Business-Level Metadata

Darryl Kerkeslager

Virginia Department of Corrections

July 10, 2014

Major Metadata Problems

Numerous redundant and conflicting silos of data

A great deal of time at project initiation defining terms, and often getting them wrong or having to redefine them

Incomplete or vague ownership of data - who make business decisions about sharing and access to sensitive data, and who controls the definitions

Additional Problems

- Not always sure what to measure, or how to measure it
- Dialogue is difficult without a common language for business and IT
- Subject matter experts have not defined vague definitions
- Aware of gaps in information reporting - but not where they are
- Actual business rules don’t always align with the rules embedded in software and databases
- Unable to supply data as quickly as the business and auditing wants
- No prioritization of data cleansing
- Regression testing sometimes overlooks the consequences of business changes
- No holistic view of the consequences and seriousness of data breaches
- Unable to score compliance with statutes, regulations, and policies
- A statutory obligation to expunge redacted data, but an unclear process
- Incomplete data retention and archiving policies
- Lack of searchable definitions makes training more difficult
- Responding manually and incompletely to the ripple effect of business, database, and application changes

Our Complexity

Multiple Lines of Business …

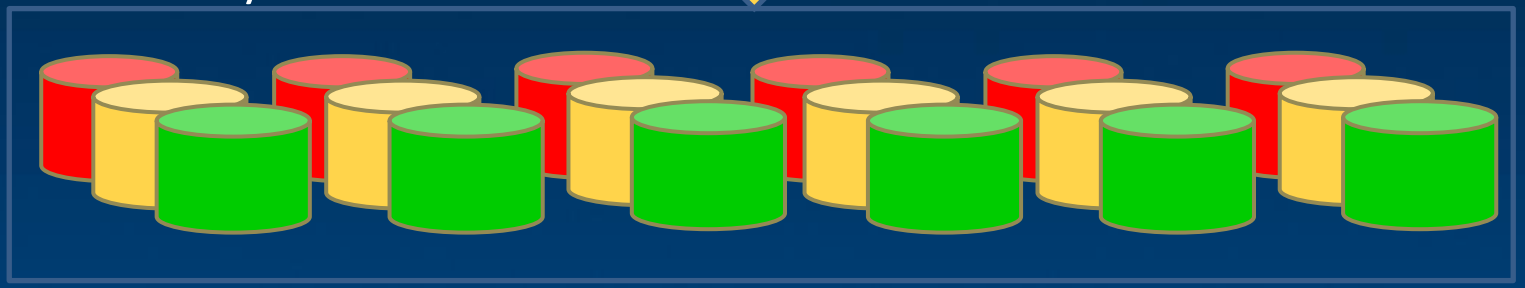
- Architecture & Engineering
- Probation Supervision
- Inmate & Facility Security
- Offender Medical Care
- Agriculture & Production
- Counseling Programs
- Accounting & Payroll
- Research & Forecasting
- Offender Education
- Human Resources
- Offender Financials
- Information Technology
- Parole Board

… but few Data Sources

Most Metadata Vendors Solve …

Few Lines of Business …

… but many Data Sources



Defining Success

- Collaboration – between business units, and IT
- Ownership – of terms, changes, and access
- User Adoption – by business users

Users Require The Ability To …

Core Functionality

- Identify business terms and database columns that have no definition
- Record business terms, definitions, owners, comments, and user discussions
- Create many-to-many relationship between terms and owners
- Create new metadata, not just discover existing metadata
- Create workflow processes for defining and approving terms, sharing, and other requirements
- Track changes over time to data attributes, business rules, report definitions, and business terms
- Extract metadata from databases, source control repositories, and BI tools of multiple vendors
- Prevent a business term from being used for multiple definitions
- Create hierarchal relationship between business terms
- Store database elements that don't yet have a business definition, and business definitions that are not yet associated with a business element

(Continued)

User Interface

- Search and retrieve any current and historical metadata
- Navigate seamlessly from business terms to other related terms and data structures
- Popup linked data using hyperlinks, without leaving the current context
- Enter free text notes with embedded hyperlinks, both Internet and within the application
- Create and use custom list of values
- Use a web based interface with current AJAX/HTML5 technology
- Interactively sort, on any and multiple columns
- Reference external links such as bug ticketing systems
- Define and control application user security and roles down to the atomic user and table level
- View a graphical data lineage interface
- Add custom fields to business and technical metadata
- Integrate with Active Directory and or Exchange

(Continued)

Low-level Functionality

- Profile data using counts, distinct values, distributions, and row samples
- View data lineage, source and target linking, and database metadata out of the box
- Analyze the impact of a proposed change to dependent data objects
- Create rules to ensure or improve data quality
- Link code snippets and standard queries to business definitions
- Create Entity Relationship Diagram based on the relational model of tables, with cardinality

Reporting & Outputs

- Use multiple reports out of the box
- Create customizable reports
- Export data to multiple file formats
- Integrate with SharePoint
- Extract data from repository via web services

How We Evaluated

We narrowed our search to three vendors/products. In the summer of 2013, we invited each for a demonstration and evaluation. Each requirement was ranked by importance and scored. Price, setup, and licensing were also large considerations.

- Collibra Data Governance Center
- IBM InfoSphere Information Server
- SAP Information Steward

We selected Collibra.

The Forrester Wave™: Data Governance Tools, Q2 2014