2

WHO OWNS INFORMATION?

Before we delve into the details of what data quality means and how it relates to knowledge management, we should establish where the responsibility for data quality falls within a company. Without a clear assignment of accountability, it is almost impossible to measure the quality of data, much less effect improvements.

This chapter examines the question of data ownership as the first step in establishing a knowledge-oriented organization. We begin by discussing data processing activity as a manufacture of information, and knowledge, which is owned by the data consumers in a business enterprise and is the final product of this factory.

Who are the data producers and data consumers in an enterprise? We look at internal data producers (internal processes like account opening, billing, marketing) and external data producers ("lead lists," consumer research, corporate structure data). We also look at the enterprise data consumers, ranging from the operational (customer service, billing, resource planning), the tactical (middle management, scheduling), and strategic consumers (directional management, strategists).

There are complicating notions with respect to data ownership. The means of dissemination, collecting data from the public domain, as well as acquiring data from data providers confuse the issues. Therefore, a set of ownership paradigms is defined, including decision makers, sellers, manipulators, guardians, and workers. These paradigms bound the "social turf" surrounding data ownership.

Finally, we try to resolve some of these issues by investigating data policy paradigms. This includes metadata ownership, governance of storage and repositories, and accountability for data policies.

2.1 THE INFORMATION FACTORY

A relatively simple analogy for processing data that we use throughout the book is the information factory. Any information processing activity can be viewed as a small factory that takes some data as raw input, processes that input, and generates some information result, potentially generating data by-products and side effects in the process. Inside the factory there may be smaller subfactories, each with its own input/output production activity. The raw input data are provided by data suppliers external to the organization or by data manufacturers within the organization. The ultimate data customers may be internal or external consumers.

2.1.1 Actors in the Information Factory

To be more precise, let's look at the different roles that exist in the context of the information factory. These roles may represent real people or automated proxies within the system.

1. *Suppliers:* Data suppliers provide information to the system.
2. *Acquirers:* Acquirers accept data from external suppliers for provision into the factory.
3. *Creators:* Internal to the factory, data may be generated and then forwarded to another processing stage.
4. *Processors:* A processor is any agent that accepts input and generates output, possibly generating some side effects.
5. *Packagers:* A packager collates, aggregates, and summarizes information for reporting purposes.
6. *Delivery Agents:* A delivery agent delivers packaged information to a known data consumer.
7. *Consumer:* The data consumer is the ultimate user of processed information.
8. *Middle Manager:* The people responsible for making sure the actors are correctly performing their jobs.
9. *Senior Manager:* The senior manager is responsible for the overall operation of the factory.
10. *Deciders:* These are senior-level managers associated with strategic and tactical decision making.

Each of these actors plays a well-defined role in the data processing operation, and each is responsible at some level for quality assurance

within each activity domain. In a perfect world, these responsibilities will all propagate up to the enterprise level, providing some degree of quality assurance overall, but in reality there are complicating factors that may prevent this from happening. It is clear, though, that at any stage of processing, it is difficult to specifically assign ownership to the information being created or processed.

2.1.2 Processing Stages

In any data processing system, it is helpful to be able to break down the entire information flow into a series of processing stages, most of which relate directly to the activity associated with one of the previously enumerated actors. Ultimately, we would like to be able to precisely identify all input and output information streams, as well as all affected data stores associated with each processing stage, and then associate responsibility with a particular actor.

When truly decomposed, the information manufacturing process contains many different stages, each of which in itself might represent an entire instance of an information factory. We end up with a hierarchical description of the information processing chain.

2.1.3 Data Producers

Data producers are those organizations that create, compile, aggregate, package, and provide information to be inserted into an information processing system. This includes organizations that generate internal data (audit data, workflow messages, intermediate processed data for incorporation into other processing stages) or external information, such as marketing data, sales reports, invoices, corporate structure data, credit reports, and so forth.

2.1.4 Data Consumers

Data consumers can be categorized into three groups: operational data consumers, tactical and strategic data consumers, and external customers. Operational data consumers are manifested as the role of processors described in Section 2.1.1. They are any internal processing

stage that requires input to operate, including any transaction processing activity, message passing or routing, or workflow activities.

The tactical and strategic data consumers are those who use processed information to make tactical and strategic decisions. This includes sales management, marketing, enterprise resource planning, mergers and acquisitions, and so on.

The external customers are those who receive information processed by the information factory. The kinds of information in this category include invoices, customer billing, sales teams data, geographic data, and data provided by government agencies.

2.2 COMPLICATING NOTIONS

What complicates the ownership question is that there are factors orthogonal to data creation or consumption that create real or artificial boundaries around information. The ownership issue is essentially a control issue — control of the flow of information, the cost of information, and the value of information.

Here is an example of how control over the flow of information can have a major impact. Some government interest rate financial data was mistakenly released early, prompting trading associated with selling bonds on the market before the official release time. The same traders, by buying back the same bonds at a discount after the official release time, made a huge profit. In this case, one party's loss of control over the flow of information allowed other parties increased control over the value of that same information!

A few of the many issues that complicate the notion of data ownership are value, privacy, turf, fear, and bureaucracy. This list is by no means inclusive, and we make no attempt to resolve them — only to expose them.

2.2.1 Value

The value of information drives a particular wedge into the question of ownership. In any environment where there is shared data ownership, how does the degree of ownership relate to the responsibility of care of that same information? Presumably, the degree of ownership may be related to more mundane aspects of the system, such as who initially

created the database or what silo currently manages the system. But at the core, the degree of ownership (and by corollary, the degree of responsibility) is driven by the value that each interested party derives from the use of that information.

2.2.2 Privacy

The issue of privacy as a complicating notion could take up an entire volume, if not more. What information should remain private, and under what circumstances? If a party willingly releases private information under one set of conditions, does that allow the receiving party the right to use that information in other situations? Consider credit information. The fact that someone applies for any particular credit card might be considered private information, although in order to receive that credit, a credit bureau must be consulted. At that point, the fact that the credit card application has been taken is now added to the credit record.

And once information is released from one party to another, who exercises control over that information? For example, if a pharmacy fills an individual's medicine prescription, can the pharmacy report back to the pharmaceutical company the information regarding which doctors prescribed the medication, which patients filled the prescriptions, and which came back for refills?

When private knowledge can be inferred from public data, does this violate any privacy constraints? For example, the amount of money one borrows on a mortgage to pay for a house might be considered a private matter between the borrower and the bank, but in fact this information is lien information that is frequently filed under various state codes and is not private information at all.

2.2.3 Turf

On a different level, the question of data ownership within an enterprise is often complicated because of the notion of "turf." In many organizations, the control of the flow of information is regarded, as are many other forms of control, as a means of job security. "As long as I am in charge of this report," thinks the middle manager, "and no one else has access to the data, I can't be fired!" Being in charge of creating,

packaging, and distributing the report naturally leads one to the conception of owning the data that makes up the report.

2.2.4 Fear

As organizational employees carve out their little fiefdoms of control, the overall ability of the organization to react to inconsistencies flowing out of these fiefdoms decreases. Any suggestion that there is a reason for an "outsider" to assess or analyze the current state of the data is taken as an attack on the fiefdom's turf and therefore can be construed as a direct challenge to the data controller's job.

An even worse fear is that any closer viewing of what goes on within one's organization will reveal that what appeared to be stellar work is actually mediocre, or worse. People can be so desperate to conceal their own mistakes that they will sabotage any attempt to uncover them. Of course, this is a prime example of conflict within the enterprise — what is good for the individual is terrible for the organization, and vice versa.

2.2.5 Bureaucracy

Another major issue that complicates data ownership is institutional bureaucracy. As organizations grow, the intermediate management structure grows as well, thereby diluting the responsibility for information as it passes from one subsystem to another. When issues regarding data quality problems arise, typically there tends to be a lot more finger pointing than problem solving.

This may be due to the fact that organizations become divided along project or system lines. Because of this, attribution of problems associated with information that passes through individually managed processing stages is hard to pin down, since each manager will pass the buck further upstream.

Another bureaucratic issue involves decisions associated with upgrades, renovations, and changes in the existing data processing infrastructure. With a highly hierarchical organization, the ability to make a decision hinges on building consensus among all interested parties both across vertical departmental lines as well as up and down the management hierarchy. Obviously, the effort involved is significant and, com-

bined with the turf and fear factors, may account for the failure of many enterprise infrastructure renovation projects.

2.3 RESPONSIBILITIES OF OWNERSHIP

What do we mean when we talk about ownership of data? The essence lies in the control of information as an enterprise asset. That control includes not just the ability to access, create, modify, package, derive benefit from, sell, or remove data but also the right to assign these access privileges to others. In this section, we discuss in greater detail some of the responsibilities associated with data ownership.

2.3.1 Definition of Data

In any data environment, the data owner is responsible for understanding what information is to be brought into a system, assigning the meanings to collections of data, and constructing the data model to hold the collected information. In addition, any modifications to the data model and any extensions to the system also fall under duties of the data owner.

2.3.2 Authorization of Access and Validation of Security

A major concern for any data system is the coordination and authorization of access. In a system that contains data that is in any way sensitive, whether it is confidential information, human resource data, or corporate intelligence, it is necessary to define a security and authorization policy and to provide for its enforcement.

2.3.3 Support the User Community

When information is provided to a user community, there is a responsibility to provide support for those users. This includes providing accessibility to new users, granting them access rights, providing documentation, training, and addressing technical needs and questions. This also includes

defining and maintaining a service level agreement, which may entail measuring system performance, and scaling or rebalancing resources to provide the agreed upon level of service.

2.3.4 Data Packaging and Delivery

In addition to standard user support, the owner also holds the responsibility for providing the data to the data consumers. This may include data preparation, packaging and formatting, as well as providing a delivery mechanism (such as a data portal or a publish/subscribe mechanism).

2.3.5 Maintenance of Data

Aside from the maintenance of the system itself, there is also the maintenance of the information. This includes managing the data input process, instituting gauges and measurements associated with the data, and creating data extraction and loading processes.

2.3.6 Data Quality

The data owner is also accountable for maintaining the quality of the information. This may include determining and setting user data quality expectations, instituting gauges and measurements of the levels of data quality, and providing reports on the conformance to data quality. This also includes defining data quality policies for all data that flows into the system and any data cleansing, standardization, or other preparation for user applications.

2.3.7 Management of Business Rules

All data processing operations have business rules. Whether these rules are embedded in application code, abstracted into a rules format, or just documented separately from their implementation, the data owner is also responsible for managing business rules.

2.3.8 Management of Metadata

Managing metadata involves the data definitions, names, data types, data domains, constraints, applications, database tables, reference repositories, and dependence rules associated with different tables and databases, users, access rights, and so forth.

2.3.9 Standards Management

Whenever information is shared between two or more parties, there must be some agreement as to a format for that data. When multiple parties agree to a representation format, that format is defined as a data standard. The owner is also responsible for making sure that all relevant data sets conform to their standard form, as well as negotiating standards on behalf of the users.

2.3.10 Supplier Management

When data sets are built as a composition of supplier-provided data, the data owner is also responsible for supplier management. This involves negotiating arrangements with each supplier, determining data delivery agreements, defining sets of data quality criteria, and enforcing these requirements and arrangements with each supplier.

2.4 OWNERSHIP PARADIGMS

We can enumerate owner responsibilities, but that does not solve the problem of assigning (or declaring) data ownership. Instead of trying to proactively dictate an ownership model, it is more helpful to explore different existing ownership paradigms. In each one of these paradigms, we will look at the question of value and how it relates to the claim of ownership.

2.4.1 Creator as Owner

In this paradigm, the party that creates or generates the data owns the data. It represents a speculative investment in creating information as a prelude to recognizing value from that information in the future.

An example of this is a geographic data consortium that analyzes geographic regions, collects latitude/longitude measures, and enters that information into a geographic database. The measurements in isolation are essentially useless; it is the collection of all the measurements that forms a useful data set. The consortium creates the information, and most likely claims ownership of that data as well.

2.4.2 Consumer as Owner

This ownership paradigm indicates that the party that consumes the data owns that data. This is a relatively broad ownership spectrum, covering all aspects of data acquisition. In this paradigm, any party that uses data claims ownership of that data. When the consumer requires a high level of confidence in the data input into a process, this ownership paradigm is very logical, since the party that cares most about the value of the data claims ownership (and thus, responsibility). In this case, the consumer derives the value from the data.

An example of this is a sales organization that uses information provided from different organizations within a company. Once the data lands at the sales staff's door, though, the information becomes integral to the proper operation of the sales team, and so the sales team will claim ownership of the data that it consumes.

2.4.3 Compiler as Owner

The operation of selecting information sources and compiling information from these different sources constitutes an ownership model. By combining data sets, the compiler is adding value and may expect to reap the benefits of ownership.

A good example of data compilation is a news item retrieval company that provides, as a service, a search for newspaper articles. By collecting and providing search capability, the data compiler has created a body of information that is more valuable than the individual pieces making up that body.

2.4.4 Enterprise as Owner

In larger corporate information technology organizations, there is a notion that all data that enter the enterprise or are created within the enterprise are completely owned by the enterprise. In effect, the company makes use of all input and generated data as fuel for its ongoing data processing needs, and therefore the value derived from the information resides with the organization as a whole.

The investment banking industry demonstrates this ownership model when it accumulates information from external market data vendors as well as data generated from internal sales and securities processing. All data are absorbed into a single operational data center that then redistributes the data, potentially with added value, out to data consumers within the organization as well as individuals (such as clients and customers) external to the enterprise.

2.4.5 Funding Organization as Owner

In this paradigm, the user that commissions the data creation claims ownership. Here there are two parties involved: the one that pays for the creation of data and the one that actually creates the data. In this case, the patron claims ownership, since the work is being done on his or her behalf.

An example is a company that commissions a research organization to prepare a competitive intelligence report covering a particular industry. The company may stipulate that the company is the sole owner of the provided data.

2.4.6 Decoder as Owner

In environments where information is "locked" inside particular encoded formats, the party that can unlock the information becomes an owner of that information. The cost of the decoding process and implementations is an investment in the value to be derived from the information.

A good example of this is embodied in the results of decoding DNA sequences to isolate specific genes. The value of decoding the DNA structure can be expressed in terms of any improvement in the discovery, prevention, or treatment of certain hereditary diseases. Bio-informatics

companies that decode genetic material can then sell the data that they decode to the medical and pharmaceutical industries.

2.4.7 Packager as Owner

As opposed to the compiler ownership paradigm, the packager paradigm focuses on the party that formats information for a particular use. There is value added through formatting the information for a particular market or set of consumers.

An example of this is authors who publish public domain information packaged as a book. The compilation process is most likely straightforward — the value added is in formatting the material to make the collected data useful.

2.4.8 Reader as Owner

This is an interesting paradigm in that it implies that the value of any data that can be read is subsumed by the reader, and therefore the reader gains value through adding that information to an information repository. The investment is in the reader's selection and consumption of data.

For example, consulting firms establish expertise practices in particular areas. In order to become a principal in one of these practices, an individual must acquire knowledge in the practice area by absorbing as much information about that area as possible. Going forward, remaining an expert requires active information gathering.

2.4.9 The Subject as Owner

This paradigm revolves around the subject data ownership issues, such as personal privacy or image copyrights. In this view, the subject of the data claims ownership of that data, mostly in reaction to another party claiming ownership of the same data.

As an example of the privacy issue, consider a pharmacy filling prescriptions. Drug companies are interested in knowing which doctors are prescribing their medicines, and doctors like to know which of their patients are refilling prescriptions as a tool to see how well their patients

follow instructions. Recently, it was revealed that a national pharmacy chain was providing to both health care providers and drug companies detailed information about who was filling which prescriptions at each of their sites. When this practice became public, naturally their customers were incensed. In effect, the individual patients were claiming ownership of personal information and insisting that the pharmacy chain had no right to sell it.

Another example is the issue of corporate image and branding. Companies will make a significant effort in establishing a connection between the quality of the products or services that it provides and a corporate image or logo. In this case, representations of the image are equated with the company, and any misrepresentation or other unauthorized use could affect the branding, so the company claims complete ownership of the image (or logo) as well as protects the use of that image.

2.4.10 Purchaser/Licenser as Owner

Similar to the funder paradigm, the individual or organization that buys or licenses data may stake a claim to ownership. In this paradigm, the purchaser assumes that the investment made in acquiring the data yields ownership. This holds for licensing as well, even if the terms of the license specify some restrictions on use.

A good example is the sale of mailing lists for direct marketing campaigns. One organization may own the lists and license their use, but once the lists are sold, the purchaser or licenser considers the data its own.

2.4.11 Everyone as Owner

The final paradigm is the model of global data ownership. Some feel that monopolization is wrong and data should be available to all with no restrictions. Clearly, in the business world, this is a radical view, and it has its benefits as well as detriments.

This ownership model is often in operation, to some degree, in scientific communities, where experimentation, following by the publishing of results, is common practice. In this situation, a common goal is the increase in the global knowledge of a particular subject, and results are subject to other experts' scrutiny.

2.5 CENTRALIZATION, DECENTRALIZATION, AND DATA OWNERSHIP POLICIES

This section explores the issues regarding the distribution of ownership across an enterprise. The question of centralization versus decentralization is orthogonal to the responsibilities of ownership, yet it is distinctly intertwined with it as well. As in our ownership model, the critical point revolves around value.

In a centralized ownership model, there is a single entity (person or group) responsible for all data ownership for the entire enterprise. Centralization implies that all ownership activities are coordinated from a single point of control, as well as coordination of metadata, information sourcing, and so forth. Centralized ownership yields the benefit of the value added — and whether the costs associated with centralization are offset by it. The costs include the increased management overhead, bureaucracy, and system integration, among others. The benefits include enterprise standardization for data and systems, the ability to make use of merged data for additional knowledge discovery, and increased leverage when dealing with external data suppliers.

In a decentralized model, the ownership roles are allocated to separate areas of interest. A decision to opt for decentralization implies that the value added from centralized control is more than offset by its associated costs. On the other hand, most organizations do not explicitly opt for decentralized control; instead, organizations evolve into it. Therefore, the real question is whether migrating from a decentralized ownership model to a centralized ownership model will increase the value of the enterprise knowledge base.

Finding the answer to this question is not simple. It involves a process of identifying the interested parties associated with all data sets, determining each party's interest, identifying the different roles associated with all data sets, and assigning roles and responsibilities to the right parties. All of these activities are embodied in an organization's **data ownership policy**, which incorporates all governance rules regarding data ownership and usage within the enterprise.

2.5.1 Creating a Data Ownership Policy

A data ownership policy is a tool used by the enterprise to establish all roles and responsibilities associated with data ownership and accountability. The goal of a data ownership policy is to finesse the kinds of

complications discussed in Section 2.2, as well as hash out the strict definitions of ownership as described in Section 2.4. The data ownership policy specifically defines the positions covering the data ownership responsibilities described in Section 2.3. At a minimum, a data ownership policy should enumerate the following features.

1. The senior level managers supporting the enforcement of the policies enumerated
2. All data sets covered under the policy
3. The ownership model (in other words, how is ownership allocated or assigned within the enterprise) for each data set
4. The roles associated with data ownership (and the associated reporting structure)
5. The responsibilities of each role
6. Dispute resolution processes
7. Signatures of those senior level managers listed in item 1

A template for describing the ownership policy for a specific data set is shown in Figure 2.1.

Data Set Name				
Primary Owner				
Data Set Location				
	Owner	Responsible Party	Reports to	Notes
Data Definition				
Access/Security				
User Support				
Data Packaging				
Data Delivery				
Maintenance				
Data Quality				
Business Rules				
Metadata				
Standards Management				
Supplier Management				

FIGURE 2.1 Template for data ownership policy

These are the steps for defining a data ownership policy.

1. Identify the interested parties or stakeholders associated with the enterprise data. This includes identifying the senior level managers that will support the enforcement of the policy.
2. Catalog the data sets that are covered under the policy.
3. Determine the ownership models in place and whether these are to continue or will be replaced.
4. Determine the roles that are and are not in place. Assign the responsibilities to each role, and assign the roles to interested parties.
5. Maintain a registry that keeps track of policies, data ownership, roles, responsibilities, and other relevant information.

2.5.2 Identifying the Stakeholders

All stakeholders in the information factory, including all the actors delineated in Section 2.1.1, should be considered interested parties. A stakeholder is anybody who expects to derive some benefit or value from the data, whether it is through the use of the data, the sale or license of the data, or beneficially through association with the data. For example, a business customer who uses the reports gets value through the data, receives monetary compensation through the sale or license of the data, and benefits from the jobs that may be dependent on continued data center operations and application development.

In a small enterprise, stakeholder identification can be relatively simple, but as the enterprise grows, the process can become extremely complex due to the degrees to which information is processed and disseminated. A good heuristic is to begin from the outside of the enterprise and work in. In other words, figure out who the end users are, look at the data they are using, and follow it backward through the information chain. While some business users may be outspoken in terms of staking their claim, others may be blind to the fact that there is any organizational process that generates the paper reports that land on their desks. Also, just because people receive the reports, they may never look at the data provided on a periodic basis.

The process of identifying the stakeholders will likely reveal areas of conflict with respect to data ownership. This is a particularly valuable part of the process, as it provides a guide to deciding how the ownership responsibilities are assigned.

2.5.3 Cataloging Data Sets

Once the stakeholders have been identified, the next step is to learn what data sets should fall under the ownership policy. The stakeholders should be interviewed to register the data sets with which they are associated and the degree to which each believes his or her stake in the data is. The goal of this step is to create a create a metadatabase of data sets to use in the enforcement of the data ownership policies. This catalog should contain the name of the data set, the location of the data set, and the list of stakeholders associated with the data set. Eventually, the catalog will also maintain information about data ownership and responsibilities for the data set.

2.5.4 Identifying and Assigning Roles

The next step is to determine the roles that are associated with each set of data in the enterprise and describe the responsibilities of each role. Here are some examples, although this list is by no means meant to be exhaustive.

Chief Information Officer The CIO is the chief holder of accountability for enterprise information and is responsible for decisions regarding the acquisition, storage, and use of data. He or she is the ultimate arbiter with respect to dispute resolution between areas of ownership and is the ultimate manager of the definition and enforcement of policies.

Chief Knowledge Officer The chief knowledge officer is responsible for managing the enterprise knowledge resource, which dictates and enforces the data sharing policies, as well as overseeing the general pooling of knowledge across the organization.

Data Trustee The data trustee manages information resources internal to the organization and manages relationships with data consumers and data suppliers, both internal and external.

Policy Manager The policy manager maintains the data ownership policy and negotiates any modifications or additions to the data ownership policy.

Data Registrar The data registrar is responsible for cataloging the data sets covered under the policy as well as the assignment of ownership, the definition of roles, and the determination of responsibilities

and assignments of each role. The data registrar also maintains the data policy and notifies the policy manager if there are any required changes to the data ownership policy.

Data Steward The data steward manages all aspects of a subset of data with responsibility for integrity, accuracy, and privacy.

Data Custodian The data custodian manages access to data in accordance with access, security, and usage policies. He or she makes sure that no data consumer makes unauthorized use of accessed data.

Data Administrator The data administrator manages production database systems, including both the underlying hardware and the database software. The data administrator is responsible for all aspects related to the infrastructure needed for production availability of data.

Security Administrator The security administrator is responsible for the creation of and the enforcement of security and authentication policies and procedures.

Director of Information Flow The director of information flow is responsible for the management of data interfaces between processing stages, as well as acting as an arbiter with respect to conflicts associated with data flow interfaces.

Director of Production Processing The director of production processing manages production processing operations, transference of data from one production source to another, scheduling of processing, and diagnosis and resolution of production runtime failures.

Director of Application Development The director of application development manages requirements analysis, implementation, testing, and deployment of new functionality for eventual turnover to the production facility.

Data Consumer A data consumer is an authorized user that has been granted access rights to some data within the enterprise.

Data Provider A data provider is an accepted supplier of information into the system.

These roles will then be integrated into a reporting structure where there are clear lines of responsibility corresponding to degrees of ownership. Note that some responsibilities are assigned to multiple roles, causing "role overlap," whose governance must be integrated into the reporting structure as well. At this point, the senior manager responsible for information (typically a chief information officer) will then assign ownership roles and responsibilities to the different organizational stakeholders.

2.5.5 Maintaining the Ownership Registry

The ownership registry is created from the data catalog and the assignment of roles. It is the enterprise log that can be queried to determine who has the ultimate responsibility for each data set. The ownership registry should be accessible by all interested parties, especially when new data requirements arise or there is a conflict that needs resolution.

Management of the ownership registry requires keeping a pulse on the organization, as it is not unusual for employee turnover to affect the data management structure. In addition, as new data sets are added to the governance by the data ownership policy, the decisions regarding the new data must be added to the registry.

2.6 OWNERSHIP AND DATA QUALITY

This brings us back to the issue of data quality. Once we have established a chain of command for the ownership of data, we can look at how the responsibility for data quality falls with respect to the policy. A major factor is the relationship between ownership and care, which is explored in this section, along with the enforcement of data policies and an introduction to data quality rules.

2.6.1 Ownership and Care

This may seem pretty obvious, but a person demonstrates a greater amount of care for an object that he or she owns than for something that belongs to someone else. To this end, it is important to consider the ownership stake associated with the parties in the enterprise responsible for the quality of an organization's data. It is less likely that the quality of data will be high when it is entrusted to someone who has no stake in the value of the data.

To be more precise, there are two ways to incorporate the ownership ideal with data quality: (1) Assign some degree of ownership to those entrusted with the data quality, or (2) assign the responsibility of data quality to the party with the highest degree of ownership.

In the first case, a person has been assigned the responsibility of maintaining the integrity of a data set but does not feel any personal attachment to the information in his or her trust. By allocating some

ownership role to that person, such as creating some kind of bonus compensation structure tied to maintaining or increasing the overall value of the data, an organization can infuse the employee with a renewed interest in data quality.

The second case reveals another common ownership issue where the end users (in many cases, business users) rely on the quality of the data but have no stake in ensuring that quality. This can be remedied by forcing the business users to get involved in the definition and assurance of data quality. This includes understanding the data that is being used, defining data quality rules that are the basis for acceptability, and enforcing data policies and data quality rules.

2.6.2 Enforcing Data Policies

It is not enough to have a set of data policies. Without a process for enforcement, the policy has no "teeth" and will be useless. It is incumbent on the drafters of a data ownership policy to incorporate the methods for enforcement as well as the process for conflict resolution. Enforcement takes the form of a means for validating that the policies are being followed and actions to be taken if it is determined that the policies are not being followed. It is the role of the policy manager to ensure that the policies are being enforced.

Validating policies is a making sure that all parties are living up to their responsibility agreements and are not overstepping the bounds of their responsibility. This involves periodic reviews by the policy manager, under the authority of the CIO, of whether each assigned role is being performed adequately. If the parties are not doing their job, it may be necessary to give them further training or decrease their responsibility (and authority) and possibly remove them completely from the position of responsibility.

No organization is immune to conflicts, and there should be a dispute resolution process in place. Typically, this will involve bringing the dispute to the next level in the management hierarchy with responsibility over the particular area of conflict (we will call this person the dispute manager). If the parties dispute an ownership role or whether the responsibilities of an ownership role have not been fulfilled properly, the dispute manager should consult the ownership policy along with the ownership registry and use that information to establish a ruling for the

dispute. If the issue is not covered in the data ownership policy, then the policy needs to be modified to incorporate the issue.

2.6.3 Data Quality Rules

As we move toward an organization with negotiated ownership roles, part of the responsibility of the management is to ensure that any data that is shared across the enterprise lives up to a standard of use. Ensuring this implies a few notions.

- There is a notion of data quality that is well defined throughout the enterprise.
- There is a means to describe data quality requirements.
- There is a means to measure conformance to the data quality requirements.
- There is an enterprisewide agreement as to the expected levels of data quality.
- There is a mechanism for improving data quality.

The concept of a definition of data quality is complex and important and will be covered in Chapter 3, and details of data quality ideas are discussed in Chapter 5. Ways to measure ongoing performance are examined Chapter 4. A large part of this book is devoted to defining and validating data quality rules.

2.6.4 Education and Training

When using data ownership policies to regulate the movement and management of information across an enterprise, it is critical that all participants understand the policies and how they work. This means that the managers defining the data ownership policy must also arrange for the education and training of staff members who are expected to use it. Data ownership training should cover at least these items.

1. An overview of data ownership, including ownership paradigms, ownership value, and data ownership responsibilities
2. A survey of the enterprise knowledge system covering the system architecture, the data architectures, the data set catalog, and all management, delivery, and presentation applications

3. An overview of the data ownership hierarchy describing the different roles involved, the responsibilities of each role, the reporting structure, and the conflict resolution process
4. A session covering enterprise metadata
5. A training session on the value of enterprise data quality, including an overview of data quality, a discussion of continuous measurement for improvement, and the definition and use of data quality and business rules

2.7 SUMMARY

This chapter introduced the concept of data ownership as a management issue. Because information is bought, used, created, modified, propagated, and sold throughout an organization, the enterprise data processing function can be contrasted to a factory where individuals or automated processes play specific roles.

Complicating notions in the enterprise make data quality management a difficult task, especially due to issues such as differing views of the value of data, privacy issues, turf wars, or standard bureaucracy. A list of the responsibilities of data ownership was provided for clarity.

There are different kinds of data ownership paradigms, and, depending on the organizational point of view, different ownership rules may apply in different situations. Ultimately, though, the management of the organization must choose between tight, centralized control or loose, decentralized control of data. These decisions are incorporated into a data ownership policy where the stakeholders, data sets, responsibilities, and dispute resolutions are all clearly defined. As stakeholders agree to subscribe to the data ownership policy, a more ordered environment enables better knowledge management overall and data quality management in particular.