Harvard University Privacy Principles Companion Guide

Last revised: January 9, 2024

Harvard strives to be a trustworthy steward of personal information. Our Privacy Principles establish a Harvard-wide framework for considering and applying a privacy-protective mindset to the work that we do at the University.

The strategy behind these Principles is threefold: 1) promote a culture that values privacy, 2) create a foundation for operationalizing privacy at Harvard, and 3) satisfy existing and anticipated regulatory compliance obligations.

The collection, use, and disclosure of personal information are unavoidable in and essential to Harvard's operations and its teaching and research mission. At the same time while doing this critical work, we need to consider privacy protections. That's where these Principles come into play. The Privacy Principles aren't hard and fast rules but rather aspirational values that we should apply when handling personal information during the course of our work activities.

This Companion Guide is intended to provide additional context and specificity to assist us in the application of the Principles. Privacy concerns should always be weighed against other University requirements and goals.

We recognize that not all of the Privacy Principles can be achieved in all situations. For example, we regularly use and share personal information in connection with our work at the University, often informally and in non-systematic ways. A teaching fellow may communicate information about a student to a faculty member, or administrators may exchange emails about an employee. By contrast, an information system used for administration at the University may generate, store, and make accessible a large data set containing information about a large number of persons—and indeed academic researchers at Harvard may obtain access to large data sets of personal information. While the Privacy Principles can inform the use of personal information at both small (one-off communication) and large (system-generated data sets) scales, the Principles will be applied differently, depending on the form and context of the information at issue.

Finally, neither the Privacy Principles, nor this Companion Guide, create any contractual or other legal obligation on Harvard's part, or any contractual or other legal right or expectation in or for any individual person.

What is the origin of the Principles? While these Principles are tailored specifically to Harvard's needs (now and future) and to anticipated regulatory developments, they are based on the internationally recognized Fair Information Practice Principles (FIPPs) that were developed in the 1970s and provide the core values underlying many federal, state and international privacy laws.

What is personal information? Personal information is any data that relates to an identifiable individual person, including their character traits, history, and activities. Personal information not only includes name, address, and other direct identifiers, but also indirect information such as purchasing history, Internet activity, and, in some cases, information about personal activities, such as resource consumption.

Owing to the ever-expanding availability of large data sets of personal information and activity logs, it has become possible to re-identify individuals from seemingly anonymized data sets. A data set may on its own be fully anonymized, in that it contains no personal identifiers, but when paired with outside information, including publicly available data, it may well be possible to identify individual subjects in the original set. Accordingly, if a piece of information could tell you something about a person, even if you need (but do not yet have) additional information to "unlock" who it is, it may be appropriate to treat it as personal information.

Do the Principles apply to all forms of personal information (or just electronic)? The Privacy Principles apply to all forms of personal information collected, stored, and processed by Harvard, including in paper or electronic form.

Do the Principles apply to all types of personal information? Yes, but of course some categories of information are necessarily more sensitive than others. For example, application of the Privacy Principles will be more important in cases involving health information, financial information, information that could be misused (such as for identity theft or surveillance), and other information that is generally accepted as or obviously private to or about an individual.

How do the Principles relate to existing Harvard policies? The Privacy Principles are not policy but instead a set of principles to consider and apply in appropriate circumstances weighing privacy against other institutional goals.

The Privacy Principles do reflect and are consistent with the conceptual underpinnings of several existing Harvard policies, including:

- Policy on Access to Electronic Information
- Policy on Installation and Use of Video Cameras
- IT Professional Code of Conduct to Protect Electronic Information

What do the Privacy Principles mean for me? How does this affect me and my work? Each of us should be mindful about how we manage and use personal information that is entrusted to us and in the systems we use and create. These Principles can inform and guide our decisions and actions as trustworthy stewards, providing a touchstone for bringing privacy to life across the breadth and depth of Harvard's activities.

How do I apply these Principles? The overarching goal is to infuse privacy considerations into all that we do at Harvard. The Principles are therefore intentionally general and succinct. When you are considering the collection, use, or disclosure of personal information, we ask that you think about the Principles in their totality, considering their purpose and spirit, and then apply your best judgment. If you are uncertain or have questions, contact the Information Security and Data Privacy (ISDP) team for advice.

- If you are establishing new business operations, research activities, technologies, or other processes involving personal information, consider whether you can incorporate the relevant principles into your system design. When we consider privacy from the earliest stages and throughout the data life cycle (*i.e.*, collection, use, retention, processing, disclosure, and destruction), we are best positioned to implement privacy protections.
- If you aim to introduce privacy protections to existing business operations, research activities, technologies or other processes involving personal information, review your data practices for alignment with the Principles. Where the practices don't align with the relevant Principles, and where practical, adjust your practices to bring them into alignment.
- In all cases, consider the principle of data minimization, as this is highly effective in reducing risks associated with privacy incidents.

What does each of these Principles mean? Below is additional context for each of the Privacy Principles to help you understand and, where appropriate, implement them.

Transparency: *Before collecting personal information, provide a notice that clearly and simply describes how Harvard plans to use it, including the specific purposes for collection. Respond candidly to questions from individuals regarding the collection and use of their personal information.*

Openness about how an individual's personal information is collected, analyzed, and used is a core privacy value that can often be accomplished through publication of a privacy statement. In most cases, privacy statements are not legally required. However, you should consider whether to offer a privacy statement, particularly in cases where personal data collected by applications or on websites may be used in ways that the individual may not expect and/or when the data collected relates to a population to which you feel you have a special obligation or relationship. If you decide to publish a privacy statement, it should be accurate, drafted in a clear and human-centric way and should be reviewed and revised annually (or any time your practices change) to ensure it remains accurate.

Ask yourself the following questions:

- Are the individuals aware that you are collecting their information, and do they understand how it will be used? If you are not providing notice, is the collection and use of the information consistent with ordinary or accepted practices for the treatment of data?
- If you are providing a notification, is it written in a way that is concise and readily understandable by the reader?
- Regarding privacy statements: Is my site or application collecting or processing data in a way that triggers a legal requirement to publish a privacy statement (contact your school GDPR

Implementation Coordinator or School PrivSec Officer for guidance)? Or absent any legal requirement, is my site or application directed toward the Harvard community and essential to university operations, or would a privacy statement be necessary to engender the level of trust needed to attract a critical mass of users to the site or application? Do I need to contact the ISDP team to help create a draft? Do I have a process in place to review the statement to ensure it remains accurate over time?

- Do I have a process for responding to individuals' inquiries into the use of their personal information, including authentication processes? Direct inquiries from non-US individuals to EEADataSubjectRequest@harvard.edu.

Minimum Necessary: *Limit the collection of, access to, and use of personal information to the minimum that is directly relevant and necessary to accomplish a legitimate institutional purpose.*

Prevent privacy problems before they start by not collecting personal data that you don't need, ensuring it is accessible to (and shared with) only those who need it, and deleting it when you no longer have a good reason to keep it. Data has the propensity to spread and persist, so don't share it among systems without a good reason. Avoid making unnecessary copies of data, such as by downloading copies to your laptop computer, and instead work with the data in the source system environment.

Ask yourself the following questions:

- Looking at each of the personal data elements, can we remove any and still fulfill the purpose?
- Who has access to the data? How can we limit the scope and time window of access?
- How can I avoid making copies of the data that do not serve a legitimate business need, especially onto other systems?
- Do I have a process for periodically reviewing who has access to the data?
- Have I considered the purpose of my collection and use of the data, and whether it is defensible in the context of Harvard's mission?

De-Identification: *To the extent practical, remove personal identifiers or use aggregation, pseudonymization, or other anonymization methodologies.*

Many data collected by Harvard will need to remain identifiable. However, particularly in the research context, some datasets may have more personally identifiable data than necessary. While it's critical to our teaching and research mission to make data available to our students and scholars, we should do so in privacy-sensitive way. De-identification is one way to accomplish that goal. We should seek to use the appropriate methodology, taking into account the risks both to the data subject and to the University's objectives, to protect that data about specific individuals from being re-identified or reconstructed.

It is important to note that no de-identification/ anonymization technique is 100% foolproof. There is always a risk of re-identification, whether by a persistent malicious attacker or by others who have

access to related data and the necessary computational resources.[1] On the flip side, careful consideration should be applied prior to re-identifying data, including the fact that you are generating personally identifiable information.

Different techniques yield different results with varying degrees of residual risk of re-identification. There is of course a tradeoff between usefulness of data and how much anonymization can be achieved: a data set anonymized past the point of reidentifiability ultimately might not contain enough information in it to provide any insight to researchers. It is therefore essential to consider how best to balance the competing values of (1) maintaining information that is useful for permissible and appropriate objectives and (2) reducing the susceptibility of this information to reidentification and subsequent misuse. An appropriately balanced de-identification program may simply make it more difficult for bad actors to reidentify information, rather than reduce to zero the possibility that it could be done. And of course, application of other Principles, including "Minimum Necessary," "Limited Sharing," and "Security Controls" help to reduce the risks of working with identified or identifiable data.

If you have determined that de-identification is appropriate after weighing privacy and other valid purposes, then ask yourself the following questions:

- Have I consulted the National Institute of Standards and Technology (NIST) guidance on the De-Identification of Personal Information?
- Have I de-identified the data to the greatest extent possible?
- Have I considered employing differential privacy tools such as Harvard's OpenDP.org?
- Is there an automated tool I could use? Should I use an external service?
- Have I assessed the probability of re-identification (given access to other readily available datasets)?

Responsible Use: *Use personal information only for the purposes for which it was collected, with the consent of the individual, or as required by law.*

Once we have collected personal information, the original purpose for the collection should remain attached to the data, and to the extent practical, we should remain true to that original purpose absent amended consent or legal requirement.

This issue often arises from a desire to use an existing research data set for a secondary, unrelated project. Access to a data set does not imply license to use it for any purpose. Before using data for a secondary purpose, consider the questions below.

Ask yourself the following questions, at the point of collecting information:

- Do we anticipate, or is it possible that, we might use the information for other purposes down the line?

[1] As an example, using public anonymous data from the 1990 census, Harvard professor Latanya Sweeney found that 87 percent of the population in the United States, 216 million of 248 million, could likely be uniquely identified by their five-digit ZIP code, combined with their gender and date of birth. About half of the U.S. population is likely identifiable by gender, date of birth and the city, town, or municipality in which the person resides. Expanding the geographic scope to an entire county reduces the reidentification probability to a still-significant 18 percent.

- Can we provide a more general statement of the purpose for which we're using the information that may cover subsequent uses, while still providing meaningful notice to the data subject?

Ask yourself the following questions, at the point of repurposing information:

- What, if anything, were the individuals originally told regarding how their data would be used?
- Is the contemplated use consistent with the original intended purpose? Would the individuals reasonably expect us to use the data for this purpose?
- Have we previously represented that we would not use information in this way or for this purpose?
- What would individuals think about how their data is being used?
- Is the desired usage consistent with existing regulations?
- Can I de-identify the data?
- Is there a tool such as OneTrust that can be used to manage the consents?

Limited Sharing: *Share personal information with third parties only where consistent with applicable regulatory and contractual requirements and when adequate privacy and security controls are in place.*

Oftentimes there may be a need to share personal information with third parties, including external research collaborators or vendors. In these circumstances it is important to apply the "minimum necessary" principle of sharing no more information than is needed, to no more persons than are needed, and for no longer than is needed to accomplish the objective. We should also ensure that regulatory and policy requirements are met, that appropriate contractual language is in place, that appropriate security protections are in place, and that the sharing does not conflict with reasonable expectations individuals may have regarding the use of their information.

Ask yourself the following questions:

- Is there a way to minimize the information shared while still accomplishing the desired objective?
- Are we sharing, disclosing or repurposing the information in a way that the individual(s) would not reasonably expect?
- Have we previously represented that we would not share information in this way or for this purpose?
- Have we worked with Strategic Procurement to negotiate privacy-preserving terms where possible?
- Does our agreement with the third party include the appropriate Harvard privacy riders?
- If the data are classified as Level 4+, has a Vendor Risk Assessment been conducted by the Information Security and Data Privacy Office to assess the third party's privacy and security controls?

Choice and Control: *To the extent practical and when doing so would not impair important institutional objectives, give individuals explicit choice and control as to how their personal information will be used, disclosed, and/or deleted.*

Once personal information has been collected, to the extent practical and consistent with Harvard's legitimate objectives, we want to provide agency to the subjects of the information regarding the continued use or storage of their data. An example of this is a request to delete one's records or prohibit or halt the use of their data for marketing purposes.

Ask yourself the following questions:

- How can I design Harvard systems to facilitate the fulfillment of this type of request?
- Have I verified the requestor's identity before acting on the request?
- Am I required to respond to the request pursuant to a regulatory requirement such as GPDR?
- Would fulfilling the request be contrary to legal requirements or other institutional reasons to retain this information?

Stewardship: *For each dataset containing personal information, designate a data owner to be responsible for ensuring that these principles are adopted, that regulatory and contractual obligations are met, that data are accurate, and for responding to questions and concerns regarding its use.*

Privacy is about trust, and we want Harvard to be a trustworthy steward of personal information. When planning a new system, managing an existing system, or handling research data, there should be a person assigned who understands the data and how it flows. The steward is responsible for the privacy of the data, ensuring that these privacy principles are infused throughout the full data life cycle.

Ask yourself the following questions:

- Is someone who has expertise about the data assigned the functional role of responsibility for the dataset? A unit or team may have multiple and different systems and may need multiple data stewards.
- Does the steward understand the data mapping and flow (how data comes in and how it moves once it is there) and the accompanying privacy risks?
- What needs to change in order to better adhere to these privacy principles so that privacy safeguards are attached to the data as it flows?
- Have those with access to the data been trained regarding these privacy principles?
- Are individuals' inquiries regarding their data being responded to in a timely fashion?

Security Controls: *Ensure that Harvard's Information Security policy is followed for systems that store, process, or transmit personal information.*

Privacy relies heavily on security—protecting the data from unauthorized access. We want Harvard to be a trustworthy steward of personal information, and nothing erodes trust faster than a data breach.

Ask yourself the following questions:

- Have I identified which data security level (DSL) applies to these data?
- Am I complying with Harvard's information security policies based on the security level?
- Do I understand the cybersecurity risks to these data? How have I managed those risks?

- Am I reviewing access to the data periodically to ensure the minimum necessary?

Retention and Deletion: *Retain or archive personal information only as long as needed (using Harvard's General Records Schedule as a guide) or as required by law or agreement. Securely delete personal information when no longer needed.*

Data is the lifeblood of Harvard's research, education, and administrative operations, but personal data shouldn't be kept longer than necessary. By getting rid of personal data when it's no longer needed, we reduce the volume of sensitive data at risk of attack, unauthorized access, and other privacy-related risks. A large percentage of privacy incidents involve personal data that are no longer needed. The less personal data we have, the better we can protect privacy.

To get started, consult the University's General Records Schedule (GRS) for policy requirements for legal, administrative, and long-term retention needs of the University, and contact the Records Management Services team for help drafting a retention plan. Be sure to ask the RMS team whether any of the records subject to deletion could be relevant to a legal claim or defense. RMS will consult with the Office of General Counsel to ensure that the University's legal interests are appropriately reflected in the plan. The GRS is Harvard's policy on how long to keep different types of records, and whether to send records to an archive or destroy them after the retention period. Learn GRS basics.

Note that there may be good reasons to retain certain data indefinitely—for example, so that the University can show it has complied with law or defend itself against legal challenge, for business continuity purposes, or for reference in support of consistency of practice. In such cases, privacy values may be supported by taking the intermediate step of moving data into long-term archival storage.

Ask yourself the following questions:

- Do I understand the retention requirements of these data and what's next?
- Have I consulted the Records Management Services team for input and guidance?
- Have I taken a Records Management workshop to better understand my retention needs?
- Is there a process for purging or archiving the data once the retention period concludes?
- Are there intermediate steps available to protect the data, short of deletion, such as reducing the user base with access to the data, removing the data from networks and/or placing the data in off-site storage? Is a phased obsolescence plan a possibility?
- If I need to retain data long-term for legal, reference, record-keeping, or institutional continuity purposes, are there offline or other access-restricted resources where I can store it, rather than on live systems for everyday access or use?
- Do I know how to securely delete the data if appropriate (and watch for eWaste events)?

What should I do if I become aware of or experience a privacy incident? *Please contact the Information Security and Data Private team immediately if you experience or are aware of any of the following*:

- Unauthorized access to personal information

- A confirmed or suspected system intrusion
- Mishandling of personal information that may put it at risk
- Lost or stolen device that contains or has access to Harvard data

If you receive a suspicious email that might be phishing, forward to phishing@harvard.edu.