

Library Privacy Glossary

Introduction	5
How to use this glossary	5
Acknowledgments	5
Glossary	5
Advertising exchange	5
Advertising technology or “adtech”	5
Algorithmic bias	6
Aggregated data	6
Anonymization	6
Artificial intelligence (AI)	6
Authentication	6
Authorization	6
Authorized user	6
Behavioral tracking / behavioral data	7
Bossware	7
Browser fingerprinting	7
Browser history	7
Brute-force attack	7
California Consumer Privacy Act (CCPA)	8
Card-swipe data	8
Clickstream data	8
Common rule	8
Confidentiality; confidential data	8
Consent	8
Cookie	9
Cookie consent banner	9
Credentials	9
Credit bureau	9
Current Research Information System (CRIS)	9
Cybercriminals	9
Data	10
Data analytics	10
Data breach	10
Data broker	10
Data cartel	10
Data controller	11

Data fusion	11
Data governance	11
Data harvesting	11
Data integrity	11
Data minimization	11
Data processing agreement	11
Data processor	12
Data transparency	12
Data warehouse	12
Deceptive patterns / Manipulative design	12
De-identification	12
Device/digital fingerprinting	12
Differential privacy	12
Direct identifier	13
Domain Name System (DNS)	13
Doxxing	13
Encryption	13
Facial recognition	13
Faculty Information System / Current Research Information System	13
Family Educational Rights and Privacy Act (FERPA)	14
Freedom of Information Acts	14
Fusion centers	14
Fuzzing	14
General Data Protection Regulation (GDPR)	14
Geolocation	15
Identifier	15
Indirect identifier	15
Information asymmetry	15
Informed consent	15
Internet Protocol Address / IP Address	15
Insider threat	16
Institutional data	16
Institutional Review Board (IRB)	16
K-anonymity	16
Land-grant institution	16
Learning management systems (LMSs)	16
Library privacy values	16
Library record	17
Local storage	17

I-diversity, ℓ -diversity	17
Malware	18
Metadata	18
Microaggregation	18
Net states	18
Nonfungible (information)	18
Opt-in/out	18
Outlier	19
Passkey	19
Password manager	19
Paywall	19
Personal data	19
Personally-identifying information (PII)	19
Platform	19
Academic platform	20
Educational technology platform	20
Research platform	20
Predictive analytics systems	20
Privacy	20
Privacy by design	20
Privacy by default	20
Privacy-enhancing technology (PET)	21
Privacy extensions/add-ons	21
Privacy impact assessment	21
Privacy policy	21
Privacy review, privacy audit	21
Privacy settings	21
Proxy variable	22
Pseudonymization	22
Ransomware	22
Real-time ad bidding (RTB)	22
Reidentification	22
Research analytics tool	23
Search history	23
Security audit	23
Sensitive information	23
Shadow profile	23
Single sign-on (SSO)	23
Stalkerware	23

Stingray	24
Student data	24
Sunshine law	24
Surveillance	24
Surveillance capitalism	24
Surveillance creep	24
Surveillance culture	24
Surveillance publisher	24
T-closeness	25
Targeted advertising	25
Threat model	25
Universal identifier	25
Usage data	25
Vendor	25
Virtual private network (VPN)	26
Web analytics	26
Web tracking	26
“Wexis”	26
References and resources	27

Introduction

How to use this glossary

Privacy and security, as substantial communities of practice both inside and outside higher education, have developed community-specific terminology that many educators will not have encountered before. This glossary therefore has two main purposes:

1. To acquaint you with specialized or unusual terminology used in the *Academic Platforms & Your Privacy* briefing documents
2. To help you understand and respond appropriately to such terminology when it is used in news, research, reports, and professional discussions.

Specialized language can be used as a tool for exclusion. This glossary is a starting point for inclusive participation in discussions about privacy and surveillance.

Acknowledgments

Much of the material in this guide was adapted from Asher et al. 2018, "[Ethics in Research Use of Library Patron Data: Glossary and Explainer](#)," available under a CC-BY 4.0 license. Our grateful thanks to all its compilers!

Contributors to this document included: Susan Barrett, Brady Beard, William Denton, Orolando Duffus, Michele Gibney, Sarah Hartman-Caverly, Lisa Levesque, Janis McKenzie, Marisol Moreno Ortiz, Dorothea Salo, and Becky Yoose.

Glossary

Advertising exchange

A vended system, where advertisers and publishers buy and sell ad space typically in real time, that programmatically decides which advertising to show which users on which websites. See also: real-time ad bidding.

Advertising technology or “adtech”

See *targeted advertising*.

Algorithmic bias

The ability of computer systems, statistical models, and even data themselves, to contain bias originating from the researchers who collected and analyzed the data or, as in many cases of Internet-derived data, from those generating the data in the first place.

Aggregated data

A method of de-identification that reduces the granularity of personal data through grouping data into categories or ranges (such as using age ranges to report on user birthdate or age data). Aggregation carries some risk of re-identifying an individual if there are outliers in the data set, if the dataset size is small overall, or if categories or ranges are granular enough to identify an individual in the dataset. See also: *de-identification*.

Anonymization

Altering a dataset such that no individual represented in it can be reidentified. Most security researchers [believe this is impossible for almost all datasets about people](#), given the large amount of publicly-available or legally-purchasable data (not to mention purchasable stolen or hacked data) available about individuals. See also: *de-identification*, *pseudonymization*.

Artificial intelligence (AI)

Umbrella term for a broad set of technologies (e.g. predictive analytics, large-language models and the chatbots based on them, image and video generators) founded on machine learning. Model training for AI often involves voluminous collection and scraping of personal data and can include various forms of surveillance.

Authentication

The process by which a system proves to its satisfaction that a user is who they say they are.

Authorization

The system privileges or information/data access that a given authenticated user is entitled to.

Authorized user

At most institutions this includes full- and part-time students regardless of physical location; full- and part-time employees (including instructors, staff, affiliate researchers, and independent contractors), regardless of their physical location; alumni (in some cases); other valid ID holders; and patrons who are physically present at library sites (“walk-ins”).

Behavioral tracking / behavioral data

Gathering and storing a dossier of individual users' online behaviors (hovers, clicks, visited pages, time on site, ratings, shares, purchases, etc.). This may occur on a single website or be aggregated across many ("cross-contextual behavioral tracking"). Comparing a de-identified behavioral-tracking dossier with a corpus of identified dossiers can result in reidentification of individuals. Similarly, a marketer (or attacker) seeking a specific individual already known to them in a corpus of de-identified dossiers is often able to use their pre-existing knowledge of that individual to seek behaviors typical of that individual.

Behavioral data can include:

- Search history
- Electronic content access histories
- Circulation histories
- Website activity
- Geolocation history

Bossware

Tools for workplace surveillance of employees, often through the computing devices they use for work. In higher education, bossware may also include academic platforms.

Browser fingerprinting

A reidentification technique that uses websites' or trackers' ability to access browser settings (location services, installed add-ons, installed fonts, etc) to uniquely identify the browser and/or associate it with its user, whether that user is logged in or not. An alternative to cookies. See also: *device/digital fingerprinting*.

Browser history

A log kept in the user's browser of web pages that the user visited. This history can be breached or recovered by forensics techniques.

Brute-force attack

An attack on a system involving multiple, often rapid-fire, trial-and-error guesses, e.g. at a password. Sometimes used against authentication systems, though almost all authentication systems today lock out accounts after a number of incorrect login attempts. May also be used to find usable credentials in stolen username/password stores, as in the 2022 LastPass breach.

California Consumer Privacy Act (CCPA)

The [California Consumer Privacy Act](#) is a data-protection law that allows California citizens to opt out of sharing/sale of personal information, especially sensitive personal information, by companies that meet certain revenue, personal-information collection, or data sale/sharing thresholds. Affected companies must also meet the laws' data transparency, retention, data minimization, purpose limitation, and cybersecurity requirements. Many library vendors would qualify under these thresholds, though most libraries themselves are not subject to the laws.

Card-swipe data

At many institutions, institution-issued student ID cards or third-party apps are used to surveil a great deal of student behavior on campus: purchases at campus stores and restaurants, entry into restricted buildings or spaces (including some libraries), event attendance, class attendance, and more. Such a dataset is often referred to as “card-swipe” data.

Clickstream data

Data derived from detailed surveillance of a user's interactions with a website or mobile app. Clickstream data may include clicks, information typed into forms or search bars, time spent on individual pages or functions, and the order of pages visited.

Common rule

Nickname for the United States “[Federal Policy for the Protection of Human Subjects](#),” which applies to all institutions accepting federal grant funding. The Common Rule outlines human-subjects research ethical requirements related to Institutional Review Boards, informed consent, and compliance obligations.

Confidentiality; confidential data

ALA's [Interpretation of the Library Bill of Rights](#) states that “confidentiality exists when a library is in possession of personally identifiable information about users and keeps that information private on their behalf.” Confidentiality is the act of holding information about an individual without disclosing it to others unless there is proper authority and necessity. Confidentiality applies specifically to a person's data, and is related to but narrower than the broader concept of privacy. See also: *privacy*

Consent

The act of giving permission for the use or disclosure of an individual's data. Explicit consent requires an affirmative action from the individual while implicit consent does not. For example, if a website user clicked a checkbox or button to agree to cookie tracking while using a website that would be explicit consent, while if they simply continued using the website which had a

posted banner stating that it tracks cookies that would be implicit consent. Modern data protection laws often seek consent that is informed, explicit/affirmative, and specific.

Cookie

A small data file sent from the website and stored on the user's computer or the user's browser. Web cookies can be used to manage user authorization and session management, as well as track users through web analytic software and other tracking software. Some cookies last until a web session ends ("session" cookies), while other cookies ("persistent" cookies) can remain on the user's computer after the end of a session. Flash Cookies are a type of persistent cookie. Additionally, a Zombie Cookie is a form of Flash Cookies that once deleted by the user from their computer the cookie automatically reappears, which makes it difficult to detect. This type of cookie can be used for persistent tracking and by malicious actors wanting to install malicious malware in user devices. A website can have web cookies from the site itself ("first-party" cookies) as well as cookies from external sites ("third-party" cookies). Because of cookies' widespread use in targeted advertising and other surveillance, some browsers block third-party cookies by default.

Cookie consent banner

An obtrusive clickthrough pop-up in a website or app, for the purpose of gaining user consent to data collection and sharing. Often used to satisfy the letter of privacy law such as *Europe's ePrivacy Directive* though perhaps not its spirit. See also: *General Data Protection Resolution*

Credentials

Shorthand for user-authentication information including usernames, passwords, biometric identifiers, etc. See also: *authentication*.

Credit bureau

A company that collects and stores financial and other information on individuals, reporting on their loan worthiness to financial institutions, employers, and others. In the United States, the biggest credit bureaus are Equifax, Experian, and TransUnion. Equifax notoriously had a significant data breach of information on individuals in 2017.

Current Research Information System (CRIS)

See: *Faculty Information System / CRIS*.

Cybercriminals

Cybercriminals are individuals that use malware to extract individuals' information. Cybercriminals use tricks as a technique to infect devices. Examples include phishing emails and infected files.

Data

Structured or semi-structured pieces of information; in a privacy context, usually about individual people.

Data analytics

Data analytics is the practice of collecting, storing, cleaning up, enriching, cross-connecting, transforming, organizing, and then analyzing data, purportedly to help make decisions and predictions. Data analytics companies collect, purchase, and aggregate large amounts of data about individual people, which they use to reidentify and make inferences about those individuals. Companies then sell the aggregated and reidentified data, the inferences, or both. Some library vendors are data analytics companies; others sell data on content use and users to them.

Data breach

When data meant to be secure are leaked or stolen. For example, if hospital records which are intended to be securely stored are made publicly available there has been a data breach.

Data broker

Business that aggregates (from credit bureaus, social-media platforms, web trackers, marketers, public information sources, etc) and resells data about individuals. Some library vendors are entwined with data brokerages.

Data cartel

Similar to manufacturing cartels, data cartels maintain high prices for their products and can restrict competition by collecting and paywalling large volumes of data to maintain dominance in the analytics market. Data cartels can collect legal and academic information along with personal data. Products they create can be sold to law enforcement, academic institutions, insurance companies, employers, landlords and healthcare servicers. Data cartels compete to provide the biggest pool of search results and seem to prioritize content quantity to quality, which can undermine quality review processes and lead data cartels to commingle vetted information with incorrect information that is then sold as trusted resources. This data influences access to services and products that reinforce financial, legal, social, political and medical entitlement structures, and can produce negative consequences to individuals who have no control over the data that is collected or shared. See *Data Cartels* by Sarah Lamdan.

Data controller

Under Europe's *General Data Protection Regulation*, which has been imitated by many other privacy laws globally, a data controller determines the purposes of any personal data and the means of processing it. See also: *data processor*, *data processing agreement* and *GDPR*.

Data fusion

The process of combining datasets from multiple sources. Such integration may reidentify formerly de-identified individuals or make possible new and possibly damaging inferences.

Data governance

The practice of inventorying data sources across an organization and defining policies for data collection, storage, access, (re)use, and deletion. Data governance policies often include classifying data by its level of required or desired confidentiality.

Data harvesting

Collecting data held by a third party, with or sometimes without consent being given to that third-party. Examples include dataset sharing (usually consensual) and screenscraping data from a website or other online service (sometimes consensual, sometimes not).

Data integrity

A measure of whether a given dataset is accurate, complete, and internally consistent.

Data minimization

Deliberately minimizing the amount, granularity, and sensitivity of data collected about people, as well as the length of time it is kept (if kept at all), on the principle that data not collected or kept cannot be leaked, stolen, or used against anyone.

Data processing agreement

Under Europe's *General Data Protection Regulation*, which has been imitated by many other privacy laws globally, a binding legal contract between a data controller and data processor that lays out "the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller" ([article 28](#)). These agreements may include specific protections for data subjects or limitations on data-processor actions on the data. See also: *data controller*, *data processor* and *GDPR*.

Data processor

Under Europe's *General Data Protection Regulation*, which has been imitated by many other privacy laws globally, a data processor processes personal data on behalf of a data controller. See also: *data controller*, *data processing agreement* and *GDPR*

Data transparency

Explaining the trajectory (collection, use, sharing/selling, security, disposal) of data clearly, thoroughly, and publicly, including via a posted privacy policy.

Data warehouse

A large collection of institutional or multi-institutional data combined from multiple smaller source datasets.

Deceptive patterns / Manipulative design

User-interface design decisions in websites or apps that convince or sometimes manipulate or coerce users to take actions not in their best interests, including but not limited to oversharing personal information. Also called “dark patterns,” though this usage is deprecated because of its perceived racism.

De-identification

The process of transforming personal data to remove identifiable aspects of the data. This may be attempted through a variety of methods, including aggregation, stripping or truncating personal data, or removing or pseudonymizing personal data. De-identification is not usually sufficient to prevent reidentification of individuals in the dataset. See also: *anonymization*, *pseudonymization*.

Device/digital fingerprinting

A reidentification technique that uses software, app, and browser access to device settings to uniquely identify a device and/or its user, even when the user has blocked access to the usual mobile identifiers. See also: *browser fingerprinting*.

Differential privacy

Differential privacy conceals whether any individual's data was present in the dataset by adding random data into the dataset. Differential privacy approaches are the subject of active research. See also: *fuzzing*.

Direct identifier

Identifier whose purpose is to identify a specific individual and tie other data to that individual.

Direct identifiers can include:

- Name
- Email or physical address
- Government or organization-issued identification numbers
- Account username and password
- Biometric information
- IP address, especially combined with a timestamp
- Device information (operating system, browser, device unique ID, etc.)

Domain Name System (DNS)

The service that associates numerical Internet Protocol (IP) addresses with domain names, which are more friendly for everyday human use. A domain name can point to another domain name through a Canonical Name (CNAME) record. A domain name mapped to the other domain name is called an *alias*.

Doxxing

Discovering and publicly revealing information – name, location, workplace, home address, family details – about the person behind a pseudonym, or a public figure. Often part of mobbing and other online harassment.

Encryption

In most data-related uses, the systematic and reversible alteration of data such that it is readable to authenticated and authorized users while remaining unintelligible to all others.

Facial recognition

An application of machine learning that aims to identify individuals captured in photographs or video. Much more likely to be inaccurate with women and people of color, due to poor diversity in training sets. Can be used from everything from payment and user authentication to employee monitoring, corporate or government surveillance, policing, and suppression of activism.

Faculty Information System / Current Research Information System

The Faculty Information System (FIS) or Current Research Information System (CRIS) is a web application that allows faculty to view, enter and track their teaching, research and service activities. These systems can streamline the personnel review process and annual faculty

activity reporting. FIS and CRIS tools can inform and influence institutional decisions related to hiring, promotion, and tenure as well as internal resource allocation decisions.

Family Educational Rights and Privacy Act (FERPA)

The [*Family Educational Rights and Privacy Act \(FERPA\)*](#) is a US federal law from 1974 that regulates student educational records. This law only applies to schools that receive funding from the US Department of Education. It puts in place some limited privacy protections over records, and gives parents some rights over their child(ren)'s school records; those rights transfer to students once they turn 18 or start higher education. It is largely considered outdated in terms of protecting student privacy in the modern era.

Freedom of Information Acts

Freedom of Information Acts (FOIA) are legislation which provide access to government records. In the USA, the [*Freedom of Information Act*](#) enables requests of information from the Executive Branch of the United States Government and records requested must be disclosed unless they are protected by one or more of the exempt categories of information found in the FOIA (U.S. Department of the Treasury). In Canada information can be requested from federal government bodies through ATIP requests, which stands for "Access To Information and Privacy" because the enabling legislation is the [*Access to Information Act*](#) and the [*Privacy Act*](#). Provincial freedom of information information requests are enabled by legislation which differs by province, including what can be requested and even the acronyms used (commonly a variation of FOI or FIPPA), as described by [*Secret Canada's Guideline to Jurisdictions*](#). See also: *Sunshine law*.

Fusion centers

Data centers used by the US federal government and the Department of Homeland Security and run by state and local governments to process big data sets for use by law enforcement.

Fuzzing

Making data in a dataset less specific, often to reduce the chance it can be used to reidentify individuals.

General Data Protection Regulation (GDPR)

The [*General Data Protection Regulation*](#) is a data-protection law which governs companies and institutions that process the personal data of citizens and residents of the European Union, no matter where they are headquartered. Under the law, processing citizens'/residents' personal data is subject to purpose and duration limitation, data minimization, confidentiality, security, and accountability requirements. Violations may incur fines of up to 4% of annual global revenue.

Geolocation

Use of cell towers, GPS, wifi access-points, IP address, bluetooth beacons, or other data to locate where a specific device or web browser (as well as its user) is geographically located. Often happens without the user's knowledge or consent, especially on mobile.

Identifier

A string of characters that uniquely identifies a person or object of interest. See also: *direct identifier*, *indirect identifier*, and *universal identifier*.

Indirect identifier

A trait or combination of traits in a dataset that, while not a direct identifier, still suffices to identify specific individuals. See also: *direct identifier* and *proxy variable*.

Indirect identifiers can include:

- Age or date of birth
- Race/ethnicity
- Gender identity
- Education level
- Field of study
- Disability status
- Veteran status
- Geographical information, such as region or ZIP code

Information asymmetry

In a privacy context, the common situation where a surveillor (such as a vendor or institution) knows immensely more about surveilled individuals than the individuals know about them, or about the surveillance and data practices in play.

Informed consent

In most human-subjects research, researchers are required to inform subjects about the researchers' aims as well as potential risks to subjects from the research. Researchers must then gain the explicit consent of subjects before the research may proceed.

Internet Protocol Address / IP Address

A usually temporary address assigned to a device so that it can communicate with the Internet. Combined with a timestamp, can allow identification of a specific user of a specific website or service at a specific time.

Insider threat

Threat to security and privacy of data from organizational insiders, who have more access to data (and opportunities to steal or misuse it) than outsiders.

Institutional data

See: *data warehouse*.

Institutional Review Board (IRB)

Committee of local researchers, including one non-scientist, at any US institution accepting federal research funding, tasked with gauging the ethical permissibility of the institution's research on human and animal subjects. Typically affiliated with a campus unit that provides staff support for the committee. Alternate terms include independent ethics committee (IEC), ethical review board (ERB), or research ethics board (REB).

K-anonymity

The number of people who can be identified individually using all variables in a dataset. For example, if a dataset has been de-identified, but includes only one row for a second-year woman graduate student in a particular department, the k-anonymity of the dataset is 1, as it is still possible to identify a single person in the dataset based on that person's unique combinations of variables.

Land-grant institution

A land-grant institution is a public university with obligations to provide education, research opportunities, and resources to the entire population of their geographical location.

Learning management systems (LMSs)

Systems that manage a great deal of activity for face-to-face, hybrid, and online courses. LMSs often manage announcements, assignments (including due dates), readings, grading, discussion and other student-student and student-instructor interactions. They also allow integrations with third-party content and educational-technology tools. Most interactions students and instructors have with LMSs are, or at least can be, identifiably tracked. Examples of LMSs include Canvas, Desire2Learn, Moodle, and Sakai.

Library privacy values

Library ethics codes have prioritized the privacy of patron information use since the first formulation of the ALA Code of Ethics in 1939. Current statements of this priority include:

- [ALA](#): We protect each library user's right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted.
- [Association of College and Research Libraries](#) (ACRL): The privacy of library users is and must be inviolable. Policies should be in place that maintain confidentiality of library borrowing records and of other information relating to personal use of library information and services.
- [Canadian Federation of Library Associations / Fédération canadienne des associations de bibliothèques](#) (CFLA / FCAB): Libraries have a core responsibility to safeguard and defend privacy in the individual's pursuit of expressive content. To this end, libraries protect the identities and activities of library users except when required by the courts to cede them.
- [International Federation of Library Associations](#) (IFLA): Librarians and other information workers respect personal privacy, and the protection of personal data, necessarily shared between individuals and institutions. The relationship between the library and the user is one of confidentiality and librarians and other information workers will take appropriate measures to ensure that user data is not shared beyond the original transaction.

Library record

A library's information about a user, traditionally including personal and contact information, currently checked-out materials, and fines. Library records often enjoy significant legal protections, restricting what a library can do with them. Such laws, however, do not presently apply to vendors or to the many other records (e.g. proxy-server logs, web-tracker data on library or publisher websites) currently collected on library patrons.

Local storage

HTML5 local storage allows for offline storing of website data in the user's web browser. Compared to cookies, local storage has a larger storage capacity to store potentially personally identifiable data, persists after the end of the user session and browser shutdowns, and does not have an automatic expiration date.

l-diversity, ℓ -diversity

The number of different values represented in a filtered ("microaggregated") group of rows in a dataset, a measure of reidentifiability. If, for example, there are 10 rows for second-year women graduate students in a particular department, and all of them have library fines, even if a researcher is unable to know which row corresponds to Tianna, who is a second-year woman graduate student in that department, the researcher is able to know that Tianna has a library fine.

Malware

Malicious software used by malicious actors, or cybercriminals, to infect a computer to gain unauthorized access and collect personal data, like account information, lock the device for ransom, and other malicious reasons. Some types of malware are phishing, spyware, trojans, and worms.

Metadata

In its broadest sense, any data describing or qualifying any object of interest. Personal metadata is any metadata about a person: name, identifiers, affiliations, etc. With respect to communication, metadata is data about a person's communication that is not the communicated content itself, but still contains information that can fuel inferences about the person. For a cell-phone call, metadata includes device location, number called (and owner of that number), date and time of call, and length of call.

Microaggregation

Combining rows of data in order to make each individual row less specific; a form of data fuzzing. This may involve creating less-accurate categories, whether by taking only the coarsest element of a piece of data (as with changing birth date to year of birth), using ranges (e.g. age ranges), or averaging pairs or triplets of rows and either retaining all rows (e.g., replacing all ages in a range with the average age from that group of records) or discarding all but one. Fuzzing does not always fully de-identify records. See also: *K-anonymity* and *I-diversity*.

Net states

This describes the power held by technology companies to act on a global stage, especially large multinational companies such as Google and Facebook. It is a play on the phrase "nation states," as with [Net States Rule the World](#).

Nonfungible (information)

Nonfungible resources are not interchangeable for one another. An example of a nonfungible resource is a book with unique information compared to other books. By contrast, water is an example of a fungible resource, where any drop from a faucet has the same value as any other. In a library context, many content subscriptions are considered nonfungible because there can be no adequate substitute for the specific content provided. The nonfungible nature of information can create power imbalances driving unsustainable price increases.

Opt-in/out

Opt-in is user choice involving active affirmation, such as checking a box or toggle, to turn on specific data sharing or collection settings. Opt-out is user choice assumed by a service provider

unless the user actively chooses otherwise. An example of opt-out is a product collecting user data until the user unchecks the box that controls the data collection setting. The American Library Association advocates for [opt-in \(“explicit consent”\)](#) rather than opt-out default settings in library systems and services. See also: consent.

Outlier

Piece (or combination of pieces) of data in a dataset that falls significantly outside the norm. Outlier characteristics in data about people often make individuals reidentifiable.

Passkey

A form of authentication credential relying on the user’s control over a specific device, such as a laptop, phone, or security key. Once fully implemented, passkeys are likely to be more secure and easier to use than passwords and multi-factor authentication.

Password manager

Software that enables a user to securely save and reuse their login information (including passwords) on their device(s) and sometimes online, as well as generate new hard-to-guess passwords. These programs usually save credentials in an encrypted database that the user accesses with a master password. Examples include LastPass (recently breached), Bitwarden, and 1Password.

Paywall

Any technology blocking access to content by those who have not paid (either individually or via library subscription) to see it.

Personal data

See: *personally-identifying information (PII)*.

Personally-identifying information (PII)

Information that directly ties data about a specific person to that person. Usually construed as direct identifiers only, but sometimes (correctly) construed as combinations of indirect identifiers or collections of behavioral data that serve to identify individuals.

Platform

The user interface through which researchers access specific academic products, namely databases. As market consolidation occurs with academic vendors, there are a smaller number of platforms which provide access to academic products. See also: *research analytics tool* and *vendor*. Specific types of platforms include:

Academic platform

An interface providing access to information required when conducting research. For instance, ProQuest is an academic platform which provides access to newspaper databases, scholarly articles databases, and other subscription collections.

Educational technology platform

An interface supporting teaching and learning in higher education, such as a learning management system or a question-and-answer platform.

Research platform

An interface for tracking faculty publications. There are many different types of higher education platforms, including platforms for managing teaching, research, assessment, or human-resources activities, and platforms may be provided by a vendor or maintained in-house.

Predictive analytics systems

These use machine-learning techniques to predict future behavior based on surveillance data. Known to be biased and inaccurate, especially on the individual level.

Privacy

ALA's statement "[Privacy: An Interpretation of the Library Bill of Rights](#)" asserts: "In a library (physical or virtual), the right to privacy is the right to open inquiry without having the subject of one's interest examined or scrutinized by others." Privacy is a significant topic in healthcare, finance, the use of technology, border crossings, and personal interactions with business and government. According to surveys about privacy, many Americans ([Pew Research Center](#)) and Canadians ([Office of the Privacy Commissioner of Canada](#)) are highly concerned about their personal privacy.

Privacy by design

Incorporating privacy into product or service design from the beginning of the design process, rather than trying to add privacy after product/service release.

Privacy by default

Assigning the highest level of privacy protections as the default settings in a system, service, or product.

Privacy-enhancing technology (PET)

According to the [US White House Office of Science and Technology Policy](#), these are “technologies that will allow researchers, physicians, and others permitted access to gain insights from sensitive data without ever having access to the data itself.”

Privacy extensions/add-ons

Software that can be added to web browsers (desktop/laptop and sometimes mobile) to enhance user privacy. These include tracker blockers, identifier blockers, and location-disclosure blockers.

Privacy impact assessment

According to the [US National Institute of Standards and Technology](#): “An analysis of how information is handled that ensures handling conforms to applicable legal, regulatory, and policy requirements regarding privacy; to determine the risks and effects of creating, collecting, using, processing, storing, maintaining, disseminating disclosing, and disposing of information in identifiable form in an electronic information system; and to examine and evaluate protections and alternative processes for handling information to mitigate potential privacy concerns. A privacy impact assessment is both an analysis and a formal document detailing the process and the outcome of the analysis.” Privacy impact assessments are often limited to a single organizational initiative or service. See also: *privacy review*.

Privacy policy

A policy explaining an organization’s or service’s privacy practices, especially though not solely around users’ personal data. Notorious for copious vague, impenetrable, easily-misunderstood legalese; [known to be poorly understood](#) by most users.

Privacy review, privacy audit

A systematic assessment of an organization’s or initiative’s privacy posture. In libraries, the [ALA offers privacy-audit guidance](#). Unlike privacy impact assessments, privacy audits typically aim to evaluate the entire organization rather than a specific initiative. In industry, audits are often tied to legal compliance. See also: *privacy impact assessment*.

Privacy settings

Privacy settings are options available for device users to control who sees their information online. This gives users more control over what information is collected, who sees it, and how long it is backed up by their device. Settings are available in social media sites, browsers, and other software.

Proxy variable

A characteristic captured in a dataset that tracks another characteristic closely enough to substitute for it. For instance, postal code can be used as a proxy for address when all addresses with that postal code are located in the same building complex. Proxies are often a way to sort people into sensitive, even legally protected, classes when direct information about class membership is not in the dataset. Information use can be a proxy variable for several demographic characteristics, including sensitive or legally-protected ones such as race/ethnicity, gender, and sexual orientation. See also: *indirect identifier*.

Pseudonymization

A data protection method in which a persistent identifier (“pseudonym”) is used instead of a person’s name or direct identifier, so that the data can no longer be attributed to a specific individual or user without the use of additional information that is maintained separately. While pseudonymization is meant to obscure identity, it is insufficient for anonymization if the pseudonym is linked to behavioral data or a direct identifier (e.g., via web tracking). See also: *anonymization*, *de-identification*, and *reidentification*.

Ransomware

A type of malware that after initial intrusion into an organization’s systems, steals and/or encrypts any data it encounters. Cybercriminals then demand ransom from the organization for the key to decrypt the data and/or for a promise not to reveal it.

Real-time ad bidding (RTB)

A live auction conducted where multiple advertisers bid on individual advertisement impressions on a web page or app. The system running the auction automatically displays the advertisement of the winning bidder. Bidders participating in RTB rely on personal data (e.g., demographic data, location, browsing history) to determine when and how to bid for advertisement impressions. See also: *advertising exchange*.

Reidentification

The process of using a de-identified dataset to identify groups or specific individuals. This typically occurs by:

1. Leveraging proxy variables or other identity clues in the data.
2. Comparing the data to overlapping datasets.

Research analytics tool

A vendor-provided or in-house service that uses bibliometric analysis to assess researcher or department productivity and prominence. Similar terms include research assessment, research-tracking system, or research metrics.

Search history

A record of searches made in a service, collected by the service provider. Under library-privacy ethics, search histories are considered confidential because they reveal a patron's topics of interest. Even de-identified or pseudonymized search histories can enable reidentification based on interests or indirectly identifying information.

Security audit

A systematic review of an organization's or service's security posture. Often includes a "penetration test," where testers actively probe for security vulnerabilities.

Sensitive information

Information or data whose collection, analysis, reuse, sharing, breach, or sale could harm the person(s) it describes, especially if the data is reidentified. Given improvements in inferences about individuals based on even seemingly innocuous data, there is a debate about whether data should still be classified as sensitive vs. not sensitive. See also: *personally-identifying information (PII)* and *reidentification*.

Shadow profile

A behavioral-surveillance dossier on individuals who do not use a service, created through cross-web/app tracking, universal identifiers, or purchased data. Facebook is notorious for [maintaining shadow profiles on non-users](#).

Single sign-on (SSO)

Technologies enabling a user to log in to multiple services with a single set of credentials. Institutions (identity providers) authenticate users and send metadata (often including personal information) to services (service providers) to grant access. Examples include Shibboleth and SAML technologies; InCommon, OpenAthens, and the UK Access Management Federation are SSO-capable federations.

Stalkerware

Software for computers and mobile devices that enables stalkers, domestic abusers, and cybercriminals to surveil their targets' behavior while using the device. Also called spyware.

Stingray

A device that pretends to be a cell-phone tower in order to capture the location and identity of individual mobile phones, often to connect them with users and track them in physical space.

Student data

Data specific to students and their behavior. Includes, but is not limited to, personally-identifying information. See also: *personally-identifying information (PII)*.

Sunshine law

A law that requires government agencies to respond to requests for (non-classified, non-confidential) government records. In the US, the federal sunshine law is the *Freedom of Information Act*; each state has its own sunshine law governing requests for state records. See also: *Freedom of Information Acts*.

Surveillance

The careful monitoring of individuals or places through behavior-observation. Surveillance is facilitated through surveillance technology, such as video cameras or the logging of digital footprints. Surveillance is often used for marketing, risk management, and website performance. Data collected during surveillance is often retained unbeknownst to individuals being surveilled. Surveillance is disproportionately conducted on and presents greater risk for members of marginalized communities.

Surveillance capitalism

A term coined by Shoshana Zuboff, referring to the commodification of personal data, typically for hyper-targeted advertising. Often associated with companies like Google and Facebook.

Surveillance creep

The repurposing of surveillance data for uses not originally intended during collection.

Surveillance culture

The normalization of surveillance practices and techniques, as applied to broad communities through widely-used systems and platforms, whether for purposes of research, control, or financial exploitation.

Surveillance publisher

A publisher that acts as a data broker and/or earns significant revenue from surveilling readers and selling data or data analysis based on that surveillance. Many content vendors to libraries

may be considered surveillance publishers or aspiring surveillance publishers, as defined by Jefferson Pooley.

T-closeness

The number representing the distance between the distribution of a particular sensitive attribute (i.e., a variable in a dataset that needs to be protected but cannot be changed without affecting the utility of the dataset) and the distribution of the same attribute in the overall dataset. The closer the two numbers are to each other, the lesser the chance of reidentifying someone through attributes in the dataset.

Targeted advertising

Ads directed at specific users or groups based on behavioral data collected online and offline. See also: behavioral tracking.

Threat model

A security or privacy risk assessment conducted for a specific person, group of people, or organization based on the nuances of their situation, e.g. what information exists about them, how much of it is sensitive, and any personal or society-created vulnerabilities they labor under.

Universal identifier

An adtech tool that tracks individuals across apps and websites without cookies, increasingly replacing browser cookie tracking.

Usage data

Any data that is collected on someone's use of a service or product, but usually referring to information gathered about online use, such as pageviews, length of time spent on a page, referrer data, even user IP addresses.

Vendor

A company selling information or platforms, including to libraries in the form of research databases, usually with profits far exceeding the cost of production. Market consolidation of vendors has occurred over time and there are now a small number of major players who not only sell information but are deeply involved in many aspects of the lifecycle of research, including publishing, aggregating, indexing, and preservation. Due to the reduction of vendor competition via consolidation of the marketplace, libraries tend to pay higher prices and have less leverage in negotiations, making it more difficult to avoid vendors that present privacy risks.

Virtual private network (VPN)

Network software that “tunnels” some or all internet traffic from a VPN-enabled device through the VPN provider before it reaches the internet, to help disguise the traffic’s origin, originating device, and originator. VPNs do not, however, eliminate web trackers or similar adtech.

Web analytics

The collection and analysis of website usage data. Web analytic applications track and capture user data from a variety of sources. Depending on the application, user data can range from search term results, page hits, and landing/exit pages to demographic data, behavioral data, and even the data of users visiting sites outside the original website.

Web tracking

Describes internet surveillance enabled by specific tools. For example, a web beacon is a web tracking method that is often undetectable by the end user. Web beacons commonly take the form of a small image file (often a single transparent pixel) loaded with a web page or email but can also be embedded in visible graphics (e.g., buttons, banners) and HTML elements. Beacons can collect user data such as behavioral data, digital fingerprints, and IP addresses. Examples include [Meta \(Facebook\) pixels](#) and [Google Analytics](#). Wifi tracking is the monitoring of individuals' movements in physical spaces by tracking device connections or connection attempts to wifi access points.

“Wexis”

A colloquialism for the vendor duopoly of Westlaw and Lexis, who control most US legal information. Their parent companies are Thomson Reuters and RELX, major data brokers. Wexis controls access to legal information that is essential to the function of the legal system, government, and the academic study of law.

References and resources

- Allan, Marshall (2018). Health Insurers Are Vacuuming Up Details About You – And It Could Raise Your Rates. ProPublica.
<https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates>.
- American Library Association (last revision 2021). ALA Code of Ethics.
<https://www.ala.org/tools/ethics>
- American Library Association (2006). Intellectual Freedom Principles for Academic Libraries: An Interpretation of the Library Bill of Rights.
<http://www.ala.org/acrl/publications/whitepapers/intellectual>.
- Bauer-Wolf, J. (2017). Big Brother: College Education. Inside Higher Education.
<https://www.insidehighered.com/news/2017/12/21/georgia-techs-monitoring-students-social-media-causes-concern>
- Beals, M. (2022). George Washington University Apologizes for Tracking Locations of Students, Faculty. The Hill.
<https://thehill.com/homenews/state-watch/594142-george-washington-university-apologizes-for-tracking-locations-of>
- Berkeley Office of Ethics (2013). Primer on autonomy privacy, information privacy, and information security. <https://ethics.berkeley.edu/privacy/pisi>
- Bevins, F., et. al. (2020). Shaping University Boards for 21st Century Higher Education in the US. Education, McKinsey & Company.
<https://www.mckinsey.com/industries/education/our-insights/shaping-university-boards-for-21st-century-higher-education-in-the-us>.
- Briney K. & Yoose B. (2022). Managing data for patron privacy: comprehensive strategies for libraries. ALA Editions.
- Caines, A. and Glass, E. (2019). Education Before Regulation: Empowering Students to Question Their Data Privacy. Educause Review.
<https://er.educause.edu/articles/2019/10/education-before-regulation-empowering-students-to-question-their-data-privacy>.
- California State University, The (2021). On the Use of Remote Proctoring Software for Assessment. Faculty Senate Resolution AS-3474-21/AA.
- Canadian Federation of Library Associations / Fédération canadienne des associations de bibliothèques (last revised 2019). Statement on Intellectual Freedom and Libraries.
<https://cfla-fcab.ca/en/guidelines-and-position-papers/statement-on-intellectual-freedom-and-libraries/>
- Colbert-Lewis D., et al. (2024). The citation economy as a site of extraction for surveillance publishing. Canadian Journal of Academic Librarianship.
<https://doi.org/10.33137/cjal-rcbu.v10.43293>
- Cornell University (2023). Privacy. Center for Teaching and Learning.
<https://teaching.cornell.edu/learning-technologies/privacy>.

- Creel, K., & Dixit, T. (2022). Privacy and Paternalism: The Ethics of Student Data Collection. *MIT Case Studies in Social and Ethical Responsibilities of Computing, Summer 2022*.
<https://doi.org/10.21428/2c646de5.b725319a>.
- Dalva, Cassidy (2022). Faculty Senate Debates Expansion of Video Surveillance, Searches for Mental Health Solutions. Campus Life, The Stanford Daily.
<https://stanforddaily.com/2022/04/21/faculty-senate-debates-expansion-of-video-surveillance-searches-for-mental-health-solutions/>.
- Dunlap, J., Palmer, I. and Wesley, A. (2021). Keeping Student Trust: Student Perceptions of Data Use Within Higher Education. Education Policy, New America.
<https://www.newamerica.org/education-policy/reports/keeping-student-trust>
- Eastman, J., et. al. (2022). University Governance in Canada: Navigating Complexity. McGill-Queen's University Press.
<https://eiko-fried.com/welcome-to-hotel-elsevier-you-can-check-out-any-time-you-like-not/>.
- Fried, E. (2022). Welcome to Hotel Elsevier: you can check-out any time you like...not. Measurement, Modeling & Complexity of Mental Health Blog.
- Germain, T. (2020). Poor Security at Online Proctoring Company May Have Put Student Data at Risk. Consumer Reports.
<https://www.consumerreports.org/digital-security/poor-security-at-online-proctoring-company-proctortrack-may-have-put-student-data-at-risk>.
- Gilliard, C. (2022). Digital Redlining, Friction-Free Racism and Luxury Surveillance in the Academy. Knowledge Equity Lab, Season 3:5.
<https://knowledgeequitylab.ca/podcast/s3-ep-5/>.
- Hai-Jew, S. (2016). A Quick Skim of Legal Considerations in Instructional Design. C2C Digital Magazine, Spring/Summer 2016.
<https://scalar.usc.edu/works/c2c-digi-magazine-spring-summer-2016/a-quick-skim-of-legal-considerations-in-instructional-design>.
- Hui-Lien C., and Chen, C. (2016). Beyond identifying privacy issues in e-learning settings – Implications for instructional designers. Computers & Education, 103, p. 124-133.
<https://doi.org/10.1016/j.compedu.2016.10.002>.
- International Coalition of Library Consortia (2002). Privacy Guidelines for Electronic Resources Vendors.
<https://www.icolc.net/statements/privacy-guidelines-electronic-resources-vendors>.
- International Federation of Library Associations (2015). IFLA Statement on Privacy in the Library Environment.
<https://www.ifla.org/publications/ifla-statement-on-privacy-in-the-library-environment/>
- Lamdan, S. (2022). Data Cartels: The Companies That Control and Monopolize Our Information. Stanford University Press, p. 224.
- Mandinach, E. B., & Gummer, E. S. (2013). A Systemic View of Implementing Data Literacy in Educator Preparation. *Educational Researcher*, 42(1), 30–37.
<https://doi.org/10.3102/0013189X12459803>

- National Information Standards Organization (2015). NISO Consensus Principles on User's Digital Privacy in Library, Publisher, and Software-Provider Systems. NISO Privacy Principles. <https://www.niso.org/publications/privacy-principles>.
- Phillips, J. and Williamson, J. (2019). UC's Learning Data Privacy Principles Gaining National Attention. University of California IT Blog. <https://uctechnews.ucop.edu/ucs-learning-data-privacy-principles-gaining-national-attention/>
- Poot, G. (2023). Learning Analytics in 5 Steps: A Guide to the GDPR. Surf. <https://www.surf.nl/en/learning-analytics-in-5-steps-a-guide-to-the-gdpr>.
- Posada, A., and Chen, G. (2018). Inequality in Knowledge Production: The Integration of Academic Infrastructure by Big Publishers. *ELPUB 2018*, Jun 2018, Toronto, Canada. [10.4000/proceedings.elpub.2018.30](https://doi.org/10.4000/proceedings.elpub.2018.30). [hal-01816707](https://hal.archives-ouvertes.fr/hal-01816707).
- Quran, L. (2019). Saudi students in U.S. say their government watches their every move. World, Public Broadcasting Service. <https://www.pbs.org/newshour/world/saudi-students-in-u-s-say-their-government-watches-their-every-move>.
- Salo, Dorothea (2021). Physical-Equivalent Privacy. *The Serials Librarian*, 81(1), 20-34 <http://digital.library.wisc.edu/1793/81297>.
- Schonfeld, R. and Esposito, J. (2021). [When Consolidation Provides Benefits as well as Market Power](#). *The Scholarly Kitchen*.
- Siems, R. (2021). When your journal reads you – user tracking on science publisher platforms. *Elephant in the Lab*. <https://doi.org/10.5281/zenodo.4683778>
- Sonnemaker, T. (2020). Students Are Protesting Tech Tools That Police Them Like Criminals. *Business Insider*. <https://www.businessinsider.com/proctorio-silencing-critics-fueling-student-protests-against-surveillance-edtech-schools-2020-10>
- SPARC (2022) [Elsevier's Acquisition of Interfolio: Risks and Responses](#). Outlines the increasing concentration in the nascent market for FIS products. The document details the ways that Elsevier's power in the FIS market may impact other markets and the subsequent risk of negative effects on the academic community, and provides concrete suggestions for how institutions can respond.
- Stewart, C. (2017). Glasgow University Staff Apologise After "Tracking Students". *Glasgow Times*. <https://www.glasgowtimes.co.uk/news/15292214.amp/>.
- Swauger, S. (2021). Leaving Surveillance Tech Behind in Higher Education; Towards Trust and Abolition. *Tech X Belonging, Other and Belonging Institute*. <https://belonging.berkeley.edu/leaving-surveillance-tech-behind-higher-education-towards-trust-and-abolition>.
- Tian, E. (2020). YaleStudents Website Generates Student Debate, Reveals Potential Data Governance Gap in University Policy. *Yale Daily News*. <https://yaledailynews.com/blog/2020/11/17/yalestudents-website-generates-student-debate-points-out-potential-governance-gap-in-university-policy/>.
- Tijerina, B. and Zimmer, M. (2018). Vendors and Privacy. *Privacy Field Guides for Libraries*. <https://libraryprivacyguides.org/vendors-and-privacy/>.

- University of Hawaii (2018). Resolution Supporting Learning Data Privacy Principles and Practices. Mānoa Faculty Senate.
<https://www.hawaii.edu/uhmfs/issues/resolution-supporting-learning-data-privacy-principles-and-practices/>.
- U.S. Department of the Treasury. Freedom of Information Act.
<https://home.treasury.gov/footer/freedom-of-information-act>
- Waldman, A. (2021). Industry Unbound: The Inside Story of Privacy, Data, and Corporate Power. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108591386>
- Yoose, B. and Shockey, N. (2023). Navigating Risk in Vendor Data Privacy Practices: An Analysis of Elsevier's ScienceDirect. SPARC. <https://zenodo.org/records/10078610>