Sensitive Database Extracts Technical Frequently Asked Questions

This Frequently Asked Questions (FAQ) document 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.

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 apples to all sensitive information, including sensitive PII.

4. **What is the purpose of the requirement?**
   The purpose of the requirement is to ensure that data extracts containing sensitive information are erased when they are no longer needed. This 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 contain sensitive information need to be logged.

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.

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 data scrubbing techniques to remove sensitive information from database records. Data scrubbing 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. Data scrubbing 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.

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.