bank would follow the Life process for accepting changes to Date of Birth
- If the party was not a Life customer, the bank would accept changes to the Date of Birth with limited documentation

Image courtesy of vlima.com via Flickr Creative Commons
## 2012 NAICS Structure

<table>
<thead>
<tr>
<th>Change Indicator</th>
<th>2012 NAICS Code</th>
<th>2012 NAICS Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>11</td>
<td>Agriculture, Forestry, Fishing and Hunting(^\dagger)</td>
</tr>
<tr>
<td>1111</td>
<td>111</td>
<td>Crop Production(^\dagger)</td>
</tr>
<tr>
<td>11111</td>
<td>11110</td>
<td>Soybean Farming</td>
</tr>
<tr>
<td>11112</td>
<td>111110</td>
<td>Soybean Farming</td>
</tr>
<tr>
<td>11113</td>
<td>111112</td>
<td>Oilseed (except Soybean) Farming(^\dagger)</td>
</tr>
<tr>
<td>11114</td>
<td>111113</td>
<td>Oilseed (except Soybean) Farming</td>
</tr>
<tr>
<td>11115</td>
<td>111140</td>
<td>Oilseed (except Soybean) Farming</td>
</tr>
<tr>
<td>11150</td>
<td>111141</td>
<td>Oilseed (except Soybean) Farming</td>
</tr>
<tr>
<td>11115</td>
<td>111115</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>1116</td>
<td>111150</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>11160</td>
<td>111161</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>11161</td>
<td>111160</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>111610</td>
<td>1111610</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>11165</td>
<td>1111611</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>1119</td>
<td>111162</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>11190</td>
<td>1111612</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>11191</td>
<td>1111613</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>11199</td>
<td>1111614</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>111995</td>
<td>1111615</td>
<td>Corn Farming</td>
</tr>
<tr>
<td>111999</td>
<td>1111616</td>
<td>Corn Farming</td>
</tr>
</tbody>
</table>

\(^\dagger\) T = trilateral agreement (United States, Canada, and Mexico)

Note for Indicator field: * = title change, no content change
** = new code for 2012 NAICS
*** = re-used code, content change (with or without title change)

### System of Record
- Inventory of code tables
- Canonical values of each code table
- Mappings across systems

### Stewardship
- Code tables will be assigned to data owners

### Workflows
- Pre-built workflows to make changes to each code table

### Propagation
- Changes in the System of Record will be propagated to source systems

Extract from 2012 NAICS structure at http://www.census.gov/eos/www/naics/
Adding Kosovo to List of Countries at Multinational

- Kosovo is not on the list of ISO 3166-1 alpha-2 country codes
- Multinational corporation wants to add Kosovo as a separate country code
- Approved workflow to add new entry to code table
  1. Regional Finance Steward (Europe) will propose the change
  2. Global Finance Steward will approve the change
  3. IT will add the new entry to the code table for Country

Source: Central Intelligence Agency's World Factbook via Wikimedia Commons
Data Governance Policies – Metadata Management

- **Business Glossary**
  Enterprise Data Management will maintain a business glossary with definitions of key business terms

- **Data Stewardship**
  Data owners will assign data stewards to manage business terms

- **Technical Metadata**
  The metadata tool will ingest key business terms and technical metadata to manage data lineage and impact analysis

- **Hierarchy**
  Policies, business rules, data rules, business terms and reference data should be linked in a hierarchy where possible

Photo by cityyear/Andy Dean via Flickr Creative Commons
Data policies **May Hierarchically Cascade to Business Rules and Data Rules** (Insurance Example)

- **Policy**
  - Data Quality should be improved for rating attributes that drive insurance premiums (this policy may not be documented)

- **Business Rule**
  - Only eligible individuals can be drivers

- **Data Rule**
  - Drivers should be at least 16 years of age (lower in some U.S. States)

- **Business Terms**
  - Date of Birth

- **Technical Metadata**
  - DOB
Data Governance Policies – Data Quality

- **Program Management**
  Enterprise Data Management will lead the overall Data Quality program

- **Data Ownership**
  Data stewards will identify critical data elements, define data rules and manage exceptions

- **Data Quality Scorecard**
  Enterprise Data Management will manage an Enterprise Data Quality Scorecard to track key data quality metrics by system and data domain

- **Data Issues Log**
  Enterprise Data Management will manage an overall data issues log that captures the status of key data issues
Data Issue Resolution – Auto Insurance Carrier

Transformation
CASE WHEN AGE <= 25 THEN
Youthful_Driver = ‘Y’ ELSE ‘N’ END

Hit Rate = 90%

<table>
<thead>
<tr>
<th>Application A Age</th>
<th>Application B Youthful_Driver</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Y</td>
</tr>
<tr>
<td>24</td>
<td>Y</td>
</tr>
<tr>
<td>55</td>
<td>N</td>
</tr>
<tr>
<td>28</td>
<td>N</td>
</tr>
<tr>
<td>40</td>
<td>N</td>
</tr>
<tr>
<td>33</td>
<td>N</td>
</tr>
<tr>
<td>Exception 83</td>
<td>Y</td>
</tr>
<tr>
<td>29</td>
<td>N</td>
</tr>
<tr>
<td>36</td>
<td>N</td>
</tr>
<tr>
<td>42</td>
<td>N</td>
</tr>
</tbody>
</table>
## Lead Data Steward Reports on Overall Status of Data Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Steward</th>
<th>Date Assigned</th>
<th>Date Resolved</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>83 year old is listed as a Youthful Driver</td>
<td>Jane Smith</td>
<td>14-Jan-2013</td>
<td>30-Jan-2013</td>
<td>Closed</td>
</tr>
<tr>
<td>Multiple life policies have past expiration dates</td>
<td>John Doe</td>
<td>29-Jan-2013</td>
<td></td>
<td>John Doe is talking to the Policy Administration department</td>
</tr>
<tr>
<td>Customer database includes a number of duplicate records</td>
<td></td>
<td>04-Feb-2013</td>
<td></td>
<td>Matter discussed at Data Governance Council on 18-Feb-2013. This issue requires a Master Data Management tool. Will be considered for 2014 budget.</td>
</tr>
</tbody>
</table>
Internal Audit Data Governance Policies

- Every data repository should have a data owner
- Approximately 25% of data repositories are subject to Internal Audit on a quarterly basis (100% on an annual basis)
- Each repository will be audited for compliance to specific data governance policies, e.g.,
  - Is there a data dictionary?
  - Have data rules been documented?
  - Who determines access controls?
Contacts

Sunil Soares
Information Asset, LLC
sunilsoares@information-asset.com

@sunilsoares1
www.information-asset.com
+1-201-693-2216