White Paper

Maturing Reference Data Management

Developing a formal Reference Data Management (RDM) practice aligned with other key functions of Enterprise Information Management (EIM) is foundational. If done well, it has a high payback.

INTRODUCTION

Reference data is found in every application used by an enterprise including back-end systems, front-end commerce applications, data exchange formats, and in outsourced, hosted systems, big data platforms, and data warehouses. It can easily be 20–50% of the tables in a data store. And the values are used throughout the transactional and mastered data sets to make the system internally consistent.

How well it is managed has a major impact on every aspect of an organization's use of data — from the integrity of its business intelligence reports to the success or failure of its system integration efforts. Managing reference data well and developing a formal Reference Data Management (RDM) practice aligned with other key functions of Enterprise Information Management (EIM) is foundational, and if done well, has a high payback. Done poorly, it has a high cost.

In this paper, we describe a RDM maturity model and one likely path an enterprise can take to mature RDM with the five levels shown in Figure 1.

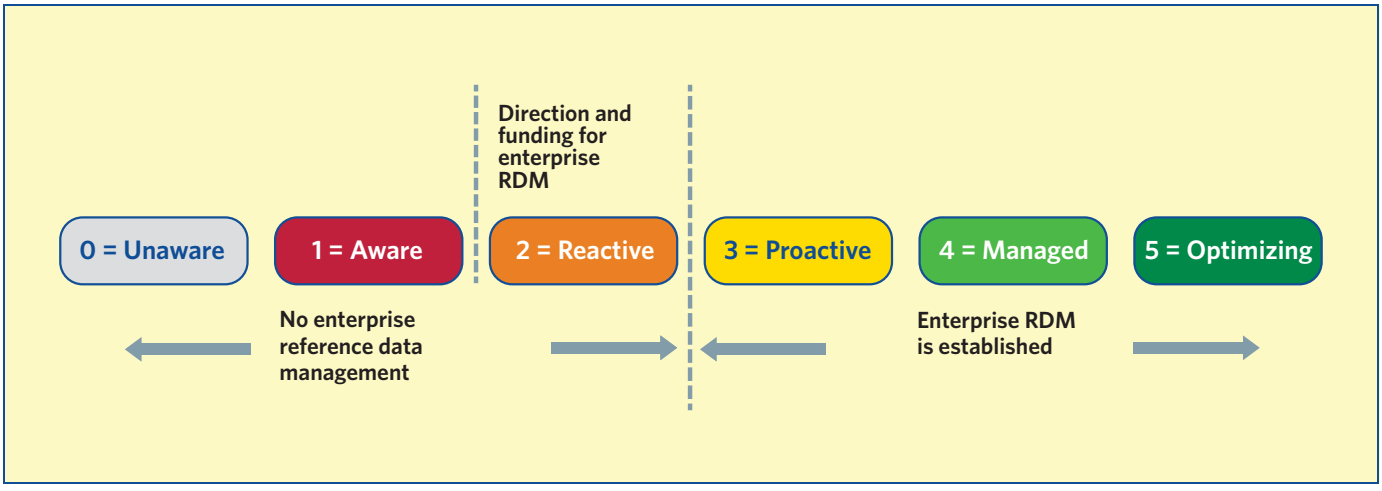


Figure 1. Maturity levels and high level maturation steps

We describe how RDM affects EIM at each maturity level. The first section of the paper sets the context of RDM in terms of the reference data sources and the required EIM capabilities. The next section provides a detailed explanation of RDM at each maturity level. It shows key activities of information value chains as they emerge.

Most medium and large enterprises today are somewhere between level 2 (Reactive) and level 4 (Managed) in their RDM maturity (see Figure 1). Most agree that they need to reach level 5 (Optimizing). Having said this, enterprises today are diverse and dynamic. It is not unusual for some of the divisions to be at level 3 and for others to be at level 1 or 2. Each enterprise will take their own path to mature RDM and EIM. This is an overview and approach that will help.

In this whitepaper, we broaden RDM to the management of "Code Sets" including any data sets which have "Code Values" as part of their representation. Even if they fall within the other categories of data. This is practical, particularly early in maturing RDM, because these different types of data are usually managed as "Reference Data" by practitioners. **(See the section, *Data, Categories, Code Sets, Reference Data*, in the Appendix, page 20.)**

Reading this white paper will help you to:

- Assess your level of RDM maturity
- Decide what level of maturity you want to be at and why
- Understand what steps you need to take to move up the maturity ladder
- Understand the relationships of RDM as it coevolves with other EIM functions

Each maturity level is described from the following perspectives:

- **People** — Who is responsible for reference data? How do roles use and manage the reference data?
- **Process** — What processes are in place for managing and governing reference data?
- **Technology** — What technology is used in reference data governance?
- **Data** — What is the quality of reference data?
- **Pain** — What are the consequences and pain points of being at this maturity level?
- **Business Costs** — What are the hard (if available) and soft costs for each level?
- **Next steps** — What an organization needs to do to move up to the next level?

As an enterprise reaches maturity level 5 (Optimizing), RDM will be executed as a key, interacting part of all business activities and be subsumed as part of EIM. There will be many benefits including:

- Decreased project development time
- Lower integration costs, better interoperability
- Less rework, more leverage
- Higher value services such as impact analysis, information directories, business services
- Faster ramp-up, smarter staff
- Improved business agility

The enterprise becomes more collaborative, efficient,and nimble

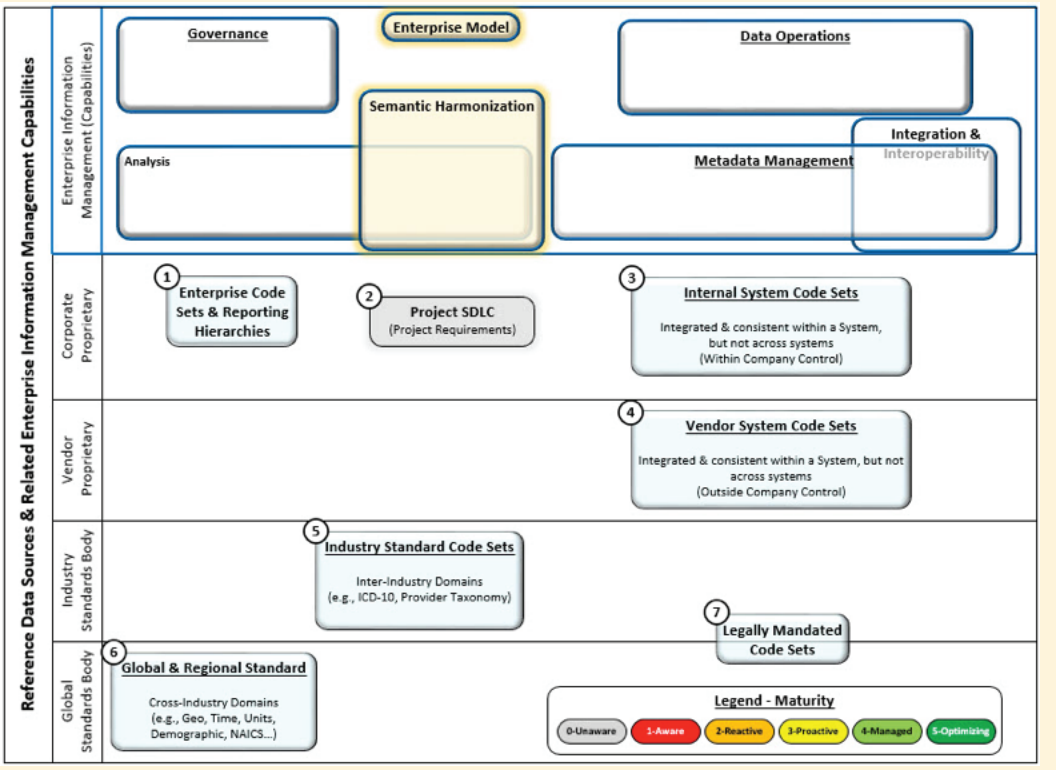
REFERENCE DATA MANAGEMENT MATURITY DIAGRAMS

Figure 2 is a context diagram that depicts different contexts for sourcing of reference data that needs to come under management:

- Top swim lane contains the key EIM functions that will come into play as RDM matures.
- Lower swim lanes provide different contexts for sourcing of reference that needs to come under management and potentially as the source for the "Enterprise Standard Code Sets."

Based on this, at each level we will show a diagram depicting maturity growth of RDM in terms of sources, data flows, EIM capabilities, and their value chains. Since RDM participates and is worked on by many cross-function practitioners, it is not labeled in the diagram — it is the diagram. Diagrams build up for each maturity level showing changes to data flows and the implementation of more formal processes as RDM matures.

To find out how TopBraid Enterprise Data Governance (EDG) provides the capabilities needed to help you mature your Data Governance and RDM practices, we invite you to download this companion whitepaper: ***Maturing Information Governance with TopBraid EDG***, (topquadrant.com/docs/whitepapers/Maturing-Information-Governance-with-TopBraid-EDG.pdf).



1 Business defined code sets for executing, managing, and analysis of the enterprise.

2-3 Locally developed and implemented code sets, usually during an SDLC. Sometimes done directly in the application

4 Vendor developed or adopted. An enterprise gets these via upgrades.

5-6 Maintained by standards bodies. Often refer to and use other standards in their composition.

7 Imposed by law. Must be added and integrated into processes by deadline

Figure 2. Reference Data Sources and Capabilities

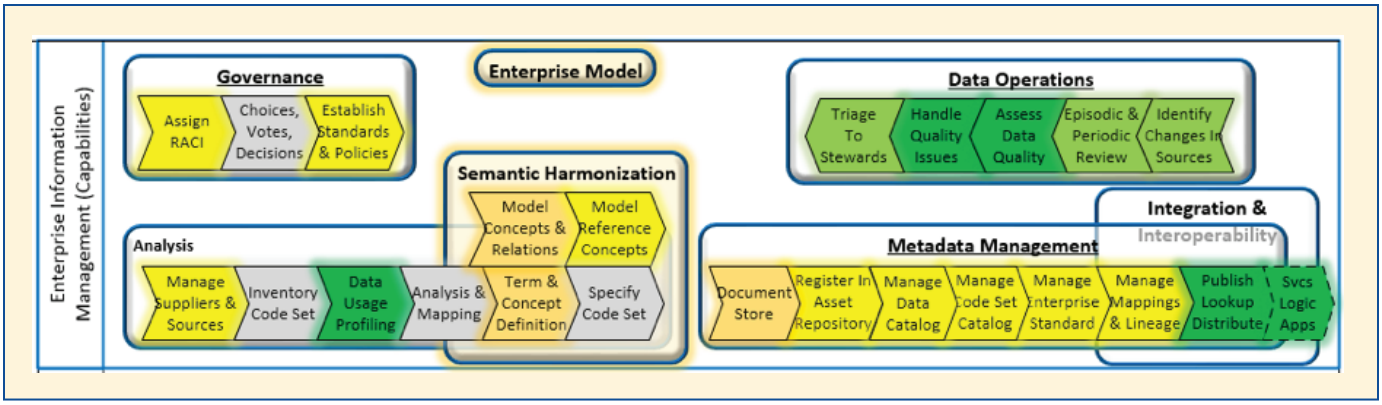


Figure 3. Important EIM capability value chains involved with Reference Data Management

Table 1, (page 5), provides more details about sources of code sets and standards. Table 2 in the Appendix, (page 18) , lists the most important best practices.

Figure 3 shows the full set of important EIM value chain activities that come into play as RDM matures.

Enterprises perform a full set of EIM activities only at level 5 of maturity. Diagrams depicting each level of maturity only show activities performed at a given level. While we recognize that the EIM activities could be "drilled down" deeper, we show them at the level of roll up sufficient for this whitepaper.

GENERAL DESCRIPTION OF REFERENCE DATA MATURATION

Internal reference datasets in proprietary enterprise systems typically start as very simple code lists or reference tables in systems. As business questions become richer and decisions more refined, demands on the quality, structure, and richness of the reference data grow. ***Even simple internal reference code sets become enriched with:***

- Drillable hierarchies used for analyzing, organizing, and reporting
- Analytic results such as scores or weighting factors
- Addition of rule factors mapped as a rule table
- Addition of relationships and other information over and beyond codes and descriptions

As organizations move up the maturity ladder, they need to understand, harmonize, and connect reference data that already exists. Manual analysis or reverse engineering at the conceptual level is often seen as too much work due to the quantity of meta-objects and their values. The task of harmonizing requires mixed manual and automated processes to scale, and ideally, a well-selected set of software tools to manage and govern the models, entities, relationships, and code set values.

Further, practitioners of modeling, data and business analysis, and stewards become mutually and increasingly responsible for ensuring the understanding, development, and management of context models, information models, conceptual reference entities and their values, and the mappings to where they are found in systems. Collaboration between these practitioners ensure the validity of data integration specifications that enable the semantic integration necessary for system interoperability.

ENTERPRISE INFORMATION MANAGEMENT CONTEXT OF REFERENCE DATA MANAGEMENT

The following sections provide a detailed look at how RDM could mature in an enterprise and become more effectively aligned with other EIM capabilities. RDM begins as ad hoc efforts within an System Development Lifecycle (SDLC). As it evolves, it becomes an intertwined set of assets and processes with other important EIM functions such as information analysis and design, stewardship, information governance, metadata management, and related activities. Rather than being separate activities, they merge and cooccur resulting in information assets such as code sets, information models, mappings, specifications, and integrations.

Reference data is used, managed, created, enumerated, refreshed, reused, modeled, harmonized, conformed, and standardized across all these activities. Table 1 summarizes the emergence and evolution of EIM capabilities as RDM matures.

TABLE 1: Summary of Reference Data Management Maturity Levels

MATURITY LEVEL	NAME	PRACTICES EMERGING	CAPABILITIES MET
0 = Unaware	• Project siloed	• Ad hoc	• SDLC
1 = Aware	• Practitioner siloed	• Ad hoc practitioner	• SDLC
2 = Reactive	• System siloed	• Multi-siloed SME, ad hoc • Information architecture • Central enterprise modeling	• Document management • Information architecture
3 = Proactive	• Enterprise harmonization	• Information governance • Central RDM business unit • Managing external reference data • Establish enterprise standard code sets • Governing internal reference data • Establish enterprise model and associated practices	• Ability to assign accountabilities for aspects of RDM per code set, especially for internal code sets • Profiling source or vendor of an external reference data standard • Capture metadata to describe an external reference data code set and each element within it • Perform semantic analysis of each permissible value of a reference data set and semantically map to business concepts • Enterprise modeling into layered ontologies to include business conceptual entities and the more detailed conceptual reference domains. • Ability to import code sets from external or internal sources into a central repository. • Ability to enrich code sets during import. • Ability to extract from sources filter transform, enrich as much as possible. Ideally, this would be meta data driven.
4 = Managed	• Enterprise RDM	• Subscription management • Business steward-driven internal code sets • Code set gaps filled, mastered and governed ahead of project need (planned work) • Cross discipline, unified mastering and harmonization	• Ability to track changes to reference data from catalog analysis • Ability to identify and fill gaps or fix mistakes
5 = Optimizing	• EIM integrated RDM	• Operational data quality practice — data profiling / DQ triage • Enterprise standard code set distribution and publication	• Ability to track changes to reference data from data profiling analysis • Ability to distribute reference data code sets

DETAILED RDM MATURITY LEVELS

0 = Unaware

0 = Unaware — Project Siloed

The use and development of reference data is done by project teams during the system development lifecycle and is usually handled as an afterthought. Each project inventories and specifies, often directly, in software code, reference data and other code sets but only enough to meet project deliverables. Decisions are made locally and are ungoverned. The assets developed are often not specified formally. They remain undocumented or are temporary documents with just enough to meet the deliverable.

Integration code is written to support only what is known now. Analysis is hurried and minimized and the knowledge used to complete the project is not captured and maintained. Shortcuts are often taken by simply adopting codes into the primary operational system, trusting vendor-based systems more than proprietary. Reference data discovered for one project is rediscovered for each subsequent project. The same is true for any integration code. All of this becomes a tangled web of complexity over time.

Faster changing code sets (such as enterprise structure data like market segment or line-of-business used for measuring performance) quickly become unsynchronized across different systems produced by projects. As a result, management decisions are often based on incoherent results. Formal practices and capabilities do not exist and the enterprise suffers, resulting in indirect costs.

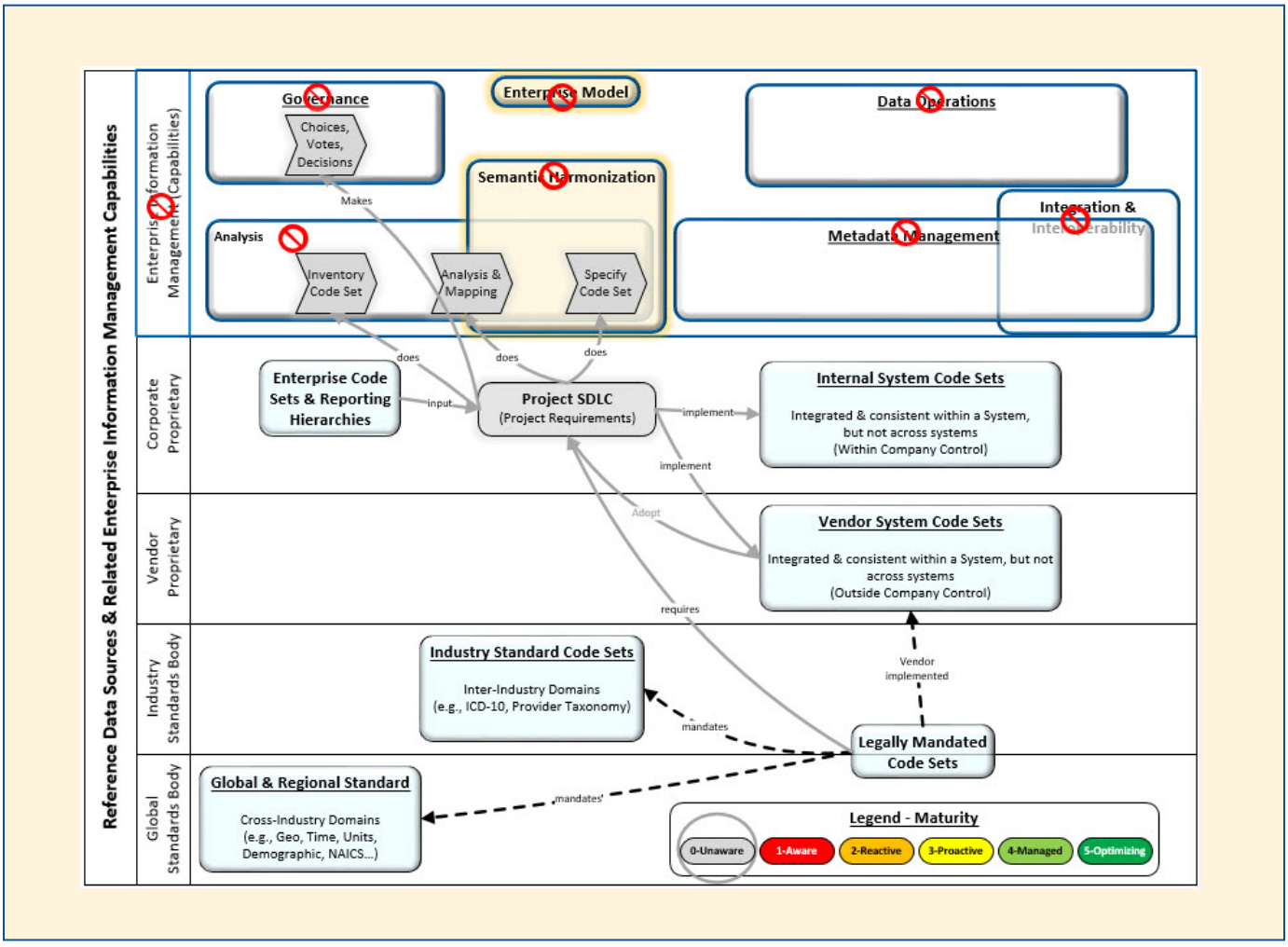


Figure 4. Project Siloed — Reference data at 0 = Unaware Maturity Level

0 = Unaware

RESULTS: REFERENCE DATA AT 0 = UNAWARE MATURITY LEVEL

Project Siloed

Reference data is an afterthought. Handled on an individual project basis, often towards the end of a project's cycle. A commonly heard question is, "Why did this come up so late in the project?"

People

Deliverable deadline focused. No view of the big picture outside the project. Reference data not thought of as a distinct set of assets — it's just data. Whoever is assigned becomes responsible for it. Low trust in prior analysis so work is repeated.

Process

"Just get it done." No effort is made to catalog or reuse reference data or other code sets. Refer only to what is necessary for the existing systems during analysis. Often new code sets are created by copying them from existing systems, including vendor managed systems. Often code sets are implemented directly in code, in file lists, or, if lucky, in databases. Efficiency is dependent on practitioner skill. Each project starts over, even on existing systems.

Technology

Use what is easiest and fastest. Text files, spreadsheets, software code, legacy system code sets. Often copies are made but no effort is made to track with original.

Data

Created each time. Don't trust prior work. Must reinterview business SMEs for requirements. Take or create within project deadline by adjusting quality and scope. Duplicated, inconsistent, forgotten. Often overlapping in meaning and values. Quality is low causing downstream issues in both operations and future project development efforts.

Pain

Continuous reinvention of the wheel. Projects are slow. Discovery and analysis is difficult. No development of business experts. Software systems become more complex and entangled. New code sets are difficult to process and are handled in an ad hoc way requiring custom software development. Any cross-cutting efforts such as business intelligence, enterprise data integration, and reporting are very costly and the end results are low quality. Processes are manual and not repeatable. Requirements are discovered in real-time during development. Missed requirements result in complex, last minute fixes in integration code. All causing more pain in the next project.

Business Costs

Business agility is compromised. Projects take too long with loss of political capital. Projects become too big and costly. Uncertainty and lack of prior knowledge capture means efforts repeat with each project touching the same data. No business memory. Integration costs can be well above 50% overall project cost. Next project will cost more and take longer getting worse with each iteration.

Next Steps

- Increase awareness of lack of knowledge reuse and the consequences
- Begin using previously created artifacts
- Analyze in-scope systems, not just within the bounds of known requirements
- Practitioners should be encouraged to capture their own work as an asset for later reuse

DETAILED RDM MATURITY LEVELS

1 = Aware

1 = Aware — Practitioner Siloed

At this level, RDM still exists only within individual projects and/or is maintained by individual practitioners that may move across projects. Practitioners are becoming aware of the different roles working on reference data, such as data modelers, analysts, and developers, but have not formally begun working together. There is still limited recognition of the importance of reference data outside of what it takes to meet the project deadline.

Many meetings happen as teams deal with contention between project deadlines and getting reference data values correct. Compromises are made that have downstream effects. There is no integration or harmonization across systems, but some analysis starts to happen when a new system is implemented with a goal to improve consistency and reuse. As skill develops in teams, there is some local reuse.

Decisions are still made locally and are ungoverned but they may be written down in the project documents. The documentation may be developed to some extent but remains in the practitioners file sets for later reuse. The practitioners, being aware of the project scope, begin to analyze the reference data within the scope of the affected systems by leveraging their prior knowledge. They try to anticipate missed requirements. Formal practices and capabilities do not exist yet, but practitioners have begun capturing and managing their own knowledge. Sharing of knowledge happens sometimes when people become aware of someone's "expertise."

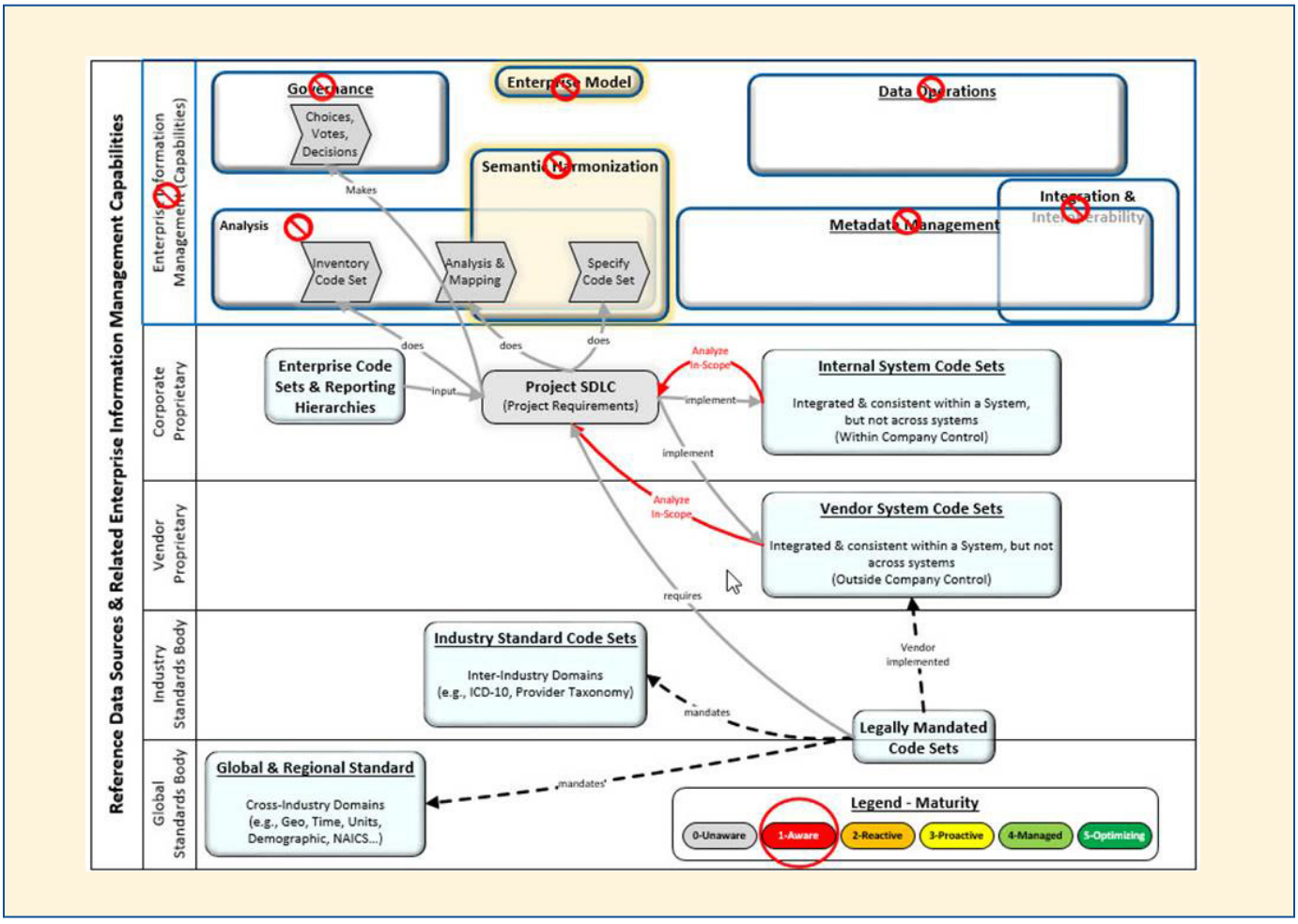


Figure 5. Practitioner Siloed — Reference data at 1 = Aware Maturity Level

1 = Aware

RESULTS: REFERENCE DATA AT 1 = AWARE MATURITY LEVEL

Project Siloed

"I have done this before, let me find the documents. I know where to look in file stores and in the systems."

People

Practitioners leverage hoarded knowledge:

- Tired of repeating the same work, they keep copies of their analysis for reuse
- Are often brought in for heroic measures to save deadlines because they "know"
- Become aware of industry and global standards and may even use them for the project. But will usually adopt them directly from existing vendor or proprietary systems
- Analyze in-scope systems to discover reference data to inform project requirements

Process

Relies heavily on practitioner knowledge. No longer just adopt vendor system reference data. Analyze "in-scope" systems. Local reuse is happening within project scope. Standards used in existing vendor and proprietary systems are leveraged. May seek actual standards within SDLC project scope. No formal project knowledge capture.

Technology

Use and build on hoarded documents. Project documents, text files, spreadsheets, software code, legacy system code sets. Analyst hard disk is used to protect prior project documents and analysis.

Data

Comes from practitioner-hoarded knowledge enriched by their analysis skill against prior sources and new requirements.

Pain

Hoarding makes experienced practitioners too valuable and they become limited resources. Building skilled IT teams is difficult as knowledge sharing is light. Prior knowledge is lost in file system. Earlier documents are untrusted because no one is sure whether they are still current and accurate. Work is repeated over and over until the pain is too obvious and practitioners lobby for a more mature process. Late arriving and poorly analyzed reference data requirements take additional time and delays projects. Outsourcing to "consultant" experts often begins but they lack sufficient domain/corporate knowledge and drive out deep corporate staff. Knowledge drain occurs as staff leave positions. Revolving door becomes expensive.

Business Costs

Projects slow down to a crawl when expert practitioners leave or become too busy. Salary for highly skilled practitioners grows rapidly. Business often must look outside to get skilled staff. Project siloes and lack of reuse causes duplication and the need for additional staff to handle it. Consulting costs often scale with belief it is a "one time" need. Knowledge loss becomes existential risk.

Managers become stressed because they have no control over increasing project time, costs, and lack of knowledgeable staff. They start asking questions on how to improve.

Next Steps

- Build awareness of cross-cutting standards and code sets
- Analyze use of reference data across in-context systems
- Encourage collaboration amongst EIM practitioners across projects
- Introduce more formal document management for storing reference data artifacts
- Begin rebuilding and retaining knowledgeable staff

DETAILED RDM MATURITY LEVELS

2 = Reactive

2 = Reactive — System Siloed

At this level, the need for different EIM capabilities, often led by the need for RDM, begin to be appreciated, driven bottom-up and within projects. The "Information Architecture" capability and its primary asset, the "Enterprise Model," emerge, begin to provide top-down information organization and structure and a formal method, collaborative space, and modeling assets to capture knowledge.

Two important processes, "Term and Concept Definition" and "Model Concepts and Relations" arise. The information architecture capability provides the collaboration among practitioners to identify and manage reference data and the associated "Conceptual Reference Domains." The other EIM capabilities, including RDM, do not yet emerge separately, but the organization and structure around Information architecture is preparing the groundwork for enterprise-wide RDM. As the "Conceptual Reference Domains" become modeled, often somewhat informally as business glossaries, practitioners also begin to define "Conceptual Value Meanings" as permissible meanings and names.

The drive is still to understand the existing reference code sets in project-scoped systems. Practitioners begin to think in terms of concepts and values, particularly in large projects that cross enterprise systems. They recognize the need to have a common concept, terms, and values for code sets such as gender or status. The need for cross-system mappings to allow the systems to interoperate using the shared meanings is also acknowledged. The organization becomes more aware of its reference data and starts to catalog and assess it along with other information assets. This is still done in a largely informal and ad hoc way. Teams with tool silos begin competing. Much work is repeated. Only large projects begin these efforts because they recognize the necessity. Reference data standards begin to be looked at but the staffing, tooling, sourcing, and governance is inadequate.

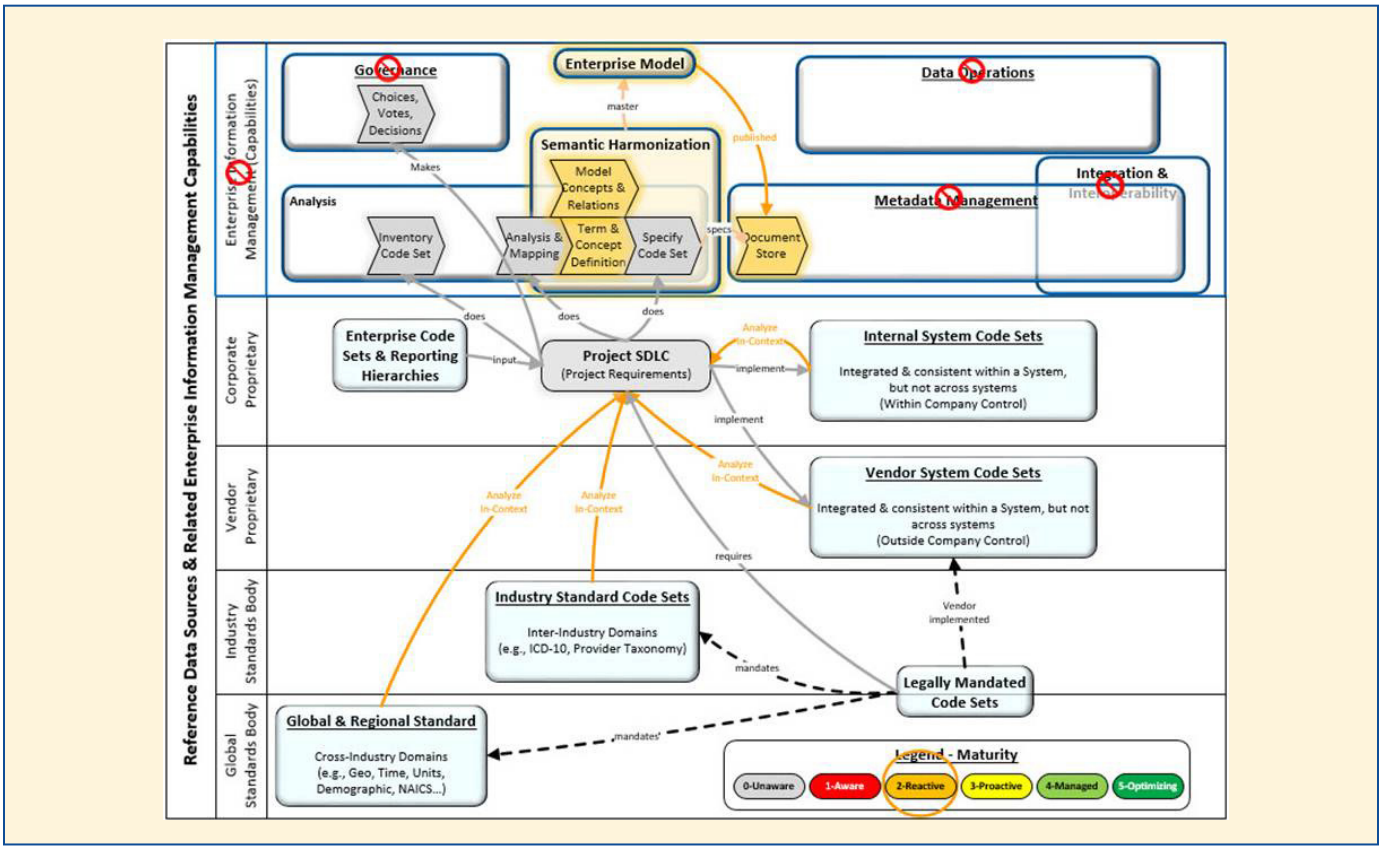


Figure 6. System Siloed — RDM at the 2 = Reactive Maturity Level

2 = Reactive

RESULTS: REFERENCE DATA AT 2 = REACTIVE MATURITY LEVEL

Project Siloed

"Let's use what is in our primary systems! They are what is real and true! Oh, but let's compare them to standards and pick the best system that has them."

People

Since the operational systems are running the business, people believe reference data in these systems should become the standard used across the enterprise. They are also beginning to see the same concept as common categories being handled differently in each system's code sets. Thus, they are becoming aware of global, industry, and system vocabularies and reference datasets. The second time a major system is replaced, practitioners start to appreciate that it is likely to happen again and there is a need to change to an enterprise approach rather than a system-to-system integration approach. An "Information Underground" of experienced practitioners' who may have worked together previously, collaborate across teams. Readiness is being built for an enterprise approach to RDM. Strategic direction bubbles up from practitioners to management until funding and management guidance is established to shift to next level.

Process

Begin to look at standards — including global and industry standards — and choose among them, in a system usage context. Still project-driven but keeping track of what exists and what was done. Much of the knowledge is captured in data models and their files. Data models and documentation, such as specifications, begin to be stored in shared drives but usually in project folders, with copies staying in Analyst C drives, and maybe into "system" organized folders. Access controls are not well defined, making it hard to get to the prior assets. Code sets are implemented by developers during an SDLC.

Technology

Local database or spreadsheets are used to manage reference code sets. There are likely to be competing "asset" management (e.g., metadata or configuration management) systems vying for adoption by various practitioner groups with some political infighting (e.g., by a data warehouse group). This repository is shared across practitioners and projects. Technology becomes a barrier as roles and access controls are held tight and not well thought through. Version control is an issue and occasional operational issues occur when an unexpected change happens to a code set in production. Data modeling software is used to capture concepts, relationships, and ultimately data models built off the conceptual models.

Data

Focus on in-scope systems and data in existing database tables. Some analyst's spreadsheets/databases may begin to manage enterprise code sets, possibly aligning with global and industry standards. Often project managers and architects will insist on starting with what is in the in-scope systems. Duplication is high and cross-system trust of data remains low.

Pain

As systems retire, it is necessary but very difficult to migrate to a new system with different reference dataset values and approaches. Integration work must be redone across touched systems. Expensive, late, and over-budget projects. Reference code sets reworked and remapped. Having no central management of code sets or mappings is an impediment. Typically, point-to-point integrations are being done, which further increases complexity and corporate inflexibility. System migrations are major efforts.

Business Costs

The complexity that has developed over the years is recognized and the business is put at risk if their market changes and they need to shift. Increasingly skilled practitioners cost more and are needed to handle the complexity and ultimately to capture the knowledge deeply enough for real reuse rather than simple data entry for knowledge capture without any other purpose than documentation.

Next Steps

- Increase awareness of lack knowledge reuse.
- Begin reusing previously created artifacts.
- Analyze in-scope systems reference data and compare to others based on concept.
- Lay the foundation and evolve needed additional EIM capabilities, especially "Information Governance" and "Information Asset Management" (aka "Metadata Management") and the related "Information Services" such as the "Enterprise Information Directory," the "Data Catalog," "System Data Dictionaries," Reference Data/Code Set Catalog," "Impact/Dependency Analysis, " "Data Flow Catalog," "Information Usage" (by Capability, by Process, by Team, by ...), and so on.
- Build a business case for enterprise RDM aligned with "Business and Data Analysis" and "Information Architecture." The junction with RDM being "Information Analysis and Design."
- Begin to delineate and understand the different kinds of code sets and establish processes and assets for managing them.
- Launch an effort to select and procure a RDM and/or an inclusive data governance solution.
- Recognize, analyze, and begin implementing enterprise RDM capabilities.

DETAILED RDM MATURITY LEVELS

3 = Proactive

3 = Proactive — Enterprise Harmonization

The "step up" between level 2 and 3 is greater than between other levels. Considerable effort must be made by an organization to reach this level. Important EIM capabilities of "Information Governance," "Business and Data Analysis," "Information Architecture" (including modeling) need to be organized, funded, and staffed. At this level, information becomes a formal business asset meant to be managed and shared. This is the beginning of EIM and the management of "code sets" in "RDM" becomes formalized centrally.

Management is done centrally with decentralization of stewards to ensure appropriate business alignment. Code sets are managed with a longitudinal and enterprise view, strategically sequenced but still in the context of projects. Initially, external standards, that cross system boundaries such as "Currency," "NAICS Codes," "Country" and other geographic code set standards are mastered. Vendors are brought under vendor management and acquisition and refresh processes are established. Information governance provides guidance for prioritization, policies, decisions, and measured accountabilities. Stewards are established for reference data assets. Tools are deployed to manage the related reference code sets and metadata.

Metadata has begun to be managed as well, ideally with an enterprise-level system. Data catalogs and associated inventories of reference data residing in different systems begin to be formally captured and managed. Initial enterprise standards begin to be harmonized and governed. The information asset foundation starts to enable advanced capabilities.

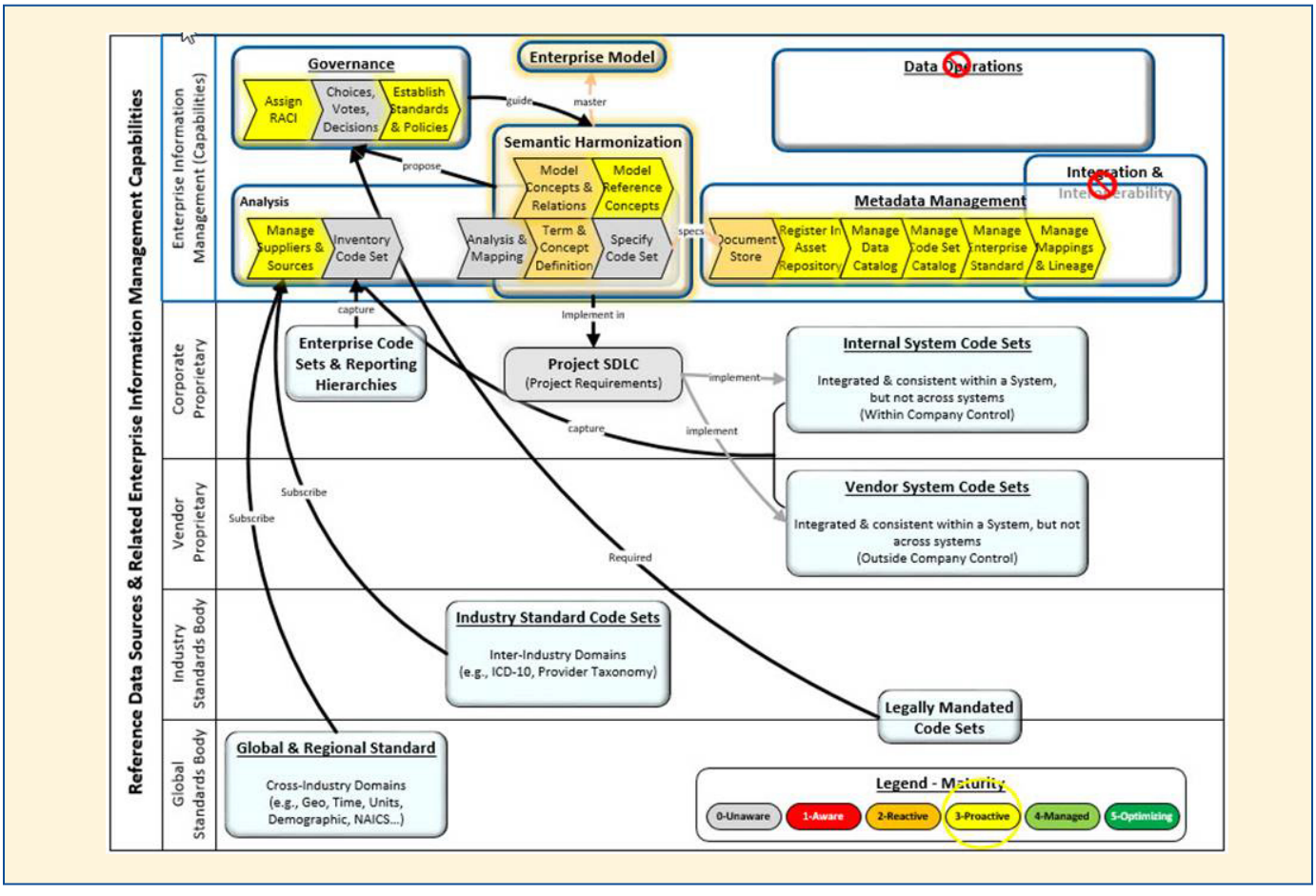


Figure 7. Enterprise Harmonization — RDM at 3 = Proactive Maturity Level

3 = Proactive

RESULTS: REFERENCE DATA AT 3 = PROACTIVE MATURITY LEVEL

Enterprise Harmonization

The RDM function has been sufficiently staffed, often in a data warehouse team, to prioritize key reference data for key projects, ideally ahead of a project's need. The focus is on common reference entities — particularly ones used in operations and reporting– such as gender, address type, age brackets, marital status, important industry or financial standards such as ICD-10 codes in health care, or geospatial standards such as state codes and county codes.

People

Key practitioners are given a role in harmonizing to a common meaning, or adopting a significant standard without analysis. The organization may have formal reference data steward roles or information steward roles, but usually highly skilled data analysts are tapped and tasked with harmonization efforts. A central reference data unit is established and staffed with an information architect and reference data analysts to support stewards and SMEs. These people work and liaise between business experts and SDLC development teams. People in business managing existing code sets resist centralization, and hold tight to their local standards. Analysts must go to them. Practitioners begin to work together toward one set of deduplicated assets. The additional tasks begin as a burden until staff gets used to them. Value is recognized each time an asset is reused. This typically occurs between projects so it takes some time. Change brings much resistance and compromise.

Process

Inventory easily accessible global and industry standards by direct download. Prioritize important code sets needed in projects. Analyze semantics and define conceptual reference domains and enterprise standard code set values. Manage harmonized and enterprise-level reference code sets as assets. SDLC teams use enterprise code sets either directly in a system, if possible, or through semantic mapping to system-level codes to achieve integrations and/or analytic capabilities. Legacy processes remain strong and resistant to change. Central processes emerge but only take on cross-project, cross-enterprise standards for harmonization. Need for dispute resolution becomes a driving need for better governance practices. Processes are being reengineered and tweaked in practice.

Technology

May or may not have specialty tooling in place for governing reference data. Leverage spreadsheets and other formats that are easy for business to use and can be imported easily into current or future tooling.

Data

Global and industry standards, then internal standards, then vendor-based code sets are captured. Semantic mapping data ties system code sets to central conceptual hub of "Conceptual Reference Domains," managed closely with the "Enterprise Model" capturing the business terms, concepts, relationships, and values as "Enterprise Standards." These "Enterprise Standards" are used directly when possible, and are mapped as the semantics of physical reference data when not. Project mistakes which must be fixed later, will be knowingly made under project pressure.

Pain

Period of significant change as practitioners adjust to new procedures and standards. Value is understood, but staffing is usually inadequate and demand outstrips the EIM and RDM team's capacity. There is a growing recognition that project teams need to do much of the work; enterprise information governance must emerge to oversee and ensure that project governance meets the RDM guidance. For organizations that have not yet deployed effective specialized RDM tooling, lack of tooling becomes a key pain point as they experience these challenges:

- Lack of tool assistance in cataloguing of data structures (metadata) and the ability to register types of datasets with focus on knowing what datasets contain reference data
- Difficulty seeing a broad picture of what exists/ how it's used

Business Costs

There is a significant investment to build the foundation on which costs can be reduced as assets become governed. There are political costs in championing this large effort but with careful prioritization, the costs can be balanced with the savings.

Next Steps

- Evaluate, select, and deploy a specialty RDM (ideally as an integrated component of an enterprise metadata solution)
- Establish automated sweep for data catalog analysis
- Create a data catalog by ingesting metadata of data sources
- Use data catalog to identify tables and datasets that hold reference data
- Ingest local reference data into RDM solution
- Start to establish "mastered" enterprise standard code sets
- Deepen the integration of RDM with related asset metadata
- Write RDM policies to guide RDM adoption
- Write/adjust other EIM policies to support RDM practices
- Establish formal accountabilities in job descriptions for stewards and practitioners for the assets they are responsible for
- Establish decision rights for RDM

DETAILED RDM MATURITY LEVELS

4 = Managed

4-Managed — Enterprise RDM

At this level, much of the negotiation and political capital spent to arrive at the "Managed" level is showing payback. Reuse from metadata and reference data systems enable projects to start with what an organization knows. Projects have faster turnaround with better understanding from project teams. Fewer mistakes are being made. Adequate knowledge has been accumulating in the EIM systems to enable automated impact and dependency analysis, and to provide a good basic understanding of information requirements ahead of a project. Gaps are easier to identify ahead of a project and can often be filled before project kickoff, greatly reducing project time and risk.

As practitioner knowledge becomes broader and deeper with the interaction of different roles, processes around EIM become fluid and quick, enabling business agility in dynamic markets. Harmonization processes now include the "mastering" of information and knowledge assets that can be reused over and over. Explicit knowledge captured in the reference and metadata system begins to enable automation of operational processes in the next stage. Projects become leaner. IT gets easier. Business begins to get its needs met more quickly.

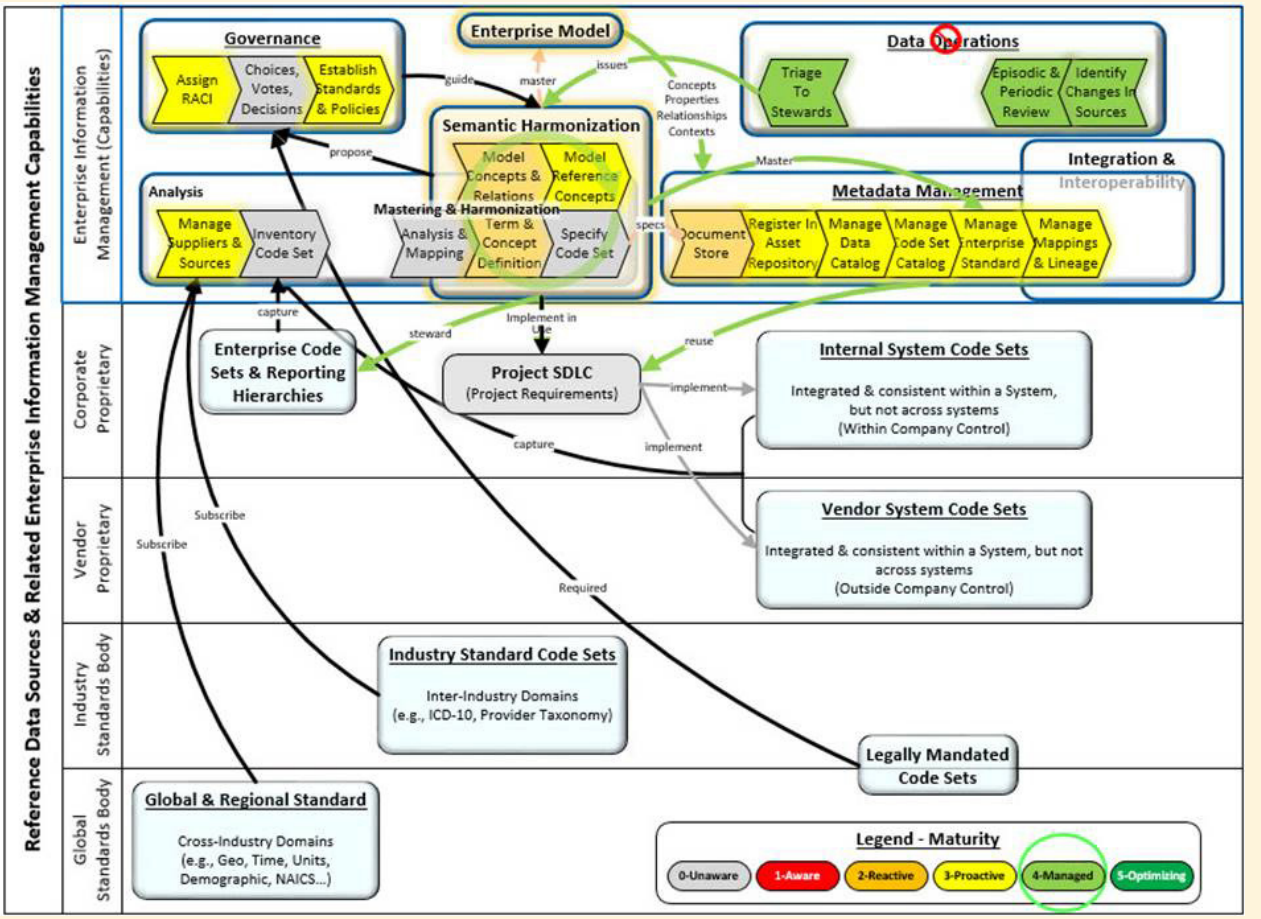


Figure 8. Enterprise RDM — RDM at the 4 = Managed Maturity Level

4 = Managed

RESULTS: REFERENCE DATA AT 4 = MANAGED MATURITY LEVEL

Enterprise RDM

RDM has been adequately budgeted, centralized, and staffed, enabling effective management. The RDM team is engaged at the right time in projects. Gaps in the code set catalog are being filled and mistakes corrected. Integration teams are beginning to rely on the code set catalog for mappings and lookups, and building information services for them. Unusual edge cases for RDM are being addressed. Governance of code sets and mappings are delegated to RDM team and the stewards they support. Escalation processes are formalized, but seldomly required.

People

Practitioners begin to rely on the code set catalog. People jumpstart project analysis by "starting with what is known." Models and categories implemented as code sets are published for any enterprise user through an "Information Directory" coming from the "Enterprise Metadata System," which includes code set management. Practitioners begin to value the benefits of managing code sets as information assets with the overall reduction of time spent harmonizing and explaining issues. EIM policies provide guidance on RDM. Staff knows their accountabilities for RDM and is measured upon the quality of their work. The additional work now seems like a contribution vs. a burden. Catalogs become more complete, harmonized, and staffs starts working on higher-value activities including helping business units leverage assets such as impact analysis, data dictionaries, and business services.

Process

Projects start with what is known. Impact analysis and scoping derived from integrated metadata and the RDM set of applications defines an "Information Asset Pack" that includes the glossary, reference data, and metadata for conceptual, logical, and physical assets in project scope. Project staff analyze existing assets, identify gaps, fill them, and get reviews by information stewards, RDM staff, with approval within EIM governance processes. Any changes/fixes/new reference data, progress through a lifecycle with an updated status and record, in appropriate systems. The project SDLC builds automation that leverages the centrally-managed reference data and its semantic mappings to system data to provide operational, metadata-driven integration within the production environment. Assets become more complete and accurate. Ultimately, the RDM team prioritizes work ahead of a project, eliminating most work in the SDLC. This shortens project time frames and ensures high quality data.

Technology

Full use of RDM/metadata tooling of enterprise assets is integrated with other EIM capabilities and tools. Leverage Information asset pack spreadsheets that are proper for business use and that can be imported into current or future tooling post analysis. Automation of data catalog sweeps include ad hoc and scheduled inventory of system reference codes, verifying what exists in a system. Automation identifies changed, new, or deleted reference data.

Data

Automation begins to identify changed, new, or deleted reference data in a system. Enterprise standard code sets are established and have mappings to existing system code sets. Where possible, these are the same, but they are kept separate to adapt to inevitable changes in data and business needs. Reference data vendor management is well established with standard evergreening processes to keep the standards current, registered, inventoried, versioned, and managed. Semantic mapping data ties system code sets to central conceptual hub of "Conceptual Reference Domains" in the "Enterprise Model." Mappings enable metadata-driven integration of key use cases, improving quality.

Pain

Pain is lessening as the standards and mappings become stable information assets, though there is a lot of work to do to capture and map all the information assets the first time. Pain is felt when vendors push back on the automation to catalog and profile their systems. There will be push back on staffing levels but these will be mitigated by the more efficient process.

Business Costs

Business costs are reducing. As knowledge is captured in the projects, the effort and time to project delivery is shortened.

Next Steps

- Shift from project-driven RDM to RDM fully integrated into other EIM processes.
- Automate Enterprise Standard Code Sets into translation and transformation processes of "Integration and Interoperability."
- Ensure business semantics of term management model management are simultaneously handled during harmonization.
- Establish metadata driven automation and automate business services using RDM and metadata assets.
- Make available reference code sets through automated publication, subscription, conformance, and lookup services.
- Establish more integrated stewardship processes between business and system stewards.
- Implement data profiling, operational inventory, and validation processes as part of overall information quality processes. Feedback issues and review through stewards to fill gaps and fix mistakes in code sets.
- Begin to treat relationship metadata as a first-order citizen in EIM including the permissible value, profiled usage, and semantic mapping relationships of reference data code sets.
- Leverage the project SDLC to focus change efforts in the enterprise information assets but, where possible for RDM, leverage operational processes to distribute them.

DETAILED RDM MATURITY LEVELS

5 = Optimizing

5-Optimizing— EIM Integrated RDM

At this stage, most of the information assets have become catalogued, harmonized, and put into EIM usage. RDM has been fully integrated into EIM capabilities. The relationships and knowledge have been explicitly captured in the "Asset Management" repositories and applications. Higher order functions (such as automated, metameta data driven integration and automated distribution of reference code sets) are in use. The distribution ensures a more consistent business meaning across all systems, semantic integrity across integration between systems, and, in the end, effective interoperability.

Data quality operations capture actual system usage including code set inventories and code usage in the transactional data itself. Issues are identified and raised to stewardship for disposition. The central management of enterprise standard code sets is necessary and at the core of system and data interoperability. Information assets and their relationships enable clear line-of-sight from any corner of the enterprise. The standard language afforded by the enterprise model and enterprise standard code sets begins to take root in corporate discussions. Business decision makers can rely on the answers to their questions. Staff understands their business and the systems that run it. IT has become simpler and easier to manage and leverage by the enterprise.

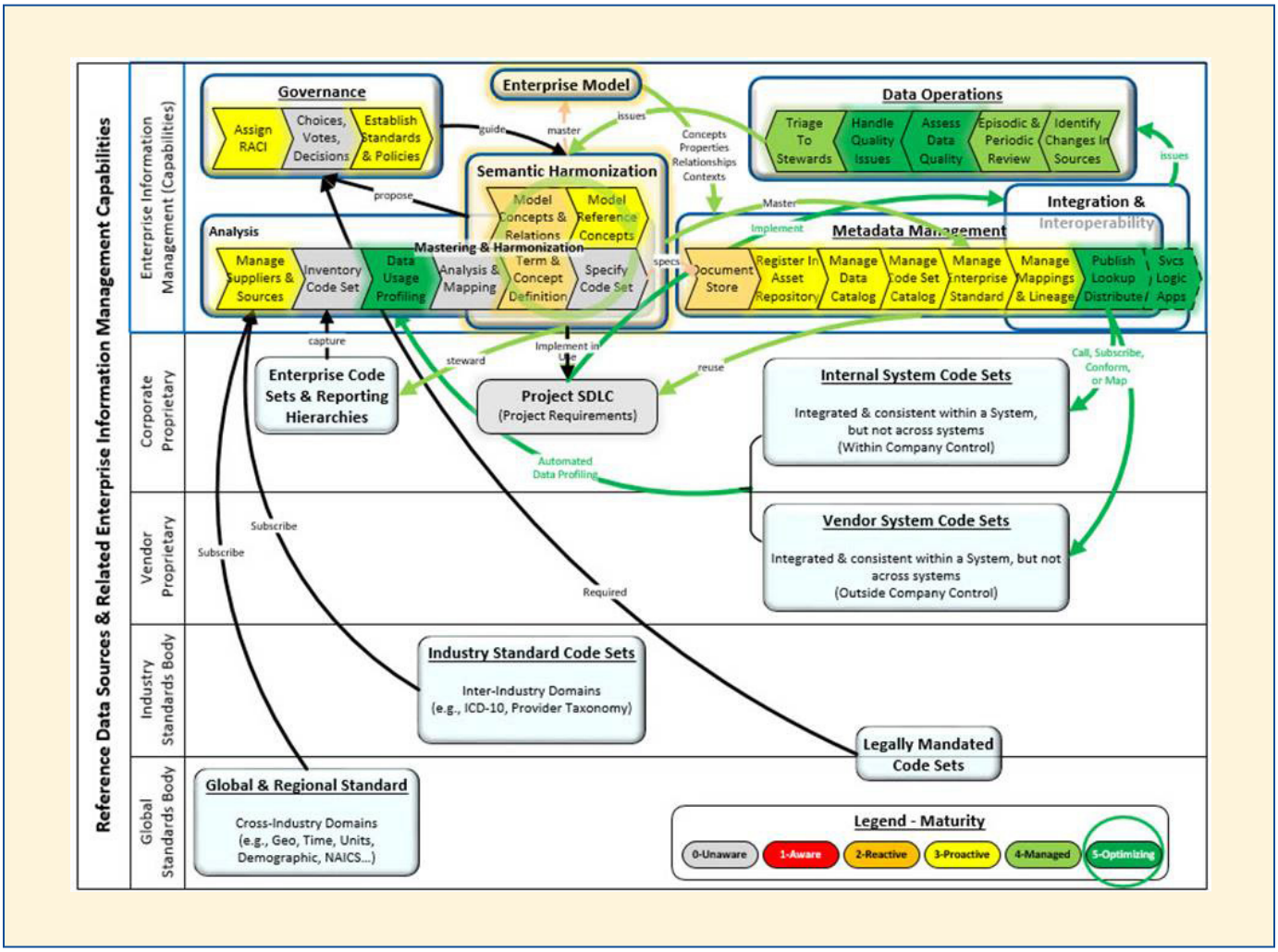


Figure 9. EIM Integrated RDM at 5 = Optimizing Maturity Level

5 = Optimizing

RESULTS: REFERENCE DATA AT 5 = OPTIMIZING MATURITY LEVEL

EIM Integrated RDM

RDM processes are complete, stable, and integrated into the broader EIM processes. Metadata-driven integrations read from the code set catalog to keep reference data complete, accurate, and up-to-date through integration processes. Categorical/reference data has generally been standardized up to enterprise standard code sets and processes are in place to evergreen and process the datasets whether they come from inside or outside the company. Mappings are durable, reliable, and used for data integration and system interoperability. Standards are evergreened within an appropriate, managed lifecycle. EIM assets are operationally managed and made available through services, automated distribution, subscription, and publication.

People

Each practitioner/role knows their "accountabilities" and executes their part in RDM. Information stewards are fully responsible for their subject area's reference data, and system stewards for semantic mapping of the system reference code sets. Business analysts rely on the metadata and reference data standards for impact analysis ahead of projects. Architects, analysts, and modelers participate with the stewards to capture and leverage knowledge about reference code sets. Enterprise code set standards are understood up and down the organizational hierarchy. People anticipate work based on the business strategy and high-level road maps reusing existing knowledge.

Process

EIM leverage information services that call into the RDM code set catalog to leverage the standards and semantic mappings for integration and interoperability. RDM fills gaps, based on anticipated needs before projects, and continues to iterate with SDLC teams as required to fill additional gaps, fix mistakes, and manage changes. Measures and processes are used to ensure quality of the enterprise code set standards and the standard semantic mappings. Vendor contracts are never made without provisions for requiring definitions of its system data and metadata, code set data, business data definitions, and vendor collaboration on mapping into enterprise semantic metadata including their reference data to the enterprise standard code sets.

Technology

RDM/metadata tooling are day-to-day necessities for practitioners and support automated operational processes. Data quality tooling becomes fully integrated. Data catalog sweeps include ad hoc and scheduled inventory of system reference codes. Instream and batch data profiling activities identify new, outlier, and improper reference data use in a system, or in the enterprise data flows, with invocation of data quality queues for disposition of the flagged issue.

Data

Fully managed enterprise standard code sets and system-specific reference datasets are catalogued and mapped in RDM tooling. Most reference code set data has been collected, reviewed, approved, and standardized. New/changed items are captured and moved quickly through RDM governance.

Pain

Pain is reduced. Business agility is enhanced. Project delivery time is shortened. Staffing levels shrink from the central RDM while the work becomes part of the internal business processes of information and system stewards. Pain comes with technology upgrades which can be mediated by selecting tools based on open standards.

Business Costs

Business costs are greatly reduced. At this stage, the explicitly captured knowledge becomes a valuable training asset to quickly deepen and broaden staff knowledge. This results in a reduction of overall staff head count and the adaptive redeployment of expertise to other enterprise initiatives.

Next Steps

- Continue intensive management and reuse.
- Stay ahead of projects.
- Continue metadata driven usage. Increase decoupling of the maintenance and quality assurance of reference data from application systems.
- Embrace data services.
- Practice controlled replication through conformance refresh cycles. Continue doing everything better.
- Plan to move operations to the cloud, reducing upgrade and technology maintenance burdens so the EIM processes can become the central focus and main activity.
- Continual improvement of enterprise RDM processes performance through both incremental and innovative organizational and technological changes/improvements.
- Rotate business and IT staff through EIM roles to deepen and broaden their expertise.
- Leverage information assets to automate more valuable services.
- Deepen the reporting and analysis based on the categories stored as code sets in RDM.

APPENDIX: MORE DETAILS

Having introduced RDM Maturity levels, this appendix provides some further details on some of the subjects covered in this whitepaper and related topics:

- *Table 2* provides additional detail on the contexts for enterprise reference data standards.
- *Table 3* lists the best practices of RDM presented in Malcolm Chisholm's whitepaper, ***The Foundations of Successful Reference Data Management*, (topquadrant.com/knowledge-assets/whitepapers/#MALCOLMRDMWP)** and maps them to the RDM maturity levels presented in this whitepaper.
- The section: ***Data, Categories, Code Sets, and Reference Data***, broadens the discussion and definition of "Code Sets" as they are treated in this whitepaper.

TABLE 2: Selections Context for Enterprise Reference Data Standards
(Higher contexts in the list have more strategic value and significance than those lower.)

CONTEXT	EXAMPLE	PROPERTIES
GLOBAL	• Units • Geography • Time • Currency	• Well analyzed, conceived, and managed by international standards bodies • Enables cross-enterprise, cross-system interoperability
REGIONAL	• North American Industrial Classification Codes (NAICS)	• Well analyzed, conceived, and managed by Regional Standards bodies but adopted via law or practice • Enables cross-system interoperability within a region • Outside corporate control
INDUSTRY	• Health: ICD-9, ICD-10, SNOMED, HL-7 • Financial: FIBO • Oil&Gas: ISO 15926	• Well analyzed, conceived, and managed • Enables cross-industry, cross-system interoperability • Mandated by law or practice • Outside corporate control
ENTERPRISE	• Line-of-business • Market segment • Chart-of-accounts • Business system partition • Custom code sets classifying standard code sets	• Established by business units within an enterprise for business reporting to measure effectiveness • Enables cross-team, cross-enterprise agreement on the categories to measure success • Mandated by policy or executive guidance • Within corporate control
VENDOR SYSTEM	• Vendor-chosen code sets	• Vendor-created or chosen code sets to operate within its application to ensure a consistent system vocabulary • Useful when trading partners use the same system • Outside corporate control
CUSTOM CORPORATE SYSTEM	• Corporate defined code sets	• Corporate proprietary, custom-created or chosen code sets to operate within its application to ensure a consistent system vocabulary to use in logic and reporting • Within corporate control • Seldomly managed across systems
LOCAL BUSINESS UNIT	• Code sets created or adopted by a business unit to manage its operations — usually these apply only locally	• Often these may overlap and/or conflict in terminology, representation, and/or meaning between local business units • Often, the short-hand name is used in local context (e.g., revenue) but means different things to business units • Often these change over time as people move from one unit or company to another resulting in apples-to-oranges confusion
SDLC CREATED CODE SETS	• Any code set can be added quickly this way but has immediate costs in the next project and in integration downstream	• Ad hoc, project-based team will often create a code set in order to deliver a project • Existing code sets, standards, or future project needs are often not considered, resulting in systemic downstream complexity, increasing integration cost and delivery time, and loss in business agility

TABLE 3: Minimal Best Practices of Reference Data Management

BEST PRACTICE	DESCRIPTION	MATURITY LEVEL
ESTABLISHING A CENTRAL REFERENCE DATA UNIT (RDU)	• A central RDU oversees RDM across the enterprise to achieve overall goals — especially standardization, quality, and operational efficiency.	• Ends Reactive • Begins Proactive
LOCATING THE CENTRAL REFERENCE DATA UNIT	• It needs to be decided where in the organization a central RDU is located. Ideally it will be close to similar functions such as data governance and master data management, perhaps in the office of the Chief Data Officer (CDO). It is less desirable to locate it in IT or in areas responsible for business intelligence.	• End of Reactive
MANAGING EXTERNAL REFERENCE DATA	• External reference data is maintained by authorities outside the enterprise. It needs to be discovered, selected, understood, and ingested. Standard practices are a major help in doing this.	• Proactive *(easiest to tackle)*
MANAGING SUBSCRIPTIONS FOR EXTERNAL REFERENCE DATA	• Once external reference data has been set up, it needs to be kept current. Subscription management does this by ensuring that changes are detected and assimilated as rapidly as possible.	• Managed
GOVERNING INTERNAL REFERENCE DATA	• Internal reference data is for business concepts that are completely specific to the enterprise. It requires a federated approach, because it is created and managed by many different SMEs. The central RDU must ensure that groups accountable for internal reference data use a standardized approach.	• Proactive *(project priority)* • Managed *(before needed)*
ESTABLISHING AND MANAGING ENTERPRISE STANDARD CODE SETS	• Begins harmonization across systems and standards with strategic choice of what to work on based on contextual scope, usefulness, project priority, and other factors.	• Proactive
PROFILING REFERENCE DATA QUALITY IN USAGE	• Profiles actual data where code values are used to categorize it to see if new, incorrect, invalid, or other non-permissible values are in actual use. Errors result in data quality triage to disposition activities which may include a cycle of analysis and management in RDM.	• Optimizing
GOVERNING REFERENCE DATA IN OPERATIONAL ENVIRONMENTS	• Operational units are challenged by changes to the business that often require rapid changes to reference data in application systems. This can create discrepancies and inconsistencies, and the central RDU must find ways to deal with local needs for change without creating difficulties at the enterprise level.	• Optimizing
DISTRIBUTION OF REFERENCE DATA	• Reference data is used widely throughout the enterprise. It is vital that all applications have synchronized copies, so distribution must be addressed. This requires a variety of approaches ranging from the fully automated to the fully manual. However, these approaches must be chosen carefully to maintain operational efficiency.	• Optimizing

DATA, CATEGORIES, CODE SETS, AND REFERENCE DATA

Unlike many information assets, reference data, or more broadly "code sets," act as a common language within systems. Figure 10 is a modified version of a figure by Malcolm Chisholm categorizing types of data.

Since reference data classifies other types of data, it represents "categorical data" generically called "code sets" because they usually have terse codes to represent them in system usage. Thus, we see the broader "code set management" as this essential capability for managing reference data. Since other kinds of "categorical data" also get identified during analysis, design, and implementation activities, RDM, especially if it has a modeling function associated with it, enables efficient capture and management of any kind of code set. For example, the work of RDM includes "enterprise structure data" (the ways the business categorizes information), rule tables with codes and description properties, etc. "Transaction structure data" (often referred to as "master data") depending on industry, may have parts of it that are "code sets." In fact, a robust RDM solution can handle many of these types of code sets well, eliminating extra silos of RDM that are hard to manage because they are in yet a different system silo.

In this whitepaper, we broaden RDM to be the management of "code sets" including any data sets which have "Code values" as part of their representation. Even if they fall within the other categories of data. This is practical, particularly early in maturing RDM because these different types of data are usually managed as "reference data" by practitioners. As RDM matures, and the subsequent experience and knowledge of practitioners deepens, these distinctions are understood and adjustments made, as needed and useful, to various aspects of EIM practices to

TABLE 4: Clarifying Reference Data Management Terminology

TERM	DEFINITION	EXAMPLE
CODE SET	• Any data set where the meaning is represented by a "code value."	• An implemented code set a system, or a standard.
CONCEPTUAL REFERENCE DOMAIN	• The modeled conceptual entity that represents a concept for which there is a list of permissible meanings (domain), usually implemented as code values in a reference table.	• This is modeled in the enterprise model with its relationships to other conceptual entities. This forms the basis for an "enterprise standard code set."
ENTERPRISE STANDARD CODE SET	• A governed code set standardized by adoption of existing standards, system-implemented code sets, enterprise-level creation through harmonization and managed as a standard. • These are always modeled in the enterprise model, have a "mastered" code value, preferred term, business definition and description, effective date, and valid through date.	• These are carefully crafted with an intent to have one value for one concept (e.g., M for Male) and no overlapping values (e.g., Both), or mixed hierarchical levels. These are used to map business meaning to implemented code set values in systems to enable semantic integration between systems.

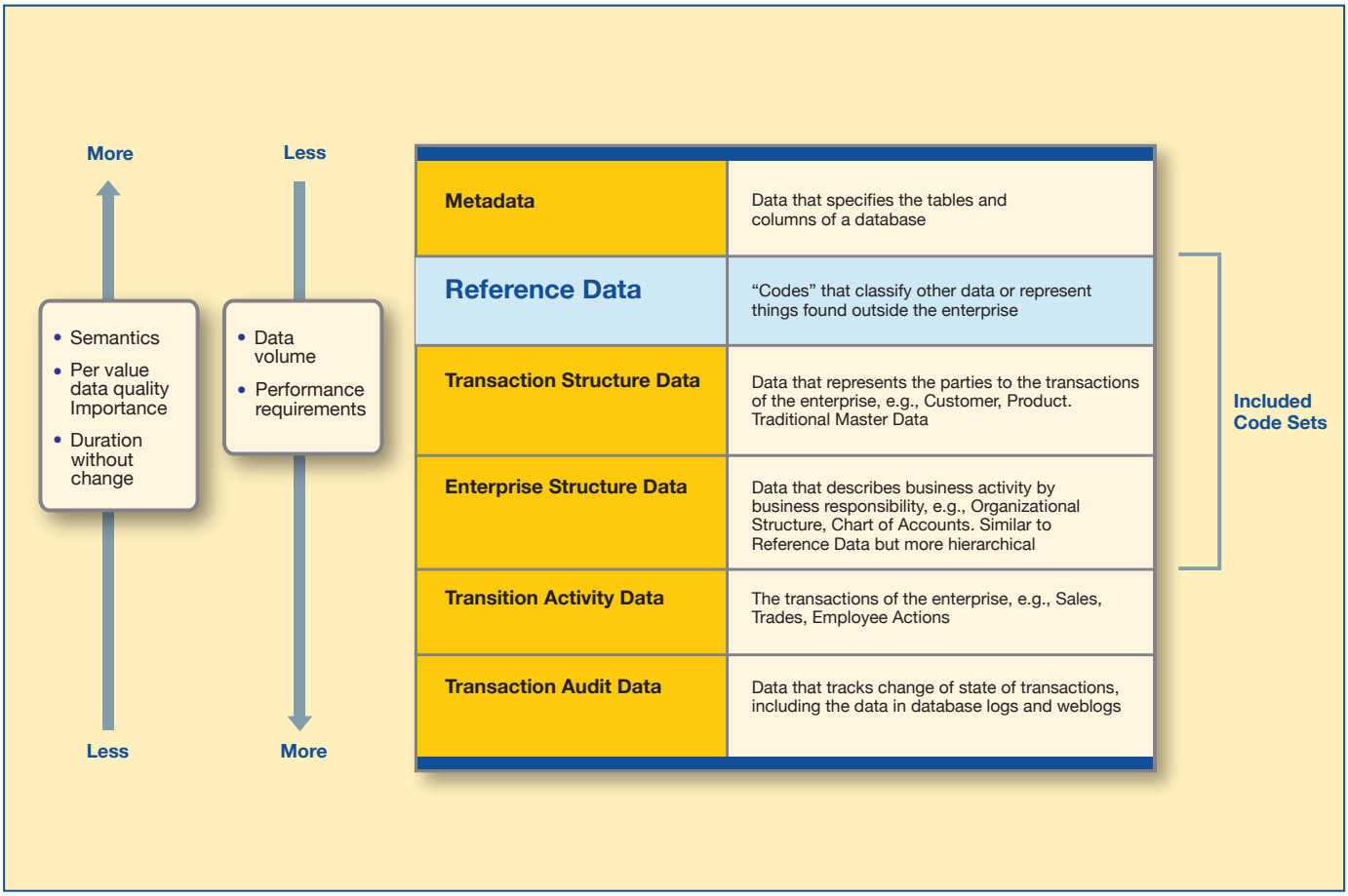


Figure 10. Reference Data in the broad sense of categorical code sets

handle the distinctions. Usually this entails ownership at different levels of organizational chart responsibility and not how they are registered and managed.

For example, "enterprise structure data" such as "line-of-business" or "market segment" or "business type" are often adjusted to measure and report business progress. When the business changes, business management changes these hierarchies. Often, these are not logical, but organized for reporting purposes and require business rules to assign them. But the "code value," "preferred term," "business definition" and other attributes still need to be managed and versioned carefully. The difference is in the rate of change, the context of use, the stewardship of creation and management, and other factors for which RDM needs to enable.

Enterprise Standard Code Sets

As an enterprise matures, code sets are brought under management, incrementally through projects, and standardized strategically by adopting and managing code sets from the broadest context to the most local context into managed "enterprise standard code sets." These are enterprise-level, business vocabularies based on "meanings" that factors in business requirements, abstraction level, coverage scope, legal mandates, cost/benefit, and reuse value to prioritize work on them. At a certain point to understand, strategically, code sets are adopted and managed based on broadest scope from global, regional, and industry standards first and then local system and business unit standards to more efficiently bring them under operational management.

For example, the product subject area, usually identifies a "product" and the many related categorizations in the form of codes, names, and descriptions. In some cases, these are explicitly managed as reference data. In other cases, they may be under master data management. In health insurance, for example, the product subject area has a constellation of categories with hierarchical and other relationships. In an enterprise, any type of data that has a "code" and "label" (aka "description"), is often managed by the same teams. This means that different types of "categorical data" must come together usefully in broader context of code set management and these will become a part of the broader activity of metadata management. But here, we call it "RDM."

Mastering Enterprise Standard Code Sets

An enterprise's span of control of reference data is only within the boundaries of the enterprise. This requires a business-level choice as to what source of reference data should be established as the standard. This then allows for it to be managed as one "enterprise standard code set" in the enterprise context. Where there are copies, say in each system, they all map to the one central greatly simplified integration between systems and prevents crippling complexity from overtaking the enterprise. It establishes the choice and usage of enterprise standard code sets. Governance of these code sets is based on practices that include assessing the correct level and source of a standard code set. These typically follow a strategic ordering of the selection to obtain the broadest and most durable value in the choice. For example, the broader the context the easier it is to achieve cross-boundary interoperability. The activity of "mastering" enterprise standard code sets including mastered code values, preferred terms, definitions, and descriptions is an enterprise-level governance activity that asserts that at an enterprise-level we will use them consistently in our "enterprise view" of information.

Enterprise View of Information

1. When possible we use the code value, preferred term, and definition in our systems, reports, analytics, business services, and common data exchange formats (CF).
2. When we can't use them, we map them to the vendor or legacy system through the data exchange formats as a standard language to achieve interoperability based on meaning.
3. Cross-discipline harmonization of new, existing, or changing code sets is achieved through collaboration and accountability between business stewards, system stewards, operational data quality analysts, project teams, appropriate SMEs, information architects, application architects, enterprise modelers (knowledge engineers, data modelers), and reference data stewards.
4. Once established as enterprise assets, everyone begins with, uses, and manages the centralized, enterprise standard code sets and their conceptual relationships into the standard enterprise models.

ABOUT THE AUTHORS

David Chasteen
Enterprise Ecologist

David Chasteen has worked in information-centric roles for over 30 years, primarily within large enterprises (e.g., health care, government, software companies, consulting companies, and as a scientist). Unlike a consultant or contractor, David has had long-term enterprise careers. As a result, he has been able to experience, grow, and iterate on Enterprise Information Management practices from conception to operation. David has led enterprise information modeling, metadata and reference data management practices and been an active participant in most other information-oriented roles and practices.

In his long career, David has wrestled through the aftermath of successful and failed projects, internal politics, consulting engagements, contractor losses, internal reorganizations, management fads, IT fads, the (wrong) belief that staff are fungible/ interchangeable, bad and good management, no funding, too much funding/no staff, bait and switch consulting, inadequate vendor contracts, knowledge loss from turnover, waves of retirements, and cloud-sourcing. He is proud of being able to keep the efforts alive through them all. Like other dedicated staff, he cared enough and maintained the fundamental belief that information assets are not optional, they are foundational. And that developing deep and wide skill in the various practices pays great dividends.

Today, David has retired into a writing and mentoring life. He is looking forward to sharing what he knows with other passionate practitioners.

Irene Polikoff
Co-founder / CEO, TopQuadrant

Irene Polikoff has more than two decades of experience in software development, management, consulting and strategic planning. Since cofounding TopQuadrant in 2001, Irene has been involved in many information management projects across a broad range of customer organizations in different industries. She has written strategy papers, trained customers on the use of the semantic technology standards, developed ontology models, designed solution architectures, and defined best practices. With the introduction of TopQuadrant's Semantic Solution Platform and TopBraid Suite, Irene has been responsible for TopQuadrant's product strategy and its data governance vision.

Before starting TopQuadrant, Irene was a principal and a senior development manager at IBM Global Services. Prior to that, Irene held IT management positions at Fortune 500 companies where she was responsible for development and deployment of enterprise-wide mission critical information systems.

Irene is a coauthor of a book on software requirements and architecture — *Capability Cases: Solution Envisioning Approach*, author of many published articles and whitepapers as well as a frequent speaker at conferences. Irene has been actively involved in technology standardization. She currently serves as a co-chair of the W3C SHACL working group that has developed a standard for describing and validating RDF data.

CONCLUSION

There are many activities and challenges to build out a successful reference data management practice.

This paper presents a Reference Data Management (RDM) maturity model based on aligning RDM practices with key functions of Enterprise Information Management (EIM). It shows a practitioner's view of a likely path into maturing RDM. Ideally, now you have some tools to:

- Assess your level of RDM maturity
- Decide what level of maturity you want to be at and why
- Understand what steps you need to take to move up the maturity ladder
- Understand the relationships of RDM as it coevolves with other EIM functions

We hope this practitioner-informed road map helps you to adopt your own RDM and EIM maturation efforts.

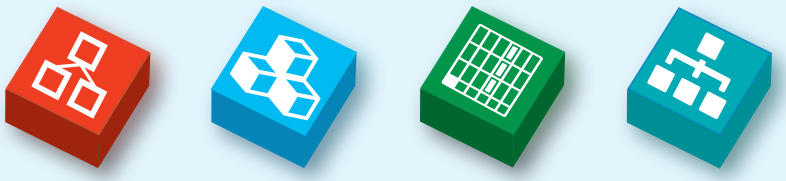
To find out how TopBraid Enterprise Data Governance (EDG) provides the capabilities needed to help you mature your Data Governance and RDM practices, we invite you to download this companion whitepaper: *Maturing Information Governance with TopBraid EDG*, (topquadrant.com/docs/whitepapers/Maturing-Information-Governance-with-TopBraid-EDG.pdf).

About TopQuadrant

TopQuadrant helps organizations succeed in data governance. Its flagship product, TopBraid EDG, delivers easy and meaningful access for all data stakeholders to enterprise metadata, business terms, reference data, data and application catalogs, data lineage, requirements, policies, and processes.

TopQuadrant's customer list includes organizations in financial services, pharma, healthcare, digital media, government and other sectors.

Governance Packages Available in TopBraid EDG



Vocabulary Management | Metadata Management | Reference Data Management | Business Glossaries

TopBraid EDG-MM is part of a comprehensive information management and governance environment where packages for other types of assets can be easily added if needed. In ramping up a data governance program, different organizations may have different starting points. With TopBraid EDG, you can start incrementally and add capabilities as you go. For details on available EDG packages and additional modules visit ***topquadrant.com/ products/topbraid-edg-gov-packs/***

For more detail or to schedule a demo, contact us at: edg-info@topquadrant.com