STATEMENT OF JASON R. BARON, PROFESSOR OF THE PRACTICE,
UNIVERSITY OF MARYLAND, COLLEGE OF INFORMATION STUDIES

SUBMITTED TO THE U.S. SENATE COMMITTEE ON HOMELAND SECURITY AND
GOVERNMENT AFFAIRS

"CORRECTING THE PUBLIC RECORD: REFORMING FEDERAL AND PRESIDENTIAL
RECORDS MANAGEMENT"

March 15, 2022

Chairman Peters, Ranking Member Portman, and distinguished members of the Committee,
thank you for the opportunity to submit testimony regarding proposed amendments to the
Presidential Records Act (PRA) and the Federal Records Act (FRA). During a 33-year career in
federal service, I served as a trial attorney and senior counsel in the Department of Justice (DOJ),
and as the first appointed Director of Litigation at the National Archives and Records
Administration (NARA). While at DOJ, I acted as the lead lawyer in litigation involving the
preservation of White House email.[1] During my time both at DOJ and NARA, I have seen first-
hand the introduction of new communications technologies that have transformed presidential
and federal recordkeeping. I also became deeply involved in providing legal and policy guidance
to federal agencies throughout the Executive branch on various aspects of recordkeeping policies
and practices. Based on my experience in government, I believe that further amendments to the
PRA and FRA are needed both to take into account recent changes in technology, and to
generally improve the preservation of and access to presidential and federal records in electronic
form.

As a recent GAO report stated:

> Records are the foundation of open government, supporting the principles of
> transparency, participation, and collaboration. Effective management of federal agency
> records is important for efficient government operations: it ensures that sufficient
> documentation is created; that agencies can efficiently locate and retrieve records needed
> in the daily performance of their missions; and that records of historical significance are
> identified, preserved, and made available to the public.[2]

In November 2011, President Obama issued a presidential memorandum on "Managing
Government Records," that recognized the fact that

> [d]ecades of technological advances have transformed agency operations, creating
> challenges and opportunities for agency records management. Greater reliance on

---

[1] *Armstrong v. Executive Office of the President*, 810 F. Supp. 335 (D.D.C. 1993), *aff'd in relevant part*, 1 F.3d
1274 (D.C. Cir. 1993).

[2] Government Accountability Office, "Selected Agencies Need to Fully Address Federal Electronic Recordkeeping
Requirements," GAO-20-59 (Feb. 2020), at 4, https://www.gao.gov/assets/gao-20-59.pdf.

electronic communication and systems has radically increased the volume and diversity of information that agencies must manage. With proper planning, technology can make these records less burdensome to manage and easier to use and share. But if records management policies and practices are not updated for a digital age, the surge in information could overwhelm agency systems, leading to higher costs and lost records.[3]

The 2011 memorandum directed the Archivist to work with the Office of Management and Budget (OMB) in spearheading recordkeeping reforms aimed at recognizing and incorporating technological progress. To that end, the Archivist and the Acting Director of OMB jointly issued a "Managing Government Records Directive" (M-12-19),[4] instructing agencies (i) that after 2019, they were to manage their permanent records in electronic formats, for eventual transfer to NARA; and (ii) that after 2016, agencies were to manage electronic mail (e-mail) in electronic formats. This memorandum was in turn updated in 2019, again by a joint directive of the Archivist and OMB (M-19-21), entitled "Transition to Electronic Records."[5] The latter issuance extended the date until after December 31, 2022, for agencies to manage permanent records in an electronic format with appropriate metadata for eventual transfer to NARA. It also required agencies to manage all temporary records electronically to the fullest extent possible.

What have these memoranda meant for recordkeeping reform? First, NARA will no longer take in paper records created after 2022, meaning that over the coming decades NARA will face the need to preserve and provide access to literally billions of Executive branch records in electronic or digital formats. Second, agencies no longer have been able to continue outdated print-and-file methods for managing email, but instead must now provide for their electronic preservation. Third, agencies themselves are expected to manage *all* forms of electronic records electronically. These policy changes are to be applauded.

However, as salutary as these actions have been, the accumulation of electronic records that NARA and the government as a whole are experiencing, coupled with the need for agencies to incorporate new forms of communications technologies into their records management policies, together create unprecedented recordkeeping and access challenges. Failure to confront the gaps in current practices, combined with failure to seize the opportunity to use new technologies in support of recordkeeping, may have profound detrimental implications for the future of government accountability and transparency.

In what follows, I wish to address three of the specific legislative proposals under discussion.

(1) *Preserving Ephemeral Communications*

The accelerating growth in the use by government officials of ephemeral messaging poses a serious threat to traditional recordkeeping practices. Ephemeral messages

---

[3] *See* https://obamawhitehouse.archives.gov/the-press-office/2011/11/28/presidential-memorandum-managing-government-records.

[4] *See* https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2012/m-12-18.pdf.

[5] *See* https://www.archives.gov/files/records-mgmt/policy/m-19-21-transition-to-federal-records.pdf.

are intended to be short-lived, [with] the applications used to generate these communications . . . designed to enable automatic disposition or expiration of the messages. The specialized functionality of ephemeral messaging applications to delete these messages automatically or after a predefined duration (most often a very short time) also eliminates the message and (in some cases) the underlying metadata residing on the user's application *and* on the applications of those who either sent or received the messages in question.[6]

For example, the Confide website states that the app allows the user to "Discuss sensitive topics, brainstorm ideas or give unfiltered opinions without fear of the Internet's permanent, digital record and with no copies left behind."[7] That is because "Confide messages self-destruct. After they are read once, they are gone."[8] In similar fashion, Wickr's ephemeral messaging page titled "How private are my Wickr messages" states that Wickr "deletes all metadata from its communications and our Secure File Shredder cleans the RAM [its military application collaborative tool] after each message or picture is opened."[9] These forms of ephemeral messaging essentially contain a "self-destructive" feature that leads to the vanishing of those messages from history.

As the Committee is well aware, there have been numerous instances of attention-grabbing headlines involving the use of WhatsApp, Wickr, Confide, and Signal by White House staff,[10] by federal officials and employees,[11] as well as by State governors and other public officials.[12] A lawsuit by the public interest group Citizens for Responsibility and Ethics in Washington was filed, challenging White House recordkeeping practices over the use of Confide and Signal.[13] In 2017, Chairman Chaffetz of the House Oversight Committee sent a letter to 55 federal agencies

---

[6] The Sedona Conference, *Commentary on Ephemeral Messaging*, 22 SEDONA CONF. J. 435 (2021) (emphasis in original).

[7] Confide, "Your Confidential Messenger," https://getconfide.com.

[8] *Id.*

[9] *Wickr,* https://support.wickr.com/hc/en-us/articles/115005145108-How-private-are-my-Wickr-messages

[10] *See* A. Parker & P. Rucker, "Upheaval is now standard operating procedure inside the White House," *Wash. Post* (Feb. 13, 2017), http://wapo.st/2lI71cl ("Staffers, meanwhile, are so fearful of being accused of talking to the media that some have resorted to a secret chat app—Confide—that erases messages as soon as they're read."); N. Fandos, "Jared Kushner and Ivanka Trump Use Private Accounts for Official Business" (March 3, 2019), https://www.nytimes.com/2019/03/21/us/politics/jared-kushner-whatsapp.html; *see also* Letter from Elijah Cummings, Chairman, H. Comm. On Oversight and Reform to Pat A.Cipollone, Counsel to the President (March 21, 2019), https://apps.npr.org/documents/document.html?id=5777681-Cummings-Letter-To-White-House-On-Kushner.

[11] A. Restuccia, et al., Federal workers turn to encryption to thwart Trump, *Politico*, Feb. 2, 2017, http://politi.co/2km4Qrb.

[12] S. Thompson, "Maryland lawmakers target Governor Hogan's self-destructing messages,: *Wash. Post* (Feb. 11, 2022), https://www.washingtonpost.com/dc-md-va/2022/02/11/hogan-wickr-deleting-messages/; C. Farivar, Judge Should Order Governor to Stop Using Ephemeral App, Lawyers Say, *Arstechnica* (Feb. 1, 2018, 6:03 AM), https://arstechnica.com/tech-policy/2018/02/lawyers-governors-secret-messaging-app use-violates-public-records-laws/; Nicole Galloway, CPA, Missouri State Auditor, Office of Attorney General, Review of Whether State Resources Were Used for Political Purposes," Report No. 2020-006 (Feb. 2020) at 5, 16, 17 & 146, https://app.auditor.mo.gov/Repository/Press/2020006600987.pdf; C. Dil, "D.C. mayor's WhatsApp use spurs stronger public records law," https://www.axios.com/local/washington-dc/2022/03/01/dc-mayor-whatsapp-public-records.

[13] *Citizens for Responsibility and Ethics in Washington v. Trump*, 302 F.Supp.3d (D.D.C. 2018), *aff'd*, 924 F.3d 602 (D.C. Cir. 2019).

on the subject of WhatsApp, Signal and Confide, stating that their use "could result in the creation of federal records that would be unlikely or impossible to preserve" and that they might end up "circumventing requirements establishing by federal recordkeeping and transparency laws."[14]

Notwithstanding the alarms that have been raised, some government agencies are continuing to go out of their way to promote the use of specific ephemeral apps. For example, a recent Customs and Border Protection (CBP) Inspector General report found that agency personnel used WhatsApp to share and then delete the information they shared with groups of hundreds of U.S. and Mexican officials.[15] And senior officials of the State Department were reported as using WhatsApp to discuss Ukraine policy in the last administration.[16] This is not surprising given that in prior years State Department officials have been quoted as giving encouragement to the use of WhatsApp to conduct diplomacy generally.[17]

It should be clear that in 2022 the ubiquity of smart phones used by ordinary federal employees, coupled with the popularity of messaging apps, effectively has enabled anyone in the federal government to communicate about government business in ways that amount to an "end-run" around the use of e-mail.[18] Yet despite Congress' best efforts to address the subject of electronic messaging generally, the emergence of the use of ephemeral apps as a substantial channel of communications nevertheless remains ineffectively regulated under existing law.

In enacting the Presidential and Federal Records Act Amendments of 2014,[19] Congress included a provision in each Act titled "Disclosure requirement for official business conducted using non-official electronic messaging accounts." Section 2911 of Title 44 provides that federal agency officers and employees "may not create or send a record using a non-official electronic messaging account unless" they copy (cc) the message to an official ".gov" account, or forward a copy of the message to a ".gov" account within 20 days of creation or sending of the message. Section 2209 of Title 44 contains parallel language governing the conduct of White House officials with respect to presidential electronic messages sent from non-official accounts.

---

[14] Letter from Rep. Jason Chaffetz, Chairman & Rep. Elijah Cummings, Ranking Member, H. Comm. On Oversight & Gov't Reform, to Kathleen McGettigan, Acting Dir., Office of Personnel Mgmt. (Mar. 8, 2017), quoted in D. Stewart, "Killer Apps: Vanishing Messages, Encrypted Communications, and Challenges to Freedom of Information Laws When Public Officials 'Go Dark'," J. OF LAW, TECH. & THE INTERNET, 10:1 (2019), at 1-2.

[15] CBP Inspector General Report, OIG-21-62 (Sept. 20, 2021), at 29 & n.73, https://www.oig.dhs.gov/sites/default/files/assets/2021-09/OIG-21-62-Sep21.pdf (also noting that CBP officials "used WhatsApp to send and receive substantive messages that may be subject to recordkeeping requirements); *see also* Joseph Cox, "Customs and Border Protection to Use Encrypted App Wickr Widely," *Vice* (Sept. 28, 2021), https://www.vice.com/en/article/dyvmjm/customs-and-border-protection-cbp-wickr.

[16] T. Robinson, "Diplomats used WhatsApp personal phones to discuss Ukraine policy," SC Media (Oct.10, 2019), https://www.scmagazine.com/news/security-news/diplomats-used-whatsapp-personal-phones-to-discuss-ukraine-policy.

[17] A. Sandre, "WhatsApp for Diplomats," *Digital Diplomacy (*Aug. 13, 2018), https://medium.com/digital-diplomacy/whatsapp-for-diplomats-c594028042f1.

[18] *See, e.g.*, U.S. Environmental Protection Agency, Office of Inspector General Report No. 17-P-0062, "Congressionally Requested Audit: EPA Needs to Improve Processes for Preserving Text Messages as Federal Records" (Dec. 21, 2016), https://www.epa.gov/sites/default/files/2016-12/documents/_epaoig_20161221-17-p-0062.pdf.

[19] Pub. L. 113-187, 128 Stat. 2014 (Nov. 26, 2014).

There are two material weaknesses in the present language of sections 2911 and 2209. First, beyond the high-profile examples noted above, reporting shows that these provisions have largely been ignored with respect to various types of electronic messaging.[20] Second, these provisions have relied entirely on self-enforcement by individuals themselves, i.e., officials and employees are expected to take on the job of ensuring that every electronic message sent is accounted for through additional steps including either copying or forwarding of messages. Yet it is completely unrealistic to rely on individuals -- especially senior officials at the White House and at the top levels of the government -- to take the time to copy or forward every one of their government-business related messages. The growing use of "non-preservable" types of electronic messaging represented by WhatsApp, Confide, Signal, and the like, only heightens the risk that individuals will fail to comply with their recordkeeping obligations before those messages self-destruct.

Sections 2911 and 2209 do allow for disciplinary actions by an appropriate supervisor at the White House or at a federal agency. However, to my knowledge, few if any disciplinary actions for violations of these provisions have taken place in the seven years since their enactment. Given the importance of memorializing presidential communications, I support creating some kind of certification and reporting process being put in place to better ensure that all White House staff are aware of their recordkeeping obligations.

As a general matter, I am against wholesale prohibitions on the use of technology out of concern for cases of misuse. As NARA has stated, "[s]imply prohibiting the use of electronic messaging accounts to conduct agency business is difficult to enforce and does not acknowledge the way employees communicate."[21] In my experience, the vast majority of federal employees simply wish to use the latest means of communicating so as to be efficient in carrying out their official business -- not to evade their legal responsibilities. It must be recognized that electronic messaging in all its forms -- including on ephemeral apps -- has become near ubiquitous; these apps are just too easy and convenient to ignore. In my view, it would be a step backward to impose an across-the-board statutory ban on federal personnel using messaging apps for the transaction of government business.

There is, however, a way to control the use of such new technologies in a manner that promotes the recordkeeping laws and improves government accountability. Agencies should proceed to designate one (or two or three) specific ephemeral apps that can be used by staff for the conduct of official business. Those apps would be configured on both government-issued and personal phones so as to capture all communications sent or received on those apps, for retention consistent with existing recordkeeping requirements including for e-mail.[22] Use of other ephemeral messaging apps would be prohibited for government business.

---

[20] *See, e.g.,* "Gone in an Instant: How Instant Messaging Threatens The Freedom of Information Act," Americans for Prosperity Foundation (2020), https://edu.americansforprosperityfoundation.org/gone-in-an-instant/; *see also* EPA OIG Report, *supra* n.18.

[21] NARA Bulletin 2015-02, https://www.archives.gov/records-mgmt/bulletins/2015/2015-02.html.

[22] *See* Everlands Sutherland, "Enforcement appears as messages disappear: The perils of personal and ephemeral messaging" (Jan. 6, 2022) ("Enterprise versions may allow companies to customize features, such as security and data retention settings, for users within the organization and may assist companies in maintaining control over communications."), https://us.eversheds-sutherland.com/NewsCommentary/Legal-Alerts/247357/Enforcement-

One further note: I am aware that in 2021 Congress enacted the Electronic Message Preservation Act (EMPA), codified in section 2912 of Title 44.[23] Section 2912 directs the Archivist to issue regulations regarding agency preservation of electronic messages constituting records, which would "require the electronic capture, management, and preservation of such electronic records . . . ." When NARA issues the regulations contemplated under EMPA,[24] they will be expected to play a positive role in ensuring that electronic messages in .gov accounts are further preserved. Section 2912 does not, however, go further in providing NARA with a mandate to prohibit federal employees from creating non-preservable records; nor is it clear whether NARA regulations will require that electronic messaging be archived in a manner consistent with its existing Capstone policy for email, as discussed in the next section.

A recent Americans for Prosperity Foundation report states the matter well: "The mechanisms that capture and preserve records must keep pace with technology for the system to be an effective tool to enforce executive branch integrity."[25] Without further legislative fixes, the accelerating growth of ephemeral applications and other forms of electronic messaging used as alternative channels of communications to "traditional" e-mail threaten to remove important records memorializing government decision making, including both in the White House and throughout the Executive branch.

Based on the above, I support making modest revisions to the existing language of sections 2209 and 2914 of Title 44 to make clear that federal officials and employees be prohibited from using non-preservable electronic messaging, unless such messages are forwarded to an official (i.e., ".gov") message account. These changes will fill in a gap in current law and promote government accountability. As stated, I also support building in a greater oversight mechanism in the case of presidential electronic messages, in the form of an enhanced certification or reporting process on the use of ephemeral communications by White House staff.

(2) *Codifying and Expanding Capstone Approach to Recordkeeping*

I have long advocated automating records management to relieve federal employees of the burden of having to take additional steps on a per-message basis to accomplish proper recordkeeping.[26] As of 2014, NARA agreed:

> [I]n many agencies, employees manage their own email accounts and
> apply their own understanding of Federal records management. This means that
> all employees are required to review each message, identify its value, and either
> delete it or move it to a recordkeeping system . . . NARA recognizes that placing
> the responsibility on employees to make decisions on an email-by-email basis

---

appears-as-messages-disappear-The-perils-of-personal-and-ephemeral-messaging.

[23] Pub. L. 116-283, div. H, title XCVI, § 9602(b)(1), 134 Stat. 4828 (Jan 1, 2021).

[24] By its terms, EMPA required NARA to issue regulations one year after its enactment, with subsequent agency reports and reports to Congress to follow. *See id.*, §§ 9602(b)(2) & (3), 134 Stat. 4828 (44 U.S.C. 2912 note).

[25] Gone in An Instant at 19, *supra* n.20.

[26] Acceptance Remarks, 2011 Emmett Leahy Award Presentation, https://emmettleahyaward.org/2011-jason-r-baron.

can create a tremendous burden.[27]

The "Capstone Approach" or "Capstone policy" for managing email was developed in response to a growing recognition that alternative strategies relying on human involvement were not working.[28] A federal agency that chooses to adopt Capstone designates those senior officials within the agency whose e-mail accounts will be deemed "permanent records" of the federal government, for eventual transfer to NARA. Those agency designations are then approved by NARA.[29] As dictated by General Records Schedule 6.1,[30] full implementation of Capstone means that senior officials' emails will automatically be archived for permanent preservation, and all other employees within a Capstone agency or agency component will have their program-related emails archived for seven years.

The Capstone Approach can be improved and expanded upon through legislative action. First, as it is currently formulated, adoption of the policy remains voluntary. Thus, there is no guarantee that agencies will continue to adhere to a policy of permanent preservation of the email records of their most senior level officials. Refraining from further continuing under Capstone would, however, be an unfortunate blow to government transparency, as Cabinet level officials, their principal deputies, and a wide variety of other agency officials create and receive emails with attached policy memoranda that provide documentation of the most important agency business in each administration.

Second, not every agency has chosen to adopt Capstone as a means for managing their email. To date, over 200 agencies and components of agencies have submitted forms to NARA indicating the senior officials being designated as Capstone account holders.[31] In the case of other agencies, while presumably they are managing electronic records pursuant to the 2012 and 2019 Directives, it isn't clear to what extent those agencies are relying on senior officials themselves to preserve their own records -- a failed policy that Capstone was intended to supersede.

For these reasons, codification of the existing voluntary Capstone Approach to managing email is desirable.

Additionally, in conjunction with the proposed amendments on electronic messaging, the Capstone policy affords a straightforward means of capturing those messages in a way that enhances government accountability. As stated above, agencies should designate specific electronic messaging apps to be used for government business. In the case of Capstone account holders (i.e., designated senior government officials), the messages they send over officially approved channels would be captured and archived for permanent preservation as part of their Capstone account.

---

[27] NARA Bulletin 2014-06, "Guidance on Managing Email," https://www.archives.gov/records-mgmt/bulletins/2014/2014-06.html.
[28] NARA, "White Paper on The Capstone Approach and Capstone GRS" (April 2015), https://www.archives.gov/files/records-mgmt/email-management/final-capstone-white-paper.pdf.
[29] NARA, "Capstone Forms," https://www.archives.gov/records-mgmt/rcs/schedules/capstone-forms
[30] NARA, "General Records Schedule 6.1: Email Managed under a Capstone Approach," https://www.archives.gov/files/records-mgmt/grs/grs06-1.pdf.
[31] NARA, "Capstone Forms," *supra* n.29.

My understanding is that NARA is open to the idea of issuing new guidance about expanding Capstone policies to include electronic messaging, and that one or more agencies may even be in the process of implementing such an expansion. However, for the same reasons as given above, I believe a statutory provision that locks in the Capstone Approach to include all forms of electronic messaging as used by senior officials will best ensure that important records across the entirety of government are permanently preserved, in furtherance of Congressional oversight, government accountability, and public access now and for future generations.

(3) *Improving Search Automation In Government*

   a) *Using Machine Learning To Perform Better Searches*

The preceding discussion has been largely focused on shoring up gaps in current recordkeeping practices, especially with respect to electronic messaging. Paradoxically, at the same time there are pressing challenges the government faces due to the *abundance* of electronic records being accumulated. Two of the most important challenges are:

   --first, a growing inability to search through electronic archives on a timely basis, including in response to Congressional requests, litigation demands, and citizen access under the Freedom of Information Act (FOIA). This includes the need to efficiently isolate exempt material from public access; and

   --second, a failure to categorize and dispose of unstructured information across agency networks and shared drives, in accordance with legacy agency records schedules.

I believe an opportunity exists for Congress to greatly help modernize federal recordkeeping by making federal agencies aware of the latest software, services and expertise offered by industry and academia in the fields of artificial intelligence, machine learning and data analytics. Technology continues to move fast in these areas, and it is worth having an advisory body of experts explore how such cutting-edge tools can help solve the recordkeeping and access challenges the federal government faces.

Perhaps the most looming challenge involving government records is the need to use machine learning to search for responsive records on a timely basis.

Consider this: as a result of the *Armstrong* case,[32] first through backup tape restoration, and then after 1994 through email archiving, White House presidential emails from the Reagan Administration through the Trump Administration have been preserved and are in the legal custody of NARA. This collection spanning all administrations is estimated at this point to consist of around 600 million e-mails, comprising on the order of three billion pages.[33] Of this vast amount of records, to date less than 1/10 of a percent have been made accessible online.[34]

---

[32] *See supra* n.1.
[33] Estimates are based on conversations with NARA staff.
[34] This estimate is based on personal knowledge and a review of NARA's websites. Emails that are publicly available include those released in response to various litigation and FOIA requests, as well as Congressional requests for the records of John Roberts, Elana Kagan, and Brett Kavanaugh in connection with their nomination

Given the enormous size of these collections, it is not surprising that NARA had difficulty searching records in time to meet a Congressional deadline to turn over all responsive documents in connection with the Kavanaugh nomination.[35]

Presidential emails at NARA are, of course, not the only enormously large and growing corpus of electronic records that need to be searched for Congressional, litigation-related, FOIA, and other access requests. As discussed above, the Archivist's call for federal agencies to transition to electronic recordkeeping by the end of 2022, coupled with what is already widespread adoption of the Capstone Approach for email preservation, all are resulting in a tidal wave of electronic records being stored within agency repositories. These repositories will grow into the tens and hundreds of millions, if not billions of records, during the next decade -- especially at the larger Cabinet departments. In my view, the time has come for federal records managers, FOIA officers, and other stakeholders throughout the Executive branch to actively consider adoption of machine learning-based search methods that are widely in use in the private sector.

It has been well known to the legal community for the past 15 years that manual or keyword searching of large amounts of electronically stored information is extremely time consuming and fails to achieve accurate results, both in terms of producing a huge number of false positive "hits" that need to be reviewed, as well as missing responsive documents to requests (false negatives).[36] Based on what was then cutting-edge research,[37] starting around 2012 courts gave their blessing to the use of machine learning methods in e-discovery -- as an alternative to keyword searching -- for the purpose of parties' finding responsive records in response to document requests.[38]

Machine learning (ML) "refers to a software programming technique that uses algorithms to autonomously improve decisions through analysis. The algorithms use statistical methods that enable machines to improve correlations as more data is used. This facilitates the machine's ability to automatically discover patterns in data which can be used to make predictions. The algorithms generally perform better as the volume of data available to analyze increases."[39]

---

hearings to the Supreme Court.

[35] Letter from Gary M. Stern to Sen. Charles Grassley, dated August 2, 2018, https://www.archives.gov/files/foia/stern-letter-to-grassley-8-2-2018.pdf (stating that NARA could not meet the deadline for turning over all responsive documents in time for the hearing in light of the need to search through 900,000 emails).

[36] *Victor Stanley, Inc. v. Creative Pipe, Inc.,* 250 F.R.D. 260, 262 (D. Md. 2008) ("there are well-known limitations and risks associated with" keywords); "The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval in E-Discovery," 15 SEDONA CONF. J. 217 (2014) (J.R. Baron & M.Grossman, eds. in chief), https://thesedonaconference.org/sites/default/files/publications/217-264%20Search%20and%20Information_0.pdf

[37] See J.R. Baron, D. Lewis, D. Oard, "TREC-2006 Legal Track Overview," https://trec.nist.gov/pubs/trec15/papers/LEGAL06.OVERVIEW.pdf; M. Grossman & G. Cormack, "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review," 17 RICHMOND J. LAW & TECHNOLOGY art. 11 (2011), http://jolt.richmond.edu/v17i3/article11.pdf.

[38] *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012) (Peck, Mag. J.), *adopted sub nom. Moore v. Publicis Groupe SA*, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).

[39] NARA, "Cognitive Technologies White Paper: Records Management Implications for Internet of Things, Robotic Process Automation, Machine Learning, and Artificial Intelligence" (Oct. 2020) at 11, https://www.archives.gov/files/records-mgmt/policy/nara-cognitive-technologies-whitepaper.pdf.

Lawyers refer to these machine learning methods under various names, including "predictive coding" and "technology assisted review" (TAR).[40] Using TAR methods, lawyers in complex litigation are able to train algorithms to find responsive records in large universes of data much more efficiently and more accurately than using older methods. This is done by training the algorithm to recognize documents containing similar word patterns throughout the document.[41]

While there are legal offices across a number of federal agencies that have access to TAR-related software for purposes of meeting litigation demands, to date few if any personnel in records management or FOIA offices have ever employed this type of technology. As the 2018-2020 FOIA Advisory Committee report made clear

> The 2019 CFO [Chief FOIA Officer] Reports show that federal agency FOIA staff do not appear to be well-versed in how AI and machine-learning technologies may improve the efficiency of FOIA searching in ever-growing digital repositories. This software has not been generally deployed in the context of FOIA searches, nor has it been developed with an eye toward the types of record content with a range of sensitivities (e.g., personally identifiable information) found within components of the federal government.[42]

Based on my own recent experience over the past year in filing FOIA requests to over 25 agencies, I can also personally attest that agencies perform searches against their growing Capstone email repositories solely using keywords. So far as I am aware, no agency employs machine learning software to search for responsive records, or to ferret out exempt material within those records for purposes of withholding.[43] With respect to the latter point, recent research I have been involved in holds the promise that machine learning methods may be used to isolate FOIA exempt material in a way that will expedite human review.[44] This type of research should be promoted.

In sum, Executive branch agencies would do well to understand the power of machine learning search tools, which hold the potential to expedite searches while making them less labor-intensive. In 2020, the FOIA Advisory Committee to the Archivist made the following recommendation:

> The Archivist should work with other governmental components and industry in promoting research into using artificial intelligence, including machine learning

---

[40] Thomson Reuters, "How to make the e-discovery process more efficient with predictive coding," https://legal.thomsonreuters.com/en/insights/articles/how-predictive-coding-makes-e-discovery-more-efficient

[41] M. Grossman & G. Cormack, "A Tour of Technology-Assisted Review," chap. 3 in J.R. Baron, R. Losey & M. Berman, PERSPECTIVES ON PREDICTIVE CODING AND OTHER ADVANCED SEARCH METHODS FOR THE LEGAL PRACTITIONER (ABA 2016), draft version of chap. 3 available at https://cormack.uwaterloo.ca/tour/tour.pdf.

[42] 2018-2020 FOIA Advisory Committee Report at 35, https://www.archives.gov/files/ogis/assets/foiaac-final-report-and-recs-2020-07-09.pdf.

[43] *See id.* at 22 (finding "no express mention by any agency [in their Chief FOIA Officer Reports] regarding the use of 'predictive coding' or 'technology assisted review' in conducting FOIA searches").

[44] J.R. Baron, M. Sayed, D. Oard, "Providing More Efficient Access to Government Records: A Use Case Involving Application of Machine Learning to Improve FOIA Review for the Deliberative Process Privilege," 15 JOURNAL ON COMPUTING AND CULTURAL HERITAGE, vol.1, art. 5: 1-19 (2022), https://dl.acm.org/doi/abs/10.1145/3481045 (pre-print at https://arxiv.org/abs/2011.07203)

technologies, to (i) improve the ability to search through government electronic record repositories for responsive records to FOIA requests and (ii) identify sensitive material for potential segregation in government records, including but not limited to material otherwise within the scope of existing FOIA exemptions and exclusions.[45]

It is long past time to tap the expertise of private industry and academia in improving the capabilities of agencies to perform better searches in response to Congressional and public requests.

b) *Automating the Categorization of Records*

"In order to effectively address NARA regulations, agencies are to establish policies and procedures that provide for appropriate retention and disposition of their electronic records." A key part of retention is through the use of federal agency record schedules.[46] These differentiate records into permanent and temporary categories of records. In turn, "[d]isposition involves transferring records of permanent, historical value to NARA for the archiving of records (preservation) and the destruction of all other records that are no longer needed for agency operations."[47]

One longstanding challenge in the area of records management is the widespread existence of legacy records schedules from as far back as the 1980s, that are still being used by agencies despite the fact that "they do not reflect current business practices."[48] There are thousands and thousands of records schedules used government-wide.[49] Many of these contain a large number of records series, each differing in their assigned retention periods. This is out of sync with what NARA and private industry consider to be best practices in records management (including with respect to flexible scheduling of records into "bigger buckets").[50] As noted by NARA's Inspector General in 2019, NARA provides guidance to agencies on the subject of updating records schedules periodically, but the agency has declined to issue regulations specifying agency time frames for doing so.[51]

The result of a failure to update records schedules is not simply one of updating forms. As an Office of Personnel Management (OPM) Inspector General Report found, outdated records schedules lead to any number of undesirable outcomes, including increased storage costs,

---

[45] 2018-2020 FOIA Advisory Committee Report at 35, *supra* n.42.

[46] 44 U.S.C. § 3303a; 36 C.F.R. § 1225.12.

[47] GAO 20-59, at 7-8*, supra* n.2.

[48] NARA Inspector General Audit of NARA's Oversight of Electronic Records Management in the Federal Government (OIG Audit Report No. 19-AUD-10) (June 11, 2019), at 24, https://www.oversight.gov/sites/default/files/oig-reports/audit-report-19-aud-10.pdf.

[49] *See* J.R. Baron, "The PROFS Decade: NARA, E-mail, and the Courts," chap. 6 in B. Ambacher, ed., THIRTY YEARS OF ELECTRONIC RECORDS (Scarecrow Press 2003), at 118 (citing to 17,000 records schedules in the Department of Defense alone).

[50] W. Saffady, "Big Bucket Retention: Objectives, Issues Outcomes," *ARMA Magazine* (Dec. 7, 2018), https://magazine.arma.org/2018/12/big-bucket-retention-objectives-issues-outcomes/; NARA, "Flexible Scheduling" (Jan. 2004), https://www.archives.gov/records-mgmt/initiatives/flexible-scheduling.html.

[51] OIG Audit Report No. 19-AUD-10 at 24, *supra* n.48. According to this report, NARA committed to notifying agencies that they were required to review schedules approved prior to 1990, but it is unclear how many agencies have taken such actions. *Id.*

decreased security where agencies keep records containing personally identifiable information (PII) longer than required, and increased FOIA response times due to the retention of unnecessary records.[52]

These problems are greatly compounded where agencies fail to enforce the specified retention periods in existing records schedules, regardless of the schedules' age. This is understandable given the volumes of electronic records that must be accounted for, and their numerous locations including on agency network and shared drives. In particular, most information on shared drives is unstructured (e.g., textual documents in Word or PDF, spreadsheets, etc.), and has been informally classified at the whim of individual employees through idiosyncratic naming conventions. But only a limited number of federal agencies have automated the disposition of electronic records on shared drives or elsewhere in accordance with their records schedules.

Although the above material weaknesses involving records schedules can be partially addressed through greater human involvement, consideration should be given to using machine learning and advanced analytical techniques in pursuit of modernizing and making electronic recordkeeping more efficient. The private sector is well-versed in using these technologies to automatically categorize records during their entire lifecycle, from creation through disposition. Using these tools, records can be initially classified based on document type and content, consistent with records schedule categories. Such tools also provide functionality for identification of sensitive or exempt material within records (e.g., PII), determining which records are subject to legal holds, and executing defensible disposition on an automatic basis in accordance with records schedules.

In Australia, at least one agency has adopted a machine learning and advanced analytics approach to records classification. According to one report, the Australian Human Rights Commission

> created a statistical model that can classify records against . . . the Commission's agency-specific records disposal authority. . . . The statistical model is developed by taking a set of records that have been manually classified and applying Natural Language Processing techniques to normalize the document content into vectors. The model is then trained using algorithms. After an initial training period, the . . . statistical model can categorize individual records with an accuracy of 80%. The Commission expects this accuracy will increase over time. [The model] also re-categorizes records each time they are edited, ensuring the classification is always current.[53]

NARA is aware of the potential of machine learning being used to further automate categorization and disposition of records. In a 2020 White Paper, NARA stated that it and records professionals should consider "[l]everaging AI and ML to identify records eligible for disposition and automating their destruction or transfer into NARA's Electronic Records

---

[52] Inspector General Final Evaluation Report of the U.S. Office of Personnel Management's Preservation of Electronic Records, Report No. 4K-CI-00-18-009 (Dec. 21, 2018) at 5, https://www.opm.gov/our-inspector-general/publications/reports/2018/evaluation-of-the-us-office-of-personnel-managements-preservation-of-electronic-records.pdf.

[53] IDM Information & Data Manager, "Australian Human Rights Commission adopts machine learning" (June 7, 2019), https://idm.net.au/article/0012509-australian-human-rights-commission-adopts-machine-learning.

Archives (ERA)."[54] However, so far as I am aware, there has only been limited engagement by federal agencies in working towards putting in place automated methods that will satisfy federal recordkeeping requirements. That is why use cases regarding the applications of these techniques are deserving of further serious study and analysis.

In recent years the use of artificial intelligence in government operations and services has received high-level attention, including through the issuance of an "Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government."[55] Consistent with the aspirational goal of using forms of AI and advanced analytics across all government operations, I strongly support the convening of an advisory committee with the expertise to advise the Executive branch on what constitute state-of-the-art machine learning and advanced analytics tools to help address the profound recordkeeping issues that the government faces, especially due to the astronomical numbers of electronic records now being created.

Additional Thoughts

I wish to congratulate the outgoing Archivist of the United States, David Ferriero, who deserves enormous credit for his role in advancing the cause of electronic records management. I would also be remiss in failing to point out that NARA provides a great amount of guidance to agencies on issues relating to electronic records management, including in recent years setting up providing best practice frameworks for agencies to follow.[56] NARA is well aware of the challenges I have addressed in my remarks here today, including a vision of the government using various forms of artificial intelligence in the future.[57] I was proud to work for 13 years at NARA, and I remain close to many of my former colleagues. My testimony today is not in any way intended to be critical of NARA's heroic efforts to reform electronic records management.

Congress can, however, play an important role in enacting recordkeeping reform measures that take into account new technologies and the evolving nature of government business. I believe that the modifications to the PRA and FRA as outlined here are a needed step to address the looming recordkeeping challenges that the government faces over the remainder of this decade.

Thank you for the opportunity to address the Committee and I look forward to answering your questions.

---

[54] NARA Cognitive Technologies White Paper at 21, *supra* n.39.

[55] EO 13960 (Dec. 3, 2020), https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal*; see also* "Executive Order on Maintaining American Leadership in Artificial Intelligence" (EO 13859), dated Feb. 11, 2019, https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence (calling for AI efforts within government to be coordinated).

[56] *See, e.g.,* the Federal Electronic Records Modernization Initiative (FERMI), https://www.archives.gov/records-mgmt/policy/fermi.

[57] NARA Strategic Plan 2018-2022, § 1.1 ("Explore cutting-edge technologies such as advanced search to automate processing of large volumes of electronic records"), https://www.archives.gov/about/plans-reports/strategic-plan/strategic-plan-2018-2022; *see also* A. Boyd, "National Archives Wants to Use AI to Improve 'Unsophisticated Search' and Create 'Self-Describing Records'," Nextgov (Apr. 16, 2021), https://www.nextgov.com/analytics-data/2021/04/national-archives-wants-use-ai-improve-unsophisticated-search-and-create-self-describing-records/173417 (AI methods to assist searches of publicly available online records).