

Going Live with Enterprise Solutions



Reference Data Management in TopBraid EDG

Irene Polikoff and Jack Spivak, TopQuadrant Inc.

February 23, 2017



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

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

ENTERPRISE CUSTOMERS



U.S. AIR FORCE



servicenow



THOMSON REUTERS™



VIDAL GROUP

syngenta®

AstraZeneca



amdocs

EURIWARE
Capgemini
CONSULTING. TECHNOLOGY. OUTSOURCING

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



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.



- 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")

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

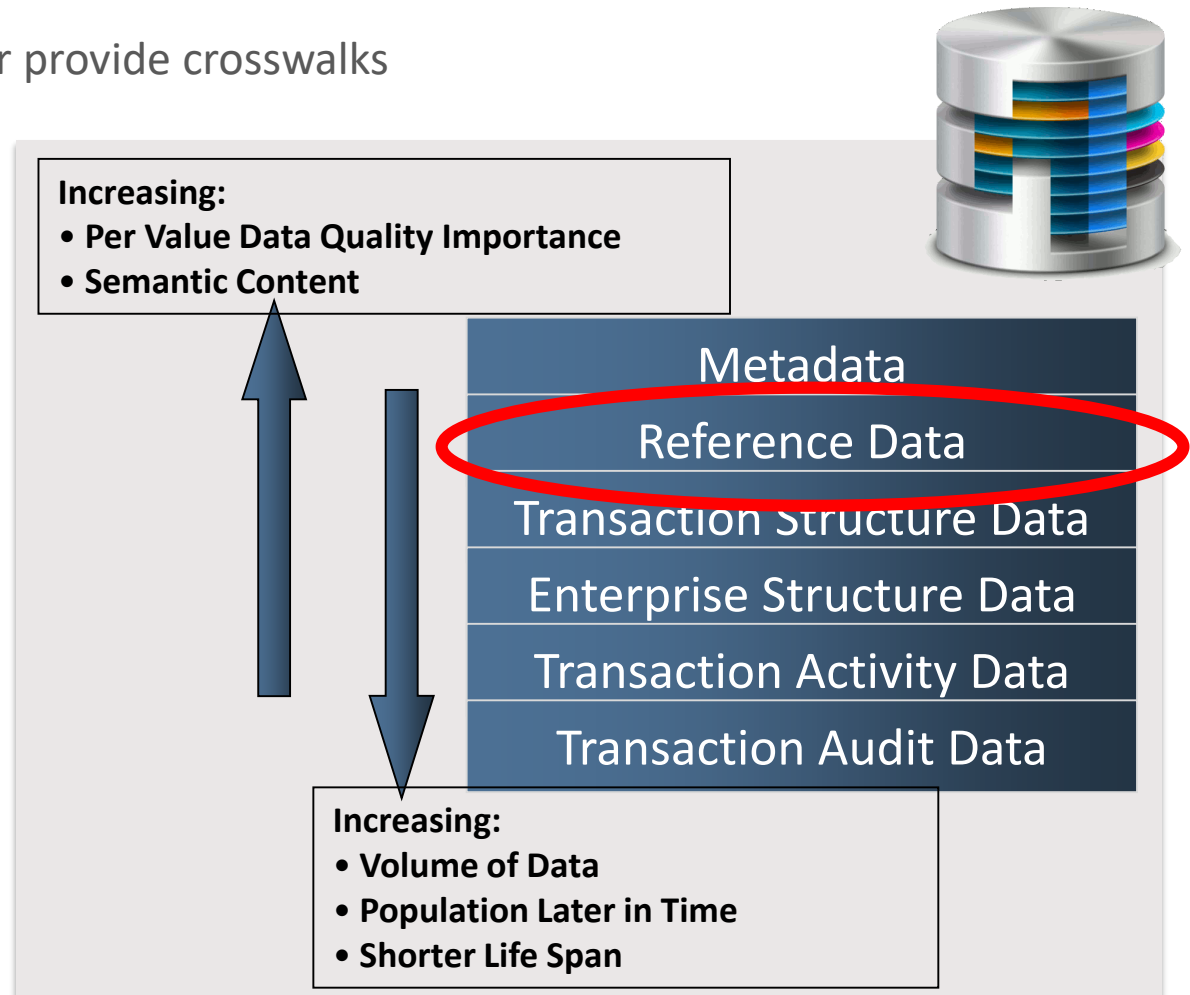
- standard definitions exist
- create policies to use these or provide crosswalks
- data structures are not complex
- strong value proposition

External Reference Data

- defined by standards bodies and industry consortiums

Internal Reference data

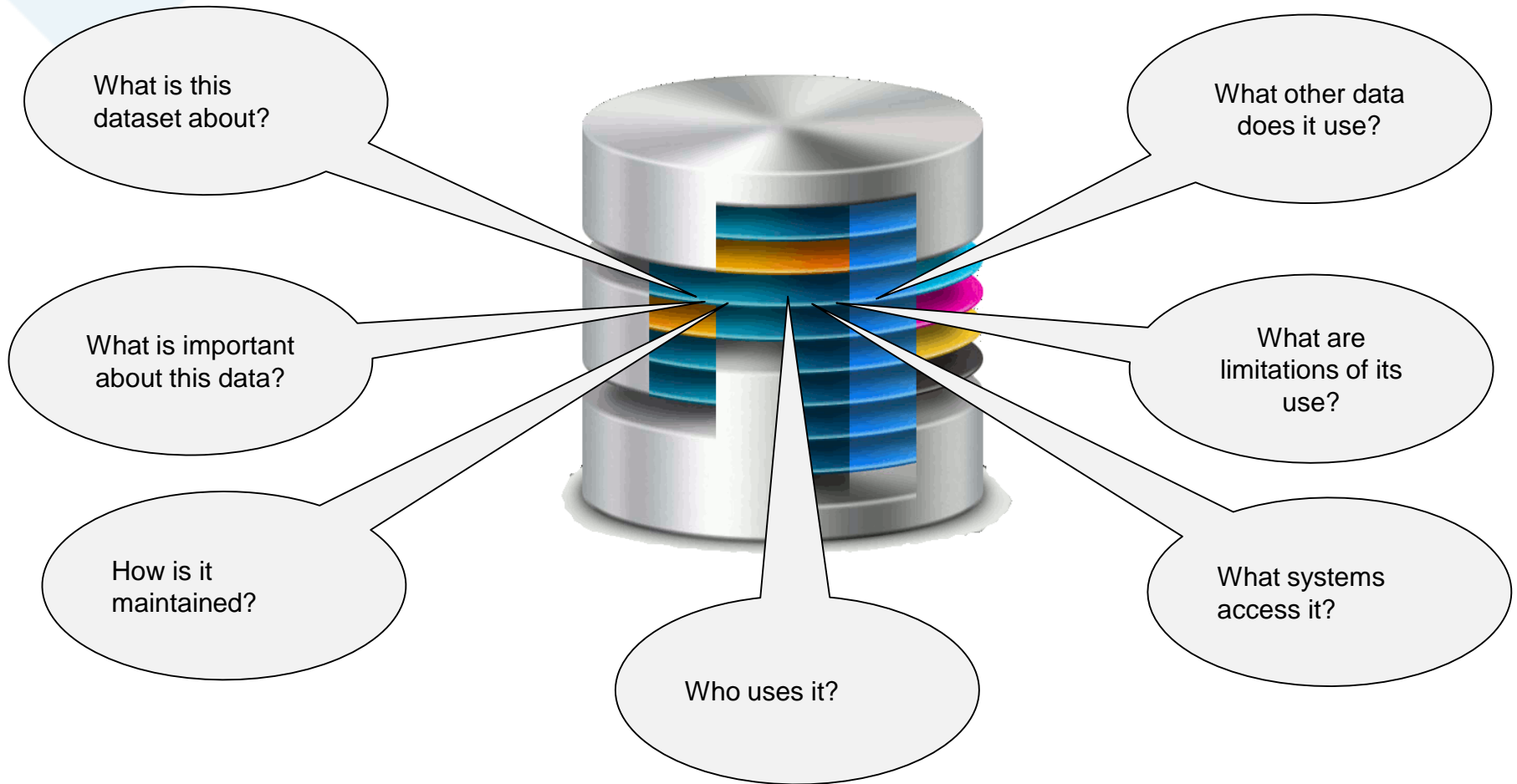
- organization specific codes, department codes, geo regions



What is Comprehensive Metadata?

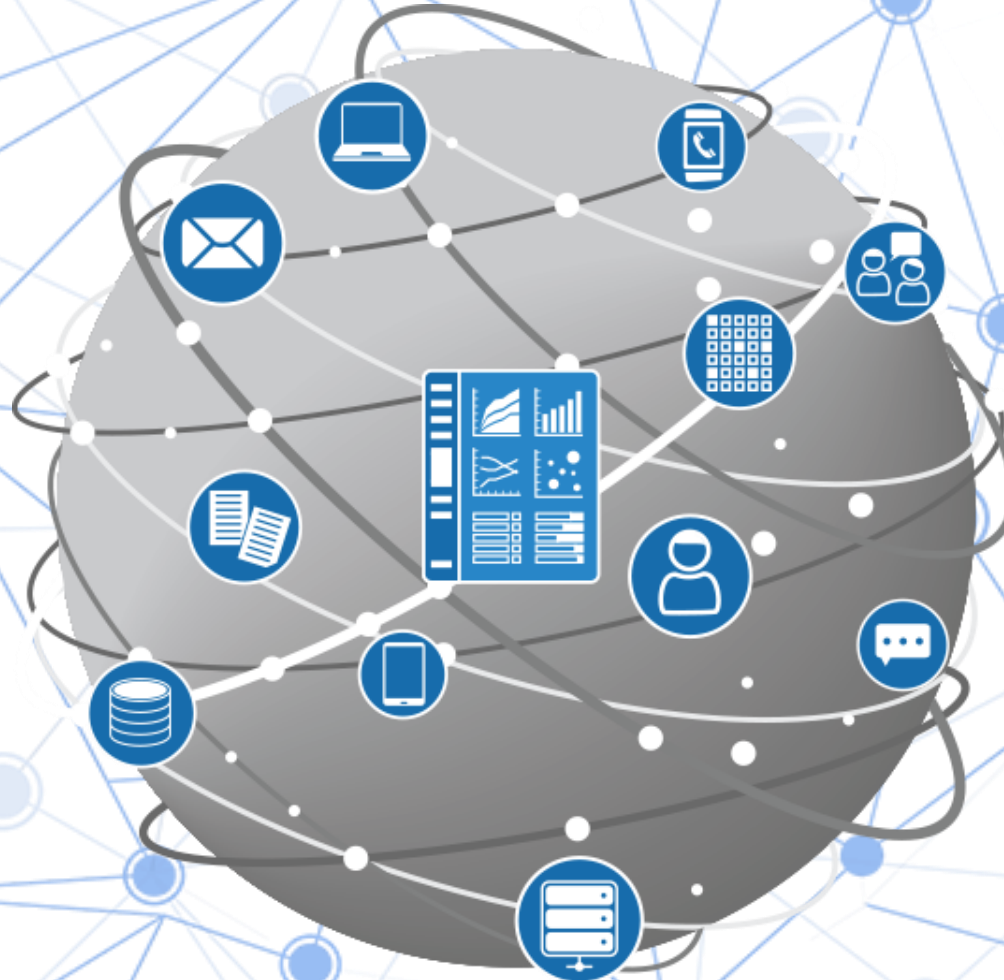
Data *ABOUT* Reference Data

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

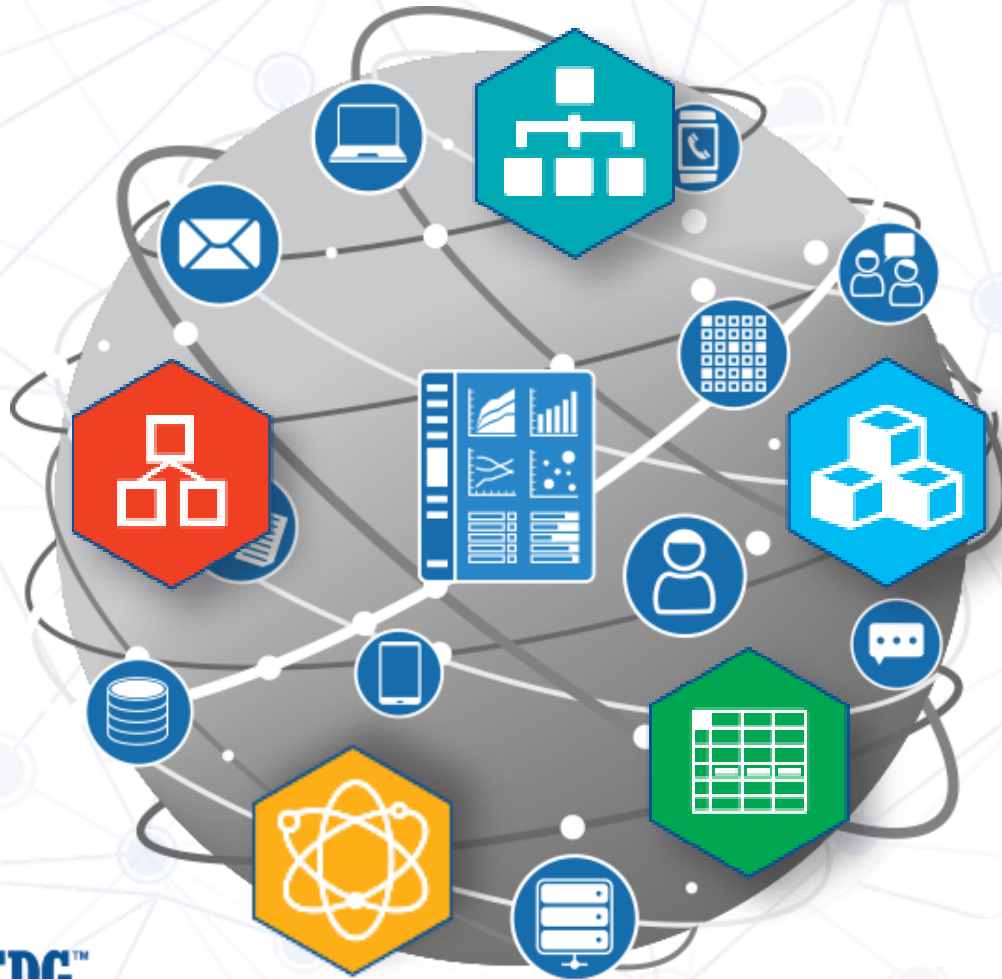


Confusion, wasted resources and errors result when users can't understand the meaning of codes and how they are used











Business Glossaries

Support for

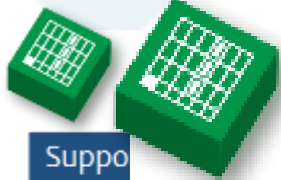
Business Glossaries



Basic Metadata Management

Support for

Data Assets



Reference Data Management

Support for

Reference Datasets

Ontologies

Crosswalks

Support for

Reference Datasets

Ontologies

Crosswalks

Management



Support for

SKOS

Ontologies

Crosswalks

For details, see:

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

- 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



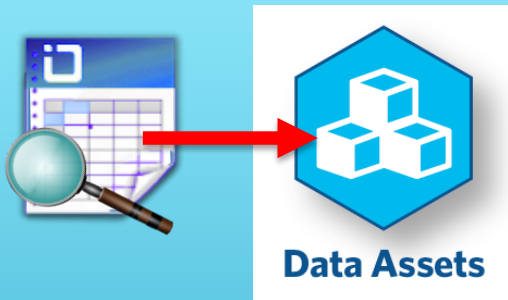
- ✓ *Identify* existing enterprise databases and systems
- ✓ *Catalog* the data landscape in TopBraid EDG





1

Import metadata for enterprise DBs into EDG DataAssets

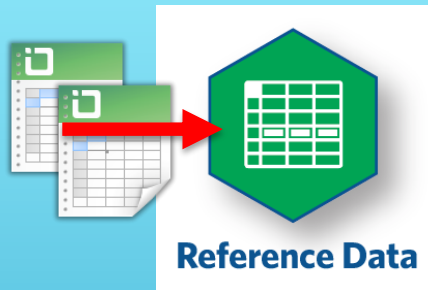


Why?

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

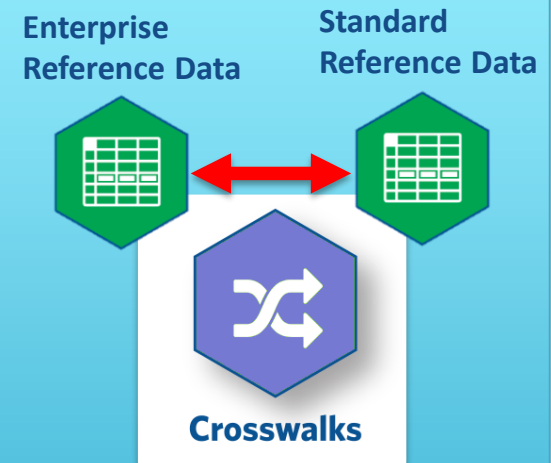


Why?

Manage, control, govern, centralize distribution

3

Create EDG Crosswalks. Connect to standard reference data.



Why?

Analyze, Compare, Understand

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



Data Assets



Reference Data



Crosswalks

UNITED STATES DEPARTMENT OF LABOR

 BUREAU OF LABOR STATISTICS


Standard Occupational Classification

A top

- 53-5011 [Able Seamen](#)
- 23-2093 [Abstractors, Title](#)
- 43-3011 [Account Collectors](#)
- 41-3011 [Account Executives, Advertising](#)
- 13-2011 [Accountants](#)
- 13-2010 [Accountants and Auditors](#)
- 13-2011 [Accountants and Auditors](#)
- 13-2011 [Accountants, Certified Public](#)
- 43-3031 [Accounting Clerks](#)
- 25-1011 [Accounting Professors](#)
- 43-3031 [Accounts Receivable Clerks](#)

DEMO

- 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


User: **Administrator** ([log out](#))
Version 5.3.0.v20161223-0031R [Server Administration](#) [Help](#)

[Home](#) > [Taxonomies](#) > [Geography Vocabulary](#)

Concept Hierarchy

- Geography
 - Africa
 - Antarctica
 - Asia
 - Atlantic Ocean
 - Europe
 - Albania
 - Andorra
 - Austria
 - Belarus
 - Belgium
 - Bosnia-Herzegovina
 - Bulgaria
 - Croatia
 - Czech Republic
 - Denmark
 - Estonia
 - Finland
 - France
 - Germany
 - Gibraltar
 - Greece
 - Greenland
 - Hungary
 - Iceland
 - Ireland
 - Italy

Germany (Country)

<http://topquadrant.com/ns/examples/geography#Germany>

Labels and Description

preferred label:

- Alemania (es)
- Allemagne (fr)
- Deutschland (de)
- Duitsland (nl)
- Germany (en)
- Niemcy (pl)
- Германия (ru)

type: [Country](#) →

Label Objects (SKOS-XL)

alternative label (object):

- [Allied-occupied Germany](#) →
- [East Germany](#) →
- [Federal Republic Of Germany](#) →
- [German Democratic Republic](#) →
- [German Empire](#) →
- [German Reich](#) →
- [Nazi Germany](#) →
- [Third Reich](#) →
- [Weimar Republic](#) →
- [West Germany](#) →

Standard Relationships

has broader: [Europe](#) →

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

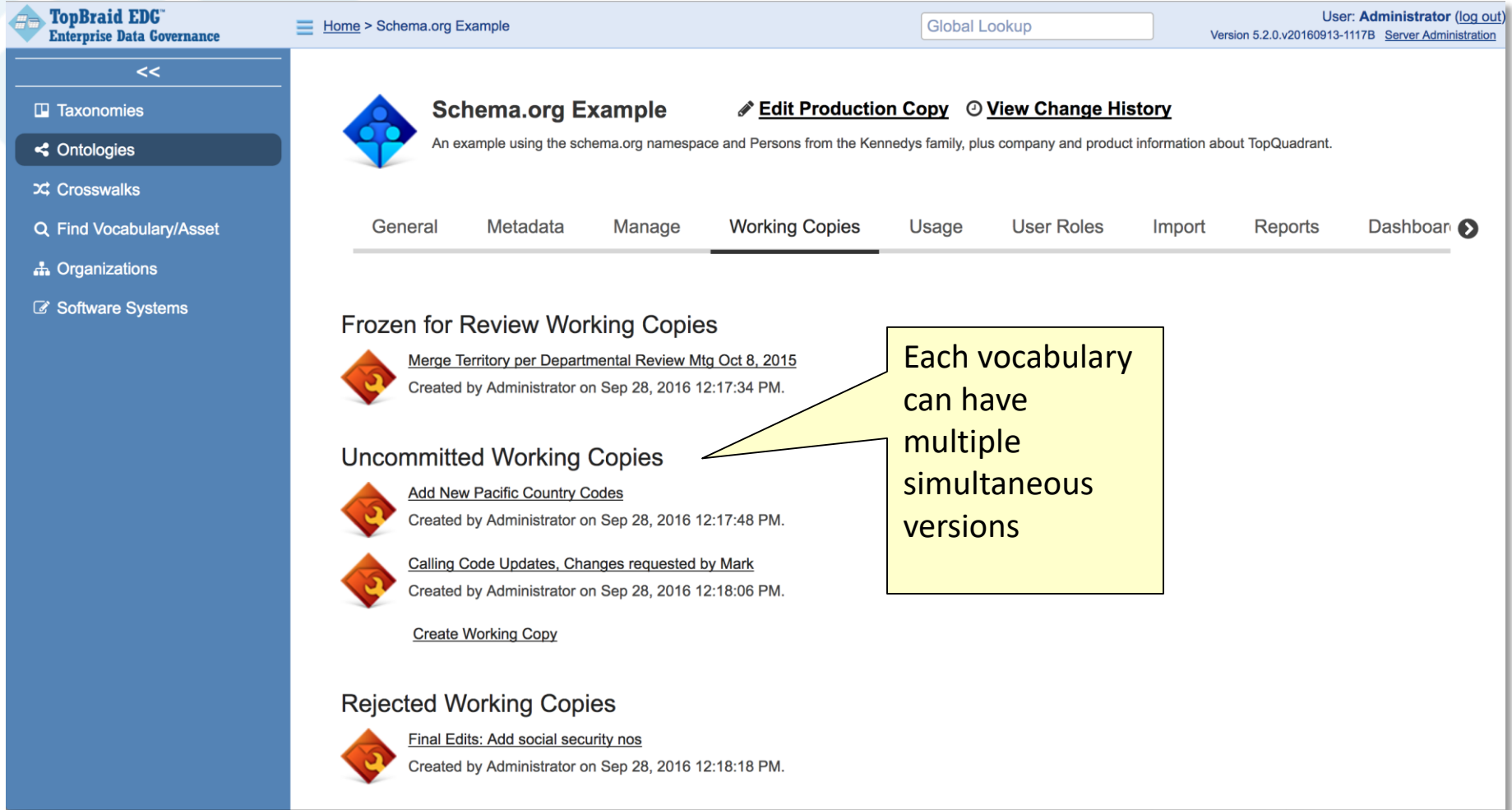
- Manager

- 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



The screenshot shows the TopBraid EDC interface. The top navigation bar includes the TopBraid EDC logo, a breadcrumb trail (Home > Schema.org Example), a search box (Global Lookup), and user information (User: Administrator (log out), Version 5.2.0.v20160913-1117B, Server Administration).

The left sidebar contains a navigation menu with the following items: Taxonomies, Ontologies (highlighted), Crosswalks, Find Vocabulary/Asset, Organizations, and Software Systems.

The main content area displays the 'Schema.org Example' page. It features a blue diamond icon and the title 'Schema.org Example'. Below the title is a description: 'An example using the schema.org namespace and Persons from the Kennedys family, plus company and product information about TopQuadrant.' There are two action links: 'Edit Production Copy' and 'View Change History'.

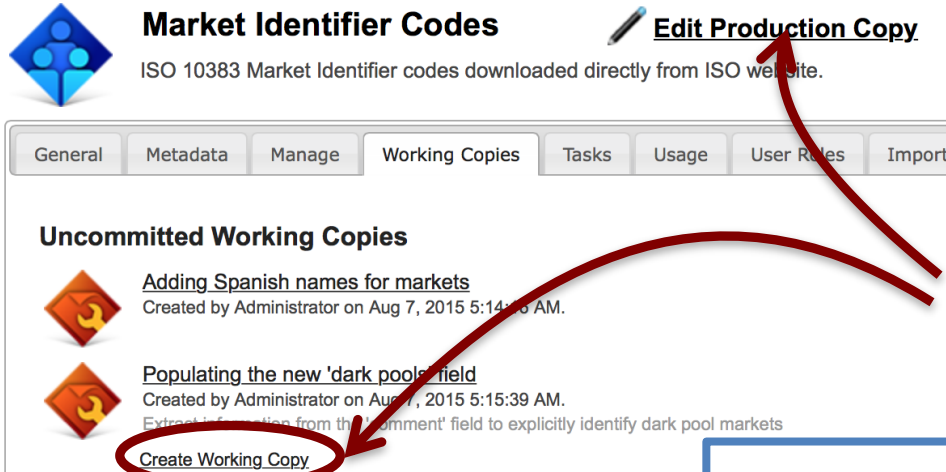
A horizontal tab bar is visible, with 'Working Copies' selected. Other tabs include General, Metadata, Manage, Usage, User Roles, Import, Reports, and Dashboard.

The 'Working Copies' section is divided into three categories:

- Frozen for Review Working Copies**
 - [Merge Territory per Departmental Review Mtg Oct 8, 2015](#)
Created by Administrator on Sep 28, 2016 12:17:34 PM.
- Uncommitted Working Copies**
 - [Add New Pacific Country Codes](#)
Created by Administrator on Sep 28, 2016 12:17:48 PM.
 - [Calling Code Updates, Changes requested by Mark](#)
Created by Administrator on Sep 28, 2016 12:18:06 PM.
 - [Create Working Copy](#)
- Rejected Working Copies**
 - [Final Edits: Add social security nos](#)
Created by Administrator on Sep 28, 2016 12:18:18 PM.

A yellow callout box with a black border points to the 'Uncommitted Working Copies' section. It contains the text: 'Each vocabulary can have multiple simultaneous versions'.

Each item has a separate audit trail from multiple users





Market Identifier Codes [Edit Production Copy](#)

ISO 10383 Market Identifier codes downloaded directly from ISO website.

General Metadata Manage Working Copies Tasks Usage User Roles Import

Uncommitted Working Copies

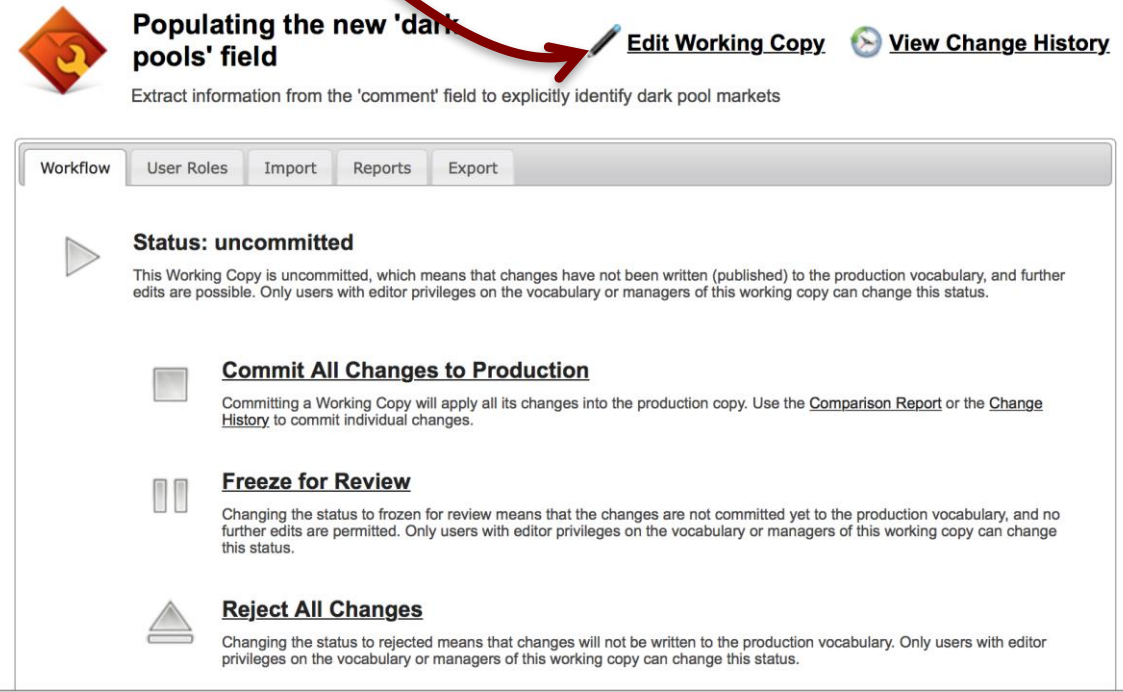
-  [Adding Spanish names for markets](#)
Created by Administrator on Aug 7, 2015 5:14:36 AM.
-  [Populating the new 'dark pools' field](#)
Created by Administrator on Aug 7, 2015 5:15:39 AM.
Extract information from the 'comment' field to explicitly identify dark pool markets


[Create Working Copy](#)

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



 **Populating the new 'dark pools' field** [Edit Working Copy](#) [View Change History](#)

Extract information from the 'comment' field to explicitly identify dark pool markets

Workflow User Roles Import Reports Export

Status: uncommitted

This Working Copy is uncommitted, which means that changes have not been written (published) to the production vocabulary, and further edits are possible. Only users with editor privileges on the vocabulary or managers of this working copy can change this status.

- Commit All Changes to Production**
Committing a Working Copy will apply all its changes into the production copy. Use the [Comparison Report](#) or the [Change History](#) to commit individual changes.
- Freeze for Review**
Changing the status to frozen for review means that the changes are not committed yet to the production vocabulary, and no further edits are permitted. Only users with editor privileges on the vocabulary or managers of this working copy can change this status.
- Reject All Changes**
Changing the status to rejected means that changes will not be written to the production vocabulary. Only users with editor privileges on the vocabulary or managers of this working copy can change this status.


General Metadata Manage Working Copies Tasks Usage User Roles Import Reports Dashboard Export

Status: Under development **Last updated:** Oct 13, 2015 4:35:39 PM

Completeness and Validity

⚠ Data Completeness: 66%


2 data entries out of 6 are incomplete, i.e. they lack required properties.



4 complete
2 incomplete

⚠ Data Validity over Time

7 reports shown.
Most recent report from Oct 15, 2015 10:57:06 AM:
1 Errors, 0 Warnings



Report #	Errors	Warnings
1	8	0
2	8	0
3	8	0
4	0	3
5	0	2
6	0	0
7	1	0

Information about the Reference Dataset

⚠ Completeness of Descriptive Metadata: 75%

Some recommended descriptive metadata is missing

- Description present
- Identifier present
- Official name present
- Purpose not present

⚠ RACI Completeness: 33%

Some recommended RACI roles are missing

- Accountability assigned
- Responsibility not assigned
- SME (consulted) not assigned

✘ 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>)

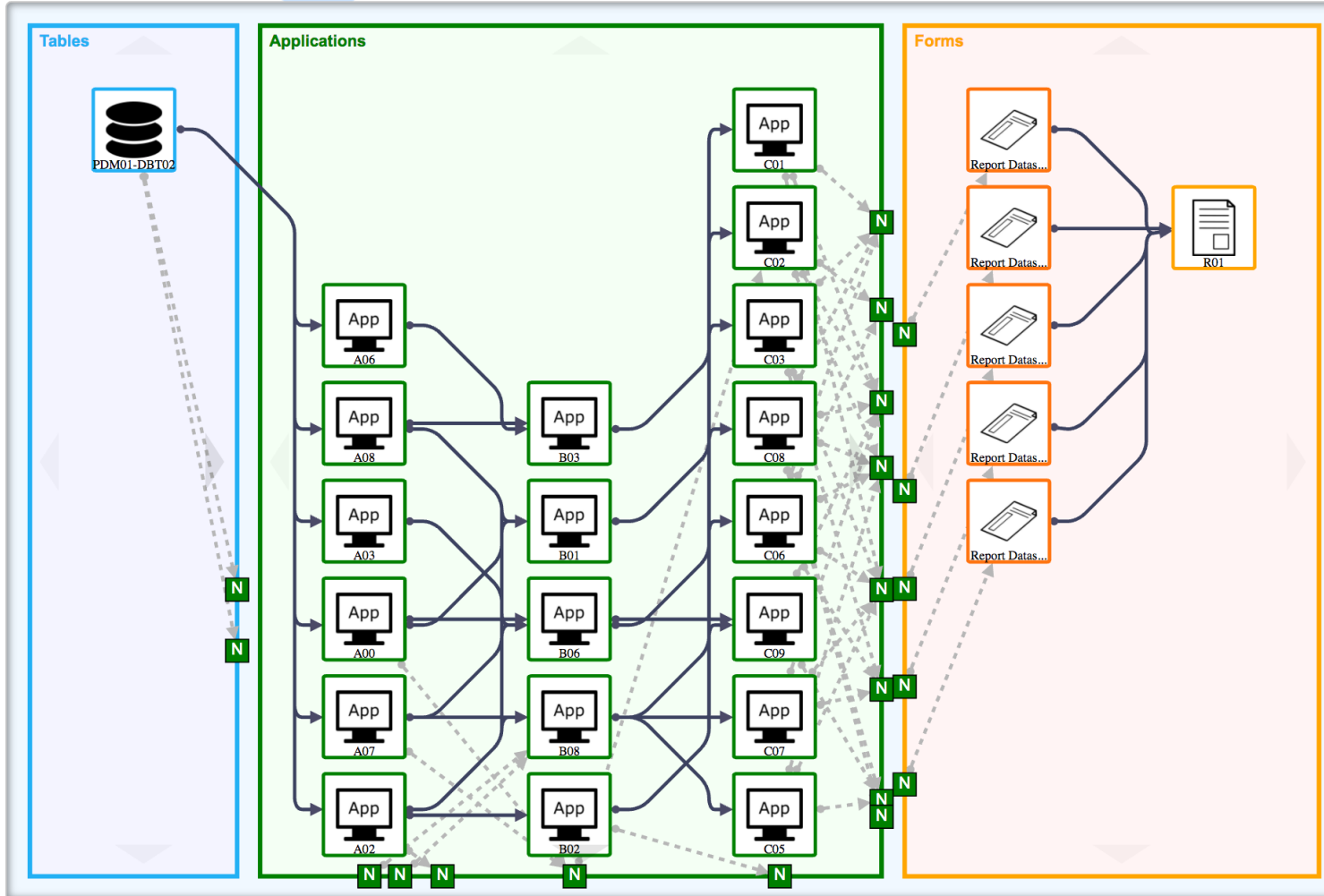
⚠ Value has more than 10 characters

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:

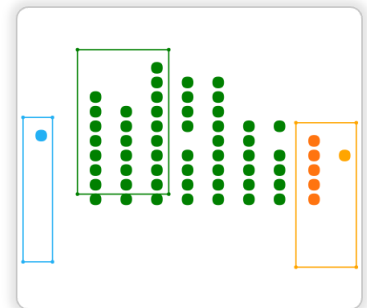
- ▶ [Delete the invalid value](#) **Apply**
- ▼ [Prune string to only 10 characters](#) **Apply**
 - ✘ MultiLabelInstance (Datatype: string)
 - ✚ MultiLabel (Datatype: string)

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

Show all termination nodes [zoom in](#) [zoom out](#)



Go to View-1



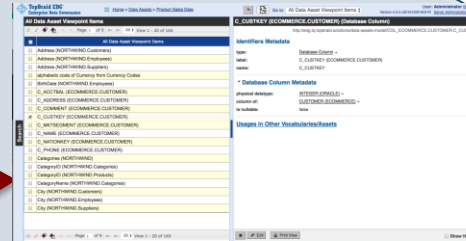
Import formats

- 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

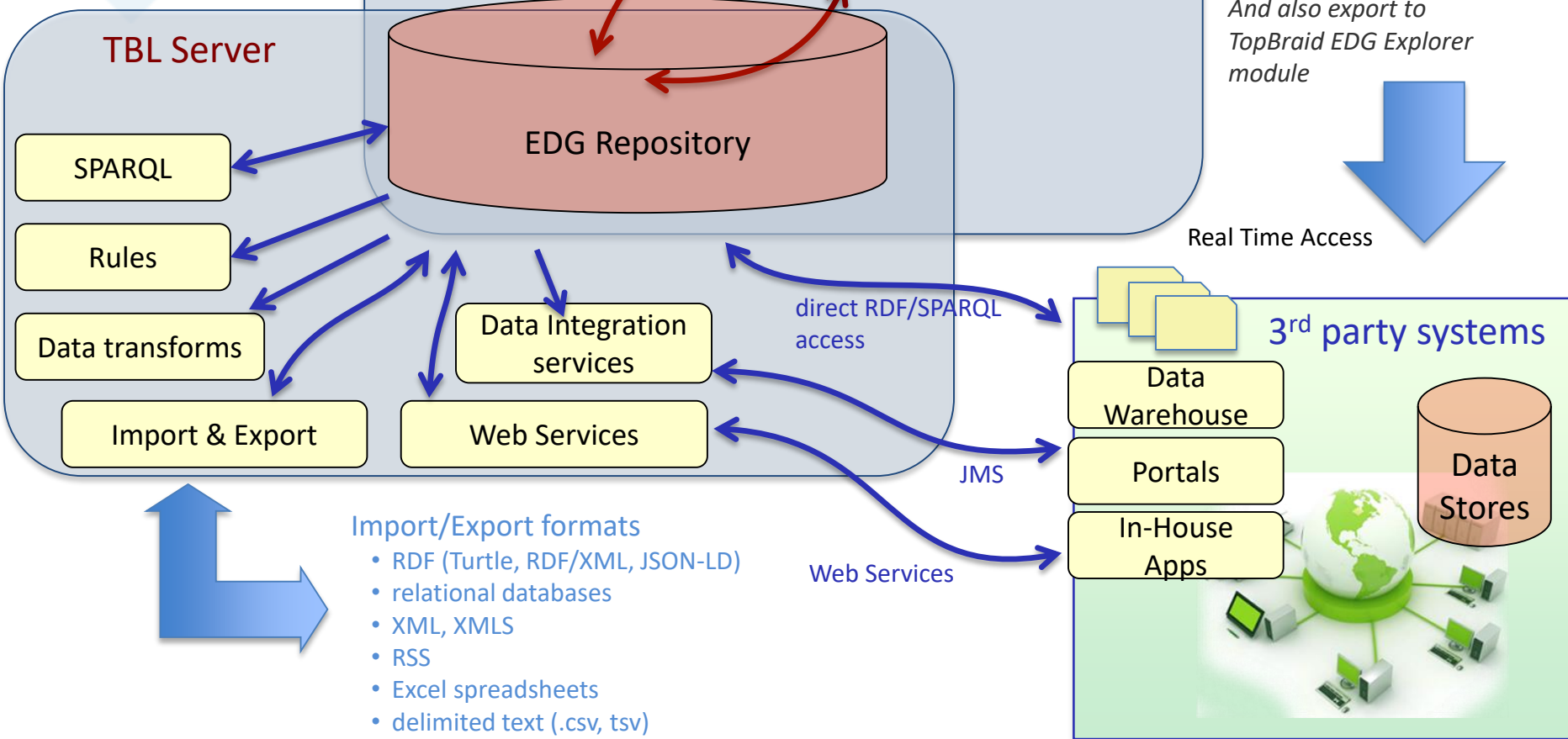
- HTML
- tab-delimited files
- RDF
- Avro, JSON
- Sharepoint, etc.



And also export to TopBraid EDG Explorer module

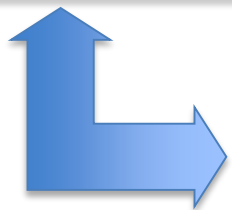


Real Time Access



Import/Export formats

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



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

Participant Poll

✓ *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.



... Questions?

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



TopBraid Products and Solutions build on a powerful platform

TopBraid Products



Data Processing Pipelines

Templates, UI Components

Inference Engines, Data Mapping

Query Templates, Functions, Rules, Data Quality Constraints

Query Engine

Model: Classes, Properties

Change Management

Data Layer: Base Infrastructure, Databases