Privacy Impact Assessment

for the

Electronic Discovery (eDiscovery) Tools

DHS Reference No. DHS/ALL/PIA-073(a)

September 17, 2020

Abstract

The U.S. Department of Homeland Security (DHS) uses commercial off-the-shelf electronic discovery (eDiscovery) software tools to aid in the collection of existing departmental records and to facilitate the review and production of electronically stored information (ESI), such as emails, computer files, and databases. eDiscovery software is a document processing tool that supports the organization of ESI for analysis, review, redaction, and production to meet DHS requirements. This Privacy Impact Assessment (PIA) Update discusses the Department's use of this technology for the purposes of internal investigations, complaint processing, Freedom of Information Act (FOIA) or Privacy Act (PA) requests, research, and other Departmental issues and requirements.

Overview

DHS increasingly relies upon ESI to conduct departmental business. Accordingly, DHS requires the use of eDiscovery software tools to preserve, tag, review, redact, and produce Department records. In some cases, DHS may have an obligation to produce these records in the same format in which it is originally compiled or maintained ("native format"), along with any associated metadata,[1] a task eDiscovery tools can accomplish. DHS uses these tools to support a variety of efforts, including internal investigations, complaint processing, FOIA or PA requests, research, and other Departmental issues.

The range of capabilities varies across the Department, but eDiscovery tools generally streamline and automate the ESI review process. As the first step in this process, users load information in various data formats (e.g., PST (emails), .doc/PDF (documents), .PPS (presentations), TIFF/JPEG (pictures), and databases), allowing document analysis using a single integrated viewer that does not require use of the original application that created the file. Second, the tools allow for the identification of duplicate or near-duplicate documents, enable the threading of emails, and maintain parent/child relationships of the documents in the review process. eDiscovery tools also allow operators to view metadata within files stored in these varying file formats. Third, the tools aid in identifying and tagging information by allowing users to conduct keyword and Boolean logic searches across the record set. The eDiscovery tools are then able to use that information to automatically flag files that contain those keywords or Boolean statements for the operator, who then reviews and determines whether the files or information therein are responsive for the purposes of their search. Finally, these tools allow operators to electronically redact protected portions of documents in the system.

[1] Metadata are the data attributes that may or may not be hidden that may reveal sensitive information about a document or file's history, such as its creator, last editor, or version history, among others.

As part of the eDiscovery process, records loaded into the tools are maintained in their original form and not modified. The tools may create new information that is associated with those records, but that data does not alter the integrity of the original records themselves. The created data consists of redactions, tags, privilege logs, and search and filter reports, which are automatically created. The software also maintains a historical record of all actions taken by users. These eDiscovery tools also can assign a unique key to the original unaltered file, which helps establish the chain of custody and proof that the content of the produced file or document has not been altered from the original version.

eDiscovery tools at DHS include the following functions:

Forensics Tools – These eDiscovery tools are used for fact-finding purposes. They retrieve and process ESI for official business and legal purposes, such as the resolution of investigations and court cases. Generally, ESI consists of email messages. However, documents, internet cookies, logon/logoff information from desktops, photos, and spreadsheets can also be uploaded to and stored in the system. Any attachments sent in email messages are saved with the respective emails and become part of searches. The content of email messages and other items saved could include personal information (e.g., custodian name, email address) on DHS employees and contractors, personnel from other federal agencies, and members of the public.

Digital Investigative Tools – These tools provide the capability to search, identify, replicate, and process potential forensics evidence on the DHS network, workstations, and mobile devices and present it to users for analysis. Digital investigative tools allow DHS to analyze and evaluate any cybersecurity misconduct and criminal activity on seized and non-seized assets. These tools may collect personally identifiable information (PII) involved in an internal or external cybersecurity threat, incident and/or activity related to DHS personnel. This data is not permanently retained within the tools as the extracted information is turned over to the appropriate office conducting the investigation.

Event Log Tools – These tools are able to pull logon/logoff information from individual workstations/desktops. In order to pull the information from individual workstations, the tool collects and retains the respective identifications (IDs) of the personnel involved. The user history is retained on an individual’s workstation until the machines are reimaged or data is erased by the system administrators.

Internet Browsing Tools – DHS uses these tools when it receives requests to pull internet browsing data (e.g., bookmarks, cached files and web pages, cookies, downloads, email addresses, favicons (shortcut icons), form history, logins, searches, session tabs, thumbnails and website visits) stored on appropriately-provisioned DHS information technology (IT) in the course of conducting investigations or for other official purposes. The history on a specific user is retained until manually deleted, or the workstation is remapped for new user profiles.

Reason for the PIA Update

This PIA is being updated to account for the use of eDiscovery tools by the Department. The previous PIA only discussed the use of eDiscovery tools for litigation purposes. The scope is now expanded to other Departmental business where the searching of large volumes of electronic documents is required and to include internal investigations, complaints, FOIA/PA requests, research, and other Departmental requirements.

Privacy Impact Analysis

Authorities and Other Requirements

DHS conforms to the requirements of the Federal Records Act,[2] which provides the basis for the federal government's policies and procedures for creating, maintaining, and disposing of federal records. The Act and its related regulations define federal records, mandate the creation and preservation of records necessary to document federal activities, establish government ownership of records, and provide the exclusive legal procedures for the disposition of records. While the Act provides the overarching framework for federal recordkeeping, the context of the records being analyzed determines the specific legal authority that permitted their original collection.

The context of the data being analyzed by the tools will determine the applicable system of records notice (SORN). For example, any records relating to an internal investigation would be covered by the DHS Internal Affairs SORN;[3] any records relating to a FOIA/PA request would be covered by the DHS Freedom of Information Act (FOIA) and Privacy Act (PA) Record System SORN;[4] and any records relating to complaints would be covered by the DHS Complaint Tracking System or Civil Rights and Civil Liberties Records SORNs.[5]

The DHS General Information Technology Access Account Records System SORN[6] covers any logs, audits, or other security data regarding the use of such information technology

[2] Federal Records Act, Pub. L. No. 81-754, 64 Stat. 578 (1950) (current version in sections of 44 U.S.C. § 3101 et seq.).

[3] *See* DHS/ALL-020 Department of Homeland Security Internal Affairs, 79 Fed. Reg. 23361 (April 28, 2014), *available at* https://www.dhs.gov/system-records-notices-sorns.

[4] *See* DHS/ALL-001 Department of Homeland Security (DHS) Freedom of Information Act (FOIA) and Privacy Act Record System, 79 Fed. Reg. 6609 (February 4, 2014), *available at* https://www.dhs.gov/system-records-notices-sorns.

[5] DHS/All-028 Department of Homeland Security Complaint Tracking System, 74 Fed. Reg. 35877 (July 21, 2009), and DHS/ALL-029 Civil Rights and Civil Liberties Records, 75 Fed. Reg. 39266 (July 8, 2010), *available at* https://www.dhs.gov/system-records-notices-sorns.

[6] *See* DHS/ALL-004 General Information Technology Access Account Records System, 77 Fed. Reg. 70792 (November 27, 2012), *available at* https://www.dhs.gov/system-records-notices-sorns.

resources. In addition, this SORN covers the access to and use of these types of tools and resources by authorized individuals.

The users of eDiscovery tools determine how long they require the data to be retained pursuant to the type of record contained and the records retention schedule set by the National Archives and Records Administration (NARA) for such records.

Characterization of the Information

eDiscovery tools store and process agency records, as necessary, to satisfy a variety of requirements. Information in eDiscovery tools could consist of any information in DHS formal or informal recordkeeping systems or any paper documents scanned into an electronic format for review. Because the tools are document processing systems, the records that may be stored and processed could pertain to any matter in the scope of DHS's mission and may contain PII or sensitive personally identifiable information (SPII) of any nature captured and stored in such records. The actual information stored and processed will always vary depending on the nature and purpose of a particular search.

Privacy Risk: There is a risk that eDiscovery tools may over-collect PII.

Mitigation: This risk is not fully mitigated. DHS only collects and processes information in these tools that it already has the authority to collect. Use of the tools is limited to those who have a need to use this technology. Furthermore, the data these users have access to can be limited to that in which they have a need-to-know. These tools can also provide role-based access to ensure that records stored and processed are only accessed to those individuals who should have access to specific data. Additionally, the eDiscovery tools have the capability to generate a robust audit trail of all user activity, including the viewing of records in the system.

Privacy Risk: There is a risk that information processed by eDiscovery is inaccurate.

Mitigation: This risk is not fully mitigated. The system contains only copies of records from other DHS recordkeeping systems, and DHS will not use the tools as an internal source of agency records about individuals. The purpose of eDiscovery tools is to support other Departmental functions such as performing internal investigations, responding to FOIA/PA requests, or conducting other Departmental activities. The processing that is conducted through eDiscovery tools does not alter the original records. Incorrect information, when identified, can be corrected in the source system.

Uses of the Information

DHS uses the information loaded into eDiscovery tools to support its broad functions and responsibilities. DHS uses the production file generated by the tools to produce portions of records in electronic and searchable form for release to the public, for internal investigations, or for Departmental functions.

Privacy Risk: There is a privacy risk of unauthorized use of the information maintained in eDiscovery tools.

Mitigation: This risk is mitigated. eDiscovery tools employ appropriate role-based access controls so only authorized personnel have access to the system and to the appropriate information based on the needs of their responsibilities. System/tool administrators also grant access to tools in the first place. An individual user would only be able to process information in eDiscovery tools that he or she already has access to through their other job responsibilities. Those users may receive training regarding the proper use of that specific data. Additionally, all users complete annual mandatory privacy and security training.

Moreover, these tools are able to maintain detailed audit logs that would capture any user's inappropriate use of information contained within the tool. These logs may be automatically generated and can be reviewed by IT security or administrators when there is evidence or reason to believe that the integrity of the data or its use has been compromised.

Notice

Because eDiscovery tools are not a primary information collection system, DHS does not provide notice to individuals prior to the tools' collection of information. This PIA Update serves as notice to the general public to the collection and use of information in these types of tools to fulfill Departmental requirements. DHS provides notice at the point of original collection wherever possible; however, in cases in which the data collection supports a law enforcement activity, opportunities for the individual to be notified of the collection of information may be limited or nonexistent. The purpose and context of the original collection of information determines whether notice is provided.

Privacy Risk: There is a risk that individuals are not provided notice that their records are processed in eDiscovery tools.

Mitigation: This risk is partially mitigated. This PIA Update, and its original, serve as public notice of the existence of eDiscovery tools, the data they collect and maintain, and the purposes for which DHS will use the data. However, because these tools support a secondary collection of information from records already compiled in existing agency recordkeeping systems, individualized notice is not possible or practical.

Data Retention by the Project

Depending on the nature of the records being processed, the retention policies will differ. All records should be maintained pursuant to the records retention schedules for the source systems' documents.

Privacy Risk: There is a privacy risk that information will be retained for longer than necessary to accomplish the purpose for which the information was originally collected.

Mitigation: This risk is mitigated. DHS maintains the records in eDiscovery tools according to the records schedules and policies specific to the type of data being processed and disposes of them accordingly. This responsibility falls on the individual user processing the information, as well as system administrators of the tools.

Information Sharing

DHS shares information stored and processed in eDiscovery tools depending on the nature of the records; and does so in accordance with routine uses in the SORNs that cover those records. As an example, for responsive requests (e.g., FOIA/PA requests), DHS may ultimately share the information stored and processed in these tools with the requester to the extent the information is not subject to withholding under an exemption or exception.

Privacy Risk: There is a risk that information collected in eDiscovery tools will be disclosed inappropriately.

Mitigation: This risk is partially mitigated. Because any sharing of information that was processed in eDiscovery tools would happen outside of the tools themselves, the tools cannot record that sharing. However, because the information processed in eDiscovery tools can be limited to those with a need-to-know, the data is maintained by those who handle this type of information during the course of their workplace responsibilities. This ensures that any sharing outside of the tools would be done by an individual who has a responsibility to share the information in the first place.

Redress

Because of the nature of eDiscovery tools as a repository for records gathered from other DHS recordkeeping systems pursuant to Departmental responsibilities, the tools are not designed to allow the individual to correct inaccurate or erroneous information about him or herself. However, individuals seeking access to any record contained in this system of records may submit a Privacy Act (for U.S. citizens and Lawful Permanent Residents) or FOIA (for all individuals) request to the respective component FOIA Office which can be found under "Contact Information" at https://www.dhs.gov/freedom-information-act-foia. One of the many purposes of eDiscovery tools is to conduct searches of responsive records for these types of FOIA/PA requests.

Additionally, depending on the nature of the records being processed, DHS may be unable to provide individual access to records contained in eDiscovery tools as they could inform the subject of an actual or potential investigation or reveal an investigative interest on the part of DHS.

Auditing and Accountability

eDiscovery tools maintain audit logs to monitor and track document review functions, as needed. This audit trail can assist investigating officials in identifying unauthorized use of the system so that DHS may take any appropriate follow-up actions.

All eDiscovery tools are required to undergo a privacy analysis (i.e., Privacy Threshold Analysis (PTA)) to ensure compliance with this PIA and other Departmental policies.

Responsible Official

Original, signed copy on file at the DHS Privacy Office.

____

Dena Kozanas
Chief Privacy Officer
U.S. Department of Homeland Security
(202) 343-1717