Data Privacy Compliance in the Application Testing Environment

Data Masking and Transformation for Privacy Compliance

WHITE PAPER

October 2006

Contents

Disclaimer

Privacy laws and regulations are the subject of continuous modification in many different jurisdictions and for many different types of data. This document is intended to provide general background information, not regulatory, legal or other advice. Princeton Softech, Inc. cannot and does not provide such advice. Readers are advised to seek competent assistance from qualified professionals in the applicable jurisdictions for the types of services needed, including regulatory, legal or other advice.

(Ref. no: 27727-3)

Why is Data Privacy a Hot Topic?

As technology evolves, more and more information is placed at our fingertips. Database and Internet capabilities allow us to access information on any topic, at any time, from any place. Organizations across industries depend on the accessibility of this information to collect customer data and gain an edge in a competitive market.

While most organizations access confidential personal information for legitimate purposes, such as providing healthcare and financial services, increasing security risks challenge a company's ability to manage and protect that data successfully. Identity theft, privacy violations and fraudulent access continue. The frequency and magnitude of these breaches has heightened social awareness of an invasive trend that undermines business practices. So what are the appropriate standards for managing sensitive customer information?

As more companies recognize the need for data privacy, and customers demand increased protection, government entities are passing increasingly stringent laws and regulations to ensure that safety. Every company must be responsible for protecting personal customer data to comply with information governance regulations and to gain the trust of customers and business partners.

Data Privacy Receives Global Attention

Data privacy is a global issue that crosses applications, databases, operating systems and platforms. The following are examples of international industry standards and data privacy legislation:

Industry	**Privacy Standard**
Banking / Credit Card	Payment Card Industry (PCI) Data Security Standard
Country	**Privacy Legislation**
Australia	Privacy Amendment Act of 2000
Canada	Personal Information Protection and Electronic Documents Act
European Union	Personal Data Protection Directive of 1998
New Zealand	Privacy Act of 1993
Hong Kong	Hong Kong Personal Data (Privacy) Ordinance of 1995
United Kingdom	Data Protection Act of 1998
United States	Gramm-Leach-Bliley Act of 1999 Health Insurance Portability and Accountability Act of 1996

Additional information on privacy legislation is provided in the Appendix.

Although the specifics may differ, these data privacy laws have a number of common elements. First, the laws are designed to protect individuals against the misappropriation and misuse of personal information. Second, the data privacy laws are complex. Compliance often requires changes to corporate policies and operating procedures, as well as the adoption of new technologies. Therefore, organizations are generally given time and grace periods to become compliant.

Finally, although enforcement may focus on education and remediation in the early stages, the laws generally impose substantial penalties for non-compliance – especially in cases of criminal misconduct.

How Does Data Privacy Compliance Affect Your Business?

Information governance is an important business initiative. Enterprise data management strategies allow you to align continuous control of application data with your business objectives to satisfy requirements for data privacy, security and retention. Companies must also apply the data protection practices that are appropriate for each stage in the information lifecycle.

Protecting data privacy is becoming an important competitive advantage that reinforces sound business practices. The Payment Card Industry (PCI) Data Security Standard is a good example. This Standard evolved in response to the overwhelming occurrences of data and identity theft, as well as fraud resulting from the misappropriation of credit card numbers.

Corporations must protect all sensitive information, whether it resides in a production system, development database or elsewhere within the enterprise. Furthermore, the same protective measures cannot be simply replicated across every environment because the methods that protect data in production may not necessarily meet the unique requirements for application testing.

How Does Data Privacy Apply to Application Testing?

Many professionals within a company access and utilize confidential customer data for legitimate business purposes. The challenge is to provide appropriate protection, while meeting business needs.

One of the most common methods for protecting data privacy in the test environment is to clone, or copy, the production database and relocate the test environment to a secure area. Security measures, such as physical access controls or network security, limit unauthorized access to application data. While these measures provide some degree of privacy protection in the testing environment, they are not sufficient.

In fact, using production data to create realistic test databases could result in a violation of privacy laws. The effects of this type of violation can be devastating for a company; resulting in negative publicity and erosion of customer confidence at a minimum, and fines or even imprisonment at the maximum. Industry analysts agree that a more effective solution would be to apply de-identification techniques when migrating data from production into a test environment.

De-Identification Techniques Provide Best Practices Solution

De-identifying test data is the process of systematically masking or transforming data elements that could be used to identify an individual. Data that has been scrubbed or cleansed in such a manner is generally considered acceptable to use in the test environment. Data de-identification enables developers and testers to use realistic test data and produce valid test results, while still complying with privacy protection rules.

However, it is important to note that the results of the data transformation have to be appropriate in the context of the application. That is, the results of data transformation must make sense to the person reviewing the test results.

Requirements for Protecting Relational Test Data

Testing activities occur throughout the application lifecycle. Therefore, companies must have procedures in place to thoroughly test applications and still comply with data privacy regulations. This challenge becomes more difficult when the application relies on relational databases. Deploying proven solutions that make de-identifying test data easy and cost effective is essential to privacy compliance.

Protecting Data Privacy in the Testing Environment

Effective privacy protection strategies ensure the confidentiality of private information and improve the security of your test and development environments. It is imperative to implement a solution that masks and transforms data so that it can no longer be used to identify an individual. In addition to masking techniques, you also need capabilities that incorporate site-specific data transformation routines, integrating the processing logic from multiple related applications, databases and platforms.

De-identified data is safe and valid to use in application testing. It still makes sense to quality assurance staff and still produces accurate, reliable test results – but is worthless to thieves and hackers. Masking confidential test data allows you to minimize disclosure risks – and force hackers to turn elsewhere.

Requirements of a Comprehensive Solution

Delivering thoroughly tested applications, while protecting privacy, requires testing capabilities that allow you to:

- Extract subsets of related data to improve productivity. Users can create and reuse realistic and manageable, “right-sized” test databases.
- Apply comprehensive data masking and transformation capabilities that allow you to de-identify or de-personalize test data, so you can test with realistic data and still protect privacy.
- Propagate masked data elements across related tables within a database, as well as across applications, databases and platforms. This capability maintains the referential integrity of your test data and ensures the validity of your test results.
- Automate comparison processing to quickly analyze differences in test results and identify problems that would otherwise go undetected.

Meeting Compliance Challenges

Princeton Softech Optim™ enables organizations to meet even the most complex application testing challenges by providing the fundamental components of effective test data management, including the capabilities to:

- Extract referentially intact subsets of data with 100 percent accuracy to create realistic test databases, no matter how many tables or relationships are involved.
- De-identify data to mask or exclude personally identifying information and comply with privacy regulations, while still providing realistic data that yields accurate test results. Apply a variety of data transformation techniques to meet specific test case requirements, and even extract and mask data in a single process.
- Propagate key masked data elements to all related tables to preserve the integrity of data relationships.
- Integrate test data from multiple database management systems (Oracle, DB2, DB2/UDB, Sybase, SQL Server and Informix) and platforms (mainframe and open systems).
- Integrate relational data (DB2) and hierarchical data (IMS, VSAM, Sequential Files) into the test environment.

Data Masking and Transformation Options

Optim provides a variety of data transformation capabilities to mask or de-identify data. The method you use will depend on the type of data you are masking and the result you want to achieve. For example, it is easy to mask customer identification numbers by simply applying a random or sequential number function, or replacing patient names with a predefined string of text. More sophisticated capabilities include a built-in Transformation Library, as well as user-defined transformation programs that provide greater flexibility to satisfy even the most complex masking requirements.

Key Propagation

If a masked data element (for example, a telephone number) is a primary or foreign key in a database table relationship, then this newly masked data value must be propagated to all related tables in the database. Key propagation ensures the referential integrity of the transformed data. Without key propagation, the relationships between parent and child tables would be severed causing the test data to be inaccurate. As a result, application tests will produce unreliable results. Optim provides full support for key propagation.

Data Transformation Techniques

Princeton Softech Optim includes table and column mapping capabilities that make it easy to mask and move data from a source to a destination database. A Table Map identifies and correlates a source table with a destination table. A Column Map provides a way to control processing data on a column-by-column basis. Specifically, you indicate how data in the source column is to be masked or transformed as it is written to the destination column.

Masking Character and Numeric Data

Masking character or numeric data provides one of the most basic examples of substituting confidential data with de-identified data that is still meaningful in the context of the application test. Some of the masking functions available with Optim are explained in the following paragraphs.

String Literals and Functions

A string literal specifies a value for a destination column that contains alphanumeric data. You can define a string literal using any combination of characters.

The Random function returns a value selected at random from within a range of user-specified values. This function can be used with character or numeric data, and can be especially helpful in de-identifying personal identification numbers, account numbers, telephone numbers, salary information and so on.

The Sequential function returns a value that is incremented sequentially. You specify the start value and the incremental step value. This masking function can be used with character or numeric data.

The Substring function returns a substring or portion of the content of a column. You specify the name of the column, the position of the first character in the string and the number of characters to use. This masking function provides precise control for de-identifying data.

Concatenated Expressions

Optim allows for masking data using concatenated expressions. These expressions enable you to define the value of a destination column by combining the values of two or more source columns or by combining a column value with some other value.

For example, suppose that bank account numbers are formatted '999-9999,' where the first three digits represent the type of account (checking, savings or money market, etc.) and the last four digits represent the customer identifier. Here, you could mask the account number by concatenating a substring using the first three digits of the actual account number with a 4-digit number derived using the sequential function.

In this example, concatenation allows you to retain the correct format of the account number column, preserve important information about the type of account and at the same time, de-identify the confidential customer information. The result is a fictionalized account number that is still valid in the context of the application test.

Masking Data Using Lookup Values

Another approach to de-identification is to transform data using lookup values. A lookup table maps the value in a source column to a corresponding value for the destination column. For example, a lookup table might transform medical diagnostic codes into fictionalized codes for testing purposes.

You can use the Lookup function to locate the desired source value in the lookup table and return the corresponding value to populate the destination table. The Random Lookup function selects values at random from a specified lookup table to insert in a destination column.

Several built-in lookup tables increase the ease of using the Lookup function for masking data:

- First Names Lookup – Contains more than 5,000 first names for de-identifying personal information.
- Last Names Lookup – Contains more than 80,000 last names for de-identifying personal information.
- Street Address/City/State/ZIP Code Lookup – Contains more than 100,000 US locations to mask complete address information.

An enhanced Random Lookup function makes it easy to transform data in any or all columns of a row in a destination table by replacing it with an entire row of data randomly selected from a lookup table. For example, instead of substituting one ZIP code for another, this feature makes it possible to mask an entire Street Address/City/State/ZIP.

Performing Custom Data Transformations

Optim's Data Privacy Transformation Library supports generating valid, masked values to de-identify social security numbers, credit card numbers and e-mail addresses:

- Social Security Numbers – Generates valid, transformed numbers that follow the rules used by the US Social Security Administration.
- Credit Card Numbers – Generates valid, transformed numbers, based on rules used by credit card issuers.

- E-Mail Addresses – Generates valid, transformed e-mail addresses using string literals or the first/last name columns and the domain.

When you need to perform custom data transformations, you can prepare a user-defined exit routine. These are simply programs or sets of instructions that perform the desired data transformation routine.

Exit routines are especially useful for generating values for destination columns that could not be defined using any other method. For example, a tester may create a customized masking algorithm to generate a value for the Customer ID code, based on the customer's geographic location, average account balance and volume of transaction activity. The Customer ID code generated using this algorithm is then used to populate a destination column.

Propagating Primary and Foreign Keys

Each of the methods described so far is effective for masking data to safeguard confidentiality. However, with relational database applications, there is an added complication. Specifically, you need the capability to propagate a masked data element to all related tables in the database in order to maintain referential integrity.

Optim provides a Propagate function that lets you assign a value to a primary key or foreign key column and propagate that value to all related tables. The value you specify can be a valid column name, string literal, expression or function. For example, assume a simple data model consisting of two related tables: Customers and Orders (see Figure 1). The Customers table is parent to the Orders table and its primary key column, *Cust_ID*, is a 5-digit numeric value.

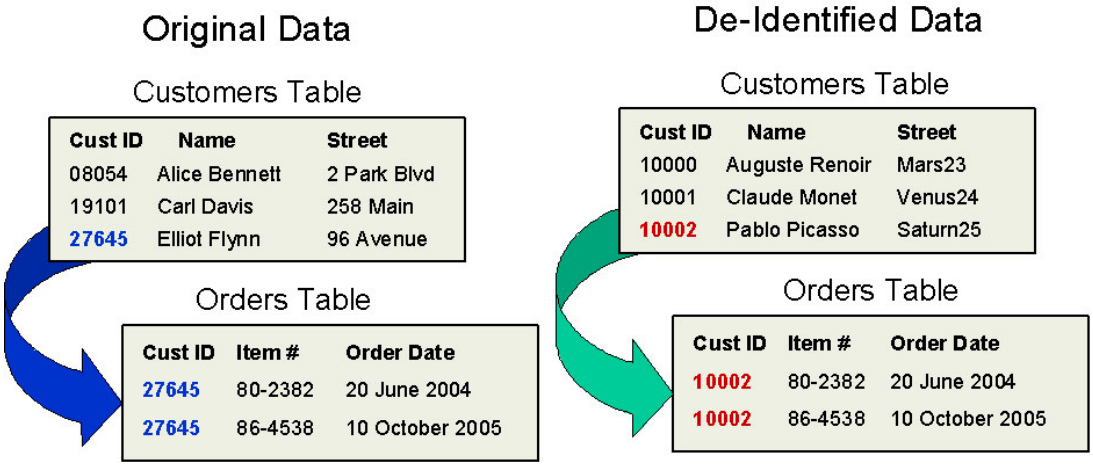

Figure 1. Optim's Key Propagation capability ensures referential integrity, even when data is masked.

If the customer identifier is masked using the sequential function, the masked values must be propagated to all related tables to ensure the referential integrity of the data.

In this example, the *Cust_ID*, *Name* and *Street* columns are masked. Notice that the name "Elliot Flynn" has been masked as "Pablo Picasso." The street name has also been masked. In particular, the sequential function was used to transform the original *Cust_ID* for Elliot Flynn from 27645 to 10002. When this masked *Cust_ID* value is propagated from the Customers table (parent) to all related tables, the key relationship between the Customers and Orders tables in the test database remains intact.

Without the ability to propagate masked values, the referential integrity of the data would be severed; creating orphan rows for the Orders table (see Figure 2).

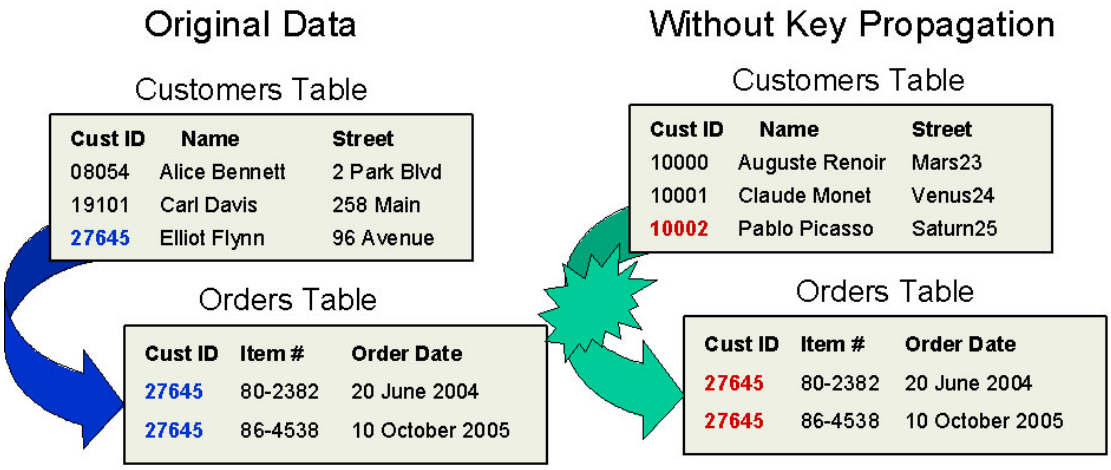

Figure 2. Without a Key Propagation capability, critical data relationships would be severed.

The ability to propagate key values ensures that the test database remains referentially intact and can produce valid test results. Imagine the complexity when there are hundreds of related tables involved, and keys must be propagated to all related tables. Without a propagate capability, many orphan tables would result and the test database would easily become corrupted.

Extending the propagation capability even further, Optim provides a new Hash-Lookup function that generates transformed replacement values for source columns and propagates the replacement values consistently and accurately across databases and platforms (see Figure 3). This data transformation function provides additional capabilities for protecting privacy across the entire enterprise.

Figure 3 illustrates Optim's capabilities for masking credit card numbers across databases.

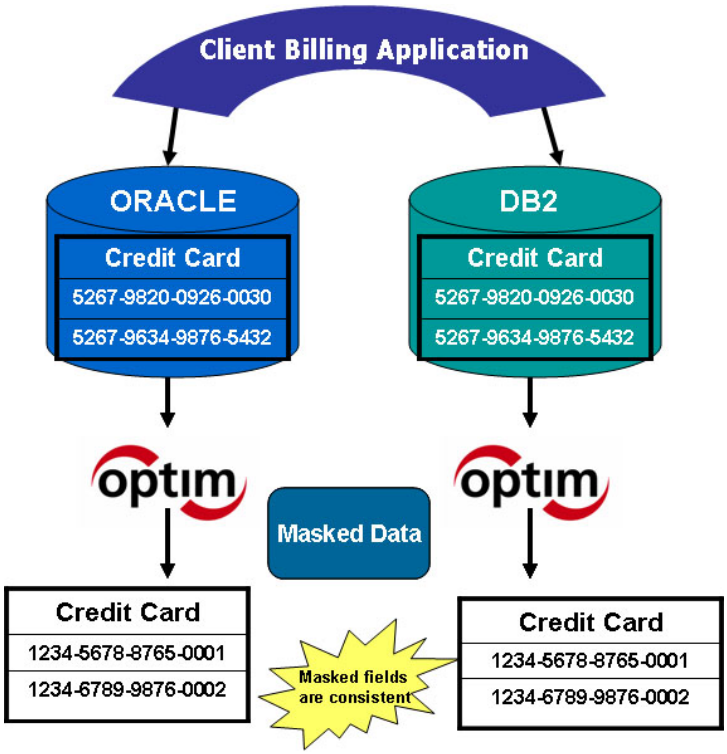

Figure 3. Optim propagates masked replacement values consistently and accurately across databases and platforms.

Best Practices Summary

The need to protect the privacy and confidentiality of customer data spans countries, industries and computing environments. And while many companies have implemented effective measures to protect data in the production environment, they are just beginning to turn their attention to the development and testing environments.

Protecting development and test environments poses a special challenge because the development and quality assurance teams need realistic data to accurately test their applications. Data de-identification or masking provides a means to remove "personally identifying information," while still providing accurate, realistic test data that is valid in the context of the application.

Princeton Softech Optim offers proven technology that allows you to create, mask and maintain realistic test databases using "right-sized," referentially intact subsets of related test data. Optim provides a variety of data transformation algorithms and built-in lookup tables, and even supports custom data masking routines.

Using the Data Transformation Library you can generate and propagate valid, masked values for social security numbers (SSNs), credit card numbers and e-mail addresses to protect privacy, while ensuring testing accuracy. Most importantly, you can propagate masked data elements across related tables to ensure the referential integrity of the database.

Optim supports the leading database management systems and provides federated access capabilities that allow you to extract and mask appropriate test data from various data sources in a single process. Optim can satisfy your current requirements and can easily adapt to changes in your database environment. Implementing Optim helps you comply with data privacy regulations, speed time to market and reduce development and test costs throughout the application lifecycle.

Princeton Softech: The Proven Leader in Enterprise Data Management

Princeton Softech provides enterprise solutions that align application data management with business objectives. Our industry leading Optim solution enables organizations to optimize performance, mitigate risks and control costs. Partnered with the market leaders in business technology, we deliver capabilities that scale and support your enterprise – applications, databases and platforms. More than 2,200 companies worldwide – including nearly half of the Fortune 500 – rely on Princeton Softech's proven solutions to maximize the business value of their enterprise applications and databases.

Appendix

Governments and industries are taking an active role to protect data privacy. Numerous government regulations and industry standards have been enacted to protect personal information from misuse. This appendix describes some of the details associated with the following:

- Payment Card Industry (PCI) Data Security Standard
- United States (US) Gramm-Leach-Bliley Act of 1999
- US Health Insurance Portability and Accountability Act of 1996
- European Union (EU) Personal Data Protection Directive of 1998
- United Kingdom (UK) Data Protection Act of 1998
- Canada's Personal Information Protection and Electronic Documents Act
- Australia Privacy Amendment Act of 2000

Payment Card Industry Standard

The Payment Card Industry (PCI) Data Security Standard was initiated by MasterCard International and Visa in January of 2005 and later endorsed by American Express. This standard is designed to protect cardholder information and must be implemented by all members, merchants and service providers that store, process or transmit cardholder information. Fines for non-compliance can range up to 500,000 USD per incident.

This standard covers a range of issues, such as maintaining a secure network, protecting cardholder information, managing risk, implementing control measures, monitoring test networks and more. In particular, one section outlines the standards for developing and maintaining secure applications throughout the development lifecycle.

The PCI Standard makes "best practice" recommendations, but especially warns against using real credit card numbers in the test and development environments. By de-identifying credit card numbers used in application development and test environments, Princeton Softech Optim can help member companies comply with the PCI Standard.

US – Gramm-Leach-Bliley Act

Also known as the *Financial Services Modernization Act of 1999*, the Gramm-Leach-Bliley Act (GLBA) came into effect in November 2000. This legislation enacted sweeping changes in the financial services industry, most significantly by allowing banks to operate in the securities business and vice versa. The GLBA provided new data privacy regulations in the following areas:

- Disclosure of privacy policies.
- Opt-out of information disclosures to third parties.
- Non-disclosure of account information.
- Standards for protecting the security and confidentiality of consumers' nonpublic information.

The GLBA applies to a broadly defined range of "financial institutions," starting with banks and securities firms and extending to insurance companies, auto leasing companies, credit card issuers and any other entity that is "significantly involved in financial activities."

In order to comply with this legislation, financial institution regulators, like the Federal Reserve, are required to establish standards to:[1]

- Ensure the confidentiality of consumer records.
- Protect against threats to the security of those records.
- Protect against unauthorized access to those records that could result in substantial harm or inconvenience to the consumer.

United States regulatory agencies can enforce the GLBA under the same sanctions used to regulate financial institutions today. For example, the Federal Deposit Insurance Corporation (FDIC) may enforce violations under the Federal Deposit Insurance Act. Penalties can range from 5,000 to 1,000,000 USD per day. Additional criminal penalties can apply to persons who gain fraudulent access to protected financial information.

US – Health Insurance Portability and Accountability Act

The Health Insurance Portability and Accountability Act of 1996, also known as HIPAA, provides comprehensive health privacy legislation designed to reduce fraudulent activity and streamline healthcare industry inefficiencies. This federal legislation requires the healthcare industry to make some major changes to protect patient information, focusing on three major areas:

- Administrative simplification standards for electronic transactions to define a single set of uniform electronic formats.
- Privacy standards to protect directly identifiable patient health information and personal data.

---

[1] Gramm-Leach-Bliley Act, Section 501, "Protection of Nonpublic Personal Information."

- Security and electronic signature standard to provide safeguards for patient information and to prevent unauthorized access.

HIPAA applies to all healthcare organizations, health plans, healthcare clearinghouses and those healthcare providers that manage certain financial and administrative transactions electronically.

The penalties for failure to comply with HIPAA standards could cost an organization 25,000 USD per violation. Breach of patient confidentiality is subject to fines up to 50,000 USD and a year in prison. Obtaining protected health information under false pretenses can result in fines of up to 100,000 USD and 5 years in prison. Disclosing protected health information with malicious criminal intent can produce fines of 250,000 USD and up to 10 years in prison.

EU – Personal Data Protection Directive

The EU Directive on Data Protection (DDP) of 1998 is a framework that stipulates the minimum data protection legislation EU member countries must have in place. The legislation is intended to protect the rights of EU citizens regarding the processing of their personal data.

The DDP applies to all member countries within the European Union and ultimately to all organizations that operate within the EU. Member countries are required to enact national legislation that provides at least the minimum level of data privacy protection, and may choose to exceed those standards. Any organization doing business in one or more EU countries must comply with the national data privacy legislation of each member country in which it operates.

It is significant that the DDP prohibits European organizations from distributing data to non-EU countries that do not have equivalent data protection standards. Recently, the European Commission and the US have enacted the "Safe Harbor Privacy Principles." These principles allow companies subject to data protection rules in the EU to now transfer personal information to US organizations that comply with Safe Harbour.

Since the DDP was enacted in 1998, most, if not all, EU countries have adopted national data protection legislation. The time frames for implementation and compliance will vary by country, as will the sanctions for noncompliance. However, noncompliance can have serious consequences, which may include imprisonment, financial penalties and business disruptions.

UK – Data Protection Act

The Data Protection Act (DPA) of 1998 applies to UK residents and UK-based organizations. It requires that all personal information, even data not stored in computerized systems, be protected from abuse and secured from unauthorized access.

Specifically under the DPA, anyone processing personal data must comply with simple principles of good practice. That is, the data must be:

- Fairly and lawfully processed for limited purposes and processed in accordance with the data subject's rights.

- Adequate, relevant, not excessive, accurate and secure.
- Not kept longer than necessary and not transferred to other countries without adequate protection.

The DPA requires that data controllers take appropriate technical and organization measures to prevent unauthorized or unlawful processing or disclosure of personal data. Data must be protected during storage, transport, transition and update.

According to the DPA, "Any person who has suffered damage as a result of an unlawful processing operation is entitled to receive compensation from the controller for damage sustained." Managers, directors and other corporate officers can be held liable for DPA offenses committed by their institutions. The DPA provides for a range of civil and criminal penalties.

Canada – Personal Information Protection and Electronic Documents Act

Canada's Personal Information Protection and Electronic Documents Act (PIPEDA) ensures the protection of personal information. Organizations are required to obtain an individual's consent when collecting, using or disclosing personal information. The personal information can only be used for the purpose for which it is collected, unless permission is again obtained from the individual. The individual has the right to access the personal information held by an organization and to challenge its accuracy.

The PIPEDA applies to "any work, undertaking or business that is under the legislative authority of Parliament."[2] Some of these federal works include, but are not limited to: inter-provincial or international transportation by land or water; airports, aircraft or airlines; telecommunications, radio and television broadcasting; banks, grain elevators, nuclear facilities and offshore drilling operations.

Organizations must protect personal information regardless of the format by:

- Developing and implementing a security policy.
- Using appropriate security safeguards, including physical measures, technological tools (passwords, encryption, firewalls and anonymizing software) and organizational controls.
- Removing or masking any personal information that has no relevance when providing copies of information.

The Privacy Commissioner works as the ombudsman and may file a complaint to the Federal Court. The Federal Court can order an organization to correct its practices and to publish a notice of any action taken or proposed to correct its practices. The Court may award an unlimited amount for damages to a complainant.

---

[2] "Your Privacy Responsibilities: Canada's Personal Information Protection and Electronic Documents Act," **www.privcom.gc.ca**, December 2000

Australia – Privacy Amendment Act of 2000

The Australian Privacy Amendment (Private Sector) amends the Privacy Act of 1988 to include private sector organizations. All private sector organizations with an annual turnover of 3 million AUD, all healthcare service providers and some small businesses must comply with the 10 "National Privacy Principles" (NPPs) originally set forth by the Act.

These NPPs govern the collection, use, disclosure and management of personal information. Organizations cannot disclose an individual's personal information for any reason other than the original reason for which it is collected. In addition, they must take reasonable steps to:

- Protect personal information from misuse and loss and prevent unauthorized access, modification or disclosure.
- Destroy or permanently de-identify data if it is no longer needed for the purpose originally collected.

The NPPs are enforced by the Federal Privacy Commissioner who has the power to enforce the principles and require an organization to compensate the complainant. If an organization does not comply after the Commissioner enforces the principles, the matter can be taken to the Federal Court. Generally, the Commissioner only enforces sanctions in more serious or repeated violations.

princetonsoftech.com

111 Campus Drive
Princeton, NJ 08540-6400
Toll free 800.457.7060
Phone 609.627.5500
Fax 609.627.7799