

Description of IRS Email Collection and Production

Over the past year, the Internal Revenue Service made a massive document production in response to Congressional and other inquiries. This activity has been challenging since processing email for production to third parties is a more complex process for the IRS than it is for many private or public organizations. Below we analyze why it is so complicated for the agency to respond to what otherwise in this modern day seem like straightforward requests, including an assessment of what is and is not currently possible. Sophisticated IRS information technology systems are designed to facilitate tax administration, cost-effective use of resources, and preserve confidential taxpayer information, not to facilitate matters related to document preservation, collection, processing, and review. The IRS faces unique challenges in producing email to third parties because of how its email is stored, the security required for IRS email, and the laws protecting confidential taxpayer information from disclosure.

I. Background

Over the past year, four Congressional committees, the Treasury Inspector General for Tax Administration (TIGTA), and the Department of Justice have conducted investigations related to the processing and review of applications for tax-exempt status as described in the May 2013 TIGTA report, *Inappropriate Criteria Were Used to Identify Tax-Exempt Applications for Review*. Congressional committees and individual members of Congress made hundreds of requests for information related to the issues raised in the TIGTA report.

In response, the IRS undertook an unprecedented document collection and production effort. As requested by investigators, electronic mail was a primary focus of IRS efforts. As of mid-March 2014, the Senate Finance Committee and the House Ways & Means Committee had received documents the IRS had identified as related to the processing and review of applications for tax-exempt status as described in the May 2013 report. Since then, IRS efforts have focused on completing the redaction of those materials for production to other committees and, in response to Congressional requests, production of email (on all topics) involving Lois Lerner, former Director of the Exempt Organizations division at the IRS.

More than 1.3 million pages of material have been produced; 750,000 pages in unredacted form to the Congressional committees authorized to receive taxpayer information protected under Section 6103¹ and another 600,000 pages in redacted format to other committees. None of the IRS systems (e.g., email storage, document collection functions) were designed to facilitate such extensive reviews and productions; as a result, the process required significant human capital and financial resources. More than 250 IRS and Chief Counsel employees have spent over 120,000 hours

¹ Under 26 U.S.C. §6103, tax “[r]eturns and return information shall be confidential,” and may not be revealed absent statutory authority. Section 6103(f) provides that upon written request by the Chairman of the Senate Finance Committee or the House Ways & Means Committee, the IRS may provide such information to the Chairman’s designee(s).

working on compliance with the investigations – at a direct cost of nearly \$10 million. Many of these employees worked on the document production and review process to the exclusion of their normal workload for months at a time. The IRS also spent an additional \$6-8 million to optimize existing information technology systems and ensure a stable infrastructure for those productions.

II. Physical Retention, Collection, and Production of Email

The IRS email system runs on Microsoft Outlook. Each of the Outlook email servers are located at one of three IRS data centers. Approximately 170 terabytes of email (178,000,000 megabytes, representing literally hundreds of millions of emails) are currently stored on those servers. For disaster recovery purposes, the IRS does a daily back-up of its email servers. The daily back-up provides a snapshot of the contents of all email boxes as of the date and time of the backup. Prior to May 2013, these backups were retained on tape for six months, and then for cost-efficiency, the backup tapes were released for re-use. In May of last year, the IRS changed its policy and began storing rather than recycling its backup tapes.²

A. Email Preservation

In late May 2013 and early June 2013, the IRS sent document retention notices to employees identified as having documents (including email) potentially relevant to the investigations. These notices instructed employees not to alter or destroy:

all communications, documents drafted or reviewed, spreadsheets created or reviewed, notes from meetings, notes relating to specific taxpayers and/or applications, information requests to applicants, training materials, or any other items that relate to the process by which selection criteria were used to identify tax-exempt applications for advocacy organizations for review, including but not limited to Be On the Look Out, from January 1, 2008 to the present.³

In that same time frame, the IRS sent similar document retention notices to all employees in the IRS Tax Exempt and Government Entities function and its Chief Counsel counterpart; the IRS Communications and Liaison function; and all employees assigned to respond to the Congressional inquiries.

B. Employees' Email Storage

The IRS has approximately 90,000 employees. Due to financial and practical considerations, the IRS has limited the total volume of email stored on its server by restricting the amount of email most individual users can keep in an inbox at any given time. This is not an uncommon practice within the government or the private sector.

² This practice of retaining rather than recycling tapes is estimated to cost approximately \$200,000 annually.

³ Litigation Hold, Attachment A.

According to estimates, it would cost well over ten million dollars to upgrade the IRS information technology infrastructure in order to save and store all email ever sent or received by the approximately 90,000 current IRS employees.

Currently, the average individual employee's email box limit is 500 megabytes, which translates to approximately 6,000 emails. See Attachment B. Prior to July 2011, the limit was lower, 150 megabytes or roughly 1,800 emails. See Attachment C. The IRS does not automatically delete email in its employees' email account to meet these limits; rather, each employee is responsible for managing and prioritizing the information stored within his/her email box.

Historically, the email of IRS employees is stored in two locations – email in an individual's active email box and therefore saved on the IRS centralized network and email archived on the individual employee's computer hard drive.⁴ If an email user's mailbox gets close to capacity, the system sends a message to the user noting that soon the mailbox will become unable to send additional messages.

When a user needs to create space in his or her email box, the user has the option of either deleting emails (that do not qualify as official records) or moving them out of the active email box (inbox, sent items, deleted items) to an archive. In addition, if an email qualifies as an official record, per IRS policy, the email must be printed and placed in the appropriate file by the employee.⁵ Archived email is moved off the IRS email server and onto the employee's hard drive on the employee's individual computer. As a result, these IRS employees' emails no longer exist in the active email box of the employee and are not backed-up as part of the daily backup of the email servers. Email moved to a personal archive of an employee exists only on the individual employee's hard drive. An electronic version of the archived email would not be retained if an employee's hard drive is recycled or if the hard drive crashes and cannot be recovered.

C. Email Collection

There have been questions from third parties about the speed of the review and production of IRS email materials; it is therefore important to understand the features of the IRS email system that make the process difficult and time-consuming. As noted above, the IRS has approximately 90,000 employees, each of whom conceivably could have responsive electronic data to any given request. There is no mechanism to allow IRS to search across its entire email system. To gather email from IRS employees, each potential custodian's mailbox and hard drive must be individually collected.

⁴ The approximately 2,000 IRS Counsel employees (as opposed to IRS employees) have a system that allows archived email to be stored on a central drive.

⁵ An official "record" is any documentary material made or received by an agency under federal law or in connection with the transaction of public business and appropriate for preservation (44 U.S.C. § 3301). Not all of the emails on IRS servers or backup tapes qualify as official records; accordingly, the agency's email system does not retain all email indefinitely. Rather, individual employees are responsible for ensuring that any email in their possession that qualifies as a "record" is retained in accordance with the requirements in the Internal Revenue Manual (IRM) and Document 12990 (Record Control Schedules).

Collecting from a hard drive involves an Information Technology employee taking physical possession of the computer and copying the contents of the computer's hard drive. These collection efforts are inherently labor-intensive and time-consuming.

The technology used by the IRS does not permit the IRS to select identifiable emails or groups of emails relating to a particular matter from a particular employee. Instead, all of an employee's email must be collected to start the processing function and limited, if at all, later in the processing function by date restrictions and search terms.

The technology used by the IRS also does not permit the IRS to search the network across multiple employees' email in connection with a particular matter. Similarly, it is impossible currently to search all IRS employees' accounts for email to a particular domain. As a result, to find literally "all email" in the custody and control of IRS on a given topic or to a particular domain (e.g., a specific government agency), every single employee's email would need to be individually collected, then processed and reviewed for responsive material.

D. Email Processing

After an employee's email is collected, it needs to be processed and properly formatted so that it can be searched and analyzed for content potentially responsive to a particular request for information. This typically involves "flattening" the email message to make it readable by, for example, an eDiscovery platform tool. This also involves decrypting the email message – for security and privacy reasons, much of IRS email is encrypted when sent. In order to decrypt an email message, the system must use the individual custodian's encryption certificate and multiple reprocessing steps, so that the email can be readable using an eDiscovery platform tool. Once properly flattened and decrypted, email can be loaded onto an eDiscovery platform tool, where it can be searched using search terms and/or date limitations when appropriate. At this point, the materials are ready for review.

E. Email Review and Redaction to Protect Taxpayer Information

The final step before email can be produced involves identifying the relevant materials and taking steps where necessary to protect confidential taxpayer return information. In the course of our productions, the IRS has reviewed and produced email collected from 83 custodians. The IRS has a unique responsibility to protect confidential taxpayer information as required by I.R.C. § 6103.⁶ All emails that might contain statutorily protected information must be reviewed and if necessary redacted.

Because when an email is sent it then exists in the author's and all recipients' email boxes, multiple copies of any one email occur frequently in document review. Although the IRS eDiscovery platform tool has a feature that eliminates certain duplicate emails before they are produced to a third party, the "deduping" feature only eliminates

⁶ Taxpayers may have a civil cause of action when these taxpayer confidentiality rules are violated and IRS employees may be subject to disciplinary action and criminal penalties if they violate these rules.

duplicate emails that are virtually identical in every respect. A slight variance in the information contained in one email versus another, e.g., the time sent and the time received, results in the emails being treated as unique documents, which translates into increased review and processing times and added volume of documents produced to a third party.

III. Investigations' Requests and IRS Production

The Congressional committees investigating issues raised in the May 2013 TIGTA report requested a broad array of materials, with date ranges that span multiple years (e.g., from 2009 through mid-2013). As outlined in the attached August 29, 2013 letter,⁷ the IRS collaborated with Congressional staff to select search terms and custodians with the goal of gathering and producing information as prioritized by Congress. This document production effort has involved hundreds of people, hundreds of thousands of hours, and millions of dollars.

The IRS followed the process described above in preserving, collecting, processing, and reviewing material in response to Congressional requests related to the processing and review of applications for tax-exempt status as described in the May 2013 TIGTA report. The IRS completed a search of 83 custodians' email using specifically identified search terms, and reviewed the results for responsiveness and for confidential taxpayer information. When email was responsive, it was produced and redacted when required by Section 6103.

Generally, the IRS produced all documents to all six investigations. There are situations, however, in which materials were produced only to investigators with authority to see information protected by Section 6103. One such situation relates to taxpayer files, which are protected in their entirety. Another is a collection of Excel spreadsheets and associated documents that were produced in native format, which format cannot be redacted for Section 6103 material.

In responding to the investigations, the IRS has not withheld relevant documents on the grounds of privilege.

Many times over the course of the year, different committees expressed interest in specific people, time periods, or events. The IRS did its best to accommodate these requests and expedite material in the priority set by investigators. In early 2014, Chairmen Camp and Issa reiterated their requests for all of Lois Lerner's email, regardless of subject matter. Because of Ms. Lerner's unavailability for Congressional interviews and in response to the Chairmen's requests, the IRS agreed to produce or make available for Congressional review all of her email.

Fulfilling the request for Ms. Lerner's emails regardless of subject matter required the IRS to load additional email beyond the email responsive to search terms originally

⁷ Attachment D, Leonard Oursler August 29, 2013 letter to Senator Baucus, also sent to Senators Hatch, Levin, and McCain and Congressmen Camp, Levin, Issa, and Cummings.

loaded for review from Ms. Lerner's custodial email box. By mid-March, IRS had produced to the tax-writing committees the Lerner-related (and other) materials it had identified as related to the processing and review of applications for tax-exempt status as described in the May 2013 TIGTA report.⁸ IRS then focused on redacting materials for the non-taxwriters and processing the rest of Ms. Lerner's email for production.

Ms. Lerner's emails were subject to the same preservation and collection process as other materials that the IRS produced to investigators.⁹ The IRS put Ms. Lerner on administrative leave as of May 23, 2013, at which date she was no longer permitted to access her computer or blackberry. On September 23, 2013 Ms. Lerner separated from the Service.

The electronic data collection for Ms. Lerner's custodial email was completed on May 22, 2013. According to personnel involved in the collection of Ms. Lerner's email, the materials on Ms. Lerner's computer were successfully captured in the data collection process.¹⁰ All of the email from 2009 through 2013 that the IRS collected from Ms. Lerner's computer has or will be produced or made available to Congressional investigators.

As part of the IRS production of materials related to the TIGTA report, email from Ms. Lerner's email box and hard drive previously had been processed using search terms. By mid-March 2014 almost 8,000 such emails from Ms. Lerner's computer and mailbox had been produced in unredacted form. Another 3,000 emails involving Ms. Lerner (as author or recipient) from other custodians also had been produced in unredacted form, for a total of approximately 11,000 produced emails involving Ms. Lerner and related to the TIGTA report.

Producing email regardless of relevance required reprocessing what had been collected from Ms. Lerner so that the email reviewed and produced was no longer limited by search terms or subject matter. As the IRS reviewed Ms. Lerner's email for production and prepared to produce to investigators the balance of 2009-2013 materials from Ms. Lerner's custodial email account (unlimited by subject matter or search terms), it determined that her custodial email (from her email box and hard drive) contains very few emails prior to April 2011, while the number of Ms. Lerner's custodial emails dated after April 2011 is more voluminous. In total, more than 43,000 Lerner custodial emails exist between January 1, 2009 and May 22, 2013, all of which have been or will be produced.

⁸ The IRS also assisted Ms. Lerner's attorneys in identifying and redacting Section 6103 information on documents located on her personal home computer and email on her personal email account.

⁹ On May 16, 2013, the IRS Chief Counsel's office sent Ms. Lerner (and many others) a litigation hold notice instructing her not to alter or destroy: "all communications, documents drafted or reviewed, spreadsheets created or reviewed, notes from meetings, notes relating to specific taxpayers and/or applications, information requests to applicants, training materials, or any other items that relate to the process by which selection criteria were used to identify tax-exempt applications for advocacy organizations for review, including but not limited to Be On the Look Out, from January 1, 2008 to the present." Litigation Hold, Attachment A.

¹⁰ TIGTA currently has custody of Ms. Lerner's computer.

Although the IRS is unable to interview Ms. Lerner to learn more, the IRS has determined that Ms. Lerner's computer crashed in mid-2011. See Attachment **Enclosure E**. At that time, the IRS Information Technology Division tried -- at Ms. Lerner's request - multiple processes to recover the information stored on her computer's hard drive. However, the data stored on her computer's hard drive was determined at the time to be "unrecoverable" by the IT professionals. Attachment **Enclosure F**. Any of Ms. Lerner's email that was only stored on that computer's hard drive would have been lost when the hard drive crashed and could not be recovered.

In order to produce as much email on which Ms. Lerner was an author or recipient as possible, the IRS:

- Retraced the collection process for Ms. Lerner's computer to determine that all materials available in May 2013 were collected;
- Located, processed, and included in its production email from an earlier 2011 data collection of Ms. Lerner's email;
- Confirmed that back-up tapes from 2011 no longer exist because they have been recycled (which not uncommon for large organizations in both the private and public sectors);
- Searched email from other custodians for material on which Ms. Lerner appears as an author or recipient, then produced such email.

As a result of these efforts, the IRS identified approximately 24,000 Lerner-related emails between January 1, 2009 and April 2011 in addition to those related to the processing and review of applications for tax-exempt status as described in the May 2013 TIGTA report, which have already been produced. All such emails have been or will be produced or made available to Congressional committees. In total, Congressional committees have received or will receive more than 67,000 emails in which Ms. Lerner was an author or a recipient.

IV. Conclusion

The Internal Revenue Service has never before undertaken a document production of this size and scope. Hundreds of employees spent thousands of hours locating, processing, reviewing, and redacting documents for the Congressional Committees and other investigators. Because of how the IRS maintains and stores its email, certain challenges were inherent in the process and we have addressed those challenges in as comprehensive a manner as possible.