THE HONORABLE HUGH NATHANIAL HALPERN
Director
United States Government Publishing Office

Testimony of Government Publishing Office Director Hugh N. Halpern before the Committee on House Administration Hearing on *"Big Data: Privacy Risks and Needed Reforms in the Public and Private Sectors"*

February 16, 2022
2:00 P.M.

U.S. GOVERNMENT PUBLISHING OFFICE
732 North Capitol Street, NW | Washington, DC 20401-0001
www.gpo.gov I facebook.com/usgpo I twitter.com/usgpo I instagram.com/usgpo

GPO

Thank you Chairperson Lofgren and Ranking Member Davis for inviting me to testify today on behalf of the more than 1,500 craftspeople and professionals who make up the Government Publishing Office to discuss our approach to the critically important issue of protecting personally identifiable information (PII) and safeguarding individual privacy. GPO is entrusted with PII belonging to our teammates, customers, and, by nature of our business, the general public. We devote considerable time, attention, and resources to securing and protecting that information, and we are always interested to share our experiences and learn from the experience of others.

Robust protection of PII is critical to building trust with our customers and stakeholders. Without that trust, we can never achieve our vision of an *America Informed*.

GPO's Approach to PII Protection

GPO's approach to protecting PII is laid out in our Privacy Program, which is overseen by our Privacy Officer, who reports to our Chief Information Officer. The Privacy Program establishes a framework for the protection of PII from unauthorized use, access, disclosure, or sharing as well as the protection of related information systems from unauthorized access, modification, disruption, or destruction.

Originally established in 2010, the Privacy Program was most recently updated in 2021 through GPO Directive 825.41B. GPO's privacy program applies to our teammates and contractors alike. It rests on a few fundamental principles:

- First, the only people who may access PII held by GPO are authorized Agency staff and contractors trained in the protocols required to protect that information;
- Second, each business unit within GPO must have a designated privacy point of contact that reports directly to the business unit leadership;
- Third, it is the affirmative obligation of any GPO employee or contractor who suspects a breach of PII security to promptly report the concern; and
- Fourth and finally, the directive provides that failure to comply with agency PII protection standards and procedures is grounds for corrective actions up to and including termination for employees, debarment for contractors, or even criminal prosecution if appropriate.

As part of his duties in administering the Privacy Program, GPO's Privacy Officer coordinates two significant processes —GPO's ongoing PII management and protection and responding appropriately to a possible PII security breach.

GPO's Privacy Officer proactively requires the periodic performance of Privacy Threshold Assessments (PTA) and Privacy Impact Assessments (PIA) by GPO business units. These assessments evaluate how those business units maintain, secure, and utilize PII in their daily operations.

The second process relates to how GPO responds to a potential breach of systems containing PII, and is referred to as the Privacy Incident Response Team (PIRT) Framework. Whenever a GPO employee or contractor suspects a breach in PII security has occurred that individual is required to report it through a Privacy Incident Report shared with his or her supervisor. The Business Unit Manager and Privacy Point of Contact are then responsible for notifying GPO's Privacy Officer, who convenes the Privacy Incident Response Team to assess the scope of the breach and determine the scope of the Agency's response, up to and including the possibility of alerting the United States Computer Emergency Readiness Team (US-CERT) within the Department of Homeland Security.

In developing and executing these policies and procedures that guide its Privacy Program, GPO has sought to follow the recommendations laid out by the National Institute of Standards and Technology (NIST) in their "Guide to Protecting the Confidentiality of Personally Identifiable Information (Special Publication 800-22)," which identifies six imperatives for Federal agencies:

- Identifying all the PII residing in their environment;
- Minimizing the use, collection, and retention of PII to what is strictly necessary to accomplish their business purpose and mission;
- Categorizing PII by confidentiality impact level;
- Applying the appropriate safeguards for PII based on the confidentiality impact levels;
- Developing an incident response plan to handle breaches involving PII; and
- Encouraging close coordination among their chief privacy officers, chief information officers, chief information security officers, and legal counsel when addressing issues related to PII.

I believe GPO's Privacy Program is responsive to each of them, and today I want to highlight two of them that have particular salience to GPO's administration of its **govinfo** trusted digital repository and GPO's overall system of providing online access to Government information. Those recommendations relate to the categorization of PII by confidentiality impact levels and the application of safeguards based on those confidentiality impact levels.

Changes in technology have driven everyone's concern about the ease with which personal information is available. On the one hand, there is a lower concern about a stray piece of PII in a printed document in a library because it is comparatively more difficult to access. Similarly, using that information in a nefarious way is also more difficult without our modern technology infrastructure. On the other, because we have made this information more accessible, the threat from the use of that information has also exponentially increased.

For example, in preparation for this hearing one of my colleagues remarked how as late as the 1990s his college used students' social security numbers as identifiers on their student ID cards, something that was completely unremarkable at the time. Another colleague remarked how a relative found her social security number because it appeared on a digitized Department of Defense document. Things that were at one time innocent practices with little impact are today serious breaches of protocol with potential wide-ranging impacts.

For GPO this amounts to a dual challenge — keeping current with the most effective strategies to protect PII in our current operations while also devising flexible solutions to PII disclosure challenges caused by the fact sensitive PII simply wasn't protected during the initial publication of tens of thousands of Government publications already in circulation.

To do this effectively without compromising the American people's access to important Government information, GPO differentiates among the various types of PII potentially housed in the collections that we either provide access to or host.

Using the NIST report and related OMB guidance to inform our approach to PII categorization, GPO considered the potential harm of disclosure to determine which forms of PII should be considered high-impact PII versus those that should be categorized as low-impact PII.

By differentiating between these two categories of PII, GPO believes it improves its ability to restrict or suppress that identifying information which is most likely to cause specific individuals harm upon disclosure without compromising the utility of the underlying public documents. GPO's definition of high-impact PII includes personal identification numbers issued by government or financial institutions printed in conjunction with an individual's name, such as full or partial social security numbers, taxpayer identification numbers, patient identification numbers, financial account numbers, or credit card numbers. In our view the unauthorized disclosure of these types of information are likely to pose the greatest risk to any individual.

Low-impact PII in our view would include a person's name, their street address, phone number, names of family members, and photographic images. Attempting to redact every such instance in which low-impact PII could be found in Government publications we believe would present a far greater challenge, with less obvious benefit, and considerably greater impact on the usability of the affected publications.

Protecting PII in GPO's Daily Operations

GPO's operations consist of two basic activities: producing Government information and identification documents and making Government information widely available. The Agency needs to protect any PII in both of those modes.

On the production side, there is no better example than what we do with PII than our work on secure identification documents. GPO's Security and Intelligent Documents operations involve significant amounts of PII and we take extraordinary steps to protect it every day. The PII data that comes into GPO is encrypted and once it is in our servers it is then decrypted and sent to the production machines. For instance, GPO manufactures and personalizes several varieties of identification cards. Specifically, the trusted traveler "smart" cards are used at both the Northern and Southern border crossings for identification and for expedited processing. By the very nature of these products, GPO must handle vast amounts of highly sensitive PII within our systems. GPO works closely with its customers and their contractors to maintain a series of firewalls that ensure that GPO only receives encrypted PII that is only decrypted when it is needed to produce and distribute those cards. When GPO finishes production of the product, that PII is scrubbed from our systems.

Much of our work to protect PII occurs as we work to make Government information more available. Since the passage of the GPO Electronic Information Access Enhancement Act of 1993, GPO has been working to broaden the availability of Government information on the Internet. What began nearly 30 years ago with the *GPO Access* website posting the daily Congressional Record and little more has now grown into the world's only ISO-certified trusted digital repository giving people worldwide free public online access to authenticated Government content. The volume of documents available online has grown to nearly 2.2 million packages, with over 320,000 packages added in FY 2021 alone. Last month we celebrated the 9 billionth retrieval of content through GPO's system of online access since the original *GPO Access* site first went live in 1994.

Today, 97 percent of all Federal Government publications are "born" digital and the growth in the electronic accessibility of Government information is significant enough that I recently established the Task Force on a Digital Federal Depository Library Program, comprised of 23 members representing the Depository Library Council, the Depository Library Community, Federal Agency partners, and library associations, to study what until recently might have been unthinkable: the feasibility and advisability of moving toward an all-digital Federal Depository Library Program.

Recognizing the importance of protecting PII that may have inadvertently found its way into our publicly accessible systems, the Superintendent of Documents adopted a policy requiring the redaction of high-impact PII whenever it is discovered within the publicly accessible files in GPO's system of online access, or within content being prepared for ingest into that system. To fulfill legal requirements of GPO's affiliate agreement with the National Archives and Records Administration, we must maintain unredacted preservation copies of those materials, but the PII can be redacted from the versions we put online.

In addition, because GPO's Partnership program provides a guide to materials that are beyond GPO's direct custodial control, GPO also takes proactive steps to alert those entities with which GPO has partnership agreements whenever GPO learns that materials in those partner's collections might reveal high-impact PII. This enables GPO's content partners to take appropriate action to redact or obscure PII that they may not have been aware existed.

In 2008, the Joint Committee on Printing authorized GPO to redact PII in the electronic access copies of congressional publications. For congressional publications published today, GPO can act immediately to suppress protected information in access copies of those publications without altering the underlying content. Where GPO's task becomes more challenging is when we are engaged in efforts to digitize historical collections of documents.

The most prominent example of this may be GPO's efforts in partnership with the Library of Congress to digitize every volume of the Congressional Record back to its inception in 1873 and make them available through **govinfo**, a project that GPO completed in 2019. The Library of Congress digitized millions of pages that contained thousands of instances of PII — particularly the entire or partial social security numbers of thousands of individuals. Many of these occurrences were related to the promotion of career staff, particularly in the Armed Forces.

While we haven't caught every single instance of PII appearing in the electronic copies of congressional publications prior to making them available to the public in electronic form, we have the infrastructure in place to systematically discover and remediate PII on an ongoing basis.

GPO maintains a contract with a vendor to provide high-impact PII redaction services that covers both historic, digitally-imaged content, as well as content already within the **govinfo** system on an as-discovered basis. Specifically, this contract covers both historic and contemporary congressional publications including the Serial Set, Digitized Hearings, the Daily Congressional Record, Bound Congressional Record, and any future digitization work that has the potential to contain PII. We also have access to software that enables us to manually redact such high-impact PII whenever we encounter it.

To date, our PII redaction contract has allowed us to find and redact 1,598 documents containing instances of PII both in publications currently found on **govinfo** as well as those we are in the process of digitizing. With significant anticipated digitization efforts planned in the future — for example the ongoing multi-year collaborative project with the Law Library of Congress to digitize the U.S. Congressional Serial Set and the possible digitization of thousands of congressionally mandated reports identified in the Congressionally Mandated Reports Act currently being considered in Congress — we expect to make considerable additional PII redaction investments in the years ahead.

Congress should be aware, however, that even if we are immensely successful in redacting such high-impact PII from those Government information collections we are digitizing ourselves, those we make available directly via **govinfo**, and those we provide access to via partnership agreements with other institutions, it is unlikely that such sensitive PII information will vanish completely from the Internet. This is because other well-intentioned institutions have undertaken efforts to make digitized copies of such Government publications electronically available in the past, and still others may choose to do so in the future.

To date, a few of the institutions who have done some of this work have demonstrated a willingness to remove high-impact PII from their digitized collections when we have made them aware of the issue, and Google Books and HathiTrust are worthy of recognition on that front. Still others lack either the resources to find and redact PII in their collections or the interest in doing so, and there's no legal authority for GPO to compel those third parties to take that action.

Additionally, it is important to remember that much of this digitization is based on existing tangible documents and those documents are distributed across the Nation, even if they aren't always easy to locate. By way of illustration, through the early 1990s GPO produced and disseminated nearly 20,000 copies of the Congressional Record each day — more than 10 times our current production level — so we are certain that there will always be some tangible copies of Government documents that contain PII that we will be unable to locate. The sheer number of tangible copies of older Government publications that contain sensitive, high-impact PII is going to continue to pose challenges for all of us for years to come.

Beyond the issue of the exposure of high-impact PII, there's the more general issue of potentially embarrassing personal information about specific individuals becoming easier to find because of the increasing electronic accessibility of Government publications. It has been my experience that people often try to conflate information which is merely unflattering with PII in order to hide that information from public view. It's important that we maintain the distinction between the two.

We can all agree that it is in the public interest to ensure that U.S. Federal Court opinions are accessible via the United States Courts Opinions (USCOURTS) collection on **govinfo**. This is a collaborative effort between GPO and the Administrative Office of the United States Courts to provide public access to opinions from select United States appellate, district, and bankruptcy courts, consistent with the E-Government Act's requirement for the substance of all written opinions, issued after April 16, 2005, to be made available in a text searchable format [Section 205 of Pub. L. No. 107-347].

However, on occasion individuals referenced in those opinions prefer that their involvement with those proceedings remain unknown and demand that GPO remove the relevant opinion from the collection. It is critical to understand that this data belongs to our customers and GPO has no independent authority to redact or remove an opinion, and per the Superintendent's policy with regard to online redactions or removal from collections, GPO will only take action in response to a request from the originating agency, the courts that issued the opinions.

When these inquiries arise, GPO refers individuals to the Administrative Office of the U.S. Courts for information about applying to the court that issued the opinion, which may decide to seal the opinion if the judge determines it warrants withdrawal from public access.

That concludes my testimony. I hope that this discussion of our practices, procedures, and capabilities was helpful to the Committee and I welcome your questions.

GPO

Hugh Nathanial Halpern,

GPO Director

Hugh Nathanial Halpern is the U.S. Government Publishing Office (GPO) Director, the agency's chief executive officer. The agency is responsible for publishing and printing information for the three branches of the Federal Government. Halpern is the 28th person to lead GPO since the agency opened its doors for business on March 4, 1861, the same day Abraham Lincoln was inaugurated as the 16th President of the United States. President Donald Trump nominated Halpern to be GPO Director on October 17, 2019, and the U.S. Senate confirmed him on December 4, 2019.

Biography

Prior to coming to GPO, Halpern held a succession of leadership positions during his 30 years on Capitol Hill. He served as the Director of Floor Operations for the Speaker of the U.S. House of Representatives. In that role, Halpern was the highest-ranking floor staffer in the House and served as Speaker Paul Ryan's Chief Advisor on all procedural matters. He managed the daily floor operations of the House, served as the liaison to all leadership offices, and oversaw legislative interactions between The White House, House and Senate. In 2018, he received the John W. McCormack Award of Excellence, the highest award given to a staff member in the House. The award recognizes a lifetime of bipartisan service to the House.

In addition to his position in the Speaker's Office, Halpern has more than a decade of experience serving on the senior leadership staff. He has a proven track record of successfully leading teams to achieve results.

During his career, he served half a dozen different committees in both policy development and procedural roles. During his 11 years on the House Committee on Rules, Halpern served as Staff Director leading the management and terms of debate on the House floor. In 2001, he was named General Counsel by Chairman Mike Oxley for the newly established House Committee on Financial Services. During his tenure, the committee provided legislation addressing terrorist financing and money laundering, improving investor confidence in the wake of the Enron and WorldCom scandals and granting consumers important new tools to fight identity theft. During the 1990s, Halpern served on the House Committee on Energy and Commerce, where he handled a variety of legislative issues, including automobile safety, insurance, FTC consumer protection and tobacco regulation. Halpern began his career in Congress as an intern for Rep. E.G. "Bud" Shuster in 1987.

Halpern served a number of temporary positions during his time on Capitol Hill. He was the Parliamentarian to the First Select Committee on Homeland Security, which created the Department of Homeland Security, General Counsel to the Select Committee to investigate the voting irregularities of August 2, 2007, and Assistant Parliamentarian to the 2008, 2012, and 2016 Republican National Conventions.

A native of Hollidaysburg, PA, Halpern received bachelor's and master's degrees in Political Science from American University in 1991 and 1992, respectively. He also received a law degree from George Mason University in 1997. Halpern has been included in Roll Call's list of 50 most powerful Congressional staffers 14 times and featured in a National Journal profile as one of "The New Power Players" on Capitol Hill.

U.S. GOVERNMENT PUBLISHING OFFICE

732 North Capitol Street, NW | Washington, DC 20401-0001

www.gpo.gov I facebook.com/usgpo I twitter.com/usgpo I instagram.com/usgpo