IBM InfoSphere Guardium Data Redaction

Document protection for regulatory compliance and risk reduction

Highlights:

- Automatically recognize and remove sensitive data in unstructured documents, forms and graphics
- Transform tedious, manual redaction alternatives into automated processes for speed, accuracy and efficiency
- Safeguard proprietary or personal information from internal/external misuse and fraud
- Support regulatory compliance and privacy requirements by applying data governance controls

The complex balancing act between openness and privacy

As information volumes expand and organizations find new ways to collaborate with partners and customers, the question arises: How secure is your sensitive information? Organizations need to make quick decisions and respond promptly to changing market and customer needs, so efficiently sharing information between people, processes and applications is no longer just a goal—it's a strategic imperative.

In addition to—and often in contrast to—this collaboration, organizations must also ensure that the right information stays private. As organizations strive to support information governance programs, they must balance the need to comply with laws and industry regulations with the need to deliver trusted information for business use. Consider the public company that needs to redact (that is, remove or black out) private or proprietary information before releasing documents or records to the press or shareholders. Or the bank that must keep a customer's credit score in a loan document hidden from an office clerk but still visible to a loan officer. In cases like these, organizations often struggle with how to effectively secure private data without sacrificing information sharing.

Further complicating matters is the fact that different types of information have different protection and privacy requirements. Therefore, organizations must take a holistic approach to protecting and securing their business-critical information:

- **Understand where the data exists:** Organizations can't protect sensitive data unless they know where it resides and how it's related across the enterprise.
- **Safeguard sensitive data, both structured and unstructured:** Structured data contained in databases must be protected from unauthorized access. Unstructured data in document and forms requires privacy policies to redact (remove) sensitive information while still allowing needed business data to be shared.
- **Protect non-production environments:** Data in non-production, development, training and quality assurance environments needs to be protected, yet still usable during the application development, testing and training processes.
- **Secure and continuously monitor access to the 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
- **Demonstrate compliance to pass audits:** It's not enough to develop a holistic approach to data security and privacy. Organizations must also demonstrate and prove compliance to third party auditors.

The challenges of securely sharing unstructured sensitive information

Traditionally, protecting unstructured information in forms, documents, graphics or XML files has been performed manually by deleting electronic content and using a virtual marking pen to delete or hide sensitive information. But this manual process can introduce errors, inadvertently omit information or leave behind hidden information within files that exposes sensitive data. Today's high volumes of electronic forms and documents make this manual process too burdensome for practical purposes and increases risk.

What if there was a way to automate the manual redaction process? Doing so would make this time-intensive activity more cost effective and help organizations better comply with regulations. It would also help to preserve an organization's competitive advantage, secure its intellectual property and safeguard its public reputation.

IBM® InfoSphere® Guardium® Data Redaction protects sensitive data buried in unstructured documents, forms, images or XML files from unintentional disclosure. The automated solution lends efficiency to the redaction process by detecting sensitive information and automatically removing it from the version of the documents made available to unprivileged readers. Alternatively, InfoSphere Guardium Data Redaction can first present this sensitive information to security professionals for review. This will allow them to explore the data before the proceeding with the redaction. Previewing sensitive data also helps security professionals make more informed choices not only about redaction policies but other privacy policies.

Based on industry-leading software redaction techniques, InfoSphere Guardium Data Redaction offers the flexibility of human review and oversight if required. InfoSphere Guardium Data Redaction is part of the IBM security framework for data and information, helping organizations meet the broader challenge of protecting sensitive data, no matter where it lies.

Industry-leading techniques for protecting sensitive unstructured data

Across the enterprise, unstructured documents come in many formats and styles, including scanned paper documents, PDF, TIFF, text, Microsoft® Word and XML files. InfoSphere Guardium Data Redaction works with each of these file types and leverages unique entity recognition techniques to identify sensitive data or "redaction candidates" in those documents and forms. Using a comprehensive set of libraries and algorithms developed in the IBM Research Labs for text extraction, InfoSphere Guardium Data Redaction raises the bar for redaction. This is accomplished via the automated steps of *search*, *analyze and extract*, making the redaction process much more reliable and efficient than alternative "virtual black marker" approaches (see Figure 1).

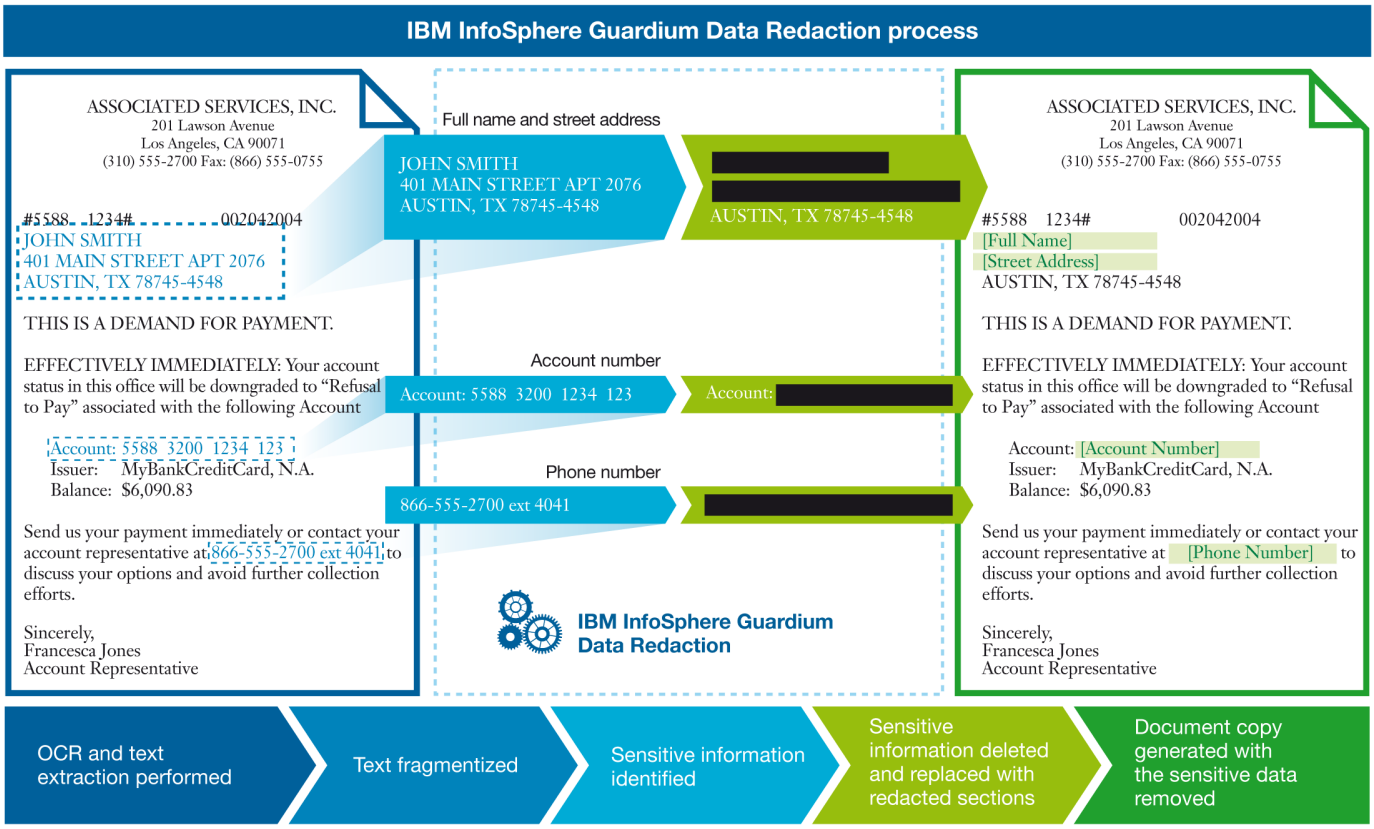

Figure 1: InfoSphere Guardium Data Redaction entity recognition techniques consist of critical steps to help ensure sensitive unstructured data is secure and safeguarded from unwanted sharing.

Organizations benefit from improved business confidence and better productivity. In short, InfoSphere Guardium Data Redaction can help transform a cumbersome, painful, manual process into a repeatable process for organizations to manage, measure and trust.

The search, analyze and extract steps are part of the process for identifying sensitive data, thereby helping to reduce the likelihood of redaction errors—saving time, money and repeated effort. A web-based review tool enables users to quickly and easily view redaction candidates that the software identifies, opt to redact them or not, include additional redactions and view previously redacted documents. This makes redaction a repeatable process and permits batch-form redaction for increased speed and consistency. InfoSphere Guardium Data Redaction has strong extensibility capabilities and can integrate with enterprise systems via its Java™ application programming interfaces (APIs) and SOAP-based architecture, making it flexible and easy to integrate with existing systems and processes.

InfoSphere Guardium Data Redaction not only removes the sensitive data you can see, but also automatically removes the hidden information or metadata that you can't see in a document. Common document editing commands like Fast Save, Undo, Track Changes and other actions can create metadata that contains copies of those document edits. Metadata removal is critical because sensitive information often resides in the metadata of documents—and in many situations it is collected without the user's awareness.

Finally, InfoSphere Guardium Data Redaction provides a secure viewer that allows viewing of sensitive data via a web browser on a need-to-know basis. A person with proper permissions can securely retrieve information as long as he or she specifies a valid business reason. InfoSphere Guardium Data Redaction logs this access to meet audit requirements. Without the flexibility enabled by this feature, redaction policies must be either overcautious and redact information that the user may need, or be too permissive, exposing information for the user's convenience but revealing more information than necessary.

Methods for meeting information governance and compliance initiatives

Today, information governance is more than just applying discipline and controls to the internal people and processes that manage data. It also includes applying data security and privacy methods to make sure information sharing supports compliance and regulatory mandates. InfoSphere Guardium Data Redaction is a key component of an information governance strategy, which helps organizations attack the challenges of information volume and variety with solutions for data protection and privacy regardless of the type of data, location or usage.

Organizations must validate the flow of trusted information by applying the appropriate business rules and privacy procedures to manage data. For example, how does one control which sensitive information is visible? Is it based solely on user role or is there a way to set up policies for more fine-grained control? InfoSphere Guardium Data Redaction tackles the tough challenges of role-based redaction.

Role-based redaction helps set InfoSphere Guardium Data Redaction apart from other solutions by allowing users with the correct privileges and policies to view the sensitive information they need, while keeping others out. This empowers users to do their jobs successfully while supporting compliance rules for data access and privacy. Why is this important? Because the cost of noncompliance with regulatory laws can result in large fines and penalties. Organizations also can suffer from damaged reputations and loss of customer loyalty.

Many types of legal documents require redaction, including tax liens, property deeds, death/birth certificates and marriage certificates. All of those documents contain sensitive personal information and are prime targets for redaction. In healthcare, discharge summaries, progress reports and patient histories contain sensitive personal health information and require redaction as they become part of the patient's electronic medical record. The need for redaction crosses almost all industries where information is distributed or exchanged.

InfoSphere Guardium Data Redaction can help organizations comply with regulatory and legal requirements including:

- **United States Freedom of Information Act (FOIA) for government:** Helps ensure that government agencies disclose or only partially disclose information as needed but continue to maintain "openness of information" for citizen requests
- **Health Insurance Portability and Accountability Act (HIPAA) for healthcare:** Protects an individual's medical records or any kind of personal health information
- **Payment Card Industry Data Security Standard (PCI-DSS) for financial services and credit card companies:** Protects credit card numbers or any kind of customer account data
- **Canada Personal Information Protection and Electronic Documents Act (PIPEDA) for general business:** Protects the privacy of any kind of personal information
- **Gramm-Leach-Bliley Act for financial institutions and insurance companies:** Protects consumers' personal financial information, such as data needed for credit counseling, transferring funds or real estate settlement
- **European Union (EU) Privacy Protection Directive for general use and business:** Protects general data privacy of EU citizens

Proven customer value: Automation with accuracy

A large private health insurer was confronted with new laws requiring it to share health and financial records with hospitals, a national health service and customers. Yet the same law that requires openness also requires the insurer to carefully tailor their leased personal health information (PHI) to the role of the document recipient and to delete credit card numbers in its archives.

The insurer also faced a strong business requirement to open up data to independent insurance agents and other business partners. However, preserving privacy is essential to maintaining the organization's good reputation for respecting its patients' rights. Previous privacy practices put the insurer at risk of regulatory violation. In some cases, it shared documents without checking them, potentially exposing private information, and its archives include millions of unredacted credit card numbers. But in other cases, the organization found it impossible to share data, knowing the compliance risks it would incur.

With InfoSphere Guardium Data Redaction, the insurer is able to smoothly share or archive documents where needed, while withholding precisely the information required by law.

Eliminate manual redaction for cost savings

InfoSphere Guardium Data Redaction transforms the way organizations identify and redact sensitive data buried in forms and documents. Key capabilities include:

- Multiple file format support, including PDF, text, TIFF, XML files and Microsoft Word documents
- Role-based redaction that enables multiple users to receive the same document with different views of information to support privacy policies
- Out-of-the-box support for IBM FileNet® P8 and IBM CM8
- Flexible APIs that support integration with existing systems and processes
- Language support for English, German, French and Spanish

Extensible entity recognition algorithms that enable organizations to customize the solution and find any sensitive data that is required to be redacted. InfoSphere Guardium Data Redaction helps organizations reduce risk and do more with less by saving time, improving accuracy and eliminating manual redaction processes involving virtual marker techniques. It supports compliance requirements and can prevent sensitive data of any type from being disclosed except in accordance with the appropriate permissions and policies—all without sacrificing information sharing and collaboration.

About IBM InfoSphere

InfoSphere Guardium 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 your systems. The InfoSphere platform provides all 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, allowing you to start anywhere, and mix and match InfoSphere software building blocks with components from other vendors, or choose to deploy multiple building blocks together for increased acceleration and value. The InfoSphere platform provides an enterprise-class foundation for information-intensive projects, providing the performance, scalability, reliability and acceleration needed to simplify difficult challenges and deliver trusted information to your business faster.

For more information

To learn more about IBM InfoSphere, contact your IBM sales representative or visit: **ibm.com**/software/data/infosphere

For more information about IBM InfoSphere Guardium Data Redaction, please contact your IBM representative or Business Partner, or visit: **ibm.com**/software/data/optim/data-redaction

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
July 2012

Please Recycle

IMS14137-USEN-03