Going Live with Enterprise Solutions

TopQuadrant™

Reference Data Management in TopBraid EDG

Irene Polikoff and Jack Spivak, TopQuadrant Inc.

February 23, 2017

TopBraid EDG™
Enterprise Data Governance

© Copyright 2017 TopQuadrant Inc.

Today's Program

Reference Data Management in TopBraid EDG

- Value of RDM and Implementation Approach
- DEMO:
  EDG Reference Data Management in Action
- Quick Poll
- Questions and Answers

Jack Spivak

Irene Polikoff

TopBraid EDG™
Enterprise Data Governance

© Copyright 2017 TopQuadrant Inc.

TOPQUADRANT COMPANY

FOUNDATION

- TopQuadrant was founded in 2001
- Standards-based approaches to data semantics — to enable connections and to avoid silos

MISSION

- Empower people and drive results — by making enterprise information meaningful

FOCUS

- Provide comprehensive data governance solutions

© Copyright 2017 TopQuadrant Inc.

ENTERPRISE CUSTOMERS

U.S. AIR FORCE

© Copyright 2017 TopQuadrant Inc.

Reference Data

- What is Reference Data?
- Why does it need to be managed?
- How to get started?
- How does EDG/RDM help?

© Copyright 2017 TopQuadrant Inc.

What is Reference Data?

Critical Data with Critical Impact

- Data that ***classifies and categorizes*** other data used by mission-critical business applications.
- Reference data has major impact – from the integrity of business intelligence reports to system integration.

© Copyright 2017 TopQuadrant Inc.

Importance of Reference Data?

- 20-50% of the tables in a database are Reference Data
- Estimated to make up 40% to 70% of the data used in financial transactions
- The same table occurs in many different applications
- *"**Faulty reference data** can cause monetary loss to a financial services organization. For example, misidentified products or data mismatching can cause transaction failures."*

(CapGemini report: "Reference Data and its Role in Operational Risk Management")

© Copyright 2017 TopQuadrant Inc.

Reference Data and Master Data Management

Reference data is a high value place to start a Master Management initiative

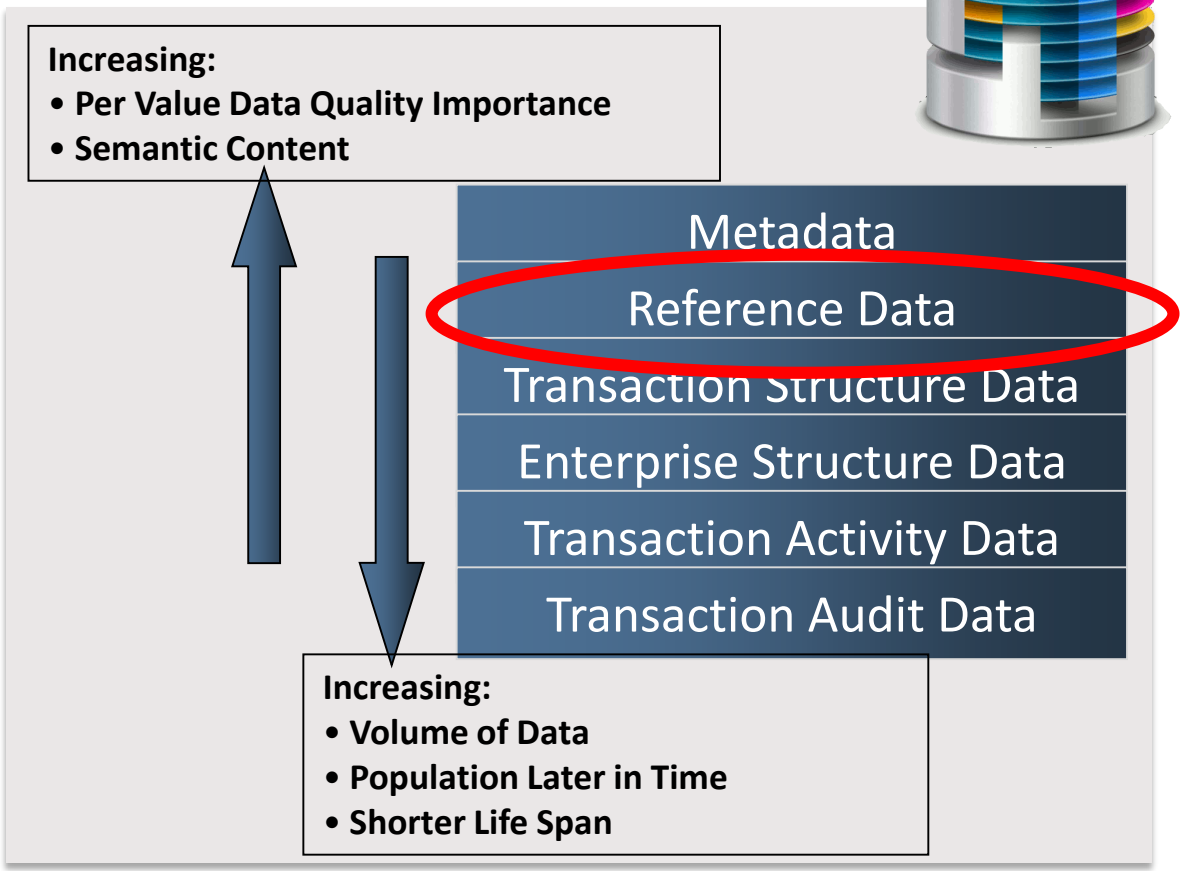
- standard definitions exist
- create policies to use these or provide crosswalks
- data structures are not complex
- strong value proposition

External Reference Data

- defined by standards bodies and industry consortiums

Internal Reference data

- organization specific codes, department codes, geo regions

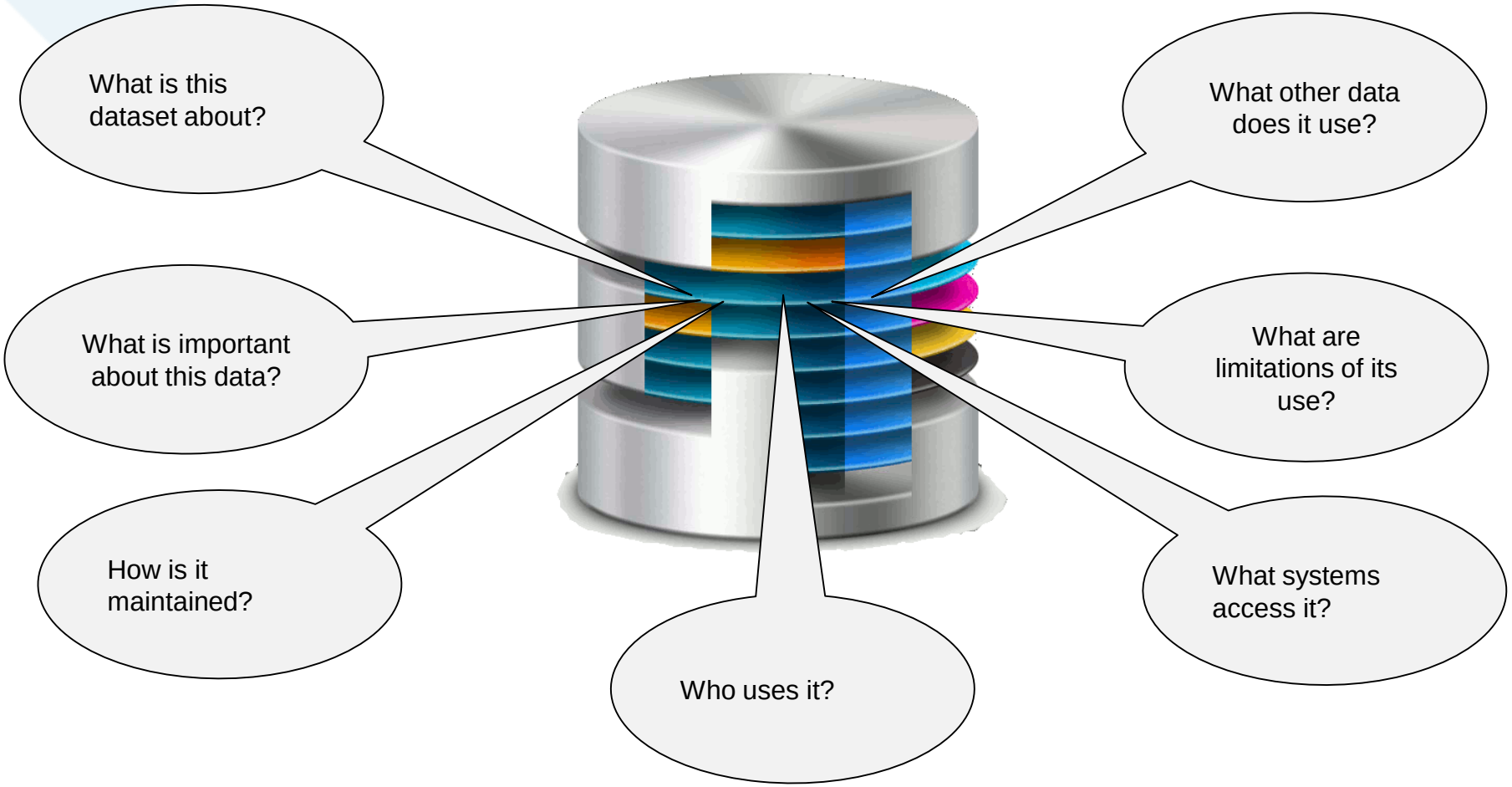


Reproduced with permission from Malcolm Chisholm

© Copyright 2017 TopQuadrant Inc.

What is Comprehensive Metadata?
Data *ABOUT* Reference Data

Metadata captures, communicates and clarifies the meaning of the reference data -- *to answer common questions about its use*



© Copyright 2017 TopQuadrant Inc.

Costs? | Business Users Need Metadata

Confusion, wasted resources and errors result when users can't understand the meaning of codes and how they are used

© Copyright 2017 TopQuadrant Inc.

TopQuadrant™

TopBraid EDG™

Enterprise Data Governance

© Copyright 2011 TopQuadrant Inc.

TopBraid Enterprise Data Governance

© Copyright 2017 TopQuadrant Inc.

TopBraid EDG™
Enterprise Data Governance

© Copyright 2017 TopQuadrant Inc.

TopBraid EDG Data Governance Packages

TopQuadrant™

Business Glossaries

Support for

Business Glossaries

Basic Metadata Management

Support for

Data Assets

Reference Data Management

Management

Support for

Reference Datasets

Ontologies

Crosswalks

Support for

SKOS

Ontologies

Crosswalks

For details, see:

http://www.topquadrant.com/products/topbraid-edg-gov-packs/

© Copyright 2017 TopQuadrant Inc.

Today's Demo

- How to start? Identifying and cataloging where your reference data is located.
- Benefits of a multi-user RDM solution that enables collaboration among all stakeholders
- Benefits of using canonical reference datasets and industry standards
- Connecting reference data management to broader data governance initiatives and goals

© Copyright 2017 TopQuadrant Inc.

How is it done? Preparation

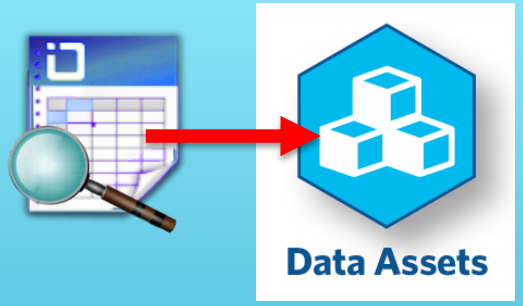
- ✓ *Identify* existing enterprise databases and systems
- ✓ *Catalog* the data landscape in TopBraid EDG

© Copyright 2017 TopQuadrant Inc.

EDG Reference Data Management

1

Import metadata for enterprise DBs into EDG DataAssets

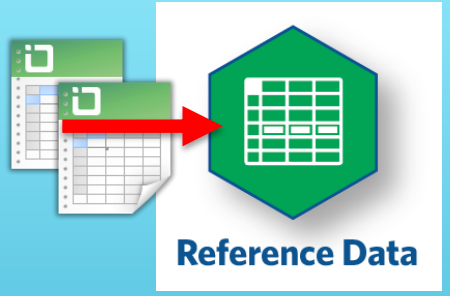


Why?

Interactively explore integrated view of enterprise data structure. Identify Reference Data and where it's used

2

Import discovered reference data into EDG Reference Data sets

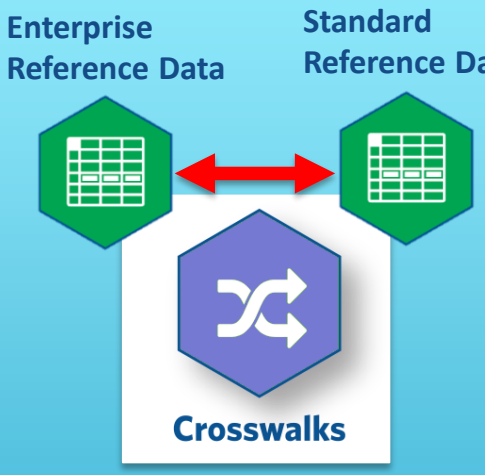


Why?

Manage, control, govern, centralize distribution

3

Create EDG Crosswalks. Connect to standard reference data.



Why?

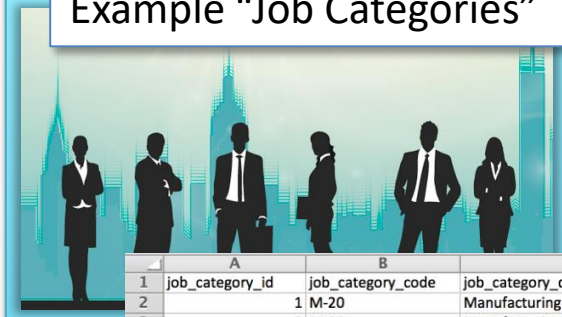
Analyze, Compare, Understand

© Copyright 2017 TopQuadrant Inc.

Today's TopBraid EDG Demo

Harvested Data:
Fictitious Corporate HR Dept.

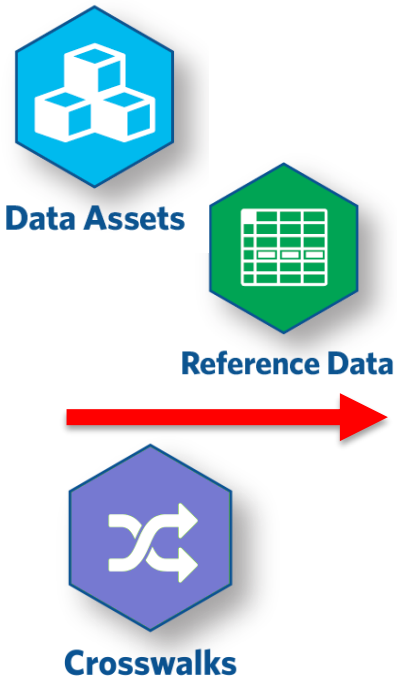
Example "Job Categories"



	A	B	C
1	job_category_id	job_category_code	job_category_desc
2	1	M-20	Manufacturing Management
3	2	M-21	Manufacturing Supervision
4	3	M-22	Manufacturing Logistics
5	4	M-23	Manufacturing QC
6	5	M-24	Manufacturing Fabrication
7	6	M-25	Manufacturing Assembly
8	7	M-26	Manufacturing Packaging
9	8	OS-1	Operations Support - Bookkeeping
10	9	OS-2	Operations Support - Accounting
11	10	OS-3	Operations Support - Procurement
12	11	OS-4	Operations Support - Planning
13	12	S-1	Sales Executive
14	13	S-2	Sales Team Lead
15	14	S-3	Sales Major Account
16	15	S-4	Sales Support Assistant
17	16	S-5	Sales Agent
18	17	S-6	Sales Associate
19	18	E-1	Mechanical Engineering
20	19	E-2	Aerospace Engineering
21	20	E-3	Electrical Engineering
22	21	E-4	HVAC Engineering
23	22	E-5	Software Engineering
24	23	E-6	Engineering Technician
25	24	E-7	Engineering Specialist
26			
27			
28			
29			
30			

Use EDG to:

- ✓ Explore Relevant Data Assets
- ✓ Create Reference Data
- ✓ Connect with Standard Occupational Codes



© Copyright 2017 TopQuadrant Inc.

DEMO

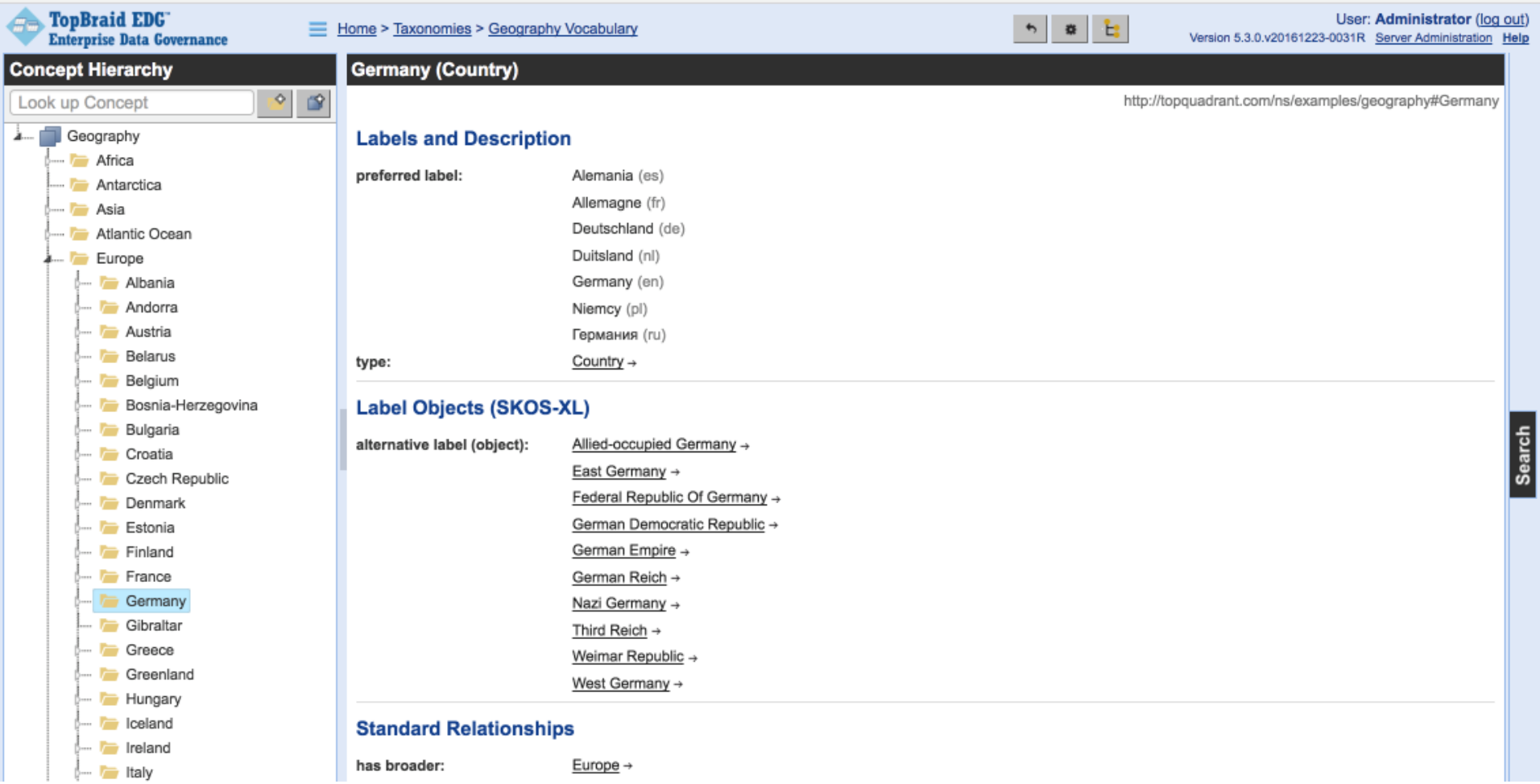
© Copyright 2017 TopQuadrant Inc.

More TopBraid EDG/RDM Capabilities

- Import of multiple RD representations
- Metadata Management
- Relationships to connect and enrich reference datasets
- Crosswalk Mappings and Traceability
- Hierarchy Management
- Documents Reference Data Provenance, On boarding and Usage across Enterprise
- Change Management
- Assignment and tracking of tasks
- RDM Notifications
- Reference Data Quality and Validation
- Reference Data Provisioning
- APIs (RESTful and JMS), integration and customization

© Copyright 2017 TopQuadrant Inc.

Multi-language support



© Copyright 2017 TopQuadrant Inc.

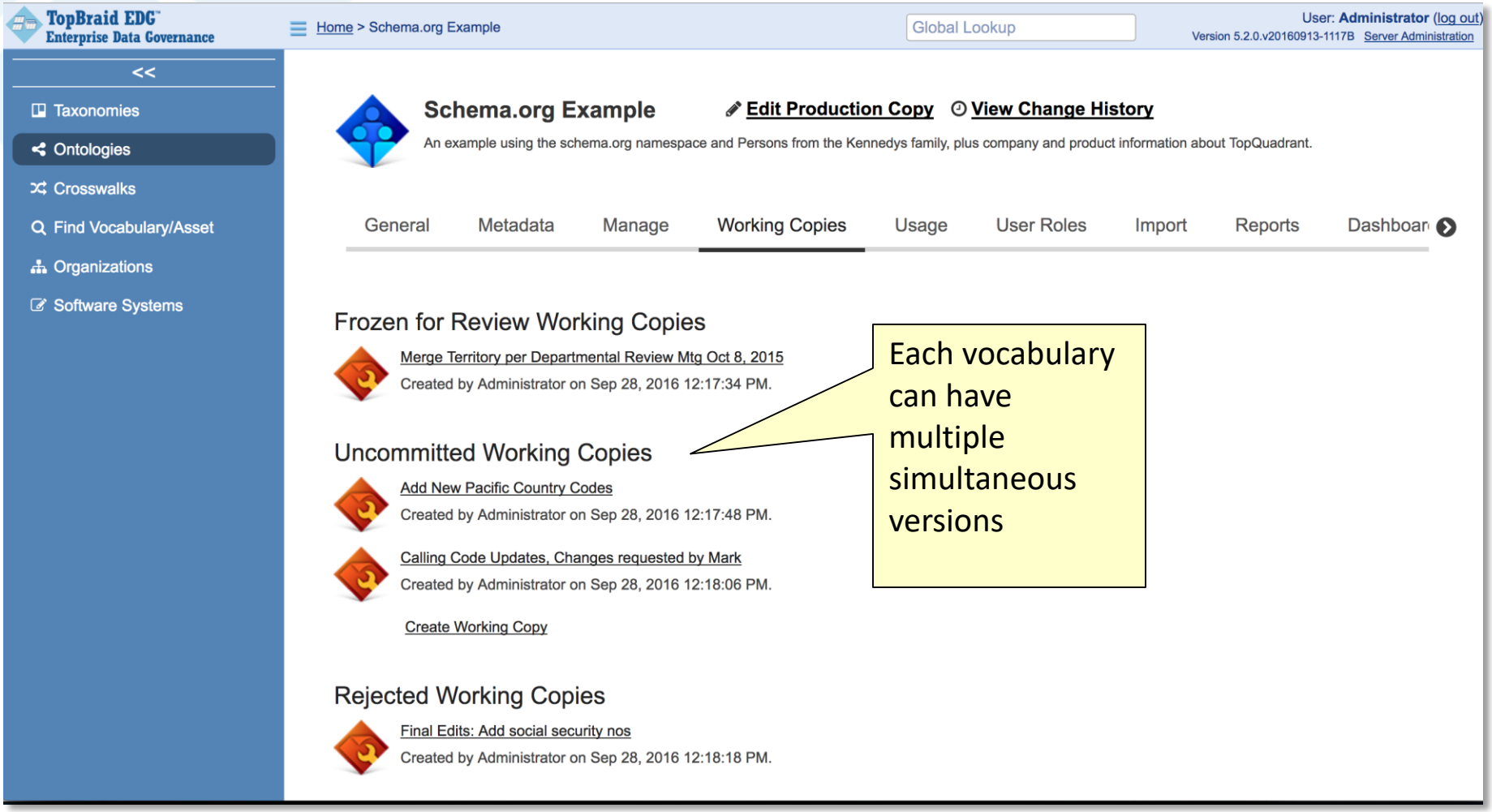
Role-based Access Control

- Assign users different roles for using the system at various levels of authority per asset:
  - Manager
  - Editor
  - Viewer

- Manager and editor are typically data stewards and curators for different assets
- System-wide read-only role supports support search and/or comment only use cases

© Copyright 2017 TopQuadrant Inc.

Workflow and Change Management



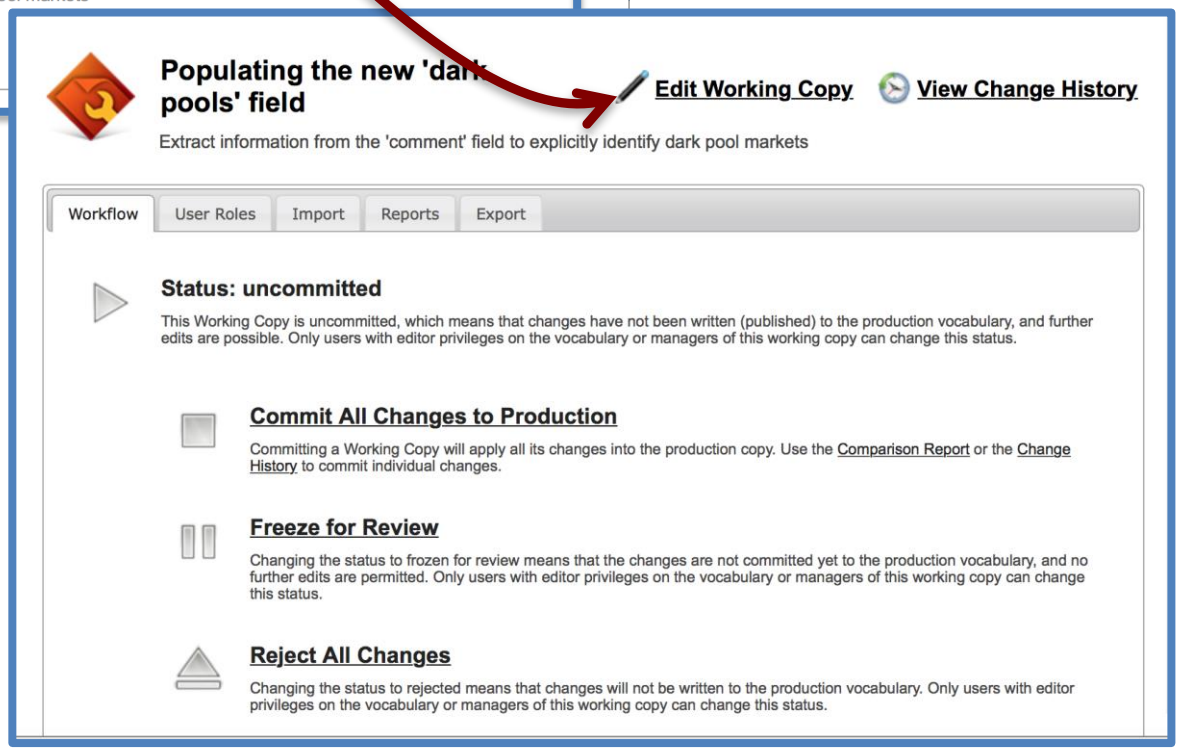
Each item has a separate audit trail from multiple users

© Copyright 2017 TopQuadrant Inc.

Workflow and Change Management

Full auditability and flexible change management:

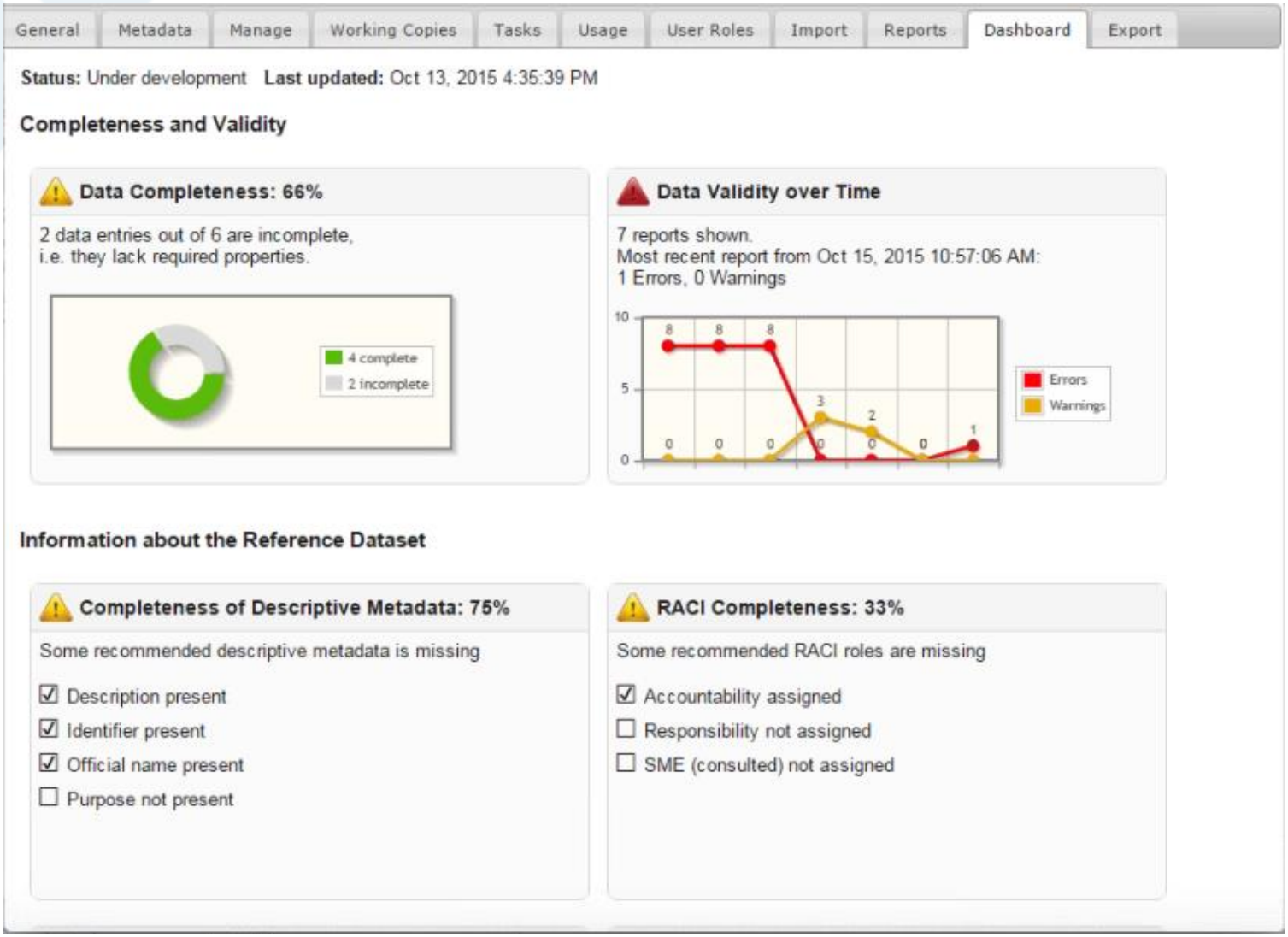
- every change is logged
- changes can be made directly to the 'production copy' or by creating virtual 'working copies' to safely sandbox changes for review and approval



- Working copies for an individual or collaborative teams
- Changes are committed individually or together 'as a whole'
- Comparison reports facilitate review and help avoid collisions

© Copyright 2017 TopQuadrant Inc.

RDM Dashboard: Data Consistency Metrics



© Copyright 2017 TopQuadrant Inc.

Data Quality Validation

Value is not in [ex:Red, ex:Orange, ex:Green]

Resource: Traffic Light 2 (URI: http://example.org/vocabulary1/Traffic_Light_2)
Property: color (URI: http://datashapes.org/tosh/tests/suggestions/in-001.test#color)
Value: Blue (URI: http://datashapes.org/tosh/tests/suggestions/in-001.test#Blue)
Suggestions:
- Delete the invalid value **Apply**
- Replace invalid value with Red **Apply**
  - ✖ Blue (URI: http://datashapes.org/tosh/tests/suggestions/in-001.test#Blue)
  - ✚ Red (URI: http://datashapes.org/tosh/tests/suggestions/in-001.test#Red)
- Replace invalid value with Orange **Apply**
- Replace invalid value with Green **Apply**

Property has more than 1 values (2)

Resource: Traffic Light 2 (URI: http://example.org/vocabulary1/Traffic_Light_2)
Property: color (URI: http://datashapes.org/tosh/tests/suggestions/in-001.test#color)
Suggestions:
- Delete 1 of the 2 values: Blue **Apply**
- Delete 1 of the 2 values: Orange **Apply**
  - ✖ Orange (URI: http://datashapes.org/tosh/tests/suggestions/in-001.test#Orange)

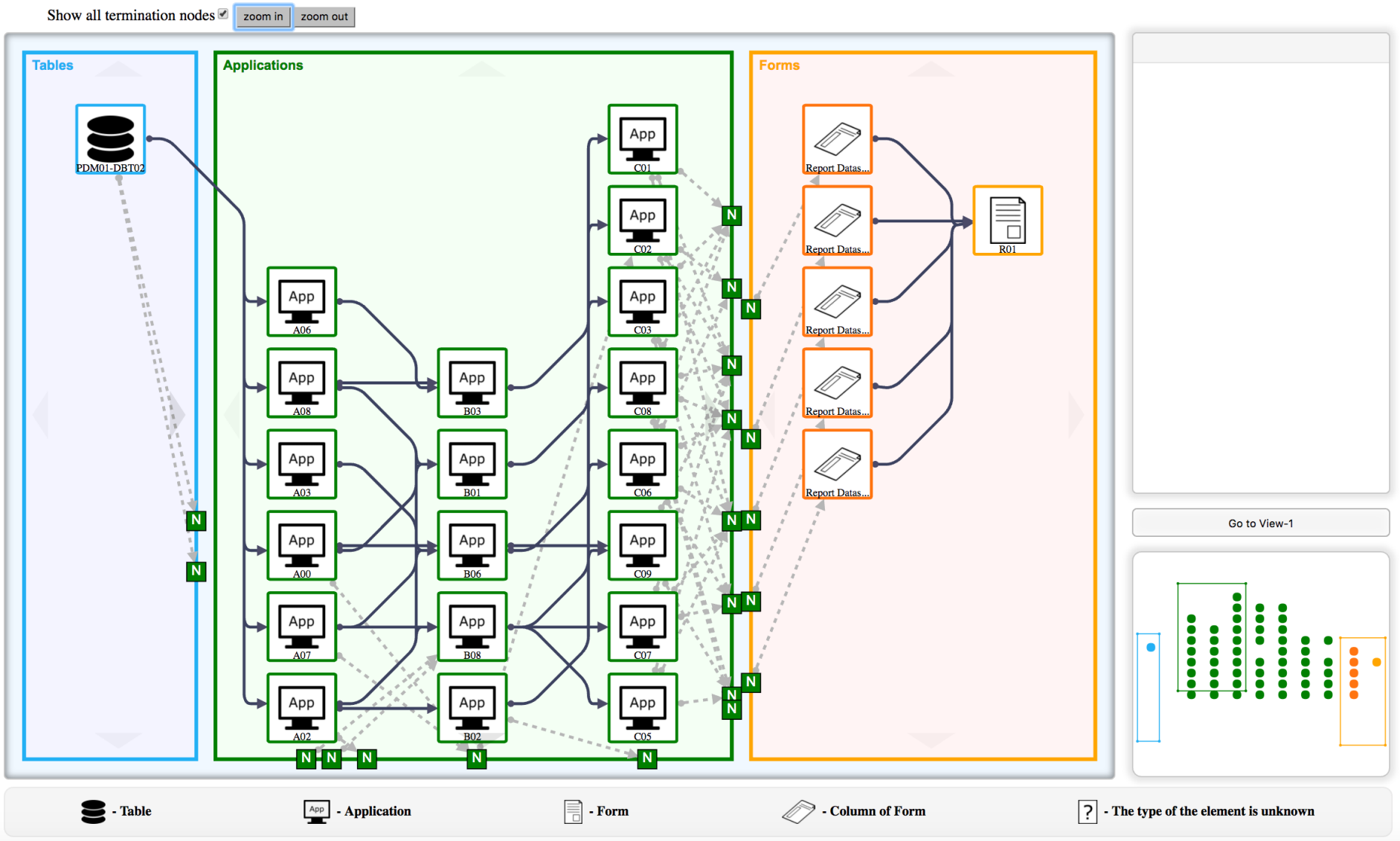
Value has more than 10 characters

Resource: MultiLabelInstance (URI: http://example.org/vocabulary1/MultiLabelInstance)
Property: label (URI: http://www.w3.org/2000/01/rdf-schema#label)
Value: MultiLabelInstance (Datatype: string)
Suggestions:
- Delete the invalid value **Apply**
- Prune string to only 10 characters **Apply**
  - ✖ MultiLabelInstance (Datatype: string)
  - ✚ MultiLabel (Datatype: string)

For information governed by EDG e.g., reference data, glossary terms, etc.

© Copyright 2017 TopQuadrant Inc.

Lineage Diagram



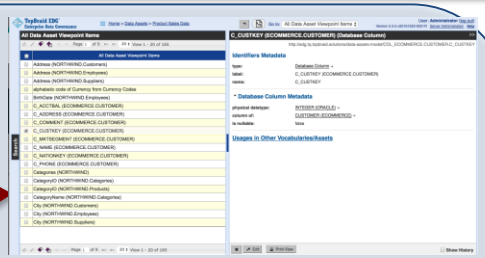
Integration with Other Systems

TopQuadrant™

Import formats
- Spreadsheets (xls, csv, tsv)
- RDF
- XML
- UML
- RDBMs and SQL DDL
- System specific outputs e.g., Collibra, IBM Bus Glossary, MultiTes

TopBraid EDG

govern information assets



Export formats
- HTML
- tab-delimited files
- RDF
- Avro, JSON
- Sharepoint, etc.

And also export to TopBraid EDG Explorer module

Real Time Access

TBL Server

EDG Repository

SPARQL

Rules

Data transforms

Import & Export

Data Integration services

Web Services

direct RDF/SPARQL access

JMS

Web Services

3rd party systems

Data Warehouse

Portals

In-House Apps

Data Stores

Import/Export formats
- RDF (Turtle, RDF/XML, JSON-LD)
- relational databases
- XML, XMLS
- RSS
- Excel spreadsheets
- delimited text (.csv, tsv)
- Web pages (RDFa, Microformats)

© Copyright 2017 TopQuadrant Inc.

Successful RDM Solutions?

Reference Data Management + Data Governance

- ✓ Central Reference Data Management Unit and Reference Data Operating Model

  "Governance is needed to assign accountabilities and enforce standard processes that drive up quality"

- ✓ Integrated process, methodology, collaboration, and organizational change management …

Models, methodology, tools and dedicated resources … ***with*** the right technology puts the right capabilities in the hands of those who govern and use reference data

© Copyright 2017 TopQuadrant Inc.

Participant Poll

© Copyright 2017 TopQuadrant Inc.

TopBraid EDG/RDM: Summary and Benefits

✓ *Reduce independent maintenance of code tables*

Provides a central system where reference data will “live” and be maintained. With this, maintenance activities like adding new codes are done once, not multiple times, reducing authoring time and effort in making changes to reference data.

✓ *Reduce data quality problems due to coding errors*

Allows end users who use reference data (codes) in the process of data entry to better understand their meaning and, as a result, avoid mistakes of using wrong codes.

✓ *Reduce the cost of designing code tables for databases*

Provides a central system database administrators can go to get exports of “codes tables” with all the fields they may need.

✓ *Reduce data integration issues due to inconsistent codes*

When code tables used by different systems are maintained independently, they become inconsistent across the systems. This lack of consistency impacts the ability to integrate data. TopBraid RDM provides an up-to-date source that can be used as input for every code table in every database in the production landscape. Consistency is achieved and integration promoted.

✓ *Make informed decisions based on reference data*

Enables the right categorization and organization of data in reporting and analytics by providing a place for report writers to go to better understand the meaning of reference data and how it is used across the different systems.

© Copyright 2017 TopQuadrant Inc.

… Questions?

TopQuadrant™

Thank You !

To Learn More …

Other EDG/RDM demos/webinar recordings:

http://www.topquadrant.com/knowledge-assets/videos/

Using TopBraid Reference Data Manager 5.0

Webinar: The Foundations of Successful Reference Data Management – with Malcolm Chisholm

Webinar: Why is Reference Data Governance Mission Critical? – with Aaron Zornes

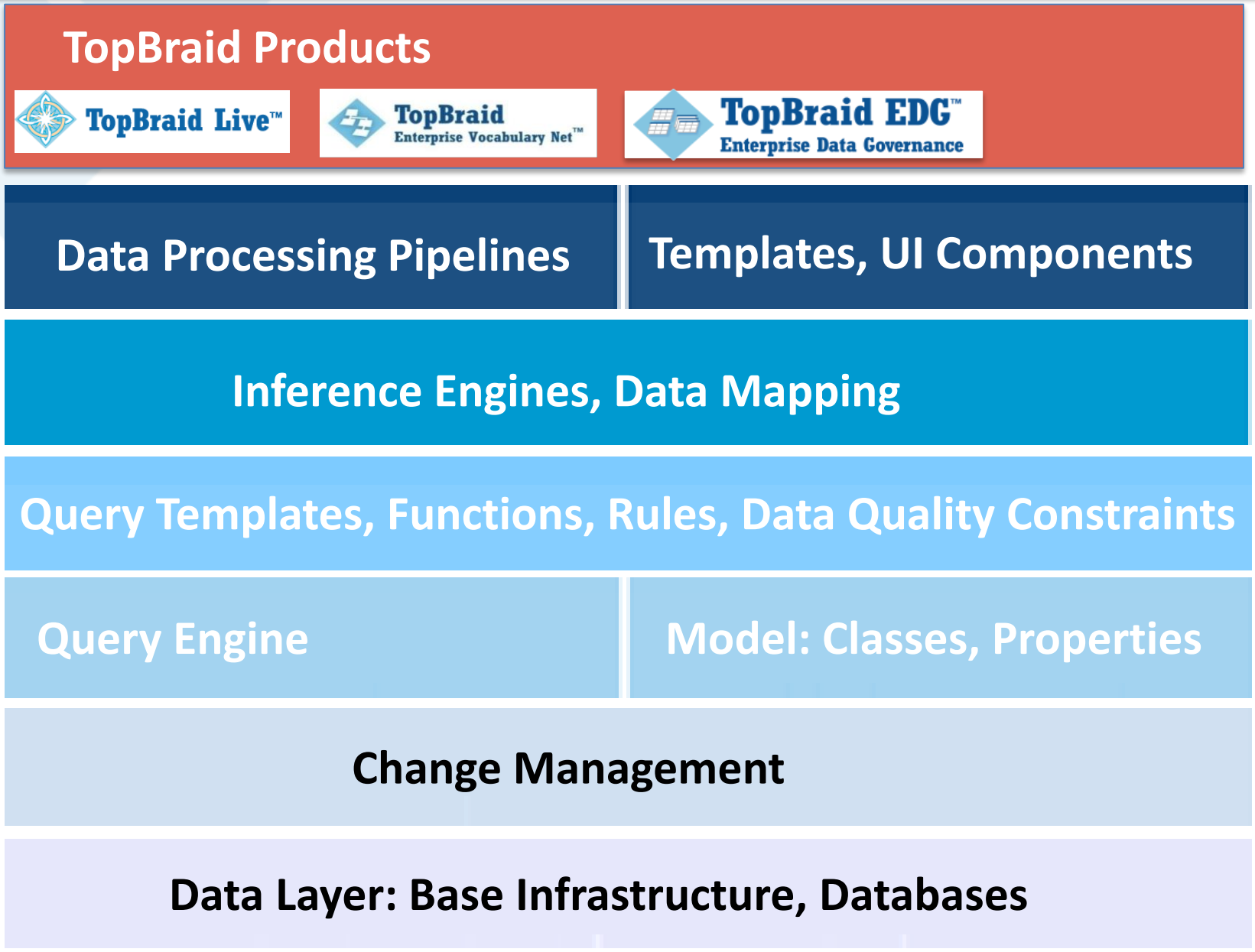
Webinar: Data Governance for the Connected Enterprise: TopBraid EDG in Action

EDG/RDM Product Info:

- http://www.topquadrant.com/products/topbraid-edg/
- http://www.topquadrant.com/products/topbraid-edg-gov-packs/

Contact us: at info@topquadrant.com to:

- Discuss data governance solutions
- Request a more targeted demo of TopBraid EDG/RDM
- Ask for a free EDG evaluation account with RDM

© Copyright 2017 TopQuadrant Inc.

TopBraid Products and Solutions build on a powerful platform