State Secrecy, Archival Negligence, and the End of History as We Know It*

Matthew Connelly

NO MATTER how heated a political debate, the bitterest of adversaries can usually agree on one thing: someday, historians will look back, judge the rights and wrongs, and render an informed verdict. What reporters and pundits say, after all, is only the first draft of history, and they do not have access to all the relevant information—especially when much of it is classified. It takes time for events to play out, for passions to cool, and for the protagonists to leave the stage. Researchers also need full access to the documentary record. True, history may be written by the winners. But if records survive and people still care about the past, we trust that eventually the truth will out.

History has often served as the ultimate court of appeal when other courts fail to uphold constitutional rights. Although eugenicists convinced eight justices of the U.S. Supreme Court in 1927 that the "feeble-minded" Carrie Buck could be compelled to undergo sterilization,[13] the patient work of scholars has shown that she was a victim of class prejudice.[14] Similarly, it took almost three-quarters of a century before the *Korematsu* decision allowing the internment of Japanese Americans was officially repudiated.[15] But long before, it had become a textbook example of how whole communities can lose their rights because of fear and racism.[16] And while efforts to defend the free speech of alleged communist "subversives" failed in a number of mid-twentieth-century cases,[17] scholars have made the Red Scare and McCarthyism object lessons in how important it is to protect unpopular speech when the nation is under threat.[18] History can supply a nearly limitless source of insight into how our government and Constitution actually work—or fail to work. As Chief Justice Earl Warren wrote in 1957: "Teachers and students must always remain free to inquire, to study and to

evaluate, to gain new maturity and understanding; otherwise our civilization will stagnate and die."[19]

But while history is our last chance to redeem constitutional liberties, it is hardly our best chance, and our chances keep getting worse. A faith in history assumes that there will be a historical record for scholars to examine. This assumption, in turn, is based on many other assumptions, starting with the idea that decision-makers will treat their records as public records, not classify them without cause, and certainly not destroy or delete them. Departments and agencies must keep these records organized on stable media and eventually entrust them to the National Archives. Archivists have to be able to identify the records that merit preservation, withhold only those containing still-sensitive national security or personal information, and put everything else on open archival shelves or on the internet—complete with finding aids to guide researchers. It is only then that we can start gaining that "maturity and understanding" and finally learn what the government has done in our name.

In fact, none of these assumptions is likely to prove valid. While politicians will continue to appeal to history when convenient, they have so neglected the National Archives and so failed to control official secrecy that future historians will have a hard time proving anything at all. It is a true crisis, one that has not garnered more public notice because we tend to pay attention only to that first link in the chain: the excessive secrecy of current officeholders, or at least officeholders from the opposing party. But that is only the beginning of our problems. The other factors contributing to the crisis are metastasizing slowly, quietly, in record centers, archives, and libraries. We will not be fully aware of what is being lost for decades to come, if ever. But the effects are already manifest in a range of scandals that have poisoned political discourse, from the destruction of interrogation videos by the Central Intelligence Agency (CIA) to the Internal Revenue Service's inability to produce emails related to the treatment of conservative groups to Secretary of State Hillary Clinton's decision to use a private server for official communications.

Government transparency was once a bipartisan cause, championed by both the left and the right. But each of these scandals was intensely polarizing, to the point that it is becoming impossible to protest such behaviors without being dismissed as politically motivated. What may be even worse is when we are *not* scandalized by even more obviously outrageous conduct, such as the recent revelation that President Donald Trump routinely rips up

his papers into tiny little pieces in violation of the Presidential Records Act, and that he fired the career government records managers who had been painstakingly taping them back together.[20] As long as we fail to see what all of these things have in common, we will keep missing the bigger picture. How will history judge a generation of government officials who not only insist on working in secrecy but also fail to protect, or even destroy, the record of their actions?

To understand how we arrived at this crisis, we have to go back in time. We need to trace two long-term developments: the growth of a postwar national secrecy complex and the simultaneous creation of a system of laws and institutions to organize and preserve official records. The first of these developments gets far more attention. But state secrecy in the United States has always been intertwined with—and to some extent legitimated by—state archiving. Over the past three decades, however, the exponential growth in state secrecy and the neglect of state archives have imperiled the whole system for organizing, protecting, preserving, and revealing the historical record. If this system collapses, America's commitment to learning from its history will become a thing of the past, because the past itself will be impossible to recover.

ARCHIVES AND THE CURRENCY OF SECRECY

Americans typically visit at most one archive and do so as tourists: the National Archives on the Mall in Washington, D.C. The impressive neo-classical facade has the aura of antiquity. In fact, the building did not open until 1935, the year after Congress established the National Archives and Records Administration (NARA). Until then, every government department managed (or mismanaged) its own records, no matter how old. The State Department had responsibility for the papers of the Continental Congress, so even the Declaration of Independence suffered from wet press-copying, direct sunlight, and botched repairs with glue and scotch tape.[21] The original, signed Constitution was lost for decades before it was found in a closet folded up in a small tin box.[22]

These documents are so faded as to be scarcely legible, but they are now displayed in the National Archives rotunda in a row of bulletproof encasements filled with argon gas. For some visitors, the physical security may be more impressive than the documents themselves. At the first hint of

danger, any one of the guards can activate a mechanism that lowers them into a custom-built armored vault. It is designed to withstand the blast of a nuclear bomb.[23]

Standing before these documents creates mixed feelings. There is a sense of democratic transparency, as all the world can see the physical manifestation of American popular sovereignty. But at the same time, one senses an acute vulnerability. These documents have barely survived and could still disintegrate or disappear.

Government archives and records centers have always been like this. They afford a glimpse of the state's inner workings. But access can be withdrawn, documents can be destroyed, and far more is hidden than what is put on display. When the documents are digital and concern sensitive matters, that sense of vulnerability is even more acute. Instead of glass cases in the National Archives, we have to imagine repositories such as the server farms operated by the National Security Agency or the National Reconnaissance Office, where enormous volumes of data can be stored, mined, and instantly deleted. We can only guess at how such data are managed because we don't have a window to look into what these agencies are actually doing. The intelligence community was built to look at *us*—all of us, around the world—recording what it sees for purposes that may not be revealed until years or decades into the future, if ever.

For its first 150 years, the U.S. government kept few official secrets, which is why it hardly bothered to keep archives. Other countries routinely intercepted the mail and operated "black chambers" to decode encrypted communications.[24] But the U.S. government had no such capacity. Mail tampering was expressly prohibited by law—indeed, it was punishable by death until the late 1800s.[25] After the first generation of revolutionaries passed from the scene, even American diplomats sent nearly all their dispatches in the clear.[26] Beginning in 1861, the State Department began to publish normally secret communications with other states on an annual basis. President Lincoln believed that this kind of radical transparency would show the world that the Union deserved support in its fight against the Confederacy.[27]

It was not until 1882 that the United States established its first foreign intelligence agency, the Office of Naval Intelligence. This office grew out of the Navy Department Library, which collected reports from naval attachés posted abroad.[28] It was only when the United States entered World War I that the Navy and Army both borrowed the British practice of systematically classifying sensitive information as "confidential" or "secret," and at the same

time began to upgrade their cryptological capabilities.[29] Shortly afterward, a former librarian, J. Edgar Hoover, began to assemble and organize surveillance files as director of the new Federal Bureau of Investigation (FBI).

But the U.S. apparatus for producing secret information was still puny in comparison to those of other great powers. The State Department continued to mark documents "confidential" or "private" almost at random, and in 1929, Secretary of State Henry Stimson shut down the department's cryptologic group on the grounds that "gentlemen do not read each other's mail."[30] There was no central agency to coordinate intelligence work. Nor was there a central archive to organize and preserve government information. In the 1930s, half a century after other great powers had created national archives, federal workers for the first time started to inventory the holdings of U.S. departments and agencies. They found War Department papers moldering in piles in the White House garage.[31]

Like so much else, this changed with the New Deal. Congress passed the National Archives Act in 1934, and President Franklin Roosevelt appointed the first Archivist of the United States. Proliferating agencies and departments produced exponential growth in government papers. By 1930, the federal government had accumulated less than half as many records as it would generate just in the ten years leading up to its entry in World War II.[32] Starting with the General Disposal Act of 1939, and continuing with the Federal Records Acts of 1943 and 1950, the National Archives was authorized to decide what should be saved. As Roosevelt's second archivist, Solon Buck, explained, "the chief reason for destroying is to save": without "weeding out useless papers," there could be no recognition or preservation of records of lasting value.[33]

Crucially, Congress made it unlawful to "alienate[] or destroy[]" any U.S. government records except in accordance with procedures established by the Archivist.[34] Even if individual employees would continue to keep their personal records to themselves, the thinking went, contemporaneous archiving systems would remove any temptation to rewrite the official record. Roosevelt so believed in the importance of archives that he created the first presidential library in 1941, offering these words when it was dedicated at his estate in Hyde Park:

To bring together the records of the past and to house them in buildings where they will be preserved for the use of men and women in the future, a Nation must believe in three things. It must believe in the past. It must

believe in the future. It must, above all, believe in the capacity of its own people so to learn from the past that they can gain in judgment in creating their own future.[35]

Roosevelt also believed in secrecy, however. He was the first president to issue an executive order systematizing the national security–related information that would be withheld from the public.[36] Wartime, moreover, revealed that archives themselves could contribute to national security. While the Declaration of Independence and the Constitution were shipped off to Fort Knox for safekeeping during World War II, archivists set to work finding detailed maps of German and Japanese cities for target intelligence.[37] Meteorological records were required for preparing amphibious landings. State Department papers from prior international negotiations were important in planning postwar settlements. Roosevelt also created the Office of Strategic Services (OSS), the precursor to the CIA. The OSS cordoned off parts of the National Archives building so that it could use the collection in secrecy.[38]

State archiving and state secrecy thus grew together out of the same ground and for some of the same reasons. Archives provided a place where secrets could be safely stored, and sometimes destroyed, but always with the idea that the most important secrets would be preserved, both as working memory for "the official mind" and for the judgment of posterity.

At the start, it was not obvious that secrecy would overpower archiving. For instance, President Roosevelt expected that the War Department would only sublet the Pentagon—after the end of hostilities, these 2 million square feet were to be turned over to the National Archives for the storage of valuable papers.[39] President Truman abolished the OSS, worried that a powerful intelligence agency might be used against Americans, and initially replaced it with a small group that merely coordinated the information flowing to the president.[40]

Ultimately, though, the war not only led to the creation of a permanent military-industrial complex; it also created what Senator Daniel Patrick Moynihan later called a "culture of secrecy."[41] On the ground that loose lips could sink ships, information was shared only on a "need to know" basis. Stamping a document "secret" made it currency that could be exchanged for other assets.[42] By 1944, so much was secret that a new classification was created, "top secret." The entire Manhattan Project was deemed a top secret, and it became the model for how to compartmentalize information.[43]

Inevitably, the currency of secrecy became debased. In 1956, during another decade of exponential growth in federal records, a Department of Defense study found that "overclassification has reached serious proportions."[44] Excessive secrecy hindered public accountability and made it harder to protect real secrets, even as it led officials to undervalue open sources. By 1961, the National Archives had taken custody of almost 100,000 cubic feet of classified records.[45] A series of high-level panels, right down to Moynihan's own 1997 Commission on Protecting and Reducing Government Secrecy, came to the same conclusion: officials found classifying information to be safe, easy, and expedient, whereas bucking the system to reduce secrecy was risky, complicated, and unrewarding.[46]

Overclassification has therefore long been notorious. But the costs are not limited to reducing accountability in the here and now. The more information is classified, the greater the cumulative burden of secrecy weighs on us for decades to come.

THE ILLUSION OF REFORM

Almost every president since Roosevelt has promised to make the government more transparent. When he tightened the security classification system, Truman claimed that in the long run his order would make more information available to the public, rather than less.[47] Even President Nixon promised "to lift the veil of secrecy which now enshrouds altogether too many papers written by employees of the Federal establishment."[48] These presidents follow the same basic playbook, pledging to "automatically" release the secrets of previous administrations after a fixed period of time, typically ten or twenty-five years; to reduce the number of people who are allowed to classify information; and to reduce the amount of information that is classified at the highest level.[49]

These reforms aspire to prevent the debasement of the currency of secrecy. But every president has allowed lots of exceptions to "automatic" declassification. And no Congress has made significant resources available for reviewing the massive backlog of classified information. Reform has always failed in the face of unrelenting inflationary pressure coming from the other direction.

Take automatic and systematic declassification, whereby presidents try to open up the secrets of their predecessors. This is the main route by

which most classified documents are declassified: some 44 million pages in fiscal year 2016.[50] But some 59 million more pages were withheld that year under automatic and systematic declassification, even though most were decades old.[51] Any department that claims an "equity" in a document can block its release. And this does not even account for all the papers exempt from automatic review. For instance, bureaucrats reportedly decided that more than 90 percent of classified documents fell under the many exemptions allowed by President Carter's executive order.[52] Executive orders have long allowed any record that might reveal intelligence sources and methods to be excluded from automatic review.[53] Even when it comes to the relatively innocuous work of research and analysis, the CIA still withholds 80 to 90 percent of the documents from the Cold War era that are reviewed for automatic declassification.[54] The CIA is also notorious for blocking the release of documents by other departments and agencies, but it is not the only one. The Department of Energy, for instance, has broad powers to keep documents classified if they might contain information related to nuclear weapons. It requires page-by-page review, and it double-checks the work of other departments.[55] So there is nothing automatic about automatic declassification.

Congress and the president sometimes direct that materials not yet accessioned to the National Archives undergo review for "discretionary" declassification. This can serve the laudable goal of advancing more immediate accountability, as when the Obama administration decided to disclose the intelligence agencies' budgets[56] and the size of the U.S. nuclear stockpile.[57] It can likewise support "truth and reconciliation" efforts, as occurred with the declassification of records on U.S. support for military dictatorships in Latin America.[58]

But discretionary declassification can also be weaponized to discredit opponents and preserve the prerogatives of whoever occupies the White House. This could be seen in the recent release (in full) of Senate Republicans' memorandum attacking the Obama administration's investigation of the Trump campaign, and even more so in the (heavily redacted) Democratic rebuttal.[59] This is an old story. Former vice president Dick Cheney criticized President Obama for releasing legal memoranda that detailed "enhanced interrogation" techniques without simultaneously declassifying documents that allegedly showed how these techniques produced valuable intelligence.[60] And decades ago, President Nixon put his "plumbers" in charge of declassifying documents on the 1961 invasion of the Bay of Pigs in order to discredit the Kennedys.[61]

Similarly, reducing the number of people in the federal bureaucracy who classify information may serve the interests of the president and the president's political appointees by helping them to centralize control of secrecy, without necessarily having any effect on the overall amount of secret-keeping. For instance, Obama promised on his first day as president to have the most transparent administration ever,[62] and during his first term he reduced the number of people with "original classification authority"— the only people who can decide that some new item or category of information must be kept secret—from 4,109 in fiscal year 2008[63] to 2,326 in fiscal year 2012.[64] But many more people can still classify documents that are related to previously classified subjects or that contain information falling into an approved category. According to the government's Information Security Oversight Office (ISOO), the number of "derivative" classification actions more than quadrupled during the same period, from approximately 23 million in 2008[65] to over 95 million in 2012.[66] In other words, three times a second, every second, some government official decided that what he or she was working on had to be shielded from public scrutiny until years or decades into the future.

The Obama administration also carried out more prosecutions for leaks to the media than all preceding administrations put together. It is not clear why. One possibility is that new technology made it easier to catch leakers; another centers on the creation in 2006 of a National Security Division within the Justice Department.[67] But whatever the cause, the effect may be to further reinforce the president's power to decide what ought to be kept secret and what ought to be revealed. Conversely, penalties for overclassification are almost never applied,[68] even though the Interagency Security Classification Appeals Panel (ISCAP) has found hundreds of cases in which officials have wrongly denied the public access to government information.[69]

In this otherwise dreary history, the Freedom of Information Act (FOIA) would appear to show what real reform looks like, given that hundreds of thousands of requests for federal agency records are processed under FOIA each year. President Lyndon B. Johnson was deeply suspicious of FOIA, but he reluctantly signed it anyway. Many in his administration did not think it would make any practical difference when it came to classified information.[70] Indeed, the original 1966 law turned out to be a "relatively toothless beast" more generally,[71] revealing a fundamental flaw: rather than regulating official secrecy at the source, FOIA placed the burden on

the individual to know which secrets should be revealed and to seek their release on a case-by-case basis.[72]

Even after Congress strengthened FOIA in 1974, FOIA has proven to be a notoriously slow and unwieldy instrument in those areas where official secrecy is most likely to be abused. With exceedingly rare exceptions—what some call "legal unicorns"[73]—judges accept at face value agencies' refusals to release documents for national security reasons, all but abdicating the role Congress gave them.[74] In 2016, Congress passed a FOIA Improvement Act that included some useful features, such as requiring that agencies post frequently requested records on the internet. But Congress did not appropriate any new funds to meet these mandates.[75] As of fiscal year 2015, the federal government was spending $480 million annually complying with FOIA requests, more than NARA's entire budget.[76] And whereas FOIA's original advocates were journalists and it is commonly assumed that FOIA requests are intended to serve the public interest,[77] most of this money constitutes corporate welfare. At numerous agencies, well over two-thirds of the FOIA requests are submitted by corporate requesters, including a slew of companies that aggregate information extracted from regulatory or contracting agencies and then keep it secret from everyone except paying customers.[78]

FOIA can still be a useful workaround for the historian when all else fails, but it is inefficient and it diverts resources and attention that might otherwise go toward constructing a more rational declassification system. Even high-profile victories can make the public believe that their government is more open and accountable than it really is. Rather than help secure the historical record, FOIA may have legitimated and strengthened the president's all-but-exclusive control over classified information, the bedrock of our national secrecy complex.[79]

Leaks to the press can be seen as another strategy for coping with overclassification.[80] Here again, though, the White House maintains its monopoly on the licit use of state secrets, as high-level officials speaking on behalf of the president are almost never prosecuted and as ambiguously authorized "pleaks" and "plants" allow such officials to shape the public record while maintaining plausible deniability.[81] Wiki-sized leaks by transparency vigilantes are no solution either. Even aside from the potentially catastrophic costs to innocent victims, these leaks, too, only represent a relatively small and biased sample of the historical record. Moreover, researchers have a hard time knowing what to do with this kind of data

because they usually cannot tell what it actually represents, or even whether it is authentic. At the same time, fears of vigilante leaks and hacks serve to justify ever more spending to protect classified information.

To be sure, both FOIA and unofficial disclosures can form useful parts of a larger transparency ecosystem.[82] But manure and decaying carcasses can also help other things grow. We need to think harder about how both FOIA and unofficial disclosures have nourished and strengthened the growth of the national secrecy complex, while the slender vine of state archiving withers in the darkness.

THE CURRENT CRISIS

In recent years, three trends have converged to create a true crisis for the U.S. historical record. The first is that a decline in appropriations and dubious management decisions have decimated the cadre of experienced archivists. NARA's annual budget dropped by nearly a quarter in inflation-adjusted terms during the Obama administration, to around $375 million.[83] Attrition is also changing the composition of NARA's workforce. In the past, newly hired archivists trained under senior colleagues while gaining deep knowledge of a particular group of records. In recent years, new hires have instead been trained in information management and rotated from one collection to the next, never long enough to acquire mastery. Archivists who previously advised researchers or processed new collections have been reassigned to serve senior management or to staff high-profile digitization projects. Once this institutional memory was lost, it was lost forever.

The result is that morale at NARA is among the worst in the entire federal government. In surveys of employee job satisfaction at government departments and agencies, NARA now routinely ranks among the unhappiest workplaces.[84] "It is so difficult," the president of the NARA workers' union has explained, "it's like trying to fight an octopus in a cave, underground, that has just squirted you with ink."[85]

At NARA's largest research facility, Archives II in College Park, Maryland, there were just forty-one archivists remaining as of 2013—the last time NARA's inspector general audited the processing of paper records—to work through over 700,000 cubic feet of unprocessed records.[86] At the presidential libraries, the situation is even worse. Archivists in these libraries had not processed the majority of the paper records they had received, and they

estimated it would take decades to reduce the backlog.[87] The inspector general's audit did not even attempt to quantify the backlog of electronic records, except to note that the volume of electronic records has "grown exponentially" since the Reagan administration, that as of 2012 the presidential libraries held over 300 million "logical data records," and that over 95 percent of these records are believed to remain unprocessed.[88] It "boggles the mind," admitted William Mayer, NARA's executive for research services, in 2014.[89] Referring to the final scene from *Raiders of the Lost Ark*, which shows how a crate stamped "top secret" would be lost among thousands of other crates in a cavernous warehouse, Mayer said he has come to realize that "Spielberg got it right!"[90]

This brings us to the second trend: the growth in the sheer volume of records that need to be reviewed for national security information. In recent years, the reported number of classification decisions has stabilized, at around 55 million annually.[91] This is self-reported data, from officials who have been told to stop classifying so much. Even so, between 2007 and 2016 over 630 million classification decisions were taken overall,[92] covering an incalculable number of documents, emails, PowerPoint presentations, and audio/video recordings.

The amount of money the government spends each year to keep this information secure provides the most tangible way to measure the growth of national security secrecy and the threat it poses to the historical record. In the three years following Edward Snowden's disclosures in 2013, government spending on such items as physical and personnel security, training, and "technical surveillance countermeasures" increased by 45 percent, to almost $17 billion per year.[93] If we had a "ministry of secrets," its budget would now be bigger than that of the Department of Commerce or the Department of the Interior. Spending on declassification, on the other hand, was less than $109 million in fiscal year 2016, notwithstanding the massive increase in the amount of classified information requiring review.[94]

So while the Obama administration promised to strike a "careful balance between protecting essential secrets and ensuring the release of once sensitive information to the public as quickly and as fully as possible,"[95] the actual "balance" is tipped toward secrecy by more than 99 to 1. And it keeps tilting further in the wrong direction. The share devoted to declassification is one-tenth as high as it was back in 1999, when spending on information security was less than a fourth of what it is now and spending on declassification was more than two times *greater* than it is now.[96]

The collapse in funding for declassification and the growth in the number of classified documents that require review have had a predictable impact on the amount of information released to the public. For a period in the late 1990s, approximately 200 million pages of documents were being declassified each year.[97] But between 2007 and 2016 (the last year for which we have data), this number averaged 31 million.[98] The last two years saw a slight recovery, with 37 and 44 million pages declassified in 2015 and 2016, respectively.[99] But the percentage of pages withheld (out of the total number of pages reviewed for declassification) also increased in those years, to over 57 percent, the highest level ever recorded.[100]

People can try to force the release of records right away by filing FOIA requests. But even apart from the special difficulties raised by requests for classified records—which are often released in extensively redacted form if they are released at all—there is also a massive and growing FOIA backlog throughout the executive branch. In 2009, President Obama ordered all departments with significant FOIA backlogs to reduce them "by 10 percent each year."[101] As of 2015, fourteen of fifteen cabinet departments had failed to meet this goal, and the average annual backlog across all federal departments had *increased* by over 8 percent.[102] The two worst offenders were the State Department and the Department of Homeland Security.[103] Moreover, it seems that many departments have reduced their backlogs simply by asking requesters if they are still interested, and then closing cases when they don't receive an immediate response.[104]

The third trend involves the shift from paper documents to digital data— really big data—and it is the one that may finally bring about a collapse. As the nonpartisan Public Interest Declassification Board warned in 2012: "The expected growth of electronic records will create new backlogs almost incomprehensible in size."[105] State Department historians have started to call it the "Big Bang." While many were amazed at the 250,000 diplomatic cables released by WikiLeaks in 2010 and 2011, these cables represented less than 1 percent of the 27 million records amassed in the State Department's Central Foreign Policy Files between 1973 and 2006.[106] But the real growth is in new media. It is estimated that the State Department is generating 2 *billion* emails every single year.[107]

Imagine for a moment that just 3 percent of these records were retained, roughly the same proportion of paper records that have historically been retained by the National Archives. What would it take to screen 60 million emails for national security and personal information? We can get a preview

by looking at how long it took the State Department to process former secretary Hillary Clinton's emails, which totaled some 54,000 messages. With a federal judge and Clinton herself urging rapid review, the government gave it top priority. A large team was assembled, with dozens of officials focusing on this one task. Still, it took about nine months to review nearly all the emails Clinton turned over and to release them to the public.[108] How long would it take to review one year's worth of State Department emails with the same urgency and the same staffing, assuming we wanted to retain roughly the same 3 percent of those emails as historically significant? More than 830 years.[109]

The State Department probably does a better job preserving its records than most of the government. It has about fifty full-time historians. Unique among federal agencies, the State Department has a congressional mandate to publish a documentary record of U.S. foreign relations.[110] Congress also created an external Historical Advisory Committee that meets regularly and reviews archival practices.[111] Contrast this with what we know (or don't know) about the rest of the federal bureaucracy. According to the Public Interest Declassification Board, a bipartisan body created by Congress to promote public access, a single (unnamed) intelligence agency produces a petabyte of classified information every eighteen months.[112] A petabyte equals approximately 20 million four-drawer file cabinets filled with textual documents. The board notes that, using current declassification methods, 2 million employees would have to work full time to review this many documents each year.[113] The fact that most intelligence records likely do not come in text form but rather as remote sensing data, communications intercepts, and so on makes this challenge qualitatively greater. It is utterly new, something that an understaffed and demoralized National Archives and an outmoded declassification system cannot possibly cope with.

A DIGITAL DARK AGE

Historians once hoped that digital sources, accessed via the internet, would democratize historical research and make it accessible to a broader audience.[114] Instead, the United States may be entering what's been called a digital dark age.[115] As discussed above, digitization has led to an exponential increase in the volume of government records and to crushing new burdens for government archivists. But historians also need to worry about the

deliberate erasure of electronic records, as when CIA official Jose Rodriguez ordered the destruction of videotapes showing how suspected terrorists were subjected to repeated waterboarding.[116] We also need to consider how darkness can fall for reasons that are more banal than evil, including overzealous pruning of archives and "bit rot," the loss of data on outmoded software and hardware. And even when electronic records survive and can be accessed, we need to confront a host of challenges raised by reviewing such records for sensitive information and using them for historical research.

Start with the unauthorized loss or destruction of federal records. The National Archives is currently investigating twenty-five cases, many involving government officials using private email or encryption services like Signal.[117] This is the first time NARA has reported such investigations. But bureaucrats have been trying to delete incriminating emails since they first began sending them. In 1989, it took a last-minute lawsuit to prevent President Reagan's outgoing National Security Council (NSC) from destroying all of its electronic messages, on the same system that had enabled investigators to uncover the Iran-Contra scandal.[118] After years of litigation, the district court eventually ruled that the government had to preserve these messages, including their metadata (the subject line, from/to fields, and so on).[119] Much of the value in electronic records comes from being able to use metadata to analyze them in the aggregate.

At that point, a whole new set of problems emerged. Recovering the NSC emails from 150 hard drives almost overwhelmed NARA's technical capabilities. Torn tapes had to be spliced, creases ironed out, and moisture baked off in ovens.[120] So how would today's NARA, with a significantly smaller budget, cope with a vastly larger data dump of DOS-era software and decades-old hardware? As it is, NARA's information technology infrastructure is inadequate for the existing workload. Back in 2013, the agency's inspector general found that it could not assure the long-term preservation of electronic records.[121] More recently, the inspector general reported that management is continuing to use legacy systems for declassification and redaction that were already due for replacement five years ago; NARA does not even have a process in place to determine the age of these systems.[122] The agency is developing a new Electronic Records Archive, but assessment of how it could be adapted to classified records "is still in the earliest stages."[123]

Nevertheless, all federal departments and agencies were ordered in 2012 to switch from paper to electronic archives by 2019.[124] Forty years too late, NARA's strategic plan calls for coordination so that these departments and

agencies can deliver the data in a form NARA will be able to preserve and process.[125] We can already see what is likely to happen by examining how NARA dealt with some of the first electronic records that came its way: the State Department's Central Foreign Policy Files. In 2007, archivists decided it would be impossible to review all 27 million records in these files to determine what to preserve permanently.[126] They began to experiment with sampling. In the case of records related to passports, visas, and citizenship, for instance, they looked at 200 documents out of almost 6 million, or 0.003 percent. They did not actually have a random sample, and in most cases they did not even read the full sample before deciding that a whole class of records should be permanently deleted. Diplomatic cables on cultural diplomacy, educational exchanges, international sport, and scientific cooperation are now among the permanent gaps in the historical record.

Records managers and archivists have always had to "weed" or "prune" routine documents and duplicate documents to make room on archival shelves. But these State Department records were small text files, millions of which easily fit on a single hard drive. Moreover, seemingly mundane records can present remarkable opportunities for research using contemporary data-mining techniques. With millions of cables on passports and visas, for instance, researchers might have been able to develop a vastly more sophisticated understanding of global migration. This opportunity is now lost forever.

Even before archivists began deleting files, the text of more than 7 percent of the cables for the years 1973 to 1978 had been lost for reasons the State Department still cannot explain. A much higher proportion of secret cables—22 percent—and cables from particular periods went missing.[127] Gone are the majority of telegrams from the beginning of December 1975, when President Ford and Secretary of State Kissinger acquiesced in General Suharto's murderous invasion of East Timor. We are also missing almost all records from the end of March 1976, when the U.S. government supported a military coup that started a civil war in Argentina. Gone, too, are the messages from June 1976, a period that included the Soweto uprising against apartheid and the Israeli raid on Entebbe.

Of the remaining files, the government is still withholding a large and growing percentage from the public. This includes virtually all the top secret cables, some of which are now almost half a century old. State Department officials have told me that declassifiers don't have access to hardware that is considered sufficiently secure. Yet most of the withheld records were

never classified to begin with, which suggests that they are being withheld because they are thought to contain "personally identifiable information." Ironically, the government's solicitous regard for decades-old addresses and phone numbers means that we may never learn the full story of how it conducted surveillance programs. Operation Boulder, for example, subjected tens of thousands of visa applicants to FBI investigation between 1973 and 1975 merely because they had Arabic-sounding last names.[128] That history might have proved instructive after 9/11, but the vast majority of these Boulder records are still unavailable to researchers.

Millions of electronic records from the State Department Central Foreign Policy Files *have* been preserved and released, and we can use keyword searching to retrieve them.[129] This is a powerful tool, but it is a poor substitute for a finding aid, which knowledgeable archivists create to provide detailed descriptions of the scope and content of historical collections. In a classic study, researchers found that people miss four out of five relevant records because they do not—and cannot—know all the keywords to use.[130] Even when researchers do find something, what then? Historians have long sought to read documents in the same way our subjects did, in the context of related materials organized in files. Only in that way can we understand how they viewed the world. In that sense, all of these diplomatic cables are now lost, floating in cyberspace.

It is not just the cables. In 1974, the State Department began microfilming paper records in the Central Foreign Policy Files and storing copies in the same electronic system, destroying the originals as well as the physical filing system. We can only order these records in the NARA reading room at College Park. But they come in a box filled with otherwise random printouts. What else might the Secretary of State have been looking at or dealing with when he held that document in his hands? It is almost impossible now to tell.

This crisis in the archives is not, therefore, just a threat for the future. It is a depressing everyday reality for the scholar of recent history. In the place of an archivist, we have a search box. Instead of finding aids, we have an FAQ webpage. Rather than gaining a sense of context and the possibility of drawing serendipitous connections, we experience vertigo, with millions of documents effectively dumped on the ground, left for us to pick through and trip over.

This is how the government treats what all agree is one of the core collections recording our national and global history, a collection that is protected

by the State Department's Office of the Historian and subject to robust oversight by a congressionally mandated advisory committee. What will happen to the millions and eventually billions of emails and other records produced by every other department and agency, records that might not have the same relevance to national security but might ultimately prove even more important?

Recently, a group of scholars gathered to consider the state of historical knowledge about just one part of our collective past, the space program. Inspiring stories like the film *Hidden Figures* have shown that, decades later, we still have much to learn about what was already celebrated as one of the greatest achievements in the history of life on earth. But in fact, the assembled historians and archivists found that, when the space program first began, officials at the National Aeronautics and Space Administration (NASA) seemed not to care whether they preserved a historical record. So great are the gaps, it is unclear whether any archivist was around when the program first reached into space, sent humans to the moon, and probed the universe beyond. Even now, NASA archivists must "constantly" justify the need for their work to agency leaders and find themselves increasingly overwhelmed by the challenge of saving digital records.[131] Will future generations of Americans travel to the stars without even knowing how we got there?

WHAT CAN BE DONE?

Martin Luther King Jr. famously said that "the arc of the moral universe is long but it bends toward justice."[132] The context makes clear that King—who was actually quoting another minister—was not talking about current events but divine judgment in the hereafter. The only part of that universe now visible to humans is what we call history. "Evil may so shape events that Caesar will occupy a palace and Christ a cross," King wrote a decade earlier, "but one day that same Christ will rise up and split history into A.D. and B.C., so that even the life of Caesar must be dated by His name."[133]

We do not need to wait for future historians, or God, to judge those rulers who try to bend historical scholarship to their own ends or bend the historical record toward oblivion. Americans need to ask themselves right now why we have allowed our elected officials to neglect their most basic duty: to preserve a record of what they do in our name.

Even rulers who do not care about how they are judged should recall that other governments, some more cynical than ours, have much longer memories. They know how to use history not just to learn from the past but to confuse and embarrass their enemies, as when Vladimir Putin recently accused the United States of breaking a promise not to extend NATO into Eastern Europe.[134] Was he right? Or how about the Iranian nuclear program: was it fair for Mahmoud Ahmadinejad to say that we only opposed it when the Iranian government was no longer our ally?[135] And what exactly is the commitment past U.S. presidents made to defend Taiwan in a conflict with China? If you don't know the answer to these questions, that is precisely the problem. As President Roosevelt realized, preserving institutional memory is not just essential for democratic accountability. It is a matter of national security.

So too is gaining control over official secrecy. When everything is secret, nothing is truly secure.[136] A culture of extreme secrecy breeds contempt, and sometimes revolt. This is a lesson we should learn from Daniel Ellsberg, Chelsea Manning, and Edward Snowden. Rightly or wrongly, each one felt that he or she owed fellow citizens a full account of official lies and misconduct. Rather than seeing the almost constant leaking from the Trump administration as an aberration, it may instead represent the culmination of a crisis that is decades in the making. And if we cannot find a way to restore control, how will any future president ever again be able to act "with secrecy and despatch," as Alexander Hamilton put it, when our national security depends on it?[137]

We therefore need to quickly take some commonsense steps, while also starting to think about more radical measures commensurate with the even more immense challenges to come.

1. Congress must raise NARA's funding to match its mission. Among federal departments and agencies, NARA is more like the Department of Veterans Affairs (VA) than the Department of Education. Its budget cannot grow and shrink depending on policy preferences any more than the VA can stop caring for wounded warriors once any given war comes to an end. Both must cope with a cumulative legacy, including conflicts and covert operations that occurred decades ago. If NARA does not have adequate resources, then the hard-won lessons of the past will never be preserved, and we will be paying for yet more national traumas. A billion-dollar budget for the National Archives—roughly half the cost of a single Navy destroyer—would be a cheap insurance policy against repeating trillion-dollar mistakes.

2. NARA must adapt archival practices to the era of big data. As NARA has acknowledged, electronic records are the agency's "single greatest challenge and opportunity."[138] Meeting this challenge does not necessarily mean diverting resources to digitizing paper collections. Why do that when we are already losing *born-digital* collections to bit rot? Meeting this challenge begins, instead, with the realization that data-mining will be an increasingly important method of historical research and taking this into account when archivists appraise whether records merit preservation. If archivists use statistical sampling, they should use rigorous methods and ensure they have a randomized sample. And if they delete a collection, they should at least preserve the sample, so future researchers will know what they are missing.

3. The rest of us must stop treating each new scandal about archival neglect or destruction as just another opportunity to score political points. All of us pay a price when public officials disregard the most basic requirements of democratic accountability. To be sure, Democrats paid the biggest price when Hillary Clinton used a private email server for official business and had her lawyers delete messages they deemed personal years after the fact. Instead of an honest reckoning, one that would acknowledge her understandable fear that her personal emails might be sought through FOIA, Clinton and her defenders continue to insist it was all just a big distraction from "the real issues."[139] But according to the Justice Department inspector general's report, one of the reasons she was not prosecuted for the concealment, removal, or destruction of records is because the relevant law has "never been used to prosecute individuals for attempting to avoid Federal Records Act requirements."[140]

On the other hand, Clinton's defenders were right to be furious when FBI director James Comey claimed she was "extremely careless" because a small fraction of her communications were later deemed to require classification.[141] Every day, government officials disagree about what information is truly sensitive—usually without even realizing it, as when they release two versions of the same document with different redactions.[142] While FBI agents were investigating Clinton for how she handled information that may or may not have been sensitive, they repeatedly handed sensitive information to reporters. Comey himself used private email for official business and leaked to the press.[143]

President Trump's habit of tearing up presidential records into little pieces[144] is only the most extreme example—so far—of a degenerative process that has been decades in the making. The more examples we have of

both Democrats and Republicans ignoring their responsibility to preserve historical records, the more obvious it should be that no issue could be more real or worthy of bipartisan attention.

4. The executive branch must establish a standard of professional responsibility for classifiers and declassifiers and prioritize the protection of truly sensitive information. All the way back in 1955, Senator Hubert Humphrey complained that the government was making a massive investment in official secrecy without establishing what, exactly, had to be kept secret.[145] More than sixty years later, it is high time the government used part of the $17 billion it spends each year on information security to identify with greater care the information that officials agree requires safeguarding and—no less important—to identify information that has been improperly withheld. Academics have carried out hundreds of studies on "intercoder agreement" regarding myriad topics, few as important as this one. This kind of research is the only way to establish a reasonable standard for recognizing truly sensitive information, and for inculcating in classifying and declassifying officials a set of shared norms of professional responsibility.

Against the status quo of rampant overclassification, government commissions have repeatedly recommended a more rational, risk-management approach to protecting sensitive information. Even when it comes to nuclear weapons, the government's design goal is not to make accidents impossible—that standard would itself be impossible and would stand in the way of the more practical goal of minimizing risk.[146] But the executive branch's classification and declassification practices too often ignore this basic principle. Congressional statutes may also require rewriting, or at least reinterpretation, starting with the Kyl-Lott Amendment to the 1999 Defense Appropriations Act. Under Kyl-Lott, millions of records—even a serviceman's application to marry a Vietnamese citizen—continue to be reviewed page by page to ensure that none has information useful for building a nuclear weapon.[147] Watchdog groups have repeatedly found that this irrational, zero-risk approach to declassification severely slows down the process while also allowing dangerous information to slip through and sit on the open shelves of the National Archives, including sabotage manuals and recipes for manufacturing explosives.[148]

5. The executive branch must employ data-science techniques to mitigate information overload and to identify state secrets and personal information. How can officials cope with millions upon millions of

electronic records and prioritize those that really do require close scrutiny and safe handling? This basic challenge is not unique to the executive branch. It is analogous to the discovery process in large-scale litigation. Defense attorneys and data scientists have devised techniques for automatically identifying records responsive to a plaintiff's subpoena and for segregating records that are privileged because they contain attorney-client communications, trade secrets, or personal information.[149] Plaintiffs, for their part, have developed methods to find suspicious patterns in electronic communications.[150] As historians confront ever larger corpora, they too are starting to team up with data scientists to develop machine learning and natural language processing techniques, such as "topic modeling" large collections of declassified records to identify anomalous language, using traffic analysis to identify bursts of activity indicative of important events, and training algorithms to automatically classify sensitive and non-sensitive communications.[151]

We can use these and other techniques to begin to build a "declassification engine." Once we have established the reliability of human classifiers and declassifiers, we will know the standard these algorithms have to meet or beat. And if we can capture and store the information that officials are generating every time they release or redact a record, we can start using this data to train algorithms to help human reviewers focus on the records that are most likely to contain sensitive information.[152] The Public Interest Declassification Board has urged pilot projects to automate and streamline declassification. But so far, this has remained an unfunded mandate, and there is "little evidence that Executive departments and agencies are employing or developing the technologies needed to meet these objectives."[153]

CONCLUSION: THE JUDGMENT OF HISTORY

In the short term, these commonsense measures could prevent a collapse in the system for keeping our government accountable. But over the longer run, we will need to consider truly creative solutions for preserving both government transparency and legitimate state secrets in the era of big data. For instance, if secrecy is a kind of currency, then perhaps officials should not be able to mint new secrets—measurable by the number of "original classifications" recorded each year—without at the same time declassifying a comparable number of old secrets. If officials are found to have wrongly

withheld information from the public, as often happens when their decisions are reviewed, it could be treated with the same severity as an unauthorized disclosure. And if the executive branch cannot reform itself and courts continue to abdicate their responsibility, Congress might consider creating an independent body, akin to the Federal Reserve, with a mandate to control official secrecy and safeguard the public record.

Historians are awakening to the danger, but we cannot do much on our own. History is bigger and more important than the historical profession. The judgment of history, the very survival of history, depends on the work of records managers, archivists, cabinet officers, members of Congress, judges, and countless ordinary citizens committed to preserving and learning from the past.

To think we can all come together to preserve the record of our times may seem naïve, but here again, we can and must learn from history. When, in the 1960s, Pennsylvania Station fell to the wrecking balls and the Cuyahoga River caught fire because it had become clogged with oil and debris, national movements rose up and passed laws to preserve landmarks and protect the environment. The danger facing the National Archives is no less grave, and to secure the historical record we need a national movement energized with similar passion and dedication. It is never naïve to let the past be our guide.