MIT Libraries Patron Data Privacy Policy

Date: 9/28/2020 **Version 1.2**	**Prepared By:** **Katie Zimmerman** **Matt Bernhardt**

Table of Contents

Revision History

Version	Version/Update Description	Author	Date
0.1	Initial working meeting to define outline and assigned sections and begin to data dump	M. Bernhardt C. Crummett K. Zimmerman	10/16/19
0.2	Working meeting to review work to date, address questions, clean up and consolidate, assign next steps	M. Bernhardt C. Crummett K. Zimmerman	11/8/19
0.3	Systematic revisions for internal cohesion, removing some notes, some additional drafting	KZ	11/25/19-11/27/19
0.4	Document clean up and formatting	CR	12/10/19
0.5	Working meeting to finalize draft	CR, KZ, MB, CQ	1/14/20
0.6	Incorporating H.B, C.C, K.S, feedback, finalizing open areas and addressing open comments	K.Z, M.B	1/21/20
0.7	Incorporating feedback group, A.N., M.A, and ADs feedback	KZ, MB	2/11/20
0.8	Incorporating feedback from MIT Libraries FCLS	K.Z.	3/9/20
1.0	Incorporated final comments from SRLT and received approval.	K.Z	6/8/20
1.1	Added exception to 6.1 for VPF financial information records retention; added update to section 9.	KZ	8/18/20
1.2	Added definition of “active use” Removed Appendix B: Language Included in Vendor Contracts Removed Section 9.1 Opting in to Data Collection	KZ	9/28/20

1 Introduction

MIT Libraries is committed to protecting your privacy and confidentiality when you visit or use services we provide. This privacy policy details how MIT Libraries handles the personally identifiable information (PII) of MIT students, faculty, staff, and other patrons of the MIT Libraries. PII is any piece of information that can be used by itself or linked with other information to identify an individual (a more full description of what we mean by PII is available here). This policy covers all aspects of the services offered to the MIT community by the MIT Libraries, including web platforms, online and print resources.

Many interactions with the Libraries or library resources result in data about you being recorded. This policy will let you know what may be collected, how it is used and protected, and when it may be shared. We try to collect only the information which is necessary to provide you with library services, and will give you as many options as possible to control your own data.

We do our best to safeguard your privacy as a user of our systems; however, there are also steps you can take to minimize the creation of personally identifiable records in your general online interactions. To learn more about your privacy generally, and find tools to help safeguard it, these resources are a good place to start: Library Freedom Project, Electronic Frontier Foundation, Cookies and You. The Libraries also maintain visitor computers with access to most library resources, which can provide a more anonymous access option.

Your privacy at MIT is also covered by the MIT Privacy Policy, which this policy supplements, and MIT Policy 11.0 Privacy and Disclosure of Personal Information, Legal protection of your privacy comes from a variety of laws (some of which, based on your citizenship, place of residence, and the type of information involved, may apply to your interactions with the Libraries) including: FERPA, HIPAA, Massachusetts regulations and data breach law, the EU General Data Protection Regulation, and data breach regulations of other US states. The MIT Libraries' privacy practices and this policy have also been informed by many library standards and guides including those of the American Library Association, International Federation of Library Associations, and the National Information Standards Organization. We encourage you to learn about and exercise your privacy rights!

2 Data Collection

The MIT Libraries tries to minimize the amount of personal information necessary to use library services. Many library services can be accessed without personal information being collected.

The patron data that the MIT Libraries do collect is classified using guidelines published by MIT's Written Information Security Program (WISP) and informed by MIT Risk Management and Compliance guidelines, which determine how long it is retained and how it is stored. This classification places systems into categories of low, medium, or high risk. Information about this program, the risk categories, and required security steps can be found at infoprotect.mit.edu. Low risk data is data which is already public or which wouldn't be harmful if released. High risk data is data which is subject to legal requirements or would cause serious safety, financial, or operational harm if released, and medium risk data is in between. The Libraries treat all information which identifies the intellectual pursuits of students - such as search logs and reference questions - as high risk, consistent with the MIT FERPA policy. Any information related to unpublished research as well as identifying information paired with an MIT ID is generally categorized as medium risk or higher. Additional guidance on risk determinations for specific information comes from MIT Risk Management.

Sections 2.1 through 2.5 below describe the sources from which we receive information about patrons. Generally, information at all risk levels may be received through all of these means.

2.1 Data provided directly by the patron

Patrons provide information to the Libraries when they hand over their MIT ID to check out materials, search for materials via a web-based discovery tool like the Barton catalog or BartonPlus, submit a web form, interact with Libraries staff directly or by chat or email, or attend Libraries events.

2.2 Data provided by the software used by the patron

The tools used by patrons when using Library services provide certain information without the patron's intervention. An example of this type of information includes a computer's IP address, and the user agent string identifying the patron's web browser. This collection can occur when a patron accesses a website or web application maintained by the Libraries.

2.3 Data provided by other systems

The Libraries receive information about patrons from other systems. This includes records from MIT, information received from Touchstone (if a user logs into a library website), and payment transaction IDs from MIT's payment processor.

We also receive aggregate data about the use of library resources when those resources are provided by third-party vendors, such as the number of total visitors to a third-party service or total download counts of specific content. This data is provided in aggregate and, wherever possible, in accordance with the Project COUNTER Code of Practice, which does not allow us to identify individual users.

2.4 Data generated by the Libraries

When a patron uses resources such as the Barton catalog or the QuickSearch application, a unique identifier is generated for use by that application which is used during that particular access session. Use of library-provided computers and equipment, for example in the GIS Lab or at library access stations, may also generate use logs.

Security cameras are in place in certain locations within the libraries, including the 24x7 spaces in Barker, Dewey and Hayden Libraries, and the Distinctive Collections Reading Room. These cameras are in place for the safety and security of library users, collections, and equipment. Security camera data is only viewable by MIT Police. It is kept as long as MIT Police determine it is needed and MIT Police make the decision if footage should be reviewed or shared.

Gate counters are installed at each library entrance and 24x7 entrance to count aggregate visits, and entry to 24x7 spaces requires use of an MIT ID. Gate counters do not capture images of individuals but rather count the “people like shapes” passing through. MIT Physical Security manages access to this data and provides the MIT Libraries with reports of total entries per day and time, on a monthly basis.

2.5 Data collection by third parties

While accessing library services, you may end up leaving the websites maintained by the MIT Libraries and instead visiting sites maintained by third parties - for example, when you access subscribed journals or ebooks hosted on a publisher’s website. Those websites may collect information about you, including through required or optional registration on their site, and will be governed by their own privacy policies. The MIT Libraries try to negotiate patron-friendly policies consistent with this privacy policy as part of their contracts with these third parties, so you may have privacy rights beyond those in their stated privacy policy. If you have questions or concerns about a particular vendor’s privacy policy, you can contact us at library-privacy@mit.edu.

Additional examples of third-party systems present in Libraries’ services include:

- Loading commonly used web content (such as fonts, icons, or javascript libraries) from a content delivery network.
- Submitting payment information to MIT’s designated payment service.

2.6 Public access computers

The Libraries have public access computers in Hayden, Barker, Dewey, Rotch and Lewis Music libraries, as well as the Department of Distinctive Collections Reading Room. Some computers are Athena Clusters maintained by MIT IS&T, and require a Kerberos ID for access. Additional computers provide access for library patrons unaffiliated with MIT (Open

Access computers) and at quick lookup kiosks; these computers do not require authentication, but have access to a limited subset of library electronic resources (based on our license agreements for that content). All of our public access computers delete user data daily, and data specific to your browsing session is deleted either when you close the internet browser (Open Access computers and quick lookup kiosks), or according to your account settings (when authentication is required).

If you are particularly privacy-sensitive, using the public access computers can provide an additional layer of anonymity to your use. Browsing activity on personal devices can sometimes be linked to you, even when you haven't volunteered your personal information. Using a public computer reduces that risk, and you can consider further anonymizing yourself by using privacy-protective tools and practices.

3 Who at MIT has access to the data we collect

User data categorized as high risk is only accessible to staff who need access to the information in order to perform the function for which it was collected - for example, if you contact the libraries for research help, library staff will use the information you provide in order to answer your questions and provide you with library resources, but will not disclose your research questions to anyone else. This includes others at MIT - so, for example, we will not share research questions identifiable to you with faculty and staff elsewhere at the Institute without your prior consent.

Data which is higher risk when directly associated with an individual user may become lower risk once de-identified (Section 7.3 describes our de-identification practices), and handled accordingly. For example, de-identified records of reference questions are retained for continued staff use (assessment of our services, identifying frequent questions in order to develop additional resources, staff training) and are then treated as medium risk data. Medium risk data may be accessible to all Libraries staff, or may be limited to specific individuals or groups as appropriate.

Low risk data may be reasonably disclosed publicly by the Libraries. Low risk data includes information made public by the originator (for example submissions to an idea bank), and could also include aggregated, de-identified information, such as library statistics we report to the Institute. In general, even if the data is low risk, it will be de-identified before public release, unless you've otherwise indicated that you are ok with public identification with the data. Low risk data may also be handled as if it were higher risk if stored alongside higher risk data.

Some Libraries records are required to be maintained as a permanent record of the Institute for insurance and security purposes. These records are only accessible to staff who need access for a legitimate purpose, as authorized by the head of the Libraries unit responsible for those records or if required by law. For example, access logs for the Distinctive Collections Reading Room are retained as permanent records in accordance with ACRL/RBMS guidelines and accepted archival practice. The permanent records of many MIT departments, labs, and centers (including the Libraries) enter the MIT Institute Archives as historical records once they are no longer actively in use, and may include PII. Access to archival records is governed by the Institute Records Access Policy.

4 Sharing data with third parties

MIT Libraries will share information with third-party vendors when vendors are used to provide you with library services. Some third-party vendors may be located outside of the United States. Information processed by third-party vendors on behalf of the Libraries will generally be governed by this Privacy Policy. Our library contracts generally require third-party services to maintain the same level of privacy and security over your information as the Libraries do.

4.1 Authentication for your visit to a third-party site

For library services where you interact directly with a third-party platform requiring MIT authentication, MIT's authentication system will pass some information to the third-party vendor in order to enable your access. For third parties that have met the criteria for the InCommon Federation's Research and Scholarship category, the information released is described by their policies. For other parties, we release only those attributes that are needed for that service - typically less than those available via InCommon.

4.2 Government Requests for Library Records

Information about individual library patrons will not be made available to any agency of state, federal, or local government except pursuant to such process, order, or subpoena as may be authorized under the authority of, and pursuant to, federal, state, or local law relating to civil, criminal, or administrative discovery procedures or investigatory powers.

In the case of court orders or subpoenas for information about an individual, that individual will ordinarily be notified of the request as soon as possible, unless a court order prohibits such notification, e.g., the USA Patriot Act. Information requested by subpoena or court order may only be released by an authorized officer of the Institute. For the Libraries, the Institute's authorized officer is the Director of Libraries.

4.3 Non-personally identifiable information

De-identified data about the use of our collections may be shared externally. Such data generally describes the Libraries overall activities. An example of this are the statistics which we provide to the annual survey of the Association of Research Libraries.

5 What we do with your data

The Libraries use the data which we have collected for a targeted set of purposes. Primarily, we use this information to provide the services which you request (such as looking up the list of materials you have currently checked out, or allowing you to renew those items).

We also use the information we receive to improve our existing services (for example to troubleshoot reported problems). We pursue these types of improvements using de-identified records wherever possible, although in some cases the work may involve access to records before that de-identification occurs.

The Libraries do not sell the information that we have collected to any other organization.

6 Data Retention

The Libraries retain the data we receive according to a life cycle that is informed by the data's risk classification and the operational need for that data. This life cycle starts at creation of the record, which is then in active use as long as it is needed (some information, if in frequent or continuous use, may therefore remain active for long periods of time). A record is defined as in active use while it is still needed for the purpose it was provided for.

Once the active need for a record has expired, we retain the record for a specified period before it is deleted. Retention periods are informed by the risk category of the information, and governed by records retention policies at MIT in accordance with MIT policy 13.4 and the records retention schedule of the MIT Libraries.

Data classified as high risk is, unless otherwise described below, retained by the MIT Libraries for 30 days or fewer after active use. For example, we retain logs of application use in order to monitor our programs for problems and optimize their performance via a third-party vendor, and this data is purged in their system after two weeks. General collections circulation records (records of what you check out) are kept for seven days before being de-identified.

Data which is considered high risk when associated with a personal identifier such as your name or email address may be retained for longer than 30 days after being disassociated with PII if there is an ongoing need for their retention (for example, records of reference questions received by the libraries may be deidentified and retained in order to assess the ongoing reference needs of the MIT community and for staff training purposes). If such data could, in theory, be re-identified with a particular individual, it may be reclassified as medium risk, and treated accordingly. If such data would be impossible to re-identify (for example, aggregate use statistics of library resources), then it may be reclassified as low risk.

Medium risk data is retained no longer than 5 years after active use. Low risk data may be retained indefinitely, or discarded according to Institute retention schedules.

6.1 Exceptions

Exceptions to the above-stated retention periods may be warranted in specific cases, and listed here. If you have questions about the retention period of specific types of records, please contact us at library-privacy@mit.edu.

1. Financial Records

Records which include financial information are retained for five years, regardless of risk classification, in accordance with MIT VPF Policy.

2. Distinctive Collections

Records of access to MIT Distinctive Collections and the Distinctive Collections Reading Room, are maintained for security and insurance purposes, and become permanent records of the Institute.

7 Data Integrity & Security

The Libraries protect the privacy of patron data through multiple avenues that are mutually reinforcing. First, we try to collect the most minimal set of information needed to provide the requested service. Second, we follow relevant security recommendations to ensure the security of the data that we do collect. Third, we discard your personal information as soon as possible and feasible.

7.1 Minimal Information

The services that you request can sometimes be provided with little or no information that can uniquely identify you. Many resources grant access solely because you are connected via the MIT network. The only information known about you from this access method is the MIT-based IP address of your device. IP addresses are sometimes traceable back to an individual user by MIT IS&T, however that information is not transmitted to the Libraries or the third party website.

Other Libraries-provided services may require you to log in via MIT's Touchstone service rather than connecting you automatically or having you log in directly to the website itself. This results in the website knowing less about you than if you had created an account on your own. For more information about how this approach can protect your personal information, please visit "How Shibboleth Works."

7.1.1 Library equipment

Some records that reflect patron actions - such as event logs - get stored directly on the hardware that generates them. For example, the public-access computers in the libraries store their event logs locally. Following the path of minimal information, these records are discarded every time the computer reboots, and are not aggregated with any other data. Information from some Libraries' equipment, such as computers in the GIS lab, may store data for longer periods, but still within the scope of this Privacy Policy.

7.2 Security Recommendations

The Libraries follow current recommended practices to ensure the security and integrity of the data that we collect, including relevant encryption standards and prompt updates to address system vulnerabilities. The extent of these practices is informed by the risk classification applicable for each type of data.

The Libraries also take care when selecting companies that provide the services which we use. When selecting vendors to contribute to our digital infrastructure, the Libraries require companies to comply with the same practices which we would follow when building a service locally.

Additionally, we avoid combining patron data unnecessarily. There is no single system of record for data about library patrons. For example, records about which items a patron has checked out are never combined with information about the web searches that the patron conducts, and both are kept separate from the reference questions that a patron asks.

7.2.1 Using cloud services

We use cloud-based service providers, where appropriate, after vetting their storage procedures and security arrangements. For example, the Libraries chose Logz.io as a log analysis service after confirming that patron data would not be accessible by Logz staff.

7.3 De-identification

When records need to be kept for reporting or analysis but no longer need to be identified with you, the Libraries take appropriate steps to remove that identification. The specific steps taken vary depending on how the records will be used as well as relevant best practices. These steps may include removal of values, the generation of aggregate summaries, or the deletion of the record. The timeframe over which this occurs is described in section 6.

8 Cookies

Some websites provided by the Libraries use cookies to assist with certain functions. This includes, at times, cookies set by third parties. These cookies are used in accordance with relevant laws, which includes providing the option to opt-out of their use. Users can also disable cookies via their web browser settings or plugins.

9 Your rights with respect to your data

The MIT Libraries is committed to protecting your data, and providing you with transparent insight into how data about you is used and protected. To the extent we can, we minimize the data that is collected and stored about you, as described above. When we do store data about you, you have the right to:

a. Access: you have the right to obtain a copy of data about you which we store
b. Rectification: you have the right to correct inaccurate information or complete incomplete information
c. Erasure: you have the right to have your personal data deleted upon request (unless certain circumstances apply)
d. Restriction or objection to processing: you can request that we limit the processing of your personal information, or cease processing your personal information (under certain conditions)
e. Data portability: you can request that we transfer data collected to another or directly to you.

If you would like to exercise these rights, you can contact us at library-privacy@mit.edu, and we will facilitate your request to the best of our ability. As described in this policy, we don't retain your information in a centralized system, so information requests may be handled on a system-by-system basis. To the extent possible, we will provide data about you securely and in a common file format, and if you request data erasure or rectification, we will, to the extent possible, also notify any other recipients of the data of such changes. To protect the personal information we hold, we may also request further information to verify your identity when exercising these rights. Some requests may be covered not by the MIT Libraries privacy policy, but by MIT data protection or a vendor's system. We will do our best to facilitate and direct your request appropriately.

Upon a request to erase information, we will maintain a core set of personal data to ensure we do not contact you inadvertently in the future. We may also need to retain some financial information for legal purposes, including US IRS compliance. In the event of an actual or threatened legal claim, we may retain your information for purposes of establishing, defending against, or exercising our rights with respect to such claim. De-identified data may also be retained, as described in this policy.

You can also contact MIT data protection by emailing dataprotection@mit.edu.

10 Accountability

The MIT Libraries will review and make any necessary updates to this policy and its implementation annually, or as necessary for changes in law or MIT policy. The policy is also subject to review by the MIT Audit Division. If a data breach is known or suspected, the Libraries will work with MIT's Infoprotect personnel through their policies and procedures.

To maintain our compliance with this policy the MIT Libraries maintains an inventory of systems that contain information about patrons in order to keep track of what information is stored. As described in Section 2 each system is classified as containing high, medium, or low risk data and each risk category has an applicable checklist describing the activities required to be compliant with the Infoprotect guidelines. Each system has a designated technical owner who is responsible for ongoing application of best practices and compliance. The data inventory will be updated as necessary and at least annually. Additionally, the Libraries will periodically review systems and practices for privacy concerns and address new threats, controls, and expectations. Libraries staff who have frequent contact with patron data also receive privacy training, and are responsible for maintaining personal practices in compliance with this policy.

Any significant changes to this policy will be accompanied by a prominent notice on the Libraries websites. If you would like to receive an email notification of policy changes or if you have any questions, concerns, comments about your privacy through the MIT Libraries you can email library-privacy@mit.edu. This policy shall apply from the date of approval and forward, although measures protective of your privacy may be applied retroactively to data currently in our systems upon approval.

This policy has been updated as follows:

Date of update	**Description of changes**	**Permalink to version**
xx/xx/2020	Privacy Policy Approved	[add]
pre-2020 policy		https://perma.cc/22T8-HPP7

Appendix A: Personally Identifiable Information Definition and Examples

We have adopted the definition of "Personal Data" from the General Data Protection Regulation (GDPR) as the basis of the Personally Identifiable Information (PII) we collect at the Libraries. That definition is "any information relating to an identified or identifiable natural person; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

Modern data capabilities make it surprisingly - disturbingly - easy to identify someone from minimal information and link that information to a preexisting profile. Because of that, we consider a broad range of information as potentially identifying.

Examples of PII (not all of which is collected or accessible to the Libraries) which could individually or in conjunction identify you include the following: names, student or employee ID numbers, email addresses, physical addresses (local and permanent), telephone numbers, dates of birth, social security numbers, race, gender, prefix or title, sexual orientation, accessibility status, names of family members or relatives, emergency contacts, driver's license numbers, credit card numbers, bank account numbers, passport numbers, citizenship status, income, financial information (e.g. fines, tuition, financial aid), transaction logs, content of transactions (e.g. emails), student coursework (anything prepared by a student for a class), library circulation records, log entries generated by a single user, or IP addresses. These categories are frequently PII, however the scope of PII covered by this policy also includes any other information that could potentially be linked back to you.