Governance through Data Controls and Data Quality Service Level Agreements

David Loshin
Knowledge Integrity, Inc.
www.knowledge-integrity.com
loshin@knowledge-integrity.com
(301) 754-6350
Agenda

- Actualizing Data Governance
- Data and Process Controls
- Modeling the Business Process
- Identifying and Managing Critical Data Elements
- Data Quality Expectations and DQ Metrics
- DQ Service Level Agreements
- DQ Issues Tracking, Monitoring, and Reporting
- Summary
Objectives of Data Governance

- Objectives of Governance:
  - Identify explicit and hidden risks associated with data expectations
  - Actualize implementation of business policy
  - Provide framework for audit
  - Oversee management of critical data elements
  - Manage conformance to data quality service level agreements
  - Oversee identification, reporting, resolution of data issues

Provide management oversight for organizational observance of different kinds of information policies
Business Policies Rely on Information

- Business policies are defined or imposed to constrain or manage the way that business is performed.
- Each business policy may loosely imply (or even explicitly define):
  - Data definitions
  - Information policies
  - Data structures and formats
- Business policy oversight is related to information policy oversight.
Example: Privacy Policy

- The company does not rent, sell, or share personal information about you with other people or nonaffiliated companies except to provide products or services you've requested, when we have your permission, or under the following circumstances:
  - We provide the information to trusted partners who work on behalf of or with the company under confidentiality agreements. These companies may use your personal information to help the company communicate with you about offers from Yahoo! and our marketing partners. However, these companies do not have any independent right to share this information.
  - We have a parent's permission to share the information if the user is a child under age 13. Parents have the option of allowing the company to collect and use their child's information without consenting to the company’s sharing of this information with people and companies who may use this information for their own purposes;
Example, continued

- “We have a parent's permission to share the information if the user is a child under age 13. Parents have the option of allowing the company to collect and use their child's information without consenting to the company’s sharing of this information with people and companies who may use this information for their own purposes;”

- Example information policies:
  - We maintain data about a user’s age
  - We verify user age before sharing user’s information
  - We check for parent’s permission for users below the age of 13
Business Policies

- Business policies drive the way the organization does business
- Internal governance oversees compliance with business policies
- Data governance reflects the way that information policies both support the business policies and impose data rules
- At the core, the information policies depend on:
  - Recognized critical data elements
  - Agreed-to business definitions
  - Measurable rules to assess conformance
  - Ability to roll-up metrics to characterize business policy compliance
- *Ability to report when data does not meet expectations before potential impacts are incurred*
Drivers for Governance Require *Control* and *Audit*

- Compliance
- Auditability
- Transparency
- Reportability
- Agility
- Risk awareness
Data and Process Controls

Controlling the process means identifying the violation of data expectations at the earliest point in the processing stream, thereby providing an opportunity to mitigate the eventual business impacts.
Every process has embedded controls

Objective: ensure that issues that might impact processing stream are identified and documented early in the processing stream

Effectiveness demonstrated when:
- Control events occur when data failure events take place,
- The proper mitigation or remediation actions are performed,
- The corrective actions to correct the problem and eliminate its root cause are performed within a reasonable time frame, and
- A control event for the same issue is never triggered further downstream
Data Quality vs. Data Control

- Data quality is about “fixing things”
  - Parsing, standardization, normalization, transformation, correction
  - *Monitors and fixes the data*
- Data control is about “observation and reporting”
  - Inspection, validation, notification, documentation, issue reporting, issue tracking
  - *Monitors and fixes the process*
Data Control

- Continually review conformance with stated expectations as data moves through a processing stream:
  - Reduce the number of errors to a reasonable and manageable level
  - Enable the identification of data flaws along with a protocol for (conditionally) halting the processing stream
  - Instituting a mitigation or remediation of the root cause within an agreed-to time frame
Types of Controls

Controls can be applied at different levels of granularity:

- **Data element** controls review the quality of the value in the context of its value assignment to a data element.
- **Data record** controls examine the quality of the set of (element, value) pairs within the context of the record.
- **Data set and data collection** controls focus on completeness of the data set, availability of data, and timeliness in its delivery.
The Control Process

- Controls measure observance of data expectations based on governance drivers/objectives:
  - Compliance
  - Auditability
  - Transparency
  - Reportability
  - Agility
  - Risk awareness

- Controls are placed at relevant locations within the process stream.

As data passes between process tasks, controls check the completeness, consistency, timeliness against defined expectations.
For any business process, “active governance” requires:
- Understanding of the business process supported by the application set
- Identification of specific governance objectives
- Determination of business policies that reflect the governance objectives
- Delineation of the corresponding information policies
- Mapping the end-to-end business processing stream
- Identifying the critical data elements and their owners
- Defining data expectations at appropriate level of control
- Identifying the relevant locations in the end-to-end process stream where data expectations can be measured
- Defining metrics for the controls
- Defined service level agreements for control metrics
Business, Process, Model

- Applications are intended to implement business operations
- There are two aspects of understanding the business:
  - Documenting how the business is *supposed* to work
  - Documenting what the applications *actually do*
- We can construct a business process model to describe the end-to-end processing stream
Business Process Modeling

- A “business process” is a coordinated set of activities intended to achieve a desired goal or produce a desired output product.
- Models are designed to capture both the high level and detail of the business process.

![Diagram showing inputs, activity, controls, and output with actors involved.]
Sharing Enterprise Data Objects

- Interactions between activities depend on shared data:

  - Instance values of common data types representing business facts communicate input and control during the business process.
Connecting Data to Business Policies

- At the conclusion of every processing stream, there are expected “side effects”
  - Transactions have completed
  - Reports have been generated
  - Notifications have been transmitted
  - Events have been triggered
- The success of the application depends on whether the side effects have completed satisfactorily within the expected tolerance levels
- Data issues that may impede this conclusion are directly tied to business policy compliance
Define/Refine Information Policies

- Embody the specification of management objectives associated with data governance
- Relate assertions to related data sets
- Articulate how business policy is integrated with information asset
- Example: Anti-money laundering
  - Establishing policies and procedures to detect and report suspicious transactions
  - Ensuring compliance with the Bank Secrecy Act
  - Providing for independent testing for compliance to be conducted by outside parties.
Data and Traceability

- Data issues that impede achievement of business objectives involve specific data sets and corresponding data elements.
- Identifying the data elements that directly contribute to business activities provides a starting point for establishing data rules that can be monitored across the processing stream.
  - Critical data elements
  - Data quality rules
  - Thresholds
Segregating Data Handoffs

- In turn, evaluating the catalog of data elements through which business processes are coordinated yields a candidate list of relevant data handoffs that are subject to control.
- Assembling that catalog enables a review of the business application services and their use of critical data elements.

*Identify application services at the business level as they access exchanged data objects.*
Information Dependency and Critical Data Elements

- Each information policy reflects data rules
- Each data rule refers to one or more data elements and assertions about their values
- Examples:
  - *We maintain data about a user’s age* implies the existence of a “birth date” attribute
  - *We verify user age before sharing user’s information* implies that we compare user birth date prior to sharing
  - *We check for parent’s permission for users below the age of 13* implies the existence of a “parent’s permission” attribute
- These also imply functional requirements in terms of data capture, use, and retirement
What are Critical Data Elements (CDEs)?

- Business facts that are deemed critical to the organization
- Example criteria:
  - Is used to support part of a published business policy
  - Is used by one or more external reports
  - Is used to support regulatory compliance
  - Is designated as Protected Personal information (PPI)
  - Is designated critical employee information
  - Is recognized as critical supplier information
  - Is designated as critical product information
  - Is designated as critical for operational decision-making
  - Is designated as critical for scorecard performance

- Critical Data Elements are the core building blocks for establishing information policy, and consequently, business policy compliance
CDEs and Information Policy

- CDEs are critical to ongoing operations, and must be subject to oversight
- Processes are defined by which critical data is governed, including:
  - Data registration
  - Monitoring and control
  - Data exchange and sharing
  - Data lifecycle events (creation, modification, retirement, etc.)
Data Ownership/Responsibility/Accountability

- Data governance policies provide guidance for how critical information is managed and monitored.
- Governance must account for:
  - “Ownership” = ultimate accountability
  - “Stewardship” = sets of responsibilities
  - “Quality” = ensuring that the data meets the combined enterprise expectations for policy compliance
- Stakeholders: Those whose business processes rely on consistent, high quality information
- Participants: Those who are involved in the definition, refinement, and endorsement of data standards as well as those involved in the development and deployment of those data standards
  - Goal: *Try to make all of them happy*
Objectives for Using CDEs

- Identify critical data elements
- Define/Refine information policies
- Describe metrics and measurements
- Create processes for:
  - Inspection
  - Measurement
  - Monitoring
  - Evaluation
  - Event generation
  - Tracking and evaluation
Critical Data Elements

- Identify enterprise metadata in use across the organization and:
  - Clarify unambiguous definitions, formats, and semantics
  - Facilitate agreement to those definitions and semantics from all stakeholders
  - Absorb replicated reference sets into a single managed repository
Managing CDE Metadata

- Business metadata registry maintains:
  - Structure (size, length, format, precision, etc.)
  - Semantics (definition, aliases, business terms, conceptual domains, etc.)
  - Connectivity (associated data elements, reverse index to use)
  - Data quality rules
  - Business rules
Infrastructure Services

<table>
<thead>
<tr>
<th>ISO-11179</th>
<th>Dictionary</th>
</tr>
</thead>
<tbody>
<tr>
<td>ebXML</td>
<td>GJXDM</td>
</tr>
<tr>
<td>Dublin Core</td>
<td>...</td>
</tr>
</tbody>
</table>

External Standards

Common Business Language

Reusable Metadata and Data Objects

Facilitation of Standards for information sharing

ODS

DW

OLAP

Data Access

Dashboards

CDE Registry

Enterprise Standards

MDM Registries
CDE Basics

- Conceptual domain: A set of valid value meanings
  - Example: US States – The set of primary governmental divisions of the United States.

- Value domain: A set of permissible values

- Data Element Concept: A concept that can be represented in the form of a Data Element, described independently of any particular representation.
  - Example: Issuing Jurisdiction – The governing body in which a license was issued

- Data Element: A unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes.
Conceptual & Value Domains

More than one Value Domain may be associated with a single Conceptual Domain

"US States"


AL, AK, AZ, AR, CA, CO, CT, DE, DC, FL, GA, OK, HI, ID, IL, IN, IA, KS, KY, LA, ME, MD, MA, MI, MN, MS, MO, MT, NE, NV, NH, NJ, NM, NY, NC, ND, OH, OR, PA, RI, SC, SD, TN, TX, UT, VT, VA, WA, WV, WI, WY

© 2007 Knowledge Integrity, Inc.
www.knowledge-integrity.com
One Value Domain may be associated with more than one Conceptual Domain.
Value Domain Reuse

One Value Domain may be used by many Data Elements

IssuingState

DriverState

FIPS 2-Character State Abbreviation
AL, AK, AZ, AR, CA, CO, CT, DE, DC, FL, GA, OK, HI, ID, IL, IN, IA, KS, KY, LA, ME, MD, MA, MI, MN, MS, MO, MT, NE, NV, NH, NJ, NM, NY, NC, ND, OH, OR, PA, RI, SC, SD, TN, TX, UT, VT, VA, WA, WV, WI, WY

DMVWorkerState

CaseState
Metadata Registries and CDEs

- A system for organizing information about data that is exchanged
- A system and tool for administering the data standards process
- A browsable gateway to access organizational “metadata”
  - Data element definitions, business terms, object structure, exchange schemas, business rules, exchange workflows
Registration

- Manage how the registry is used to administer the data standards process:
  - Oversee data standards submissions
  - Assign identifiers and review names
  - Propagate appropriate information to data stewards
  - Facilitate the progress of submittals through the registration process
  - Manage the data standards workflow status
  - Alert submitters to additional requirements and decisions
Data Quality Expectations and DQ Metrics

- Data expectations for business reflect level of satisfactory conformance to data quality rules
- Data quality rules are defined in terms of “dimensions of data quality”
What are “Dimensions of Data Quality”? 

- The concept of a dimension evokes thoughts of measurement
- Different dimensions are intended to represent different measurable aspects of data quality
  - Used in characterizing relevance across a set of application domains and to ensure an enterprise standard of data quality
  - Measurements are taken to review data quality performance at different levels of the operational hierarchy
  - Monitoring overall both line-of-business and enterprise performance
- Each group within the organization has the freedom to introduce its own dimensions with customized characteristics.
Categorization of Dimensions

- Governance
- Conformance
- Application
  - Policies
  - Operations
  - Standards
Objective Data Quality Dimensions

- Characteristics
- Criteria
- Metrics
- Thresholds
Dimensions of Data Quality

- Intrinsic
  - Accuracy
  - Provenance
  - Semantic
  - Structure
  - Rating

- Contextual
  - Timeliness
  - Currency
  - Completeness
  - Consistency
  - Aggregation
  - Transformation

- Conformance
  - Fitness
  - Compliance
Caveat

- This list is intended as guidance and as a starting point for defining the dimensions that are relevant within the organization.
- The methods for measurement should be identified before agreeing to the selection of a metric.
- The metrics should be correlated with the impacts identified during the impact analysis.
Sample Dimensions

- Timeliness
- Currency
- Completeness
- Consistency
Timeliness

- Accessibility
  - Newly posted records should be available to enterprise applications within a specified time period
  - Policies specifying acceptable time delays must be provided.
- Response time
  - Ensure that requested data is provided within the acceptable time period
  - Expectations for response time must be specified
Currency

- **Age/Freshness**
  - The acceptable time period lifetime between updates for each data element is defined - Expiry date

- **Time of release**
  - The date/time upon which the data becomes available is defined
  - If data is expected to be delivered to specified participants, the release date/time should be specified

- **Synchronization of replicas**
  - Data synchronizations and replication policies between systems must be specified

- **Correction/update promulgation**
  - Policies for promulgation of corrections and updates, must be specified.

- **Temporal**
  - Temporal Consistency rules are valid
Completeness

- Population Density
  - Specify the minimum degree of population for each data element

- Optionality
  - Mandatory attributes are expected to have assigned values in a data set
  - Optionality must be specified for all data elements

- Null validation
  - Null value rules for all data elements are defined
  - Null value rules are conformed to
Consistency

- Presentation
  - *Common presentation formats for each data element are defined*

- Presentation completeness
  - *Each data presentation format can convey all information within the attributes*

- Null presentation
  - *Standards for the presentation of missing information for each data type are defined*

- Capture and collection
  - *Data entry edits and data importation rules should be defined for each data element*
Defining Data Expectations as Rules

- Data quality rules can be used to monitor conformance to data expectations as dictated by information policies.
- Conformance can be measured, thresholded, and reported at each handoff location in the processing stream.
- Specific failures can generate events as directed by Data Quality Service Level Agreements.
DQ Service Level Agreements

- An SLA is a “contract” between service providers and their customers detailing the specific services provided as well as levels of:
  - Availability
  - Performance
  - Operation
  - Cost
  - Duration
  - Time for resolution

In order to meet the customer’s business needs
SLA Monitoring

- The SLA will also describe:
  - How service levels are measured
  - How service levels are tracked
  - Roles
  - Responsibilities
  - Metrics
  - Thresholds
  - When events are generated
  - Notification strategies
  - Escalation strategy for identified issues
Inspection & Control

- **task**

- **Task**
  - Data set controls
  - Data record controls
  - Data element controls

- **issues**
  - DQ Issue Tracking

© 2007 Knowledge Integrity, Inc.  
www.knowledge-integrity.com
DQ Issue Tracking

- Alerts are generated when inspection shows that control indicates missed objectives
- Process to log and notify:
  - Description
  - Characterization
  - Prioritization
  - Routing
  - Start the clock...
- Measures added to metrics repository
- Resolution terms dictated by SLA
Data Quality Control

- Data Quality SLA
- DQ Issue Tracking

Issues flow into the Data Quality SLA system, which then tracks and manages issues.
Metrics and Measurement

- Decompose information policies into specific measurable data rules
- Apply tools and techniques for measuring conformance to data rules (think: data profiling)
- Metrics can be “rolled up” from data rules defined as a by-product of analyzing the information policy
- Institute protocol for alerting key staff members when controls trigger data quality events
- Establish agreements for resolving issues within a reasonable time frame
- Monitor conformance to service level agreements
Monitoring and Evaluation

- One business policy can encompass multiple information policies
- Each information policy may encompass multiple data rules
- Each data rule, therefore, contributes to monitoring compliance with business policy!
What Makes a Good Metric?

- Clarity of Definition
- Measurability
- Business Relevance
- Controllability
- Representation
- Reportability
- Trackability
- Drilldown Capability
Metrics repository captures collected metrics from assorted measurement systems in a single repository for reporting and analysis.
Data Quality Dashboard

- Identify key performance indicators
  - Select them from the list of characteristics of data quality dimensions
- Provide a reporting scheme
  - Use existing reporting and dashboard applications, if possible
- Integrate with the data governance framework
1. What are the most effective methods for overseeing compliance with stated performance goals?
2. What currently absent measures/metrics are needed to fully support achieving the business objectives?
3. Are higher level metrics dependent on the proper set of supporting metrics?
4. What future business objectives and performance goals are not covered by existing metrics?
5. Are all business terms in metric descriptions defined?

---

1. What processes are in place for notification and mitigation of issues raised by the metrics?
2. What actions do you take to make sure that your team meets stated performance goals?
3. What barriers exist to meeting stated performance goals?
4. What is the process for modifying and/or deleting a metric?
5. What is the process for creating a new metric?

---

1. Are there expected metrics that have no data source?
2. Does metrics reporting depend on any other data management resources? (e.g., CDE registry, metadata repository, CDW)
3. Do metric scores indicate when and which actions are to be taken to meet stated performance goals?
4. Who are the consumers of each metric?
5. Is each metric associated with a control/decision point in a business process?

---

1. Is the process for collecting the metric score clearly defined?
2. Are you tracking metric score/performance against stated performance goals on a regular (defined periodic) basis?
3. Are the measurement dimension values at the proper level of granularity?
4. Does each metric accurately reflect (and does the presentation accurately report) the aspect of the business that is being measured?
5. Are all aspects of each metric properly reported? (e.g., meeting SLAs within the agreed-to timeframe)
6. Are metric thresholds and evaluation criteria clearly defined?
Summary

- Collect business artifacts
- Articulate business governance objectives and business policies
- Compose a business process model
- Identify critical data elements
- Assess information policies
- Define data quality expectations
- Define data quality rules
- Agree to a DQ SLA
- Institute inspection
- DQ Issues tracking
- Monitoring, reporting: Governance Dashboard
- Process and Protocols for mitigation and remediation
Questions?

- If you have questions, comments, or suggestions, please contact me

  *David Loshin*
  
  (301) 754-6350
  
  *loshin@knowledge-integrity.com*