Chapter 1

The Time Has Come for Change

1.1 What to Look for in This Chapter

- Why data protection is important
- What data protection is
- Why the right framework for data protection is necessary
- Why organizations should ride the sea change in data protection
- How to read this book

1.2 Why Data Protection Is Important

The protection of electronically stored information—in all its different expressions—should be at or near the top of the *have-to* issues for any business. First, data protection seeks to protect that information without which businesses cannot function well—if at all. Indeed, electronic information is now a primary source of many businesses' competitive advantage. The permanent physical loss of key information (such as customer account information) or the loss of confidentiality of sensitive information (such as the theft of a trade secret) could have a severe negative impact on a business (such as loss of revenue or capital value of the firm). So data protection is a cornerstone of any organization's management of risk, and risk management is now recognized as one of the fundamental tasks of any enterprise.

Moreover, businesses have other obligations to protect their data, apart from the risk of loss of usability for normal business purposes. Compliance is—or should be—at the top of consciousness for nearly every organization today. As one of the many facets of compliance, the well-known data security threat that loss of confidentiality of information through a data breach can bring, such as a loss of data privacy, is paramount. And the need for better accountability has led to the need for better governance, notably from a data protection perspective, in how to deal with information with respect to requirements for managing the civil litigation process.

Today, data protection touches a wide spectrum of business issues, including but by no means limited to:

- Backup and restore
- Disaster recovery
- Business continuity
- High availability
- Compliance
- Governance
- Data privacy
- Data security
- eDiscovery

How all these pieces of the data protection puzzle fit into a comprehensive data protection framework is the subject of this book.

1.3 What Data Protection Is

Data protection is mitigation of the risk of loss of or damage to an enterprise's data. That loss can take many forms. One is physical loss of the data itself, either temporarily or permanently. Another is the loss of confidentiality of sensitive data. Still another is loss of the ability to be able to use the data because of a loss of access to the data for any reason or a loss of responsiveness in which the data cannot be retrieved for use (even if it is technically *available*) within a reasonable period of time.

Data protection, as a *have-to* function, means that it is a cost of doing business, and not a *want-to* function, which directly carries out the mission of any organization. This means that managing the costs of data protection is important, since spending more money on data protection generates fewer profits for for-profit businesses or requires more tax dollars for governmental organizations. However, data protection can be thought about in a different way than most other cost functions.

Think of data protection as an insurance policy. In that sense, the aim of data protection is not to maximize profits or revenues, or to minimize costs, but to minimize worst-case losses. Like other insurance, data protection insurance is a necessary cost of the prudent business, and it balances the costs of unplanned outages against the costs of the insurance policy. A side effect of data protection may be more cost-effective use of information assets; but users should not require profits from their data protection solutions, any more than from their life insurance policies on key executives.

Unlike the traditional insurance markets, the data protection market offers no "third-party" insurers (with the possible exception of Lloyd's of London). Enterprises are "self-insured" today, and should expect to be self-insured tomorrow. Insurance "premiums" are paid internally, in the form of additional hardware, software, and people. One principle remains the same, however: When payment is made for data protection insurance, the goal is to minimize its cost and maximize its value.

One principle remains the same, however: When payment is made for data protection insurance, the goal is to minimize its cost and maximize its value.

As noted above, data protection seeks to ensure not only the preservation and availability of data, but also its confidentiality, privacy, and availability to regulators. This is still insurance—the legal costs of failure to protect confidentiality and privacy, or to fail to supply appropriate information to regulators, are high, as are the competitive disadvantages of leaking proprietary information.

1.4 Data Protection Has to Be Placed in the Right Framework

Businesses are actively examining how to improve the data protection function from the perspectives of people, processes, and technology. And many data protection technologies, both old and new, are vying for attention as enablers of data protection processes. Trying to sort through the myriad of choices can be difficult.

The key to choosing any of these technologies is understanding the overall context, the overall "data protection infrastructure portfolio," into which individual data protection technologies should fit. Otherwise, what appear to be individually sound decisions may not lead to the necessary levels of data protection. Among the problems that can occur are

- Failure to protect data adequately, which can lead to negative consequences, such as loss of revenue from lost customer orders.
- Making the wrong allocation decision (spending too much on areas that do not really require that level of protection and too little on areas that require greater protection)
- Straining the administrative resources assigned to data protection even further and with less results than necessary

Without the right model, enterprises cannot know where to place their longer-term data protection technology investment bets or how much they should place on each bet. And that means that any model has to take into account the changing world of data protection technology.

1.5 Evolving to the Governance, Risk Management, and Compliance Framework

Data protection means many things to many people. Yet what is data protection really, and what does it cover? The depth and breadth of data protection can be daunting. Exploration of data protection starts with defining the first principles of data protection and then expanding to get a more detailed and comprehensive view. That process starts with traditional risk management but eventually moves on to include the compliance and governance-related aspects of data protection.

Getting to an overall understanding of the breadth and depth of data protection was an evolutionary process. Finding a concept that offered a way of tying the pieces of the data puzzle together was necessary. That organizing principle would simplify thinking about data protection at the highest level and then allow a drill-down to deeper levels of understanding.

The organizing principle that eventually seemed to fit the best was built around the concepts of governance, risk management, and compliance (GRC). The most visible advocate of GRC is the Open Compliance and Ethics Group (OCEG). OCEG has promulgated the concepts of governance, risk management, and compliance from a corporate perspective. OCEG promotes what it calls *principled performance,* so it is a strong advocate of businesses operating with the highest ethical standards.

How the general concept of GRC applies to data protection has been independently derived, but hopefully the application to data protection with the overall goal of proper conduct by all organizations is consistent with the broader corporate perspective.

1.6 Ride the Sea Change in Data Protection

Change that affects the requirements for data protection is coming from several directions. One of the directions is extending and improving what is already being done. An example of this from a technology perspective is disk-to-disk backup that improves on the traditional backup/restore process.

A second direction is change in the basic way that the movement and storage of information is carried out in an organization. For example,

information lifecycle management (ILM) is not only about moving information from one tier of storage to another, but also about managing stored information differently—and a major effect of the difference in information management is in better data protection. Moreover, ILM leads to an overall change in the mix of data protection technologies (e.g., data replication versus data backup) that are used within an enterprise.

A third direction of change comes about from changing business requirements. A key illustration is a new emphasis on business-governance/compliance policies, which require organizations to understand and implement new policies, processes, procedures, and practices as well as possibly new hardware and software data protection technologies.

The rest of this book examines the basic principles of data protection in light of these changing business requirements and in light of existing and emerging data protection technologies. The key takeaways that should be kept in mind are these:

1. Determine where overinvestment and underinvestment in data protection are taking place, so future investments can be directed to shore up the weak spots.
2. Determine what the effects of changing business requirements and technology advances on the data protection investment are.
3. Gain a sense of how the major categories of data protection technologies interact, so that a determination of the proper mix and delivery of the proper level of service can take place.

1.7 How to Read This Book

The starting point for understanding data protection is risk management. Chapters 2 through 6 build the story of data protection from a risk management perspective.

Chapter 2 starts off the exploration of data protection with a familiar subject—business continuity as part of risk management. Disaster recovery and operational recovery are the two key components of business continuity. A key distinction is made between logical data protection (such as protecting against data corruption) and physical data protection (such as protecting against the failure of a storage device). Chapter 3 uses a simple matrix of the disaster-operational-physical-logical first principles as a reference point in describing where key problems of data protection lie for business continuity. Chapter 4 discusses how the concept of high availability is

important for data protection, but that there are three other primary objectives—preservation, confidentiality, and responsiveness—that have to be met as well. Chapter 5 introduces the need to have multiple degrees (or layers) of data protection to prevent failures from destroying the ability to protect data.

Chapter 6 introduces how information lifecycle management changes the data protection game dramatically, because ILM leads to the need for active archiving. Active archiving not only affects what data is stored where, but how different data is managed differently, such as for data retention purposes.

Chapter 7 on compliance and Chapter 8 on governance introduce the two other pillars of the GRC framework. Chapter 8 also shows how the data protection objectives match up with each of the GRC responsibilities.

Chapter 9 expands earlier mentions of data retention into the greater depth and detail that is necessary to discuss this pivotal issue in data protection.

Chapter 10 gives a data security perspective of data protection. Data security is integral to data protection. Many data protection issues are often viewed under the rubric of data security. This chapter focuses on the issues related to the loss of confidentiality for sensitive information, including data privacy and encryption, but also touches on a number of other topics, including information assurance and nation-state attacks.

Chapters 11 through Chapter 15 focus on data protection technologies, primarily from a risk management perspective. This part of the book includes Chapter 12 on traditional technologies, such as backup/restore software. Chapter 13 discusses technologies that do not perform data protection functions directly, but that support the ability of data protection to work better and more efficiently, such as data deduplication, WAN acceleration, and disaster recovery testing. Chapter 14 describes how disk and tape technologies complement and compete with each other, including virtual tape libraries. Chapter 15 covers high-availability and low (or no)-data loss technologies, including point-in-time copying, continuous data protection, and replication technologies.

Chapter 16 discusses the special technology requirements for compliance, governance, and data security, and Chapter 17 covers the importance of eDiscovery for civil litigation for the governance pillar of the GRC framework.

Chapter 18 dwells on the issues surrounding the impact of the use of third-party services in conjunction with data protection. This impact is growing in importance. Notably, cloud computing, software-as-a-service, and storage-as-a-service take center stage in this discussion.

Chapter 19 covers a number of other considerations that have to be taken into account when performing data processing. The role of tiering in data protection, from flash computing to tape, is an important consideration. So is the impact of server and storage virtualization on data protection. Interestingly, better data protection can lead to better overall information management, such as master data management, which can yield benefits derived from the ability to use information more effectively. And, of course, the role of data protection in green computing deserves attention.

Chapter 20 describes a kick-start planning model to help businesses get started in the planning process to improve their data protection as well as summing up and giving suggestions on redesigning data protection.

1.8 An Aside on Process Management

Although data protection technologies are an important part of the overall data protection picture, data protection is much more than a collection of technologies. Technology is not a *deus ex machina.* That is, users should not expect technology to fall from the sky and magically lead to the design, implementation, and ongoing carrying out of the activities that exemplify chosen data protection strategies. Instead, the 4 Ps of process management—*p*olicy, *p*rocess, *p*rocedures, and *p*ractices—have to be put in place. (Technology enables the 4 Ps, but it does not replace them.) *Policy* defines a course of action but does not actually carry out the necessary actions. *Processes* are the actions that are necessary to reach the ends directed by a policy; they make the policy actionable. *Procedures* define the steps in any process. *Practices* ensure that the procedures with the processes that fulfill a policy are actually carried out.

Each of the 4 Ps requires conscious effort and thought on the part of any business for each of the pieces of data protection. Think of what needs to be done—who, what, where, how, and when—for each aspect of data protection separately and integrated as a whole. Throughout this book, think how people need to use the 4 Ps to ensure the proper use of technology.

1.9 Key Takeaways

- Protection of electronically stored information is essential for an organization, to meet not only risk management requirements, but also those of compliance and governance.
- Data protection is the self-insurance policy that an organization takes to mitigate the risk of loss (in a number of ways) of its data.

- If data protection is not set in the right framework, organizations are exposed to consequences from the failure to protect data adequately, misallocation of funds spent on data protection, and unnecessarily high costs to administer data protection.
- Change in data protection is coming about because of new business requirements, new and evolving data protection technologies to meet those business requirements, and a change in the basic way that information is moved and managed. Together the changes amount to a sea change that organizations have to align themselves with in order to avoid being swamped.