ORACLE
FUSION MIDDLEWARE
DATA INTEGRATION SUITE

Data Governance with Oracle

Defining and Implementing a Pragmatic Data Governance Process with Oracle Metadata Management and Oracle Data Quality Solutions

ORACLE WHITE PAPER | SEPTEMBER 2015

Disclaimer

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.

Table of Contents

.

Introduction

Data governance is the formulation of policy to optimize, secure, and leverage information as an enterprise asset by aligning the objectives of multiple functions. Data governance programs have traditionally been focused on people and process. In this whitepaper, we will discuss how key data governance capabilities are enabled by Oracle Enterprise Metadata Manager (OEMM) and Oracle Enterprise Data Quality (EDQ). This paper will introduce specific examples of how to leverage OEMM and EDQ tools within a well-defined data governance process and will also provide specific technical examples of how to deploy these solutions on Oracle and Intel® based engineered systems.

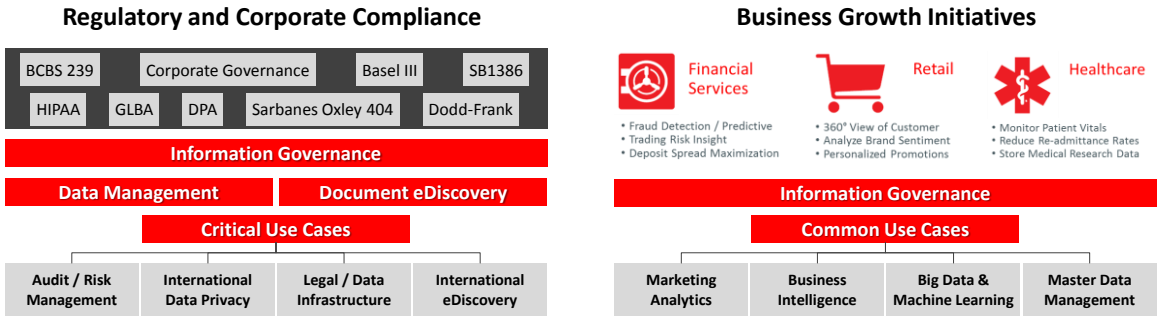

Figure 1: Key business drivers for data governance solutions

Data governance is more important today than it has ever been before. While always critical for Regulatory, Legal and Corporate Governance initiatives, the role of data governance has expanded to become a foundation for driving Business Growth initiatives as well. Since 2010, the industry focus on Analytics, Business Intelligence, and Big Data programs has elevated the impact of data governance by driving requirement to make data easier to find, associate data with applications and business line growth programs, and improve the confidence and trust that data is accurate enough for business decision-making.

For the remainder of this paper we'll take a look at a pragmatic step-by-step approach you can use in order to get started with an operational data governance program using Oracle tools.

1. Define the Business Problem
2. Identify the Business and Executive Sponsors
3. Manage and Classify the Business Terms
4. Identify the Critical Data Elements for the Business
5. Setup Policies, Rules and Allowable Values
6. View Metadata Lineage: Business and Technical
7. Manage Stewardship Process and Workflows
8. Integrating Glossary and Metadata from Big Data and NoSQL

9. Linking Data Quality Rules to Business Policies
10. Tying it Together with a Quality Dashboard
11. Handling Remediation and Repair of Bad Data

After we explain the steps for using the Oracle Enterprise Metadata Management (OEMM) and Oracle Enterprise Data Quality (EDQ) tools with a data governance process, we will also specifically explain how these solutions can fit together with other Oracle tools such as Oracle Data Integrator, Oracle GoldenGate, Oracle Data Relationship Management, and Oracle Master Data Management. Finally, we will conclude by providing specific guidance on how to deploy OEMM and EDQ into Oracle and Intel® environments such as Exadata, Exalogic, Big Data Appliance, Oracle Public Cloud and Oracle Managed Cloud.

Let's get started!

First Define the Business Problem

The first step in any data governance program[1] is to define the business problem that needs to be solved. The business problem may be characterized as follows:

1. Data scientist: "I spend too much time looking for definitions."
2. Chief Financial Officer: "I received two reports yesterday with different numbers for total sales last month."
3. Chief Information Security Officer: "We need a comprehensive repository of applications so that we can apply information security protection."
4. Chief Executive Officer: "I asked for a report on the total number of customers. However, finance and marketing gave me different numbers."

A medical device manufacturer implemented a data governance program focused solely on the "ship-to" information for its business customers. The company's customers experienced delayed shipments, as products would bounce from one department to another before they reached the correct destination. The company deployed data stewards within the commercial (sales) organization to improve the quality of the ship-to addresses. The company soon experienced improved customer satisfaction as equipment began arriving at the correct destination faster. In addition, the finance department noticed an uptick in revenues from after-sale consumables as customers began to deploy the equipment earlier.

Table 1 provides a description of the business benefits.

[1] Selling Information Governance to the Business, Sunil Soares (MC Press, 2010)

A. Number of ship-to addresses in the customer master database	100,000
B. Estimated percentage of inaccurate customer ship-to addresses	5%
C. Estimated number of inaccurate ship-to customer addresses (AxB)	5,000
D. Average delay in weeks where the product is not installed at the customer site due to inaccurate ship-to addresses	2
E. Average sales of aftermarket consumables per customer per week	$100
F. Total potential increase in aftermarket consumables based on improved ship-to customer addresses (CxDxE)	$1,000,000

Table 1: Business benefits from improved ship-to addresses at a medical device manufacturer.

Identify Executive Sponsor

The business problem drives the executive sponsor for the data governance program[i2]. Sponsors for the data governance may include the following roles:

1. Vice president of marketing to increase customer cross-sell and up-sell.
2. Vice president of engineering to drive clean product data and consistent hierarchies.
3. Vice president of human resources to ensure correct employee data during the on-boarding process.
4. Chief financial officer to improve the quality of vendor data to get the best payment terms.

For example, the vice president of credit risk at a bank sponsored a data governance initiative to reduce the amount of capital that needed to be set aside to meet regulatory requirements. The data governance team profiled the columns in the collateral database. It found that a number of secured loans actually had null values in the COLLATERAL_VALUE field. For example, the bank might have approved a loan for $10,000 against an automobile with a resale value of $20,000. However, due to clerical errors, the value of the collateral was never entered into the COLLATERAL_VALUE field. As a result, the bank had to consider the loan to be unsecured, which meant that it had to assign a higher level of regulatory capital than if the loan were fully secured. Once the data governance program identified the list of loans with data quality issues, it was able to engage the operations team to enter the correct value of the collateral. As a result, the bank was able to reduce its regulatory capital by five percent with respect to those loans.

In another example, a European telecommunications operator had first-hand experience with inconsistent business terms when different functional areas within the organization could not agree on a consistent definition for the term “active subscriber.” For example, the billing department defined an active subscriber as someone who had received a bill in the previous 30 days. The network department defined an active subscriber as someone who had used the carrier’s network in the previous 30 days. The key difference between these two definitions related to subscribers

2 Selling Information Governance to the Business, Sunil Soares (MC Press, 2010)

who had signed up for service but who had switched SIM cards and were roaming outside the carrier's network on a full-time basis. The data governance team was able to establish a consistent set of definitions in this case. It got people out of their silos to share information and see the broader business concepts and drivers. The data governance organization then re-used these common definitions across others parts of the organization as well.

Manage Glossary of Business Terms

The starting point for most data governance programs is usually a glossary of business terms. Although the starter glossary for many data governance teams is often an Excel file, this becomes unmanageable once the number of business terms exceeds 100 or so. Once the number of business terms crosses that threshold, data governance teams should look at automated tooling. Tools like Oracle Enterprise Metadata Management (OEMM) can import directly from Excel as a way to seed business terms from pre-existing efforts. As shown in Figure 2, OEMM maintains a definition for the term "Registration Type," and we can also see that Jack Smith is the data steward.

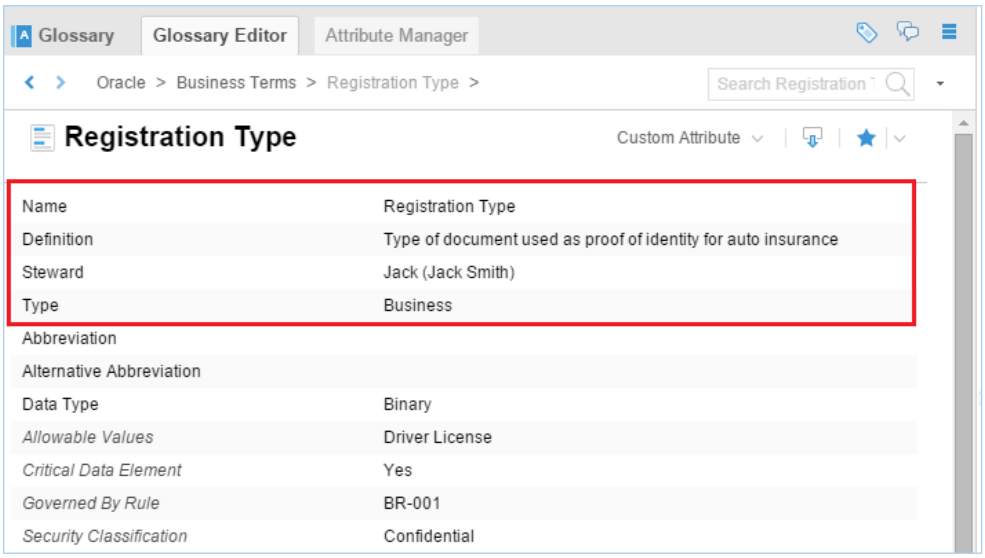

Figure 2: OEMM business glossary contains the definition for the term "registration type."

Another reason that organizations will frequently consider tools rather than spreadsheets is that audit requirements demand to be able to track and report the history of changes to business terms that are used to drive reporting and operational decisions. For example, OEMM maintains a change log of data steward activities on the business objects. Beyond simple audit ability requirements, this type of tool capability is absolutely crucial for maintaining and verifying the organizational and staff accountability of specific KPIs, reports, or reference data. Among the most important of all data objects to maintain accountability and audit ability for are Critical Data Elements.

Identify Critical Data Elements

Critical Data Elements (CDEs) are metrics and attributes that are of high importance to the business. CDEs have a significant impact on regulatory reporting, operational performance, and business intelligence. The rule of thumb is that CDEs should constitute only five to ten percent of organizational data. CDEs will drive data quality, security, and other data improvement programs. Since CDEs will often have the most impact to business success, it is vital to

identify these and focus efforts on supporting the sustainability and the effective use of these assets. Data stewards often have a significant role in identifying CDEs. Figure 3 shows that registration type has been flagged as a CDE in OEMM.

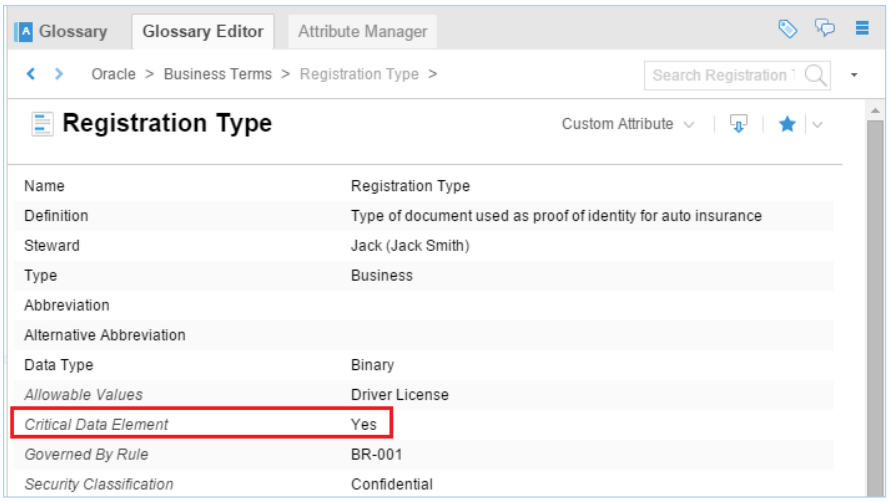

Figure 3: Registration Type is identified as a CDE in OEMM.

Classify Data from an Information Security Perspective

The data governance tool should also support the classification of data from an information security perspective. Sample information security classifications may be “Public,” “Internal,” “Confidential” and “Highly Confidential.” The information covered by each information security classification has specific handling instructions. For example, Social Security Number may be classified as Highly Confidential. This classification may mean that data stewards would need to restrict access to applications and tables that contain Social Security Numbers. Figure 4 shows that Registration Type has been classified as confidential information in OEMM

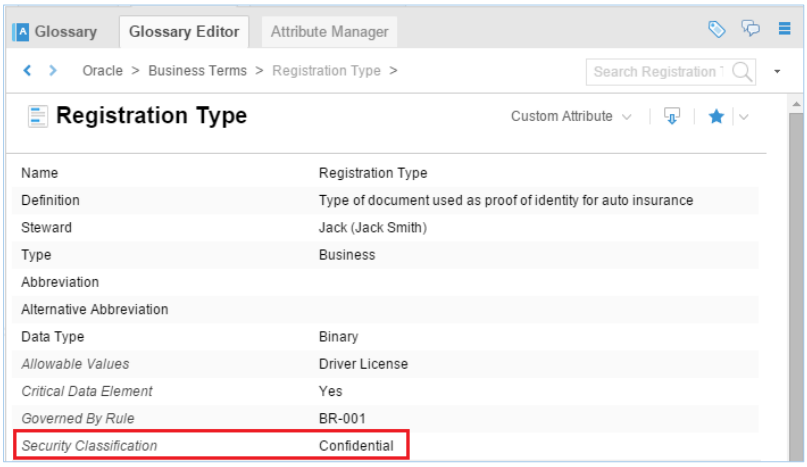

Figure 4: Registration Type is classified as confidential information in OEMM.

Manage Business Rules

The data governance tool should also support the governance of business rules. Figure 5 shows business rule **BR-001** in OEMM that states that “a party who is a person must have a registration type that is a driver license, national ID or passport.”

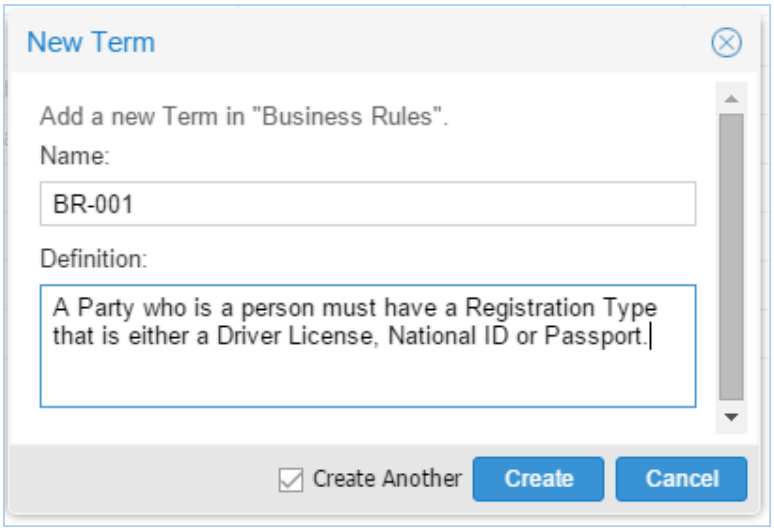

Figure 5: Business rule in OEMM.

The data governance tool should also support the mapping of business terms to the associated business rules. As shown in Figure 6, the data steward links the **Registration Type** business term to business rule **BR-001** in OEMM. Documenting the business rules in this manner is the first step in enforcing compliance with business processes when data is gathered in upstream applications. In this case, the upstream application should implement a data validation rule that driver license, national ID and passport are the only acceptable forms of evidence for registrations.

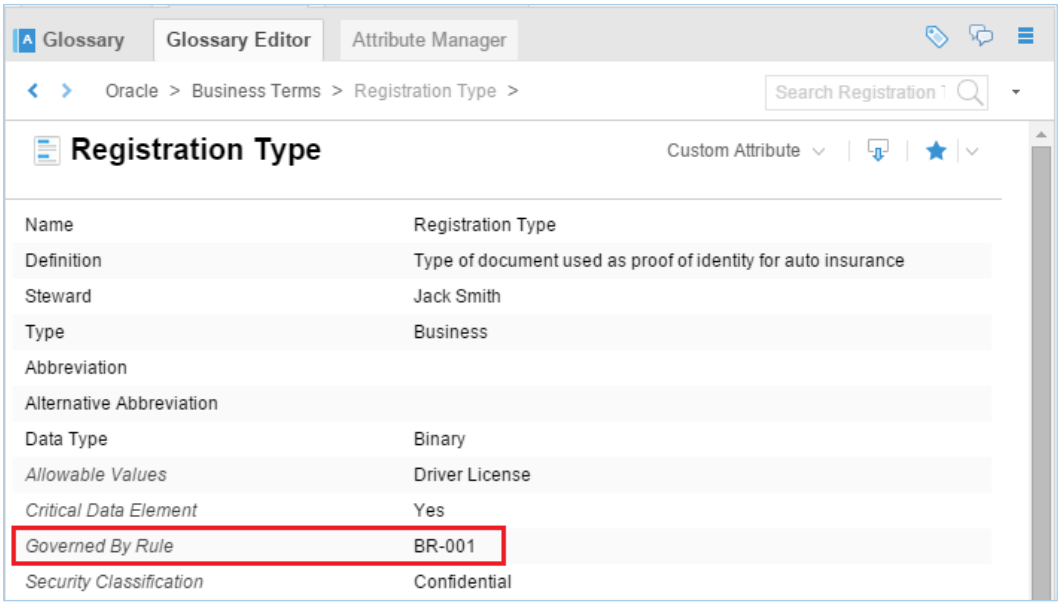

Figure 6: Mapping of business rule to business term in OEMM.

For larger and more complex business rule sets, data stewards can create a new domain in OEMM and associate those terms in relation to it. This way, you'll have a hierarchy that is more searchable and traceable. Additionally, this will restrict the user to select only from the available domains. Within OEMM, this approach will also leverage the mapping-specification functionality, which is the ability to semantically map those terms to actual physical attributes; this way, one is able to make a semantic analysis for a specific physical attribute (or entity), and make use of the definitions and business rules mapped to it.

Manage Allowable Values for Business Terms

The data governance tool should also display the allowable values for a business term. This approach helps to answer questions such as, "Is Puerto Rico considered a state within our address table?" Having an agreed upon set of allowable values also supports data quality efforts. For example, Figure 7 shows that the allowable values for **registration type** in OEMM.

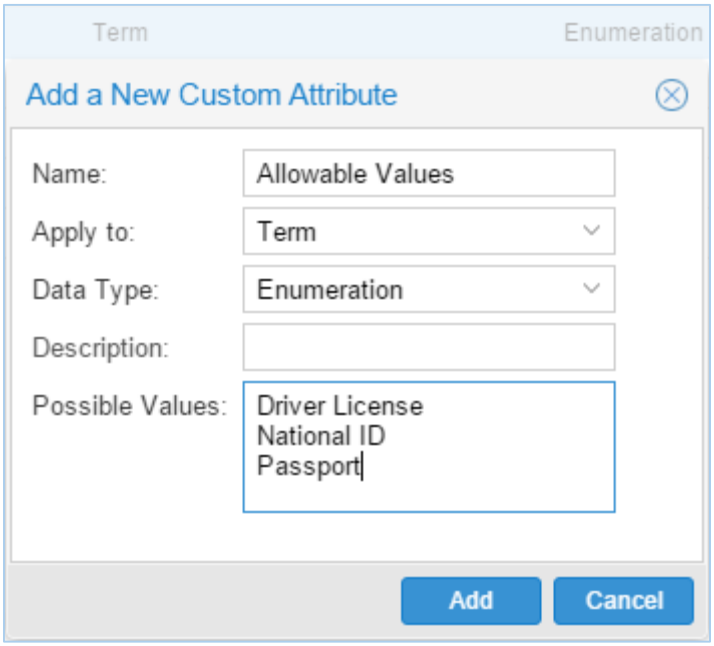

Figure 7: Allowable values for registration type in OEMM.

Support for Data Lineage and Impact Analysis

Business users often have to answer questions like these:

1. Where did this data come from?
2. Where is it going?
3. What happens to it along the way?
4. What is the impact if we drop this column?

Organizations need a strong metadata foundation to be able to answer these questions. Once metadata tools have ingested technical metadata from a variety of data sources, data stewards should be able to view four types of data lineage:

1. ### Report to source lineage

   A data steward is able to view the lineage from a field in a report to the source table. Figure 8 shows an example of data lineage in OEMM from the Products Report all the way back to the source tables.

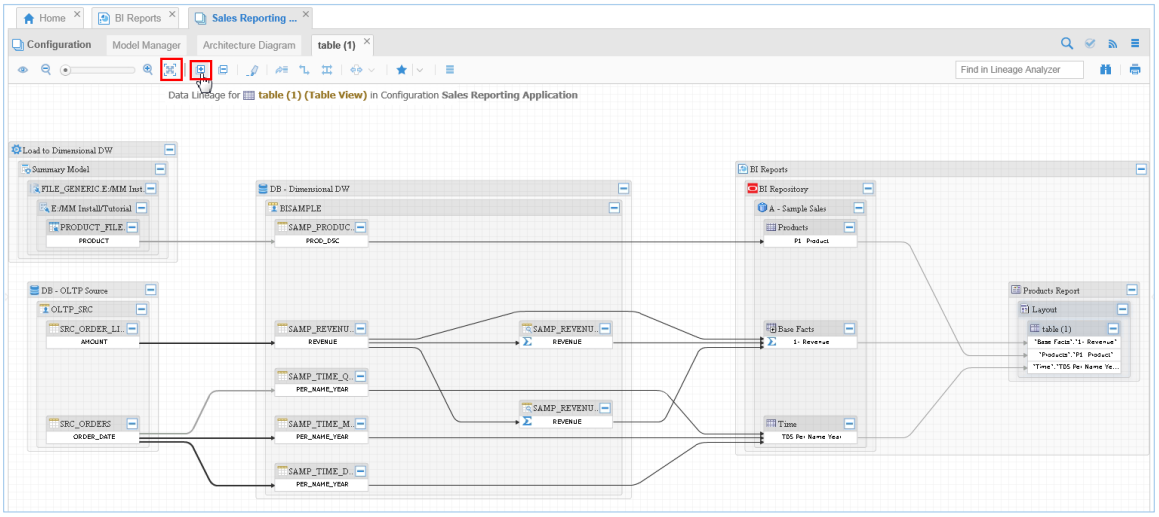

Figure 8: Data lineage in OEMM.

2. Glossary to column lineage

A data steward is able to view the lineage in OEMM from a business term or CDE to the source tables. Taking our earlier example, the data steward can see that the Registration Type CDE is represented by the REG_TYPE column in the CUSTOMERS table.

3. Business lineage

A business user can view the lineage from business terms and CDEs to report semantics. For example, the data steward should be able to use OEMM to trace the Profit per Customer metric in the profitability report to the definitions of profit and customer. The definition of profit must also be traced to the business terms representing individual line items in the company's profit and loss statement. This same functionality can be used to show where terms and rules are used across many business reports. For example, given a term in the repository, show all business reports which use a data element classified by (semantically linked to) that term in the glossary. The business user may search for terms in the glossary. Once they find what they are interested it, one click shows all the reports which have fields that relate to the term by (a) direct 'Glossary to Report' mappings and (b) 'Glossary to Data Element' classification where that data element is then connected to the report field by pass-through (or very simple) data flow

4. Data value lineage

A data steward or business user should also be able to trace source records to aggregate values. This solution combines elements of the three types of data lineage covered earlier. For example, the data steward should be able to view the lineage from the Profit per Customer metric to the REVENUE_AMOUNT column in the financial application.

Manage Data Stewardship Workflows[3]

Data governance involves a lot of manual work and meetings. Figure 9 shows a sample workflow to add a new business term. It is obvious that the approval of a single business term may involve multiple meetings and emails. Data stewards can save valuable time by leveraging automated workflow capabilities within OEMM.

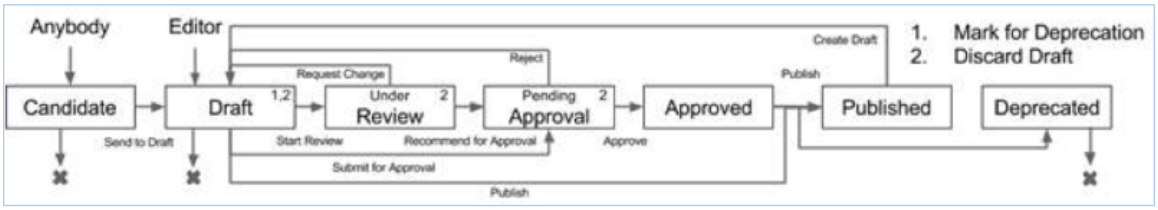

Figure 9: Sample workflow to add a new business term.

Figure 10 shows a new business term called **SAMP_REVENUE_F** that is in draft status in OEMM.

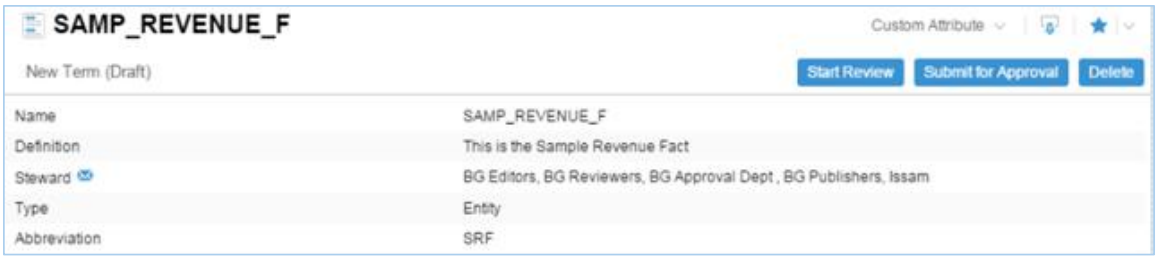

Figure 10: SAMP_REVENUE_F is a business term in draft status in OEMM.

As shown in Figure 11, the OEMM workflow routes the business term to the data steward who adds an approval.

Figure 11: Data steward approves the business term as part of an OEMM workflow.

3 Oracle Metadata Management – Business Glossary, Issam Hijazi (June 8, 2015)
https://www.linkedin.com/pulse/oracle-metadata-management-omm-business-glossary-issam-hijazi

As shown in Figure 12, the status of the business term has been changed to approve in OEMM.

Figure 12: Status of business term changed to approved in OEMM.

Govern Big Data

Big Data is especially hard to govern because information is in structured and semi-structured formats. As enterprises manage mission-critical applications within Hadoop, the metadata repositories also need to follow suit. As shown in Figure 13, OEMM's Big Data connectors allow the user to view the metadata relating to the Hive columns.

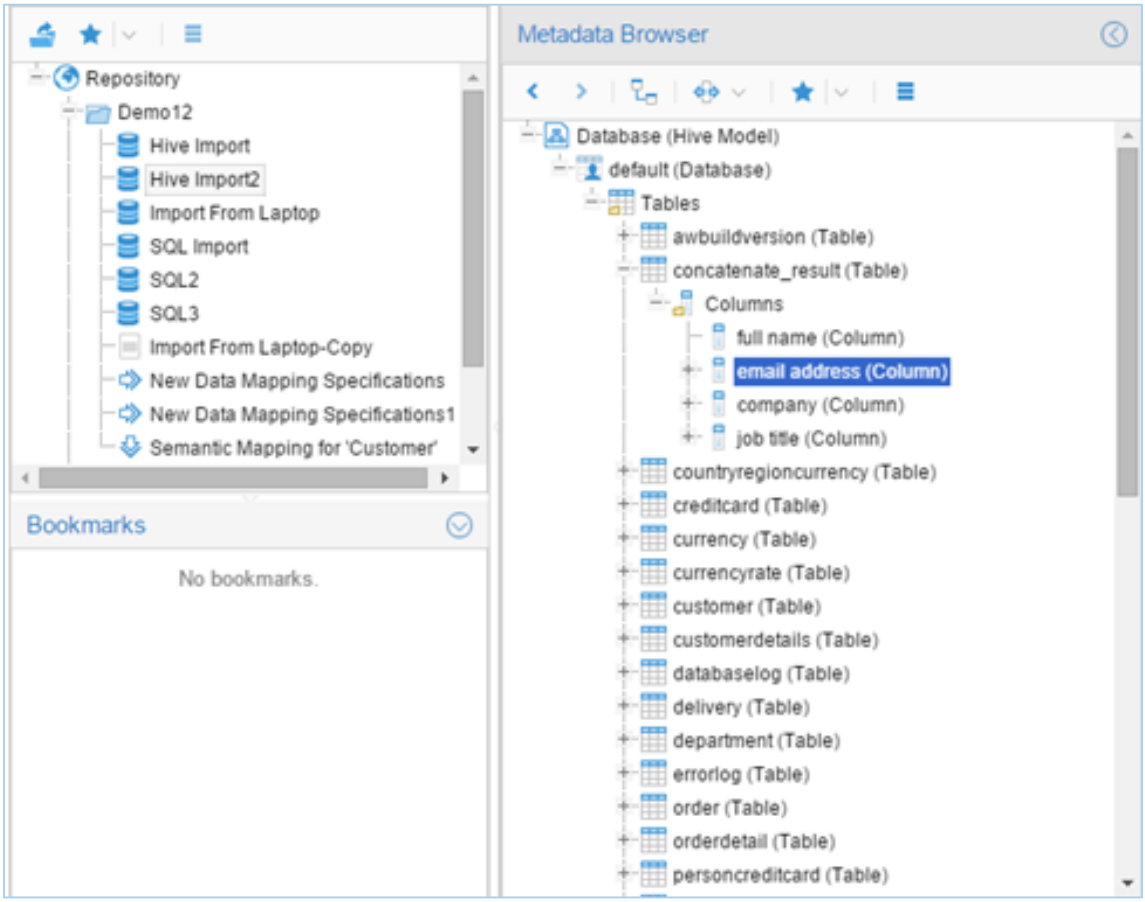

Figure 13: Viewing imported Hive database and tables in OEMM.

Since Hadoop is very often used to store "raw data" that is by definition ungoverned, a critical first step to governance is cataloging the data. OEMM provides the metadata harvesting capabilities to harvest metadata from Apache, Cloudera, Hortonworks and MapR Hadoop distributions. Metadata may be harvested from HDFS, Hive, Impala and HBase among others.

Other governance functionality such as data profiling, data cleansing, and data enrichment are covered by other Oracle tools such as Enterprise Data Quality and Big Data Preparation.

Manage Data Quality Rules

OEMM and EDQ provide a tightly integrated approach to data governance. Data stewards can view “plain English” data quality rules in OEMM. These data quality rules are then encoded and executed in EDQ. In Figure 14, the data steward has cataloged a data quality rule that **First Name** should not have null values.

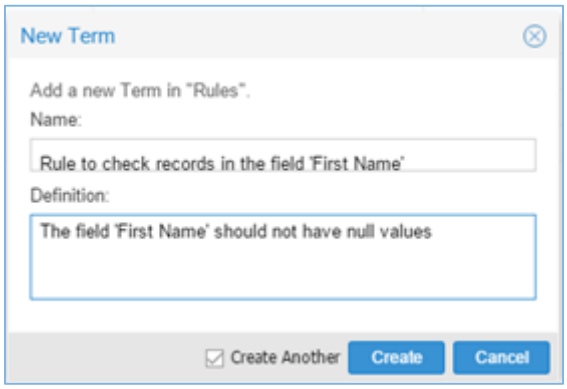

Figure 14: Data quality rule for First Name in OEMM.

In similar fashion, the data steward created two additional data quality rules in OEMM associated with an “invalid character check for Last Name” and “a valid email address format in the Email Address field.”

Execute Data Quality Rules

The data steward then uses EDQ to execute the data quality rules that have been cataloged in OEMM. In Figure 15, the data steward drags and drops three processors called **No Data Check**, **Invalid Character Check** and **Email Check** from the tool palette in EDQ.

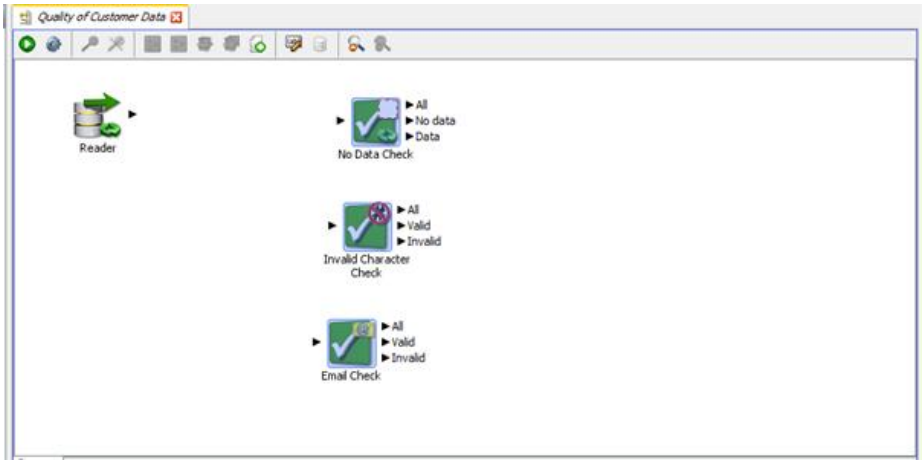

Figure 15: Tool palette in EDQ.

In Figure 16, the data steward then associates **First Name** with the **No Data Check** processor in EDQ.

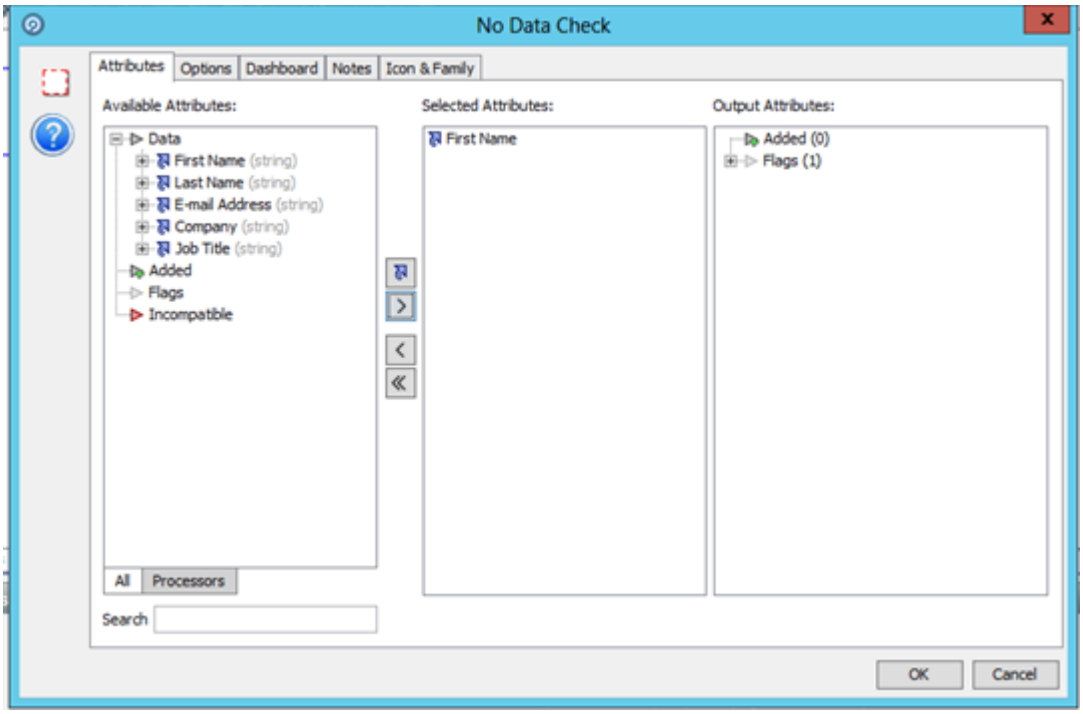

Figure 16: Associate First Name with No Data Check processor in EDQ.

In Figure 17, the data steward then associates **Last Name** with the **Invalid Character Check** processor in EDQ.

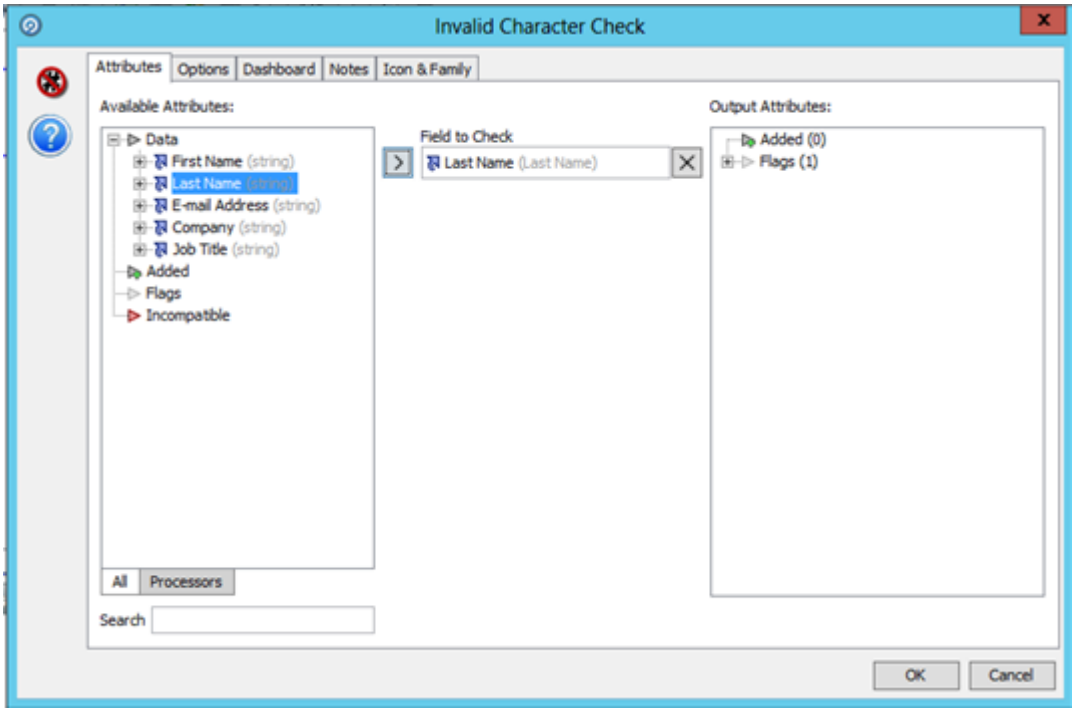

Figure 17: Associate Last Name with Invalid Character Check processor in EDQ

Finally, the data steward associates **Email Address** with the pre-built **Email Check** processor in EDQ as shown in Figure 18.

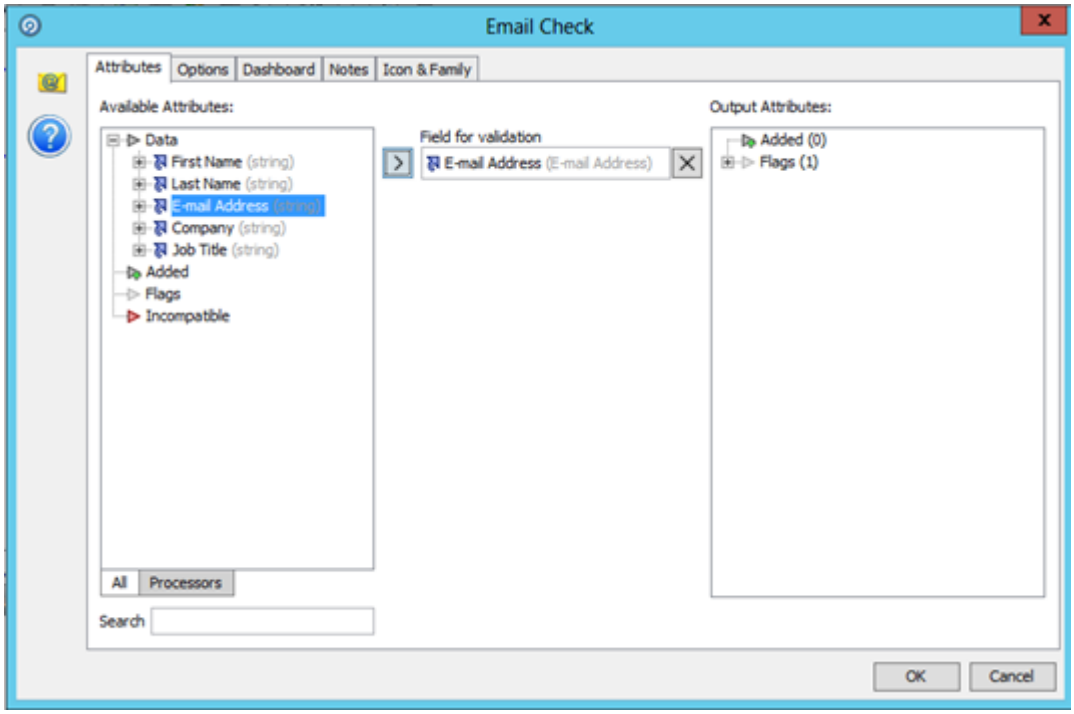

Figure 18: Associate Email Address with Email Check processor in EDQ.

The data steward executes the processes in EDQ. As shown in Figure 19, the Invalid Character Check processor shows that 4,362 records and 278 records have valid and invalid last names respectively.

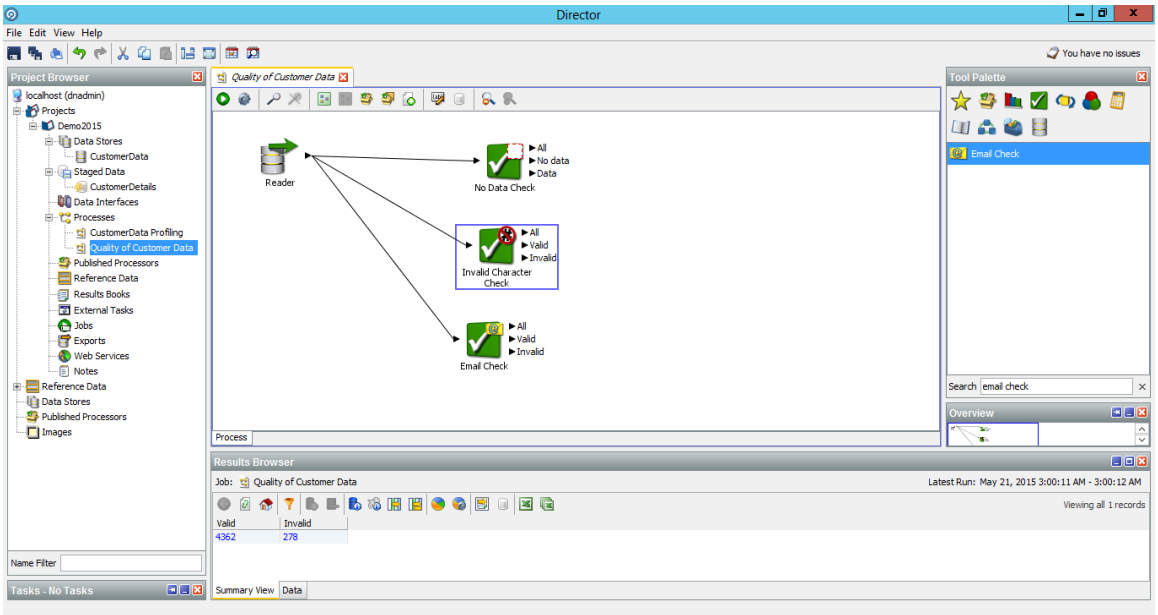

Figure 19: View results of Invalid Character Check on Last Name in EDQ.

View Data Quality Dashboard

The data steward then displays a data quality dashboard in EDQ as shown in Figure 20. The **Invalid Character Check for Last Name** had a pass rate of 94 percent. The **email format** check had a pass rate of 99.9 percent while the **First Name** check had a pass rate of 100 percent.

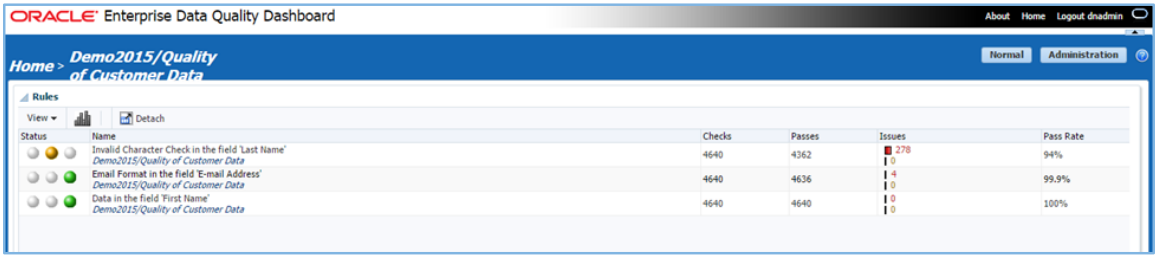

Figure 20: Enterprise Data Quality dashboard in EDQ.

Data Quality Remediation

Depending on the business rules associated with the data element, missing or invalid data may need to be corrected before being passed on to a system of record. This may be accomplished by preventing acceptance of an incomplete or invalid record in the original data capture system through the execution of a real-time data quality check. Where this is not possible and poor quality data has already been captured, the records in question may be reviewed and managed, through the use of the case management sub-system of EDQ. As shown in Figure 21, the business rule **BUS_VENDOR_TYPE_INVALID** states that the attribute **VENDOR_TYPE_LOOKUP_CODE** should be valid. In this case, vendor number 35 has an invalid lookup code. The data steward can view all the details of the issue in EDQ Case Management. The data steward can take a number of actions including assigning the case to another steward, adding a comment, adding an attachment, changing the state from open to closed, or fixing the data directly with EDQ.

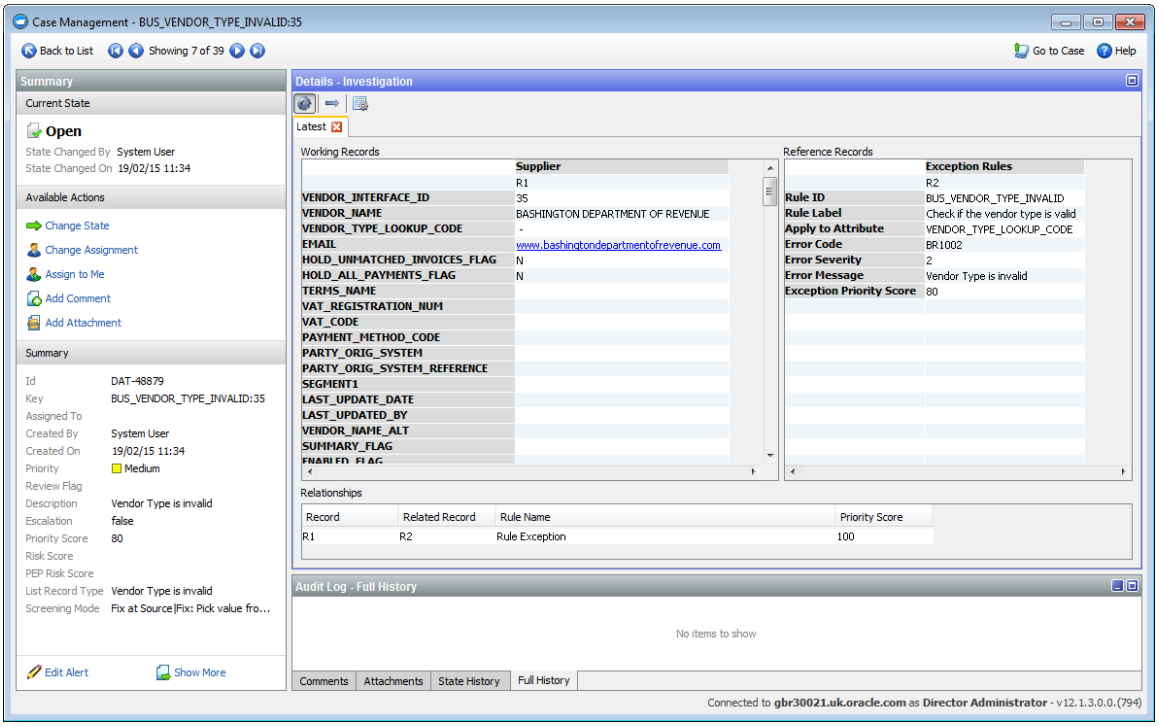

Figure 21: EDQ Case Management.

As discussed so far, data governance involves a number of processes and artifacts including business glossaries, CDEs, alignment with information security, business rules, allowable values, business and technical lineage, data stewardship workflows, Big Data Governance, data quality rules, data quality dashboards and data issue resolution. OEMM and EDQ provide a robust set of software tools to support the creation of these data artifacts as part of the data governance process.

Data Privacy and Security

Our focus in this paper has been on the elements of data governance associated with business glossary, metadata management and data quality. Additionally, a crucial ingredient to data governance that is not covered at depth in this paper is data privacy and data security. Data privacy is the frameworks, methodologies and controls to assure proper encryption and masking of data within enterprise systems. Data security are the multi-level controls for preventing unwanted access to data and are typically handled via Access Control Lists (ACLs) and Authentication techniques such as Kerberos. Data privacy and security are incredibly important topics well deserving of their own focus for data governance and it is not our intent to cover them in depth from within this paper.

Ingredients for Data Governance Success

Although the focus of this paper is around the tools that can enable a solid data governance foundation, it would be wrong to assume that data governance can be achieved simply through the implementation of software; rather, the software is there to assist in the implementation of previously determined business rules and policies. And because data governance is inherently a cross-functional discipline, it is also necessary to have a proper agreement on the

roles and responsibilities of the different organizations involved. This interconnection of "people, process and technology" can requires some dedication and focus, but when implemented successfully, can deliver some significant business benefits.

The following best practices illustrate how these different elements come together in a successful data governance program:

- Strong business ownership and IT support

  There has been considerable debate on whether data governance programs should report up through the line of business, or through IT. Nascent data governance programs may start in IT, but to expand and deliver their full potential, they will need to move into the business once they have achieved a level of maturity. It should also be noted that a data governance program will often come under the purview of the Chief Data Officer (CDO) organization – a role that is being implemented in an increasing number of companies. The reason for strong business ownership is that the business rules to be implemented and monitored by the program must be determined by the business, and most if not all of the enforcement effort and support must also come from the business.

  That having been said, it is clear that IT and the implementation of the right software tools is a critical ingredient of data governance success. A strong technology platform is critical to successful data governance programs because it can promote collaboration, boost productivity, and enable the repeatable processes that makes up data governance.

- Focus on critical data elements

  Because it is impossible to successfully govern all the data across the enterprise, mature data governance programs must focus on CDEs which, as stated earlier, have a significant impact on the business. Once selected, CDEs should be assigned to an 'owner' and business rules and policies developed for how the CDE will be governed. Ownership will likely be with a data steward, while recording the business rules will be done in OEMM and enforcement carried out through EDQ.

- Emphasis on data artifacts

  Successful data governance programs generate a number of valuable data artifacts. These data artifacts include business terms, business rules, code tables, policies, standards, processes, and data quality scorecards. Over time, successful data governance programs generate a rhythm associated with production, discussion and approval of these artifacts, and many of these artifacts will be stored by or created in either the metadata management system or the data quality system.

- Alignment around metrics and policy enforcement

  As with any program, data governance must focus on metrics that are important to the business. Some metrics will focus on the implementation of the program itself, for example, the number of fully defined, documented and implemented CDEs with full end-to-end lineage. Others will focus on the quality and usefulness of the data generated by the process, for example, the percentage of correct, complete and valid phone numbers captured on new customer records. Setting and reviewing targets for these metrics drives the implementation and improvement of the overall program and ensures that all participants are working towards the same goals. Most of these metrics can be derived from the metadata and data quality systems.

- Celebration of quick wins around the long-term roadmap

  Successful data governance programs are able to point to quick wins within weeks of inception. These quick wins may be creation of a glossary for key business terms, or a data quality report. However, these programs are also tied into a long-term roadmap over 12 to 18 months.

There is no right or wrong way to organize around and implement data governance and whole books can, and have, been written on the subject. Each enterprise needs to consider its own unique circumstances before deciding on the approach that works best in its environment. Significant factors may include industry, geography, history, culture and, of course, the specific business objectives at hand. However it is clear that to achieve the speed, scalability and breadth required of most data governance programs, the selection and incorporation of appropriate software technology will be an essential foundational element.

Governance with Any Enterprise System

Most businesses and IT organizations have invested in many software tools and technologies over the years. For a data governance solution to be truly effective it must work seamlessly with the systems and technologies that are in use within operational environments that run the business and IT systems. Oracle recognizes this fundamental reality and has done more than any others to support a wide array of Oracle and non-Oracle software solutions in the metadata management environment. The figure below describes the technologies directly compatible with OEMM today:

- Adaptive
- Altova
- Apache HCatalog
- Apache Hive/HQL
- Apache HDFS
- Borland
- CA ERwin
- Cloudera Impala
- Cloudera Apache
- COBOL Copybook
- DataStax
- Embarcadero
- EMC ProActivity
- GentleWare
- Google BigQuery
- Grandite
- Hadapt Hive
- Hortonworks Hive
- IBM Cognos
- IBM DB2
- IBM DataStage
- IBM Discovery
- IBM Federation Server
- IBM Lotus Notes
- IBM Netezza
- IBM Rational Rose
- CoSORT
- ISO SQL Standard (DDL)
- MapR Hadoop Hive
- MicroFocus
- Microsoft Access
- Microsoft Office Excel
- Microsoft Visio
- Microsoft SQL Server
- Microsoft SSIS
- 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
- QlikView
- SAP BO Crystal Reports
- SAP BO Designer
- SAP BO Desktop Intelligence
- SAP BO Repository
- SAP BO Data Integrator
- SAP BO Data Steward
- SAP Master Data Management
- SAP Sybase PowerDesigner
- SAP Sybase ASE Database
- SAS Data Integration Studio
- SAS BI Server
- SAS Information Map
- SAS Metadata Management
- SAS OLAP

▪ IBM Rational Architect ▪ Informatica Metadata Manager ▪ Informatica PowerCenter ▪ Informatica Developer	▪ Oracle Database ▪ Oracle Designer ▪ Oracle Hyperion Applications ▪ Oracle Hyperion Essbase ▪ Oracle Warehouse Builder ▪ Pivotal Greenplum ▪ PostgreSQL	Server ▪ Select ▪ Sparx Architect ▪ Syncsort ▪ Tableau ▪ Talend ▪ Teradata ▪ Tigris ▪ Visible ▪ W3C DTD & XSD Schema

Table 2: OEMM harvesting bridges for importing technical metadata

Looking beyond the vast number of metadata harvesting bridges that are directly supported by and compatible with Oracle Enterprise Metadata Management it also becomes necessary to support a rich import/export API for bringing glossary and technical metadata into the governance environment. The OEMM tool provides an exceptionally rich Excel/XLS formatted means to import complex metadata and begin governing metadata from any system – even those that are not formally certified with the Oracle tools.

Align with Other Oracle Solutions

One very important key to a successful data governance program is to connect it directly into the operational data and data flows that drive applications and integration points within IT and business intelligence use cases. Without this direct connection to production, the data governance solutions will simply become a separate and disconnected environment and run the risk of being out of date with the actual truth of what is happening in production systems. To make this kind of direct connection easier, there are several ways that OEMM and EDQ can be leveraged together with other Oracle products as part of a comprehensive data governance program, let's take a look at how:

- **Oracle Data Integrator (ODI)** – ODI is a data integration tool for executing data transformations where the data resides. Frequently during data transformations it is necessary to execute data quality rules, and ODI is integrated with EDQ both via the ODI Studio user interface, as well as via the runtime execution of the data flows. Additionally, ODI metadata is fully integrated with OEMM so that all data transformation logic and data movement activity can be documented as part of an end-to-end data lineage.
- **Oracle GoldenGate (GG)** – GG is a real-time replication tool for capturing and moving data transactions with high speed and reliability. Typically data cleansing operations would be performed at the application layer or in batch on the data warehouse, but integration with EDQ is possible using the GG user exits. Additionally, GG replication data flows may be included in OEMM for visually depicting data lineage along with other data sources.
- **Oracle Data Relationship Management (DRM)** – DRM is used to manage changes in reference data and business hierarchies across operational systems. EDQ can leverage DRM reference data as input to EDQ cleansing rules via lookup tables and/or via application APIs (e.g., MDM or CRM). DRM reference data may also be exchanged with OEMM via import/export formats as a way of seeding business taxonomies in either tool.
- **Oracle Master Data Management (MDM)** – MDM tools are used to govern customer, product or citizen data records for operational and analytic reporting. EDQ is pre-integrated with Oracle MDM via a pre-built

connector that allows EDQ to provide the name and address cleansing features for MDM. OEMM can be leveraged with Oracle MDM as a way to externalize and govern the CDEs of MDM attributes from directly within the OEMM Business Glossary.

Oracle has always been committed to an Engineered Systems strategy that aims to make it easier to use Oracle products together; this is part of the core investment philosophy in the data governance area as well.

Consider Deployment Options

These days more than ever customers are looking for flexibility in how they choose to consume and deploy enterprise software. Fortunately, Oracle recognizes this and provides options on how to deploy data governance that include normal software installation, deployment on engineered systems, and deployment on both public and privately managed cloud ecosystems.

- **Bare Metal Deployments** – both OEMM and EDQ support software installations on Intel®-based Windows and Linux platform servers from commodity hardware providers.
- **Oracle Exalogic Deployment** – Oracle Exalogic Elastic Cloud is hardware and software engineered together to provide extreme performance, reliability, and scalability. It runs both Intel® x86 and Solaris chipsets, and works with Oracle, Java, and other applications, to deliver low total cost of ownership, reduced risk, higher user productivity, and one-stop support. Both OEMM and EDQ runtime processes can run directly on the Exalogic system.
- **Oracle Exadata Deployment** – Oracle Exadata is the highest performing and most available platform for running Oracle Database. Its Intel® x86 and Solaris architecture features scale-out industry-standard database servers and intelligent storage servers, and a high-speed InfiniBand internal fabric that connects all servers and storage. Both OEMM and EDQ are certified to use an Exadata host for their internal database repository requirements.
- **Oracle Big Data Appliance Deployment** – Oracle Big Data Appliance is an integrated, optimized, and tuned Hadoop environment running Intel® x86 architecture to rapidly deliver comprehensive and secure big data capabilities to enterprises. The appliance offers dramatically shortened deployment times and reduced risk result in an overall lower total cost of ownership. OEMM is certified to harvest metadata from the Cloudera Hadoop instance running on the Big Data Appliance and EDQ may be used in combination with Oracle Data Integrator to cleanse any data being used in the big data environment.
- **Oracle Managed Cloud Deployments** – Oracle delivers dedicated managed cloud services across its broad portfolio of business applications, middleware, database, and hardware. Oracle Managed Cloud Services are complete, secure, and customized to fit specific customer business requirements. Both OEMM and EDQ may be deployed, hosted and managed within the Oracle Managed Cloud environment.
- **Oracle (Public) Cloud Deployment** – Oracle Cloud is the industry's broadest and most integrated public cloud, offering best-in-class services across software as a service (SaaS), platform as a service (PaaS), infrastructure as a service (IaaS), and data as a service (DaaS). Oracle Enterprise Data Quality is certified to be deployed, hosted and managed within the Oracle Cloud environment as part of the Oracle Java Cloud Service.

Benefits of an Engineered Systems Approach

In the examples above, Oracle Exalogic, Exalytics, Exadata and Big Data Appliance are examples of engineered systems powered by Intel®. When customers decide to use Oracle and Intel® engineered systems they receive the benefits of a high performance and scalable platform along with the advantages of extremely fast deployment and provisioning of new solutions. Further, there are additional security and transparency benefits that accrue because of the shared access controls derived from hardware and software such as Oracle Database and Oracle Big Data SQL. Operating data governance solutions such as OEMM and EDQ with engineered systems powered by Intel® benefit from these performance, time-to-value, and security advantages.

Intel® latest Xeon® processors allows for repurposing the server by varying core counts and frequencies to meet the needs of different workloads. It also simplifies system configuration process allowing dynamic repurposing of assets. Intel®'s engineers have collaborated with Oracle on enhancements that include threading support for overall better performance, improved hash calculation for fewer conflicts, less time running unnecessary code executions, and faster data encryption and decryption with Intel®Advanced Encryption Standard New Instructions (Intel®AES-NI).

Conclusion and Next Steps

Successful data governance programs have strong people, process and technology components. In this whitepaper, we reviewed how Oracle and Intel-based engineered systems provide a solid technology backbone to support data governance programs. For more information about OEMM, please visit http://www.oracle.com/us/products/middleware/data-integration/enterprise-metadata-management/overview/index.html. For more information about EDQ, please visit http://www.oracle.com/us/products/middleware/data-integration/enterprise-data-quality/overview/index.html. To download evaluation software, please go to https://www.oracle.com/downloads/index.html .

About the Author

Sunil Soares is the Founder and Managing Partner of Information Asset, a consulting firm focused on Data Governance and Enterprise Data Management. He is the author of several books, including *Selling Information Governance to the Business* and *Big Data Governance*. For more information, please visit www.information-asset.com.

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US


blogs.oracle.com/oracle


facebook.com/oracle


twitter.com/oracle


oracle.com

Data Governance with Oracle
September 2015
Author: Sunil Soares
Contributing Authors: Jeff Pollock


Oracle is committed to developing practices and products that help protect the environment