



IBM. **Information Management** software

## IBM® Optim™ Data Privacy Solution

---

### Highlights

---

- ***Safeguard personally-identifiable information, trade secrets and other sensitive data***
- ***Easily mask confidential data using predefined transformations and site-specific routines***
- ***Discover hidden instances of private data so that they can be fully protected***
- ***Support compliance with privacy regulations and corporate governance standards***

### **Data privacy: the “untold story”**

Data privacy protection continues to be a tremendous focus for the IT community today. Organizations are making great strides to protect sensitive data in live application environments. But the “untold story” of implementing protection strategies in non-production (testing, development and training) environments remains a critical risk. As data breach headlines continue to mount, organizations must begin to address the most vulnerable areas of IT infrastructure – non-production environments.

So, what makes non-production environments so unique? The answer lies in the methods used to create non-production databases. Commonly, live production systems are cloned (copied) to a test environment – confidential data and all. Developers and QA testers find it easy to work with live data because it produces test results that everyone can understand. But do non-production environments actually require live data? The answer is, “no.” Using realistic data is essential

to testing, but live data values are not specifically necessary. Capabilities for “de-identifying” or masking production data offer a best practice approach for protecting sensitive data while supporting the testing process.

### **Data masking offers a best practice approach**

Data masking is the process of systematically transforming confidential data elements such as trade secrets and personally-identifying information (“PII”) into realistic but fictionalized values. Data that has been scrubbed or cleansed in such a manner is considered acceptable to use in non-production environments. Masking enables developers and QA testers to use “production-like” data and produce valid test results, while still complying with privacy protection rules.

Data masking represents a simple concept, but it is technically challenging to execute. Most organizations operate within complex, heterogeneous IT environments, consisting of multiple, interrelated

applications, databases and platforms. IT managers do not always know where confidential data is stored or how it is related across disparate systems. The ideal solution must both discover sensitive data across related data stores, and mask it effectively.

The IBM® Optim™ Data Privacy Solution provides comprehensive capabilities for masking sensitive data effectively across non-production environments. You can take the necessary steps to protect privacy, while still providing realistic data for use in development, testing or training. When you use Optim to mask confidential data, you protect privacy and safeguard shareholder value.

**Implement proven data masking techniques**

Optim Data Privacy Solution users can apply a variety of proven data transformation techniques to mask sensitive real data with contextually accurate, realistic data. Users can mask data in a single database or across multiple related systems. Some simple examples of Optim’s masking techniques include substrings, arithmetic expressions, random or sequential number generation, date aging and concatenation.

Optim’s context-aware masking capabilities ensure that masked data resembles the look and feel of the original information.

These capabilities make it easy to de-identify many types of sensitive information, such as birth dates, bank account numbers, street address and postal code combinations, and national identifiers (like Canada’s Social Insurance numbers or Italy’s Codice Fiscale).

Optim’s Transformation Library™ routines allow for accurately masking complex data elements, such as, credit card numbers and e-mail addresses. You can also incorporate site-specific data transformation routines that integrate processing logic from multiple related applications and databases. Optim offers the greatest flexibility to support even the most complex data masking requirements.

**Discovering sensitive data**

Some sensitive data is easy to find. For instance, credit card numbers in a column named “credit\_card\_num”

will be simple to recognize. Most application databases though, are more complex. Sensitive data is sometimes compounded with other data elements, or buried in text or comment fields. Subject matter experts can sometimes offer insight, but only if they fully understand the system.

Figure 2 illustrates an example. Table A contains telephone numbers in the “Phone” column. In Table B however, the telephone number is obscured within a compound field in the “Transaction Number” column. Both instances represent confidential information that must be protected. But while data analysts can clearly recognize the telephone number in Table A, they may well overlook it in Table B. And every missed occurrence of private information represents a risk to the organization. What is the alternative?

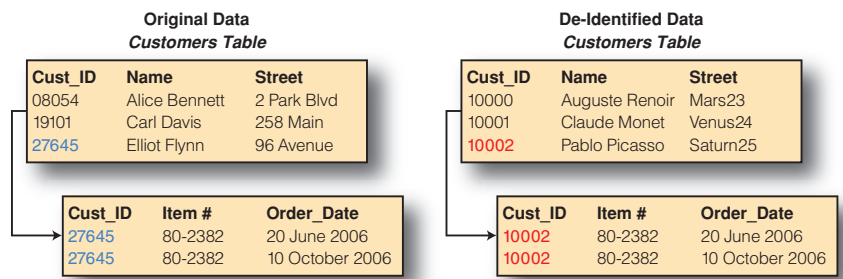


Figure 1: Optim offers a variety of data masking techniques to protect the confidentiality of private information and propagate it throughout the system.

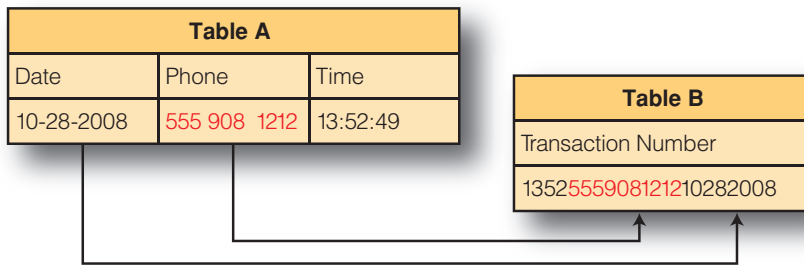


Figure 2: Confidential information hidden in compound fields poses a privacy risk to the organization

IBM® InfoSphere™ Discovery enables organizations to identify all instances of confidential data across the environment – whether clearly visible or obscured from view. InfoSphere Discovery works by examining data values across multiple sources to determine the complex rules and transformations that may hide sensitive content. It can locate confidential data items that are contained within larger fields, as described in the prior example, or that are separated across multiple columns. InfoSphere Discovery delivers automated capabilities that offer greater accuracy and reliability than manual analysis. When used together, the Optim Data Privacy Solution and InfoSphere Discovery provide the most effective, enterprise-scale solution for locating and masking sensitive data across complex, heterogeneous environments.

### Ensuring data integrity

Finding and masking data is part of the solution. However, there is an added complication. You need the capability to propagate masked data elements to all related tables in the database and across databases to maintain referential integrity. For example, if a masked data element, such as 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 or across data sources. If the data is a portion of another row's data, it must be updated with the same data as well.

InfoSphere Discovery not only discovers hidden sensitive data, it also provides a full range of data analysis capabilities to discover hidden

relationships and bring them clearly into view. By leveraging the combination of InfoSphere Discovery and Optim, all relationships will be uncovered and replacement values will be masked consistently and accurately across multiple data sources.

Non-production data is created in multiple ways, but all must be protected. To minimize risk, data should be masked as close to its source system as possible. In some scenarios data is copied directly from a live system. In this case, data must be masked “in place” to ensure that the newly created test database is protected for use. In other scenarios, specific subsets of data are extracted using test data management products like the IBM® Optim™ Test Data Management Solution. Here, data is masked during the extract process to ensure that private information is never exposed.

### Support compliance initiatives

To support industry, government and internal compliance initiatives, data masking is a must. The European Union has established the Personal Data Protection Directive as the framework for privacy protection governing its

member countries. And many other countries have similar regulations around the world. The U.S. Department of Health and Human Services has enacted the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule for the privacy of individually identifiable health information. Additionally, industry coalitions are developing sector-specific governance standards. For instance, the Payment Card Industry Data Security Standard (PCI DSS), initiated by Visa® and MasterCard®, Implementing Optim helps you comply with these data privacy regulations by protecting the confidentiality of sensitive information across your enterprise.

Optim provides a scalable data privacy solution with flexible capabilities that can be easily adapted to your current and future requirements. You also benefit from knowing that Optim supports all leading enterprise databases and operating systems, including IBM DB2®, Oracle®, Sybase®, Microsoft® SQL Server®, IBM Informix®, IBM IMS™, IBM VSAM®, Teradata®, Adabas®, Microsoft Windows®, UNIX®, Linux® and IBM z/OS®. In addition to providing data management support for all custom and packaged

applications, Optim has the meta-model knowledge to support the key ERP and CRM applications in use today: SAP® Applications, Oracle® E Business Suite, PeopleSoft® Enterprise, JD Edwards® EnterpriseOne, Siebel® and Amdocs® CRM.

### **About IBM Optim Integrated Data Management Solutions**

IBM Optim Integrated Data Management Solutions offer proven, integrated capabilities to manage enterprise application data from requirements to retirement. With Optim, teams can share data artifacts (like models, policies and metadata) to align data management with business goals and improve collaboration. Today, organizations of all types leverage Optim to improve performance, streamline database administration, speed application development, and enable effective governance. Optim delivers better business outcomes, at lower cost, with less risk, while providing capabilities that scale across enterprise applications, databases and platforms.

### **For more information**

To learn more about IBM Optim Integrated Data Management Solutions, contact your IBM sales representative or visit: [www.ibm.com/software/data/optim-solutions/](http://www.ibm.com/software/data/optim-solutions/)



© Copyright IBM Corporation 2008

IBM Software Group  
[www.optimsolution.com](http://www.optimsolution.com)

Produced in the United States of America  
12-08  
All Rights Reserved

DB2, IBM, the IBM logo, IMS, Informix, Optim, VSAM and z/OS are trademarks or registered trademarks of the IBM Corporation in the United States, other countries or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. UNIX is a registered trademark of The Open Group in the United States and other countries. Windows and SQL Server are registered trademarks of Microsoft Corporation in the United States and other countries. 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.

Each IBM customer is responsible for ensuring its own compliance with legal requirements. It is the customer's sole responsibility to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.