

1 Introduction

Chapter 1 introduces the importance of data assets in the information age, the data management function, the data management profession, and the goals of the DAMA-DMBOK Guide. It sets the stage for presenting a Data Management Overview in the next chapter.

1.1 Data: An Enterprise Asset

Data and information are the lifeblood of the 21st century economy. In the Information Age, data is recognized as a vital enterprise asset.

“Organizations that do not understand the overwhelming importance of managing data and information as tangible assets in the new economy will not survive.”

Tom Peters, 2001

Money and people have long been considered to be enterprise assets. Assets are resources with recognized value under the control of an individual or organization. Enterprise assets help achieve the goals of the enterprise, and therefore need to be thoughtfully managed. The capture and use of such assets are carefully controlled, and investments in these assets are effectively leveraged to achieve enterprise objectives.

Data, and the information created from data, are now widely recognized as enterprise assets.

No enterprise can be effective without high quality data. Today’s organizations rely on their data assets to make more informed and more effective decisions. Market leaders are leveraging their data assets by creating competitive advantages through greater knowledge of their customers, innovative uses of information, and operational efficiencies. Businesses are using data to provide better products and services, cut costs, and control risks. Government agencies, educational institutions, and not-for-profit organizations also need high quality data to guide their operational, tactical, and strategic activities. As organizations need and increasingly depend on data, the business value of data assets can be more clearly established.

The amount of data available in the world is growing at an astounding rate. Researchers at the University of California at Berkeley estimate that the world produces between 1 and 2 billion bytes of data annually. It often seems we are drowning in information.

Yet for many important decisions, we experience information gaps – the difference between what we know and what we need to know to make an effective decision. Information gaps represent enterprise liabilities with potentially profound impacts on operational effectiveness and profitability.

Every enterprise needs to effectively manage its increasingly important data and information resources. Through a partnership of business leadership and technical

expertise, the data management function can effectively provide and control data and information assets.

1.2 Data, Information, Knowledge

Data is the representation of facts as text, numbers, graphics, images, sound or video. Technically, data is the plural form of the word Latin word *datum*, meaning “a fact.” However, people commonly use the term as a singular thing. Facts are captured, stored, and expressed as data.

Information is data in context. Without context, data is meaningless; we create meaningful information by interpreting the context around data. This context includes:

1. The business meaning of data elements and related terms.
2. The format in which the data is presented.
3. The timeframe represented by the data.
4. The relevance of the data to a given usage.

Data is the raw material we interpret as data consumers to continually create information, as shown in Figure 1.1. The resulting information then guides our decisions.

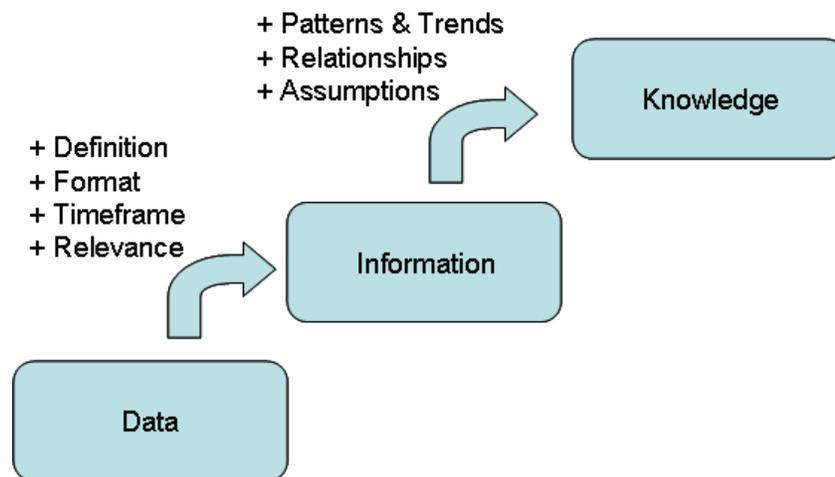


Figure 1.1 Data, Information, and Knowledge

The official or widely accepted meanings of commonly used terms also represent a valuable enterprise resource, contributing to a shared understanding of meaningful information. Data definitions are just some of the many different kinds of “data about data” known as *meta-data*. Meta-data, including business data definitions, helps establish the context of data, and so managing meta-data contributes directly to improved information quality. Managing information assets includes the management of data and its meta-data.

Information contributes to *knowledge*. Knowledge is understanding, awareness, cognizance, and the recognition of a situation and familiarity with its complexity. Knowledge is information in perspective, integrated into a viewpoint based on the recognition and interpretation of patterns, such as trends, formed with other information and experience. It may also include assumptions and theories about causes. Knowledge may be explicit—what an enterprise or community accepts as true—or tacit—inside the heads of individuals. We gain in knowledge when we understand the significance of information.

Like data and information, knowledge is also an enterprise resource. Knowledge workers seek to gain expertise through the understanding of information, and then apply that expertise by making informed and aware decisions and actions. Knowledge workers may be staff experts, managers, or executives. A learning organization is one that proactively seeks to increase the collective knowledge and wisdom of its knowledge workers.

Knowledge management is the discipline that fosters organizational learning and the management of intellectual capital as an enterprise resource. Both knowledge management and data management are dependent on high quality data and information. Knowledge management is a closely related discipline, although in this document, knowledge management is considered beyond the scope of data management.

Data is the foundation of information, knowledge, and ultimately, wisdom and informed action. Is data truth? Not necessarily! Data can be inaccurate, incomplete, out of date, and misunderstood. For centuries, philosophers have asked, “What is truth?”, and the answer remains elusive. On a practical level, truth is, to some extent, information of the highest quality – data that is available, relevant, complete, accurate, consistent, timely, usable, meaningful, and understood. Organizations that recognize the value of data can take concrete, proactive steps to increase the quality of data and information.

1.3 The Data Lifecycle

Like any asset, data has a lifecycle, and to manage data assets, organizations manage the data lifecycle. Data is created or acquired, stored and maintained, used, and eventually destroyed. In the course of its life, data may be extracted, exported, imported, migrated, validated, edited, updated, cleansed, transformed, converted, integrated, segregated, aggregated, referenced, reviewed, reported, analyzed, mined, backed up, recovered, archived, and retrieved before eventually being deleted.

Data is fluid. Data flows in and out of data stores, and is packaged for delivery in information products. It is stored in structured formats—in databases, flat files, and tagged electronic documents—and in many less structured formats—e-mail and other electronic documents, paper documents, spreadsheets, reports, graphics, electronic image files, and audio and video recordings. Typically, 80% of an organization’s data assets reside in relatively unstructured formats.

Data has value only when it is actually used, or can be useful in the future. All data lifecycle stages have associated costs and risks, but only the “use” stage adds business value.

When effectively managed, the data lifecycle begins even before data acquisition, with enterprise planning for data, specification of data, and enablement of data capture, delivery, storage, and controls.

Projects accomplish the specification and enablement of data, and some of the planning for data. The System Development Lifecycle (SDLC), shown in Figure 1.2, is not the same as the data lifecycle. The SDLC describes the stages of a project, while the data lifecycle describes the processes performed to manage data assets.

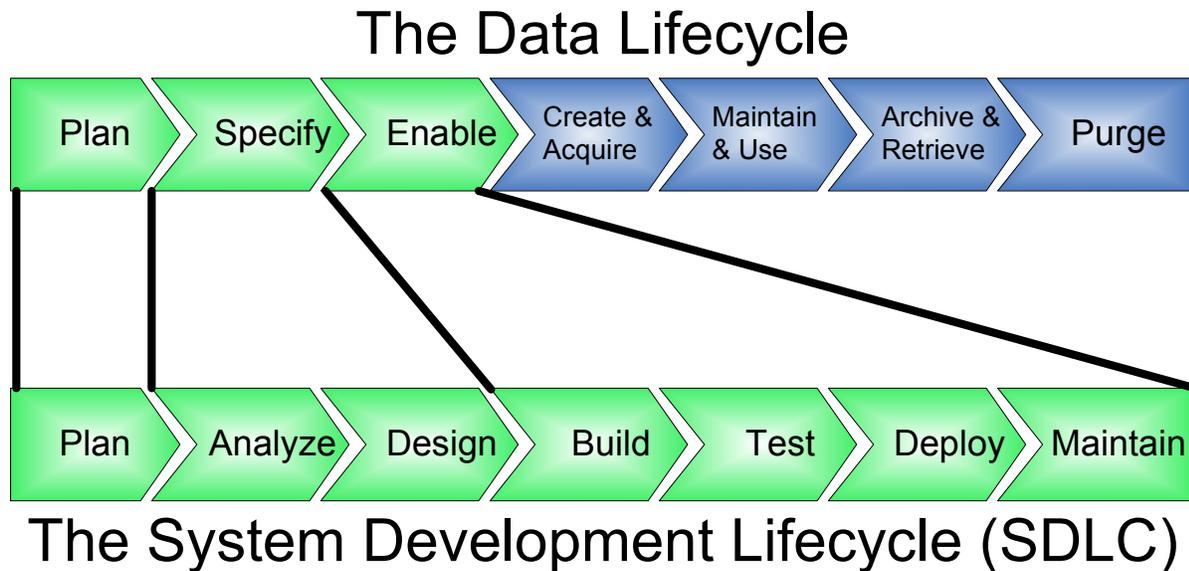


Figure 1.2 The Data Lifecycle and the System Development Lifecycle

However, the two lifecycles are closely related because data planning, specification and enablement activities are integral parts of the SDLC. Other SDLC activities are operational or supervisory in nature.

1.4 The Data Management Function

Data management (DM) is the business function of planning for, controlling and delivering data and information assets. This function includes

- The disciplines of development, execution, and supervision
- of plans, policies, programs, projects, processes, practices and procedures
- that control, protect, deliver, and enhance
- the value of data and information assets.

Data management is known by many other terms, including:

- Information Management (IM).
- Enterprise Information Management (EIM).
- Enterprise Data Management (EDM).
- Data Resource Management (DRM).
- Information Resource Management (IRM).
- Information Asset Management (IAM).

All these terms are generally synonymous, but this document consistently refers to Data Management.

Often the word “enterprise” is included in the function name to emphasize the enterprise-wide focus of data management efforts, i.e., Enterprise Information Management or Enterprise Data Management. Enterprise-wide data management is a recommended best practice. However, data management may also be performed effectively in a local context without an enterprise-wide mandate, although with less business benefit.

The data management function includes what is commonly referred to as database administration—database design, implementation, and production support—as well as “data administration”. The term “data administration” was once a popular way to vaguely refer to all the functions of data management except database administration. However, as the data management function matures, its specific component functions are better understood. The data management function is important to enterprises regardless of their size and purpose.

The scope of the data management function and the scale of its implementation vary widely with the size, means, and experience of organizations. The nature of data management remains the same across organizations, even as implementation details widely differ.

1.5 A Shared Responsibility

Data management is a shared responsibility between the *data management professionals* within Information Technology (IT) organizations and the *business data stewards* representing the collective interests of data producers and information consumers. Data stewards serve as the appointed trustees for data assets. Data management professionals serve as the expert curators and technical custodians of these data assets.

Data stewardship is the assigned accountability for business responsibilities in data management. Data stewards are respected subject matter experts and business leaders appointed to represent the data interests of their organizations, and take responsibility for the quality and use of data. Good stewards carefully guard, invest, and leverage the

resources entrusted to them. Data stewards ensure data resources meet business needs by ensuring the quality of data and its meta-data. Data stewards collaborate in partnership with data management professionals to execute data stewardship activities and responsibilities.

Data management professionals operate as the expert technical custodians of data assets, much like bank employees and money managers serve as the professional custodians of financial resources for their owners and trustees. While data stewards oversee data assets, data management professionals perform technical functions to safeguard and enable effective use of enterprise data assets. Data management professionals work in Data Management Services organizations within the Information Technology (IT) department.

Data is the content moving through the information technology infrastructure and application systems. Information technology captures, stores, processes, and provides data. The IT infrastructure and application systems are the “pipes” through which data flows. As technological change has exploded over the past fifty years, IT organizations have traditionally focused primarily on maintaining a modern, effective hardware and software infrastructure, and a robust application system portfolio based on that infrastructure. Most IT organizations have been less focused on the structure, meaning, and the quality of the data content flowing through the infrastructure and systems. However, a growing number of IT executives and business leaders today recognize the importance of data management and the need for effective Data Management Services organizations.

1.6 A Broad Scope

The overall data management function, shown in Figure 1.3, encompasses ten major component functions:

- Data Governance: Planning, supervision and control over data management and use.
- Data Architecture Management: Defining the blueprint for managing data assets.
- Data Development: Analysis, design, implementation, testing, deployment, maintenance.
- Data Operations Management: Providing support from data acquisition to purging.
- Data Security Management: Insuring privacy, confidentiality and appropriate access.
- Data Quality Management: Defining, monitoring and improving data quality.
- Reference and Master Data Management: Managing golden versions and replicas.

- Data Warehousing and Business Intelligence Management: Enabling reporting and analysis.
- Document and Content Management: Managing data found outside of databases.
- Meta-data Management: Integrating, controlling and providing meta-data.

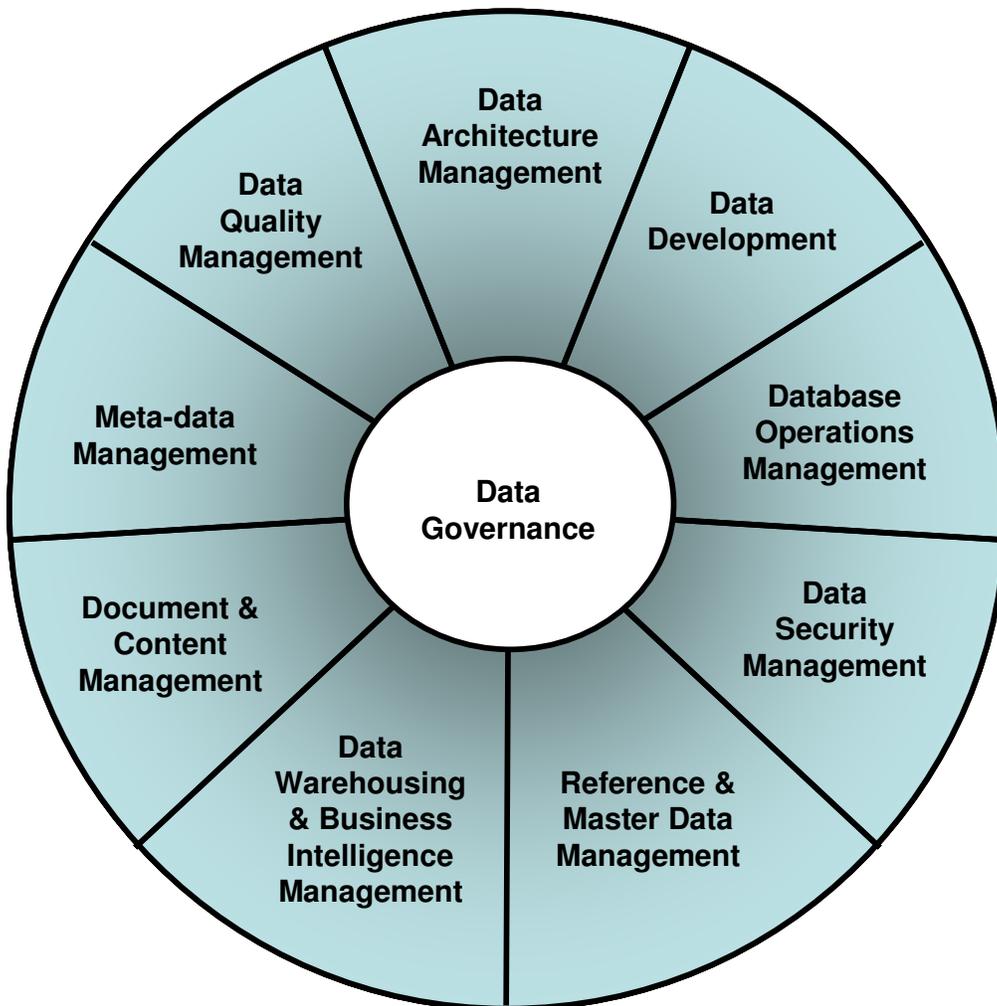


Figure 1.3 Data Management Functions

1.7 An Emerging Profession

The management practices for established assets like money and people have matured over many years. Data management is a relatively new function and its concepts and practices are evolving rapidly.

Within the IT community, data management is an emerging profession—an occupational calling requiring specialized knowledge and skills. Specialized data management roles require unique skills and experienced judgments. Today’s data management professionals demonstrate a sense of calling and exceptional commitment to managing data assets.

Creating a formal, certified, recognized, and respected data management profession is a challenging process. The current environment is a confusing mixture of terms, methods, tools, opinion, and hype. To mature into an established profession, the data management community needs professional standards: standard terms and definitions, processes and practices, roles and responsibilities, deliverables and metrics.

Standards and recognized best practices can improve the effectiveness of data stewards and data management professionals. Moreover, standards help us communicate with our teammates, managers, and executives. Executives especially need to fully understand and embrace fundamental data management concepts in order to effectively fund, staff and support the data management function.

1.8 A Growing Body of Knowledge

One of the hallmarks of an emerging profession is the publication of a guide to a recognized consensus body of knowledge. A “body of knowledge” is what is generally accepted as true in a professional field. While the entire body of knowledge may be quite large and constantly growing, a guide to the body of knowledge introduces standard terms and best practices.

1.9 DAMA—The Data Management Association

The Data Management Association (DAMA International) is the *Premiere organization for data professionals worldwide*. DAMA International is an international not-for-profit membership organization, with over 7500 members in 40 chapters around the globe. Its purpose is to promote the understanding, development, and practice of managing data and information to support business strategies.

The DAMA Foundation is the research and education affiliate of DAMA International, dedicated to developing the data management profession and promoting advancement of concepts and practices to manage data and information as enterprise assets.

The joint mission of DAMA International and the DAMA Foundation, collectively known as DAMA, is to *Lead the data management profession toward maturity*. DAMA promotes the understanding, development, and practice of managing data, information, and knowledge as key enterprise assets, independent of any specific vendor, technology, and method.

DAMA International seeks to mature the data management profession in several ways. A few of these efforts include:

- DAMA International conducts the annual DAMA International Symposium, now the Enterprise Data World, the largest professional data management conference

in the world, in partnership with Wilshire Conferences. Workshops, tutorials, and conference sessions at the Symposium provide continuing education for data management professionals.

- DAMA International conducts the annual DAMA International Conference Europe, the largest professional data management conference in Europe, in partnership with IRMUK. Workshops, tutorials, and conference sessions at the Conference provide continuing education for data management professionals.
- DAMA International offers a professional certification program, recognizing Certified Data Management Professionals (CDMP), in partnership with the Institute for Certification of Computing Professionals (ICCP). CDMP certification exams are also used by The Data Warehouse Institute (TDWI) in the Certified Business Intelligence Professional (CBIP) program.
- The DAMA International Education Committee's Data Management Curriculum Framework offers guidance to US and Canadian colleges and universities regarding how to teach data management as part of any IT and MIS curriculum in the North American higher education model.

1.10 Purpose of the DAMA-DMBOK Guide

DAMA International produced this document, The Guide to the Data Management Body of Knowledge (the DAMA-DMBOK Guide), to further the data management profession. The DAMA-DMBOK Guide is intended to be a definitive introduction to data management.

No single book can describe the entire body of knowledge. The DAMA-DMBOK Guide does not attempt to be an encyclopedia of data management or the full-fledged discourse on all things related to data management. Instead, this guide briefly introduces concepts and identifies data management goals, functions and activities, primary deliverables, roles, principles, technology and organizational / cultural issues. It briefly describes commonly accepted good practices along with significant alternative approaches.

1.11 Goals of the DAMA-DMBOK Guide

As a definitive introduction, the goals of the DAMA-DMBOK Guide are:

1. To build consensus for a generally applicable view of data management functions.
2. To provide standard definitions for commonly used data management functions, deliverables, roles, and other terminology.
3. To identify guiding principles for data management.
4. To overview commonly accepted good practices, widely adopted methods and techniques, and significant alternative approaches, without reference to specific technology vendors or their products.

5. To briefly identify common organizational and cultural issues.
6. To clarify the scope and boundaries of data management.
7. To guide readers to additional resources for further understanding.

1.12 Audiences of the DAMA-DMBOK Guide

Audiences for the DAMA-DMBOK Guide include:

- Certified and aspiring data management professionals.
- Other IT professionals working with data management professionals.
- Data stewards of all types.
- Executives with an interest in managing data as an enterprise asset.
- Knowledge workers developing an appreciation of data as an enterprise asset.
- Consultants assessing and helping improve client data management functions.
- Educators responsible for developing and delivering a data management curriculum.
- Researchers in the field of data management.

1.13 Using the DAMA-DMBOK Guide

DAMA International foresees several potential uses of the DAMA-DMBOK Guide, including:

- Informing a diverse audience about the nature and importance of data management.
- Helping standardize terms and their meanings within the data management community.
- Helping data stewards and data management professionals understand their roles and responsibilities.
- Providing the basis for assessments of data management effectiveness and maturity.
- Guiding efforts to implement and improve their data management function.
- Pointing readers to additional sources of knowledge about data management.
- Guiding the development and delivery of data management curriculum content for higher education.

- Suggesting areas of further research in the field of data management.
- Helping data management professionals prepare for CDMP and CBIP exams.

1.14 Other BOK Guides

Several other professions have published a Body Of Knowledge document. Indeed, the existence of a Body of Knowledge document is one of the hallmarks of a mature profession (see Chapter 13).

The primary model for the DAMA-DMBOK Guide is A Guide to the Project Management Body of Knowledge (PMBOK® Guide), published by the Project Management Institute (PMI®). PMI® is a professional organization for project managers. Among its many services, PMI® conducts the Project Management Professional (PMP) certification program.

Other Body of Knowledge documents include:

- A Guide to the Software Engineering Body of Knowledge (SWEBOK), published by the Institute of Electrical and Electronic Engineers (IEEE). IEEE has begun to offer a certification program for software engineers.
- The Business Analysis Body of Knowledge (BABOK), published by the International Institute of Business Analysis.
- The Common Body of Knowledge (CBK) published by the International Information Systems Security Certification Consortium ((ISC). The CBK is the information tested to achieve the Certified Information Systems Security Professional (CISSP) designation.
- The Canadian Information Technology Body of Knowledge (CITBOK) is a project undertaken by the Canadian Information Processing Society (CIPS) to outline the knowledge required of a Canadian Information Technology Professional.

1.15 The DAMA Dictionary of Data Management

The DAMA Dictionary of Data Management is a companion volume to the DAMA-DMBOK Guide. Originally developed as an extensive Glossary for the DAMA-DMBOK Guide, DAMA International published it separately due to its size and business value. Definitions for terms found in the Dictionary are consistent with their usage in the DAMA-DMBOK Guide. The Dictionary is available for purchase as a CD-ROM.