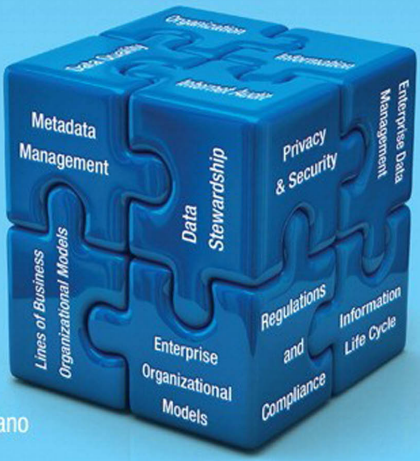
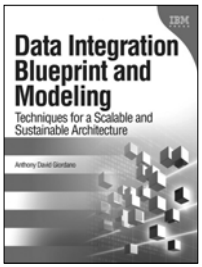
Performing Information Governance

A Step-by-Step Guide to Making Information Governance Work



Anthony David Giordano

Related Books of Interest



Data Integration Blueprint and Modeling

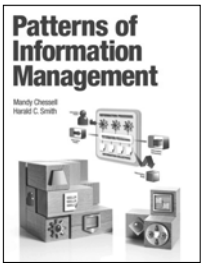
Techniques for a Scalable and Sustainable Architecture

By Anthony David Giordano

ISBN-13: 9780137084937

Making Data Integration Work: How to Systematically Reduce Cost, Improve Quality, and Enhance Effectiveness

As businesses increasingly rely on analytics, the need for a blueprint for data integration is increasing now more than ever. This book presents the solution: a clear, consistent approach to defining, designing, and building data integration components to reduce cost, simplify management, enhance quality, and improve effectiveness. Leading IBM data management expert Tony Giordano brings together best practices for architecture, design, and methodology, and shows how to do the disciplined work of getting data integration right.



Patterns of Information Management

By Mandy Chessell, Harald Smith

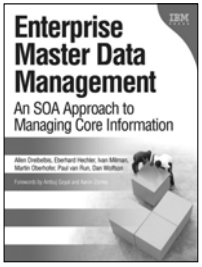
ISBN-13: 9780133155501

Use Best Practice Patterns to Understand and Architect Manageable, Efficient Information Supply Chains That Help You Leverage All Your Data and Knowledge

In the era of "Big Data," information pervades every aspect of the organization. Therefore, architecting and managing it is a multi-disciplinary task. Now, two pioneering IBM® architects present proven architecture patterns that fully reflect this reality. Using their pattern language, you can accurately characterize the information issues associated with your own systems, and design solutions that succeed over both the short- and long-term.

Building on the analogy of a supply chain, Mandy Chessell and Harald C. Smith explain how information can be transformed, enriched, reconciled, redistributed, and utilized in even the most complex environments. Through a realistic, end-to-end case study, they help you blend overlapping information management, SOA, and BPM technologies that are often viewed as competitive.

Related Books of Interest



Enterprise Master Data Management

An SOA Approach to Managing Core Information

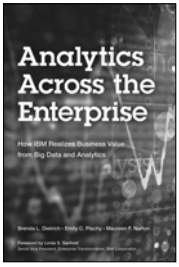
By Allen Dreibelbis, Eberhard Hechler, Ivan Milman, Martin Oberhofer, Paul Van Run, Dan Wolfson

ISBN-13: 9780132366250

The Only Complete Technical Primer for MDM Planners, Architects, and Implementers

Enterprise Master Data Management provides an authoritative, vendor-independent MDM technical reference for practitioners: architects, technical analysts, consultants, solution designers, and senior IT decision makers. Written by the IBM data management innovators who are pioneering MDM, this book systematically introduces MDM's key concepts and technical themes, explains its business case, and illuminates how it interrelates with and enables SOA.

Drawing on their experience with cutting-edge projects, the authors introduce MDM patterns, blueprints, solutions, and best practices published nowhere else—everything you need to establish a consistent, manageable set of master data, and use it for competitive advantage.



Analytics Across the Enterprise

How IBM Realizes Business Value from Big Data and Analytics

By Brenda L. Dietrich, Emily C. Plachy, Maureen F. Norton

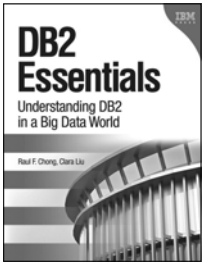
ISBN-13: 9780133833034

How to Transform Your Organization with Analytics: Insider Lessons from IBM's Pioneering Experience

Analytics Across the Enterprise demystifies your analytics journey by showing you how IBM has successfully leveraged analytics across the enterprise, worldwide. Three of IBM's pioneering analytics practitioners share invaluable real-world perspectives on what does and doesn't work and how you can start or accelerate your own transformation. This book provides an essential framework for becoming a smarter enterprise and shows through 31 case studies how IBM has derived value from analytics throughout its business.

Whatever your industry or role, whether a current or future leader, analytics can make you smarter and more competitive. *Analytics Across the Enterprise* shows how IBM did it—and how you can, too.

Related Books of Interest



DB2 Essentials

Understanding DB2 in a Big Data World, 3rd Edition

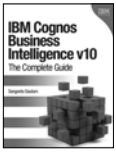
By Raul F. Chong, Clara Liul

ISBN-13: 9780133461909

The Easy, Visual Introduction to IBM DB2 Version 10.5 for Linux, UNIX, and Windows

DB2 Essentials covers everything you need to get productive with the latest version of IBM DB2 and apply it to today's business challenges. It discusses key features introduced in DB2 Versions 10.5, 10.1, and 9.7, including improvements in manageability, integration, security, Big Data support, BLU Acceleration, and cloud computing.

Raul F. Chong and Clara Liu illuminate key concepts with examples drawn from their extensive experience with DB2 in enterprise environments. They explain how DB2 has evolved, what's new, and how to choose the right products, editions, and tools and then walk you through installation, configuration, security, data access, remote connectivity, and day-to-day administration.

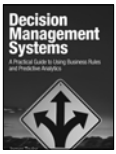


IBM Cognos Business Intelligence v10

The Complete Guide

By Sangeeta Gautam

ISBN-13: 9780132724722

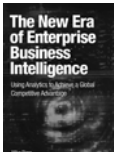


Decision Management Systems

A Practical Guide to Using Business Rules and Predictive Analytics

By James Taylor

ISBN-13: 9780132884389

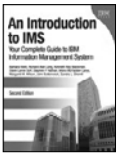


The New Era of Enterprise Business Intelligence

Using Analytics to Achieve a Global Competitive Advantage

By Mike Biere

ISBN-13: 9780137075423



An Introduction to IMS

Your Complete Guide to IBM Information Management System

By Barbara Klein, Richard Alan Long, Kenneth Ray Blackman, Diane Lynne Goff, Stephen Paul Nathan, Moira McFadden Lanyi, Margaret M. Wilson, John Butterweck, Sandra L. Sherrill

ISBN-13: 9780132886871

IBM Cognos 10 Report Studio

Practical Examples

By Filip Draskovic, Roger Johnson

ISBN-13: 9780132656757

This page intentionally left blank

Performing Information Governance

This page intentionally left blank

IBM PRESS

Performing Information Governance:

A Step-by-Step Guide to Making Information Governance Work

Anthony David Giordano, PMP, CCP

IBM Press Pearson plc
Upper Saddle River, NJ • Boston • Indianapolis • San Francisco
New York • Toronto • Montreal • London • Munich • Paris • Madrid
Cape Town • Sydney • Tokyo • Singapore • Mexico City

ibmpressbooks.com

IBM Press Program Managers: Steven M. Stansel, Ellice Uffer

Cover design: IBM Corporation

Executive Editor: Bernard Goodwin

Marketing Manager: Stephane Nakib

Publicist: Heather Fox

Development Editor: Michael Thurston

Managing Editor: Kristy Hart

Designer: Alan Clements

Project Editor: Andy Beaster

Copy Editor: Keith Cline

Indexer: Heather McNeill

Senior Compositor: Gloria Schurick

Proofreader: Sarah Kearns

Manufacturing Buyer: Dan Uhrig

Published by Pearson plc
Publishing as IBM Press

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact international@pearsoned.com.

Library of Congress Control Number: 2014943598

ISBN-13: 978-0-13-338565-6
ISBN-10: 0-13-338565-5

Text printed in the United States on recycled paper at R.R. Donnelley in Crawfordsville, Indiana.

First printing: September 2014

Contents

Part III: Performing Ongoing Information Governance Processes

Preface

This text provides a practitioner's guide on how to apply information governance principles and processes in both projects and ongoing enterprise information management (EIM) operations. Information governance has been recognized as a critical organizational process for both information management best practices and regulatory compliance. However, of all the EIM functions, information governance cannot be performed successfully as a standalone discipline. Information governance when practiced in a vacuum has not brought the value expected and intended. Although there are many sources for why information governance is important, few go into detail as to how to, step by step, execute information governance processes in both development projects and in ongoing organization processes.

This book introduces you to the core components of information governance and how they "thread" into the various functions of EIM. It also covers in detail how to pragmatically and practically execute information governance functions on development projects and in on-going organizational processes.

Intended Audience

This text serves many different audiences. Experienced data management professionals can use it to confirm the activities, tasks, and best practices for performing information governance. College students can use this book as a textbook in an upper-level information management college curriculum. The intended audience includes the following:

- Chief information officers
- Chief data officers
- Business and technical data stewards
- Data quality analysts and auditors
- Metadata management professionals
- Master data management professionals

- Program/project managers
- Data modeling and database practitioners
- Other information management practitioners
- Information management-focused college students

Scope of the Text

This book stresses the implementation and operational aspects of implementing information governance using the six core components of information governance and how they thread into other EIM functions such as transactional processing, business intelligence (BI), and master data management (MDM).

With that goal in mind, this book:

- Reviews the functions of EIM and the components of information governance
- Provides a step-by-step approach to performing project-level information governance activities within each of the EIM functions
- Provides a step-by-step approach to ongoing information governance activities within the organization
- Reviews case studies for each of the project-level and operational information governance activities

Organization of the Text

The text is organized into three parts:

- **Part I: Overview of Information Governance & Enterprise Information Management**

  The first part of this text provides an overview of EIM. Because of the broad nature of information governance and its potential implications in all facets of information management, it is important to understand the core functions within the scope of EIM, such as transactional processing, BI, and MDM, as well as the emergence of "big data." Then to clarify how information governance is a "horizontal thread" to the other EIM functions, the core components of information governance are defined.

- **Part II: Performing Information Governance Activities in Enterprise Information Management Projects**

  The second part of the text covers the systems development life cycle (SDLC) activities of an information governance project in terms of the activities, tasks, and deliverables with each of the three EIM functions. For example, it covers the data stewardship, metadata management, and data quality tasks in the definition, design, and deployment of a data warehouse within a BI environment. It also covers transactional processing and MDM. It concludes with thoughts on big data and the information life cycle management (ILM).

- **Part III: Performing Ongoing Information Governance Processes**

  The third part of this text starts with a discussion on the evolution of information governance organization models, optimal models for how to align information governance organizations within the broader information, and the typical roles in an information governance organization. It then covers the roles of data stewardship with business and technical organizations. Next, it reviews ongoing data quality processes in terms of the types of ongoing data quality projects. It concludes by reviewing ongoing metadata management functions.

Also in this text, each chapter provides a set of questions on the core concepts in the book to test your comprehension of the material. You can find answers to the questions for each chapter in Appendix B, "Chapter Exercise Answers."

Much of the supporting materials to the text can be found in the appendices, which include the following:

- **Appendix A, "Organizational Change Management in Information Governance"**—This appendix discusses techniques on how to overcome the organizational challenges of implementing an information governance process and organization.
- **Appendix B, "Chapter Exercise Answers"**—This appendix contains answers to the questions found at the end of each chapter.
- **Appendix C, "Enterprise Information Management Systems Development Life Cycles"**—This contains the SDLCs and the example of how they have been applied from the chapters.
- **Appendix D, "Glossary of Information Governance Terms"**—This appendix contains the glossary of terms used in the book.
- **Appendix E, "Case Study Models"**—This appendix reviews the information governance case studies found throughout the text.

Acknowledgments

Information governance is a broad and pervasive information management function with many perspectives and points of view. The processes defined in this book are a result of actual project work that is a product of countless hours defining processes, testing processes, and using these processes in many organizations. These efforts can only be performed in collaboration with knowledgeable, dedicated, and experience practitioners. In particular, I want to acknowledge Glenn Finch, Sharon Hodgson, Ron Shelby, Linc Markham, Brian Scheld, Dan Sutherland, and Brett Gow, all of whom played an integral part in the development of this book.

Dedication

I want to dedicate this book to my daughters, Katie and Kelsie; they are my greatest joy and blessing.

Also, I have to thank my two Irish wolfhounds, Rupert and Switters, who kept me company during the long hours writing this book. They are truly man's best friends.

About the Author

Anthony Giordano is an experienced executive with more than 25 years of global professional services experience in the information management field in the areas of business intelligence, data warehousing, customer relationship management, and program management. He has significant work experience in the financial services, life sciences, and automotive sectors. Mr. Giordano has functioned in various roles, including positions such as IBM Japan BAO Service Line Leader, Financial Services Practice Partner, and other consulting roles. He has extensive international experience in Australia, Thailand, Turkey, England, and living in Japan. Tony has been dedicated to implementing customer-focused business analytics and information management platform that deliver business outcomes.

In his role at Merkle, Tony is responsible for the Technology Solutions Group, which defines designs, develops, and deploys the Foundation Marketing Platform's that Merkle's customer run their multi-channel campaigns and marketing analytics.

Tony joins Merkle from IBM's Business Analytics and Optimization Practice, where he spent 18 years in a variety of senior-level positions. Most recently, Tony led IBM's BAO Japan Service Line. He also held IBM's BAO Global Leader for Banking and Financial Markets, Enterprise Information Management Service Line for North America, and Financial Services BAO Partner roles. Prior to IBM and PricewaterhouseCooper's, Tony held consulting roles in Oracle.

Tony holds a Master of Business Administration degree from Wayne State University in Detroit, majoring in accounting, and a Bachelor of Business from Walsh College, majoring in Computer Information Systems.

Tony is the author of *Data Integration Blueprint and Modeling* (IBM Press).

Introduction: Why Another Book on Information Governance?

Information governance is not a new discipline; in fact, it has been an acknowledged practice going back to the 1980s. Every 2 to 3 years, there is a renewed interest in information governance due to a new technology or set of government regulations. It has evolved in the past 20 years from multiple areas of the business enterprise. For example, data stewardship and data quality evolved from data management in areas such as data modeling, where the need to define the business and technical metadata is required to understand the optimal design blueprint for the target data model.

The need to redefine how the enterprise's data is protected is a result of the ever-increasing blur between the enterprise and the external world. This has developed many of the security and privacy aspects of information governance—resulting in a broad discipline with many interfaces, control points, and processes. Figure I.1 shows how this involves both business and technology.

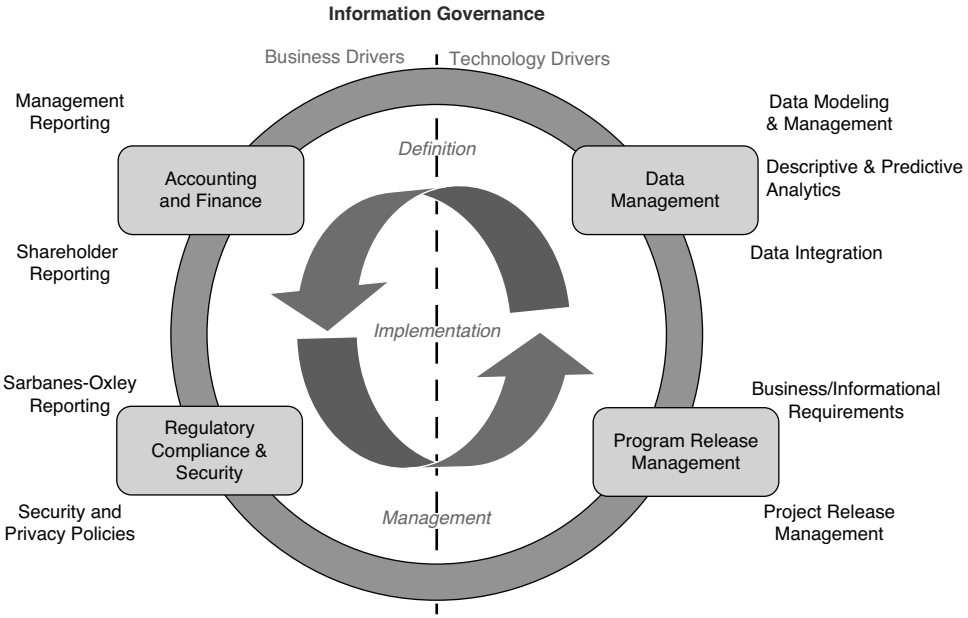


Figure I.1 The business and technology drivers for information governance

What Is Not Working in Information Governance

Many enterprises have started programs or projects covering all or some aspects of information governance, but many others have also abandoned their efforts, frustrated with their inability to sustain the implementation and achieve results. Enterprises fail in their information governance efforts for many reasons, including a lack of

- **Clarity of purpose—**Many organizations have bought into the "what" of information governance but have struggled to understand the "why." Success, in some companies, may be defined as implementing information governance processes. Here, companies may struggle to rigorously implement every component of a process without recognizing how information governance processes help the organization achieve operational and performance benefits. In these cases, implementing a process is less the measure of success than is the achievement of well-defined and well-understood performance benefits of that process improvement.

- **Organizational commitment—**Some organizations believe that they need to have an information governance process or organization to meet regulatory requirements or government mandates. After fulfilling that obligation, they perceive that the need for continuing information governance no longer exists and so they discontinue/disband the process/organization. Other organizations face a change in leadership, where the new leadership has not bought into the value of having an information governance process. Or in other words, these organizations never bought into the sustained business value of information governance (the "why" of information governance).
- **Political tradeoffs and organizational change management—**Developing an information capability and competency requires the balancing of current practices against practices optimized for enterprise performance. This often requires consideration of mid- to longer-term adjustments that, although less efficient in the near term, produce more consistent and predictable results. Some of these adjustments include adjusting to new roles, learning new processes, and adapting to broader and more integrated definitions of performance. New decision-making processes are some of the more challenging adjustments in that decisions benefiting individuals and specific lines of business or functional areas are balanced against broader enterprise performance requirements.
- **Understanding how to design and coordinate enterprise information governance—**These enterprises have bought into the "why" and "how" of information governance. They have bought the books, went to the seminars, brought in the "experts" and have sold information governance to their management, staff, and stakeholders, and may have even instantiated an organization. However, they fail due to a lack of actual performance. They simply do not integrate their information governance activities within development projects, and ongoing activities create additional project and process overhead. They are invariably shut down from a cost/benefit perspective.

Providing the "How" of Information Governance

The scope of this text is not on the "why" and "what" of information governance. This book provides prescriptive instructions, a step-by-step "how" to perform information governance. It discusses how to build activities on ongoing enterprise information management projects and ongoing processes, as portrayed in Figure I.2.

I. "Why" Information Governance?

- ✓ Lost productivity due to reconciling, correcting, rationalizing, combining data
- ✓ Financial restatements due to incorrect definitions of aggregated data
- ✓ Poor data quality costs
- ✓ Fraud due to inadequate privacy and security controls

II. "What" Is Information Governance?

- ✓ Information governance organization component
- ✓ Data stewardship component
- ✓ Data quality management component
- ✓ Metadata management component
- ✓ Information governance change management component
- ✓ Privacy and security component
- ✓ Information life cycle management component

III. "How" to Implement & Perform Information Governance

- ✓ Performing information governance tasks in business intelligence
- ✓ Performing information governance tasks in master data management
- ✓ Performing information governance tasks in transactional projects
- ✓ Performing information governance tasks in information life cycle management projects
- ✓ Performing information governance tasks in "big data" projects
- ✓ Ongoing data stewardship processes
- ✓ Ongoing data quality processes
- ✓ Ongoing metadata management processes

Figure I.2 The focus on performing information governance

It covers how to integrate information governance within existing efforts rather than create duplicative activities and tasks. Duplication unnecessarily burdens projects and creates a negative environment for information governance. Instead, this book places key information governance tasks and roles with the systems development life cycle of the various enterprise information management disciplines and facilitates and hardens the development of the various information management components.

After an information governance organization has been created, the text covers how to practically sustain and audit the information environment in such a way to ensure the highest ongoing data quality, data security, and data understanding.

This page intentionally left blank

CHAPTER 1

Information Governance & EIM Overview Governance

Part I of this text provides an overview of enterprise information management (EIM) and information governance. It reviews and details the various functions in EIM, such as business intelligence, master data management, and transactional data management, with a discussion on the emerging function of "big data." It next introduces and details the seven components of information governance. Although this is not a "what" is information governance book, it is necessary to establish a foundation for what each component is and how to deploy them in both development and ongoing efforts.

Part I sets the stage for the major focus of the book: how information governance components thread into EIM projects. For example, what data stewardship tasks are necessary when building a data mart data model? What types of metadata are captured and managed in a master data management project? What are the ongoing privacy and security tasks needed for connecting your intranet to a social network? These are presented at a high level and then defined and explained at a detailed level in subsequent sections and chapters of the book.

This chapter focuses on each of the functions within EIM in terms of their purpose, unique artifacts that are used to design and maintain the applications, and how they are built. It explains how these different functions interconnect to provide an entire information environment for an organization.

What Is Enterprise Information Management?

The major focus of this book is performing information governance within the context of enterprise information management (EIM.) To understand information governance, it is important to start with an understanding of EIM area within an organization. EIM is the information management foundation for both transaction and analytic processing. It is a series of functions that are segmented based on the type of transactional, operational, and analytic processes they support (see Figure 1.1).

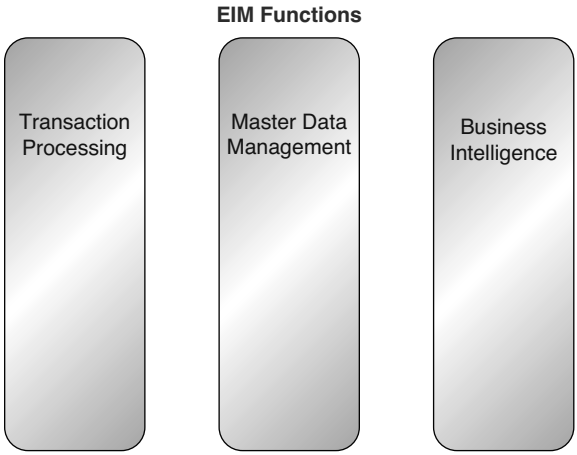


Figure 1.1 Enterprise information management functions

The Functions of EIM

These EIM functions have similar design patterns around data and provide the context for process areas such sales and marketing, finance, and production. For the purpose of this text, each is defined as follows:

- **The transaction processing function—**Centers on the creation and maintenance of the core business transactions in the business. This function is still 60% to 70% of all information technology (IT) budgets and defines what the transactions mean and how it is used within the organization.
- **The master data management (MDM) function—**Concentrates on the creation and maintenance of the core domain definitional information of an organization. It provides the context for our transactional and analytic data. For example, it provides the definition of what a customer is and what a product is. These definition and instantiated data elements are used in creating transactions and determining the measures needed to analyze what is a customer or how much of a product is used.
- **The business intelligence (BI) function—**Focuses use of data for different types of information analysis. A BI environment is the most data-centric of all EIM functions. It captures, collates, and conforms data from many disparate sources into a set of repositories in various structures for the many different types of reporting, descriptive, and predictive analytics used by disparate end users. A BI environment now offers their organizations a centralized environment to provide financial and marketing reporting and analytics.

Other authors and organizations may have different perspectives of what EIM consists of and the functional processes that it covers. This book, though, focuses on how to perform

information governance activities and tasks within the development and ongoing operations in these three EIM functions.

Data Management: EIM's Technical Development and Management Discipline

To understand how information governance interacts in EIM functions, it is important to understand how EIM functions are developed and maintained. This section discusses the technical discipline of data management. Common patterns exist in the data-driven aspects of the three EIM functions. They have similar requirements and patterns in the blueprints, development life cycles, and maintenance of the applications. Over the past 30 years, IT has evolved a technical discipline known as data management.

Data management is the development and maintenance of architectures, best practices, and procedures that manage the full data life cycle of an organization. It is within data management that data architecture artifacts such as data models, data integration models, and information access patterns are developed and maintained.

The best example of a well-known data management process is data modeling. The systems development life cycle (SDLC) details how data models capture business requirements of an organization. It determines how to best structure those requirements into the different types of technical structures that are available: transactional, operational, and analytic (data warehouse, dimensional), as shown in Figure 1.2.

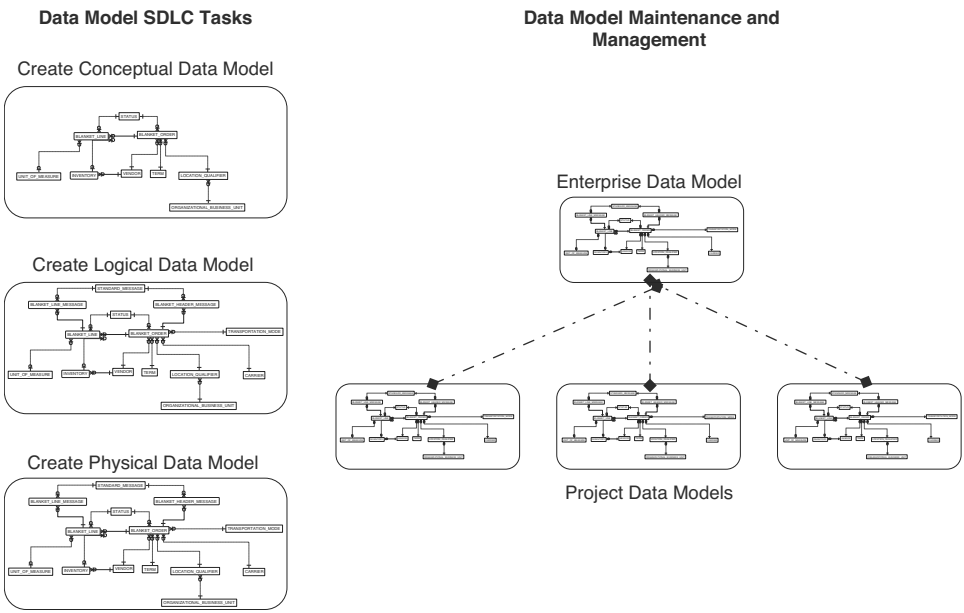


Figure 1.2 Data modeling in SDLC and maintenance tasks

The Relationship Between Data Management and Information Governance

A very tight relationship exists between data management and information governance. Often, the two are confused as the same discipline or overlap in areas such as metadata and data quality management. Within the development of the data management artifacts such as data models, there are information governance tasks such as business definitions of the entities, attributes, and relationships. Chapter 4, "Performing Information Governance Tasks in Transactional Projects," explores these relationships in much greater detail. For this section, you just need to understand that a data management artifact is the blueprint for a database or data integration process and that the information governance aspects give it business context.

What Is Information Governance?

There are many definitions and points of view on what information governance is and what it is not. For this book, the formal definition of information governance is as follows:

> Information governance is the orchestration of people, process, and technology to enable an organization to leverage data as an enterprise asset.

Although information governance spans both business and technology (as shown in Figure I.1 in the Introduction), it is truly a business function with its primary directive to establish the policies for the creation and usage of data with an organization. It is an integral aspect of the understanding of an organization, which leads to the position that information governance should be considered an ongoing organizational function on par with accounting or marketing.

The Information Governance Component Framework

Information governance is a broad discipline that encompasses the definition, creation, use, security, ownership, and deletion of all organizational data. The information governance component framework covers the organizational models and roles needed to define and manage the policies and processes that affect the business and technical creation, maintenance, and usage of data within the enterprise. These components of information governance include the following:

- Information governance organization component
- Data stewardship component
- Data quality management component
- Metadata management component
- Privacy and security component
- Information life cycle management component

Information Governance Organization Component

The information governance organizational component is the "people" aspect of the discipline. It sets the policies for information governance and maintains a staff to support those policies in managing the development and ongoing usage of corporate information. Because information governance is an organizational process similar to accounting or marketing, as mentioned earlier, it requires a staffing structure capable of performing both project and ongoing activities and tasks and fitting into the broader organization. Many information governance programs and organizations fail because of their inability to work within the corporate culture and, when necessary, modify that corporate culture. It also depends on how the organization is organized and how successful its reporting chain is.

The Information Governance Organizational Model

There are several information governance organizational models, all of which revolve around certain key functions, including the information governance council and the data stewardship community function.

The information governance council (IGC) function focuses on setting the vision and goals, providing alignment within the broader organization, and setting the direction for the information governance process. The IGC function includes establishing the policies and procedures for information governance, such as the following:

- Data as a corporate asset policy
- Data creation and usage policy
- Data security requirements policies
- Data regulatory requirements (e.g., Sarbanes-Oxley) policies
- Data quality audit policies

The IGC is traditionally organized with key stakeholders from core functional areas such as accounting, marketing, research and development, sales, and production. Often, the department leaders will chair the IGC or provide an advocate. The IGC is traditionally led by a chief data officer (CDO).

The CDO has been an evolving role over the past 15 years and originally was only responsible for the information governance organization in terms of overall direction and day-to-day oversight. In recent years, many organizations have been expanding the responsibilities of the CDO to include oversight and day-to-day management of data development and maintenance functions. The CDO role and responsibility often includes the following:

- Owning and driving the organization's data strategy and enterprise-level data vision
- "Selling" information governance, by driving data ownership and accountability in the business
- Directing data quality practices across the organization

- Aligning business and IT to support data quality through consistent business definitions and well-defined data quality ranges in transactional and analytic applications
- Providing leadership by sitting on executive committees where data programs and projects are approved and sponsored to ensure information governance processes are embedded into those programs
- Working with other business executives to understand their data quality requirements, objectives, and issues
- Providing leadership and support to members of the data stewardship community as they define data and metadata
- Working closely with the information council's business liaisons to evangelize enterprise data governance within the organization

The success of an information governance organization depends on having the right candidate at the right level in the CDO role.

The data stewardship community function focuses on implementing the information governance policies and processes. It works with the end users to define the business and technical metadata, provides the data quality measures and ranges to be managed to (and performs data quality audits), and ensures that the end users are getting as much value as possible out of the data. The next section of this chapter covers those responsibilities for data stewards in greater detail.

The data stewardship community can be tightly aligned as a group or aggregated by organizational areas, as shown in Figure 1.3.

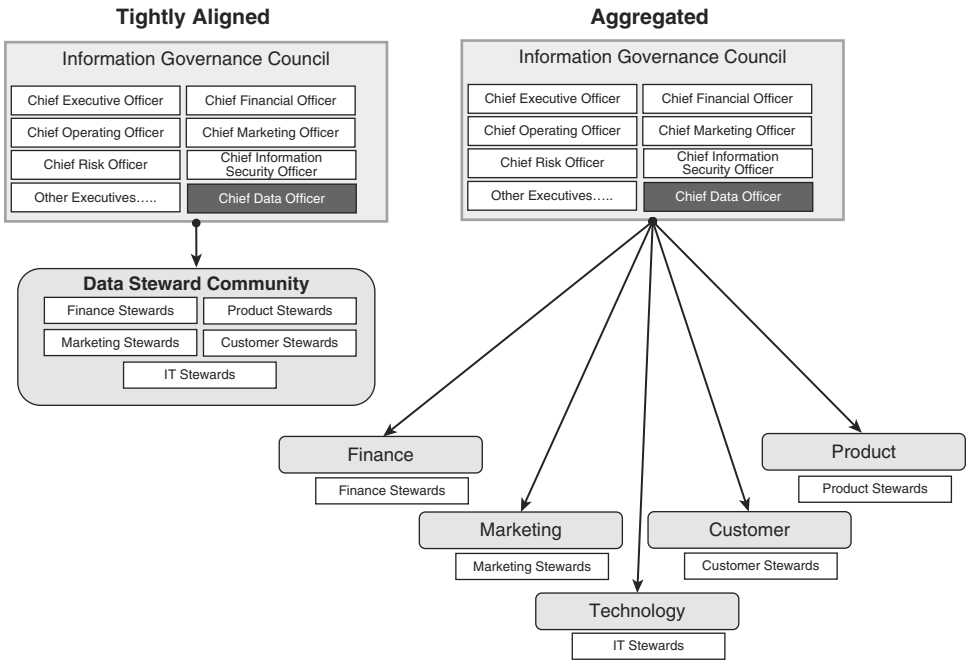


Figure 1.3 Tightly aligned versus aggregated data stewardship community alignment

Where the data stewards are aligned to the organizational area, they are often "solid" lined (direct reporting) to that organizational area, and "dotted" line (indirect reporting) to the IGC.

The Information Governance Reporting Models

The success (or failure) of information governance initiatives are often a direct result of the alignment of the information governance organization within the enterprise. There are many reporting models, but there three are typically found:

- **Aligned to the chief financial officer (CFO)—**In this reporting model, the CDO and the IGC direct report to the CFO. This model has been used in both the manufacturing and life science industry. There are many benefits to this alignment that include tying budgets to adherence to information governance standards, tight alignment to financial management reporting (business metadata management), and the usage of financial information (data security).
- **Aligned to the chief risk officer (CRO)—**This model is most prevalent in the financial services industry, where adherence to government regulatory requirements and mandates is tightly tied to the common set of data definitions and the ability to demonstrate data lineage (e.g., Sarbanes-Oxley).

- **Aligned to the chief information officer (CIO)—**In this reporting model, the CDO and the IGC direct report to the CIO. One of the advantages of reporting to the CIO is the tight alignment to the development and maintenance of the data assets within the organization. Among the disadvantages of information governance organizations aligning within the IT is that business functions tend to view those organizations as technical only and discount the importance of the discipline. This leads to issues in the enforcement of (and lack there of) information governance standards and guidelines within the business functions.

Data Stewardship Component

Data stewardship is the "people" aspect of information governance that directly interfaces with the creators and users of data. Data stewards support, maintain, and execute the policies and procedures instituted by the IGC. Data stewards are often organized in communities that are aligned either (or both) by functional areas, such as customer or product, (and) or by departmental areas, such as accounting or marketing. Most information governance tasks discussed in this text are either directly performed by or influenced by data stewards.

Typical Data Stewardship Responsibilities

A data steward's responsibilities vary widely from organization to organization based on the structure of the information governance process, the maturity of information governance within the enterprise (e.g., perceived important and authority granted to the information governance organization), and how the enterprise has organized its IT function. Typical data stewardship responsibilities are shown as follows and categorized by how data is created, organized, managed, and monitored. These responsibilities include the following:

- **Data stewardship creation responsibilities:**
  - Work with the business stakeholder and technologies in the business and technical definitions of data requirements
  - Ensure that the planned data has defined data quality criteria and ranges for critical data entities
  - Ensure that those definitions are captured and stored as metadata
  - Collaborate with IT data architects and modelers to ensure that the captured data requirements are structured correctly so that the intended business users gain the intended value
  - Collaborate with the business users and corporate security on data privacy requirements, user access control procedures, and data-retention policies

- **Data stewardship management responsibilities:**
  - Review and approve potential changes to the definitions and structures of the data, ensuring that those changes are appropriately maintained in the metadata management environment
  - Provide ongoing communications on the information governance organization, its policies, and processes
  - Assist/perform "road shows" on evangelizing information governance
  - Work with data management organizations on embedding information governance activities into ongoing processes and activities
- **Data stewardship monitoring responsibilities:**
  - Manage and communicate changes to data quality and security controls to business and technical stakeholders
  - Perform ongoing data quality and security audits on critical subject areas and application systems within the organization
  - Manage issues due to technical data quality and definitional understanding inconsistency, including data quality renovation projects

The breadth of information governance within the processes of an organization has led to the development of several types of data stewards and data stewardship-type roles. Most of these are segmented between business and technology roles, each with certain characteristics and responsibilities. The following sections provide a noncomprehensive list of the types of data stewards.

Business Data Stewards

Business data stewards focus more on the interaction of data with the executives and end users of a business function. They tend to focus on the data definition of base and aggregated data. For example, the business definition and calculation of return on net assets (RONA) can be a hotly contested definition between functional areas of an organization and a source of considerable time and effort for data stewards to develop common understandings and agreed to definitions to avoid perceived data quality issues and erroneous reporting. These business data stewardship roles include the following:

- **Departmentally focused data stewards—**These stewards tend to align into organizational areas such as accounting, finance, and marketing. They narrowly focus on the definition, creation, maintenance, and usage of data only within an organizational area. Often these data stewards are aligned closer to the executive of that organizational area than with the information governance organization (for example, finance data stewards that report directly to the CFO).

- **Functionally focused data stewards—**These stewards tend to align closer to the information governance organizations and are responsible for the definition, creation, maintenance, and usage of data for a functional area such as customer or product that may span many different organizations. For example, the customer domain may cross finance, accounting, marketing, production, and distribution. It requires an understanding of how the definition and process events that impact the concept of customer as a customer are processed from potential to purchaser of the organization's goods and services. This broader organizational view almost always needs an information governance process to reconcile all the different organizational perspectives.

Technical Data Stewards

Technical data stewards focus more on the technical definition, creation, and maintenance of the data. They tend to report to IT, often the data management group, and provide the interface between IT and the business functional areas. These roles include the following:

- **Analytic data stewards—**These data stewards focus on the definition, maintenance, and usage of data generated from BI environments. Because much of this data has been transformed from its raw state through calculations and aggregations, one of the major tasks of these stewards is ensuring that the stakeholders agree to the common definitions and calculations of this data. They often work with the IT developers and end users in the definitions of the key performance measurements, calculations, and aggregations that make up the reporting. These are also the data stewards that work very closely to ensure that the information used for regulatory reporting meets the technical requirements of correctness and security.
- **Metadata management stewards—**These individuals have a very specific data stewardship focus on capture, maintenance, and versioning of the various types of business and technical metadata. They play a role that transcends IT's data management organization and the IGC in managing the metadata environment. For those organizations that have established a commercial or homegrown metadata management repository, these data stewards are responsible for the capture, versioning, and maintenance of the different types of metadata. Later this chapter provides a broader definition of the different types of metadata that are created and managed.
- **Data quality analysts—**These specific-purpose data stewards concentrate on the data quality aspects of a functional or organization area within an information governance organization. They assist in the definition of the data by focusing on what the data quality criteria are for critical data elements (for example, what the technical and business domains and ranges are). They also approve the critical data elements to meet the project's data quality requirements. They manage and perform the ongoing data quality audits and renovation projects on behalf of the information governance organizations.

Note that these are simply types of roles; in certain organizations, the same individual will perform any number of these data stewardship roles. The number and definition of the types of roles are also a function of the information governance maturity within an organization. The more mature the information governance, the more delineation will be found within the types of data stewardship roles.

Common Characteristics of Data Stewards

Regardless of type, certain common characteristics are found in all data stewards, such as a deep understanding of the underlying data and the processes and business rules that create that data; they are usually the data experts. Good data stewards tend to have deep industry expertise; they are very experienced practitioners in the industries that they work in. For example, a healthcare data steward understands the critical nature of ICD-10 codes, whereas a banking data steward is familiar with the regulatory requirements of the Dodd-Frank Act. They are by nature data evangelists, often with a deep passion for the data and its definition. Good data stewards tend to be 40% trained and 60% passion.

Understanding that the data steward is the performing "people" part of information governance ensures that when information governance activities and tasks are performed in development and ongoing operations, data stewards will in most instances be a primary or secondary performer.

Data Quality Management Component

Data quality management is the definition, supervision, and when necessary, renovation of data to the business and technical ranges. Data quality management is one of the most visceral aspects of information governance. It is also "threads" through each of the "people, process, and technology" aspects of information governance. For example, organizational reactions to perceived or real data quality issues have cost organizations millions of dollars in regulatory fines, cost executives their positions, and are one of the primary reasons companies start information governance initiatives. However, despite all the press, it is still one of the least understood areas of information governance.

What Is Data Quality?

Data quality is the commonly understood business and technical definitions of data within defined ranges. It is measured by how effectively the data supports the transactions and decisions needed to meet an organization's strategic goals and objectives, as embodied in its ability to manage its assets and conduct its core operations.

The level of data quality required to effectively support operations will vary by information system or business unit, depending on the information needs to conduct that business unit's operations. For example, financial systems require a high degree of quality data because of the importance and usage of the data, but a marketing system may have the latitude to operate with a lower level of data quality without significantly impacting the use of the information in

measuring marketing success. Because the purpose varies, so does the bar used to measure fitness to purpose.

Causes of Poor Data Quality

Causes for bad data quality can be categorized as business-process and technology-process data quality issues, as demonstrated in Figure 1.4.

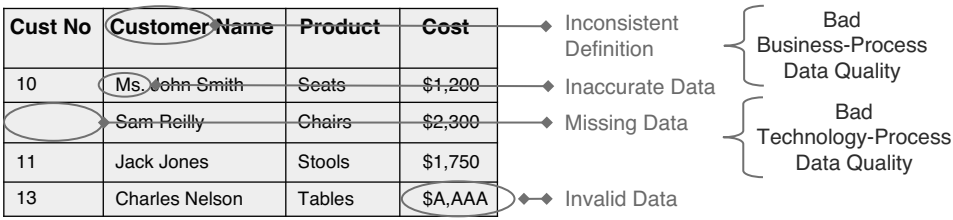


Figure 1.4 Examples of bad data quality types

Technology-driven poor data qualities are those types that are caused by not applying technology constraints on either the database or data integration. These types include the following:

- **Invalid data—**Data that in incorrect in that field. For example, by not applying constraints, alphanumeric data is allowed in a numeric data field (or column).
- **Missing data—**Data that is missing in that field. For example, by not applying key constraints in the database, a not-null field has been left null.

Business-driven bad data qualities are those types that are caused by end users inaccurately creating or defining data. Examples include the following:

- **Inaccurate data—**Invalid data due to incorrect input by business users. For example, by inaccurately creating a record for Ms. Anthony Jones, rather than for Mr. Anthony Jones, poor data quality is created. Inaccurate data is also demonstrated by the "duplicate data" phenomenon. For example, an organization has a customer record for both Anthony Jones and Tony Jones, both the same person.
- **Inconsistent definitions—**Inconsistent data is where stakeholders have different definitions of the data. By having disparate views on what the definition of poor data quality is, perceived bad quality is created. For example, when the sales department has a different definition of customer profitability than the accounting department.

The Data Quality Framework

Most EIM functions have an architecture or framework by which to understand that function; data quality is no exception. The data quality framework illustrated in Figure 1.5 is a multidimensional

reference model to explain and define data different dimensions of data quality. The first dimension defines the key data quality elements, or what data within an organization or application is important to measure quality. The business and technical dimensions provide the rules that measure how well a data element meets a company's data quality goals and ultimately provides trusted and critical information.

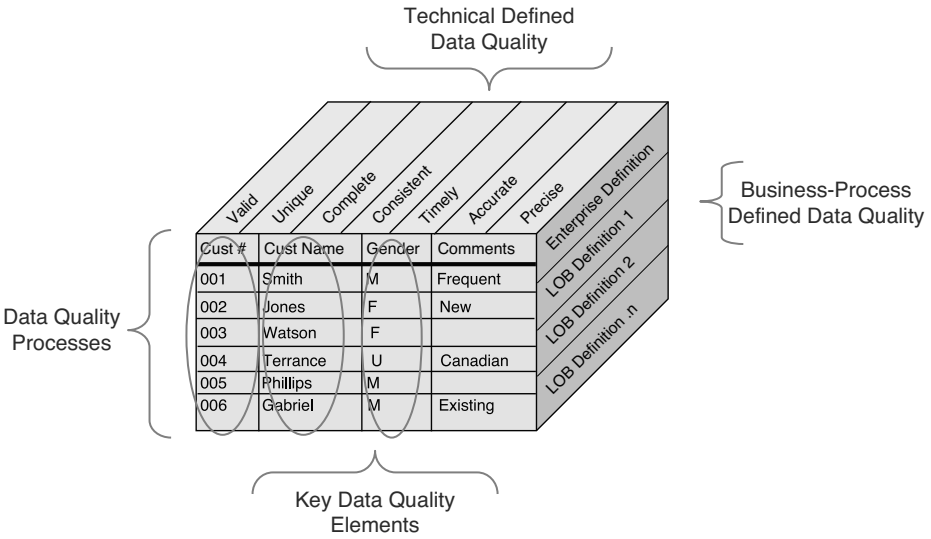


Figure 1.5 The dimensions of the data quality framework

Understanding all four aspects of this framework will help you determine what information governance activities and tasks must be performed to ensure the levels of data quality desired by an organization.

Key Data Quality Element Dimension

Within an organization, certain data elements are critical to the business and so the data quality of such should be identified, defined, and measured. These key data elements can be both base element data (for example, customer name) as well as derived data (for example, net profit).

These key data quality elements are often defined as such during data definition activities such as data modeling. Once identified as a key data quality element, the technical and business data quality criteria for that element are identified and defined in terms of ranges of compliance to requirements of a business. For instance, the key data quality element birth date has a business data quality criteria defined as a date range, as follows:

Birth date = Range: from 0 to 140

This business user-defined range reflects the probability that most people simply do not live beyond 140 years.

Although a relationship exists between relational key constraints, mandatory data, and key data quality elements, that relationship is not one to one. Not all mandatory and constraint data is necessarily key data quality data.

For instance, a customer ID column may be both mandatory and a primary key constraint, but not a key data quality element based on that element's importance to the organization.

Business-Process Data Quality Dimension

The business-process data quality dimension refers to the data quality criteria based on the business definition and business rules defined within the data. It contains the business defined ranges and domains that are a direct result of a business decision.

It is the lack of formal definition or misunderstanding of the different interpretations that create the inconsistent definitions and different business rules for similar data within each line of business (LOB), with each LOB having its own understanding of what that data element is. For example:

- Marketing definition of net assets = Assets – Expenses
- Finance definition of net assets = Assets – Expenses + Owners equity

Hence, with disparate views on what the definition and business rules of a data quality element are, when information is compared from different LOBs, the perception of bad quality is created, as shown in Table 1.1.

Table 1.1 Business Dimension of Data Quality

Name	Description	Examples of Poor Business Data Quality
Definitional	The data element has a commonly agreed-upon enterprise business definition and calculations.	Return on net assets (RONA), net present value (NPV), and earnings before interest, taxes and amortization of goodwill (EBITA) are calculated using different algorithms/equations and using different source data for each algorithm/equation for multiple departments within an enterprise.

Applying a consistently agreed-upon *common* business definition and rules against the data elements provides the insurance against inconsistent data quality issues.

It is the management of the common understanding of business definitions throughout the data stewardship community that is so critically important to not have misunderstood reporting issues.

Technical-Process Data Quality Dimension

The technical-process data quality dimension refers to the data quality criteria found in the technical definition of the data (for example, as defined in both the entity integrity and referential integrity relational rules found in logical data modeling). Table 1.2 describes key aspects of this dimension.

Table 1.2 Technical Dimensions of Data Quality

Name	Description	Examples of Poor Technical Data Quality
Valid	The data element passes all edits for acceptability.	A customer record has a name that contains numbers. The Social Security Number field should be a numeric integer but is populated with alphanumeric characters instead.
Unique	The data element is unique; there are no duplicate values.	Two customer records have the same Social Security number.
Complete	The data element is always required or required based on the condition of another data element.	A product record is missing a value such as weight. Married (y/n) field should have a non-null value of y or n, but is populated with a null value instead.
Consistent	The data element is free from variation and contradiction based on the condition of another data element.	A customer order record has a ship date preceding its order date.
Timely	The data element represents the most current information resulting from the output of a business event.	A customer record references an address that is no longer valid.
Accurate	The data element values are properly assigned (e.g., domain ranges).	A customer record has an inaccurate or invalid hierarchy.
Precise	The data element is used only for its intended purpose, i.e., the degree to which the data characteristics are well understood and correctly utilized.	Product codes are used for different product types between different records.

Each of these technical data quality rules are enforced against the key data quality elements with different methods. Many of the rules are enforced with simple relational database rules such as entity and referential integrity. For instance, the precise dimension is enforced in the relational database by applying the primary key constraint.

Within each of these dimensions, technical data quality rules are applied against key data quality elements, as shown in Figure 1.6.

Data Quality Criteria Work Book								
Table: Customer Column Name	Technical Dimension							Business Dimension
	Valid	Unique	Complete	Consistent	Timely	Accurate	Precise	Enterprise Business Definition
	Data element passes all edits for acceptability	Data element is unique —there are no duplicate values	Data element is (1) always required or (2) required based on the condition of another data element	Data element is free from variation and contradiction based on the condition of another data element	Data element represents the most current information resulting from the output of a business event	Data element values are properly assigned. E.g., Domain ranges.	Data element is used only for its intended purpose	
Cust_Id	Must Be Numeric	Primary	Not Null	Relational rules on Primary Keys	Last Update within the past month	Is a part of an Involved Party	Must be Marketing or Sales to Create	The unique identifier assigned to a Customer.
Cust_First_Name	**N/A**	Manditory	Not Null	Cust_Id must exist	Last Update within the past month	Is a part of an Involved Party	Must be Marketing or Sales to Create	Specifies the first name of the Party
Cust_Last_Name	**N/A**	Manditory	Not Null	Cust_Id must exist	Last Update within the past month	Is a part of an Involved Party	Must be Marketing or Sales to Create	Specifies the last name of the Party
Gender	Yes	Manditory	Not Null	It must be "**Male**", "**Female**", or "**Unknown**"	Last Update within the past month	Is a part of an Involved Party	Must be Marketing or Sales to Create	Gender of the customer. Data Quality Criteria: Male, Female, Unknown

Figure 1.6 The applied technical data quality rules in a data quality workbook

Data quality is not just about the structure and content of individual data attributes. Often, serious data quality issues exist because of the lack of integrity between data elements within or across separate tables that might be the result of a business rule or structural integrity violations. Ultimately, the degree to which the data conforms to the dimensions of the data quality framework that are relevant to it dictates the level of quality achieved by that particular data element.

Data Quality Processes Dimension

The data quality framework provides the structure to instantiate the policies and procedures developed and agreed to by the IGC and provide the basis for data stewards and development teams to define the processes to capture and prevent bad data quality. Examples of these processes are found in the next section.

Data Quality Checkpoints

Capturing and renovating bad data that has been defined in the context of the data quality framework can be prevented by determining key data quality criteria and building those rules into data quality checkpoints. There are two types of data quality checkpoints:

- **Technical data quality checkpoints—**Technical data quality checkpoints define the data quality criteria often found in both the entity integrity and referential integrity relational rules found in logical data modeling. They address the invalid and missing data quality anomalies. Technical data quality criteria are usually defined by IT and

information management subject matter experts (SMEs). An example includes the primary key null data quality checkpoint.

- **Business data quality checkpoints—**The business data quality checkpoints confirm the understanding of the key data quality elements in terms of what the business definition and ranges for a data quality element are and what business rules are associated with that element. Business data quality checkpoints address the inaccurate and inconsistent data quality anomalies. The classic example of a business data quality check is gender. A potential list of valid ranges for gender is Male, Female, or Unknown. This is a business definition, not an IT definition; the range is defined by the business. Although many organizations find the three values for gender sufficient, the U.S. Postal Service has seven types of gender, so their business definition is broader than others.

Types of Data Quality Processes

The final aspect of the data quality framework are those processes that ensure good data quality or prevent bad quality from being created and those that find bad data quality for renovation.

Ensuring data quality is typically a *result* of solid adherence to the definition of data quality criteria from both a business process and data design perspective. As a result, there are *preventive* data quality best practices that focus on the development of new data sources and integration processes, and there are *detective* data quality best practices that focus on identification and remediation of poor data quality. Both of these types are found in the tasks and steps of the data quality life cycle, which is discussed in Chapter 11, "Ongoing Data Quality Management Processes."

The understanding of what data quality is, the framework for which it is defined, and how to capture data quality is critical to understanding one of the important "process" components of information governance, especially in terms of ensuring the right data quality processes are built and then monitored in ongoing operations.

Metadata Management Component

The metadata management component is one of the process and technology aspects of information governance that captures, versions, and uses metadata to understand organization data. It is the "database" for data stewards and other types of users to store, maintain, and use the business and technical definitions of the organization's data.

What is metadata? Metadata is defined as "data about data," but it can also be explained as another layer of information created to help people use raw data as information. Metadata provides context to raw data; it is the business and technical rules that provide that particular data element meaning, as illustrated in Figure 1.7.

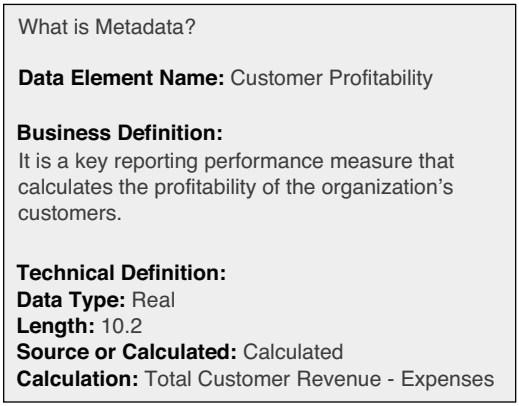


Figure 1.7 Types of metadata: Business and structural

Metadata is created whenever data is created, either in transaction processing, master data management (MDM) consolidation, or BI aggregations. Each event creates a type of metadata that often needs to be captured and managed. For example, when a data element is created, it contains information about what process was used to create it, along with rules, formulas, and settings, regardless of whether it is documented. The goal is to capture this metadata information at creation to avoid having to rediscover it later or attempt to interpret it later.

The discipline of metadata management is to capture, control, and version metadata to provide users such as data stewards the ability to manage the organization's data definitions and data processing rules in a central location. The application to capture, store, and manage metadata is a metadata repository, which is a metadata "database" for use by stakeholders such as data stewards.

Metadata can be composed of any information that describes the actual data itself. For data warehousing purposes, metadata has been classified based on the purpose created and the functions it is used for and can be classified into the types or categories. In each of these categories, there are relationships. For example, navigational, structural, and analytic all require the business definitions in the business metadata to provide context to the data, as demonstrated in Figure 1.8.

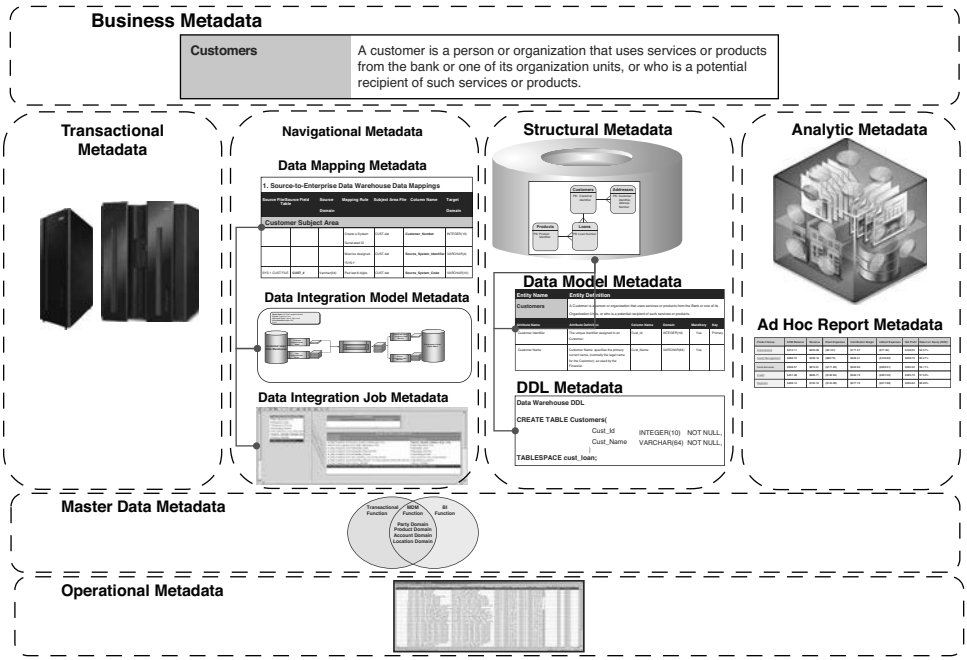


Figure 1.8 The categories of metadata

Business Metadata

The business category of metadata defines the information that the data provides in a business context. Examples of business metadata include subject area definitions (e.g., product), entity concept definitions, business attribute names, business attribute definitions, business attribute valid values, data quality rules, and business rules. Business metadata is found in transactional data master data. One of the primary sources of business metadata includes conceptual data models, logical data models, and business process rules engines.

Transactional Metadata

Transactional metadata contains the business and technical data definitions and business rules used in creating transactional systems. Transactional metadata is the source of all downstream uses of information, and when it is poorly defined or enforced, it is the major source of data quality issues.

Structural Metadata

Structural metadata contains the logical and technical descriptions of the permanent data structures within the EIM infrastructure. This metadata includes structures such as flat files and hierarchical and relational databases. Structural metadata contains both logical and technical metadata, as shown in Figure 1.9.

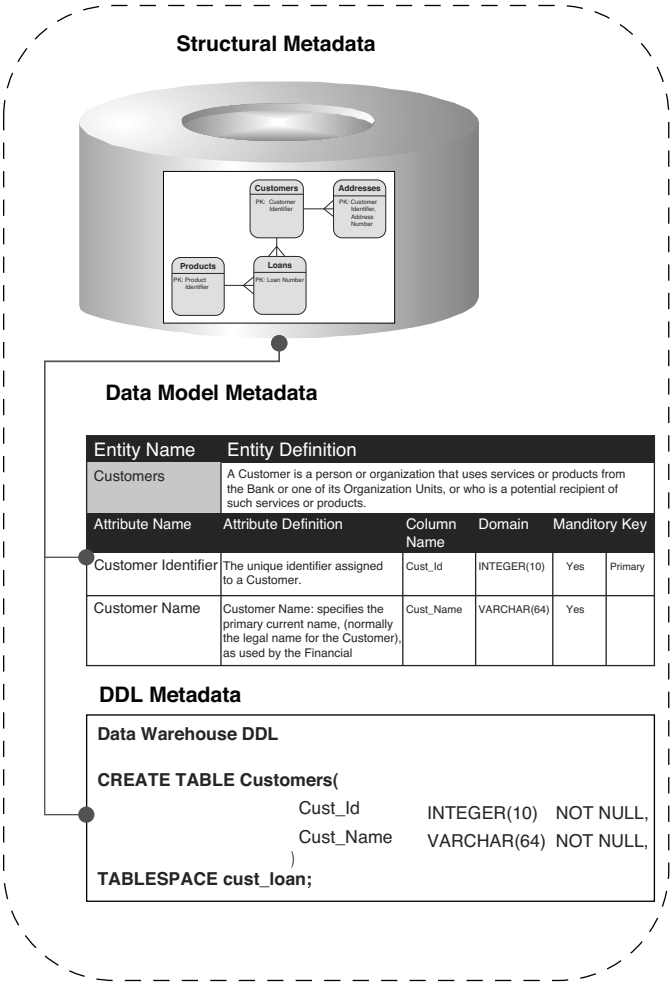


Figure 1.9 Structural metadata example

Logical metadata consists of data models and entity, attribute, and relationship metadata. A level of overlap exists between business and logical metadata (for example, business attributes and physical attributes). Business attributes are defined by the business to describe an aspect of

an entity. A physical attribute is defined by a data modeler or application database administrator to describe an aspect of the physical store of data. Some organizations only retain and manage the one type.

The technical metadata is the physical structures themselves (for example, databases/file groups, tables/views/files, keys, indices, columns/fields, source columns/fields, and target columns/fields). Often this type of information is found in Database Definition Language (DDL).

Navigational Metadata

Navigational metadata describes the process rules and data formats of the data extraction, transformation, and movements, as illustrated in Figure 1.10. Examples of navigational technical metadata are derived fields, business hierarchies, source columns and fields, transformations, data quality checkpoints, target columns and fields, and source and target locations. Primary sources of navigational metadata include data profiling results, data mappings, logical/physical data integration models, and data quality criteria workbooks.

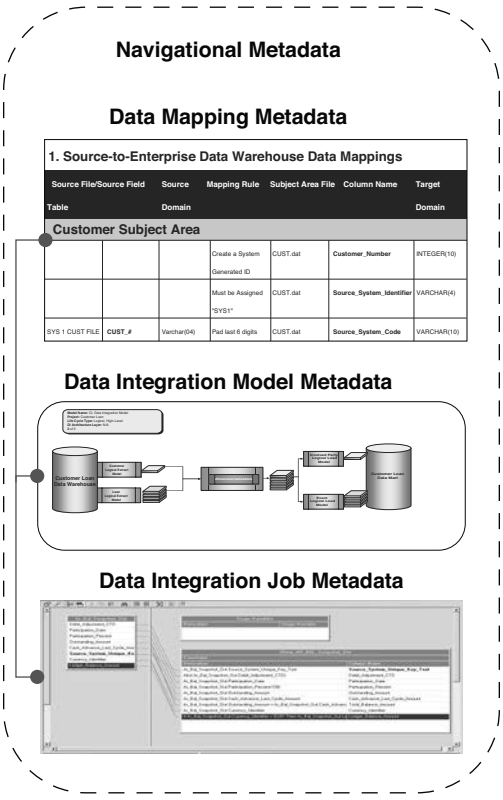


Figure 1.10 Navigational metadata example

Commercial data integration software vendors have addressed navigational metadata from two perspectives:

- **Integrated software suites—**IBM, Ab Initio, and Informatica have integrated profiling and data analysis tools into their design and development suites. This includes data mapping.
- **Metadata repositories—**The same vendors have metadata repositories for navigational metadata as well as the capabilities to integrate other types, which is discussed later in the chapter.

Analytic Metadata

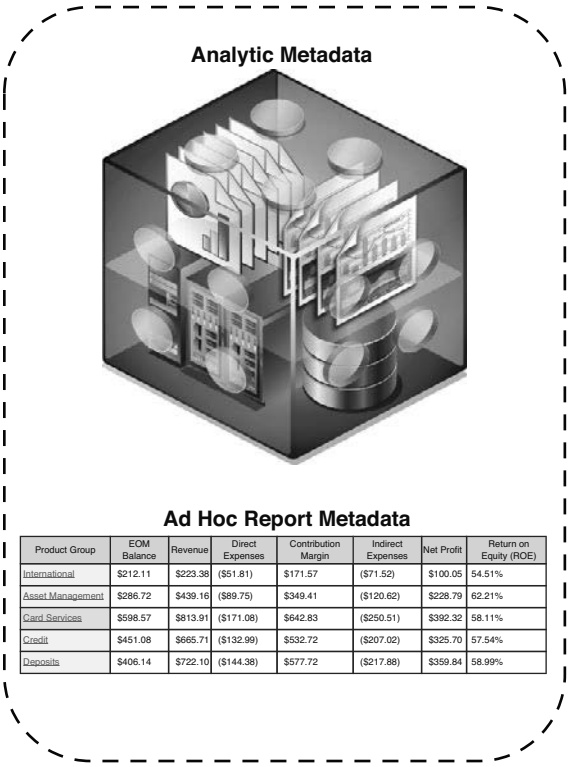
Analytic metadata, shown in Figure 1.11, consists of the metadata that is used in a reporting and ad hoc environment and includes the following:

- **Report data elements—**Within the report itself, the definition of the report-level data elements displayed on the report or in the ad hoc query environment is metadata to be created and managed. These elements are often the same technical and business definitions as the data warehouse or dimensional data mart.

Note

However, these data elements often have changed technical and business metadata that is different from the data warehouse environment, leveraging the ability of the commercial analytic tool metadata capabilities. These changes should be captured and documented from both a data stewardship and metadata management perspective.

- **Report-level aggregations and calculations—**Most commercial analytic tools provide the ability to build aggregations and calculations at the report level.
- **Report layout and report navigation metadata—**This technical metadata describes the layout of the report, the fonts to be used, and how the data should be shown and navigated.



Product Group	EOM Balance	Revenue	Direct Expenses	Contribution Margin	Indirect Expenses	Net Profit	Return on Equity (ROE)
International	$212.11	$223.38	($51.81)	$171.57	($71.52)	$100.05	54.51%
Asset Management	$286.72	$439.16	($89.75)	$349.41	($120.62)	$228.79	62.21%
Card Services	$598.57	$813.91	($171.08)	$642.83	($250.51)	$392.32	58.11%
Credit	$451.08	$665.71	($132.99)	$532.72	($207.02)	$325.70	57.54%
Deposits	$406.14	$722.10	($144.38)	$577.72	($217.88)	$359.84	58.99%

Figure 1.11 Analytic metadata example

Primary sources of analytic metadata include OLAP and reporting packages metadata environments.

Master Data Metadata

Master data metadata crosses both transaction and analytic application definitions that describe the core business domains of an organization. Master data provides transaction and analytic data the context of the organization for core domains such as party-customer, product, and account, as shown in Figure 1.12.

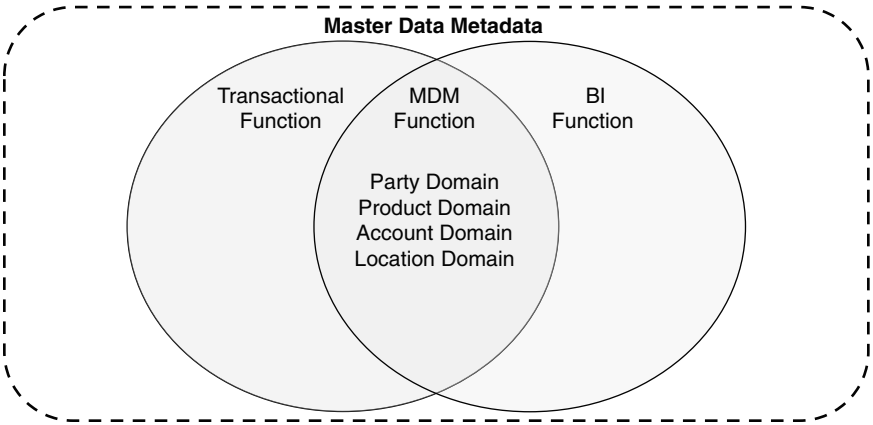


Figure 1.12 Master data metadata example

Operational Metadata

The operational category of metadata describes the transaction and data integration application's job description through statistics, giving a full technical view of the environment. Examples of operational metadata include jobs statistics and data quality check results.

Whereas the prior categories are primarily used by business users, data stewards, and data management professionals, operational metadata is used by production support and systems administration for troubleshooting and performance tuning.

Sources of operational metadata include transaction and data integration job logs being generated either by the data integration jobs or the production scheduler.

Metadata Users

Metadata provides value at a variety of levels to a range of users but can typically be divided into three categories:

- **Business users—**Business users of metadata need to understand the business meaning of the data in the systems they use. In addition, they need to know the business rules and data access rules that apply to the data. Data stewards (either business or technology) are usually classified as business users due to the creation, maintenance, and usage patterns of metadata.
- **Technology users—**IT professionals who are responsible for planning and building the transaction and analytic systems need to understand the end-to-end picture of the data to manage change. These users leverage the technical metadata for the technical information about the data environment, such as physical data structures, extract-transform-load

rules, reporting information, and impact analysis. Examples of technology users include data modelers, service-oriented architecture (SOA) architects, data-integration architects, BI architects, designers, and developers.

- **Operational users—**IT operational professionals are those who are responsible for day-to-day operation of the data environment and are users of operational metadata. Operational metadata can assist them in identifying and resolving problems as well as managing change in the production environment by providing data information about the data integration processing and job processing impact analysis.

Managing Metadata

Because metadata is created in many places during the development of a system, it is important to understand and govern all the categories of metadata in the metadata life cycle. Information management professionals have had the goal of a centrally managed metadata repository that governs all metadata, but that vision is difficult to achieve for a variety of factors. The reality is that metadata is created in many different tools used to develop data structures and process that data, as shown in Figure 1.13.

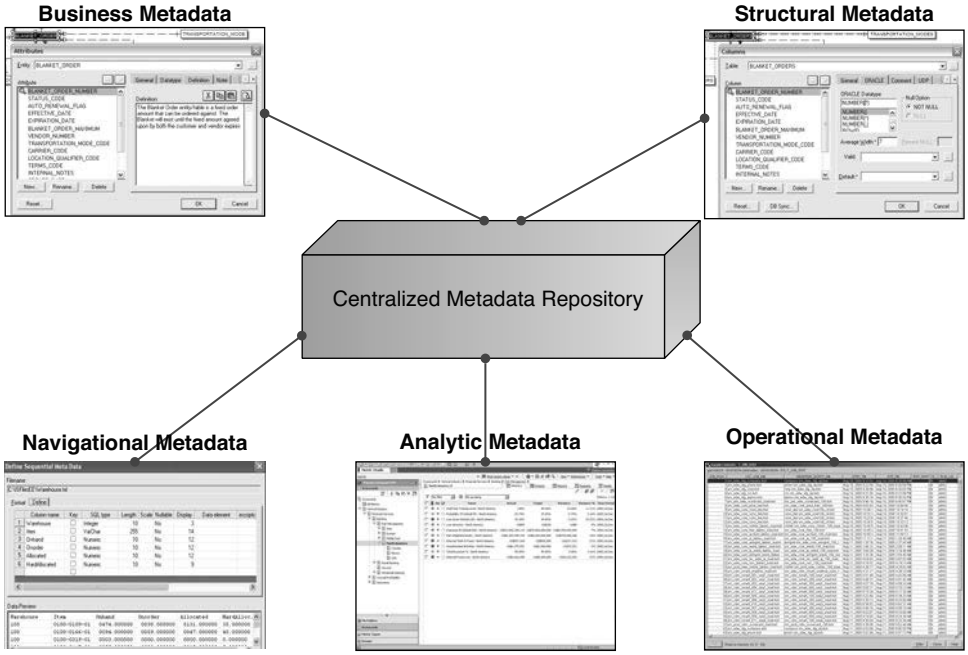


Figure 1.13 Centrally managing sources of metadata

At best, a centralized metadata repository should enhance metadata found in local repositories. A metadata repository should consider the following:

- **Where it will be stored—**Identify the data store requirements (e.g., commercial metadata repository, homegrown relational database).
- **What will be stored—**Identify metadata sources.
- **How it will be captured—**Identify load mechanism, CRUD (create, read, update, delete) requirements, administration requirements, and audit and retention requirements.
- **Who will capture the data—**Identify the roles and responsibilities for managing the repository and levels of users.
- **When it will be captured—**Identify capture frequency, history, and versioning considerations.
- **Why it will be captured—**Identify the benefits of the requirements and the specific questions this metadata will answer and provide reporting/browsing requirements.

Metadata is an organization's "encyclopedia" of business and technical definitions for use by data stewards and other key users. Capturing and updating metadata is a very visible an important set of activities in performing information governance.

Understanding that the data steward is performing the "people" part of information governance ensures that when information governance activities and tasks are performed in development and ongoing operations, data stewards will in most instances be a primary or secondary performer.

Privacy and Security Component

The privacy and security component covers all three of the people, process, and technology aspects of information governance to address who has create, read, update, and delete privileges of organizational data. There have been security requirements for data since the beginning of IT, with access and file security on mainframes with ACF2 and RACF security packages. This was further refined with the advent of relational database technologies with role- and column-level security and "locking data" down with schema-level security roles.

Privacy has taken on an equal if not more important (from a legal liability perspective) role with the integration of organizations' intranets with the external Internet. The ability for nonstakeholders to access critical financial, customer, and employee data has spawned legislation such as personally identifiable information (PII) laws on how data can and cannot be used to identify, contact, or locate an individual. Another example is in the healthcare industry in the Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy and security law, which seeks to ensure the privacy and security rights of an individual's health information. These and other such laws have made the role of information governance even more prominent.

A Broader Overview of Security

Information governance security "interlocks" with the broader IT security and general security functions at the data definition and usage level, as shown in the classic Venn diagram in Figure 1.14.

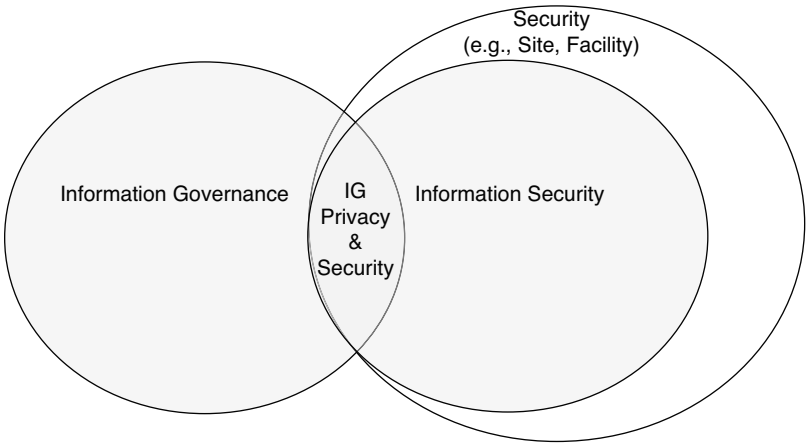


Figure 1.14 Information governance security and privacy in the context of a broader security function

As with other information governance components, there is a framework that best describes how security and privacy "threads" into EIM functions, as shown in Figure 1.15.

EIM Functions	Transactional Processes			Master Data Management		Business Intelligence		
	Data	Process	Objects	Operational	Analytical	Data Integration	Database	Analytics
Business Privacy & Security Requirements	✓	✓	✓	✓	✓	✓	✓	✓
Technology Privacy & Security Requirements	✓	✓	✓	✓	✓	✓	✓	✓
External Requirements	*PII*			*HIPAA*		*Other Regulatory Agencies*		

Figure 1.15 Security and privacy framework

Each EIM functional component of the framework in Figure 1.15 requires a thoughtful analysis and implementation approach for each of the dimensions for the business, technical, and external requirements for privacy and security. For example, a healthcare organization's member data that is collected through a website needs to consider the following:

- **Business privacy and security requirements—**Who with the healthcare organization is allowed to access that data?
- **MDM technical requirements—**What are the business, technical, and HIPAA (regulatory) rules for integrating this member data with other stores of member data?
- **Privacy and security requirements in analytic analysis—**How can the member data collected from the Web be used for member profiling without violating HIPAA?
- **Technical privacy and security requirements for the data warehouse—**What technical solution, such as database security, schema security, and user roles, will meet HIPAA requirements for healthcare member data?

Each EIM "functional layer" of data should be determined through stewardship processes in conjunction with the chief information security officer.

Chief Information Security Officer

The critical nature of security and privacy has placed the chief information security officer (CISO) in the IGC as a board member, as shown in Figure 1.3. The CISO works with the CDO in setting security and privacy policies and often works directly with data stewards on project and operational issues surrounding security and privacy. For example, a data steward may need to review proposed security standards with the CISO to ensure that they meet HIPAA requirements.

Understanding how privacy and security is defined for data based on the business, technical, and regulatory requirements is critical in performing information governance.

Information Life Cycle Management Component

Information life cycle management (ILM) covers the process and technology aspect of information governance that addresses the entire life cycle of a set of data, including creation, retention, and deletion. It covers the business rules on how long data is to be kept and in what format. Due to the very technical nature of ILM, it is as much a data management discipline as it is a component of information governance. Despite the commoditization of computing CPU and disk storage, retaining vast amounts of data that can easily be into the hundreds of petabytes can run in the range of $50 million to $100 million per year. Based on usage and legal requirements, data can be cycled from traditional "hot" storage to cheaper archived storage that can still be accessed as needed (thus saving considerable amounts of money).

It is important for data stewards to consider both the usage and the legal requirements in determining whether to archive or delete old data. For example, a telecommunications company's

data warehouse is required to store 4 years of billing data in its data warehouse. However, for tax compliance, it is required for 7 years; so, a potential life cycle management plan for billing data would be 4 years online and then 3 years offline/archived. After 7 years, the data could be deleted. So, in most cases, the following formula can be used:

Data must be retained for whichever is greater: organizational retention requirements or regulatory retention requirements.

This area of information governance has become much more focused since it provides a much more manageable and cost-effective approach to storing vast amounts of data.

Information life cycle management is one more dimension to consider when defining data and performing data stewardship audits.

Information Governance in the Context of EIM

With each of the information governance components defined, the next step is to understand the relationship of information governance in the context of EIM in order to document how to perform these information governance activities and tasks in both project and ongoing operations.

Observe how the information governance components thread into each of the different EIM functions, as documented in Figure 1.16.

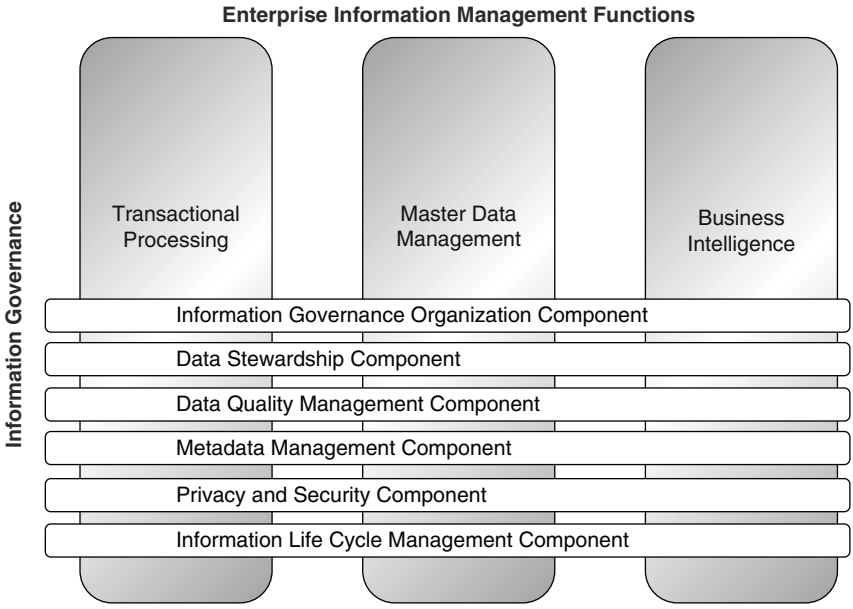


Figure 1.16 information governance threading into EIM functions

A fully functional information governance organization would be deeply embedded into the development and maintenance cycle of each of the EIM functions in order to determine, define, and audit the information governance aspects of the information environment.

The goal will be to ensure that each of the information governance component requirements are integrated into the EIM functions' common design and development themes.

Summary

This chapter provided a review of each of the different components of information governance, as follows:

- **Information governance organizational component—**The information governance organizational component is the "people" aspect of the discipline that sets the policies for information governance and maintains a staff to support those policies in managing the development and ongoing usage of corporate information.
- **Data stewardship component—**Data stewards support, maintain, and execute the policies and procedures instituted by the IGC.
- **Data quality management component—**Data quality management is the definition, supervision, and when necessary, renovation of data to the business and technical ranges.
- **Metadata management component—**The metadata management component is one of the process and technology aspects of information governance that captures, versions, and uses metadata to understand organization data.
- **Privacy and security component—**The privacy and security component covers all three of the people, process, and technology aspects of information governance to address who has create, read, update, and delete privileges of organizational data.
- **Information life cycle management component—**Information life cycle management covers the process and technology aspect of information governance that addresses the entire life cycle of a set of data, from creation, retention, and deletion.

Each of these components is threaded in each of the EIM functions, transactional processing, MDM, and BI.

The policies and requirements for each of the components must be instantiated to achieve the context and quality of the data needed for transaction and analytic processing within the organization.

This chapter defined each of the information governance components that need to be performed to achieve the context and quality of the data needed for transaction and analytic processing within the organization. Chapter 2, "Other Core EIM Functions," reviews, at a high-level, the transaction, operational, and analytic functions of EIM so that you can understand where the requirements for the information governance components will interface.

End-of-Chapter Review Questions

1. What is the formal definition of information governance?
2. Fill in the blank: The information governance organizational component is the ____ aspect of the discipline that sets the policies for information governance and maintains a staff to support those policies in managing the development and ongoing usage of corporate information.
3. What are some of the reasons why organizations fail in their information governance efforts?
4. What types of data stewards are discussed in this chapter?
5. What are the four dimensions of the data quality framework?
6. What are the four causes of poor data quality?
7. When is metadata created?
8. What are the three types of privacy and security requirements that need to be considered when determining the creation and usage of data?
9. What is the definition of information life cycle management?
10. True or false: Information governance security "interlocks" with the broader IT security and general security functions.

Essay Question

Explain what information governance components are in place in your organization today.

For Further Reading

General Data Management
http://www.amazon.com/Guide-Management-Knowledge-DAMA-DMBOK-Edition/dp/1935504029/ref=tmm_pap_title_0?ie=UTF8&qid=1354892108&sr=1-1

Enterprise Information Management
http://www.amazon.com/Practical-Implementing-Enterprise-Information-Management/dp/1934938920/ref=sr_1_1?ie=UTF8&qid=1354891982&sr=8-1&keywords=stephen+t+boschulte

http://www.amazon.com/The-Enterprise-Information-Architecture-Systems-Based/dp/0137035713/ref=sr_1_sc_1?s=books&ie=UTF8&qid=1354892060&sr=1-1-spell&keywords=the+art+of+enterprise+information+architecture

Master Data Management
http://www.amazon.com/MASTER-DATA-MANAGEMENT-GOVERNANCE/dp/0071744584/ref=sr_1_1?s=books&ie=UTF8&qid=1354892176&sr=1-1&keywords=mdm

http://www.amazon.com/Enterprise-Master-Data-Management-Information/dp/0132366258/ref=sr_1_4?s=books&ie=UTF8&qid=1354892176&sr=1-4&keywords=mdm

Enterprise Architecture

http://www.amazon.com/Enterprise-Architecture-Planning-Developing-Applications/dp/0471599859/ref=sr_1_4?s=books&ie=UTF8&qid=1354892283&sr=1-4&keywords=enterprise+architecture

http://www.amazon.com/Patterns-Enterprise-Application-Architecture-Martin/dp/0321127420/ref=sr_1_3?s=books&ie=UTF8&qid=1354892283&sr=1-3&keywords=enterprise+architecture

This page intentionally left blank

Index

B

C

E

F

G

H

M

P

Q

R

T