Best Practices for Ensuring Data Privacy in Production and Nonproduction Systems
# Table of Contents

Executive Summary ....................................................... 2

Data Privacy Best Practices Are Sound Business Practice .......... 3

Protection of Sensitive Data .............................................. 3

  Best Practices for Data Classification ................................... 4
  Structured and Unstructured Data ....................................... 5
  Data Leakage ........................................................ 5

  Best Practices for Data Leakage Prevention .......................... 5
  Sensitive Data and Data Masking ....................................... 6
  Best Practices for Data Masking .................................... 6

Special Regulatory and Industry Requirements ....................... 7

  Best Practices for Regulatory and Industry Requirements .......... 7
  Regulatory Requirements ........................................... 7
  Health Insurance Portability and Accountability Act (HIPAA) ..... 9
  General Compliance Best Practices .................................. 10

Risk Assessment ......................................................... 10

  Ongoing Compliance Assessment ..................................... 11

Securing Applications and Data in Development ..................... 11

  Best Practices for Sensitive Data in Development ................ 12
  Documentation ..................................................... 12

  Best Practices for Documentation .................................. 12
  System Administrators and Developers ............................. 13
Executive Summary

There is a growing need to protect sensitive employee, customer, and business data across the enterprise wherever such data may reside. Until recently, most data theft occurred from malicious individuals hacking into production databases. With a number of well-publicized and costly thefts creating both tremendous legal liability and bad publicity for the effected organizations, business has quickly grown more sophisticated in protecting against such attacks, but so have the attackers.

While the industry deals with the most egregious aspects of data theft, many computer systems still remain vulnerable to attack at some level. An important tier of data remains practically untouched and unprotected by today’s new data security procedures: nonproduction systems used for development, testing, and training purposes. These systems are generally less protected and leave a large hole in the data privacy practices at organizations of all sizes. These environments leverage real data to test applications, housing some of the most confidential or sensitive information in an organization, such as national identification numbers, bank records, and other financial information.

This white paper discusses best practices for creating data privacy procedures in both production and nonproduction environments. These procedures include creating a comprehensive set of policies to classify datatypes that need to be protected, integrating these policies into day-to-day business processes, providing ongoing compliance reviews, using a proven commercial solution for masking sensitive data in all production and nonproduction environments, and integrating these privacy processes and technology across the enterprise.
Data Privacy Best Practices Are Sound Business Practice

Confidentiality, integrity, and availability are the cornerstones of data privacy, as well as sound business practices. They are essential for:

- Compliance with existing regulations and industry standards
- Reliable, accurate, high-performance services
- Competitive positioning
- Reputation of the firm
- Customer trust

“Best practice” varies widely from one situation to another, even for a specific kind of control such as passwords. “Best” is not used here in a literal way. Rather, it is used to combine such notions as “good”, “commonly used”, “prudent”, “industry standard”, or “generally accepted”.

Note that frameworks such as ISO 27001, COSO, COBIT, and ITIL provide a broad range of control objects, but do not provide specific information protection controls. Although there is no official framework for best practices that your organization can simply adopt, a variety of data protection controls have come to be widely accepted as sensible, baseline, and sound practices. In the end, the true authority for what is right for your organization comes from your organization’s management team, regulatory examiners, and industry standards.

Protection of Sensitive Data

Every organization has sensitive data: trade secrets, intellectual property, critical business information, business partners’ information, or customers’ information. All of this data must be protected based on company policy, regulatory requirements, and industry standards. This section will cover several important elements of protecting this data.

Any organization that collects, uses, and stores sensitive information should establish an information classification policy and standard. This classification policy and standard should contain a small number of classification levels that will meet the organization’s needs. Most organizations will have at least three categories such as public, internal use only, and confidential.
Many organizations have long-established data classification guidelines. However, with the increasing number of new regulations and advancement in industry standards, the mere presence of a corporate policy is no longer sufficient. Some organizations have spent a lot of time and effort operationalizing their data protection policy into the information technology (IT) infrastructure by deploying various controls and tools to minimize the risk of noncompliance. Data leakage detection, prevention, and protection technologies have emerged over the past few years and are now being widely used within IT organizations.

Data governance, risk management, compliance, and business requirements should drive the number and definition of each data category as well as the requirements for labeling, storage, distribution, disclosure, retention, and destruction. Regulatory and industry rules and standards will clearly play a significant role in the definition process. Other data will require protection, including trade secrets, research, formulas, pre-patent discovery, and various forms of customer and employee information.

Another important aspect of data protection is to understand how data is used within the operations of the organization and in what form they reside (e.g., hard copy, electronic documents, within database). In addition, protection requirements will vary by type of operational environment, such as production, production support, development, quality assurance (QA), or third party.

The requirements for protecting sensitive or confidential data must be clearly defined and reflect the specific requirements within the appropriate regulatory and industry rules and standards or business policies. Specific data elements must be labeled as sensitive and should never be used within their factual state in development, QA, or other nonproduction environments. The data classification policy should clearly identify the requirements for data masking.

Finally, the organization must implement an audit process that will provide periodic independent review to ensure that best practices are followed.

**Best Practices for Data Classification**

Organizations must create a comprehensive set of policies and procedures for the classification of all private, sensitive, and confidential data to adequately protect the organization’s critical data assets. In addition, organizations should implement the following steps:

1. Supply periodic awareness training for employees, contractors, and third-party service providers
2. Integrate procedures into the day-to-day business processes and automate as much of the process as feasible
3. Obtain periodic independent audits to review and report the results to senior management
Structured and Unstructured Data

Sensitive data occurs in two forms: structured and unstructured. Structured sensitive data resides within business applications, databases, enterprise resource planning (ERP) systems, storage devices, third-party service providers, back-up media, and off-site storage. Unstructured sensitive data is dispersed throughout the firm’s infrastructure including desktops, laptops, thumb-drives, and other endpoints.

Organizations must define, implement, and enforce their data classification policy and provide procedures and standards to protect both structured and unstructured sensitive data. For unstructured, they can use end-point security tools to control the use of portable devices and media, content analysis tools to detect the presence of sensitive data, and encryption tools to protect unauthorized access to these devices. For structured data, organizations can use encryption and data masking software.

Data Leakage

Data leakage is the intentional or unintentional release or loss of data to an untrusted third party. Business partners, customers, and employees trust that organizations that hold data about them will take reasonable measures to protect the confidentiality and integrity of their sensitive data and that those organizations must foresee and prevent intentional or unintentional misuse, breach, or theft of the sensitive data.

Available technologies vary from simple blocking devices, paths, ports, other forms of egress and access, and mass encryption of devices, media, and connections to more complex or selective blocking. Technology now exists to monitor content in real time to identify selected information, conditions, people, entitlements, and actions to block, quarantine, encrypt, log, alert, or sanitize data. There are currently two methodologies: scanning data at rest and analyzing data in motion. These technologies can be deployed in many parts of the infrastructure, but are more commonly found at end-point devices and external gateways. End-point devices typically contain removable digital storage devices, hard-copy devices, and various forms of network connectivity that provide access to many internal network resources and in some cases could by-pass internal managed network gateways to get outside the organization. These often become conduits for data leakage.

Some of the major drivers to prevent data leakage come from regulations such as the Gramm-Leach-Bliley Act (BLBA), Health Insurance Portability and Accountability Act (HIPAA), and 37 state breach laws or from industry requirements such as the payment card industry (PCI-DSS) or from national security such as NERC Cyber Security Standard (CIP), DHS, NIST; and corporate policy.

Best Practices for Data Leakage Prevention

Deploy and integrate technology and processes throughout the infrastructure to detect and/or protect sensitive data from leaking out of your enterprise. These steps will require physical and logical controls and technology, changes in routine business and operational processes, and ongoing monitoring and assessment of personnel who have access to sensitive information.
Sensitive Data and Data Masking

It is important to have a common definition of “real data” and what we will refer to as “nonfactual but real data” or “masked data.”

For example, in SAP ERP, data elements define the characteristics of the data including the type, length, and business term such as “first name”, “last name”, or “city”.

The data contained with the actual tables can be factual (e.g., an actual Social Security number) or it can be nonfactual (e.g., a random collection of numerals conforming to the data definition for that particular data element).

Data elements, such as “customer” or “order”, are often related to each other through the use of a key field. When there is an association with many data elements, protecting individual data elements becomes complex. Some data elements alone may not contain sensitive data; however, once an association with other data elements occurs, they all become sensitive data. So data masking software must quickly become very sophisticated to ensure that all the sensitive data is protected (masked), the data is still contextually valuable, and referential integrity is maintained.

Organizations sometimes develop their own data masking tools, which are effective to varying degrees. However, with the continuing onslaught of regulations and risk of fines, negative impact on reputation, and possibility of criminal convictions, organizations have moved toward third-party data masking technologies that are regularly updated according to evolving standards and regulations.

Best Practices for Data Masking

Use a proven commercial solution for masking of sensitive data in both production and nonproduction environments, including development, quality assurance, sandbox systems, training, production support, and production. Choose a vendor that includes

- Support for a wide variety of databases and applications
- A proven track record
- Data discovery capabilities
- Out of the box data masking metadata to expedite project timelines
- Straight-forward easy to learn transformation logic
- A scalable high performing data masking server
- Simple to use and reuse data masking rules
- Built in separation of duties
- Auditing and validation capabilities

Include data masking in your standard data provision process so that sensitive data never resides in your nonproduction environments.

Never provide third parties or offshore teams with sensitive data that is not masked.

Never allow developers or other unauthorized personnel to access production data without dynamically masking sensitive data.
Special Regulatory and Industry Requirements

Depending on the countries that your organization does business in, requirements for the protection of sensitive data will vary. Nearly every country has data privacy laws, so you should always perform a comprehensive review of each of the pertinent laws and their supporting requirements. You will find that there is substantial commonality among these regulations—at least in their spirit or intent. This white paper will focus on some of the more common regulations in North America.

Best Practices for Regulatory and Industry Requirements

Although there are unique aspects to each regulatory item, complying with one may help you comply with (or minimize your exposure to) another. In general, data security techniques are broadly applicable across these regulations and also across a wide variety of industries. What one industry devises has a good chance of being useful to others.

Many financial services firms have been taking data security seriously for a long time. The majority of controls called for by the regulations have simply been considered sound business practices for earning customer trust. The latest regulations only add a few new concepts. Many firms will find themselves well down the road toward compliance.

Because testing and monitoring of controls should be done by parties not directly involved in the design or operation of those controls, there has been a substantial increase of a new business function called IT risk management or IT data governance as a new best practice. The financial services industry has taken the lead, but other industries also have adopted this new function.

Many of the regulations and industry standards have special requirements for third-party service providers. These service providers include IT outsource, third party software vendors as well as providers that fulfill elements of certain business processes (e.g., a business that does promotional mailings for the company and must receive names and addresses).

Your organization is not relieved of its responsibility to protect sensitive information just because the covered information moves into someone else’s hands.

Regulations and industry standards either imply or specify that your organization must adapt to change. Your organization must keep up with changes in the business risk profile, business process, employee training, all forms of threats, technologies, software bugs, and the never-ending flow of software application patches.

Regulatory Requirements

The following three examples of government regulation and industry standards indicate how confidential data should be handled.

Gramm-Leach-Bliley Act (GLBA)

The Gramm-Leach-Bliley Act (GLBA) applies to financial institutions that offer financial products or services such as loans, financial or investment advice, or insurance to individuals. Compliance is mandatory for all nonbank mortgage lenders, loan brokers, financial or investment advisers, tax preparers, debt collectors, and providers of real estate settlements. The law requires that financial institutions protect information that is collected about individuals; it does not apply to information that is collected in business or commercial activities.
**Best Practices for GLBA**

There are three basic rules to understand to ensure compliance with GLBA:

1. Ensure the security and confidentiality of customer records and information.
2. Protect against any anticipated threats or hazards to the security or integrity of such records.
3. Protect against unauthorized access to or use of such records or information that could result in substantial harm or inconvenience to any customer.

Let's look into these three rules further. First, what does “ensure the security and confidentiality of customer records and information” mean?

In general terms, you should take all reasonable measures to guarantee that the privacy of nonpublic personal information in all forms (electronic, hard copy, verbal) is protected from unauthorized access and disclosure.

Second, what does “protect against any anticipated threats or hazards to the security or integrity of such records” mean? Let us explain in layperson terms:

- “Protect against anticipated threats or hazards”—this requires the application of a risk assessment process to foresee the possible known or unknown threats and vulnerabilities in any form (physical, logical, human, or disaster) that may compromise the security and integrity of nonpublic personal information.

- “to the security or integrity of such records”—in this case, security can be defined as anything that would compromise the confidentiality of nonpublic personal information; integrity can be defined as anything that could compromise the trustworthiness, reliability, accuracy, or soundness of nonpublic personal information.

Finally, what does “protect against unauthorized access to or use of such records or information that could result in substantial harm or inconvenience to any customer” mean?

The term “unauthorized access” is very familiar, but who authorizes access? We as consumers and customers play a role in defining who has access to our “nonpublic personal information” and under what conditions they can use this information. We sign and click on agreements every day that define the terms and conditions of granting access to and use of our nonpublic personal information. However, the service providers that we grant these privileges to are required to follow security practices to ensure that “need-to-know” concepts are followed and all unauthorized parties are denied access to our nonpublic personal information.

Another very important best practice is to keep up to date with agency guidance on GLBA; it’s like keeping up with operating system patches, but with less frequency. You must protect personally identifiable information in all environments, including development, QA, and test environments; the regulations do not differentiate among these environments. Remember: static data masking is the safest way to protect sensitive information in development and QA environments.
Health Insurance Portability and Accountability Act (HIPAA)

The HIPAA Security and Privacy Standard defines administrative, physical, and technical safeguards to protect the confidentiality, integrity, and availability of electronic protected health information (PHI), sometimes referred to as personal health information. HIPAA has three major purposes:

- To protect and enhance the rights of consumers by providing them access to their health information and controlling the inappropriate use of that information
- To improve the quality of health care in the United States by restoring trust in the health care system among consumers, health care professionals, and the multitude of organizations and individuals committed to the delivery of care
- To increase the efficiency and effectiveness of health care delivery by creating a national framework for health privacy protection that builds on efforts by states, health systems, and individual organizations and individuals

Best Practices for HIPAA

Understanding the HIPAA Security and Privacy Standard requirements is the key to interpreting what the covered entities must do:

1. Ensure the confidentiality, integrity, and availability of all electronic protected health information that the covered entity creates, receives, maintains, or transmits
2. Protect against any reasonably anticipated threats or hazards to the security or integrity of such information
3. Protect against any reasonably anticipated uses or disclosures of such information
4. Ensure compliance by the workforce

Payment Card Industry Data Security Standard (PCI DSS)

PCI DSS originally began as five different programs by VISA, MasterCard, American Express, Discover, and JCB. Each company’s intentions were roughly similar: to create an additional level of protection for customers by ensuring that merchants meet minimum levels of security when they store, process, and transmit cardholder data. In December 2004, these five companies aligned their individual policies and created the Payment Card Industry Data Security Standard.

The first PCI DSS was introduced in January 2005. The standard is intended to allow merchants, card issuers, card processing companies, and other third-party service providers to demonstrate compliance with a common agreement for information security due care, rather than requiring them to comply with differing requirements from each payment processing company.

All of the five founding members have agreed to incorporate the PCI DSS as the technical requirements of each of their data security compliance programs. Each founding member also recognizes the Qualified Security Assessors (QSAs) and Approved Scanning Vendors (ASVs) certified by the PCI Security Standards Council as being fit to validate compliance to the PCI DSS. The PCI Security Standards Council is an open global forum for the ongoing development, enhancement, storage, dissemination, and implementation of security standards for account data protection.
The PCI DSS is considered one of the more comprehensive data security standards in a cluster of regulations that have emerged over the past decade, and it is regarded as being relatively more prescriptive than other laws and regulations. It covers 6 overall areas and 12 requirements, each supported by lower-level requirements.

**General Compliance Best Practices**

One of the most important aspects of data privacy is to use a risk management approach. If your organization deals with health, financial, or other personal information, then your risk model should be risk averse and therefore, your interpretation of these requirements should lean toward a higher bar for your controls.

Privacy is a subset of confidentiality and, in the spirit of these regulations and industry standards, must be protected from unauthorized access by using the latest industry security technology products and solutions.

Some best-practice technologies are content extrusion at the host and network gateways, encryption or masking of data at rest, network segmentation of sensitive data storage, and logging all access attempts to sensitive data at infrastructure and application levels. Data at rest must be encrypted in all environments and masked in nonproduction environments. Use of tamper-proof technologies in data storage environments is one of the new methods of best practice.

Encryption of portable devices is now an industry standard best practice, and failure to encrypt portable devices containing sensitive data can be seen as being negligent by the regulatory agencies, the courts, and the general public.

Include real and frequent testing of your security infrastructure and IT control environment. Disaster recovery and business continuity plans should be tested by actually switching over to the respective sites at least annually. Some organizations test their disaster recovery plan every quarter.

**Risk Assessment**

Any risk is assumable so long as the risk assumption decision is made by the right person(s) and so long as they are adequately informed.

Risks pose potential consequences that can increase the cost of doing business. The same is true for controls, which can add obvious costs (such as new processes, IT equipment, or software licenses) and can also introduce qualitative costs (such as inconvenience to customers or employees or processing overhead). The cost of a control is justified only if it is less than the avoided costs of compromise.

No control is perfect, no matter how much you spend; it’s another balancing act. Strong controls are usually more costly and almost always much more intrusive into processes and people’s experiences. Least privileged and need to know are well-accepted best practices that are not always easy to implement. If you limit a user’s privileges to the minimum he or she needs to do his or her job, you will have done what you can to minimize the risks associated with that user’s access to your environment. Some companies believe that every employee should be empowered...
to serve the customer in any way. This business choice makes limitation of privileges somewhat moot. People make or break security; no amount of technology can make up for poor practices and behaviors.

Every technical control ultimately relies on some form of fallible human process: to build it, configure it, administer it, and use it.

**Ongoing Compliance Assessment**

Compliance assessment is not a one-time task. It requires a repetitive process to be reasonably effective. Ideally, compliance is fully integrated into daily operations. Our definition of best practice compliance assessment is fairly broad. It starts with understanding the controls:

- First, you need a clear understanding of the purpose of the control (what risk or risks it is mitigating), the specific attributes of the control (settings, parameters), the significance of the control (from a risk perspective), whether there are secondary controls (back-up control that will perform the same level of risk mitigation), and the type of control (technology, process, people).

- There must be specific documented test criteria to ensure that the control, as defined in the first step, has been implemented correctly. The control must, most importantly, be tested to determine its effectiveness in mitigating the risk as defined in step 1.

- The test results must be recorded (a permanent record of the test results must be retained based on your regulatory requirements). There should be an accountability process for this step. In some cases, the recording may be technological. If you are using a software tool to test your controls, then the recording should be protected to ensure the integrity of the results.

- The test results must then be collected into a central repository for correlation and reporting.

- Periodic reporting of the state of controls should be conducted to assess whether compliance has been met or not.

**Securing Applications and Data in Development**

The data processed by and/or housed in applications and databases must be protected. Therefore, applications and databases must be secured. Infrastructure security is necessary for application and database security, but in itself is insufficient. Applications and databases themselves ought to contribute to the overall security model, but augmented by dynamic data masking software.

In nonproduction environments, applications and databases are more open where application level security is not applicable. Some might think that it is enough to strongly secure the infrastructure that hosts the application, but this is simply not true. The applications are at the top of the IT food chain: they are the heart of the business lifeline and the channels to the consumers, business clients, and business partners. The controls start in IT, but in development and QA it is not practical to put in strict controls due to the nature of the user’s job functions. Developers, testers and trainers will require access to many different types of data throughout the organization in order to perform their duties.
**Best Practices for Sensitive Data in Development**

Developing and testing applications demands access to data—lots of it and as real as possible. Copies of real databases usually contain sensitive or confidential data that must be protected. Here are some selected best practices:

- Minimize points where real data is required
- Mask real data to remove personal identification
- Apply very strict controls—similar to production controls—when real data must be used
- Never send sensitive “personally identifiable” data offshore without masking it. U.S. regulations are typically unenforceable outside the United States

Except under rare circumstances, developers and administrators have no defensible reason to view sensitive or confidential personal information. This rule applies to structured data and unstructured data. If they do, then it is probably a clear violation of regulations, maybe even the organization’s privacy policy. Corporate policy should state that even if they can, they must not. Severe penalties should accompany violations of this principle.

**Documentation**

Documentation is not free. It takes legwork upfront and maintenance: time, effort, and an investment.

However, it is undoubtedly a regulatory requirement. The Auditing Standards Board (on which the Public Company Accounting Oversight Board and SEC are likely to rely) has said that the “lack of documentation” is a clear regulatory deficiency in the context of the Sarbanes-Oxley Act. Whether the controls are effective or not, documentation is required. If not documented, then how can management assert that they know the state of control versus the intended state?

**Best Practices for Documentation**

Good documentation is a valuable business/technical resource and best practice. Here are some examples of what to document:

- Document procedures and processes, both business and IT
- Keep records of controls and the decision processes behind them
- Furnish support for IT and security architectures and strategies
- Document monitoring and oversight activities and control status
- Document sensitive data definitions, policies to carry out for that sensitive data and where sensitive data is located
System Administrators and Developers

System administrators who can also administer security components (e.g., add, change, and delete users and give them rights/privileges) can do anything on that platform and/or the applications it supports. If they can also alter the logs, then they can cover up any unauthorized access to sensitive data.

Developers who also test their own code may be tempted to overlook certain faults or avoid certain tests that might make their work look bad or cause them to miss deadlines. Or they could abuse this combined role to move malicious code through the test phase and possibly into production.

Testing in the production environments is of course not a best practice. Such testing has a high risk of breaking things or launching bogus transactions. The same principle applies for testing against live networks (intranet or Internet). Best practice: run the final test of your applications in a segregated preproduction environment that is a replica of your production environment (network, computing, database, and application) infrastructure, and also ensure that your sensitive data is masked.

Developers who have access to production have the power to unilaterally cause harm perhaps without an audit trail and can make mistakes with extremely serious consequences. There are rare situations when developers really do have to patch production code. Best practice: a fire call (“break glass”) is the accepted way to manage this. The problem/change ticket must be completed ASAP following fire call access. Fire call passwords are changed after use. In addition, dynamic data masking will ensure that during these fire calls, the developer or other privileged IT user will not see any sensitive data.

Developers who are also administrators may not only be able to create malicious code but could also use their admin rights to alter logs that might hold clues to the attack. The broader the powers and rights you have, the more likely that an error or mistake will have serious consequences. Small organizations may be forced to combine roles due to the limited number of qualified staff. This is a great example of why “best practice” is hard to define. Yet with the advent of both static and dynamic data masking, organizations can go a long way to ensure that sensitive data is protected.

Best Practices for System Administrators and Developers

Developers and administrators must often have root access to many systems (trusted by other systems) to do their job. They possess the technical skill to use and abuse privileges. Some selected best practices:

• Isolate development environments and monitor them closely
• Limit test Internet connections
• Mask all sensitive data in nonproduction
• Configure development platforms securely
• Motivate developers to use good security practices
• Introduce dynamic data masking in all production or near production environments
• Find alternatives to shared accounts for administrators and support functions

ABOUT INFORMATICA

Informatica enables organizations to operate more efficiently in today’s global information economy by empowering them to access, integrate, and trust all their information assets. As the independent data integration leader, Informatica has a proven track record of success helping the world’s leading companies leverage all their information assets to grow revenues, improve profitability, and increase customer loyalty.