IBM Software

Business-driven data privacy policies

Establish and enforce enterprise data privacy policies to support compliance and protect sensitive data

IBM

Contents

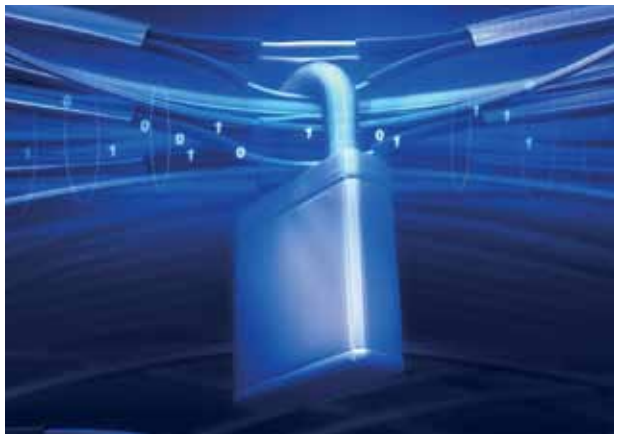
Introduction

Enhancing information security is no longer solely a task for IT—it's a strategic, enterprise-wide priority. According to interviews conducted by IBM with 138 security leaders, two-thirds of senior executives are paying closer attention to enterprise data privacy than they were two years ago.[1]

Recent statistics on data security breaches and data loss help explain why. In 2011, the corporate world experienced the second-highest data loss total since 2004.[2] A 2012 study conducted by Symantec calculated that global cybercrime cost USD114 billion annually and claimed more than one million victims per day.[3]

In response to growing threats, organizations are increasing their security budgets. Two-thirds of security leaders expect spending on information security to rise over the next two years. Of those, almost 90 percent anticipate double-digit growth. One in 10 expects increases of 50 percent or more.[4]

Protecting data privacy is no longer optional—it's the law. Failure to comply with privacy regulations can dramatically impact a business. Corporations and their officers could be fined from USD5,000 to USD1 million per day. Officers could be sentenced to jail time if data is misused. According to a study conducted by the Ponemon Institute, the average organizational cost of a data breach in 2011 was USD5.5 million.[5]



Monetary penalties and jail time are just two examples of how organizations can be affected by data privacy failures. Data breaches can foster investor concern and negative publicity, leading to the erosion in a company's share price. But perhaps most alarming is the risk of irreparable brand damage if investors, customers and partners believe that the company cannot be trusted.

Organizations must have procedures in place to protect data in databases, applications and reports in both production and nonproduction systems to comply with data privacy regulations and avoid damages.

Establish a holistic data protection strategy

Different types of data have different protection requirements, so organizations must take a holistic approach to safeguarding information no matter where it resides. This approach includes:

- **Understanding where the data exists.** Organizations cannot protect sensitive data unless they know where it resides and how it is related across the enterprise.
- **Safeguarding sensitive data, both structured and unstructured.** Structured data contained in databases must be protected from unauthorized access. Unstructured data in documents and forms requires privacy policies to redact (remove) sensitive information while still allowing required business data to be shared.
- **Protecting nonproduction environments.** Data in nonproduction, development, training and quality-assurance (QA) environments needs to be protected yet still usable during application development, testing and training processes.
- **Securing and continuously monitoring access to data.** Enterprise databases, data warehouses and file shares require real-time insight to ensure data access is protected and audited. Policy-based controls are required to rapidly detect unauthorized or suspicious activity and alert key personnel. In addition, databases and file shares need to be protected against new threats or other malicious activity and continually monitored for weaknesses.
- **Demonstrating compliance to pass audits.** It is not enough to develop a holistic approach to data security and privacy. Organizations must also demonstrate compliance and prove it to third-party auditors.

IBM® InfoSphere® solutions for data security and privacy are designed to support this holistic approach to data protection. They incorporate intelligence that enables organizations to proactively address IT threats and enterprise risks.

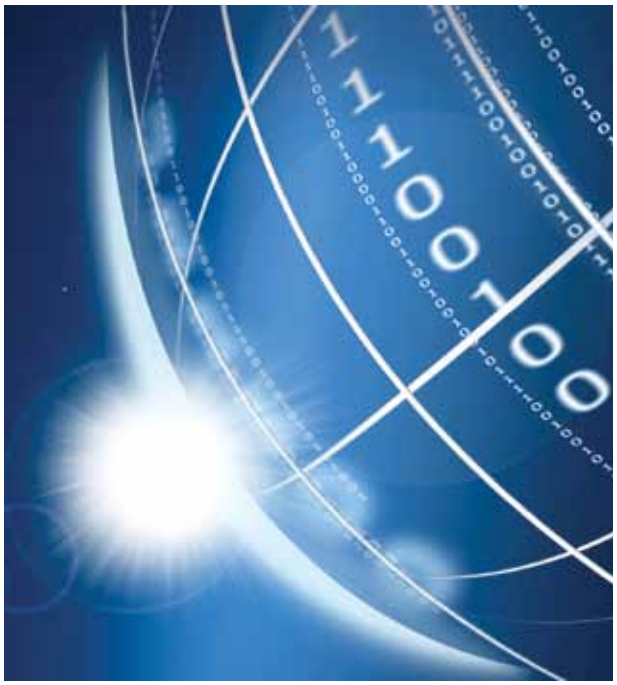
Define data privacy policies to support business goals

A policy-driven, on-demand masking approach enables organizations to proactively protect data privacy and support compliance. This approach is especially important in the current era of computing, where data is everywhere and is continuously growing in volume, variety and velocity.

A focus on privacy fundamentally changes how big data platforms are adopted. The end goal is to provide aggregated sensitive data to an analytics platform while protecting privacy. By selecting big data platforms that integrate intelligent data masking, organizations can conduct analytics while keeping private information out of sight.

Policy-driven, on-demand data masking should be built into an overall information integration and governance platform. Information integration and governance policies should be well defined and championed by the business owners responsible. In addition, they should be easily accessible to the people and processes responsible for implementing and enforcing them.

Persistent, effective governance is no longer optional. Compliance drivers now require organizations and their executives to demonstrate meaningful protocols and processes to guide organizational decision making while also ensuring data is available to support real-time decision making. The expectation is that all data is available at any time for analytics.



Today's organizations must also take a hard look at how to embed meaningful governance principles into their daily operations, processes and culture. Governance needs to be consistent and flexible enough to adjust to an organization's business, but the overall framework and its principles should endure over time.

Let's look at three steps required to set up data privacy policies.

Step 1: Define sensitive data

Effective data privacy begins with an agreement that outlines the purpose, accountabilities and participants in the data privacy strategy. Not all data has to be protected in the same way; some may be considered low risk and not worth the time and effort required to secure it.

The first step is to define sensitive data—but defining it is not solely an IT function. A cross-functional team including marketing, sales, line-of-business (LOB), operations and IT should work together to create the definition. Keep in mind that high-value data such as design specifications or corporate secrets might not require protection under legal mandates, but organizations will most certainly want to protect it with stringent privacy controls. The team should decide what data to protect based on business priorities.

The end goal is to create a clear definition of what needs to be protected in light of regulatory mandates. The questions that follow provide an example of the considerations that should guide this part of the definition phase:

- What data is considered sensitive?
- What constitutes a corporate secret?
- What are the components of personally identifiable information?
- What is the definition of high-risk data?
- What data is affected by legal mandates and what is not?
- Where is sensitive data copied across the enterprise?
- Is sensitive data being shared with third parties or outsourced for testing, development or QA work?
- Who has a valid business need to know sensitive data?

The outcome should be a business glossary—an authoritative dictionary of data privacy terms and relationships employed across the enterprise. Designed to be accessible across the enterprise, the business glossary defines the terms used to build enterprise data privacy policies. All employees can leverage this central source for standard definitions of sensitive data, which helps remove reactionary processes and guesswork.

Step 2: Understand where sensitive data resides

After defining what should be protected, the cross-functional team should determine where that sensitive data resides across the enterprise. Most of the world's data is stored in commercial databases or data warehouses, such as those based on IBM DB2®, IBM Informix®, Oracle Database, Microsoft SQL Server, Sybase, MySQL, IBM Netezza® and Teradata solutions.

However, many organizations do not have a complete understanding of their enterprise data stores. They rely too heavily on system and application experts for this information. A deeper understanding of enterprise data and data relationships is required to protect it.

In addition, organizations tend to neglect non-production environments. What sensitive data is replicated and used for testing, development, QA, training or demonstration purposes? Understanding sensitive data means understanding where data is copied and identifying which data has left the premises.

Organizations should consider an automated process to identify data and data relationships since this can take months of manual analysis—with no assurance of completeness or accuracy.

Step 3: Mask sensitive data

Data masking is the process of systematically transforming sensitive data elements, as defined and identified in steps 1 and 2, into realistic but fictionalized values. Masking enables recipients of the data to use "production-like" information while ensuring compliance with privacy protection rules (see Figure 1).

Data masking is a simple concept, but it is technically challenging to execute. Many organizations operate complex, heterogeneous IT environments comprising multiple, interrelated applications, databases and platforms. A data masking solution must consistently and effectively mask sensitive data across related data sources while also preserving application and database integrity.

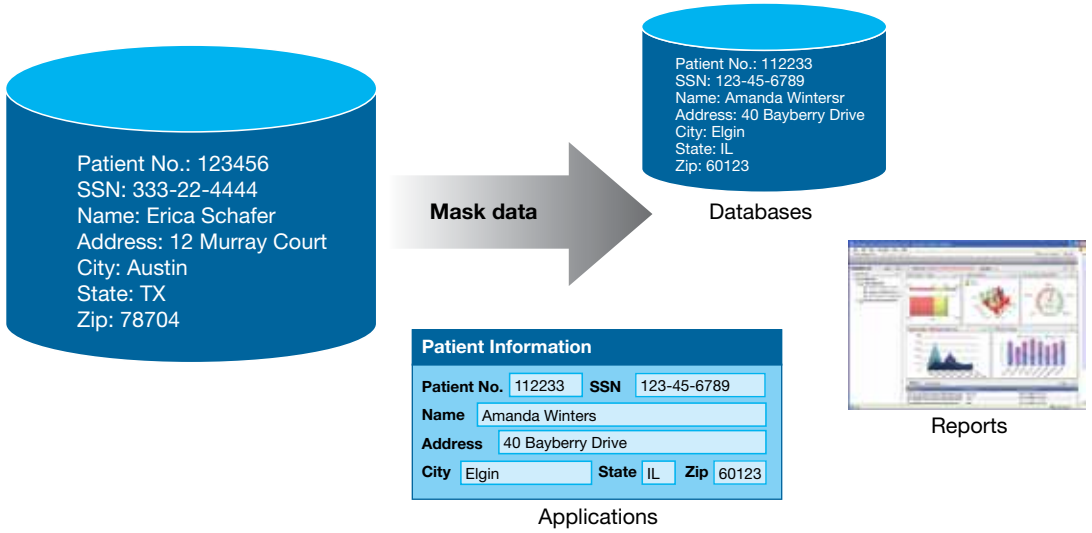


Figure 1: Data masking transforms sensitive data (such as customer information) into fictionalized values to protect privacy

Create and enforce enterprise-wide privacy policies with IBM InfoSphere Optim

IBM InfoSphere Optim™ Data Privacy is designed to standardize data protection and validate data privacy policies throughout the data lifecycle. The solution provides required capabilities to define sensitive data, determine where data resides and mask sensitive information. With InfoSphere Optim Data Privacy, organizations can maintain compliance by establishing enterprise-wide policies that span distributed, heterogeneous systems and can be accessed by teams across the enterprise.

InfoSphere Optim Data Privacy supports standard ways to define policies, execute privacy policies, validate compliance and report results, including:

- **Naming**—Use standard words, acronyms and naming patterns.
- **Meaning**—Associate words with shared meaning through business glossaries.
- **Values**—Define appropriate values or ranges for attributes.
- **Privacy**—Specify standards for masking rules and associate them with specific attributes.

InfoSphere Optim is a key part of the IBM InfoSphere portfolio

IBM InfoSphere software is an integrated platform for defining, integrating, protecting and managing trusted information across systems. The InfoSphere platform provides the foundational building blocks of trusted information, including data integration, data warehousing, master data management and information governance, all integrated around a core of shared metadata and models.

The portfolio is modular. Organizations can start anywhere, and then mix and match InfoSphere software building blocks with components from other vendors. Or they can deploy multiple building blocks together for increased acceleration and value. The InfoSphere platform offers an enterprise-class foundation for information-intensive projects, providing the performance, scalability, reliability and acceleration needed to simplify difficult challenges and deliver trusted information faster.

InfoSphere Optim Data Privacy enables organizations to determine what information to mask by establishing enterprise-wide rule definitions and identifying the proper business content for masking policies. For example, companies can mask sensitive data effectively across applications, reports and databases in both production and nonproduction environments. Data masking capabilities de-identify data anywhere a contextually accurate yet fictionalized value is appropriate, such as:

- **Masking data in development, Q/A and testing environments** to enable use of production-like data without jeopardizing privacy
- **Masking data on demand** for applications and business report environments to support real-time decision making
- **Masking data in extract, transform, load (ETL) or data movement solutions** to maintain compliance in data warehouse environments

By enabling data masking in several distinct use cases, InfoSphere Optim Data Privacy can deliver a variety of important business benefits (see Table 1).

	Use case	Business benefits
Masking for databases	Protect data across both nonproduction environments (test, development, Q/A and training) and in production	• Ensure only those with a valid purpose see sensitive data • Protect data from misuse by outsourced personnel or third parties
Masking for warehouses	Protect data during the ETL process or while testing data integration code	• Deliver security and ensure compliance in data warehouse environments
Masking for reports	Protect sensitive data in reports without inhibiting business processes	• Distribute reports across teams to facilitate information sharing without compromising security
User-defined masking routines	Mask data in applications	• Deliver masking to production environments
Masking for Hadoop-based systems	Protect sensitive data as it moves into and out of Hadoop-based systems	• Build security into Hadoop-based systems from the start

Table 1: Benefits of masking data across the enterprise

Data masking: A real-world example

Large US-based insurance company protects privacy in Oracle PeopleSoft application

A large US health insurance company offers HMOs, PPOs, point-of-service (POS) plans, health savings accounts and traditional indemnity coverage, along with dental, vision, behavioral health and Medicare plans. **With more than 15 million health plan members, 13 million dental plan members and 10 million pharmacy members, the company manages, processes and is responsible for a greater amount of sensitive data than most organizations.**

Like health insurance companies around the world, this company needs to comply with government laws aimed at protecting its members' personal information. Operating in the United States, the insurer must demonstrate compliance with the US Health Insurance Portability and Accountability Act (HIPAA) and pass HIPAA audits.

To protect sensitive data in accordance with HIPAA, the company needed to better understand where sensitive data resided and to identify which types of sensitive data were contained in its vast data sources. The company also needed to protect specific data types that were part of an Oracle PeopleSoft implementation.

Using InfoSphere Optim Data Privacy, the insurer was able to discover and mask the following sensitive data types:

- Credit card information (CCI)
- Personal health information (PHI)
- Personally identifiable information (PII), including Social Security numbers, addresses, ages and other member information

InfoSphere Optim Data Privacy helped the company establish consistent, enterprise-wide data masking policies for both custom applications and the Oracle PeopleSoft application. Maintaining integrity in the Oracle PeopleSoft application was essential because it services 30 downstream applications. InfoSphere Optim ensured consistent masking across these systems and maintained application and database integrity during the masking process.

A key to the company's success was forming a cross-functional team to establish the right masking policies. The group—which included project managers, customer subject-matter experts (SMEs), database administrators (DBAs) and business analysts—made sure the right masking policies were implemented in accordance with business objectives. IBM mentored the company's IT group throughout the project, leaving the group with the skills and tools needed to maintain and use the new system.

Accommodate more data while helping to ensure privacy

As the volume, variety and velocity of data increases, the importance of securing data privacy will increase as well. The potentially devastating business impact of data breaches is a strong motivator, pushing organizations to invest in new data privacy solutions.

InfoSphere Optim Data Privacy enables organizations to address the challenges of securing data privacy with a holistic, business-driven data privacy strategy. Organizations can define sensitive data, understand where it resides and mask sensitive data across a wide variety of environments, from databases, warehouses and big data environments to business report applications and nonproduction environments. By helping to protect data privacy, InfoSphere Optim Data Privacy plays an important role in facilitating compliance and minimizing business risks.

Resources

To learn more about data privacy and protection, as well as more about IBM InfoSphere Optim Data Privacy, visit:

- **ibm.com**/software/data/optim/protect-data-privacy
- **ibm.com**/software/data/information-governance/overview.html

[1] IBM Center for Applied Insights, "Finding a strategic voice: Insights from the 2012 IBM Chief Information Security Officer Assessment," May 2012, http://public.dhe.ibm.com/common/ssi/ecm/en/cie03117usen/CIE03117USEN.PDF

[2] Verizon Business, "2012 Data Breach Investigations Report," www.verizonbusiness.com/resources/reports/rp_data-breach-investigations-report-2012_en_xg.pdf

[3] "Norton Study Calculates Cost of Global Cybercrime: $114 Billion Annually," Sept. 7, 2011, www.symantec.com/about/news/release/article.jsp?prid=20110907_02

[4] IBM Center for Applied Insights, "Finding a strategic voice."

[5] Ponemon Institute LLC, "2011 Cost of Data Breach Study," March 2012, www.symantec.com/content/en/us/about/media/pdfs/b-ponemon-2011-cost-of-data-breach-us.en-us.pdf?om_ext_cid=biz_socmed_twitter_facebook_marketwire_linkedin_2012Mar_worldwide__CODB_US

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
February 2013

IMM14108-USEN-00