CON-6628
Data Governance

How to govern your data with Enterprise Data Quality and Enterprise Metadata Management

Mike Matthews & Jayant Mahto
Product Management
Data Integration

Neha Kaptan, Data Governance and Reference Data Leader, Cummins

September 2016
Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
What do we mean by Data Governance and why do we need to do it?

“Data governance encompasses the people, processes, and information technology required to create a consistent and proper handling of an organization's data across the business enterprise.” [Wikipedia]
Data Governance is Not Easy, there is No Silver Bullet!

...people and process first,

...tools and capabilities next,

...and, there is no magic!

“...the overall impact of poor-quality data on the whole dataset remains the same. In addition, much of the data that organizations use in a big data context comes from outside, or is of unknown structure and origin. This means that the likelihood of data quality issues is even higher than before. So data quality is actually more important in the world of big data.”

- Ted Friedman, Gartner

http://www.gartner.com/newsroom/id/2854917
It’s all about...

TRUST

...in your data
Effective Data Governance requires closely coordinated effort across disparate teams and systems:

Three Layers of Data Governance

- **Business**
  - Is the data complete and accurate?
  - Is the data what it claims to be?
  - Can it be understood?

- **Technical**
  - Are the field mappings right?
  - Are we connecting to all the right systems?
  - Can we simplify the architecture?

- **Operational**
  - Have all the processes run correctly?
  - Is the data complete and up-to-date?

**Focus on:**

- Data Contents
- Metadata
- Delivery
Enterprise Data Quality
Is the data ‘right’? Is it useful?

Focus on:

Business
- Is the data complete and accurate?
- Is the data what it claims to be?
- Can it be understood?

Technical
- Are the field mappings right?
- Are we connecting to all the right systems?
- Can we simplify the architecture?

Operational
- Have all the processes run correctly?
- Is the data complete and up-to-date?
Quality of Data Impacts Everything

Quality = Fitness for purpose

Complete – Valid – Consistent – Timely – Accurate

Poor data quality is the primary reason for 40% of all business initiatives failing to achieve their targeted benefits” Gartner

“Only 30% of BI/DW implementations fully succeed. The top two reasons for failure? Budget constraints and data quality.” Gartner

“More than 50 percent of data warehouse projects will have limited acceptance, or will be outright failures, as a result of a lack of attention to data quality issues” Gartner

“Through 2016, 25% of organizations using consumer data will face reputation damage due to inadequate understanding of information trust issues” Gartner

“It does not really matter how good your management sponsorship or your business-driven motivation is. If you do not have the data, or the data does not have sufficient quality, any BI implementation will fail.” Kimball

Copyright © 2016, Oracle and/or its affiliates. All rights reserved.
Oracle Enterprise Data Quality

Govern
Monitor effectiveness & resolve problems

Match
Identify & merge duplicates

Standardize
Drive conformance to standards

Profile
Quickly understand data content

Enterprise DQ Platform

Market-leading usability for all types of data

Unparalleled time-to-value

High performance engine

Out-of-the-box global knowledge-base

Foundation for governance program

Common Access/UI
EDQ for Data Quality Governance

Data Analysts → Data Stewards → Dashboard

Director

Case Management

Data Stakeholders
Enterprise Metadata Management
Where is all my data? What is all my data? How does it move between systems?

Focus on:

Business
- Is the data complete and accurate?
- Is the data what it claims to be?
- Can it be understood?

Technical
- Are the field mappings right?
- Are we connecting to all the right systems?
- Can we simplify the architecture?

Operational
- Have all the processes run correctly?
- Is the data complete and up-to-date?

Data Contents
Metadata
Delivery
Value of Enterprise Metadata Management
Solves significant pain points for wide variety of business consumers and technical staff

- Which reports use this customer data?
- What reports use the mainframe data?
- How was the sales figure calculated?
- How do I organize my DW and Reports?
- Where did this data come from?
- What reports use the mainframe data?
- I want to design an experiment to measure the success of a signup page. What data do I have?
- Can I trust the sources of this customer data?

---

ETL
BI
Dashboards
App

ETL
Enterprise
Architect

ETL
Sys Admin

ETL
Data Steward

ETL
Data Reservoir

ETL
Data Scientist

ETL
BI
Dashboards

GG

Oracle

Copyright © 2016, Oracle and/or its affiliates. All rights reserved.
Oracle Enterprise Metadata Management (OEMM)

- **Metadata Management** – horizontal and semantic data lineage for all data sources
- **Business Glossary** – simple tools to catalog, link and collaborate on business terms

- **Business Data Catalog**
- **Report to Source Lineage**
- **Impact Analysis**
- **Audit, Versioning & Diff Reports**
- **Social/Collaboration Features**
- **Annotations and Tagging**
- **Comprehensive Harvesting**
  - 3rd Party BI Metadata
  - 3rd Party ETL Metadata
  - 3rd Party DB Metadata
  - 3rd Party Modeling Tools
  - Big Data Metadata
  - Metadata Standards
Customer Initiatives & Project Drivers

All Data Integration Use Cases

- **Oracle GoldenGate** (Migration)
- **Oracle Data Integrator** (Transformation)
- **Data Service Integrator** (Publication)

Cloud

- Apps
- Database
- Big Data

Realtime BI and Data Warehousing

- Top Performance, Integrated & “Red Optimized”
- Support Any Legacy Data Feeds
- Non-Invasive Source Capture
- Support traditional use cases: Data Warehousing, Reporting, Data Distribution, etc.
- Most Cost-Effective and High-Performance Exadata Data Loading

Big Data Reservoir

- Continuous Staging + ELT Fast Load to DW

Cloud Integration

- **Oracle Data Integration – Pragmatic Solutions for Cloud**
  - Cloud Database Sync
    - Database Consolidation
    - DBaaS Data Synchronization
  - Cloud SaaS to Mart/EDW
    - Bring SaaS Application data into on-prem data warehouses
    - Synchronize reference data or master data with SaaS Apps
  - Cloud BI / Analytics
    - Oracle Data Integration supports Oracle and non-Oracle BI/Analytics platforms

Oracle Enterprise Metadata Mgmt
metadata management aids in...

business user understanding
- lineage analysis for reports
- semantic analysis
- browsing/searching/organizing of metadata across the enterprise
- business term maintenance and understanding its usage

simplifying enterprise complexities
- understanding past implementations by various teams on various platforms
- fills the knowledge gap due to lack of documentation & employee turn-around
- liberates metadata from its platform silos and consolidates for centralized access

addressing the data mess
- identifies duplicate metadata/data flows
- checks for semantic consistency in data flows
- analyzes metadata for detecting errors /standards

providing structure to change management
- assists in application migration efforts by understanding old application metadata
- assists etl & bi development by metadata discovery of source & target systems
- provides metadata standardization across the enterprise
- keeps track of metadata changes over time
Data Lineage in Three Crucial Varieties

**Business Lineage**
- For the Business Users
- Links a business friendly set of Glossary terms to the BI reporting models (eg; Semantic Models, Universe, etc.)
- Links Business Glossary, Taxonomy, Ontology, Conceptual Models to Reporting Fields

**Vertical Lineage**
- For the Data Stewards
- Links a business friendly set of terms to the IT metadata and operational assets
- Links Business Glossary, Taxonomy, Ontology, Conceptual Models to Technical Metadata (Columns, Files, Objects etc)

**Horizontal Column Level**
- For the Architects
- Links the data fields from Business Intelligence Dashboards or Reports back to the Source DB/App Columns
- Schemas, BI View Layers, ETL Transformations, Calculations, etc.
Simple Screens for both Business and IT User Profiles

Search Driven Business Access

Simple to Navigate All Metadata

Comprehensive Data Lineage for IT

Business / IT Collaboration
End-to-End Data Flow Architecture Views

See how systems are connected while visualizing annotations and system boundaries
Keyword Based Search on all Metadata

Full index text search provides simple navigational access to very large metadata models.
Graphical Browser for Data Model Diagrams

Birds-eye-view metadata viewer enables zoom-in / zoom-out around complex metadata models
Versioning with Model Comparisons

Easily group and compare data models from similar databases to determine crucial differences.
Social Collaboration for Stewardship Teams

Team-based approach is designed in from the start
Find Data Provenance from the Report Layers

Visualize column and calculation level dependencies across systems, databases and reporting views.
Business Glossary is Built In (not a separate add-on)

Going beyond database dictionaries, link business terminology across metadata models and systems.
Perform Semantic Analysis

For a Business Term, find all the physical implementations where the term is used
Organize and Document Metadata

Organize Metadata, Use custom attributes, Use Labels and custom reporting

<table>
<thead>
<tr>
<th>Name</th>
<th>Entity Type</th>
<th>Context</th>
<th>Physical Name</th>
<th>Description</th>
<th>Comment</th>
<th>Position</th>
<th>Datatype</th>
<th>Lab...</th>
<th>Data Sensitivity</th>
<th>CustomAttribute 01</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADDR</td>
<td>Column</td>
<td>ADDR</td>
<td>ADDR</td>
<td></td>
<td></td>
<td>4</td>
<td>char</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL_ADDR</td>
<td>Column</td>
<td>BL_ADDR</td>
<td>BL_ADDR</td>
<td></td>
<td></td>
<td>5</td>
<td>char</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMAIL</td>
<td>Column</td>
<td>EMAIL</td>
<td>EMAIL</td>
<td></td>
<td></td>
<td>1</td>
<td>char</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM</td>
<td>Column</td>
<td>NM</td>
<td>NM</td>
<td></td>
<td></td>
<td>2</td>
<td>char</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP_ADDR</td>
<td>Column</td>
<td>SP_ADDR</td>
<td>SP_ADDR</td>
<td></td>
<td></td>
<td>6</td>
<td>char</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USR_NM</td>
<td>Column</td>
<td>USR_NM</td>
<td>USR_NM</td>
<td></td>
<td></td>
<td>3</td>
<td>char</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Business Glossary workflow

Workflow for the entire lifecycle from drafting to the final approval of business terms
Open Source, Standards and Third Party Integrations

**Key Standards**
- Apache HDFS
- Apache Hive Metastore
- Apache Hive Server
- Borland
- CA Component Modeler
- CA COOL (previously Sterling)
- CA Erwin
- CA Gen
- Cloudera Enterprise
- Cloudera Impala
- Database (JDBC)
- DataStax
- Embarcadero
- EMC ProActivity
- GentleWare
- Google BigQuery
- Grandite
- Hadapt Hive
- Hortonworks Hive
- IBM Cognos
- IBM DB2
- IBM Data Architect
- IBM DataStage
- IBM Discovery
- IBM Federation Server
- IBM Information Server
- IBM Lotus Notes

**Metadata Harvesting (Glossary, Lineage & Impact Analysis)**
- IBM Netezza
- IBM Rational Rose
- IBM Rational Architect
- IBM Telelogic
- Informatica Developer
- Informatica PowerCenter
- MapR Hadoop Hive
- MicroFocus
- Microsoft Office Excel
- Microsoft Visio
- Microsoft SQL Server
- Microsoft SSAS
- Microsoft SSIS
- Microsoft SSRS
- Microsoft Visual Studio
- Microstrategy
- Magic Draw
- OMG CWM Standard
- OMG UML Standard
- Oracle BI Answers
- Oracle BI Enterprise Edition
- Oracle BI Server
- Oracle DAC
- Oracle Data Integrator
- Oracle Data Modeler
- Oracle Database
- Oracle Designer
- Oracle Hyperion EIS
- Oracle Warehouse Builder
- Pivotal Greenplum
- PostgreSQL
- SAP BO Crystal Reports
- SAP BO Designer
- SAP BO Desktop Intelligence
- SAP BO Information Design Tool
- SAP BO Repository
- SAP Sybase PowerDesigner
- SAP Sybase ASE Database
- SAS OLAP Server (via ODBO)
- Select SE
- Sparx Architect
- Tableau
- Talend
- Teradata
- Tigris
- Visible
- W3C XSD Schema
But don’t forget to put People & Process first!

...people and process first,

...tools and capabilities next,

...and, there is no magic!

“...the overall impact of poor-quality data on the whole dataset remains the same. In addition, much of the data that organizations use in a big data context comes from outside, or is of unknown structure and origin. This means that the likelihood of data quality issues is even higher than before. So data quality is actually more important in the world of big data.”

- Ted Friedman, Gartner

http://www.gartner.com/newsroom/id/2854917
Customer Spotlight
Neha Kaptan – Data Governance and Reference Leader, Cummins
DATA FUELS THE BUSINESS

Taking Data Governance To The Next Level with OEMM
Who Are We?

- **Headquarters**: Columbus, Indiana
- **2015 revenue**: $19.1 billion
- **2015 Fortune 500 rank**: 154
- **Employees**: 55,000 worldwide
- **Presence**: 190+ countries and territories with 600 distributors and 7,200+ dealers
Our Data Governance Journey

• Started in Q4 2013
• Grown from 3 working councils to 8
• Agile and iterative
• Top-down vs bottom-up approach

What does that exactly mean?
Bottom-Up Approach...

- Identify all systems in scope
- Discover and analyze data
- Define data standard based on profiling and analysis
- Get everyone to agree on standard

Months of effort
Endless meetings
...vs Top-Down Approach

- Identify an owner
- Define data standard based on business need
- Define control plan for data quality
- Drive implementation through planned IT roadmaps

Quick Results
Easier Adoption
Our Data Governance Process

- Identify Data Owners
- Define Data Governance Scope
- Prioritize Data Quality Needs
- Measure Compliance
- Cleanse and Standardize Data
- Define Data Standards and Controls
- Implement Data Standards and Controls
- Publish Standards
- Train Data Stewards
- Resolve Data Issues
- Publish Standards
How Do We Operate Today?

Identify Data Owners → Prioritize Data Quality Needs → Define Data Standards and Controls → Publish Standards

Define Data Governance Scope → Resolve Data Issues → Measure Compliance → Cleanse and Standardize Data → Implement Data Standards and Controls

Train Data Stewards
How Can OEMM Take Us Further?

Identify Data Owners

Prioritize Data Quality Needs

Define Data Standards and Controls

Publish Standards

Define Data Governance Scope

Train Data Stewards

Resolve Data Issues

Data usage and traceability

Which data is most widely used across my reports and subject areas?

What data attributes make up my most critical KPIs?
How Can OEMM Take Us Further?

What is the business definition of my data attributes?

What are the business rules, formatting rules and quality rules that define their standard?
How Can OEMM Take Us Further?

Where does this data reside? How does it flow?

If I were to change a data standard, which systems will be impacted?
So We Have OEMM. Now What?

Some Questions to Consider:
- How can we best fit the tool and its capabilities within our data governance process?
- Who needs to learn and use this tool across our teams? How to make them capable?
- How can we make metadata management, data quality management and data governance management processes and tools seamless for our councils?
Data Integration Solutions Program - tinyurl.com/DISOOW16

Presentations on:

- Oracle Data Integrator
- Oracle GoldenGate
- Oracle Enterprise Data Quality
- Oracle Enterprise Metadata Management
- Oracle Big Data Preparation Cloud Service

Hands-on labs:

- Oracle Enterprise Data Quality HOL7466
- Oracle GoldenGate Deep Dive HOL7528
- ODI and OGG for Big Data HOL7434
- Oracle Big Data Preparation Cloud Service HOL7432

Demo Stations:

- Middleware Demoground - Moscone South
- Database Demoground - Moscone South
- Big Data Showcase - Moscone South
Data Integration Solutions Program - tinyurl.com/DISOOW16

Monday, Sept 19
- Oracle Data Integration Solutions – Platform Overview and Roadmap [CON6619]
- Oracle Data Integration: the Foundation for Cloud Integration [CON6620]
- A Practical Path to Enterprise Data Governance with Cummins [CON6621]
- Oracle Data Integrator Product Update and Strategy [CON6622]
- Deep Dive into Oracle GoldenGate 12.3 New Features for the Oracle 12.2 Database [CON6555]

Tuesday, Sept 20
- Oracle Big Data Integration in the Cloud [CON7472]
- Oracle Data Integration Platform: a Cornerstone for Big Data [CON6624]
- Oracle Data Integrator and Oracle GoldenGate for Big Data [HOL7434]
- Oracle Enterprise Data Quality – Product Overview and Roadmap [CON6627]
- Self Service Data Preparation for Domain Experts – No Programming Required [CON6630]
- Oracle Big Data Preparation Cloud Service: Self-Service Data Prep for Business Users [HOL7432]
- Oracle GoldenGate 12.3 Product Update and Strategy [CON6631]
- New GoldenGate 12.3 Services Architecture [CON6551]
- Meet the Experts: Oracle GoldenGate Cloud Service [MTE7119]

Wednesday, Sept 21
- Data Quality for the Cloud: Enabling Cloud Applications with Trusted Data [CON6629]
- Oracle Enterprise Data Quality for All Types of Data [HOL7466]
- Oracle GoldenGate for Big Data [CON6632]
- Accelerate Cloud On-Boarding using Oracle GoldenGate Cloud Service [CON6633]
- Oracle GoldenGate Deep Dive and Oracle GoldenGate Cloud Service for Cloud Onboarding [HOL7528]

Thursday, Sept 22
- Best Practices for Migrating to Oracle Data Integrator [CON6623]
- Best Practices for Oracle Data Integrator: Hear from the Experts [CON6625]
- Dataflow, Machine Learning and Streaming Big Data Preparation [CON6626]
- Data Governance with Oracle Enterprise Data Quality and Metadata Management [CON6628]
- Faster Design, Development and Deployment with Oracle GoldenGate Studio [CON6634]
- Getting started with Oracle GoldenGate [CON7318]
- Best Practice for High Availability and Performance Tuning for Oracle GoldenGate [CON6558]
Connect with Oracle Data Integration

Oracle Data Integration

@OracleDI

Blogs.oracle.com/DataIntegration/

Oracle Data Integration
Integrated Cloud
Applications & Platform Services