National Institute of
Standards and Technology
U.S. Department of Commerce

Special Publication 800-122
(Draft)

Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) (Draft)

Recommendations of the National Institute of Standards and Technology

Erika McCallister
Tim Grance
Karen Scarfone

NIST Special Publication 800-122 (Draft)

Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) (Draft)

Recommendations of the National Institute of Standards and Technology

Erika McCallister
Tim Grance
Karen Scarfone

C O M P U T E R S E C U R I T Y

Computer Security Division
Information Technology Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899-8930

January 2009

U.S. Department of Commerce

Carlos M. Gutierrez, Secretary

National Institute of Standards and Technology

Dr. Patrick D. Gallagher, Deputy Director

Reports on Computer Systems Technology

The Information Technology Laboratory (ITL) at the National Institute of Standards and Technology (NIST) promotes the U.S. economy and public welfare by providing technical leadership for the nation's measurement and standards infrastructure. ITL develops tests, test methods, reference data, proof of concept implementations, and technical analysis to advance the development and productive use of information technology. ITL's responsibilities include the development of technical, physical, administrative, and management standards and guidelines for the cost-effective security and privacy of sensitive unclassified information in Federal computer systems. This Special Publication 800-series reports on ITL's research, guidance, and outreach efforts in computer security and its collaborative activities with industry, government, and academic organizations.

National Institute of Standards and Technology Special Publication 800-122 (Draft)
Natl. Inst. Stand. Technol. Spec. Publ. 800-122, 58 pages (Jan. 2009)

Acknowledgements

The authors, Erika McCallister, Tim Grance, and Karen Scarfone of the National Institute of Standards and Technology (NIST), wish to thank their colleagues who reviewed drafts of this document and contributed to its technical content. Of particular note are the efforts of Joseph Nusbaum of Innovative Analytics & Training, Deanna DiCarlantonio of CUNA Mutual Group, and Michael L. Shapiro and Daniel I. Steinberg of Booz Allen Hamilton, who contributed significant portions to previous versions of the document. The authors would also like to acknowledge Ron Ross, Kelley Dempsey, and Arnold Johnson of NIST; Michael Gerdes, Beth Mallory, and Victoria Thompson of Booz Allen Hamilton; and Julie McEwen and Aaron Powell of MITRE, for their keen and insightful assistance throughout the development of the document. Additional acknowledgements will be added to the final version of the publication.

Note to Reviewers

This publication contains several examples of determining the PII confidentiality impact level to assign to various instances of PII. These examples are intended to illustrate the factors to consider when deciding how to protect the confidentiality of PII, and are not intended to define how certain types of data should always be protected. Every situation has unique characteristics that may affect the assigned impact level and the corresponding protective measures applied to the PII. An organization's legal counsel and privacy officer should be consulted when determining whether there are legal obligations to protect the confidentiality of PII. The authors welcome feedback on the examples, such as different opinions on the appropriate impact levels and suggestions for additional examples that would be helpful to readers. Finally, the authors are also seeking suggestions for feasible technical solutions for logging and verifying sensitive database extracts, as described in Appendix E. NIST thanks the reviewers in advance for sharing their expertise and valuable time to perform this public service.

Table of Contents

Executive Summary **ES-1**

1. Introduction **1-1**

1.1 Authority 1-1
1.2 Purpose and Scope 1-1
1.3 Audience 1-1
1.4 Document Structure 1-1

2. Introduction to PII **2-1**

2.1 Identifying PII 2-1
2.2 Examples of PII Data 2-2
2.3 PII and Fair Information Practices 2-2

3. PII Confidentiality Impact Levels **3-1**

3.1 Impact Level Definitions 3-1
3.2 Factors for Determining PII Confidentiality Impact Levels 3-2
3.2.1 Distinguishability 3-3
3.2.2 Aggregation and Data Field Sensitivity 3-3
3.2.3 Context of Use 3-3
3.2.4 Obligation to Protect Confidentiality 3-4
3.2.5 Access to and Location of the PII 3-4
3.3 PII Confidentiality Impact Level Examples 3-5
3.3.1 Example 1: Incident Response Roster 3-5
3.3.2 Example 2: Intranet Activity Tracking 3-6
3.3.3 Example 3: Fraud, Waste, and Abuse Reporting Application 3-6

4. PII Confidentiality Protection Measures **4-1**

4.1 General Protection Measures 4-1
4.1.1 Policy and Procedure Creation 4-1
4.1.2 Education, Training, and Awareness 4-2
4.2 Privacy-Specific Protection Measures 4-3
4.2.1 Minimizing Collection and Retention of PII 4-3
4.2.2 De-Identifying Information 4-4
4.2.3 Anonymizing Information 4-5
4.3 Security Controls 4-6

5. Incident Response for Breaches of PII **5-1**

5.1 Preparation 5-1
5.2 Detection and Analysis 5-2
5.3 Containment, Eradication, and Recovery 5-3
5.4 Post-Incident Activity 5-3

Appendices

Appendix A— Scenarios for PII Identification and Handling A-1

A.1 General Questions A-1
A.2 Scenarios A-1

Appendix B— Frequently Asked Questions (FAQ) B-1

Appendix C— Definitions of Private Information C-1

Appendix D— Fair Information Practices D-1

Appendix E— Sensitive Database Extracts Technical Frequently Asked Questions E-1

Appendix F— Glossary F-1

Appendix G— Acronyms and Abbreviations G-1

Appendix H— Resources H-1

Executive Summary

Breaches of personally identifiable information (PII) have increased dramatically over the past few years and have resulted in the loss of millions of records.[1] Breaches of PII are hazardous to both individuals and organizations. Individual harms may include identity theft, embarrassment, or blackmail. Organizational harms may include a loss of public trust, legal liability, or high costs to handle the breach. To appropriately protect the confidentiality of PII, organizations should use a risk-based approach; as McGeorge Bundy[2] once stated, "If we guard our toothbrushes and diamonds with equal zeal, we will lose fewer toothbrushes and more diamonds." This document provides guidelines for a risk-based approach to protecting the confidentiality[3] of PII.

The recommendations in this document are intended primarily for U.S. Federal government agencies and those who conduct business on behalf of the agencies,[4] but other organizations may find portions of the publication useful. Each organization may be subject to a different combination of laws, regulations, and other mandates related to protecting PII, so an organization's legal counsel and privacy officer should be consulted to determine the current obligations for PII protection. For example, the Office of Management and Budget (OMB) has issued several memoranda with requirements for how Federal agencies must handle and protect PII.

To effectively protect PII, organizations should implement the following recommendations.

Organizations should identify all PII residing in their environment.

An organization cannot properly protect PII it does not know about. This document uses the broad definition of PII from OMB Memorandum 07-16[5] to identify as many potential sources of risks related to PII as possible. OMB defined PII as "information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc." Examples of PII include, but are not limited to:

- Name, such as full name, maiden name, mother's maiden name, or alias
- Personal identification number, such as social security number (SSN), passport number, driver's license number, taxpayer identification number, or financial account or credit card number
- Address information, such as street address or email address
- Personal characteristics, including photographic image (especially of face or other distinguishing characteristic), fingerprints, handwriting, or other biometric image or template data (e.g., retina scans, voice signature, facial geometry).

---

1 Government Accountability Office (GAO) Report 08-343, *Protecting Personally Identifiable Information*, January 2008, http://www.gao.gov/new.items/d08343.pdf

2 Congressional testimony as quoted by the New York Times, March 5, 1989. McGeorge Bundy was the U.S. National Security Advisor to Presidents Kennedy and Johnson (1961-1966). http://query.nytimes.com/gst/fullpage.html?res=950DE2D6123AF936A35750C0A96F948260

3 For the purposes of this document, confidentiality is defined as "preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information." 44 U.S.C. § 3542. http://uscode.house.gov/download/pls/44C35.txt.

4 For the purposes of this publication, both are referred to as "organizations".

5 OMB Memorandum 07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information*, http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf

Organizations should categorize their PII by the PII confidentiality impact level.

All PII is not created equal. PII should be evaluated to determine its PII confidentiality impact level so that appropriate safeguards can be applied to the PII. The PII confidentiality impact level—*low, moderate, or high*—indicates the potential harm that could result to the subject individuals and/or the organization if the PII were inappropriately accessed, used, or disclosed. This document provides a list of factors an organization should consider when determining the PII confidentiality impact level. Each organization should decide which factors it will use for determining impact levels and then create and implement the appropriate policy, procedures, and controls. The following are examples of factors:

- **Distinguishability.** Organizations should evaluate how easily the PII can be used to distinguish particular individuals. For example, an SSN uniquely identifies an individual, whereas a telephone area code could map to many people.
- **Aggregation and Data Field Sensitivity.** Organizations should evaluate the sensitivity of each individual PII data field, as well as the sensitivity of the PII data fields when combined. For example, an individual's SSN or financial account number is generally more sensitive than an individual's phone number or zip code. Similarly, the combination of an individual's name and financial account number is more sensitive than the individual's name alone.
- **Context of Use.** Organizations should evaluate the context of use, which is the purpose for which the PII is collected, stored, used, processed, disclosed, or disseminated. The context of use may cause identical PII data elements to be assigned different PII confidentiality impact levels based on their use. For example, suppose that an organization has two lists that contain the same PII data fields (e.g., name, address, phone number). The first list is people who subscribe to a general-interest newsletter produced by the organization, and the second list is people who work undercover in law enforcement. The potential impacts to the affected individuals and to the organization are significantly different for each list.
- **Obligations to Protect Confidentiality.** An organization that is subject to any obligations to protect PII should consider such obligations when determining the PII confidentiality impact level. Obligations to protect generally include laws, regulations, or other mandates (e.g., Privacy Act, OMB guidance). For example, some Federal agencies, such as the Census Bureau and the Internal Revenue Service (IRS), are subject to specific legal obligations to protect certain types of PII.[6]
- **Access to and Location of PII.** Organizations may choose to take into consideration the nature of authorized access to and the location of the PII. When PII is accessed more often or by more people and systems, or the PII is regularly transmitted or transported offsite, then there are more opportunities to compromise the confidentiality of the PII.

Organizations should apply the appropriate safeguards for PII based on the PII confidentiality impact level.

Not all PII should be protected in the same way. Organizations should apply appropriate safeguards to protect the confidentiality of the PII based on the PII confidentiality impact level. Some PII does not need to have its confidentiality protected, such as information that the organization has permission or authority to release publicly (e.g., an organization's public phone directory). NIST recommends using

[6] The Census Bureau has a special obligation to protect based on provisions of Title 13 of the U.S. Code, and IRS has a special obligation to protect based on Title 26 of the U.S. Code. There are more agency-specific obligations to protect PII, and an organization's legal counsel and privacy officer should be consulted.

general protection measures, privacy-specific protection measures, and security controls[7] used for other types of information, such as:

- **Creating Policies and Procedures.** Organizations should develop comprehensive policies and procedures for protecting the confidentiality of PII.
- **Conducting Training.** Organizations should reduce the possibility that PII will be accessed, used, or disclosed inappropriately by requiring that all individuals receive appropriate training before being granted access to organization information systems.
- **De-Identifying PII.** Organizations can de-identify records by removing enough PII such that the remaining information does not identify an individual and there is no reasonable basis to believe that the information can be used to identify an individual. De-identified records can be used when full data records are not necessary, such as for examinations of correlations and trends.
- **Using Access Enforcement.** Organizations can control access to PII through access control policies and access enforcement mechanisms (e.g., access control lists).
- **Implementing Access Control for Mobile Devices.** Organizations can prohibit or strictly limit access to PII from portable and mobile devices, such as laptops, cell phones, and personal digital assistants (PDA), which are generally higher-risk than non-portable devices (e.g., desktop computers at the organization's facilities). Organizations may choose to forbid all telework and remote access involving higher-impact instances of PII so that the information will not leave the organization's physical boundaries through telework activities.
- **Providing Transmission Confidentiality.** Organizations can protect the confidentiality of transmitted PII. This is most often accomplished by encrypting the communications or by encrypting the information before it is transmitted.
- **Auditing Events.** Organizations can monitor events that affect the confidentiality of PII, such as inappropriate access to PII.

Organizations should minimize the collection and retention of PII to what is strictly necessary to accomplish their business purpose and mission.

The likelihood of harm caused by a breach of PII is greatly reduced if an organization minimizes the amount of PII it collects and stores. Organizations should limit PII collection and retention to the least amount necessary to conduct their business purpose and mission. For example, an organization should only request PII on a new form if the PII is absolutely necessary. Also, an organization should regularly review its holdings of previously collected PII to determine whether the PII is still relevant and necessary for meeting the organization's business purpose and mission. For example, organizations could have an annual PII purging awareness day.[8]

OMB M-07-16 specifically requires agencies to:

- Review current holdings of PII and ensure they are accurate, relevant, timely, and complete
- Reduce PII holdings to minimum necessary for proper performance of agency functions
- Develop a schedule for periodic review of PII holdings

[7] This document provides some selected security control examples from NIST SP 800-53.

[8] Disposal of PII should be conducted in accordance with the retention schedules approved by the National Archives and Records Administration (NARA).

- Establish a plan to eliminate the unnecessary collection and use of SSNs.

Organizations should develop an incident response plan to handle breaches of PII.

Breaches of PII are hazardous to both individuals and organizations. Harm to individuals and organizations can be contained and minimized through the development of effective incident response plans for PII breaches. Organizations should develop plans[9] that include elements such as determining when and how individuals should be notified, when and if a breach should be reported publicly, and whether to provide remedial services, such as credit monitoring, to affected individuals. Organizations should integrate these additional policies into their existing incident handling policies.

Organizations should encourage close coordination among their privacy officers, chief information officers, information security officers, and legal counsel[10] when addressing issues related to PII.

Protecting the confidentiality of PII requires knowledge of information systems, information security, privacy, and legal requirements. Decisions regarding the applicability of a particular law, regulation, or other mandate should be made in consultation with an organization's legal counsel and privacy officer because relevant laws, regulations, and other mandates are often complex and change over time. Additionally, new policies often require the implementation of technical security controls to enforce the policies. Close coordination of the relevant experts helps to prevent PII breaches by ensuring proper interpretation and implementation of requirements.

[9] OMB M-07-16 requires agencies to develop and implement breach notification policies.

[10] Some organizations are structured differently and have different names for roles. These roles are examples, used for illustrative purposes.

1. Introduction

1.1 Authority

The National Institute of Standards and Technology (NIST) developed this document in furtherance of its statutory responsibilities under the Federal Information Security Management Act (FISMA) of 2002, Public Law 107-347.

NIST is responsible for developing standards and guidelines, including minimum requirements, for providing adequate information security for all agency operations and assets, but such standards and guidelines shall not apply to national security systems. This guideline is consistent with the requirements of the Office of Management and Budget (OMB) Circular A-130, Section 8b(3), "Securing Agency Information Systems," as analyzed in A-130, Appendix IV: Analysis of Key Sections. Supplemental information is provided in A-130, Appendix III.

This guideline has been prepared for use by Federal agencies, also referred to as organizations in the guide. It may be used by nongovernmental organizations on a voluntary basis and is not subject to copyright, though attribution is desired.

Nothing in this document should be taken to contradict standards and guidelines made mandatory and binding on Federal agencies by the Secretary of Commerce under statutory authority, nor should these guidelines be interpreted as altering or superseding the existing authorities of the Secretary of Commerce, Director of the OMB, or any other Federal official.

1.2 Purpose and Scope

The purpose of this document is to assist Federal agencies in protecting the confidentiality of a specific category of data commonly known as personally identifiable information (PII). PII should be protected from inappropriate access, use, and disclosure. This document provides practical, context-based guidance for identifying PII and determining what level of protection is appropriate for each instance of PII. The document also suggests safeguards that may offer appropriate levels of protection for PII and provides recommendations for developing response plans for breaches involving PII. Organizations are encouraged to tailor the recommendations to meet their specific requirements.

1.3 Audience

The primary audience for this document is the individuals who apply policies and procedures for protecting the confidentiality of PII on Federal information systems, as well as technical and non-technical personnel involved with implementing system-level changes concerning PII protection methods. Individuals in many roles should find this document useful, including chief privacy officers and other privacy officers, privacy advocates, privacy support staff, compliance officers, system administrators, chief information system security officers, information system security officers, information security support staff, computer security incident response teams, and chief information officers.

1.4 Document Structure

The remainder of this document is organized into the following sections:

- Section 2 provides an introduction to PII and lists some basic requirements involving the collection and handling of PII.

- Section 3 describes factors for determining the potential impact of inappropriate access, use, and disclosure of PII.
- Section 4 presents several methods for protecting the confidentiality of PII that can be implemented to reduce PII exposure and risk.
- Section 5 provides recommendations for developing an incident response plan for breaches involving PII and integrating the plan into an organization's existing incident response plan.

The following appendices are also included for additional information:

- Appendix A provides samples of PII-related scenarios and questions that can be adapted for an organization's exercises.
- Appendix B presents frequently asked questions (FAQ) related to protecting the confidentiality of PII.
- Appendix C contains definitions of common general terms related to private information.
- Appendix D provides additional information about the Fair Information Practices that may be helpful in understanding the framework underlying most privacy laws.
- Appendix E contains a FAQ pertaining to logging and verifying sensitive database extracts.
- Appendix F provides a glossary of selected terms from the publication.
- Appendix G contains a list of acronyms and abbreviations used within the publication.
- Appendix H presents a list of resources that may be helpful to individuals in gaining a better understanding of PII, PII protection, and other related topics.

2. Introduction to PII

One of the most broadly used terms to describe personal information about individuals is PII. Examples of PII range from an individual's name or email address to an individual's financial and medical records or criminal history. Unauthorized access, use, or disclosure of PII can seriously impact both individuals, by contributing to identity theft, and the organization, by reducing public trust in the organization. In many cases, it may not be clear to the professionals responsible for protecting information which instances of PII need additional confidentiality protection and at what level. This section explains how to identify and locate PII[11] maintained within an organization's environment and/or under its control, and it provides an introduction to the Fair Information Practices. Sections 3 and 4 discuss factors for assigning PII impact levels and selecting protection measures, respectively. Section 5 discusses incident response for breaches involving PII.

2.1 Identifying PII

This publication uses the definition of PII from OMB Memorandum 07-16,[12] which is "information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc."

To distinguish an individual[13] is to identify an individual. Some examples of information that could distinguish an individual include, but are not limited to, name, passport number, social security number, or biometric image and template. In contrast, a list containing only credit scores does not have sufficient information to distinguish a specific individual.

Information elements that are linked or linkable are not sufficient to distinguish an individual when considered separately, but which could distinguish individuals when combined with a secondary information source. For example, suppose that two databases contain different PII elements and also share some common PII elements. An individual with access to both databases may be able to link together information from the two databases and distinguish individuals. If the secondary information source is present on the same system or a closely-related system, then the data is considered *linked*. If the secondary source is available to the general public or can be obtained, such as from an unrelated system within the organization, then the data is considered *linkable*. Linked data is often de-identified in some way (as described in Section 4), and information that makes re-identification possible is available to some system users. Linkable data is also often de-identified, but the remaining data can be analyzed against other data sources, such as telephone directories and other sources available to large communities of people, to distinguish individuals.

Organizations should use a variety of methods to identify all PII residing within their organization or under the control of their organization through a third party (e.g., a system being developed and tested by a contractor). Privacy threshold analyses (PTAs), also referred to as initial privacy assessments (IPAs), are often used to identify PII.[14] Some organizations require a PTA to be completed before the

---

[11] Even if an organization determines that information is not PII, the organization should still consider whether the information is sensitive or has organizational or individual risks associated with it, and determine the appropriate protections.

[12] OMB Memorandum M-07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information*, http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf.

[13] The terms "individual" and "individual's identity" are used interchangeably throughout this document. For additional information about the term *individual*, see Appendix B.

[14] For example PTA/IPA templates, see: http://www.usdoj.gov/opcl/initial-privacy-assessment.pdf or http://www.dod.mil/pubs/foi/privacy/DHS_PTA_Template.pdf.

development or acquisition of a new information system and when a substantial change is made to an existing information system. PTAs are used to determine if a system contains PII, whether a Privacy Impact Assessment is required, whether a System of Records Notice (SORN) is required, and if any other privacy requirements apply to the information system. PTAs should be submitted to an organization's privacy office for review and approval. PTAs are often comprised of simple questionnaires that are completed by the system owner. PTAs are useful in initiating the communication and collaboration for each system between the privacy officer, the information security officer, and the information officer. Other examples of methods to identify PII include reviewing system documentation, conducting interviews, conducting data calls, or checking with system owners.

2.2 Examples of PII Data

The following list contains examples of information that may be considered PII.[15]

- Name, such as full name, maiden name, mother's maiden name, or alias
- Personal identification number, such as SSN, passport number, driver's license number, taxpayer identification number, patient identification number, and financial account or credit card number[16]
- Address information, such as street address or email address
- Asset information, such as Internet Protocol (IP) or Media Access Control (MAC) address or other host-specific persistent static identifier that consistently links to a particular person or small, well-defined group of people
- Telephone numbers, including mobile, business, and personal numbers
- Personal characteristics, including photographic image (especially of face or other distinguishing characteristic), x-rays, fingerprints, or other biometric image or template data (e.g., retina scans, voice signature, facial geometry)
- Information identifying personally owned property, such as vehicle registration or identification number, and title numbers and related information
- Information about an individual that is linked or linkable to one of the above (e.g., date of birth, place of birth, race, religion, weight, activities, or employment, medical, education, or financial information).

2.3 PII and Fair Information Practices

The protection of PII and the overall privacy of records are concerns both for individuals whose personal records are at stake and for organizations that may be liable or have their reputations damaged should such PII be inappropriately accessed, used, or disclosed. Treatment of PII is distinct from other types of data because it needs to be not only protected, but also collected, maintained, and disseminated in accordance with Federal law.[17] The Privacy Act, as well as other privacy laws, is based on the widely-recognized Fair Information Practices, also called Privacy Principles. There are five core Fair

---

[15] As discussed in Section 3, the risk posed by these examples and the appropriate protections needed for each vary on a case-by-case basis.

[16] Partial identifiers, such as the first few digits or the last few digits of SSNs, are also often considered PII because they are still nearly unique identifiers and are linked or linkable to a specific individual.

[17] This document focuses on protecting the confidentiality of PII. Protecting the privacy of PII is a broader subject, and information about the Fair Information Practices is provided to increase reader awareness.

Information Practices[18] that are based on the common elements, or privacy principles, of several international reports and guidelines. These core practices are as follows:

- **Notice/Awareness**—Individuals should be given notice of an organization's information practices before any personal information is collected from them.
- **Choice/Consent**—Individuals should be given a choice about how information about them is used.
- **Access/Participation**—Individuals should have the right to access information about them and request correction to ensure the information is accurate and complete.
- **Integrity/Security**—Data collectors should ensure that information is protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data.
- **Enforcement/Redress**—Data collectors should be held accountable for complying with measures that give effect to the practices stated above.

For more information on the Fair Information Practices, including a summary of variations of the Fair Information Practices, see Appendix D.

[18] See: http://www.ftc.gov/reports/privacy3/fairinfo.shtm.

3. PII Confidentiality Impact Levels

This publication focuses on protecting PII from losses of confidentiality. The security objective of confidentiality is defined by law as "preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information."[19] The security objectives of integrity and availability may also be important for PII, and organizations should use the NIST Risk Management Framework to determine the appropriate integrity and availability impact levels. Organizations may also need to consider PII-specific enhancements to the integrity or availability impact levels. For example, malicious alterations of medical test results could endanger individuals' lives.

The confidentiality of PII should be protected based on its risk level. This section outlines factors for determining the PII confidentiality impact level for a particular instance of PII, which is distinct from the confidentiality impact level described in Federal Information Processing Standards (FIPS) Publication 199, *Standards for Security Categorization of Federal Information and Information Systems.*[20] The PII confidentiality impact level takes into account additional PII considerations and should be used to determine if additional protections should be implemented. The PII confidentiality impact level—*low, moderate, or high*—indicates the potential harm that could result to the subject individuals and/or the organization if the PII were inappropriately accessed, used, or disclosed. Once the PII confidentiality impact level is selected, it should be used to supplement the provisional confidentiality impact level, which is determined from information and system categorization processes outlined in FIPS 199 and NIST Special Publication (SP) 800-60, *Volumes 1 and 2: Guide for Mapping Types of Information and Information Systems to Security Categories.*[21]

Some PII does not need to have its confidentiality protected, such as information that the organization has permission or authority to release publicly (e.g., an organization publishing a phone directory of employees' names and work phone numbers so that members of the public can contact them directly). In this case, the PII confidentiality impact level would be *not applicable* and would not be used to supplement a system's provisional confidentiality impact level. PII that does not require confidentiality protection may still require other security controls to protect the integrity and the availability of the information, and the organization should provide appropriate security controls based on the assigned FIPS 199 impact levels.

3.1 Impact Level Definitions

The harm caused from of a loss of confidentiality should be considered when attempting to determine which PII confidentiality impact level corresponds to a specific set of PII data. *Harm* for the purposes of this document, includes any adverse effects that would be experienced by an individual whose PII was the subject of a loss of confidentiality, as well as any adverse effects experienced by the organization that maintains the PII. Harm to an individual includes any negative or unwanted effects (i.e., that may be socially, physically, or financially damaging). Examples of types of harm to individuals include, but are not limited to, the potential for blackmail, identity theft, physical harm, discrimination, or emotional distress. Organizations may also experience harm as a result of a loss of confidentiality of PII maintained by the organization—including but not limited to administrative burden, financial losses, loss of public reputation and public confidence, and civil liability.

[19] 44 U.S.C. § 3542, http://uscode.house.gov/download/pls/44C35.txt
[20] http://csrc.nist.gov/publications/PubsFIPS.html
[21] http://csrc.nist.gov/publications/PubsSPs.html

The following describe the three impact levels—low, moderate, and high—defined in FIPS 199, which are based on the potential impact of a security breach involving a particular system:[22]

> "The potential impact is **LOW** if the loss of confidentiality, integrity, or availability could be expected to have a **limited adverse effect** on organizational operations, organizational assets, or individuals. A limited adverse effect means that, for example, the loss of confidentiality, integrity, or availability might (i) cause a degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is noticeably reduced; (ii) result in minor damage to organizational assets; (iii) result in minor financial loss; or (iv) result in minor harm to individuals.
>
> The potential impact is **MODERATE** if the loss of confidentiality, integrity, or availability could be expected to have a **serious adverse effect** on organizational operations, organizational assets, or individuals. A serious adverse effect means that, for example, the loss of confidentiality, integrity, or availability might (i) cause a significant degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is significantly reduced; (ii) result in significant damage to organizational assets; (iii) result in significant financial loss; or (iv) result in significant harm to individuals that does not involve loss of life or serious life threatening injuries.
>
> The potential impact is **HIGH** if the loss of confidentiality, integrity, or availability could be expected to have a **severe or catastrophic adverse effect** on organizational operations, organizational assets, or individuals. A severe or catastrophic adverse effect means that, for example, the loss of confidentiality, integrity, or availability might (i) cause a severe degradation in or loss of mission capability to an extent and duration that the organization is not able to perform one or more of its primary functions; (ii) result in major damage to organizational assets; (iii) result in major financial loss; or (iv) result in severe or catastrophic harm to individuals involving loss of life or serious life threatening injuries."

Harm to individuals as described in these impact levels is easier to understand with examples. A breach of the confidentiality of PII at the low impact level would not cause harm greater than inconvenience, such as changing a telephone number. The types of harm that could be caused by a breach of PII at the moderate impact level include financial loss due to identity theft or denial of benefits, public humiliation, discrimination, and the potential for blackmail. Harm at the high impact level involves serious physical, social, or financial harm, resulting in potential loss of life or inappropriate physical detention.

3.2 Factors for Determining PII Confidentiality Impact Levels

Determining the PII confidentiality impact level should take into account relevant factors. Several important factors that organizations should consider are described below. It is important to note that relevant factors should be considered together; one factor by itself might indicate a low impact level, but another factor might indicate a high impact level, and thus override the first factor. Also, the impact levels suggested for these factors are for illustrative purposes; each instance of PII is different, and each organization has a unique set of requirements and a different mission. Therefore, organizations should determine which factors, including organization-specific factors, they should use for determining PII confidentiality impact levels and should create and implement policy and procedures that support these determinations.

[22] This document pertains only to the confidentiality impact and does not address integrity or availability.

3.2.1 Distinguishability

Organizations should evaluate how easily the PII can be used to distinguish particular individuals. For example, PII data composed of individuals' names, fingerprints, and SSNs uniquely identify individuals, whereas PII data composed of individuals' phone numbers only would require the use of additional data sources, such as phone directories, and would only allow some unique individuals to be identified (for example, unique identification might not be possible if multiple individuals share a phone or if a phone number is unlisted). PII data composed of only individuals' area codes and gender would not allow any unique individuals to be identified.[23] PII that is easily distinguishable may merit a higher impact level than PII that cannot be used to distinguish individuals without unusually extensive efforts.

Organizations may also choose to consider how many individuals can be distinguished from the PII data. Breaches of 25 records and 25 million records may have different impacts, not only in terms of the collective harm to individuals but also in terms of harm to the organization's reputation and the cost to the organization in addressing the breach. For this reason, organizations may choose to set a higher impact level for particularly large PII data sets than would otherwise be set. However, organizations should not set a lower impact level for a PII data set simply because it contains a small number of records.

3.2.2 Aggregation and Data Field Sensitivity

Organizations should evaluate the sensitivity of each individual PII data field, as well as the sensitivity of the PII data fields together. For example, an individual's SSN or financial account number is generally more sensitive than an individual's phone number or zip code, and the combination of an individual's name and SSN is less sensitive than the combination of an individual's name, SSN, date of birth, mother's maiden name, and credit card number. Organizations often require the PII confidentiality impact level to be set to at least moderate if a certain sensitive data field, such as SSN, is present. Organizations may also consider certain combinations of PII data fields to be more sensitive, such as name and credit card number, than each data field would be considered without the existence of the others.

3.2.3 Context of Use

Context of use is defined as the purpose for which the PII is collected, stored, used, processed, disclosed, or disseminated, as well as how that PII is used or could potentially be used. Examples of context include, but are not limited to, statistical analysis, determining eligibility for benefits, administration of benefits, research, tax administration, or law enforcement. Organizations should assess the context of use because it is important to understanding how the disclosure of data elements can potentially harm individuals and the organization. Organizations should consider what harm is likely to be caused if the PII is disclosed (either intentionally or accidentally) or if the mere fact that the PII is being collected or used is disclosed could cause harm to the organization or individual. For example, law enforcement investigations could be compromised if the mere fact that information is being collected about a particular individual is disclosed.

The context of use may cause multiple instances of the same types of PII data to be assigned different PII confidentiality impact levels. For example, suppose that an organization has three lists that contain the same PII data fields (e.g., name, address, phone number). The first list is people who subscribe to a general-interest newsletter produced by the organization. The second list is people who have filed for retirement benefits, and the third list is individuals who work undercover in law enforcement. The potential impacts to the affected individuals and to the organization are significantly different for each of

[23] Section 4.2 discusses how organizations can reduce the need to protect PII by removing PII from records.

the three lists. Based on context of use only, the three lists are likely to merit impact levels of low, moderate, and high, respectively.

Examples of topics that are relevant to context of use as a factor for determining PII confidentiality impact level are abortion; alcohol, drug, or other addictive products; illegal conduct; illegal immigration status; information damaging to financial standing, employability, or reputation; information leading to social stigmatization or discrimination; politics; psychological well-being or mental health; religion; same-sex partners; sexual behavior; sexual orientation; taxes; and other information due to specific cultural or other factors.[24]

3.2.4 Obligation to Protect Confidentiality

An organization that is subject to any obligations to protect PII should consider such obligations when determining the PII confidentiality impact level. Many organizations are subject to laws, regulations, or other mandates[25] governing the obligation to protect personal information,[26] such as the Privacy Act of 1974, OMB memoranda, and the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Additionally, some Federal agencies, such as the Census Bureau and the Internal Revenue Service (IRS), are subject to additional specific legal obligations to protect certain types of PII.[27] Some organizations are also subject to specific legal requirements based on their role. For example, organizations acting as financial institutions by engaging in financial activities are subject to the Gramm-Leach-Bliley Act (GLBA).[28] Also, some agencies that collect PII for statistical purposes are subject to the strict confidentiality requirements of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA).[29] Violations of many of these laws can result in civil or criminal penalties. Organizations may also be obliged to protect PII by their own policies, standards, or management directives.

For example, a database with PII for beneficiaries of government services that retrieves information by SSN would be considered a System of Records under the Privacy Act of 1974, and the organization would be required to provide appropriate administrative, technical, and physical safeguards for the database. Decisions regarding the applicability of a particular law, regulation, or other mandate should be made in consultation with an organization's legal counsel and privacy officer because relevant laws, regulations, and other mandates are often complex and change over time.

3.2.5 Access to and Location of the PII

Organizations may choose to take into consideration the nature of authorized access to the PII. When PII is accessed more often or by more people and systems, there are more opportunities for the PII's confidentiality to be compromised. Another element is the scope of access to the PII, such as whether the PII needs to be accessed from teleworkers' systems and other systems outside the direct control of the organization. These considerations could cause an organization to assign a higher impact level to widely-

---

24 See *Guide to U.S. Census Bureau Data Stewardship/Privacy Impact Assessments (DS/PIAs)*, http://www.census.gov/po/pia/Guide_to_PIAs.doc

25 See Appendix H for additional resources.

26 Personal information is defined in different ways by different laws, regulations, and other mandates. Many of these definitions are not interchangeable. Therefore, it is important to use each specific definition to determine if an obligation to protect exists for each type of personal information. See Appendix C for a listing of common definitions of personal information.

27 The Census Bureau has a special obligation to protect based on provisions of Title 13 of the U.S. Code, and IRS has a special obligation to protect based on Title 26 of the U.S. Code. There are more agency-specific obligations to protect PII, and an organization's legal counsel and privacy officer should be consulted.

28 For additional information, see GLBA, 15 U.S.C. § 6801 et seq.

29 CIPSEA is Title 5 of the E-Government Act of 2002, Pub.L. 107-347, 116 Stat. 2899, 44 U.S.C. § 101 et seq. CIPSEA covers all types of data collected for statistical purposes, not just PII. For additional information, see the OMB Implementation Guidance for CIPSEA, http://www.whitehouse.gov/omb/fedreg/2007/061507_cipsea_guidance.pdf.

accessed PII than would otherwise be assigned to help mitigate the increased risk caused by the nature of the access.

Additionally, organizations may choose to consider whether PII that is stored or regularly transported off-site by employees should be assigned a higher PII confidentiality impact level. For example, surveyors, researchers, and other field employees often need to store PII on laptops or removable media as part of their jobs. PII located offsite is more vulnerable to unauthorized access or disclosure because it is more likely to be lost or stolen than PII stored within the physical boundaries of the organization.

3.3 PII Confidentiality Impact Level Examples

The following are examples of how an organization might assign PII confidentiality impact levels to specific instances of PII. The examples are intended to help organizations better understand the process of considering the various impact level factors, and they are not a substitute for organizations analyzing their own situations. Certain circumstances within any organization or specific system, such as the context of use or obligation to protect, may cause different outcomes.

Obligation to protect is a particularly important factor that should be determined early in the categorization process. Since obligation to protect confidentiality should always be made in consultation with an organization's legal counsel and privacy officer, it is not addressed in the following examples.

3.3.1 Example 1: Incident Response Roster

An organization maintains a roster (in both electronic and paper formats) of its computer incident response team members. In the event that an IT staff member detects any kind of security breach, standard practice requires that the staff member contact the appropriate people listed on the roster. Because this team may need to coordinate closely in the event of an incident, the contact information includes names, professional titles, office and work cell phone numbers, and work email addresses. The organization makes the same types of contact information available to the public for all of its employees on its main Web site.

Distinguishability: The information directly identifies a small number of individuals (fewer than 20).

Aggregation and data field sensitivity: Although the roster is intended to be made available only to the team members, the individuals' information included in the roster is already available to the public on the organization's Web site.

Context of use: The release of the individuals' names and contact information would not likely cause harm to the individuals, and disclosure of the fact that the organization has collected or used this information is also unlikely to cause harm.

Access to and location of the PII: The information is accessed by IT staff members that detect security breaches, as well as the team members themselves. The PII needs to be readily available to teleworkers and to on-call IT staff members so that incident responses can be initiated quickly.

Taking into account these factors, the organization determines that unauthorized access to the roster would likely cause little or no harm, and it chooses to assign the PII confidentiality impact level of *low*.

3.3.2 Example 2: Intranet Activity Tracking

An organization maintains a Web use audit log for an intranet Web site accessed by employees. The Web use audit log contains the following:

- The user's IP address
- The Uniform Resource Locator (URL) of the Web site the user was viewing immediately before coming to this Web site (i.e., referring URL)
- The date and time the user accessed the Web site
- The amount of time the user spent at the Web site
- The web pages or topics accessed within the organization's Web site (e.g., organization security policy).

Distinguishability: By itself, the log does not contain any distinguishable data. However, the organization has another system with a log that contains domain login information records, which include user IDs and corresponding IP addresses. Administrators that can access both systems and their logs and took the time to correlate information between the logs could distinguish individuals. Potentially, information could be gathered on the actions of most of the organization's users involving Web access to intranet resources. The organization has a small number of administrators that have access to both systems and both logs.

Aggregation and data field sensitivity: The information on which internal Web pages and topics were accessed could potentially cause some embarrassment if the pages involved certain human resources-related subjects, such as a user searching for information on substance abuse programs. However, since the logging is limited to use of intranet-housed information, the amount of potentially embarrassing information is minimal.

Context of use: The release of the information would be unlikely to cause harm, other than potentially embarrassing a small number of users if their identities could be distinguished. The fact that the logging is occurring is generally known and assumed and would not cause harm.

Access to and location of the PII: The log is accessed by a small number of system administrators when troubleshooting operational problems and also occasionally by a small number of incident response personnel when investigating internal incidents. All access to the log occurs only from the organization's own systems.

Taking into account these factors, the organization determines that a breach of the log's confidentiality would likely cause little or no harm, and it chooses to assign the PII confidentiality impact level of *low*.

3.3.3 Example 3: Fraud, Waste, and Abuse Reporting Application

A database contains Web form submissions by individuals claiming possible fraud, waste, or abuse of organizational resources and authority. Some of the submissions include serious allegations, such as accusing individuals of accepting bribes or accusing individuals of not enforcing safety regulations. The submission of contact information is not prohibited, and individuals sometimes enter their personal information in the form's narrative text field. The Web site is hosted by a server that logs IP address, referring Web site information, and time spent on the Web site.

Distinguishability: By default, the database does not request distinguishable data, but a significant percentage of users choose to provide distinguishable information. A recent estimate indicated that the database has approximately 30 records with distinguishable information out of nearly 1000 total records. The Web log does not contain any distinguishable information, nor could it be readily linked with the database or other sources to identify specific individuals.

Aggregation and data field sensitivity: The database's narrative text field contains user-supplied text and frequently includes information such as name, mailing address, email address, and phone numbers. The organization does not know how sensitive this information might be to the individuals, such as unlisted phone numbers or email addresses used for limited private communications.

Context of use: Because of the nature of the submissions—reporting claims of fraud, waste, or abuse—the disclosure of individuals' identities would likely cause some of the individuals making the claims to fear retribution by management and peers. The ensuing harm could include blackmail, severe emotional distress, loss of employment, and physical harm. A breach would also undermine trust in the organization by both the individuals making the claims and the public.

Access to and location of the PII: The database is only accessed by a few people who investigate fraud, waste, and abuse claims. All access to the database occurs only from the organization's own systems.

Taking into account these factors, the organization determines that a breach of the database's confidentiality would likely cause catastrophic harm to some of the individuals and chooses to assign the PII confidentiality impact level of *high*.

4. PII Confidentiality Protection Measures

PII should be protected through a combination of measures, including general protection measures, privacy-specific protection measures, and security controls. Organizations should use a risk-based approach for protecting the confidentiality of PII. The PII protection measures provided in this section are complementary to other general protection measures for data and may be used as one part of an organization's comprehensive approach to protecting the confidentiality of PII.

4.1 General Protection Measures

This section describes two types of general PII protection: policy and procedure creation; and education, training, and awareness.

4.1.1 Policy and Procedure Creation

Organizations should develop comprehensive policies and procedures for handling PII at the organization level, the program or component level, and occasionally the system level.[30] Some types of policies include foundational privacy principles, privacy rules of behavior, policies that implement laws and other mandates, and system-level policies. The organizational privacy principles act as the foundation upon which the overall privacy program is built and reflect the organization's privacy objectives. Foundational privacy principles may also be used as a guide against which to develop additional policies and procedures. Privacy rules of behavior policies provide guidance on the proper handing of PII, as well as the consequences for failure to comply with the policy. Some policies provide guidance on implementing laws and OMB guidance in an organization's environment based upon the organization's authorized business purposes and mission. Organizations should consider developing privacy policies and associated procedures for the following topics:

- Development of Privacy Impact Assessments (PIAs) and coordination with System of Records Notices (SORNs)
- Access rules for PII within a system
- PII retention schedules and procedures
- Redress
- Individual consent
- Data sharing agreements
- PII incident response and data breach notification
- Privacy in the System Development Life Cycle Process
- Limitation of collection, disclosure, sharing, and use of PII

[30] There are laws and OMB guidance that provide agency requirements for policy development. For example, OMB Memorandum 05-08 requires that a "senior agency official must…have a central policy-making role in the agency's development and evaluation of legislative, regulatory and other policy proposals which implicate information privacy issues…" Additionally, the Privacy Act requires agencies to "establish rules of conduct for persons involved in the design, development, operation, or maintenance of any system of records, or in maintaining any record, and instruct each such person with respect to such rules and the requirements of…" the Privacy Act "including any other rules and procedures adopted…and the penalties for noncompliance."

- Consequences for failure to follow privacy rules of behavior.

If the organization permits access to or transfer of PII through interconnected systems external to the organization or shares PII through other means, the organization should implement the appropriate documented agreements for roles and responsibilities, restrictions on further sharing of the information, requirements for notification to each party in the case of a breach, minimum security controls, and other relevant factors. Also, Interconnection Security Agreements (ISA) should be used for technical requirements, as necessary.[31] These agreements ensure that the partner organizations abide by rules for handling, disclosing, sharing, transmitting, retaining, and using the organization's PII.

PII maintained by the organization should also be reflected in the organization's incident response policies and procedures. A well-defined incident response capability helps the organization detect incidents rapidly, minimize loss and destruction, identify weaknesses, and restore IT operations rapidly. OMB Memorandum M-07-16 sets out specific requirements for reporting incidents involving the loss or inappropriate disclosure of PII. For additional information, see Section 5.

4.1.2 Education, Training, and Awareness

Education, training, and awareness are distinct activities, each critical to the success of privacy and security programs. Their roles related to protecting PII are briefly described below. Additional information on privacy education, training, and awareness is available in NIST SP 800-50, *Building an Information Technology Security Awareness and Training Program*.

Awareness efforts are designed to change behavior or reinforce desired PII practices. The purpose of awareness is to focus attention on the protection of PII. Awareness relies on using attention-grabbing techniques to reach all different types of staff across an organization. For PII protection, awareness methods include informing staff of new scams that are being used to steal identities, providing updates on privacy items in the news such as government data breaches and their effect on individuals and the organization, providing examples of how staff members have been held accountable for inappropriate actions, and providing examples of recommended privacy practices.

The goal of training is to build knowledge and skills that will enable staff to protect PII. Laws and regulations may specifically require training for staff, managers, and contractors. An organization should have a training plan and implementation approach, and an organization's leadership should communicate the seriousness of protecting PII to its staff. Organizational policy should define roles and responsibilities for training; training prerequisites for receiving access to PII; and training periodicity and refresher training requirements. To reduce the possibility that PII will be accessed, used, or disclosed inappropriately, all individuals that have been granted access to PII should receive appropriate training and, where applicable, specific role-based training. Depending on the roles and functions involving PII, important topics to address may include:

- The definition of PII
- The basic privacy laws, regulations, and policies that apply to a staff member's organization
- Restrictions on data collection, storage, and use of PII
- Roles and responsibilities for using and protecting PII
- Having the organization's legal counsel or privacy officer determine legal obligations to protect PII

[31] See NIST SP 800-47, *Security Guide for Interconnecting Information Technology Systems*, http://csrc.nist.gov/publications/PubsSPs.html

- Appropriate disposal of PII
- Sanctions for misuse of PII
- Recognizing a security or privacy incident involving PII
- Retention schedules for PII
- Roles and responsibilities in responding to PII-related incidents.

Education develops a common body of knowledge that reflects all of the various specialties and aspects of PII protection. It is used to develop privacy professionals who are able to implement privacy programs that enable their organizations to proactively respond to privacy challenges.

4.2 Privacy-Specific Protection Measures[32]

Privacy-specific protection measures are controls for protecting the confidentiality of PII. These controls provide types of protections not usually needed for other types of data. Privacy-specific protection measures provide additional protections that help organizations collect, maintain, use, and disseminate data in ways that protect the confidentiality of the data.

4.2.1 Minimizing Collection and Retention of PII

The practice of minimizing the collection and retention of PII is a basic privacy principle.[33] By limiting PII collections to the least amount necessary to conduct its mission, the organization may limit potential negative consequences in the event of a data breach involving PII. Organizations should consider the total amount of PII collected and maintained, as well as the types and categories of PII collected and maintained. This general concept is often abbreviated as the "minimum necessary" principle. PII collections should only be made where such collections are essential to meet the authorized business purpose and mission of the organization. If the PII serves no current business purpose, then the PII should no longer be collected.

Also, an organization should regularly review[34] its holdings of previously collected PII to determine whether the PII is still relevant and necessary for meeting the organization's business purpose and mission.[35] If the PII is no longer relevant and necessary, then the PII should be properly destroyed. The destruction or disposal of PII must be conducted in accordance with the Federal Records Act and records control schedules approved by the National Archives and Records Administration (NARA).[36] The

---

32 Portions of this section were submitted as contributions to the ISO/IEC 29100 *Framework for Privacy* draft standard.

33 Fair Information Practices are also referred to as privacy principles. See Appendix D for additional information.

34 The frequency of reviews should be done in accordance with laws, regulations, mandates, and organizational policies that apply to the collection of PII.

35 The Privacy Act requires that Federal agencies only maintain records relevant and necessary to their mission. Also, OMB directed Federal agencies to review their PII holdings annually and to reduce their holdings to the minimum necessary for proper performance of their missions. OMB Memorandum M-07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information*, http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf

36 The Federal Records Act, 44 U.S.C. § 3301, defines records as "[a]ll books, papers, maps, photographs, machine-readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government or because of the informational value of the data in them." Agencies are required to create and maintain "adequate and proper documentation" of their organization, mission, functions, etc., and may not dispose of records without the approval of the Archivist of the United States. This approval is granted through the General Records Schedules (GRS) and agency specific records schedules.

effective management and prompt disposal of PII, in accordance with NARA-approved disposition schedules, will minimize the risks of unauthorized disclosure.

4.2.2 De-Identifying Information

Full data records are not always necessary, such as for some forms of research, resource planning, and examinations of correlations and trends. The term *de-identified information* is used to describe records that have had enough PII removed or *obscured*, also referred to as masked or obfuscated, such that the remaining information does not identify an individual and there is no reasonable basis to believe that the information can be used to identify an individual.[37] De-identified information can be re-identified (rendered distinguishable) by using a code, algorithm, or pseudonym that is assigned to individual records. The code, algorithm, or pseudonym should not be derived from other related information about the individual, and the means of re-identification should only be known by authorized parties and not disclosed to anyone without the authority to re-identify records. A common de-identification technique for obscuring PII is to use a one-way cryptographic function, also known as a hash function, on the PII. De-identified information can be assigned a PII confidentiality impact level of *low*, as long as the following are both true:

- The re-identification algorithm, code, or pseudonym is maintained in a separate system, with appropriate controls in place to prevent unauthorized access to the re-identification information.
- The data elements are not linkable, via public records or other reasonably available external records, in order to re-identify the data.

For example, de-identification could be accomplished by removing account numbers, names, SSNs, and any other identifiable information from a set of financial records. By de-identifying the information, a trend analysis team could perform an unbiased review on those records in the system without compromising the PII or providing the team with the ability to identify any individual. Another example is using health care test results in research analysis. All of the distinguishable PII fields can be removed, and the patient ID numbers can be obscured using pseudo-random data that is linked to a cross-reference table located in a separate system. The only means to reconstruct the original (complete) PII records is through authorized access to the cross-reference table.

Additionally, de-identified information can be aggregated for the purposes of statistical analysis, such as making comparisons, analyzing trends, or identifying patterns. An example is the aggregation and use of multiple sets of de-identified data for evaluating several different types of education loan programs. The data describes characteristics of loan holders, such as age, gender, region, and outstanding loan balances. With this dataset, an analyst could draw statistics showing that 18,000 women in the 30-35 age group have outstanding loan balances greater than $10,000. Although the original data sets contained distinguishable identities for each person and is considered to be PII, the de-identified and aggregated dataset would not contain linked or readily distinguishable data for any individual.

[37] For the purpose of analysis, the definition for de-identified information used in this document is loosely based on the Standard for de-identified data defined in the HIPAA Privacy Rule, and it is generalized to apply to all PII. This definition differs from the HIPAA definition in that it is applied to all PII and does not specifically require the removal of all 18 data elements described by the HIPAA Privacy Rule. The HIPAA Privacy Rule recognizes two ways to de-identify data such that it is no longer considered to be protected health information (PHI). First, 18 specific fields can be removed, such as name, SSN, and phone number. Second, the anonymity of the data can be proven statistically. 45 CFR §164.514, http://www.hhs.gov/ocr/hipaa/finalreg.html

4.2.3 Anonymizing Information

Anonymous is defined as something that cannot be "named or identified." It derives from a Greek word meaning "without a name." [38] Similarly, *anonymized information* is defined as previously identifiable information that has been de-identified and for which a code or other link no longer exists.[39] Anonymized information differs from de-identified information because anonymized information cannot be re-identified. A re-identification algorithm, code, or pseudonym does not exist or has been removed and is not available. Anonymizing information usually involves the application of statistical disclosure limitation techniques[40] to ensure the data cannot be re-identified, such as: [41]

- **Generalizing the Data—**Making information less precise, such as grouping continuous values
- **Suppressing the Data—**Deleting an entire record or certain parts of records
- **Introducing Noise into the Data—**Adding small amounts of variation into selected data
- **Swapping the Data—**Exchanging certain data fields of one record with the same data fields of another similar record (e.g., swapping the zip codes of two records)
- **Replacing Data with the Average Value—**Replacing a selected value of data with the average value for the entire group of data.

Using these techniques, the information is no longer PII, but it can retain its useful and realistic properties.[42]

Anonymized information is useful for system testing.[43] Most systems that are newly developed, newly purchased, or upgraded require testing before being introduced to their intended production environment. Testing generally should simulate real conditions as closely as possible to ensure the new or upgraded system runs correctly and handles the projected system capacity effectively. If PII is used in the test environment, it is required to be protected at the same level that it is protected in the production environment, which can add significantly to the time and expense of testing the system.

Randomly generating fake data in place of PII to test systems is often ineffective because certain properties and statistical distributions of the PII may need to be retained to effectively test the system. There are tools available that substitute PII with synthetic data generated by anonymizing PII. The anonymized information retains the useful properties of the original PII, but the anonymized information is not considered to be PII. Anonymized data substitution is a privacy-specific protection measure that

---

38 Merriam Webster Dictionary Online, http://www.merriam-webster.com/dictionary/anonymous.

39 Based on the Common Rule, which governs confidentiality requirements for research, 45 CFR 46.

40 Both anonymizing and de-identifying should be conducted by someone with appropriate training. It may be helpful, as appropriate, to consult with a statistician to assess the level of risk with respect to possible unintended re-identification and improper disclosure. For additional information on statistical disclosure limitation techniques, see OMB's Statistical Policy Working Paper #22, http://www.fcsm.gov/working-papers/spwp22.html. See also Census Bureau, *Report on Confidentiality* and Privacy 1790-2002, http://www.census.gov/prod/2003pubs/conmono2.pdf.

41 The Federal Committee on Statistical Methodology provides a checklist to assist in the assessment of risk for re-identification and improper disclosure. For additional information, see the Federal Committee on Statistical Methodology, Confidentiality and Data Access Committee's Checklist on Disclosure Potential of Data Releases, http://www.fcsm.gov/committees/cdac/.

42 The retention of useful properties in anonymized data is dependent upon the statistical disclosure limitation technique applied.

43 Anonymization is also commonly used by agencies to release data sets to the public for research purposes.

enables system testing while reducing the expense and added time of protecting PII. However, not all data can be readily anonymized (e.g. biometric data).

4.3 Security Controls

In addition to the PII-specific protection measures described earlier in this section, many types of technical and operational security controls are available to safeguard the confidentiality of PII. These controls are often already available on a system to protect other types of data processed, stored, or transmitted by the system. The security controls listed in NIST SP 800-53 address general protections of data and systems. The items listed below are some of the NIST SP 800-53 controls that can be used to help safeguard the confidentiality of PII. Note that some of these controls may not be in the recommended set of security controls for the baselines identified in NIST SP 800-53 (e.g., a control might only be recommended for high-impact systems). However, organizations may choose to provide greater protections than what is recommended; see Section 3.1 for a discussion of characteristics to consider when choosing the appropriate controls. In addition to the controls listed below, NIST SP 800-53 contains many other controls that can be used to help protect PII, such as incident response controls.

- **Access Enforcement (AC-3).** Organizations can control access to PII through access control policies and access enforcement mechanisms (e.g., access control lists). This can be done in many ways. One example is implementing role-based access control and configuring it so that each user can access only the pieces of data necessary for the user's role. Another example is only permitting users to access PII through an application that tightly restricts their access to the PII, instead of permitting users to directly access the databases or files containing PII.[44] Encrypting stored information is also an option for implementing access enforcement.[45] OMB M-07-16 specifies that Federal agencies must "encrypt, using only NIST certified cryptographic modules, all data on mobile computers/devices carrying agency data unless the data is determined not to be sensitive, in writing, by your Deputy Secretary or a senior-level individual he/she may designate in writing".

- **Separation of Duties (AC-5).** Organizations can enforce separation of duties for duties involving access to PII. For example, the users of de-identified PII data would not also be in roles that permit them to access the information needed to re-identify the records.

- **Least Privilege (AC-6).** Organizations can enforce the most restrictive set of rights/privileges or accesses needed by users (or processes acting on behalf of users) for the performance of specified tasks. Concerning PII, the organization can ensure that users who must access records containing PII only have access to the minimum amount of PII data, along with only those privileges (e.g., read, write, execute) that are necessary to perform their job duties.

- **Remote Access (AC-17).** Organizations can choose to prohibit or strictly limit remote access to PII. If remote access is permitted, the organization can ensure that the communications are encrypted.

- **Access Control for Mobile Devices (AC-19).** Organizations can choose to prohibit or strictly limit access to PII from portable and mobile devices, such as laptops, cell phones, and personal digital assistants (PDA), which are generally higher-risk than non-portable devices (e.g., desktop computers

---

[44] For example, suppose that an organization has a database containing thousands of records on employees' benefits. Instead of allowing a user to have full and direct access to the database, which could allow the user to save extracts of the database records to the user's computer, removable media, or other locations, the organization could permit the user to access only the necessary records and record fields. A user could be restricted to accessing only general demographic information and not any information related to the employees' identities. More information on restricting extracts from PII databases is available in Appendix E.

[45] Additional encryption guidelines and references can be found in FIPS 140-2: *Security Requirements for Cryptographic Modules*, http://csrc.nist.gov/publications/PubsFIPS.html.

at the organization's facilities). Some organizations choose to forbid all telework and remote access involving higher-impact instances of PII so that the information will not leave the organization's physical boundaries. If access is permitted, the organization can ensure that the devices are properly secured and regularly scan the devices to verify their security status (e.g., antivirus software enabled and up-to-date, operating system fully patched).

- **Auditable Events (AU-2).** Organizations can monitor events that affect the confidentiality of PII, such as unauthorized access to PII.

- **Audit Monitoring, Analysis, and Reporting (AU-6)**. Organizations can regularly review and analyze information system audit records for indications of inappropriate or unusual activity affecting PII, investigate suspicious activity or suspected violations, report findings to appropriate officials, and take necessary actions.

- **User Identification and Authentication (IA-2).** Users can be uniquely identified and authenticated before accessing PII.[46] The strength requirement for the authentication mechanism depends on the impact level of the PII and the system as a whole. OMB M-07-16 specifies that Federal agencies must "allow remote access only with two-factor authentication where one of the factors is provided by a device separate from the computer gaining access," and also must "use a 'time-out' function for remote access and mobile devices requiring user re-authentication after thirty minutes of inactivity."

- **Media Access (MP-2).** Organizations can restrict access to information system media containing PII, including digital media (e.g., CDs, USB flash drives, backup tapes) and non-digital media (e.g., paper, microfilm). This could also include portable and mobile devices with a storage capability.

- **Media Marking (MP-3).** Organizations can label information system media and output containing PII to indicate how it should be distributed and handled. The organization could exempt specific types of media or output from labeling so long as it remains within a secure environment. Examples of labeling are cover sheets on printouts and paper labels on digital media.

- **Media Storage (MP-4).** Organizations can securely store PII, both in paper and digital forms, until the media are destroyed or sanitized using approved equipment, techniques, and procedures. One example is the use of storage encryption technologies to protect PII stored on removable media.

- **Media Transport (MP-5).** Organizations can protect digital and non-digital media and mobile devices containing PII that is transported outside the organization's controlled areas. Examples of protective measures are encrypting stored information and locking the media in a container.

- **Media Sanitization (MP-6).** Organizations can sanitize digital and non-digital media containing PII before it is disposed or released for reuse.[47] An example is degaussing a hard drive—applying a magnetic field to the drive to render it unusable.

- **Transmission Confidentiality (SC-9).** Organizations can protect the confidentiality of transmitted PII. This is most often accomplished by encrypting the communications or by encrypting the information before it is transmitted.[48]

---

[46] More information on authentication is available from NIST SP 800-63, *Electronic Authentication Guideline.*

[47] For more information on media sanitization, see NIST SP 800-88, *Guidelines for Media Sanitization.*

[48] NIST has several publications on this topic that are available from http://csrc.nist.gov/publications/PubsSPs.html.

5. Incident Response for Breaches of PII

Handling breaches involving PII is different from regular incident handling and may require additional actions by an organization. Breaches involving PII can receive considerable media attention, which can greatly harm an organization's reputation and reduce the public's trust[49] in the organization. Moreover, affected individuals can be subject to embarrassment, identity theft, or blackmail as the result of a breach of PII. Due to these particular risks of harm, organizations should develop additional policies, such as determining when and how individuals should be notified, when and if a breach should be reported publicly, and whether to provide remedial services, such as credit monitoring, to affected individuals. Organizations should integrate these additional policies into their existing incident handling response policies.

FISMA requires Federal agencies to have procedures for handling information security incidents, and it established a Federal information security incident center to coordinate and share information about incidents, which resulted in the creation of U.S. Computer Emergency Readiness Team (US-CERT). Additionally, NIST provided guidance on security incident handling in NIST SP 800-61 Revision 1, *Computer Security Incident Handling Guide*. In 2007, OMB issued M-07-16, which provided specific guidance to Federal agencies for handling incidents involving PII.

Incident response plans should be modified to handle breaches involving PII. Incident response plans should also address how to minimize the amount of PII necessary to adequately report and respond to a breach. NIST SP 800-61 Revision 1 describes four phases of handling security incidents. Specific policies and procedures for handling breaches involving PII can be added to each of the following phases identified in NIST SP 800-61: preparation; detection and analysis; containment, eradication, and recovery; and post-incident activity. This section provides additional details on PII-specific considerations for each of these four phases.

5.1 Preparation

Preparation requires the most effort because it sets the stage to ensure the PII breach is handled appropriately. Organizations should build their PII breach response plans into their existing incident response plans. The development of PII breach response plans requires organizations to make many decisions about how to handle PII breaches, and the decisions should be used to develop policies and procedures. The policies and procedures should be communicated to the organization's entire staff through training and awareness programs. Training programs should inform employees of the consequences of their actions for inappropriate use and handling of PII.

The organization should determine if existing processes are adequate, and if not establish a new incident reporting method for employees to report suspected or known breaches of PII. The method could be a telephone hotline, email, or a management reporting structure in which employees know to contact a specific person within the management chain. Employees should be able to report any PII breach immediately at any day or time. Additionally, employees should be provided with a clear definition of what constitutes a PII breach and what information needs to be reported. The following information is helpful to obtain from employees who are reporting a known or suspected PII breach:[50]

- Person reporting the incident

[49] According to a 2007 Government Privacy Trust Survey conducted by the Ponemon Institute, a Federal department fell from being a top five most trusted agency in 2006 to just above the bottom five least trusted agencies after the highly publicized breach of millions of PII records in 2006. http://www.govexec.com/dailyfed/0207/022007tdpm1.htm

[50] U.S. Department of Commerce, *Breach Notification Response Plan*, September 28, 2007

- Person who discovered the incident
- Date and time the incident was discovered
- Nature of the incident
- Description of the information lost or compromised
- Storage medium from which information was lost or compromised
- Controls in place to prevent unauthorized use of the lost or compromised information
- Number of individuals potentially affected
- Whether law enforcement was contacted.

Federal agencies are required to report all known or suspected breaches of PII, in any format, to US-CERT within one hour.[51] To meet this obligation, organizations should proactively plan their breach notification response. A PII breach may require notification to persons external to the organization, such as law enforcement, financial institutions, affected individuals, the media, and the public.[52] Organizations should plan in advance how, when, and to whom notifications should be made. Organizations should conduct training sessions on interacting with the media regarding incidents. Additionally, OMB M-07-16 requires federal agencies to include the following elements in their plans for handling breach notification:

- Whether breach notification is required[53]
- Timeliness of the notification
- Source of the notification
- Contents of the notification
- Means of providing the notification
- Who receives the notification; public outreach response.

Additionally, organizations should establish a committee or person responsible for using the breach notification policy to coordinate the organization's response.

The organization should also determine what circumstances require the organization to provide remedial assistance to affected individuals, such as credit monitoring services. The PII confidentiality impact level should be considered for this determination because it provides an analysis of the likelihood of harm for the loss of confidentiality for each instance of PII.

5.2 Detection and Analysis

Organizations may continue to use their current detection and analysis technologies and techniques for handling incidents involving PII. However, adjustments to incident handling processes may be needed, such as ensuring that the analysis process includes an evaluation of whether an incident involves PII.

---

[51] In M-07-16, OMB required Federal agencies to report all known or suspected PII breaches to US-CERT within one hour.

[52] For additional information about communications with external parties, such as the media, see NIST SP 800-61 Revision 1, *Computer Security Incident Handling Guide*, http://csrc.nist.gov/publications/PubsSPs.html.

[53] For Federal agencies, notification to US-CERT is always required.

Detection and analysis should focus on both known and suspected breaches of PII. In the event that a suspected breach of PII is detected, Federal agencies should notify US-CERT within one hour.

5.3 Containment, Eradication, and Recovery

Existing technologies and techniques for containment, eradication, and recovery may be used for breaches involving PII. However, changes to incident handling processes may be needed, such as performing additional media sanitization steps when PII needs to be deleted from media during recovery. Particular attention should be paid to using proper forensics techniques[54] to ensure preservation of evidence for intentional criminal acts. Additionally, it is important to determine whether PII was accessed and how many records or individuals were affected.

5.4 Post-Incident Activity

As with other security incidents, information learned through detection, analysis, containment, and recovery should be collected for sharing within the organization and with the US-CERT to help protect against future incidents. The PII breach response plan should be continually updated and improved based on the lessons learned during each incident.

Additionally, the organization should use its PII breach response policy to determine whether the organization should provide affected individuals with remedial assistance, such as credit monitoring. When providing notice to individuals, organizations should make affected individuals aware of their options, such as obtaining a free copy of their credit report, obtaining a freeze credit report, placing a fraud alert on their credit report, or contacting their financial institutions.

[54] For additional information, see NIST SP 800-86, *Guide to Integrating Forensic Techniques into Incident Response*, http://csrc.nist.gov/publications/nistpubs/800-86/SP800-86.pdf.

Appendix A—Scenarios for PII Identification and Handling

Exercises involving PII scenarios within an organization provide an inexpensive and effective way to build skills necessary to identify potential issues with how the organization identifies and safeguards PII. Individuals who participate in these exercises are presented with a brief PII scenario and a list of general and specific questions related to the scenario. After reading the scenario, the group then discusses each question and determines the most appropriate response for their organization. The goal is to determine what the participants would really do and to compare that with policies, procedures, and generally recommended practices to identify any discrepancies or deficiencies and decide upon appropriate mitigation techniques.

The general questions listed below are applicable to almost any PII scenario. After the general questions are scenarios, each of which is followed by additional scenario-specific questions. Organizations are encouraged to adapt these questions and scenarios for use in their own PII exercises. Also, additional scenarios and questions specific to PII incident handling are available from NIST SP 800-61 Revision 1, *Computer Security Incident Handling Guide*.[55]

A.1 General Questions

1. What measures are in place to identify, assess, and protect the PII described in the scenario?

2. Which individuals have designated responsibilities within the organization to safeguard the PII described in the scenario?

3. To which people and groups within the organization should questions about PII or the possible misuse of PII be reported?

4. What could happen if the PII described in the scenario is not safeguarded properly?

A.2 Scenarios

Scenario 1: A System Upgrade

An organization is redesigning and upgrading its physical access control systems, which consist of entry-way consoles that recognize ID badges, along with identity management systems and other components. As part of the redesign, several individual physical access control systems are being consolidated into a single system that catalogues and recognizes biometric template data (a facial image and fingerprint), employee name, employee identification number (an internal identification number used by the organization) and employee SSN. The new system will also contain scanned copies of "identity" documentation, including birth certificates, driver's licenses, and/or passports. In addition, the system will maintain a log of all access (authorized or unauthorized) attempts by a badge. The log contains employee identification numbers and timestamps for each access attempt.

1. What information in the system is PII?

2. What is the PII confidentiality impact level? What factors were taken into consideration when making this determination?

[55] SP 800-61 Revision 1 is available at http://csrc.nist.gov/publications/PubsSPs.html.

3. By consolidating data into a single system, does it create additional vulnerabilities that could result in harm to the individual? What additional controls could be put in place to mitigate the risk?

4. Is all of the information necessary for the system to function? Is there a way to minimize the information in the system? Could PII on the system be replaced with operating data that is not PII?

5. Is the organization required to conduct a PIA for this system?

Scenario 2: Protecting Survey Data

Recently, an organization emailed to individuals a link to an online survey, which was designed to gather feedback about the organization's services. The organization identified each individual by name, email address, and an organization-assigned ID number. The majority of survey questions asked individuals to express their satisfaction or dissatisfaction with the organization, but there were also questions asking individuals to provide their zip code along with demographic details on their age, income level, educational background, and marital status.

The following are additional questions for this scenario:

1. Which data elements collected through this survey should be considered PII?

2. What is the PII confidentiality impact level? What factors were taken into consideration when making this determination?

3. How are determinations made as to which data from the survey are relevant to the organization's operations? What happens to data that is deemed unnecessary?

4. What privacy-specific protection measures might help safeguard the PII collected and retained from this survey?

5. What other types of protection measures for safeguarding data (that are not necessarily specific to safeguarding PII) might be used to protect the data from the responses?

Scenario 3: Completing Work at Home

An organization's employee needed to leave early for a doctor's appointment, but the employee was not finished with her work for the day and had no leave time available. Since she had the same spreadsheet application at home, she decided to email a data extract as an attachment to her personal email address and finish her work at home that evening. The data extract was downloaded from an access controlled human resources database located on a server within the organization's security perimeter. The extract contained employee names, identification numbers, dates of birth, salary information, manager's name, addresses, phone numbers, and positions. As she was leaving, she remembered that she had her personal USB flash drive in her purse. She decided the USB drive would be good to use in case she had an attachment problem with the email she had already sent. Although much of the USB drive's space was taken up with family photos she had shared with her coworkers earlier in the day, there was still enough room to add the data extract. She copied the data extract and dropped it in her purse as she left for her appointment. When she arrived home that evening, she plugged the USB drive into her family's computer and used her spreadsheet application to analyze the data.

The following are additional questions for this scenario:

1. Which data elements contained in this data extract should be considered PII?

2. What is the PII confidentiality impact level? What factors were taken into consideration when making this determination?

3. What privacy-specific protection measures might help safeguard the PII contained in the data extract?

4. What should the employee do if her purse (containing the USB drive) is stolen? What should the organization do? How could the employer have prevented this situation?

5. What should the employee do with the copies of the extract when she finishes her work?

6. Should the emailing of the extract to a personal email address be considered a breach? Should storing the data on the personal USB drive be considered a breach?

7. What could the organization do to reduce the likelihood of similar events in the future?

8. How should this scenario be handled if the information is a list of de-identified retirement income statistics? Would the previous questions be answered differently?

Scenario 4: Testing Systems

An organization needed to test an upgrade to its fingerprint matching system before the upgrade could be introduced into the production environment. Because it is difficult to simulate fingerprint image and template data, the organization used real biometric image and template data to test the system. In addition to the fingerprint images and templates, the system also processed the demographic data associated with each fingerprint image, including name, age, sex, race, date of birth, and nationality. After successful completion of the testing, the organization upgraded its production system.

1. Which data elements contained in this system test should be considered PII?

2. What is the PII confidentiality impact level? What factors were taken into consideration when making this determination?

3. What privacy-specific protection measures might help safeguard the PII used in this test?

4. Is a PIA required to conduct this testing? Is a PIA required to complete the production system upgrade?

5. What should the organization do with the data used for testing when it completes the upgrade?

Appendix B—Frequently Asked Questions (FAQ)

Privacy and security leadership and staff, as well as others within organizations, may have questions about identifying, handling, and protecting the confidentiality of personally identifiable information (PII). This appendix contains frequently asked questions (FAQ) related to PII. Organizations are encouraged to customize this FAQ and make it available to their user community.

1. What is personally identifiable information (PII)?

PII is defined in OMB Memorandum M-07-16 as "information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc."

2. How does this apply to foreign nationals?

OMB defined the term *individual,* as used in the definition of PII, to mean a citizen of the United States or an alien lawfully admitted for permanent residence, which is based on the Privacy Act definition.[56] For the purpose of protecting the confidentiality of PII, organizations may choose to administratively expand the scope of application to foreign nationals without creating new legal rights. Expanding the scope may reduce administrative burdens and improve operational efficiencies in the protection of data by eliminating the need to maintain separate systems or otherwise separate data. Additionally, the status of citizen, alien, or legal permanent resident can change over time, which makes it difficult to accurately identify and separate the data of foreign nationals. Expanding the scope may also serve additional organizational interests, such as providing reciprocity for data sharing agreements with other organizations.

Agencies may also, consistent with individual practice, choose to extend the protections of the Privacy Act to foreign nationals without creating new judicially enforceable legal rights. For example, DHS has chosen to extend Privacy Act protections (e.g., access, correction), to foreign nationals whose data resides in mixed systems, which are systems of records with information about both U.S. persons and non-U.S. persons.[57]

Organizations should also consult with legal counsel to determine if they have an additional obligation to protect the confidentiality of the personal information relating to foreign nationals, such as the Immigration and Nationality Act, which requires the protection of the confidentiality of Visa applicant data.[58]

3. What does it mean to "distinguish" an individual?

To *distinguish* an individual is to identify an individual. Some examples of information that could distinguish an individual include, but are not limited to, names, passport number, social security number, or biometric image and template.

---

[56] OMB M-03-22, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002,* http://www.whitehouse.gov/omb/memoranda/m03-22.html#1.

[57] *See DHS Privacy Policy Regarding Collection, Use Retention, and Dissemination of Information on Non-U.S. Persons,* http://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2007-1.pdf.

[58] Immigration and Nationality Act, 8 U.S.C. 1202.

4. What does it mean when a record is "linked"?

A record is *linked* to an individual when it contains information that cannot distinguish an individual when considered separately, but which could distinguish an individual when combined with other data elements present on the same system or a closely-related system. For example, an individual could be identified only by ID #12345 in one database, and another database on the same system could map that ID # to the individual's name and social security number. The records in the first database would be considered "linked" if users were likely to have access to both databases, or could obtain access with minimal effort.

5. What does it mean when a record is "linkable"?

A record is *linkable* to an individual when it contains information that cannot distinguish an individual, but that may be matched or compared with other data elements from a source available to the general public or that is otherwise obtainable. For example, individuals might be identified in a database by home telephone number. The identities of some of the individuals could be determined by comparing this information to publicly available telephone directories.

6. Is personally identifiable information (PII) the same as information in identifiable form (IIF)?

No, the terms PII and IIF are not the same. Their definitions are distinct, cannot be used interchangeably, and have different requirements associated with them.

OMB defines *PII* in OMB Memorandum 07-16 as "information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc."

OMB defines *IIF* in OMB Memorandum 03-22 as "information in an IT system or online collection: (i) that directly identifies an individual (e.g., name, address, social security number or other identifying number or code, telephone number, email address, etc.) or (ii) by which an agency intends to identify specific individuals in conjunction with other data elements, i.e., indirect identification. (These data elements may include a combination of gender, race, birth date, geographic indicator, and other descriptors)."[59]

The following two distinctions exist between these terms:

- *Indirect identification*. The terms are distinct in how they address the issue of indirect identification. If information does not—by itself—identify an individual, it must be "intended by the agency" to identify an individual in conjunction with other data to meet the definition of IIF. The definition of PII does not address the agency's intent and states that information can be PII even if the data is merely "linkable" to the individual, whether the agency intends to actually link it or not.

- *Scope of policies*. IIF is defined only in terms of identifying the IT systems for which a federal agency must complete a privacy impact assessment (PIA). PII is defined broadly to include personal information stored in any format, both electronic and paper-based.

[59] OMB M-03-22, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*

7. How did the need for guidelines on protecting PII come about? Why is this important?

The protection of PII is important to maintain public trust and confidence in an organization, to protect the reputation of an organization, and to protect against legal liability for an organization. Organizations have always considered trust, confidence, and reputation as motivating factors in protecting PII. Recently, organizations have become more concerned about the risk of legal liability due to the enactment of many federal, state, and international privacy laws.

In the United States, Federal privacy laws are generally sector-based. For example, the Health Insurance Portability and Accountability Act of 1996 applies to the health care sector, and the Gramm-Leach-Bliley Act of 1999 applies to the financial services sector. In contrast, many states have enacted their own generally applicable privacy laws, such as breach notification laws. Some U.S.-based organizations that conduct business abroad must also comply with international privacy laws, which vary greatly from country to country. Organizations are responsible for determining which laws apply to them based on sector and jurisdiction.

For Federal government agencies, the need to protect PII was first established by the Privacy Act of 1974. It required Federal agencies to protect PII and apply the Fair Information Practices to PII. Also, the Privacy Act required agencies to "establish appropriate administrative, technical and physical safeguards to ensure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained."

In response to the increased use of computers and the Internet to process government information, the E-Government Act of 2002 was enacted to ensure public trust in electronic government services. It required Federal agencies to conduct Privacy Impact Assessments (PIAs) and to maintain privacy policies on their web sites. The E-Government Act also directed the OMB to issue implementation guidance to Federal agencies. In 2003, OMB issued M-03-22 to provide guidance on PIAs and web site privacy policies. OMB has continued to provide privacy guidance to Federal agencies on many PII protection topics such as remote access to PII, encryption of PII on mobile devices, and breach notification (see Appendix H for additional information).

Additionally, Federal agencies are required to comply with other privacy laws, such as the Children's Online Privacy Protection Act (COPPA) and HIPAA (only if the agency acts as a health care provider or other covered entity as defined by the statute).

8. What is the Privacy Act?

The Privacy Act of 1974 is the foundation of public sector privacy law in the U.S. It applies only to Federal agencies and provides a statutory basis for the required use of Fair Information Practices. The Privacy Act pertains only to data maintained within a System of Records (SOR), which means any "group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."[60] Record is defined broadly to include any item of information about an individual, both paper and electronic.

The basic provisions of the Privacy Act include the following:

[60] 5 U.S.C. § 552a (a)(5).

- Provide notice to individuals that explains:[61]
  - The authority for the data collection
  - The purpose of the data collection
  - Routine uses for the data
  - Effects, if any, of not providing the information
- Limit collection of data to the minimum necessary to accomplish the purpose of the agency
- Collect information directly from the person about whom the information pertains, if possible
- Maintain accuracy and completeness of the data
- Disclose the data to only those who need access for proper purposes, such as sharing for an identified routine use or to perform agency work
- Allow individuals to access data pertaining to them, request correction of wrong or incomplete data, and make an appeal for denials of requests for access and correction
- Maintain appropriate administrative, technical, and physical safeguards to ensure the security and confidentiality of the records.

Violations of the Privacy Act can result in civil and criminal liability.

Information contained within a Privacy Act System of Records usually will be considered PII. Organizations that seek to protect systems (e.g., via security controls) containing PII may be able to realize efficiencies by coordinating with efforts to comply with the Privacy Act, as these activities will often be similar.

9. What is a Privacy Impact Assessment (PIA)? When do I need to conduct a PIA?

The E-Government Act of 2002 required Federal agencies to conduct PIAs, which are processes for identifying and mitigating privacy risks within an information system. If used effectively, a PIA should address risk at every stage of the system development life cycle (SDLC). Most organizations have established their own templates that provide the basis for conducting a PIA. The E-Government Act of 2002 requires Federal agencies to conduct PIAs when:

- Developing or procuring information technology that collects, maintains, or disseminates information that is in an identifiable form; or
- Initiating a new collection of information that—
  - Will be collected, maintained, or disseminated using information technology; and
  - Includes any information in an identifiable form permitting the physical or online contacting of a specific individual, if identical questions have been posed to, or identical reporting requirements imposed on, 10 or more persons, other than agencies, instrumentalities, or employees of the Federal Government.

[61] The Privacy Act also requires publication of general notice in the Federal Register, which is called a System of Records Notice (SORN).

The E-Government Act authorized OMB to provide Federal agencies with guidance on conducting PIAs, which resulted in OMB Memorandum 03-22. The Memorandum provided examples of system changes that create new privacy risks and trigger the requirement for a new PIA:

- **Conversions**—when paper-based records are to be converted to electronic systems
- **De-Identified to Identifiable**—when functions applied to an existing information collection change de-identified information into information in identifiable form
- **Significant System Management Changes**—when new uses of an existing IT system, including application of new technologies, significantly change how information in identifiable form is managed in the system
- **Significant Merging**—when agencies adopt or alter business processes so that government databases holding information in identifiable form are merged, centralized, matched with other databases, or otherwise significantly manipulated
- **New Public Access**—when user-authenticating technology (e.g., password, digital certificate, biometric) is newly applied to an electronic information system accessed by members of the public
- **Commercial Sources**—when agencies systematically incorporate into existing information systems databases of information in identifiable form purchased or obtained from commercial or public sources
- **New Interagency Uses**—when agencies work together on shared functions involving significant new uses or exchanges of information in identifiable form, such as the cross-cutting E-Government initiatives
- **Internal Flow or Collection**—when alteration of a business process results in significant new uses or disclosures of information or incorporation into the system of additional items of information in identifiable form
- **Alteration in Character of Data**—when new information in identifiable form added to a collection raises the risks to personal privacy (for example, the addition of health or financial information)

The E-Government Act requires publication of PIAs,[62] which must analyze and describe the following information:

- What information is to be collected
- Why the information is being collected
- The intended use of the information
- With whom the information will be shared
- What opportunities individuals have to decline to provide information (i.e., where providing information is voluntary) or to consent to particular uses of the information (other than required or authorized uses), and how individuals can grant consent
- How the information will be secured

[62] An agency may exempt itself from this requirement if publication of the PIA would raise national security concerns or reveal classified or sensitive information.

- Whether a system of records is being created under the Privacy Act, 5 U.S.C. 552a
- What choices the agency made regarding an IT system or collection of information as a result of performing the PIA.

10. What is the Paperwork Reduction Act?

The Paperwork Reduction Act (PRA)[63] was passed in 1995, and created a process for the review and approval of Federal government information collections from the public. The purpose of the Act is to minimize the burden of paperwork on the public, minimize the cost of information collections, and increase the quality of Federal information.[64] The PRA requires Federal agencies to get clearance from OMB when an agency plans to collect information from ten or more persons using identical reporting, recordkeeping, or disclosure requirements. The term *persons* is defined broadly to include people, organizations, local government, etc., but it does not include Federal agencies or employees of Federal agencies. Agencies must also provide notice of the collection in the Federal Register before submitting the information collection to OMB for clearance. OMB reviews the proposed information collection and assigns a control number to the collection, which must be displayed on the collection form. A PIA is required for any electronic collection of information that includes PII and requires OMB clearance pursuant to the PRA.

11. What are the general risks to individuals and the organization if PII is misused?

Depending on the type of information lost, an individual may suffer social, economic, or physical harm. If the information lost is sufficient to be exploited by an identity thief, the person can suffer, for example, from a loss of money, damage to credit, a compromise of medical records, threats, and/or harassment. The individual may suffer tremendous losses of time and money to address the damage. Other types of harm that may occur to individuals include denial of government benefits, blackmail, discrimination, and physical harm.

Organizations also face risks to their finances and reputation. If PII is misused, organizations may suffer financial losses in compensating the individuals, assisting them in monitoring their credit ratings, and addressing administrative concerns. In addition, recovering from a major breach is costly to many organizations in terms of time spent by key staff in coordinating and executing appropriate responses. If a loss of PII constitutes a violation of relevant law, the organization and/or its staff may be subject to criminal or civil penalties, or it may have to agree to receive close government scrutiny and oversight. Another major risk to organizations is that their public reputation and public confidence may be lost, potentially jeopardizing the organizations' ability to achieve their missions.

12. What should I consider when reviewing restrictions on collecting PII?

Key considerations to review are any legal requirements that could impact PII collections. One should ask what laws, regulations, and guidance are applicable to the organization considering the type of PII that is collected (e.g., Privacy Act, Paperwork Reduction Act, and the E-Government Act for general PII; HIPAA for health PII; Gramm-Leach-Bliley Act (GLBA) for financial PII; COPPA for children's PII). An organization's legal counsel and privacy officer should always be consulted to determine whether there are restrictions on collecting PII.

---

63 PRA, 44 U.S.C. § 3501 et seq.

64 For additional information, see: http://ocio.os.doc.gov/ITPolicyandPrograms/Information_Collection/dev01_003742.

One could more specifically ask if the collected PII is absolutely necessary to do business (i.e., does it support the business purpose of the system or the organization's mission?) If it does not serve a viable business purpose, then federal agencies may not collect that PII. If the collection of PII does serve a business purpose, then it should be collected, used, shared, and disseminated appropriately.

13. What are examples of PII?

The following examples are meant to offer a cross-section of the types of information that could be considered PII, either singly or collectively, and is not an exhaustive list of all possibilities. Examples of PII records include financial transactions, medical history, criminal history, and employment history. Examples of individual data elements of PII include an individual's name, social security number, passport number, driver's license number, credit card number, vehicle registration or ID number, x-ray, patient ID number, and biometric image and template data (e.g., retina scans, voice signature, facial geometry).[65]

14. What is different about protecting PII compared to any other data and how should PII be protected?

In many cases, protection of PII is similar to protection of other data and includes protecting the confidentiality, integrity, and availability of the information. Most security controls used for other types of data are also applicable to the protection of PII. Also, there are several privacy-specific protection measures, such as anonymization, minimization of PII collection, and de-identification.

In addition to protection requirements for PII, there are other requirements for the handling of PII. The Fair Information Practices provide an overview of these requirements, which include, but are not limited to, notice, consent, access, correction, integrity, and enforcement. Moreover, the factors for assigning a confidentiality impact level to PII are different than other types of data. Breaches to the confidentiality of PII harm both the organization and the individual. Harm to individuals should be factored in strongly because of the magnitude of the potential harm, such as identity theft, embarrassment, and denial of benefits.

[65] Organizations may want to consider how PII relating to deceased individuals should be handled, such as continuing to protect its confidentiality or properly destroying the information. Organizations may want to base their considerations on any obligations to protect, organizational policies, or by evaluation of organization-specific risk factors. With respect to organization-specific risk factors, there is a balancing act because PII relating to deceased individuals can both promote and prevent identity theft. For example, making available lists of deceased individuals can prevent some types of fraud, such as voter fraud. In contrast, PII of a deceased individual also could be used to open a credit card account or to set up a false cover for criminals. Organizations should consult with their legal counsel and privacy officer.

Appendix C—Definitions of Private Information

Various Federal laws, regulations, and guidance documents describe data elements or records about individuals; other terms they define include "information in identifiable form (IIF), "private information," "systems of records," "protected health information (PHI)," and "directory information." Some of these are similar to the definition used in this document for "PII." However, the definition of PII provided in this section should not be confused with any of these other terms, and readers should not assume that the definition used for PII may be used interchangeably with any of these other terms. The table below provides definitions for some of these terms. The table is not intended to be comprehensive, and considers only a few of the Federal authorities with broad effects and applicability.

Term	Defining Authority	Definition	Comments
Information in Identifiable Form (IIF)	E-Government Act of 2002, 44 USC § 208(d).	Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means." Information "permitting the physical or online contacting of a specific individual" (see section 208(b)(1)(A)(ii)(II)) is the same as "information in identifiable form."	See Appendix B for further information about the differences between PII and IIF.
Information in Identifiable Form (IIF)	OMB Memorandum 03-22, § II.A.2.	Information in an IT system or online collection: (i) that directly identifies an individual (e.g., name, address, social security number or other identifying number or code, telephone number, email address, etc.) or (ii) by which an agency intends to identify specific individuals in conjunction with other data elements, i.e., indirect identification. (These data elements may include a combination of gender, race, birth date, geographic indicator, and other descriptors).	OMB 03-22 limits the definition of "individual" to "a citizen of the United States or an alien lawfully admitted for permanent residence," mirroring the Privacy Act definition.
Individually Identifiable Health Information (IIHI)	Health Insurance Portability and Accountability Act of 1996 (HIPAA), The Privacy Rule, 45 CFR § 160.103.	Under the HIPAA Privacy Rule, IIHI is information that is a subset of health information, including demographic information: - Collected from an individual - Created or received by a health care provider, health plan, employer, or health care clearinghouse; - Relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual; and - That identifies the individual; or with respect to which there is a reasonable basis to believe the information can be used to identify the individual.	Describes a term applicable only to the HIPAA Privacy and Security Rules; subject to a number of exemptions not made for PII

Term	Defining Authority	Definition	Comments
Protected Health Information (PHI)	Health Insurance Portability and Accountability Act of 1996 (HIPAA), The Security Rule, 45 CFR § 160.103.	Under the HIPAA Security Rule, PHI is individually identifiable health information (IIHI) that is: - Transmitted by electronic media; - Maintained in electronic media; or - Transmitted or maintained in any other form or medium. Protected health information excludes individually identifiable health information in: - Education records covered by the Family Educational Rights and Privacy Act, as amended, 20 U.S.C. 1232g; - Employment records held by a covered entity in its role as employer.	Describes a term applicable only to the HIPAA Privacy and Security Rules; subject to a number of exemptions not made for PII
Systems of Records	Privacy Act of 1974, 5 U.S.C. § 552a(a)(5).	A group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual. A record is defined as any item, collection, or grouping of information about an *individual.*	Applies only to Federal agencies. Provides some exemptions for certain types of records. Defines *individual* as limited to citizens of the United States or aliens lawfully admitted for permanent residence.
Education Records	Federal Education Rights Privacy Act, 20 USC § 1232g (a)(4)(A).	Records, files, documents, and other materials which: - contain information directly related to a student; and - are maintained by an educational agency or institution or by a person acting for such agency or institution, subject to some exceptions.	Applies only to educational institutions receiving funds from the Federal government. Exceptions exist for some records maintained for purposes of law enforcement, health, administration, student employment, and others.
Financial Records Non-public personal Information (NPPI)	Gramm-Leach-Bliley Act (GLBA), 15 USC § 6801-6810.	Information collected about consumers: - When consumer is obtaining credit - When entity is performing services in relation to financial product for consumer	Applies only to Financial Institutions, defined as “an entity that regularly provides financial products or financial services to consumers.”

Appendix D—Fair Information Practices

The Fair Information Practices, also known as Privacy Principles, are the framework for most modern privacy laws around the world. Several versions of the Fair Information Practices have been developed through government studies and international organizations. These different versions share common elements, but the elements are divided and expressed differently. The most commonly used versions are discussed in this appendix.[66]

In 1973, the U.S. Department of Health, Education, and Welfare (HEW) (now the Department of Health and Human Services) issued a report entitled *Records, Computers, and the Rights of Citizens* (commonly referred to as the *HEW Report*). The report was the culmination of an extensive study into data processing in the public and private sectors. The HEW Report recommended that Congress enact legislation adopting a "Code of Fair Information Practices" for automated personal data systems. The recommended Fair Information Practices became the foundation for the Privacy Act of 1974. The HEW Report Fair Information Practices included the following:

- There must be no personal data record-keeping systems whose very existence is secret.
- There must be a way for an individual to find out what information is in his or her file and how the information is being used.
- There must be a way for an individual to correct information in his or her records.
- Any organization creating, maintaining, using, or disseminating records of personally identifiable information must assure the reliability of the data for its intended use and must take precautions to prevent misuse.
- There must be a way for an individual to prevent personal information obtained for one purpose from being used for another purpose without his or her consent.

In 1980, the Organisation for Economic Co-operation and Development (OECD)[67] adopted *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, which provide a framework for privacy that has been referenced in U.S. Federal guidance and internationally. The OECD Guidelines, along with the Council of Europe Convention,[68] became the foundation for the European Union's Data Protection Directive.[69] The OECD Guidelines include the following Privacy Principles:

- **Collection Limitation—**There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
- **Data Quality—**Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.

---

66 Portions of this appendix were contributed to and published in the Executive Office of the President, National Science and Technology Council's *Identity Management Task Force Report 2008*, see http://www.ostp.gov/galleries/NSTC%20Reports/IdMReport%20Final.pdf.

67 The U.S. is an OECD member country and participated in the development of the OECD Privacy Guidelines, see http://www.ftc.gov/speeches/thompson/thomtacdremarks.shtm.

68 In 1981, the Council of Europe enacted the *Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data*, which also recognized the Fair Information Practices.

69 In 1995, the European Union enacted the *Data Protection Directive*, Directive 95/46/EC, which required member states to harmonize their national legislation with the terms of the Directive, including the Fair Information Practices. For additional information, see Jody R. Westby, *International Guide to Privacy*, American Bar Association Publishing, 2004.

- **Purpose Specification—**The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.
- **Use Limitation—**Personal data should not be disclosed, made available or otherwise used for purposes other than those specified, except with the consent of the data subject or by the authority of law.
- **Security Safeguards—**Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data.
- **Openness—**There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.
- **Individual Participation—**An individual should have the right: (a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him; (b) to have communicated to him, data relating to him within a reasonable time; at a charge, if any, that is not excessive; in a reasonable manner; and in a form that is readily intelligible to him; (c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and (d) to challenge data relating to him and, if the challenge is successful, to have the data erased, rectified, completed, or amended.
- **Accountability—**A data controller should be accountable for complying with measures which give effect to the principles stated above.

In 2004, the Asia-Pacific Economic Cooperation (APEC) ministers officially endorsed the Privacy Framework[70] developed within one of its committees. The APEC Privacy Framework was based on the OECD Privacy Guidelines, and was developed to encourage electronic commerce among the member states and to build trust with the international community. The Privacy Framework includes the following Privacy Principles:

- **Preventing Harm—**Recognizing the interests of the individual to legitimate expectations of privacy, personal information protection should be designed to prevent the misuse of such information. Further, acknowledging the risk that harm may result from such misuse of personal information, specific obligations should take account of such risk, and remedial measures should be proportionate to the likelihood and severity of the harm threatened by the collection, use and transfer of personal information.
- **Notice—**Personal information controllers should provide clear and easily accessible statements about their practices and policies with respect to personal information.
- **Collection Limitation—**The collection of personal information should be limited to information that is relevant to the purposes of collection and any such information should be obtained by lawful and fair means, and where appropriate, with notice to, or consent of, the individual concerned.
- **Uses of Personal Information—**Personal information collected should be used only to fulfill the purposes of the collection and other compatible related purposes, except with the consent of the

[70] http://www.apec.org/etc/medialib/apec_media_library/downloads/taskforce/ecsg/pubs/2005.Par.0001.File.v1.1.

individual, when necessary to provide a product or service requested by the individual, or by authority of law.

- **Choice—**Where appropriate, individuals should be provided with clear, prominent, easily understandable, accessible and affordable mechanisms to exercise choice in relation to the collection, use and disclosure of their personal information. It may not be appropriate for personal information controllers to provide these mechanisms when collecting publicly available information.
- **Integrity of Personal Information—**Personal information should be accurate, complete and kept up-to-date to the extent necessary for the purposes of use.
- **Security Safeguards—**Personal information controllers should protect personal information that they hold with appropriate safeguards against risks, such as loss or unauthorized access to personal information, or unauthorized destruction, use, modification or disclosure of information or other misuses. Such safeguards should be proportional to the likelihood and severity of the harm threatened, the sensitivity of the information and the context in which it is held, and should be subject to periodic review and reassessment.
- **Access and Correction—**Individuals should be able to obtain from the personal information controller confirmation of whether the personal information controller holders personal information about them, have the information provided to them at a reasonable charge and within a reasonable time, and challenge the accuracy of the information, as well as have the information corrected or deleted. Exceptions include situations where the burden would be disproportionate to the risks to the individual's privacy, the information should not be disclosed due to legal or security concerns, and the privacy of other persons would be violated.
- **Accountability—**A personal information controller should be accountable for complying with measures that give effect to the Principles stated above.

Appendix E—Sensitive Database Extracts Technical Frequently Asked Questions

This Frequently Asked Questions (FAQ) document[71] addresses technical aspects associated with implementing the Office of Management and Budget (OMB) requirement to log and verify sensitive database extracts, which was required by OMB Memorandum M-07-16, "Safeguarding Against and Responding to the Breach of Personally Identifiable Information" which reiterates the log and verify requirement set forward in M-06-16, "Protection of Sensitive Agency Information," issued in June 2006. Topics covered in this FAQ include data extract logging, restrictions, verification, and erasure.

NIST is particularly interested in reviewer suggestions for feasible technical mechanisms for the log and verify requirements. NIST encourages Federal agencies to provide feedback during the public comment period on the possible solutions described in the FAQ and to suggest additional technical solutions.

GENERAL

1. What is the requirement in the OMB memorandum?

The text of the requirement, as stated on page 7 of OMB M-07-16, is "Log all computer-readable data extracts from databases holding sensitive information and verify each extract, including whether sensitive data has been erased within 90 days or its use is still required."

2. What is a computer-readable data extract from a database?

This involves retrieving data from a database through a query and saving the data into a separate computer-readable entity such as another database, a spreadsheet, or a text file.

3. What types of information does the requirement apply to?

Although much of M-07-16 focuses on personally identifiable information (PII), the log and verify requirement applies to all sensitive information, including sensitive PII.

4. What is the purpose of the requirement?

The purpose of the requirement is to prevent data extracts containing sensitive information from being accessed by unauthorized parties. This is primarily a concern for mobile devices and removable user media. Ensuring that extracts with sensitive data are erased when they are no longer needed reduces the likelihood of sensitive information being breached.

LOGGING DATA EXTRACTS

5. Which data extracts need to be logged?

All data extracts from databases that are specifically performed by a human, saved to a separate file, and contain sensitive information need to be logged. Machine-to-machine transactions and any transactions that do not result in saving extracts to a file, such as an extract temporarily held in memory, do not need to be logged. If data extracts are well-protected using compensating controls—for example, data extracts

[71] This version of the FAQ is based on the original, which was posted on March 3, 2008 at http://csrc.nist.gov/drivers/documents/OMB/OMB-M-07-16-Data-Extract-FAQ.pdf. The version presented in this appendix includes clarifications from OMB on the intended scope of the requirement, affecting the answers to questions 4, 5, 8, and 12.

are stored on a logically well-secured server in a physically well-secured data center, or stored on properly encrypted media—then log and verify actions may not be necessary.

6. What information should be logged for each extract?

NIST Special Publication (SP) 800-53 Revision 2, *Recommended Security Controls for Federal Information Systems*, specifies an Audit and Accountability (AU) family of technical security controls, which encompasses audit logging requirements. Control number AU-3, Content of Audit Records, states that "audit record content includes, for most audit records: (i) date and time of the event; (ii) the component of the information system (e.g., software component, hardware component) where the event occurred; (iii) type of event; (iv) user/subject identity; and (v) the outcome (success or failure) of the event." In addition to logging this information for each extract, agencies may also log other types of information. For example, agencies may log whether each data extract contains sensitive information, for future use in determining which extracts need to be erased. Agencies may also describe the purpose and length of time for which extracted sensitive information will be used.

7. What recommendations does NIST provide for logging?

In addition to the audit logging-related security controls specified in NIST SP 800-53 Revision 2, NIST has developed SP 800-92, *Guide to Computer Security Log Management*. SP 800-92 provides recommendations for developing, implementing, and maintaining log management practices throughout an enterprise.

RESTRICTING DATA EXTRACTS

8. How can my agency reduce the number of data extracts that are subject to the requirement?

This can be accomplished by reducing the amount of sensitive information, including sensitive PII, in its databases and by limiting users' ability to perform extracts from databases with sensitive information. Also, as discussed in the answer to question 5, another option is the use of compensating controls.

9. What are some examples of how an agency can reduce the amount of sensitive PII in its databases?

As stated in OMB M-07-16, agencies must collect and retain only the minimum sensitive PII necessary. Agencies may also use de-identification and anonymization techniques to remove sensitive information from database records. These techniques can remove sensitive information permanently, such as replacing PII values with pseudonyms that provide the ability to sort and quantify populations as groups but not individuals. De-identification can also remove PII temporarily, such as mapping PII values to pseudonyms, storing the mappings in a separate file, and replacing the PII values in the database with the pseudonyms. Only an individual with access to both the database and the mapping file could match the individuals' actual identities with the corresponding database records.

10. How can an agency limit users' ability to perform data extracts from databases with sensitive information?

Agencies may grant only authorized users the least access necessary to such databases and to the sensitive information within each database. This could include restricting the types of queries that users can perform and the database fields (for example, social security number) that users can view and include in extracts. Another method is to permit users to access sensitive information in databases only through applications that tightly restrict the users' access to the sensitive information, instead of permitting direct

database access. Such applications could manage the data extract process by permitting extracts only when necessary, scrubbing sensitive information, such as sensitive PII, during extraction, forcing all extracts containing sensitive information to be stored centrally, and interacting with centrally-stored extracts on behalf of users so that the users cannot directly access extracts. Agencies may also use other options for limiting data extracts.

11. What technical methods are available for restricting where sensitive extracts are stored?

In addition to the application-based method mentioned above, there are other methods that agencies may use to limit where sensitive extracts are stored. For example, agencies may configure their remote access solutions so as not to permit access to sensitive information databases from mobile devices and non-organization computers (e.g., personally-owned home computers). Agencies could also permit extracts to be stored only on media protected by storage encryption technology. Other methods are more complex and may require considerable planning and deployment time. One example is requiring that sensitive extracts be stored within and accessed only through encrypted virtual machines, which may be set to expire after 90 days. Another example is implementing centralized processing for access to sensitive databases, where the data never leaves the centralized servers and the applications that access the data are run only through thin client solutions.

VERIFYING AND ERASING SENSITIVE DATA EXTRACTS

12. What is required for verifying a sensitive extract?

Agencies may accomplish extract verification through simple checks. An example of such a solution is ad hoc attestation. This involves implementing one or more systems to log the creation of extracts containing sensitive information and to send each extractor a message after 90 days that requires that the extractor either attests to having erased the extract or justifies why the extract is still needed. Agencies may implement more rigorous and formal verification processes than ad hoc attestation to achieve greater confidence in extracts being erased. An example of a more rigorous verification process is storing all extracts on a well-secured centralized system, prohibiting users from directly accessing the extracts, and running a utility that automatically erases extracts 90 days after creation. This assumes that the useful life of the extract ends on the day that it is created; the intent of the requirement is that extracts should be destroyed within 90 days after their useful life ends, which is not necessarily within 90 days of the extract creation date.

13. What is required for erasing a sensitive extract?

The actions needed to erase an extract vary based on the system or media where the extract has been stored. For example, erasing an extract stored on read-only removable media may necessitate physical destruction of the removable media, whereas erasing an extract on a centralized server may involve deleting the extract file and logically sanitizing the portions of the server media that held the file, as well as ensuring that all copies of the extract are properly erased from server backups. Data artifacts from extracts, such as temporary files, may also need to be erased. The procedures for erasing sensitive extracts can result in a significant operational impact on agencies.

14. What other types of technical solutions could be used for sensitive extract verification and erasure?

In addition to the solutions described above, agencies can also implement long-term solutions that automate most of the verification and erasure processes, thus reducing operational impact. Such solutions generally require at least a few years' effort to implement, so agencies that choose to implement one or

more of the long-term solutions may implement one or more of the currently available solutions described above in the meantime. Examples of possible long-term solutions are as follows:

- Use a trusted Digital Rights Management (DRM) platform or similar solution to manage extracts. Such technologies could be used to permit access to each extract for a certain number of days and by particular users, as well as to restrict how each extract can be used (e.g., preventing an extract from being copied or printed). Designing and implementing scalable DRM-type infrastructures and supporting systems for database extract management, including the deployment of client and server applications and platforms that support the chosen technology, is likely to require significant time and resources (at least two years).

- Implement centralized processing for access to sensitive databases using dumb terminals. This is similar to the thin client solution described earlier, except that the dumb terminals have no memory or storage, which prevents any data from being stored locally. Today's versions of "dumb terminals" are actually emulations that run on general-purpose computers, which means that sensitive data could be stored locally. This solution cannot be implemented on a large scale in the near term using current off-the-shelf components.

- Automatically encrypt each extract, centrally manage all the keys, and destroy the keys at the appropriate times to expire the extracts. Identity-based cryptography could extend this scheme to provide finer-grained access control. These methods are currently in the research stage and cannot be implemented in the near term.

Appendix F—Glossary

Selected terms used in the publication are defined below.

Aggregated: Information elements collated on a number of individuals, typically used for the purposes of making comparisons or identifying patterns.

Anonymized Information: Previously identifiable information that has been de-identified and for which a code or other link no longer exists.

Confidentiality: "Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information." [44 U.S.C., Sec. 3542, http://uscode.house.gov/download/pls/44C35.txt]

Context of Use: The purpose for which PII is collected, stored, used, processed, disclosed, or disseminated.

De-identified Information: Records that have had enough PII removed or obscured such that the remaining information does not identify an individual and there is no reasonable basis to believe that the information alone can be used to identify an individual.

Distinguishable Information: Information that can be used to identify an individual.

Harm: Any negative or unwanted effects that would be experienced by an individual (i.e., that may be socially, physically, or financially damaging) or an organization if the confidentiality of PII were breached.

Linkable Information: Information that is not sufficient to allow the recipient to distinguish any individual, but that may be matched or compared to information from a secondary data source that is available to the general public or can be otherwise obtained, in order to link together information and potentially distinguish individuals.

Linked Information: Information that is not sufficient to distinguish an individual when considered separately, but which could distinguish an individual when taken collectively or if considered in conjunction with other data elements in the same system or in a closely-related system.

Obscured Data: Data that has been distorted by cryptographic or other means to hide information. It is also referred to as being masked or obfuscated.

Personally Identifiable Information (PII): "Information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc." [OMB Memorandum 07-16]

PII Confidentiality Impact Level: The level of impact on organizations and individuals should there be a breach of confidentiality involving PII. The possible levels are low, moderate, and high.

Privacy Impact Assessment (PIA): An analysis of how information is handled that ensures handling conforms to applicable legal, regulatory, and policy requirements regarding privacy; determines the risks and effects of collecting, maintaining and disseminating information in identifiable form in an electronic

information system; and examines and evaluates protections and alternative processes for handling information to mitigate potential privacy risks.

System of Records: A group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual.

Appendix G—Acronyms and Abbreviations

Selected acronyms and abbreviations used in the publication are defined below.

AC	Access Control
APEC	Asia-Pacific Economic Cooperation
CD	Compact Disc
CFR	Code of Federal Regulations
CIPSEA	Confidential Information Protection and Statistical Efficiency Act
COPPA	Children's Online Privacy Protection Act
DRM	Digital Rights Management
FAQ	Frequently Asked Questions
FIPS	Federal Information Processing Standards
FISMA	Federal Information Security Management Act
GAO	Government Accounting Office
GLBA	Gramm-Leach-Bliley Act
HEW	U.S. Department of Health, Education, and Welfare
HIPAA	Health Insurance Portability and Accountability Act
IA	Identification and Authentication
ID	Identification
IIF	Information in Identifiable Form
IIHI	Individually Identifiable Health Information
IP	Internet Protocol
IPA	Initial Privacy Assessment
IRS	Internal Revenue Service
ISA	Interconnection Security Agreement
IT	Information Technology
ITL	Information Technology Laboratory
MAC	Media Access Control
MP	Media Protection
NARA	National Archives and Records Administration
NIST	National Institute of Standards and Technology
NPPI	Non-Public Personal Information
OECD	Organisation for Economic Co-operation and Development
OMB	Office of Management and Budget
OPM	Office of Personnel Management
PDA	Personal Digital Assistant
PHI	Protected Health Information
PIA	Privacy Impact Assessment
PII	Personally Identifiable Information
PRA	Paperwork Reduction Act

PTA	Privacy Threshold Assessment
SC	System and Communications Protection
SDLC	System Development Life Cycle
SOR	System of Records
SORN	System of Records Notice
SP	Special Publication
SSN	Social Security Number
URL	Uniform Resource Locator
USB	Universal Serial Bus
U.S.C.	United States Code
US-CERT	United States Computer Emergency Response Team

Appendix H—Resources

Those personnel involved with protecting PII and concerned about individual and organizational impact may want to review the following privacy laws and requirements that apply to Federal agencies.[72] Additionally, OMB has issued several memoranda that provide policy guidance and instructions for the implementation of privacy requirements.

Document	URL
Children's Online Privacy Protection Act (COPPA)	http://www.ftc.gov/ogc/coppa1.htm
Confidential Information Protection and Statistical Efficiency Act (CIPSEA)[73]	http://www.whitehouse.gov/omb/inforeg/cipsea/cipsea_statute.pdf
Confidential Information Protection and Statistical Efficiency Act (CIPSEA) Implementation Guidance	http://www.whitehouse.gov/omb/fedreg/2007/061507_cipsea_guidance.pdf
Consolidated Appropriations Act of 2005, Section 522	http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_bills&docid=f:h4818enr.txt.pdf
E-Government Act of 2002, Section 208	http://thomas.loc.gov/cgi-bin/query/z?c107:H.R.2458.ENR:
Federal Information Security Management Act (FISMA)[74]	http://csrc.nist.gov/drivers/documents/FISMA-final.pdf
Intelligence Identities Protection Act of 1982 (50 U.S.C. 421 et seq.)	http://caselaw.lp.findlaw.com/casecode/uscodes/50/chapters/15/subchapters/iv/sections/section_421.html
FIPS 140-2, *Security Requirements for Cryptographic Modules*	http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf
FIPS 199, *Standards for Security Categorization of Federal Information and Information Systems*	http://csrc.nist.gov/publications/fips/fips199/FIPS-PUB-199-final.pdf
Freedom of Information Act (FOIA)[75]	http://www.justice.gov/oip/amended-foia-redlined.pdf
Gramm-Leach-Bliley Act (GLBA)	http://thomas.loc.gov/cgi-bin/query/z?c106:S.900.ENR:
Health Insurance Portability and Accountability Act (HIPAA)	http://aspe.hhs.gov/admnsimp/pl104191.htm
Implementing Recommendations of the 9/11 Commission Act of 2007	http://www.govtrack.us/congress/bill.xpd?bill=h110-1
NIST SP 800-30, *Risk Management Guide for Information Technology Systems*	http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf
NIST SP 800-37 Revision 1 (draft), *Guide for the Security Certification and Accreditation of Federal Information Systems*	http://csrc.nist.gov/publications/drafts/800-37-Rev1/SP800-37-rev1-IPD.pdf
NIST SP 800-47, *Security Guide for Interconnecting Information Technology Systems*	http://csrc.nist.gov/publications/nistpubs/800-47/sp800-47.pdf

[72] This list is provided for reference only and is not an exhaustive list. For additional information, an organization's legal counsel and privacy officer should be consulted.

[73] CIPSEA is Title V of the E-Government Act of 2002.

[74] FISMA is Title III of the E-Government Act of 2002.

[75] FOIA was recently amended by the *OPEN Government Act of 2007*, Pub. L. No. 110-175 (2007).

Document	URL
NIST SP 800-53 Revision 2, *Recommended Security Controls for Federal Organizations and Information Systems*	http://csrc.nist.gov/publications/nistpubs/800-53-Rev2/sp800-53-rev2-final.pdf
NIST SP 800-60 Revision 1, *Volume 1: Guide for Mapping Types of Information and Information Systems to Security Categories*	http://csrc.nist.gov/publications/nistpubs/800-60-rev1/SP800-60_Vol1-Rev1.pdf
NIST SP 800-61 Revision 1, *Computer Security Incident Handling Guide*	http://csrc.nist.gov/publications/nistpubs/800-61-rev1/SP800-61rev1.pdf
NIST SP 800-63 Version 1.0.2, *Electronic Authentication Guidelines*[76]	http://csrc.nist.gov/publications/nistpubs/800-63/SP800-63V1_0_2.pdf
Office of Personnel Management (OPM), *Guidance on Protecting Federal Employee Social Security Numbers and Combating Identity Theft*, June 2007	http://www.chcoc.gov/Transmittals/TransmittalDetails.aspx?TransmittalID=847
OMB Circular A-130, *Management of Federal Information Resources*	http://www.whitehouse.gov/omb/circulars/a130/a130.html
OMB Memorandum M-01-05, *Guidance on Inter-agency Sharing of Personal Data – Protecting Personal Privacy*	http://www.whitehouse.gov/omb/memoranda/m01-05.html
OMB Memorandum M-03-22, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*	http://www.whitehouse.gov/omb/memoranda/m03-22.html
OMB Memorandum M-04-04, *E-Authentication Guidance for Federal Agencies*	http://www.whitehouse.gov/omb/memoranda/fy04/m04-04.pdf
OMB Memorandum M-05-08, *Designation of Senior Agency Officials for Privacy*	http://www.whitehouse.gov/omb/memoranda/fy2005/m05-08.pdf
OMB Memorandum M-06-15, *Safeguarding Personally Identifiable Information*	http://www.whitehouse.gov/omb/memoranda/fy2006/m-06-15.pdf
OMB Memorandum M-06-16, *Protection of Sensitive Agency Information*	http://www.whitehouse.gov/omb/memoranda/fy2006/m06-16.pdf
OMB Memorandum M-06-19, *Reporting Incidents Involving Personally Identifiable Information and Incorporating the Cost for Security in Agency Information Technology Investments*	http://www.whitehouse.gov/omb/memoranda/fy2006/m-06-19.pdf
OMB Memorandum M-06-20, *FY 2006 Reporting Instructions for the Federal Information Security Management Act and Agency Privacy Management*	http://www.whitehouse.gov/omb/memoranda/fy2006/m06-20.pdf
OMB Memorandum M-07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information*	http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf

[76] NIST 800-63-1 was released as a draft in December 2008, http://csrc.nist.gov/publications/drafts/800-63-rev1/SP800-63-Rev1_Dec2008.pdf.

Document	URL
OMB Memorandum M-07-19, *FY 2007 Reporting Instructions for the Federal Information Security Management Act and Agency Privacy Management*	http://www.whitehouse.gov/omb/memoranda/fy2007/m07-19.pdf
OMB Memorandum M-08-09, *New FISMA Privacy Reporting Requirements for FY 2008*	http://www.whitehouse.gov/omb/memoranda/fy2008/m08-09.pdf
OMB Memorandum, September 20, 2006, *Recommendations for Identity Theft Related Data Breach Notification*	http://www.whitehouse.gov/omb/memoranda/fy2006/task_force_theft_memo.pdf
OMB Memorandum, July 2007, *Common Risks Impeding the Adequate Protection of Government Information* (developed jointly with DHS)	http://csrc.nist.gov/pcig/document/Common-Risks-Impeding-Adequate-Protection-Govt-Info.pdf
Paperwork Reduction Act	http://www.archives.gov/federal-register/laws/paperwork-reduction/
President's Identity Theft Task Force, *Combating Identity Theft: A Strategic Plan*, April 2007	http://www.idtheft.gov/reports/StrategicPlan.pdf
Privacy Act of 1974	http://www.usdoj.gov/oip/privstat.htm
Sensitive Database Extracts Technical Frequently Asked Questions	http://csrc.nist.gov/drivers/documents/OMB/OMB-M-07-16-Data-Extract-FAQ.pdf