Reassembling Scholarly Communications

Histories, Infrastructures, and Global Politics of Open Access

OA Funding Provided By:

- Arcadia Fund
- Birkbeck, University of London

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

The title-level DOI for this work is:

doi:10.7551/mitpress/11885.001.0001

16 Accessing the Past, or Should Archives Provide Open Access?

István Rév

The Archive, as we knew it for a long time, seemed to consist of static repositories based on a read-only paradigm.[1] Once documents were accessioned and processed, described and entered into finding aids, they were usually expected to remain dormant, except when read, consulted by the researcher. In recent decades the situation has changed: the Archive is now considered to be key to the understanding of an individual or a collective past, of future memory, of private and official secrets that provide explanations for either historical or quotidian—but nonetheless important—events.[2] Thus, archives became targets for openness, to shed light on the darkness of the depths of depositories, to reveal secrets, to gain access to the documents in custody of these solid, locked, dusty, unhealthy institutions. The public, instead of waiting for the researcher to find the relevant documents in the cellar, demanded immediate, free, digital access to all documents that are deemed important.

When, in December 2001, we invited a dozen or so scholars to a meeting—out of which came the Budapest Open Access Initiative (BOAI), one of the founding documents of the Open Access Movement—we were convinced that not only scholarly reports, the transactions of the learned societies, but also documents stored in the archives should become freely and openly accessible. In hindsight, it was a naïve and mistaken expectation.

* * *

Open and free access to documents is now conventionally understood as the right to have unimpeded access to documents with political, historical, or cultural significance for either the relevant community or the individual citizen concerned. The assumption is that the public has or should have the right to gain access to the information contained in documents

that are produced with direct or indirect public funding, that are legally no longer constrained by acceptable national security and secrecy provisions, are free of intellectual property or copyright restrictions, and that do not disproportionately harm the privacy of specific, nameable corporations or private individuals. Open data initiatives, providing free access to public or nonsensitive information, are now treated as a natural part of the widening concept of basic human rights. On the basis of this interpretation of rights, secrecy provisions, intellectual property and copyright restrictions, and archival laws and rules began to be disputed and challenged.

As part of such efforts to achieve openness, access, and transparency, legislatures have been urged to pass freedom of information acts, to change archival laws, and to make publicly available historical documents (especially documents of recent reprehensible government actions or incriminating documents of overturned repressive regimes). The public, often in the wake of regime change, wants to know not only what has happened, but also the specific legally or morally unjustifiable acts of named individuals. The publics in Argentina, Chile, Columbia, South Africa, Germany, Poland, and Russia demanded openness and public access to documents of the overthrown regimes. Archival or legal concerns about privacy, the informational rights of either implicated individuals, or third parties—individuals whose names were recorded in the documents, but who did not play any incriminating role in the events described in the sources—were treated by the public mostly as alibis for keeping the shameful acts of the past locked up in the dark.

In the course of the first decades of the twenty-first century, the situation of archives and archival documents has, thus, radically changed. The assumption cannot be made any more that there is a clear, strictly definable distinction between public and private information. As a growing body of empirical research shows:

> The degree to which information is thought to be accessible does not drive judgments about the appropriateness of accessing that information. … The immediate source of information matters to the perceived appropriateness of the data flows, even for information contained in public records. … Considering the respondents' strong judgments about the appropriate uses of information, the term "public data" may be not only inaccurate, but also misleading. The term "public" is often conflated with "not private" thereby leading policy makers to believe that individuals have no privacy concerns or expectations around the access and use of these public records. However, our study suggests the opposite. The data

> presented shows that individuals have deep concerns about who should have access to public records data and how it should be used.[3]

The relative value of information, its contextual meaning and sensitivity, are perceived differently in the open digital era and can have dramatically different consequences than under a previous information regime. The meaning, value, and significance of the documents in the care of the archive could undergo radical changes, depending on changes in the historical, political, and cultural context. For instance, until the dawn of the twenty-first century, one's gender was considered a nonsensitive item of public information, contained in every birth certificate. No longer: in a growing number of countries, individuals have the possibility and the right to choose their gender and to decide to keep that information (and identity) private or public. On the other hand, in some countries, one's sexual orientation, once a highly sensitive private item of information, has ceased to be a personal matter.

Around 1989, at the time of the political changes in Eastern and Central Europe, the archives of the former secret services were treated as depositories of denunciations, the repositories of lies, the material evidence of collaboration. Legislatures and archivists had to weigh the possible harm the accessibility of the obvious lies might cause to the individuals concerned, on the one hand, and the right of the public to get to know the real, until then secret, face of the previous regimes. In radical illiberal states, among them Russia, Poland, and Hungary, so called institutes of "remembrance and national memory," the official agents of historical revisionism, now use these records as reliable historical documents, giving credit to the allegations of the informers in order to denounce historical actors, former members of the democratic oppositions, and present adversaries. Sensitive documents, including medical records, information about past forcible psychiatric treatment (an often-used tool to isolate, lock up, and compromise the adversaries of the autocratic regimes) are now customarily made available to the public as information of genuine "public interest."

The change of the cultural milieu can lead to retroactive redescriptions of the past that, in turn, change the status of archival documents, and thus the way archivists and historians should handle them. Des Browne, the UK Secretary of State for Defence, announced in September 2006:

> The Government [plans] to seek parliamentary approval for a statutory pardon for service personnel executed for a range of disciplinary offences during the First World War. … Although this is a difficult issue it is right to recognize the exceptional

> circumstances that gave rise to these executions and to show compassion to the families who have had to live with the associated stigma over the years. …
>
> Rather than naming individuals, the amendment will pardon all those executed following conviction by court martial for a range of offences likely to have been strongly influenced by the stresses associated with this terrible war; this will include desertion, cowardice, mutiny and comparable offences committed during the period of hostilities from 4 August 1914 to 11 November 1918. Over 300 individuals from the UK, her dominions and colonies were executed under the 1881 Army Act. We will also seek pardons for those similarly executed under the provisions of the 1911 Indian Army Act. …[4]

The philosopher Ian Hacking, when commenting on a draft of the bill, a decade before it was finally passed by the British Parliament, asserted that "the author of the private member's bill states that today the men would be judged to be suffering from post-traumatic stress disorder and to be in need of psychiatric help not execution."[5] The new bill changed the status of both the dead and also the documents related to them: for about ninety years they had been treated as traitors and/or deserters, the documents of their story as part of military history, including legal documents of court martial procedures. As the law redescribed them as sick persons, victims of post-traumatic shock syndrome, the related documents should be treated (at least in part) as medical records, sensitive medical information, and handled as such in the archive. Different jurisdictions treat protected health information differently, providing privacy protection even for the dead for a varying period, sometimes well beyond the 50 years mandated under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule in the US.

As Hacking pointed out, the private member's bill had changed not only the status of the dead, but the status and perception of the surviving relatives, and the public at large. In the course of the Great War, court-martialed soldiers were described, treated, and stigmatized as traitors, and most probably the wider public saw them as such. Following the war, after the first literary reflections, such as Erich Maria Remarque's *All Quiet on the Western Front* (adapted to an Academy Award-winning film in 1930), Hemingway's *A Farewell to Arms* or Charles Yale Harrison's *Generals Die in Bed*, became available, the perception could have changed, and the executed soldiers might have turned into conscientious objectors, pacifists, who did the only thing one could expect of sane and courageous people. The law passed finally in 2006 in the British Parliament twisted the story one more time, and medicalized the conscientious objectors into sick individuals, who were not in

charge of their fate, whom the surviving relatives could not remember with a certain pride, but in the best case, with melancholy compassion. This is an instance of retroactive intervention in the past.

In 2012 a historian was confronted with a similar problem, although from the opposite angle of the private member's bill. Sydney Halpern was conducting research on federally funded human hepatitis experiments that ran in the US between 1942 and 1972:

> In the process, she has turned up names of many experimental subjects. Halpern had no intention of naming the vast majority of them, especially the mentally disabled and prisoners since they are now considered vulnerable populations. … Her problem was … what to do with the conscientious objectors during World War II who freely agreed to participate in experiments on hepatitis as an option for alternative service: "The COs weren't just research subjects. They were also historical actors making a statement. They were speaking through their actions … I think it's a mistake to apply a no-names convention without considering the situation of particular subjects. Leaving COs nameless robs them of a voice in the narrative—it silences them, and they wanted to be heard."[6]

* * *

In 2013, my archive, the Open Society Archives, one of the largest repositories of grave violations of human rights, received a letter from a Rwandan woman who was living in the US. Fearing deportation based on an archival description on our website, she demanded that her name be erased from the online finding aid. As part of our human-rights related film collection, our archive holds a copy of a short BBC documentary, *Rwanda, Master Conform*, directed by a British journalist, Lindsey Hilsum, who lived in Rwanda during the first weeks of the genocide.[7] She decided to return to Rwanda to investigate the fate of the people she once knew. The film features interviews with former acquaintances, some of them in an internment camp, among them a woman, who tells the reporter in French—subtitled in English—that she had been accused of having taken part in the genocide. The detailed archival description included both the names of the interviewees and a short summary of the interviews. The film was shown on the BBC. In the letter demanding the erasure of her name, the woman claimed that although she told the reporter that she had been accused of genocide, she was innocent, but now in danger of deportation from the US.

We knew that only a tiny minority of the perpetrators had been identified in Rwanda. We also knew that people with questionable pasts managed

to receive entry visas to the US, among them another woman who had received permission to enter the US; but when it was discovered that the Rwanda Gacaca Courts had convicted her for human rights violations in absentia, the US authorities deported this second woman back to Kigali in November 2011. Still, after careful consideration, the Archive decided to remove this woman's name from the description because archives, although custodians of information about the past, are not legal authorities, and thus cannot—when describing documents—judge or implicate individuals.

This was an unusual case: it was the subject herself, answering a question from the filmmaker, who stated that she had been accused of genocide. As Judge Posner of the United States Court of Appeals for the Seventh Circuit stated in a ruling in 1993, it is not easy to "bury the past" by claiming invasion of privacy when information comes from the public record.[8]

According to the UK's Rehabilitation of Offenders Act (1974), some, mostly relatively minor, criminal convictions can be ignored after a defined rehabilitation period.[9] Serious crimes, though, punished with over four years in prison—even according to the 2014 amendment of the Act—cannot be considered "spent," and thus cannot be erased from the records.

This Rehabilitation of Offenders Act has been considered one of the precursors of the so-called and now-prevalent "right to be forgotten." From the early 2000s, activists of strict privacy protection have been arguing for the "right to be forgotten" to be treated as a basic human right. Advocates of free speech, on the other hand, have reason to fear that a broad interpretation of the right might lead to suppression of free speech and to a widening censorship of the internet. In 2014, the Court of Justice of the European Union decided in one of its rulings that "if, following a search made on the basis of a person's name, the list of results displays a link to a web page which contains information on the person in question, that data subject may approach the operator directly and, where the operator does not grant his request, bring the matter before the competent authorities in order to obtain, under certain conditions, the removal of that link from the list of results."[10] Although the ruling invoked respect for private and family life, besides the requirements of protecting personal data, the decision of the court was widely interpreted as upholding the right to be forgotten, even without explicit reference to this right.

Indeed, according to the General Data Protection Regulation (GDPR) adopted by the European Union (and enforced since May 25, 2018), "data subjects" have the right to request erasure of personal data related to them

on certain defined grounds. The "right of erasure" is similar to but more limited than the right to be forgotten:

> Personal data must be erased immediately where … the data subject has withdrawn his consent and there is no other legal ground for processing, the data subject has objected and there are no overriding legitimate grounds for the processing. … The controller is therefore on the one hand automatically subject to statutory erasure obligations, and must, on the other hand, comply with the data subject's right to erasure. In addition, the right to be forgotten is found in Art. 17(2) of the GDPR.
>
> The right to be forgotten is not unreservedly guaranteed. It is limited especially when colliding with the right of freedom of expression and information. Other exceptions are if the processing of data which is subject to an erasure request is necessary to comply with legal obligations, for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes or for the defence of legal claims.[11]

Between May 2014 (the ruling of the Court of Justice in a case against Google) and March 2019, Google received more than 3 million erasure requests, and decided to remove 780,265 search results from its search engine.[12]

Based on the precedent established by the 2014 ruling of the Court of Justice of the European Union, a case is now pending in front of the Court in Luxembourg. In this case, the French data regulator is seeking to extend the right of state authorities to request so-called data controllers, such as Google, to erase information deemed unacceptable for state authorities. Extending the applicability and interpretation of the 2014 ruling, so as to include state actors, might have far-reaching consequences for freedom of information. Thomas Hughes, the executive director of Article 19, an NGO that monitors free speech, claimed that

> This case could see the right to be forgotten threatening global free speech. European data regulators should not be allowed to decide what internet users around the world find when they use a search engine. The [court] must limit the scope of the right to be forgotten in order to protect the right of internet users around the world to access information online. … If European regulators can tell Google to remove all references to a website, then it will be only a matter of time before countries like China, Russia and Saudi Arabia start to do the same. The [ECJ] should protect freedom of expression, not set a global precedent for censorship.[13]

* * *

The GDPR contains provisions related to archives, and provides certain exemptions and derogations in cases of personal data processed for

archiving purposes.[14] Still, as far their freely and globally available digitized documents are concerned, archives should be considered data controllers, for according to the definition of "data controller" under Article 4 of the Regulation: "controller means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data."[15]

Archives exist not only for collecting, storing, and preserving documents but also in order to make the documents available, retrievable, and usable for all those who—for whatever reason—decide to study, consult or scrutinize the documents deposited in the archive. Archives should thus provide retrievable access to the documents they keep. However, the way the documents can be accessed makes an important difference of type, rather than just of degree. Electronic copies of documents accessible on the website of the archive become available without control to the public at large, for anyone, without the mediation of a known (re)searcher, who could and should bear ethical and moral—not just legal—responsibility for the way personal data are made public in (print or electronic) publication. While it is in the public interest that (historically, socially, economically, legally, and so forth) relevant information—even that containing named, identifiable individuals—should become available, it is also in the public interest that archives should retain their status as trusted institutions.

Trust depends not only on the respect *des fonds*, the guarded and provable authenticity and integrity of the documents in the archival collection, but on the demonstrated care with which the archive also handles sensitive personal information. Records of the same provenance should not be mixed with documents of a different provenance, since without the context in which records were created, the original intention or meaning of the records would, supposedly, be lost. As Anne J. Gilliland-Swetland puts it: "the principle of provenance has two components: records of the same provenance should not be mixed with those of a different provenance, and the archivist should maintain the original order in which the records were created and kept. The latter is referred to as the principle of original order."[16] Trust springs from the assumption that the archive preserves the authentic documents, guarding their integrity, and would not "deaccession" or destroy them. It comes from an understanding that the archive makes such items retrievable but would not mishandle sensitive personal information either; that it would handle them in a legally and ethically foreseeable way.

In the spirit of its responsibility to the public, the Archive should make the documents, unrestricted by the donor, but containing sensitive personal information related to third parties and nonpublic figures, available on their premises, while exercising great care and discretion when making personal information openly and freely available on its websites. Archives are expected to engage in a never-ending balancing act between their responsibility to the public, which has the right to know, and to private individuals, who have the right to be protected.

My archive has two large Russian collections that demonstrate this dilemma: the so-called Red Archive of official reports by Soviet party and government sources, and the "Samizdat Archive," containing unofficial, underground documents produced by generations of anti-Soviet opposition. Documents in the "Red Archive" mention the name of a Russian psychiatrist, who, in the official sources "having betrayed his country," defected from the Soviet Union in order to live in the West. The name of the same person surfaced in samizdat publications, as one of those who had been engaged in the forcible psychiatric treatment of members of the opposition, and who having arrived in London as a self-styled critic of Soviet psychiatry, was offered a position at the famous Tavistock Clinic.

As it is the obligation of the Archive to preserve the integrity of the documents, it is unimaginable to redact the name in either of the collections. Whenever a researcher wants to consult one or both sources, the archive does not anonymize the documents. Being neither able nor inclined to judge the authenticity of the claim in any of the documents, the Archive does not and should not take a stand in the truthfulness of the sources.

Indeed, since we are the custodians of one of the largest propaganda archives in the world[17] our repository is obviously full of unsubstantiated claims, *ad hominem* accusations, and blatant lies about identifiable private citizens, not just public figures. The Cold War was fought with mutual lies and fantasies, the fabrications are the authentic sources of the times, as the title of a collection of essays on Cold War science says: *How Reason Almost Lost Its Mind*.[18] In lies there lies the truth.

The Archive is also the repository of forensic documents, testimonials, witness reports, the sources of which—victims, witnesses, accidental observers—could suffer retribution, even grave physical harm, were their identities made public. As we are an archive of both recent history and recent violations of human rights, tens of thousands of people implicated

in the documents under our care are still alive, among them victims and witnesses of mass rapes of Bosnian women or mass atrocities during the Balkan war in the 1990s. The Archive is obliged to protect not only the informational rights of private citizens but also the complete anonymity of legal and forensic sources.

There are in fact whole groups of archival documents in our repository, such as the antemortem questionnaires used in the course of the exhumation and identification of the victims of the Srebrenica massacre, that it would be ethically improper to make public, even in an anonymized form. Relatives can consult the documents, and researchers the anonymized sources—that contain sensitive personal information—but out of respect for the victims of the tragedy and their relatives, it would be unacceptable to make even the redacted documents public, or to upload them to the public web.

* * *

The authority of the archive as an institution traditionally rests on trust in the authenticity and integrity of the documents housed inside the walls of the archive, as well as trust in the integrity of the archivists, the custodians of the documents. From 1840 onward, the notion of archival integrity has been based on and connected to the principle of the chain of custody, the chronological documentation of the movement of the records, and the principle of provenance, which stipulates that records that originate from a common source are kept together, if not physically, at least intellectually with the help of the archival finding aids, in order to prove and to substantiate the authenticity and integrity of the records.

The archive, however, in the course of its daily routine of professional archival work endangers the authenticity and integrity of the documents; the archive could not exist without harming the integrity of the documents that it keeps. The institution that is supposed to guard the privacy and the information rights of people, especially of private persons, whose names and acts are recorded in the sources, contributes every single day to the violation of these rights.

Even in traditional archives, documents did not remain completely unaltered. Keepers of the archives, minor officials, monks, scribes, learned antiquarians copied, rescribed, translated, and annotated the documents. The Library of Alexandria, one of the first known archives—in Ptolemaic Alexandria, the librarian, "the guardian of the books" was considered to

be the "keeper of the archives"—contained tens of thousands of papyrus scrolls, a large number of which were confiscated from the ships in the harbor of the city and copied in the library, after which the copy was given back to the owner. In the course of copying the original, the text was frequently altered, involuntarily, as a mistake of the scribbler, or consciously in order to "improve" the original. The archivists or philologists ("the lovers of words") of the Ptolemaic museum were engaged in conserving, "rectifying," restoring a past (corpus) that had, supposedly, become altered, distorted, contaminated, or corrupted. In the words of the philologist Daniel Heller-Roazen, the practice, the guiding consideration, the figure of the library (of Alexandria), the notion of the library and the archive, demonstrates and stands for the understanding "of history as catastrophe."[19] The ongoing daily activity of the Archive is a heroic attempt to preserve or restore the presumed "the original," and to prevent the worst from happening: the flood, fire, invasion of mice or worms, sudden technological changes, digital decay, and so on, that make retrieval impossible.

Libraries and archives have been set up in order to collect under one roof, and thus preserve, otherwise dispersed texts: to prevent the disappearance and destruction of important records. The materiality of the documents has always been highly vulnerable: the majority of the papyrus scrolls of the Library of Alexandria most probably would have disappeared even without the fire that allegedly destroyed the library. Papyri survive more than two or three hundred years only in exceptional climatic circumstances, and even then, bugs and mice might finish off what the climate left intact. Papyri, like other manuscripts, had to be copied in order to be preserved, the corrected documents then often became reattributed, and named individuals in the copied documents reappear in new contexts with the possibility of their deeds being redescribed, thus posing new concerns for privacy.

Archives have never been completely immune from the suspicion of having forged documents in the interests of the archives, external authorities, or private individuals. Monastic archives in the West started with massive selective remembrance, by discarding documents deemed contrary to the interests of the monastery, or by producing fake documents to strengthen the spiritual, legal, or economic standing of the house. The forgeries implicated benefactors, legal heirs, dead or still alive, and their past deeds. Revisiting and rectifying the past was a double process of creation and destruction. In most cases, the original documents were destroyed in order to cover the

traces of alterations. The archive of the Abbey of St. Denis, which reaches "back to the dawn of institutional archival formation, was systematically pillaged and destroyed [already in the eleventh century] in order to build from its fragments a more useful and appropriate past," to make alternative interpretations inaccessible.[20]

As the documents in the archive have always been prone to both material and textual deterioration, they had to be moved, reshelved, reboxed, transcribed, altered, reattributed and, in consequence, recontextualized. With the emergence of digitization, however, dangers to authenticity and privacy became more pervasive. Digitization might affect the text and its readability as the yet far-from-perfect optical character recognition software cannot faithfully recognize the printed text, the manuscript or longhand. My archive has contracted unemployed Cambodians to fix digitized and OCR-ed text collections, but the nonnative, though highly conscientious, English readers came up with versions that barely resemble the originals.

Digitized information is always in movement: from one server to another, from one format to another, uploaded to the cloud and then copied, and stored on multiple servers. Cloud architectures necessitate the replication of data, which are in constant, automated movement from one location to another, without the consent or the knowledge of the administrator, the data specialist or the archivist.[21] Multiple storage locations increase the leakage of data, which could become public even without the malicious efforts of unfriendly hackers.

Archivists working in a digital environment are confronted, then, with the so-called Collingridge dilemma, named after the British academic, David Collingridge, who came to the conclusion that "when change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult and time consuming."[22] Archivists are not able to foresee the impact of technological changes on issues related to privacy. Had they been able to understand the future implications at the time when the new technologies were introduced, before they became embedded and widely distributed, there would then still have been a chance to take into consideration such concerns, and to modify the technology or its parameters. By the time the full impact of the new technology became apparent, however, it was too late: there are now strong corporate and/or political forces with vested interests in the insistence on keeping such profitable technologies, even when they have obvious high social costs.

Digitized archival documents could be connected to the holdings of other archives that store specialized data, placing the original documents and their subjects in a new and completely different frame. Descriptive documents can now be related to sensor or geospatial data, radio-frequency identification, social data to images obtained from surveillance cameras, and data originating from the Internet of Things. Billions of individuals voluntarily provide, share, and transmit data that finally end up on the servers of a few big data companies, state or private surveillance organizations. Relating and connecting archived records, and data coming from different—historical, social, commercial, surveillance—repositories, results in a deep layer of recursivity: the collectors or keepers of the original records are not able to predict where the aggregation of the data might lead. For, "when analysts can draw rules from the data of a small cohort of consenting individuals that generalize to an entire population, consent loses its practical import."[23] Indeed, information related to specific individuals that seems harmless from the perspective of the Archive, "may implicate others who happen to share … observable traits that correlate with the traits disclosed."[24]

* * *

Archives are institutions entrusted with the task of collecting and preserving records, even when recognizing that preservation and conservation endangers the very documents that the archive was meant to save for posterity. Archives are responsible for protecting the privacy and information rights of those mentioned and implicated in the documents; however, the archival workflow itself undermines the safeguards that are supposed to provide privacy protection. For a historian, some of the most important data are (or used to be until recently) the set of proper names, names of individuals, connected to certain events, since "sentences containing proper names can be used to make identity statements which convey factual and not merely linguistic information," as the philosopher of language John Searle stated.[25]

In a specific and limited sense, there is no difference between the natural sciences and the historical profession: both require experiments that can be repeated and then checked, verified, confirmed, or falsified using the same data.

Since the end of the 1960s, when Searle wrote his essay, the situation has changed: in the contemporary world, aggregated sets of metadata, including geospatial information, provide factual information on the basis

of which identity claims—even without mentioning the name—could be made. Still, "the thread of Ariadne that leads the researcher through the archival labyrinth is the same thread that distinguishes one individual from another in all societies known to us: the name."[26]

While, for data companies, specific information and traits are more important than proper names because personal identities can be reconstructed from cross-referenced data without knowing the name of the user (for Google, the personal name is just noise), historians go back to the archives, sources, and documents to find and check the names in order to analyze them one more time in a new context. Proper names are rigid designators (that is, in every possible world they designate the same person). If, in the effort to protect personal data privacy, archivists were to start erasing names, anonymizing documents, they would prevent historians from practicing their profession.

* * *

Archives are thus trusted custodians, appointed by the present on behalf of future generations, but functioning in such a way that fulfilling one part of their mandate—protecting privacy—would force the archive to delete larger and larger parts of its collection; to limit the period of data retention, to prevent connections between metadata sets, and in this way to make the work of the researchers more difficult and complex, or even impossible. Archives are trafficking in sensitive, dangerous material. Newly available digital technology, the ease and carelessness of voluntary, individual data production, the willingness of individuals to sell themselves by offering their data free to huge, nontransparent, data monopoly companies, in the business of targeted advertising or data mining ("if something is free it must be you that is being sold")[27] makes the archived material highly explosive. Surveillance and intelligence organizations, and obviously commercial data companies, are able—and willing—to collect all the data digitally produced by anyone, including archives. Although millions, even billions of individuals are voluntarily willing to share with the wider public even sensitive personal information on social networking sites, this does not absolve archives from their responsibilities as institutions of trust. Individuals with information kept in the archives have the right to expect trusted institutions to handle their information according to widely shared public norms, despite the private practices of the same individuals. Even in the midst of rapid technological change, archives cannot disregard the norms that

distinguish everyday practices from the responsibilities of trusted institutions. In order to guard the remaining and ever-shrinking authority and integrity of the institution, archives cannot open up all their secrets to the public at large on their websites. Public archives, or archives serving the public, should serve the interest of the citizens, both as members of the community and as private individuals.

Helen Nissenbaum, the American media scholar and privacy expert, is an advocate of the Principle of Respect for Context.[28] The Principle was included in the Obama administration's 2012 Privacy Bill of Rights as its third principle. That Bill of Rights, however, interpreted context specificity in a very limited way: with the naïve expectation that "companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data."[29] When consumers, companies, or archives make data openly available today, the future trajectory of the data remains unknown, and thus future contextual integrity cannot be guaranteed. As we are witnessing now, when consenting to disclosure of personal data we do not know the possible consequences of our consent: we cannot foresee the possible impact of interrelated media; we do not know in what ways data and attributes collected from others would disclose additional sensitive data about ourselves; or how a limited quantity of information would be amplified by the connected data sources.

Issues of privacy, according to the notion of contextual integrity, are not private, but social matters. In their practices, the Archive should consider both the interests and the preferences of all the affected parties, which include the public, present and future researchers, and nonpublic figures whose sensitive data the documents contain, and the archivists' control. Individuals have differing expectations about how their private data will be handled depending on the context: our expectations and behaviors at airport security are different from those we expect from a professional archive. Public interest archives are in the business of serving the public good by sustaining ethical, political, and scholarly principles, even when these principles might conflict with each other. Archives should be aware that they are expected to promote complex contextual functions, even when the different functions (promoting and enabling research, protecting sensitive information, transmitting historical knowledge but protecting the personal dignity of individuals) might be in competition with each other. Archives, where they exist as not-for-profit institutions, are in the position

to experiment with and demonstrate to commercial companies trafficking in data, context-specific substantive norms that constrain what information websites can collect, with whom they can share it, and under what conditions it can be shared.[30]

In *De Doctrina Christiana* Augustine wrote: "Because it is shameful [*flagitiose*] to strip the body naked at a banquet among the drunken and licentious, it does not follow that it is shameful [*flagitium*] to be naked in the baths. …" As the historian Carlo Ginzburg noted: "Augustine carefully traced a distinction between criminal *facinus* and shameful *flagitium*, the latter a sphere which, he insisted, had to be evaluated according to circumstances. We must, therefore, consider carefully what is suitable to times and places and persons, and not rashly charge men with sins [*flagitia*]."[31] Since privacy is a complex non-private issue, archives should think twice and act in a careful, differentiated way, taking the needs of context specificity into consideration before making archival documents openly accessible. This has been an issue for all of history, ever since we kept archives, but it is an especially complicated quandary in our open, digital era, when even public information, when placed, analyzed, aggregated, and used in a new context for previously unforeseen purposes, can have sometimes seriously harmful private consequences.

Notes

1. I here and throughout use the capitalized form of Archive to refer to the idealized instantiation, rather than any concrete, actually existing space.

2. See: Aleida Assmann, "Canon and Archive," in *Media and Cultural Memory*, ed. Astrid Erll and Ansgar Nünning (Berlin: Walter de Gruyter, 2008), 97–108, https://kops.uni-konstanz.de/handle/123456789/13382; Anthea Josias, "Toward an Understanding of Archives as a Feature of Collective Memory," *Archival Science* 11, no. 1 (2011): 95–112, https://doi.org/10.1007/s10502-011-9136-3; Marianne Hirsch and Leo Spitzer, "The Witness in the Archive: Holocaust Studies/Memory Studies," *Memory Studies* 2, no. 2 (2009): 151–170, https://doi.org/10.1177/1750698008102050; Michelle Caswell, "Khmer Rouge Archives: Accountability, Truth, and Memory in Cambodia," *Archival Science* 10, no. 1 (2010): 25–44, https://doi.org/10.1007/s10502-010-9114-1; Ann Laura Stoler, *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense* (Princeton, NJ: Princeton University Press, 2009); Diana Taylor, *The Archive and the Repertoire: Performing Cultural Memory in the Americas* (Durham, NC: Duke University Press, 2003).

3. Kirsten Martin and Helen Nissenbaum, "Privacy Interests in Public Records: An Empirical Investigation," *Harvard Journal of Law & Technology* 31, no. 1 (2017): 116, 141.

4. Des Browne, "House of Commons Hansard Ministerial Statement," UK Parliament, September 18, 2006, https://publications.parliament.uk/pa/cm200506/cmhansrd/vo060918/wmstext/60918m0187.htm.

5. Ian Hacking, *Rewriting the Soul: Multiple Personality and the Sciences of Memory* (Princeton, NJ: Princeton University Press, 1998), 241.

6. Susan C. Lawrence, *Privacy and the Past: Research, Law, Archives, Ethics* (New Brunswick, NJ: Rutgers University Press, 2016), 107–108.

7. Lindsey Hilsum, "Rwanda, Master Conform" (BBC, October 30, 1996), Box 374, Videocassette RW038, International Monitor Institute. Rwanda Videotapes and Audiotapes, David M. Rubenstein Rare Book & Manuscript Library, Duke Universities.

8. Quoted by Lawrence, *Privacy and the Past*, 59.

9. "Rehabilitation of Offenders Act 1974," The National Archives Legislation, 1974, https://www.legislation.gov.uk/ukpga/1974/53.

10. Court of Justice of the European Union, "Judgment in Case C-131/12 Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González," May 13, 2014.

11. Intersoft Consulting, "Right to Be Forgotten," *General Data Protection Regulation (GDPR)* (blog), accessed April 29, 2019, https://gdpr-info.eu/issues/right-to-be-forgotten/.

12. Google, "Search Removals under European Privacy Law," Google Transparency Report, 2019, https://transparencyreport.google.com/eu-privacy/overview.

13. Owen Bowcott, "'Right to Be Forgotten' Could Threaten Global Free Speech, Say NGOs," *The Guardian*, September 9, 2018, sec. Technology, https://www.theguardian.com/technology/2018/sep/09/right-to-be-forgotten-could-threaten-global-free-speech-say-ngos.

14. See, among other provisions: Under Article 9: 1. "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited. 2. Paragraph 1 shall not apply if one of the following applies : … (j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes … Article 89: Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes …" The European Parliament, "Regulation (EU) 2016/679 of The European Parliament and of The Council," European Union Law, April 27, 2016, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02016R0679-20160504.

15. The European Parliament, "Regulation (EU) 2016/679."

16. Anne J. Gilliland-Swetland, "Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment" (CLIR, 2000), 12.

17. We house the former archive of the Research Institute of Radio Free Europe/Radio Liberty, perhaps the most important propaganda organization in the Cold War era, and also the propaganda materials of the former Communist countries.

18. Paul Erickson et al., *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality* (Chicago: University of Chicago Press, 2015).

19. See Daniel Heller-Roazen, "Tradition's Destruction: On the Library of Alexandria," *October* 100 (2002): 133–153.

20. Patrick J. Geary, *Phantoms of Remembrance: Memory and Oblivion at the End of the First Millennium* (Princeton, NJ: Princeton University Press, 1996), 107.

21. Copying entire digital collections seems to be a reasonable foresight today. The LOCKSS (Lots of Copies Keep Stuff Safe) Program at the Stanford University Library developed and provides open source tools for libraries and archives to copy, and thus to preserve their content. COAR, the Confederation of Open Access Repositories keeps multiple copies of the collections of its members. The Internet Archive, based in San Francisco—as its storage is in constant danger of destruction, since the Archive sits literally on top of the San Andreas Fault—set up a mirror site in the new Library of Alexandria. Following the November 2016 US election, the Internet Archive, which held, as of October, 2016, 273 billion webpages from over 510 billion web objects, and grows by over 500 million webpages a week, taking up 15 petabytes of storage, decided to move its backup data to Canada, in order "to keep the Archive free, accessible and reader private." Brewster Kahle, "Help Us Keep the Archive Free, Accessible, and Reader Private," *Internet Archive Blogs*, November 29, 2016, https://blog.archive.org/2016/11/29/help-us-keep-the-archive-free-accessible-and-private/. See also Tung-Hui Hu, *A Prehistory of the Cloud* (Cambridge, MA: The MIT Press, 2015) for a set of theoretical provocations around cloud infrastructures.

22. David Collingridge, *The Social Control of Technology* (New York: St. Martin's Press, 1980), 11.

23. Solon Barocas and Helen Nissenbaum, "Big Data's End Run around Procedural Privacy Protections," *Communications of the ACM* 57, no. 11 (2014): 32, https://doi.org/10.1145/2668897.

24. Barocas and Nissenbaum, "Big Data's End Run," 32.

25. John R. Searle, *Speech Acts: An Essay in the Philosophy of Language* (Cambridge: Cambridge University Press, 1969), 165.

26. Carlo Ginzburg and Carlo Poni, "The Name and the Game: Unequal Exchange and the Historiographic Marketplace," in *Microhistory and the Lost Peoples of Europe*,

ed. Edward Muir and Guido Ruggiero (Baltimore, MD: Johns Hopkins University Press, 1991), 5.

27. Tim Worstall, "Facebook Is Free Therefore It Is You Getting Sold," Forbes, November 10, 2012, https://www.forbes.com/sites/timworstall/2012/11/10/facebook-is-free-therefore-it-is-you-getting-sold/.

28. See Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford, CA: Stanford University Press, 2009).

29. Quoted in Helen Nissenbaum, "Respecting Context to Protect Privacy: Why Meaning Matters," *Science and Engineering Ethics* 24, no. 3 (2018): 834, https://doi.org/10.1007/s11948-015-9674-9.

30. See Helen Nissenbaum, "A Contextual Approach to Privacy Online," *Daedalus* 140, no. 4 (2011): 32.

31. Quoted in Carlo Ginzburg, "The Bond of Shame," in *Passionen. Objekte—Schauplätze—Denkstile*, ed. Corina Caduff, Anne-Kathrin Reulecke, and Ulrike Vedder (Munich: Wilhelm Fink, 2010), 24, http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/year/2017/docId/44333.