Business Data Glossary & Metadata Management

Arturo Salazar & Vincent Rejany – SAS SWE
DM BSM
Incubator for sustainable data governance
IT IS ALL ABOUT DATA

EVERYWHERE!
EVEN MORE WITH ANALYTICS!
HOW WEIRD?

PEOPLE

FINANCE

LEGAL

DATA

HR DEPARTMENT

FINANCE DEPARTMENT

LEGAL DEPARTMENT

DATA GOVERNANCE
DO YOU GOVERN YOUR DATA? HOW?

“We have been trying but it is too complex to set up?”

“I don’t know it should be managed by Business Users”

“I don’t know it should be managed by IT”

“Business Glossaries, yes we have many in Excel, but nobody use them”

“Each department is managing its own data”

“We have Metadata … everywhere”

“Of course, We will start a corporate MDM project soon”

“SAP is our ERP, we do not need Data Governance”

“Why? IT is setting up a Data Lake”

Come on, what is the ROI?
DATA GOVERNANCE IS NOT A RESOLUTION RATHER A CULTURE SHIFT
ALWAYS START DATA GOVERNANCE WITH A PLAN

ALWAYS GO TO THE GYM WITH A PLAN.

IFUCKINGLOVEFITNESS.COM
WHAT DOES DATA GOVERNANCE NEED FOR BEING SUCCESSFUL

**SPONSORSHIP**
- Get an executive-level mandate for sponsoring the culture shift

**PEOPLE**
- Business & IT collaboration
  - Committed profiles

**PROCESSES**
- Remediation & escalation workflows
  - Project policies

**QUICK WINS**
- Achievable objectives
  - Move step by step

**TECHNOLOGY**
- Data Quality Controls
- Business Glossary
- Metadata Management
LET’S TAKE AN EXAMPLE

The G20 endorsed at the Los Cabos Summit in June 2012 a Global Legal Entity Identifier System (GLEIS) which uniquely identifies parties to financial transactions.

Key building block for improvements in the quality of financial data across the globe and is to be assigned to all entities that are counterparties to financial transactions.

20-character alphanumeric code, defined in the ISO 17442 standard

J.P.Morgan Trust III: 5493000D71UBB8X2RY09
CONTEXT - LEGAL ENTITY IDENTIFICATION (LEI)

Retail
Wholesales
Investment

Ability to identify parties involved in associated financial transactions

LEI: 5493000D71UBB8X2RY09
J.P.Morgan Trust III  “Morgan”  “JP Morgan”

Customer On-boarding
Processing and settlement efficiency

Risk Calculations
Strengthen the accuracy, integrity and aggregation of data across entities & subsidiaries

Regulatory Reporting
Easier to combine data for aggregation & analysis
WHAT DOES IT MEAN IN TERMS OF DATA GOVERNANCE

What is the definition? ground rules?

Who is Responsible? Accountable? Consulted? Informed?

What is the risk, remediation process if missing/incorrect?

Which system of reference?

What are the applications /systems impacted

How to enrich our existing these systems/applications?

Who will be able to access/create /modify it?

How to monitor compliancy?

Are there any future or ongoing projects impacted?
DATA GOVERNANCE– COLLABORATION IS FUNDAMENTAL

**BUSINESS USERS**

- Define & Document Business Terms
- Describe Business Terms Requirements & Data Controls to be performed

**STEWARDS**

- Leverage Business Terms to enable data alignment
- Lead data assessment for validating top down analysis
- Review data monitoring reports and dashboards
- Coordinate remediation process between business users & IT

**IT TEAM**

- Implement & execute the data quality controls
- Relate data management operations and content to business terms
- Ensure data controls execution and alignment with business requirement
Create a common language for business data so business users can collaborate on the definition and management of business terms across the enterprise relate to risk data aggregation and reporting processes.

Building a Framework for a Data Dictionary/Business Glossary including:
- Roles & Responsibilities
  - Ownership of terms
  - Term maintenance
- Information Gathering & Taxonomy Management
  - Building a library of categorized terms
  - Process to contribute term information
- Reference data centralization
  - Lists of Values, Look Up tables, Hierarchies
THE VALUE OF BUSINESS METADATA

- The Language of the Business, independent of technology—used to:
  - Define authoritative meaning
  - Increase and share understanding throughout the enterprise
  - Establish responsibility, accountability, & traceability
  - Represent business hierarchies
  - Document business descriptions, examples, requirements, valid values,
  - Find relevant information assets

- Tie business terminology to physical objects that use the terminology

- Understand relationships in an enterprise
BUSINESS GLOSSARY - SAS BUSINESS DATA NETWORK DEMO
SAS BUSINESS DATA NETWORK -

- SAS Business Data Network is a patented business data glossary, providing to organizations with a single point of entry to develop, document and share common vocabulary

- Enables data governance
  - Common language supports compliance regulations such as Basel II
  - Represent and expose business relationships
  - Track history of changes

- Accountability and responsibility
  - Document and communicate ownership
  - Notify interested parties on changes
  - Users can send terms into workflow for review and approval before publishing

- Supports better collaboration
  - Capture and share annotations between team members
  - Greater understanding of the context of information
  - Use and reuse of trusted information
Demo SAS BUSINESS Data Network
DATA GOVERNANCE LINE – ASSESS DATA QUALITY

Enables to perform a **Bottom-Up** approach, business rules are discovered and defined based on existing data.
Allows to validate or detail **Top-Down** business rules coming from organization knowledge and data strategy.

**Discover & Profile Data**
- Allows to collect complete information about risk data and create a clear picture of which data quality issues to tackle.
- Supports seamless data discovery and analysis, retrieving standard statistics (null, blank, patterns, frequency distribution, foreign key ...) or custom metrics
- Supports data issues investigation with uniqueness measures and redundant data analysis
- Schedule Risk Data assessment and share reports for continuous audit.
ASSESS DATA QUALITY
COMMUNICATE BOTTOM-UP DATA ASSESSMENT RESULTS

Quickly analyze data source completeness
Identify data outside normal limits easily with Outlier detection
DATA GOVERNANCE LINE – EXECUTE DATA QUALITY CHECKS

Apply, enforce and automate risk data quality standards. Check consistently quality applied for external and reference data through business rules and monitoring.

Address Data Management Challenges through Data Cleansing and Monitoring Actions
- Create and enforce business rules that govern the quality of your data.
- Detect when data meets pre-set limits
- Recognize and correct problematic data before it enter your systems
- Reuse Data Quality operations across the organization
- Integrate Data Quality operations with business applications in batch or real-time
EXECUTE DATA QUALITY CHECKS

Rule 1 – Completeness
A LEI is required for each legal entity.

Rule 2 – Structure
LEI is 18-character alphanumeric + 2-character numeric code (Control Key) (ISO 17442)

Rule 3 – Consistency
LEI checksum algorithm compliant with control key

Rule 4 – Accuracy
Related LEI belongs to the LEI corporate reference database.
Tackle critical risk data management issues
Understand and refine mission-critical processes with reporting tools that allow you to see and address trends
Define remediation process

Properly understand through appropriate reports the costs associated with risk data quality and business rule violations and to ensure organization avoids those costs:
• Monitor data quality controls results through data stewardship console
• Generate scorecards that validate risk data governance and data improvement initiatives
• Broadcast reference data & quality indicators across the organization
• Link risk data quality score with business reports
VALUE ADDED REPORTING & CONSOLIDATION FOR DATA GOVERNANCE

Detailed Operational Reporting
• Analyze data quality controls triggers, by sources, business objects, date …

Consolidated Operational Reporting
• Dashboard Business Rules Results
• Calculate a Global DQ Score and Monitor TRENDS
• Consolidate According to Data Quality Dimensions (Accuracy/Integrity, Completeness, Timeliness)

Key Value Indicators
• Assess incompliant information impact either in terms of risk, severity, or cost.
• Link risk data quality metrics with business reports

Path to Data Governance Excellence
CONSOLIDATED OPERATIONAL REPORTING
AGGREGATE & SHARE DATA MONITORING RESULTS
DATA GOVERNANCE LINE - IMPROVE DATA GOVERNANCE PROCESS

Trace data from source to consumer and all the steps in between as well as document what has been done to the data and how it has been transformed.
Loop back for rescoping data governance initiative.

Link data management operations to business terms definition:
- Trace data from source to consumer and all the steps in between.
- Document what has been done to data and how it has been transformed.
- View relationships across data elements and the impact of a change within the entire glossary of business terms.
- Adapt data controls and specifications according to business priorities.
THE VALUE OF DATA ASSET – FILL THE GAP

Lack of Collaboration
SAS Lineage is a centralized GUI for viewing and exploring the relationships of all business and technical metadata within an organization.

- **Centralized Metadata Viewer GUI**
  - View and Explore the relationships of all business and technical metadata within an organization.
  - Common lineage viewer for all SAS products and solutions

- **Trace metadata from business definition to table field**
  - Search for metadata such as business terms, table or column names, or anything that can be represented by the SAS Metadata Repository.
  - Perform Impact Analysis
  - Create and manage custom views, filter the contents of a view

- **Integrate 3rd Party metadata**
  - REST Services
  - Erwin: data models including tables, columns, column type, formats, UDF information
  - CWM (common metadata model format), supports over 200 different metadata exchange formats
METADATA RELATIONSHIPS - SAS LINEAGE DEMO
Demo SAS LINEAGE
SAS DATA GOVERNANCE

- **Visualization tools** to help data stewards and enterprise architects understand relationships and how changes might impact various reports or other technical assets
- **Role-based Web user interface** for both business and IT users to help in aiding communication and collaboration
- Governance policy enforcement and data remediation via granular **role-based security** and **workflow** to help meeting compliance regulatory requirements
- **Business glossary** for capturing business metadata and relating it to policies, owners and technical metadata (i.e., reports, tables or analytical models)