



Data privacy best practices: time to take action!

Contents

2 Executive summary

3 Why is it important to protect privacy? Understand the facts.

3 What areas are most vulnerable? Understand the risks.

4 Considerations for planning a data privacy project

13 Additional points to consider

13 Optim implementation – a user's perspective

Executive summary

In a technology-driven world, data breaches are not only common, they can also be costly. Privacy violation statistics indicate that the number of incidences and costs associated with data breaches are increasing steadily, proving that organizations across industries need to take a more pragmatic approach for protecting information, especially in highly vulnerable non-production (development, testing and training) environments. Data in non-production can be more susceptible to a breach when it is used in development and testing activities, accessed by mobile employees or outsourced.

In the midst of unprecedented security breaches, the best way to ensure that confidential information remains protected is to develop and implement a comprehensive privacy and security strategy. Once organizations realize that protecting privacy is no longer optional, most ask, “Where do we start?”, “What are the requirements?” and “What steps should our organization take to implement an enterprise data privacy and security strategy?”

This white paper explains the steps you need to consider when developing your privacy strategy and implementing your first data privacy project. Using proven data masking techniques, such as those provided in the IBM® Optim™ Data Privacy Solution, can help your organization implement best practices in privacy protection and make your privacy project successful from start to finish. Lastly, learn how a large retail company implemented Optim to develop a best practice strategy and a successful privacy project, overcoming many of the challenges that occur when implementing a privacy project on an enterprise scale.

Why is it important to protect privacy? Understand the facts.

With hackers on the loose and identity theft on the rise, data privacy breaches are impacting our personal and business lives in ways we never dreamed of before. The proof is in the numbers. According to the Privacy Rights Clearing House, since January 2005 in the U.S. alone, the total number of records containing sensitive personal information involved in security breaches was 230,441,730.¹ This number is increasing daily.

In this technology age, much of the confidential information that is targeted for theft resides in the business applications and computer systems that drive enterprise business initiatives. Without appropriate measures in place to protect privacy and prevent the severity of a breach, the next company affected could be yours.

Data privacy begins with protecting different types of sensitive application data, no matter where it resides across your organization, in both production and non-production (development testing, and training) environments. However, companies are realizing that the methods for protecting privacy in production environments may not be practical or appropriate for managing data in non-production environments.

What areas are most vulnerable? Understand the risks.

The methods for protecting privacy in production versus non-production environments should be different. For example, most production environments have established security and access restrictions to protect against data breaches. Standard security measures can be applied at the network, application and database levels. Physical entry access controls can be extended by implementing multi-factor authentication schemes, such as key tokens or even biometrics. However, these

protective measures cannot simply be replicated across every environment. The methods that protect data in production may not meet the unique requirements for protecting non-production environments, where developers, testers and trainers need more access to realistic data, not less.

A 2007 survey, conducted by the Ponemon Institute and Compuware, showed that an overwhelming number of organizations use live data for testing and development purposes (69 percent use live data for testing of applications and 62 percent of respondents use live data for software development).² For example, companies are using live data, such as customer, employee and vendor records, consumer lists, and payment card, business partner and other types of confidential information, for development and testing purposes, ultimately heightening the risk of exposure.³ Many of the companies surveyed indicated that live data used in software development is not protected.

The study also indicates that about half of the companies surveyed outsource their application testing and share live data. Most times, these companies have no real way of knowing if the live data used in outsourced testing environments has been compromised. The only viable solution is to disguise the data through de-identification.

De-identifying data in non-production environments is the process of systematically removing, masking or transforming data elements that could be used to identify an individual. Data de-identification enables developers, testers and trainers to use realistic data and produce valid results, while still complying with privacy protection rules. De-identified data is generally considered acceptable to use in non-production environments and ensures that even if the data is stolen, exposed or lost, it will be of no use to anyone.

Considerations for planning a data privacy project

Industry privacy laws and regulations no longer make protecting data privacy an option. So how can companies ensure that the information they send overseas, keep on laptops, or use for internal development and testing activities stays protected? To

help protect the sensitive information entrusted to them, companies must consider implementing data de-identification practices as a priority in their overall privacy and security strategy.

Companies should also decide on a type of development methodology to use in accordance with the privacy project. Baseline project requirements and assumptions may evolve over the course of the project. Because of the size and complexity of the privacy project, as well as compliance pressures, it is important for organizations to develop a methodology that allows for flexibility along the way. So what must be considered when undertaking a significant privacy project? Table 1 presents six best practices for the implementing a successful privacy project.

Table 1. Managing a successful privacy project

Step	Description
Get organized	Form a cross-functional privacy team to help guide your endeavor.
Define requirements	Define the requirements of your privacy project and identify the types of applications/hardware/data that must be protected.
Perform data inventory	Analyze and catalog your data stores, flows, processes, dependencies and business rules to help simplify the scope of your privacy project.
Select a solution	Choose and implement a data privacy solution that provides the techniques needed to protect privacy in all environments.
Test, test, test	Develop a prototype and methodology for your project and then test the prototype for validation.
Widen the scope	Expand your data privacy project to encompass other applications across your organization.

These steps present a high-level overview for managing your data privacy project. Let's take a look at each step in more detail.

Step 1 - Get organized. The first step in creating, planning and managing a data privacy project is to get organized. Establishing some base privacy directives and guidelines can help you visualize the scope of the project and stay on track. To manage your project and make sure directives are addressed, you will need to appoint a privacy project leader or team. The team should include: *application and business owners*, who directly use the applications; *compliance managers*, who ensure that your company complies with privacy rules and regulations; *IT managers*, whose teams will implement technology to support your privacy initiatives; *operation managers and QA managers*, who may be involved if automated processes and test cases are altered throughout; and anyone else who directly impacts the privacy project. A cross-functional privacy team will require and stimulate inter-departmental cooperation, so that all involved areas are represented.

Throughout the duration of the project, the privacy team will collaborate in making project decisions and finding answers to questions along the way, helping to keep the project focused and on target. Having an established privacy team will also foster support for any additional projects, as you expand data privacy initiatives across your enterprise.

Step 2 – Define requirements. Once you have set up your basic privacy directives and guidelines, and have established a privacy team to spearhead the project, the next step is to understand and define your organization’s privacy and de-identification requirements. First, identify your compliance goals, with respect to national or industry privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), the Payment Card Industry Data Security Standard (PCI DSS) and many others. Each regulation has specific requirements that must be considered in your privacy project.

Next, compile a list of targeted applications, including physical location, business area, supporting databases and hardware platforms. Applications that manage or store confidential customer, employee and corporate business data must be a high priority. Understanding the purpose of each application will help you identify the different types of data that must be masked across program areas and why. From there, identify the specific types of data that require de-identification and estimate how much needs to be masked and propagated across applications, databases and operating environments.

Lastly, when masking data for use in non-production environments, know how closely the subsets of masked replacement data need to accurately reflect the application logic in the original, live data. Each type of data may have a different masking requirement to consider. For example, if your production data contains U.S. Social Security numbers, does the masked replacement data have to match the Social Security format?

Ultimately, de-identifying data provides the most effective way to protect privacy and support compliance initiatives. The capabilities for de-identifying confidential data must allow you to protect privacy, while still providing the necessary “realistic” data for use in development, testing, training environments or for other legitimate business purposes.

Step 3 – Perform data inventory. This step involves some “data forensics,” as you analyze the metadata (the information about your data that makes it easy to understand, use and share). Browsing your applications to determine and validate the content will give a more reliable description of the data managed in your applications, especially in legacy data stores or sequential files, where personally identifiable data can be “hidden.”

Analyzing the data flow within your application portfolio can also help you classify data into groups and hierarchies. Any changes made to data in the upper tiers of the hierarchy cascade down to the lower tiers, so you can reduce the time, size and complexity of your project. Examining privacy requirements at such a specific level can simplify your project, helping you create a more realistic project roadmap to meet your final goal.

Step 4 – Select a solution. In evaluating data privacy technology, look for a solution that meets your requirements, as defined in Steps 2 and 3. In addition, plan for change. As new privacy legislation is enacted and current regulations are enforced, you must have the capabilities to accommodate change. Similarly, there will be changes to your business applications with each new upgrade and enhancement. To keep up with these changes, your data privacy solution must be scalable and provide flexible capabilities that allow you to modify data masking and privacy protection routines, as needed.

For example, the IBM® Optim™ Data Privacy Solution provides comprehensive capabilities for de-identifying application data that can be used effectively across non-production environments. Delivering scalability and flexibility, Optim's data masking techniques are consistent and repeatable, and can be deployed across applications, databases, operating systems and hardware platforms to meet current and future needs.

Optim has the breadth and depth to provide data masking across your enterprise, and is not a point solution designed to resolve one concentrated privacy issue. As an enterprise solution, you can incorporate Optim into your existing and ongoing business processes.

To enable organizations to meet even the most complex data privacy challenges, Optim provides the following fundamental components of effective data masking:

- **Preserving application logic.** Optim's application-aware data masking capabilities understand, capture and process data elements accurately so that masked data does not violate application logic. For instance, surnames are replaced with random, valid surnames, not with meaningless text strings. Numeric fields retain the appropriate structure and pattern. If diagnostic codes are four digits, and range in value from 0001 to 1000, then a masked value of 2000 would be invalid in the context of the application test. Checksums remain valid, so that functional tests pass application validity checks. Optim also propagates all masked data elements consistently throughout a test database, and to other related applications and databases.

For example, Direct Response Marketing Company, Inc. is testing its order fulfillment system and needs to de-identify customer names to ensure safe testing. Optim's Random Lookup function allows the company to arbitrarily generate first and last names from a predefined Customer Information table. "Lucille Ball" would become "Elena Wu" each time it appears and so on. Ultimately, masking will be repeatable and predictable so that the same change appears consistently as appropriate to meet your requirements across non-production environments.

- **Masking key data elements.** Optim's context-aware, prepackaged data masking routines de-identify key data elements, providing a variety of proven data masking techniques that can be used to de-identify many types of sensitive information. For example, birth dates can be masked to accurately reflect a person's correct age. Similarly, you can mask bank account numbers, national identifiers (like Canada's Social Insurance numbers or Italy's Codice Fiscale), benefits information and so on.

Prepackaged Transformation Library™ routines allow for accurately masking complex data elements, such as Social Security numbers, payment card numbers and e-mail addresses. Built-in lookup tables support masking names and addresses. You can also incorporate site-specific data transformation routines that integrate processing logic from multiple related applications and databases and provide greater flexibility and creativity in supporting complex data masking requirements.

For example, Green Bill Bank's account numbers are formatted "999-9999," where the first three digits represent an account type (checking, savings, or money market) and the last four represent a customer identification number. For testing purposes, these account numbers must be masked. Optim can utilize the first three digits of the actual account number and then generate a sequential four-digit number to replace the last four in the actual account number. So, "001-4570" would become "001-1000" and so on. The result is a fictionalized account number that still retains the bank's valid account number format.

- **Propagate masked data elements accurately.** Optim's persistent masking capability generates transformed replacement values for source columns and propagates the replacement values consistently and accurately across applications, databases, operating systems and hardware platforms. Persistent data masking capabilities ensure scalability for protecting privacy across multiple non-production environments. Propagating masked primary key values to all related tables is necessary to help maintain the referential integrity of the data – even after it is masked – keeping complete subsets of related data intact (see Figure 1).

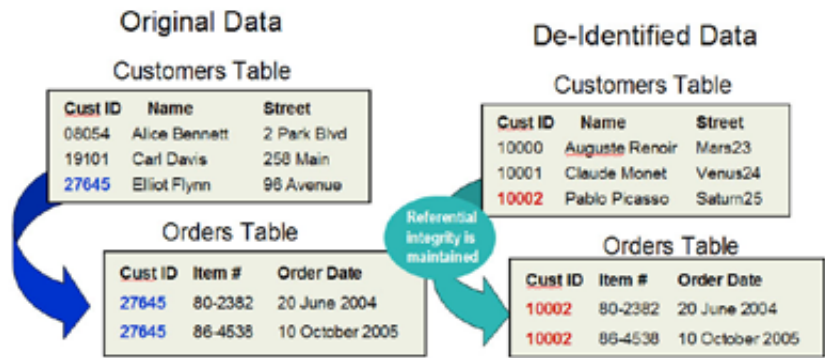


Figure 1. Optim's key propagation capabilities help preserve data's referential integrity.

Figure 1 depicts a simple example consisting of two related tables, where the Customers table is parent to the Orders table, and its primary key column, Cust_ID, is a 5-digit numeric value. The Cust_ID, Name and Street columns are masked. The name "Elliot Flynn" has been masked as "Pablo Picasso." The sequential masking technique was used to transform the original Cust_ID for Elliot Flynn from 27645 to 10002. The masked Cust_ID value is propagated from the Customers table to all related tables and the key relationship between the Customers and Orders tables in the test database remains intact.

Once you have designed your privacy strategy and have selected a solution, implementation is the next step. Optim's unique application-aware and context-aware techniques, as well as its propagation capabilities, provide the powerful data masking resources you will need for a successful implementation.

Step 5 – Test, test, test. Once you have developed your data privacy strategy, the next step is to determine, "Does it work?" Build a prototype masking scenario for a

selected application or set of applications to ensure that your masking techniques and testing activities will execute as expected. This scenario should correlate to your application testing requirements. Verify that your privacy strategy has been designed effectively to meet defined requirements. For instance, compare your original data with the de-identified data to ensure that your changes have been applied. Finally, validate that your privacy strategy is appropriate for your applications and non-production environments.

Then, test the prototype to validate that the appropriate masking techniques are applied. Have you ensured that all areas containing sensitive data are masked? Will all sensitive data be masked across applications and testing environments? Will your privacy strategy keep you in compliance with industry and federal privacy laws and regulations?

Step 6 – Widen the scope. Post-implementation, expand your data privacy project to encompass other sets of applications across your organization. Being proactive about protecting privacy across your organization is key. Follow through with implementing a data privacy strategy across other areas within your organization that store or manage sensitive information. Are these areas at risk of becoming the target of a data breach?

Chances are, the need to protect privacy is not limited to one area of your company. Expanding the reach of your privacy project to include various areas that contain sensitive information will be a more effective way to secure your information assets.

Protecting privacy represents a best practice for managing sensitive data and makes good business sense. Data de-identification provides a “safe sandbox” for mission-critical application testing, helping companies protect privacy and safeguard customer loyalty. Optim provides flexible capabilities that scale to support current and future data privacy requirements, helping to make data de-identification part of your overall privacy strategy. With Optim, an organization is better positioned to satisfy local, state, national and international privacy regulations, as well as legal and industry requirements.

Additional points to consider

Privacy projects – inclusive of privacy project teams, processes, environments and formulas – will vary from site to site and not every one is the same. Ensuring efficient data protection and security can depend on many factors. Here are some other points you may want to keep in mind as you embark on your own data privacy project:

- Consider physically separating the privacy process from the testing process. Implementing the privacy process in a production-like test environment will help promote the highest level of security available in your organization.
- Consider enforcing a strict “separation of roles” on your project team. If de-identified data is accessed only on a need-to-know basis, then there is less risk of someone reversing the masking process.
- Consider taking steps to protect internal security by slightly modifying your data privacy “recipe.” Making slight changes to the data before moving it to production can add an additional layer of protection.
- Consider a third-party audit to certify that your privacy processes are tamper-proof, for instance, you enforce separation of roles, your roadmaps are appropriate and your timeline provides sufficient “cushioning” to remediate any gaps in the process.

Optim implementation – a user’s perspective

As a growing discount retailer, Marzan Corporation owns and operates a chain of retail stores throughout the United States and the United Kingdom. With its first store opening in New Jersey in 1979, Marzan has made a name for itself over the years as a “cheap yet chic” one-stop shop for your everyday needs. From household items to clothing, and electronics to automotive equipment, Marzan Retail Stores provide customers with a convenient, stylish and fun – yet affordable – place to shop.

Diverse application portfolio supports the business. To help support and drive its ongoing business initiatives and support its vendor relationships, Marzan relies on a PeopleSoft® Enterprise supply chain application, internally known as IZZI, that manages vendor orders and inventory activities. IZZI contains sensitive information, like vendor names and IDs, and runs on an Oracle® database.

Marzan also uses a Siebel® CRM application, referred to as Jasper, to handle customer orders. Jasper, which runs in a DB2 open systems environment, contains sensitive customer order information, such as customer names, addresses, telephone numbers and payment card numbers. Some of this information is essential in helping Marzan do promotional activities to expand its customer base.

Finally, Marzan utilizes a custom Financials application, called Centz, which runs on IBM System z™. This customer billing application captures and stores Marzan credit card application information, as well as customer billing data, which may include Social Security numbers, names and addresses.

Data privacy challenges. Offering top-quality products and customer service has always been a high priority for Marzan. To help deliver better customer service, Marzan wanted to give its customers enhanced capabilities for accessing account information online and performing online order entry. Achieving this goal required improvements to both the existing Jasper and Centz applications. Each application contains sensitive data that is collected from online entries — payment card numbers, Social Security numbers, names, street addresses and telephone numbers. Enhancements would allow for more sophisticated online, customer-facing account functionality, accessibility and security.

Further, since the IZZI application houses vendor order information, such as vendor names and IDs, Marzan requires that the information be masked to ensure that the information is safe to use in testing and development activities. Knowing that it would have to perform testing across its databases and applications, Marzan also wanted to ensure that the referential integrity of the data remained intact.

Finally, as a large retail chain, Marzan had to maintain compliance with PCI DSS (Payment Card Industry Data Security Standard) regulations. PCI DSS requires large retailers and businesses that process payment cards to mask personally identifiable customer information used in application testing environments. Marzan would have to guarantee that any data from its Centz and Jasper applications, used for development and testing purposes, was masked and safe to use in non-production environments.

The search for a solution. Marzan needed a solution that would meet all of its privacy needs across the enterprise. In order to support and implement its overall data privacy initiatives, Marzan decided to purchase and implement the IBM Optim Data Privacy Solution. The decision to purchase an enterprise data management masking solution was initiated by a small privacy team, comprised of application developers and testers, IZZI, Jasper and Centz users and privacy professionals.

Marzan established a privacy team to spearhead the multi-faceted project. The team wanted to be proactive in safeguarding customer information. Optim had the capabilities to protect sensitive information across Marzan's enterprise. Optim's unique masking and transformation capabilities for mainframe and open systems application data helps to protect privacy across applications, databases, operating systems and hardware platforms.

Successful implementation gives way to business benefits. After a successful Optim implementation, Marzan's de-identification project team deployed the privacy project methodology that it worked so hard to render. Beginning with IZZI, Marzan took steps to de-identify sensitive data in non-production environments across the enterprise.

As an approach to de-identify vendor names in IZZI, Marzan used Optim's Lookup techniques to transform data using substitution values. The team could mask a value in a source column by returning a corresponding masked value to a destination column, transforming real vendor names into fictionalized names for testing and development purposes. Optim's Lookup tables ensured that "Dave Acme" from Acme Pencil Company would be de-identified as "Michael Craft" in non-production.

The vendor IDs in IZZI are six characters in length, with the first being an alphabetic character based on the first letter of the vendor name and the second being an alphabetic character based on the vendor city. The last four digits are numeric and must range from 1000 to 6999. So, the Acme Pencil Company in Tucson, AZ would have a vendor ID like AT1453. Optim's application-aware capabilities ensure that a masked vendor ID preserves application logic and returns a masked vendor ID like CD2047, not CD8945.

Next, Marzan applied de-identification techniques to Jasper, the Siebel CRM application. Optim's context-aware data masking routines de-identified key data elements, such as Social Security numbers, credit card numbers and birth dates, in the application testing and development environments. To perform accurate development and testing activities, Marzan's Development staff needed to work with real sixteen-digit credit card numbers. Optim's intelligent masking features devised contextually valid, but masked credit card numbers. In the event of a data breach, masking renders credit card numbers useless to thieves, but ensures they are valid for use in non-production environments (see Figure 2).



Figure 2. Optim generates valid and unique, yet de-identified, payment card numbers according to the issuer's format requirements.

Similar context-aware masking techniques were applied to the Centz financial application. This application captures and stores customer billing information, so Marzan used Optim's built-in Lookup tables to mask names and addresses that are stored in Centz. So, any instance of the name "Beth K. Smith" would become "Claire P. Hamill" once de-identified. Optim propagated these changes appropriately to maintain referential integrity and ensure privacy protection across Marzan's non-production environments.

Finally, in order to make application enhancements for its online customer account access program, Marzan needed to create federated testing and development environments. Optim provided federated access capabilities that allowed developers to extract and mask appropriate test data from various data sources in a single process. Optim's subsetting capabilities provided automated and repeatable capabilities for processing federated extracts from both the IZZI and Jasper applications.

With Optim, Marzan created realistic, right-sized subsets of application data for its testing and development environments. Optim masked sensitive customer information, like customer names, addresses, phone numbers, payment card numbers and payment history details. De-identified data was acceptable to use in

testing and development environments. The data was then propagated accurately across non-production environments, while preserving the referential integrity to support reliable testing.

Optim's capabilities for protecting customer information in the development and testing environments also helped satisfy regulatory PCI DSS requirements. Since personally identifiable customer and payment card information resides in Marzan's supply chain applications, Optim provided the capability to mask confidential data. As a result, Marzan was able to reduce legal risks that could have resulted in financial penalties, as well as the loss of the customer loyalty and trust the company worked so hard to build.

Using substrings, random or sequential number replacements, arithmetic expressions, date aging and other techniques, Marzan substituted actual customer data with contextually accurate, but fictionalized data to produce accurate test results. This data was acceptable and safe to use in non-production, but was useless to thieves or hackers.

Ultimately, Optim helped reduce the risks and costs associated with potential privacy breaches, which upheld and reinforced Marzan's first-rate reputation. Marzan's customers and business users are benefiting from more reliable and feature-rich applications. More importantly, Marzan has maintained its high levels of customer service and is reaping the benefits of expanded revenue opportunities.

About IBM Optim

IBM® Optim™ enterprise data management solutions focus on critical business issues, such as data growth management, data privacy compliance, test data management, e-discovery, application upgrades, migrations and retirements. Optim aligns application data management with business objectives to help optimize performance, mitigate risk and control costs, while delivering capabilities that scale across enterprise applications, databases and platforms. Today, Optim helps companies across industries worldwide capitalize on the business value of their enterprise applications and databases, with the power to manage enterprise application data through every stage of its lifecycle.

For more information

To learn more about IBM Optim enterprise data management solutions, contact your IBM sales representative or visit: www.optimsolution.com.



© Copyright IBM Corporation 2008

IBM Software Group
111 Campus Drive
Princeton, NJ 08540-6400
USA
www.optimsolution.com

Produced in the United States of America
09-08
All Rights Reserved.

¹ *Privacy Rights Clearing House*, <http://www.privacyrights.org/ar/ChronDataBreaches.htm#2008>.

² *Compuware and The Ponemon Institute LLC*.
"The Insecurity of Test Data: The Unseen Crisis."
Compuware.com. December 2007: Page 3.

³ *Ibid*, Page 4.

DB2, IBM, the IBM logo, Optim and Transformation Library are trademarks or registered trademarks of the IBM Corporation in the United States, other countries or both.

All other company or product names are trademarks or registered trademarks of their respective owners.

References in this publication to IBM products, programs or services do not imply that IBM intends to make them available in all countries in which IBM operates or does business.