

# Data Governance Preparedness for Reference Data Management



**June 10, 2015**

Presented by:

Malcolm Chisholm – Askget.com Inc.

Robert Coyne – TopQuadrant Inc.

[mchisholm@askget.com](mailto:mchisholm@askget.com)

[robert@topquadrant.com](mailto:robert@topquadrant.com)

# Agenda

---

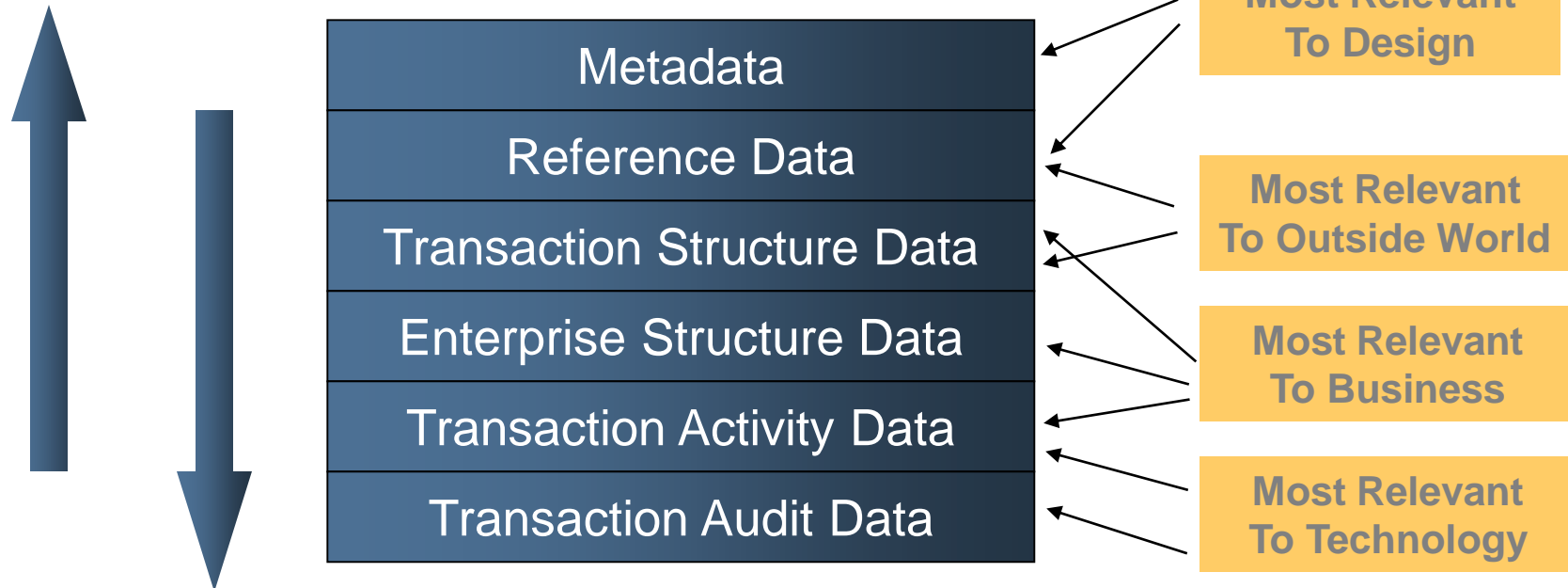
- What is Reference Data?
- Semantics and Reference Data
- External vs Internal Reference Data
- Governing External Reference Data
- Reference Data Operating Model
- Implementing Reference Data Management

# What is Reference Data?

# Reference Data in Context

Increasing:

- Per Value Data Quality Importance
- Semantic Content



Increasing:

- Volume of Data
- Population Later in Time
- Shorter Life Span

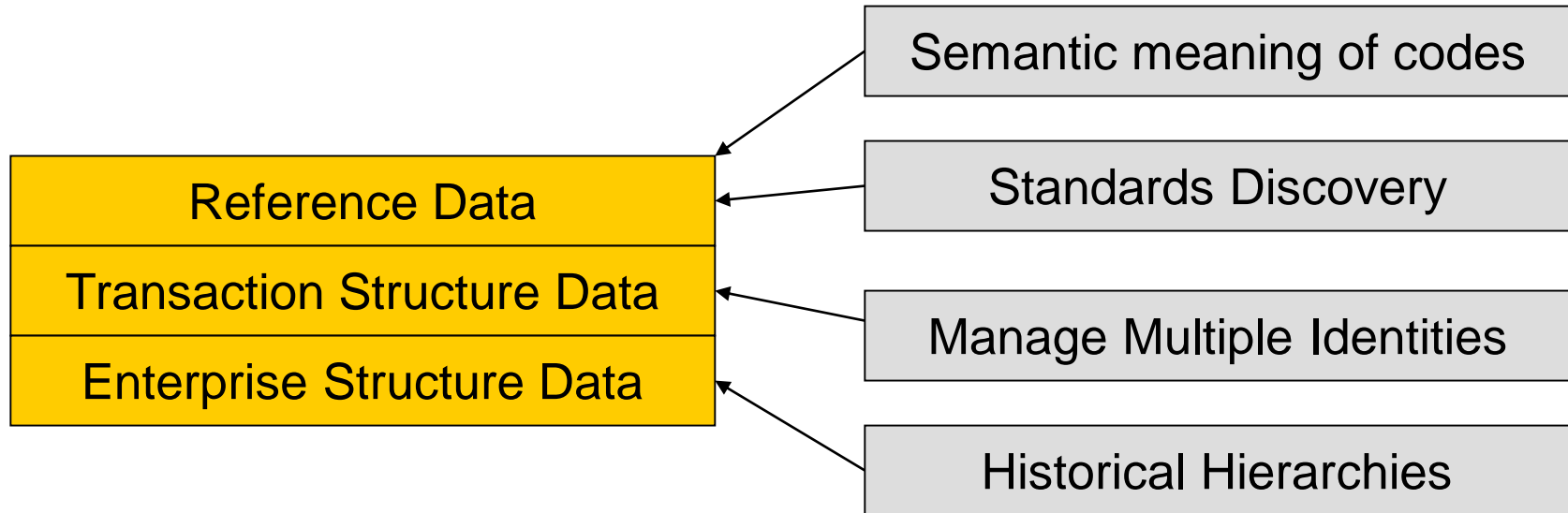
# Layers of Data

Metadata	<p>The data that describes all aspects of an enterprise's information assets, and enables the enterprise to effectively use and manage these assets.</p> <p>Here it is confined to the structure of databases. Found in a database's system catalog. Sometimes included in database tables.</p>
Reference Data	<p>Any kind of data that is used solely to categorize other data found in a database, or solely for relating data in a database to information beyond the boundaries of the enterprise.</p> <p>Codes and descriptions. Tables containing this data usually have just a few rows and columns.</p>
Transaction Structure Data	<p>Data that represents the direct participants in a transaction, and which must be present before a transaction fires.</p> <p>The parties to the transactions of the enterprise. E.g. Customer, Product.</p>
Enterprise Structure Data	<p>Data that permits business activity to be reported and/or analyzed by business responsibility.</p> <p>Typically, data that describes the structure of the enterprise. E.g. organizational or financial structure.</p>
Transaction Activity Data	<p>Data that represents the operations an enterprise carries out</p> <p>Traditional focus of IT – in many enterprises the only focus.</p>
Transaction Audit Data	<p>Data that tracks the life cycle of individual transactions.</p> <p>Includes application logs, database logs, web server logs.</p>

# Layers of Data

Metadata	Metadata
Master Data	Reference Data
	Transaction Structure Data
	Enterprise Structure Data
Event Data	Transaction Activity Data
	Transaction Audit Data

# Specific Data Governance Tasks Exist for Each Category of Data, E.g.



These tasks mean many different governance issues exist for MDM

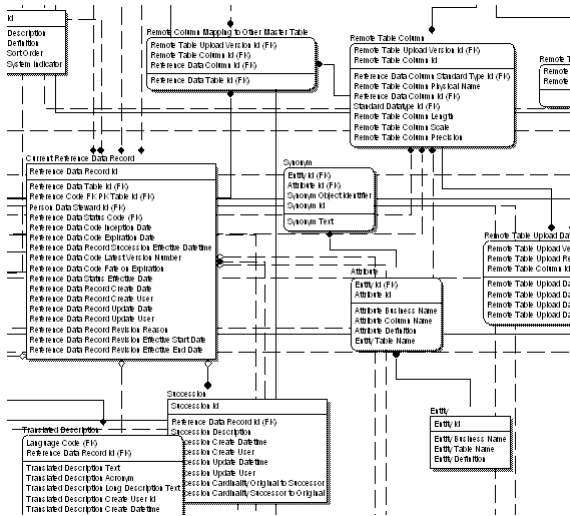
Determine specific management tasks for each category of data

“One size fits all” approaches to data management will not work

# Semantics and Reference Data



# Where Are Semantics?



Customer Type	
C	Corporate Customer
I	Individual Customer
G	Government Customer

Data Model
Entities (Tables)
Attributes (Columns)
Relations (Constraints)

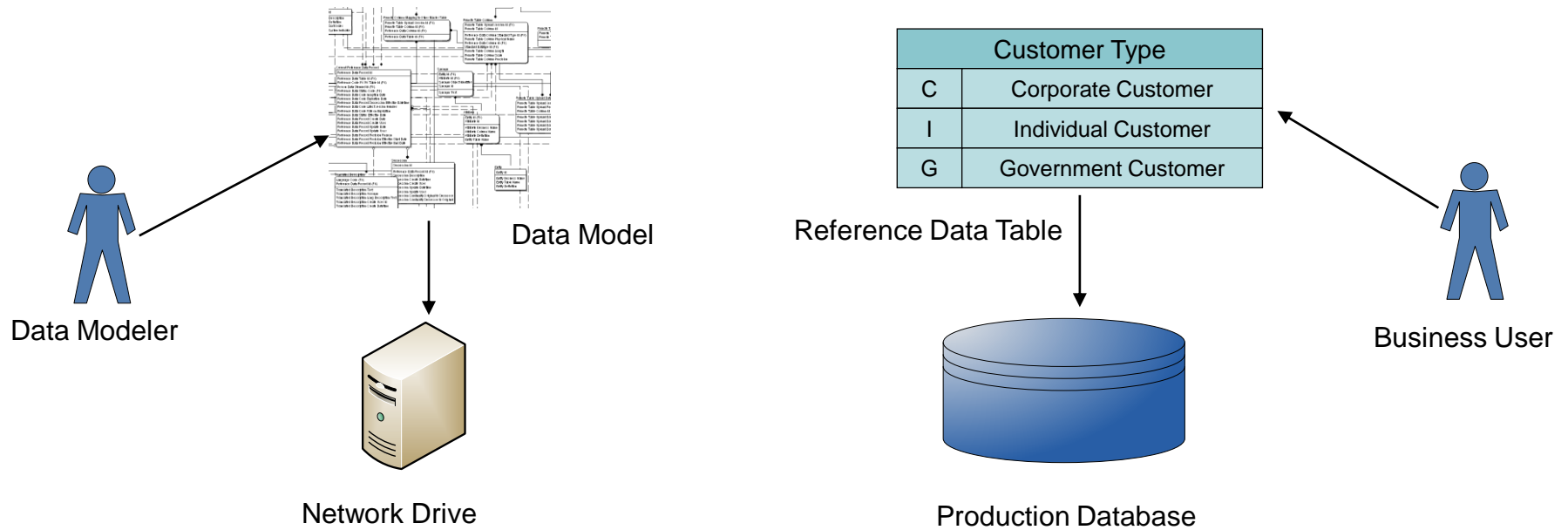
Reference Data Table
Codes
Business Terms
Busiess Definitions?

# Need for Definitions in Reference Data

<b>Country Code</b>	<b>Country Name</b>	<b>Description</b>
HKG	Hong Kong	The former British crown colony. Now part of the People's Republic of China, it is considered a separate territory for our enterprise
PRC	People's Republic of China	Includes Macao and Hong Kong, but excludes Taiwan.
ROC	Republic of China	Taiwan. Not recognized as an independent country by the People's Republic of China, but treated as a separate territory for our enterprise.

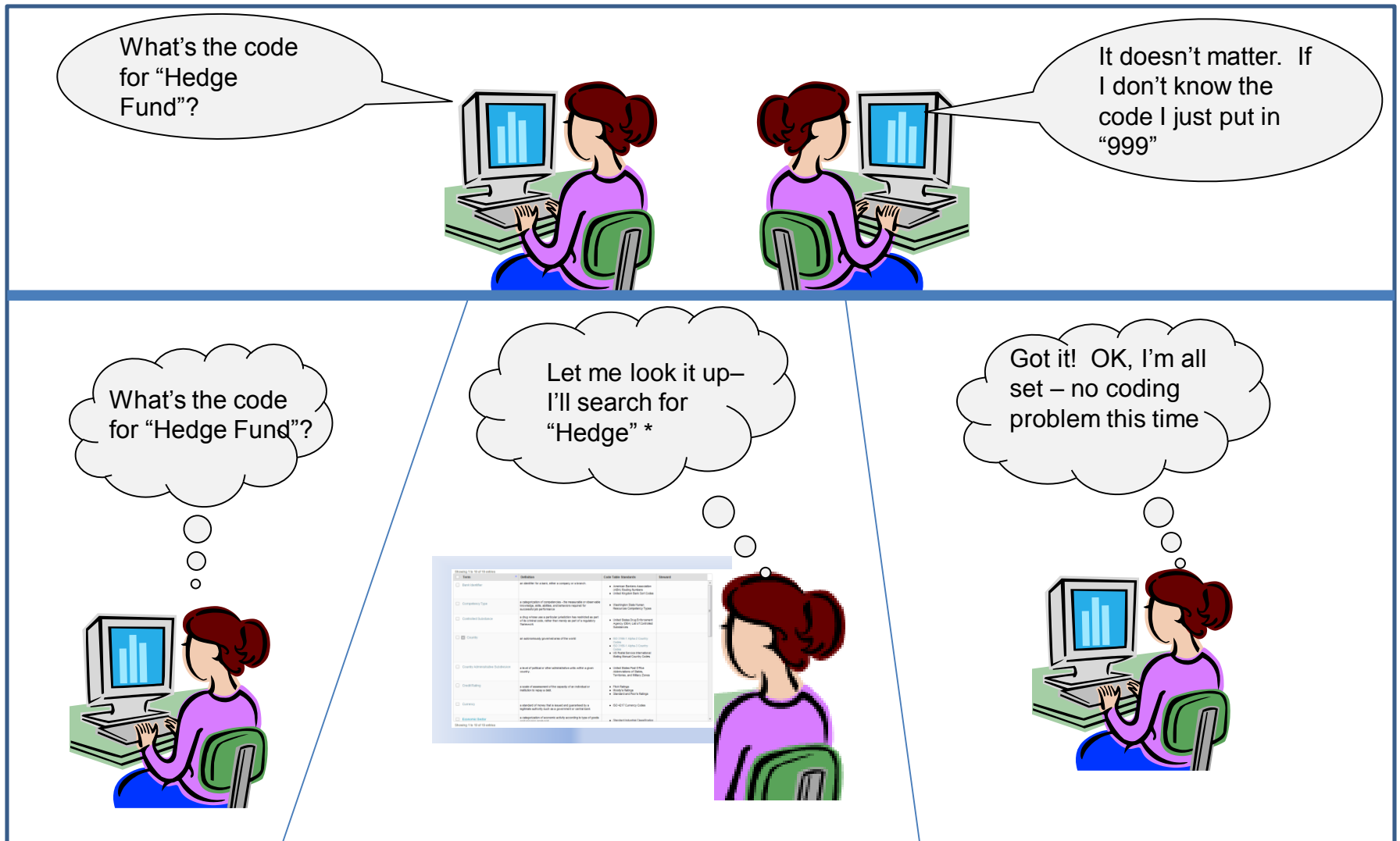
Without definitions we really cannot understand Reference Data

# Where Are Semantics Stored



- Although difficult to estimate it seems that 50% of business semantics (or more) exist in Reference Data vs Data Models
- The Reference Data tables have Business Terms (“Descriptions”) – the Data Models have logical “Data Naming Conventions”.

# Lack of Semantics Lead to Coding Errors

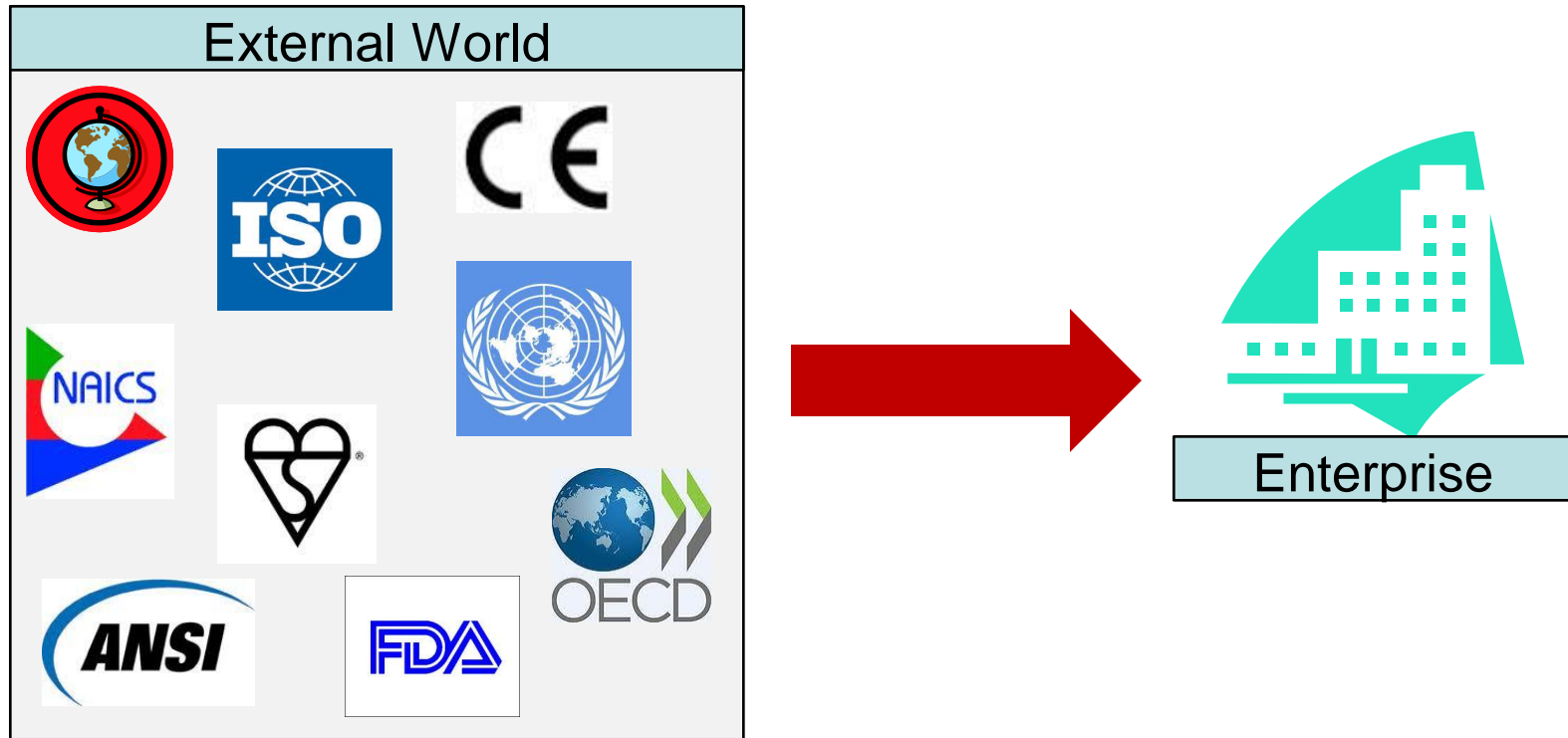


# Lack of Semantics Leads to Bad Business Decisions



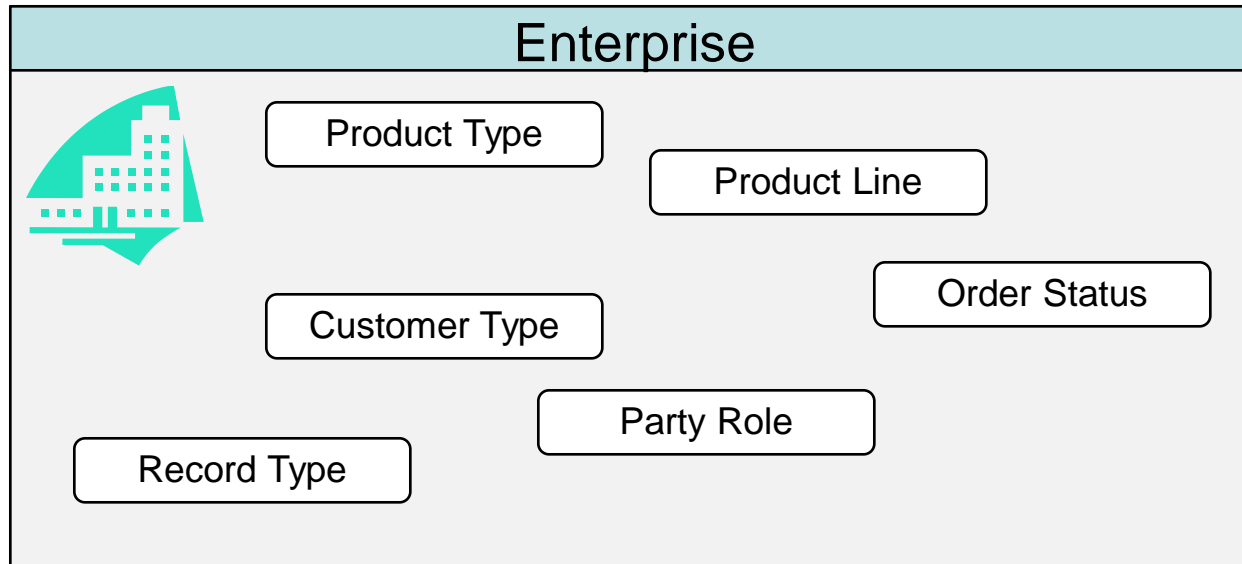
# External vs Internal Reference Data

# Origin: External Reference Data



- The enterprise has to get at least some Reference Data from outside itself.
- These are “standards” – they allow the enterprise to communicate with other organizations outside itself (e.g. regulators).
- The enterprise does not have to do the hard work of creating the Reference Data.
- The enterprise does not have to do the hard work of revising the Reference Data.
- This class of Reference Data has its own concerns, and these affect architecture.

# Origin: Internal Reference Data

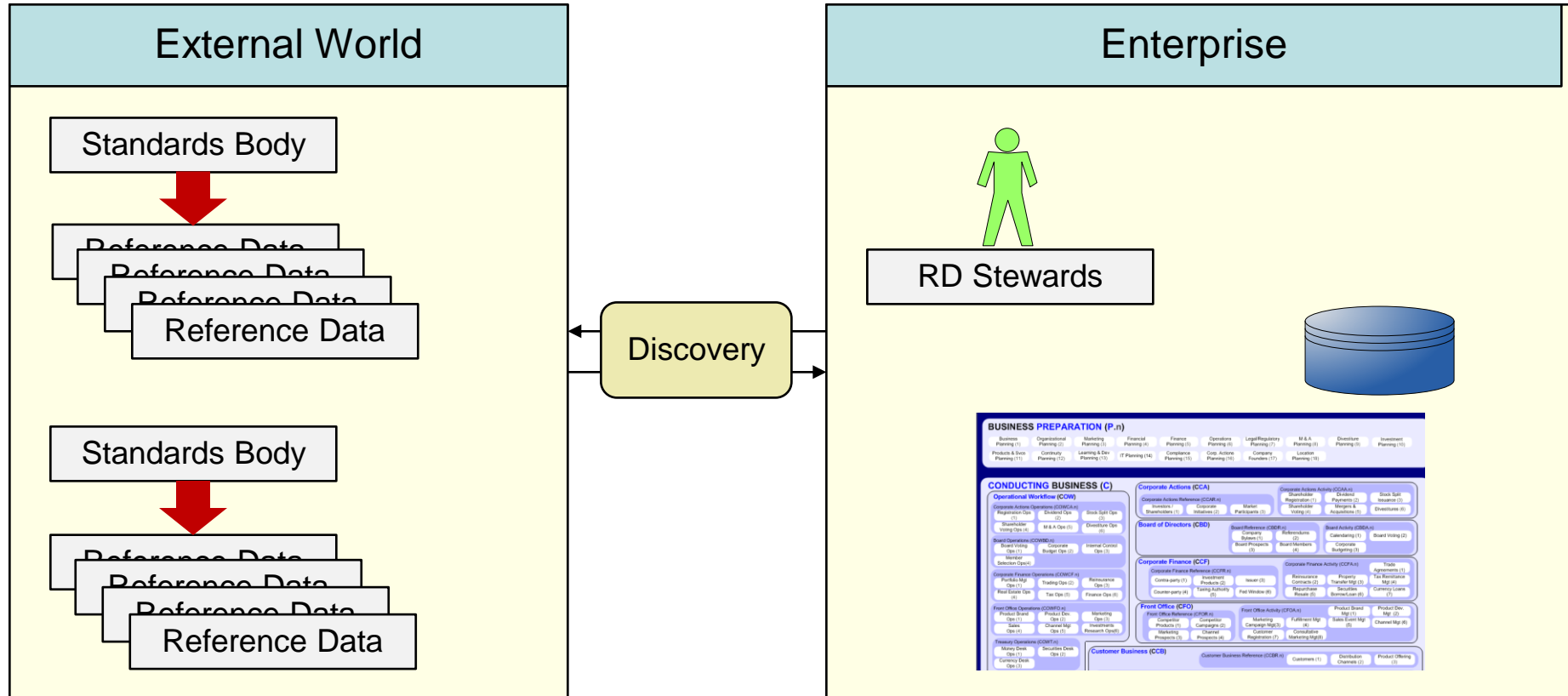


- A great deal of Reference Data is managed within the enterprise itself.
- It is typically governed poorly.
- It is often internally inconsistent.
- Much of it is administered by IT



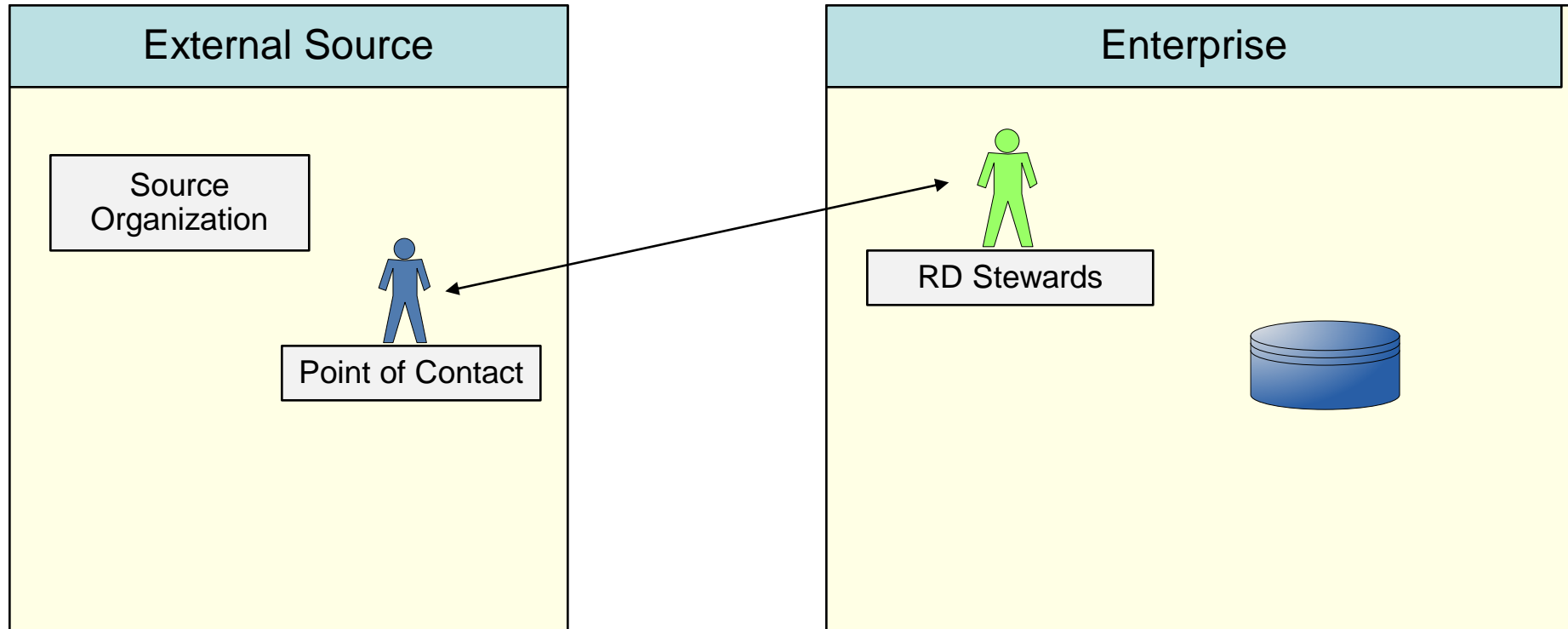
# Governing External Reference Data

# Discovery and Conceptual Mapping



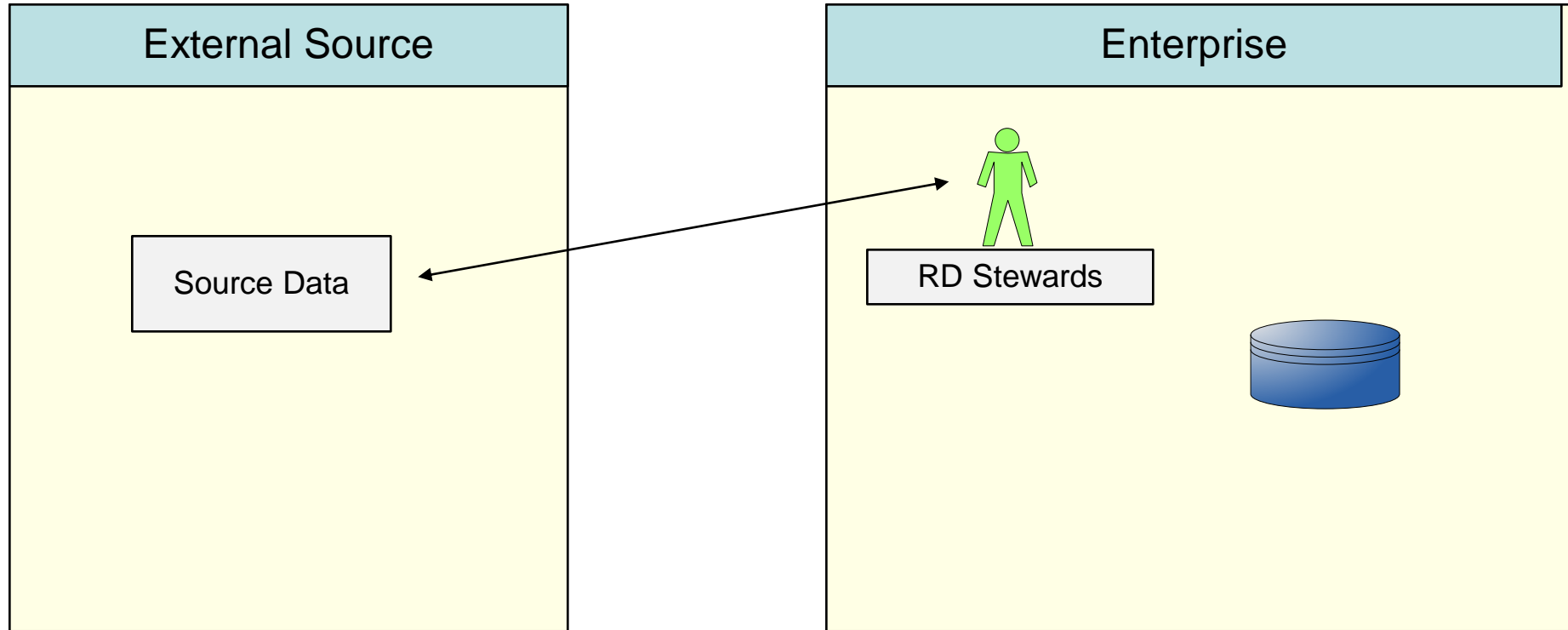
- The enterprise must make sense of (a) what it wants to capture as Reference Data; and (b) what external standards-setting bodies maintain this
- Thus a formal discovery process is required, and a conceptual map (ontology) that must be kept updated, and a database to capture the results of the discovery efforts
- Discovery will be ongoing.
- All of these translate to important architectural components

# External Source Organization Profiling



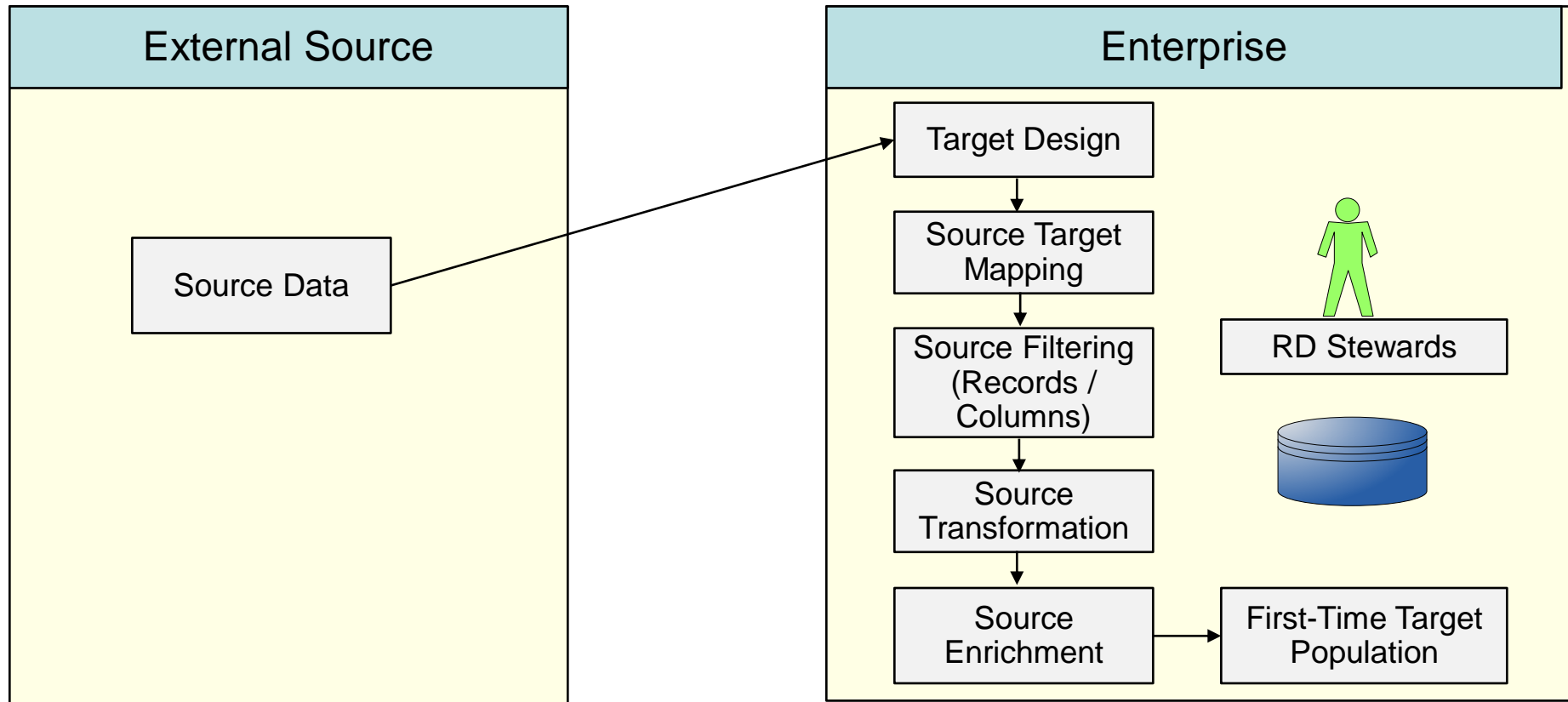
- Every Source Organization needs a profile.
- This will facilitate communications,
- This information needs to be stored in some kind of repository and kept up to date
- Again, important architectural components

# External Source Reference Data Profiling



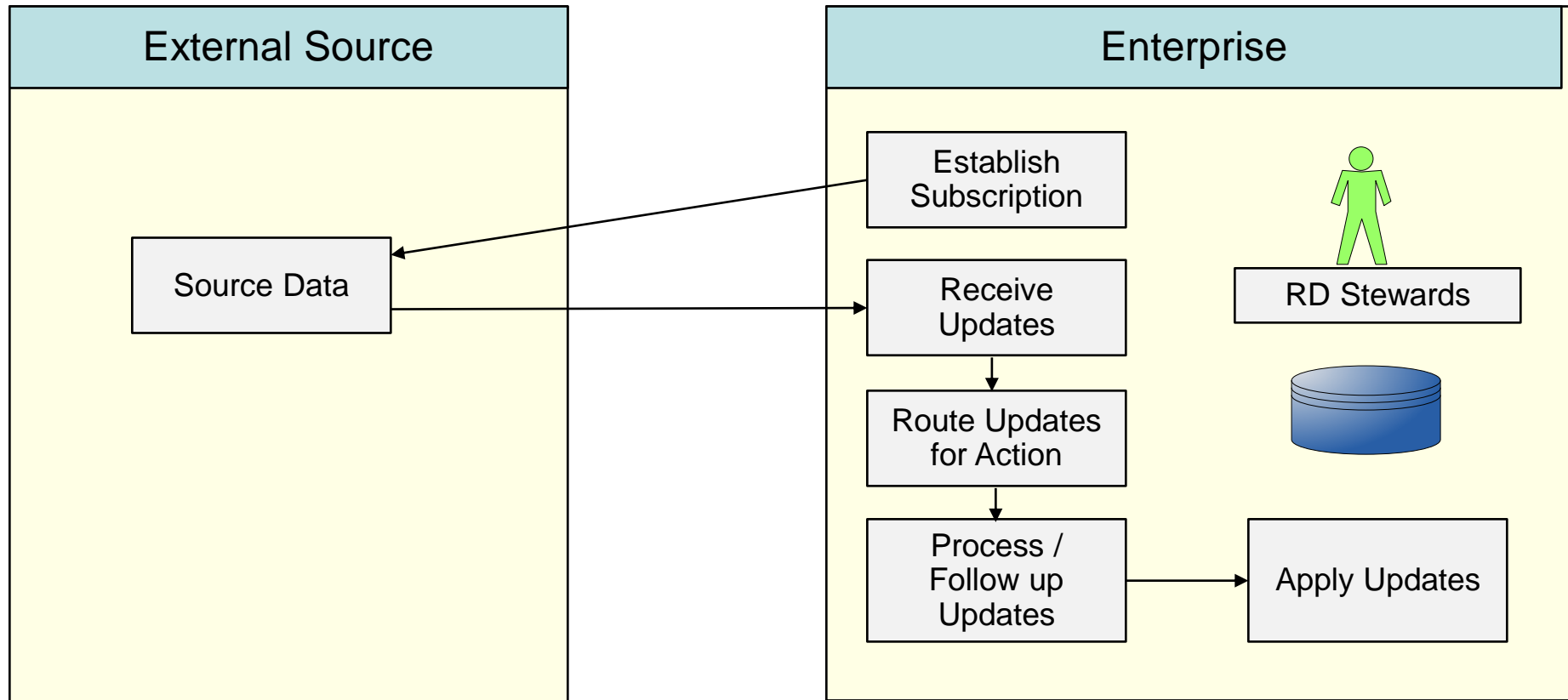
- Every Source dataset needs profiling
- This is to understand the source Reference Data, and where it is to be found
- This includes file layouts, semantic analysis - This can be quite complex
- Again, important architectural components

# External Reference Data Onboarding



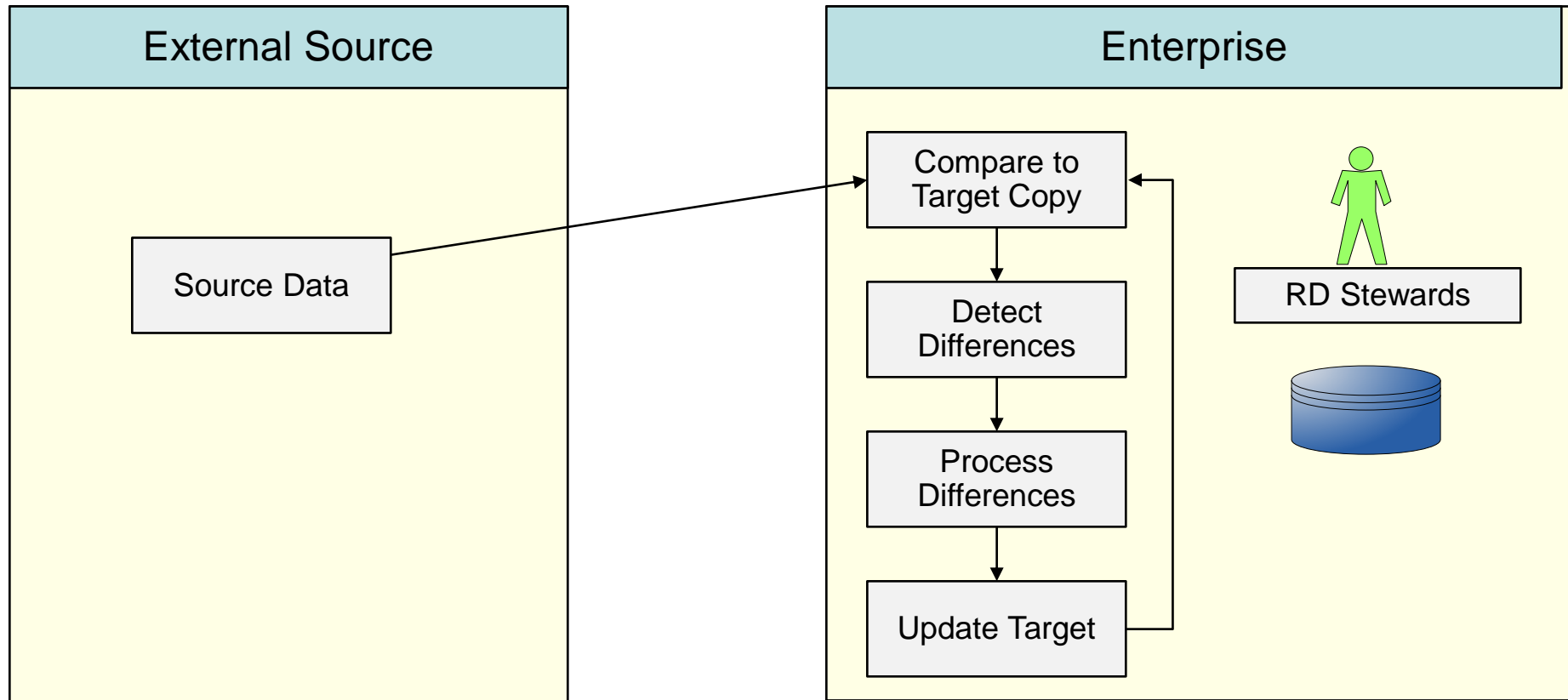
- Onboarding is getting the source Reference Data set into the enterprise for the first time.
- There are many steps, and these reflect the need for distinct data and metadata stores to be built
- Also the processes may require distinct architectural components
- We now begin to see what solution architectures might look like
- Try to reuse existing architectural components, rather than build new ones
- Note that the complexity extends to governance as well as architecture.

# Subscription Management



- Many external sources offer some kind of subscription.
- These need to be established, and the metadata needs to be captured
- Infrastructure needs to be set up to receive updates from the subscription
- Routing and processing of updates requires additional architectural components
- Also, we much check that updates are being processed. Typically, they are received centrally and distributed to stewards. That way the center can assess if they are being actioned. This adds to architectural components.

# Periodic Reconciliation

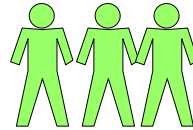


- Even with subscriptions, the enterprise copy may drift out of synch with the source
- Periodic reconciliation is needed to detect material differences and provide assurance that the enterprise's copy is still in synch with the source.
- The Onboarding Guide, produced during Source Onboarding is vital to understand the differences that will inevitably be detected

# Reference Data Operating Model



# Central Reference Data Management Unit



## Why Specialized?

Reference Data has unique properties

E.g. It has meaning, and is added to production environments

Reference Data has unique challenges

E.g. It has to be synchronized across many applications

Reference Data has unique risks

E.g. It is often misunderstood leading to “miscodings” etc.

## Why Centralized?

Need for standardization

E.g. Which Country Code will be use – GENC, ISO Alpha-2, ISO Alpha-3...

Need for one place in enterprise to deal with external authorities

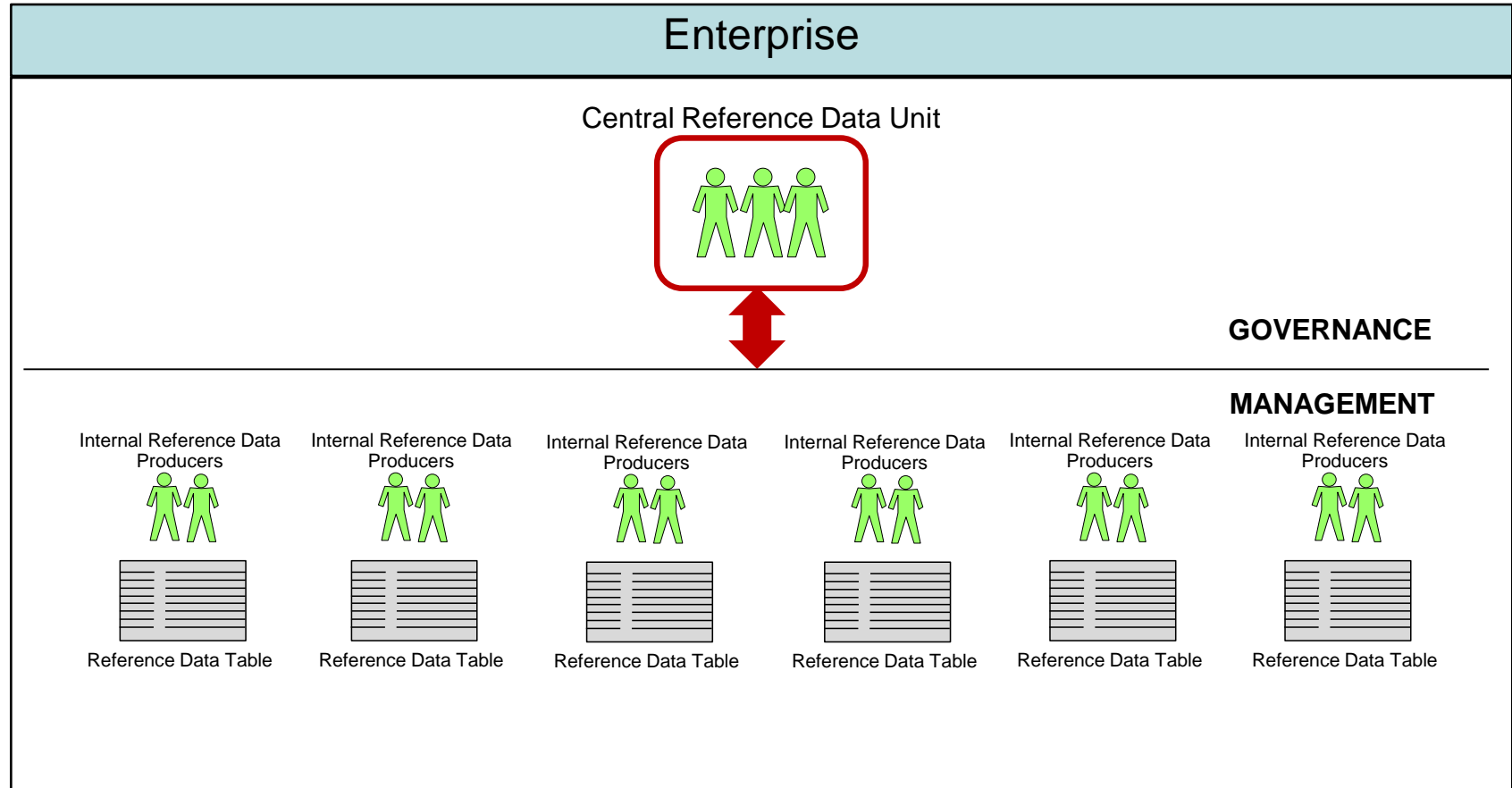
E.g. Who ensures we get the NAICS updates

Need to set up governance for internal reference data mgmt.

E.g. How are Customer Type, Product Line managed?

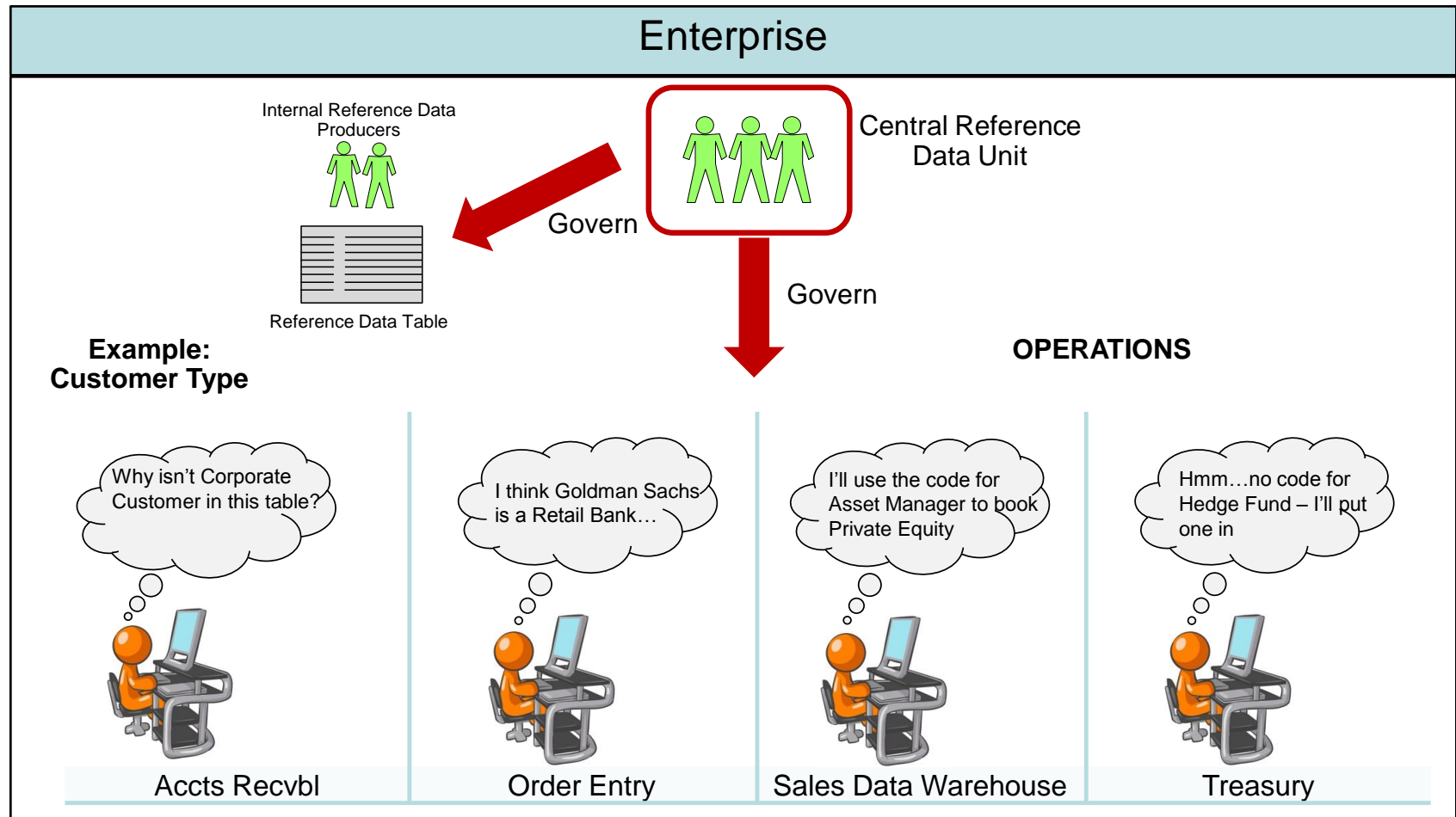
There are a number of reasons why enterprises should set up a central unit for RDM

# Governing and Managing Internal Reference Data



- Typically Internal Reference Data tables are managed poorly and have no governance
- Governance is needed to assign accountabilities and enforce standard processes that drive up quality

# Governing Reference Data in Operational Environments



- Producers of Internal Reference Data may be well governed, but both Internal and External Reference Data can be misunderstood, misused, and abused in operational environments
- This impacts downstream use, data integrity across the enterprise
- Governance is required

# Summary and Capabilities to Consider in Solutions

- Profile an External Authority
- Profile an External Reference Dataset
- Support Semantic Analysis of each Element in Reference Dataset
- Document Semantic Analysis
- Import Reference Data into a Repository
- Assign Accountabilities for RDM Tasks
- Track Changes to Reference Data
- Support Distribution of Reference Data

• There are other capabilities to consider in Reference Data solutions, but these are fundamental.

# Implementing Reference Data Management

# Summary and Capabilities to Consider in Solutions

- Profile an External Authority
- Profile an External Reference Dataset
- Support Semantic Analysis of each Element in Reference Dataset
- Document Semantic Analysis
- Import Reference Data into a Repository
- Assign Accountabilities for RDM Tasks
- Track Changes to Reference Data
- Support Distribution of Reference Data

• There are other capabilities to consider in Reference Data solutions, but these are fundamental.

## Focus:

- TopQuadrant was founded in 2001
- Our focus is to harness emerging technology to build practical but innovative business applications.

## Foundation:

- We continue our strong commitment to standards-based approaches to data semantics

## Our Mission:

- Empower people  
—by making enterprise information meaningful

[www.topquadrant.com](http://www.topquadrant.com)

# An RDM Solution Should Emphasize Flexibility and Extensibility

- Provide rich information structures that enable users to capture anything they need
- Enable faster responsiveness to new data requirements by allowing extension of models
- Capture data quality rules -- by those who do the governance
- Make it easy to create exports of tailored views of reference datasets as web services



# Summary and Capabilities to Consider in Solutions

- Profile an External Authority
- Profile an External Reference Dataset
- Support Semantic Analysis of each Element in Reference Dataset
- Document Semantic Analysis
- Import Reference Data into a Repository
- Assign Accountabilities for RDM Tasks
- Track Changes to Reference Data
- Support Distribution of Reference Data

- There are other capabilities to consider in Reference Data solutions, but these are fundamental.

# ISO 10383 - Market Identifier Codes

## Codes for exchanges and market identification (MIC)

This International Standard specifies a universal method of identifying exchanges, trading platforms, regulated or non-regulated markets and trade reporting facilities.

Market Identifier Code	name-institution description	acronym	operating MIC	creation date	website	comment	city
AATS	ASSENT ATS		AATS	1-Oct-11	WWW.SUNGARD.COM	ASSENT ATS IS A US EQUITY DARK POOL OPERATED BY ASSENT, LLC - A SUBSIDIARY OF SUNGARD.	JERSEY CITY
ALDP	NYSE ALTERNEXT DARK	AMEXDARK	XNYS	1-May-11	WWW.EURONEXT.COM/LANDING/EQUITIESOP-21363-EN.HTML	DARK POOL.	NEW YORK
AMXO	NYSE AMEX OPTIONS	NYSE	XNYS	1-Nov-08	WWW.NYSE.COM		NEW YORK
AQUA	AQUA EQUITIES L.P.		AQUA	1-Sep-08	WWW.AQUAEQUITIES.COM	REGISTERED MARKET FOR BLOCK EQUITIES	NEW YORK
ARCD	ARCA DARK	ARCADARK	XNYS	1-May-11	WWW.NYSE.COM	DARK POOL.	NEW YORK

### Publication and modifications implementation timeline

The MIC list is published on the second Monday of the month or the following business day if it falls on a public holiday in the country of the ISO 10383 Register.

### Next publication dates

2015	June 8
	July 13
	August 10
	September 14
	October 12
	November 9
	December 14

The published List of MICs includes all requests introduced by applicants until the first Monday of the month, after successful validation by the RA. Valid requests are incorporated in the publication of the month on a best effort basis.

# Information About Reference Data: An Illustration – Market Identifier Codes

## Overview

<b>identifier:</b>	market-identifier-codes
<b>official name:</b>	ISO 10383 - Market Identifier Codes
<b>purpose:</b>	This International Standard specifies a universal method of identifying exchanges, and related information in order to facilitate automated processing.
<b>main entity (class):</b>	<u>Market Identifier Code</u>
<b>related entity (class):</b>	<u>ISO Country</u>

- Support Semantic Analysis of each Element in Reference Dataset

### ▼ Property Definitions (Semantic Analysis)

<b>MIC Country:</b>	Country where the market is registered. Linked directly from Country Codes defined in 'Country Codes'.
<b>acronym:</b>	Known acronym for the market. From <u>ISO 10383</u> .
<b>city:</b>	City where the market is located. From <u>ISO 10383</u> .
<b>comment:</b>	Any additional information worth mentioning to help users with identifying the exchange. From <u>ISO 10383</u> .
<b>description:</b>	Description of the market. From <u>ISO 10383</u> .
<b>name-institution description:</b>	Name of the market. From <u>ISO 10383</u> .
<b>o/s:</b>	O (Operating) or S (Segment) indicating whether the MIC is an operating MIC or a market segment. From <u>ISO 10383</u> .
<b>operating MIC:</b>	MIC representing the market at operating level. An operating MIC identifies the entity or facility in a specific country; it is the 'parent' MIC. From <u>ISO 10383</u> .
<b>status:</b>	Active status of the Market Identifier Code. active, modified (since last publication), deleted. From <u>ISO 10383</u> .
<b>status date:</b>	date indicating when the MIC was last modified. From <u>ISO 10383</u> .
<b>website:</b>	Website of the market. From <u>ISO 10383</u> .

- Document Semantic Analysis

### ▼ Subscription

# Information About Reference Data: An Illustration –*Market Identifier Codes-2*

## ▼ Subscription

**standardized by:** ISO

**maintained by:** ISO

**sourced from:** SWIFT SCRL

**URL:** [http://www.iso15022.org/MIC/ISO10383\\_MIC.xls](http://www.iso15022.org/MIC/ISO10383_MIC.xls)

**access payment type:** Free

**change frequency:** Monthly

**access to update payment  
type:** Free

**update file URL:** [http://www.iso15022.org/MIC/ISO10383\\_MIC.xls](http://www.iso15022.org/MIC/ISO10383_MIC.xls)



- Profile an External Authority



- Profile an External Reference Dataset

## ► RACI



Meaningful Metadata Guides the Proper Use and  
Governance of Reference Data



## ▼ RACI



**responsible: +** North America's Data Steward  



**accountable:** Chief Data Officer  

- Assign Accountabilities for RDM Tasks

**informed: +** Marketing Strategy Program Manager  

VP of Business Analytics  

**informed: +** Flight Search Application Owner  

Flight Tracker Application Owner  

- Each Reference Dataset Requires Specific Information in the Dataset
  - More is needed than just a code and a description
- For each dataset you need to be able to define what data is in the set
- Take for example three different reference datasets
  - Market Identifier Codes
  - Country Codes
  - Currency Codes

## Market Identifier Codes

Market Identifier Code	name-institution description	acronym	operating MIC	creation date	website	comment	city
AATS	ASSENT ATS		AATS	1-Oct-11	WWW.SUNGARD.COM	ASSENT ATS IS A US EQUITY DARK POOL OPERATED BY ASSENT, LLC - A SUBSIDIARY OF SUNGARD.	JERSEY CITY
ALDP	NYSE ALTERNEXT DARK	AMEXDARK	XNYS	1-May-11	WWW.EURONEXT.COM/L21363-EN.HTML	<b>DARK POOL.</b>	NEW YORK
AMXO	NYSE AMEX OPTIONS	NYSE	XNYS	1-Nov-08	WWW.NYSE.COM		NEW YORK
AQUA	AQUA EQUITIES L.P.		AQUA	1-Sep-08	WWW.AQUAEQUITIES.COM	REGISTERED MARKET FOR BLOCK EQUITIES	NEW YORK
ARCD	ARCA DARK	ARCADARK	XNYS	1-May-11	WWW.NYSE.COM	<b>DARK POOL.</b>	NEW YORK

## Country Codes

ISO Country	iso 3166 2-alpha country code	iso 3166 3-alpha country code	numeric code	ISO 3166 status	independent	former country
Trinidad and Tobago	TT	TTO	780	Officially assigned	TRUE	
Tunisia	TN	TUN	788	Officially assigned	TRUE	
Turkey	TR	TUR	792	Officially assigned	TRUE	
Turkmenistan	TM	TKM	795	Officially assigned	TRUE	USSR
Turks and Caicos Islands	TC	TCA	796	Officially assigned	FALSE	

Currency code	Currency name	Countries using this currency	Numeric code	Issuing country
AED	United Arab Emirates dirham	United Arab Emirates	784	United Arab Emirates (the)
AFN	Afghan afghani	Afghanistan	971	Afghanistan
ALL	Albanian lek	Albania	008	Albania
AMD	Armenian dram	Armenia	051	Armenia
ANG	Netherlands Antillean guilder	Curaçao (CW), Sint Maarten (SX)	532	Curacao, Sint Maarten (Dutch Part)
AOA	Angolan kwanza	Angola	973	Angola
ARS	Argentine peso	Argentina	032	<a href="#">Argentina</a>

## Currency Codes



- Reference data quality – correctness, consistency, accuracy – is important
- Different code sets require different sets of rules
- For example:

**label:** ISO Currency

**constraint:** Values of numeric code must have between 3 and 3 characters

## Currency Code Info

**currency:** + Argentine Peso

**alphabetic code:** ARS

**numeric code:** + 32

**country using currency:** + Argentina

**entity:** ARGENTINA

### 1 Constraint violations:

- Value 32 must have between 3 and 3 characters  
Are you sure you want to make these edits?

Cancel

OK

- Adding and changing rules needs to be an end user capability



## I. Market Identifier Codes (for US only)

Market Identifier Code	name-institution description	acronym	operating MIC	website	city
AATS	ASSENT ATS		AATS	WWW.SUNGARD.COM	JERSEY CITY
ALDP	NYSE ALTERNEXT DARK	AMEXDARK	XNYS	WWW.EURONEXT.COM/LANDING/EQUITIES	NEW YORK
AMXO	NYSE AMEX OPTIONS	NYSE	XNYS	WWW.NYSE.COM	NEW YORK
AQUA	AQUA EQUITIES L.P.		AQUA	WWW.AQUAEQUITIES.COM	NEW YORK
ARCD	ARCA DARK	ARCADARK	XNYS	WWW.NYSE.COM	NEW YORK
ARCO	NYSE ARCA OPTIONS			WWW.NYSE.COM	NEW YORK
ARCX	NYSE ARCA	NYSE			NEW YORK

• Support Distribution of Reference Data

## II. Market Identifier Codes (worldwide)

Market Identifier Code	name-institution description	acronym	operating MIC	website	MIC Country	city
360T	360T		360T	WWW.360T.COM	Germany	FRANKFURT
ACEX	ACE DERIVATIVES & COMMODITY EXCHANGE LTD		ACEX	WWW.ACEINDIA.COM	India	MUMBAI
AFET	AGRICULTURAL FUTURES EXCHANGE OF THAILAND		AFET	WWW.AFET.OR.TH	Thailand	BANGKOK
AIXE	AIXECUTE		AIXE	WWW.BEKB.CH	Switzerland	BERNE
ALTJ	ALTERNATIVE EXCHANGE	ALTJ	XJSE	WWW.JSE.CO.ZA/MARKETS/ALTJ.ASPX	South Africa	JOHANNESBURG

# Implementation Checklist

---

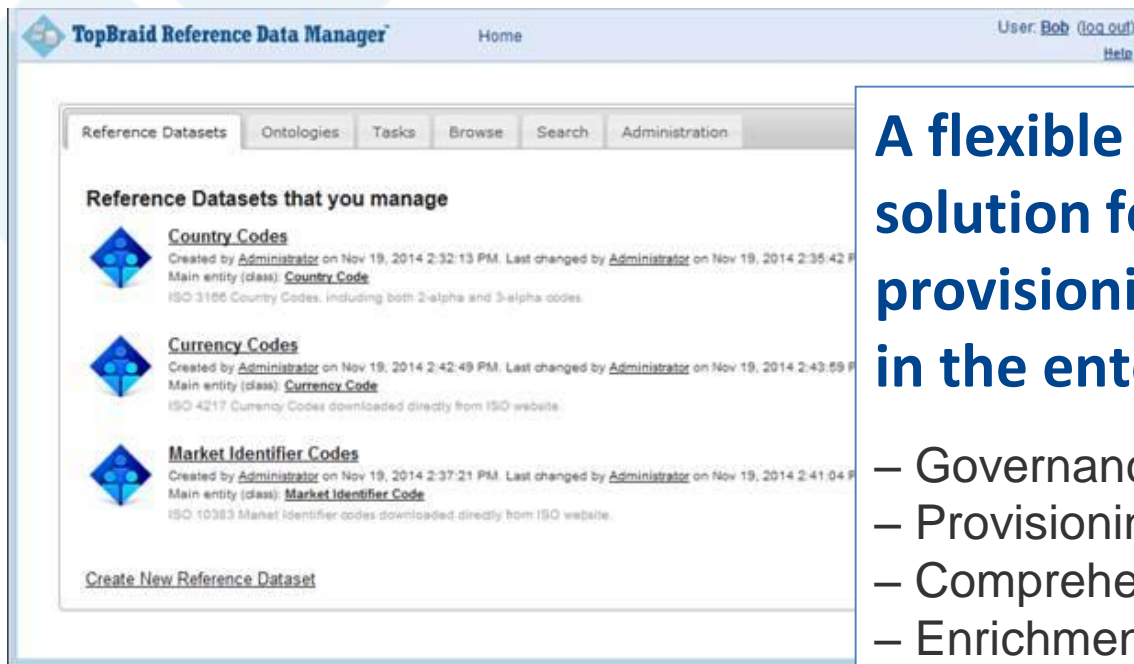
- Manage information about reference datasets as well as the reference data
- Document the meaning (semantics) of reference data
- Include extensible models for reference data and metadata
- Make it easy to create exports of tailored views of reference datasets as web services
- Semantically integrate management of business concepts (models) reference data and metadata

Put the power directly in the hands of Data Stewards, Application Architects and other stakeholders who govern and use reference data

# What do Comprehensive RDM Solutions Require?

---

- Much more than a tool – no matter how good, flexible and extensible it is
- A Central Reference Data Management Unit and Reference Data Operating Model are needed
  - “Governance is needed to assign accountabilities and enforce standard processes that drive up quality”
- Integrated process, methodology, collaboration, and organizational change are essential ...
- *together with* the right technology that puts the right capabilities in the hands of those who govern and use reference data



[topquadrant.com/rdm](http://topquadrant.com/rdm)

**A flexible web-based solution for governing and provisioning reference data in the enterprise:**

- Governance
- Provisioning
- Comprehensive metadata
- Enrichment

- Recent Whitepaper by Malcolm:  
***“The Foundations of Successful Reference Data Management”***
- [robert@topquadrant.com](mailto:robert@topquadrant.com)



# Questions and Answers

---

## Data Governance Preparedness for Reference Data Management



**June 10, 2015**

Presented by:

Malcolm Chisholm – Askget.com Inc.

Robert Coyne – TopQuadrant Inc.

[mchisholm@askget.com](mailto:mchisholm@askget.com)

[robert@topquadrant.com](mailto:robert@topquadrant.com)