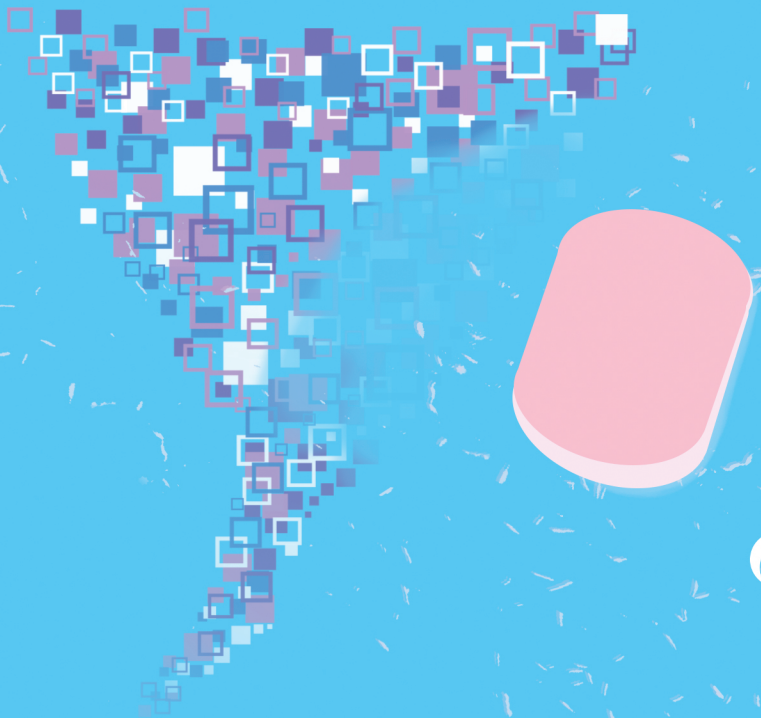
BALANCING THE SCALE OF

# STUDENT DATA DELETION AND RETENTION IN EDUCATION

MARCH 2019



A SERIES OF PAPERS ON STUDENT PRIVACY

CENTER FOR DEMOCRACY & TECHNOLOGY

## ABOUT CENTER FOR DEMOCRACY & TECHNOLOGY

The Center for Democracy & Technology is a 501(c)(3) working to promote democratic values by shaping technology policy and architecture, with a focus on the rights of the individual. CDT supports laws, corporate policies, and technological tools that protect privacy and security and enable free speech online. Based in Washington, D.C., and with a presence in Brussels, CDT works inclusively across sectors to find tangible solutions to today's most pressing technology policy challenges. Our team of experts includes lawyers, technologists, academics, and analysts, bringing diverse perspectives to all of our efforts.

*Learn more about our experts and the issues we cover: https://cdt.org/*

## ABOUT STUDENT PRIVACY

CDT's vision for the *Student Privacy Project* is to create an educated citizenry that is essential to a thriving democracy by protecting student data while supporting its responsible use to improve educational outcomes. To achieve this vision, CDT advocates for and provides solutions-oriented resources for education practitioners and the technology providers who work with them, that center the student and balance the promises and pitfalls of education data and technology with protecting the privacy rights of students and their families.

## AUTHORED BY

***Elizabeth Laird, CDT Student Privacy Senior Fellow***
***Hannah Quay-de la Vallee, Senior Technologist***
***Research support by Adarsh Mahesh***

# Balancing the Scale of Student Data Deletion and Retention in Education

## Executive Summary

Schools collect a lot of data about students, which can be valuable for improving student outcomes. For example, this information can assist with identifying students who are at risk of dropping out, allowing teachers to intervene early on. However, that same information can pose a substantial risk to students and their families if it is not managed well. Deleting data is much more complicated than one might think, with a number of important policy, legal, and technical considerations. This issue brief offers three recommendations and related best practices to assist the education sector in achieving the right balance of retaining useful data that can serve students with deleting information that is no longer needed. They are:

- Conduct comprehensive inventory of student data;
- Create an organizational student data retention policy; and
- Implement technical best practices when deleting student data.

To assist education leaders and the companies with which they work, the brief provides practical resources that can be adapted to implement these recommendations, including samples of a student retention policy, a student data inventory template, a deletion certificate, and an initiative kick-off letter. Striking the right balance between data retention and deletion is not an easy task and is never finished; however, the goal of this issue brief is to empower education practitioners and the companies they work with to adopt a student-centered approach that maximizes the value of data and technology while protecting privacy rights.

## Introduction

Historically, the education system has erred on the side of retaining data – keeping student data indefinitely with the intent of better serving students and their families. There are many ways that data can can be useful in improving student outcomes like ensuring data access by former students and supporting longitudinal research. However, indefinite data retention does not come without risk or cost. For example, data that has been retained but is no longer relevant has the potential to be used out of context. Education leaders and the companies they work with have to balance student data deletion and retention. This brief explores this issue and offers recommendations that protect student privacy while supporting effective data use that improves student outcomes.

## The Problem with Indefinite Data Retention

Schools, districts, and state education agencies collect a great deal of information about students and their families to improve their practices and provide better outcomes for students. However, large data stores come with risks and drawbacks. On the logistical side, larger data sets are more expensive to maintain, search, and store. If the data is not carefully managed to ensure it is consistently high-quality, it can be difficult to glean useful insights from the noisy or missing data.

Most importantly, this data can pose a threat to students and their families. Data used maliciously as a result of a data breach or exposure, or even just used outside of its intended context, can do significant harm to students. Take, for instance, the Dark Overlord attacks against Johnston Community School District in Iowa.[1] A hacking group that calls itself Dark Overlord obtained information from school records, such as student names and parent phone numbers, and used this information to text threats against the children to their parents. Although law enforcement ultimately considered the threats non-credible, several schools closed for one to two days to protect their students. A similar attack occurred in Montana's Flathead County, where over 30 schools were closed for three days.[2]

In the face of the increasing attention showed to schools by malicious hackers,[3] it is important that educational institutions protect their students' digital privacy and wellbeing. A key way institutions can do that is by "minimizing" the data they maintain on their students, whether by deleting data once it is no longer needed or by limiting the amount of data collected with which to begin. This "data minimization" approach is an effective way to mitigate the potential harms that arise from maintaining an excess of student data. Smaller data sets are less expensive to maintain, allowing schools to preserve resources. Ultimately, the fundamental value of minimizing data is that data that does not exist cannot be misused. Thus, minimizing data, whether by deleting unneeded data or by limiting unnecessary data collection, protects students from the harms that data could cause.

---

[1] Cox, J. (2017, Oct 5) '*Dark Overlord' Hackers Text Death Threats to Students, Then Dump Voicemails From Victims.* Retrieved from: www.thedailybeast.com/dark-overlord-hackers-text-death-threats-to-students-then-dump-voicemails-from-victims.
Hilyard, H. (2017, Oct 6) *Here's why hackers are targeting Iowa schools, children.* Retrieved from www.kcci.com/article/threats-force-johnston-schools-to-cancel-classes/12769814.
*Iowa Schools Closed by Threats to Reopen Wednesday.* (2017, Oct 3) Retrieved from: www.usnews.com/news/best-states/iowa/articles/2017-10-03/threats-force-suburban-des-moines-district-to-cancel-classes.

[2] Larson, S. (2017, Oct 18) *Hackers are targeting schools, U.S. Department of Education warns.* Retrieved from: https://money.cnn.com/2017/10/18/technology/business/hackers-schools-montana/index.html.

[3] Campbell, S. (2018, Aug 28) *Why schools are prime targets for data breaches.* Retrieved from: https://www.wpri.com/back-to-school/why-schools-are-prime-targets-for-data-breaches/1400415386#.

## Stakeholder Engagement

Education leaders will be more successful in creating and executing a balanced student data deletion and retention strategy if they involve diverse stakeholders in this process. Education leaders will need to identify the most critical stakeholders but should, at a minimum, consider the following perspectives:

| *Stakeholder* | Why They Care |
|---|---|
| *Parents and Students* | Parents and students have the most at stake when it comes to the decisions that are made about their data. As broader trends in privacy are aimed at empowering the consumer to make decisions about their data, the education system should also consider how to meaningfully engage parents and students in these discussions. |
| *Policymakers* | Policymakers often rely on data that is collected and analyzed by school districts and states to inform policy decisions. They should be aware of any changes to what is collected and maintained about students to prevent any surprises in the future when they ask for data that is no longer available. |
| *Educational Technology Vendors* | The Family Educational Rights and Privacy Act (FERPA) requires that educational technology (EdTech) vendors delete student data when there is no longer a purpose for it, including when a contract or data sharing agreement expires. Education leaders should ensure that they have communicated expectations related to the technical practices that should be applied when they delete student data. |
| *Privacy Advocates* | Privacy advocates are often in touch with concerns from parents and educators and may have expertise that extends beyond education and provide useful feedback about best practices and trends in other industries. |
| *Researchers* | Similar to policymakers, researchers rely on data that is collected and analyzed by school districts and states to inform policy decisions, so they should be informed if data is no longer available. Additionally, similar to EdTech vendors, researchers are required to delete student data once the research has concluded. Education leaders should ensure that they have communicated expectations related to the technical practices that should be applied when they delete student data. |
| *Other Government Agencies* | Students may be served by other state agencies that could benefit from integrating student information. In determining the usefulness of student data, other government agencies should be consulted as they may have legitimate use cases that could improve outcomes for students that should be considered when making decisions about whether to delete student data. |
| *Public* | Members of the public frequently request information from school districts and states. Education leaders should consider how to communicate what information is available and if the requested information has been deleted, including the rationale for those decisions. This can help manage data and Freedom of Information Act (FOIA) requests that cannot be fulfilled. |

CENTER FOR DEMOCRACY & TECHNOLOGY

## How Do Federal & State Laws Address Student Data Deletion & Retention?

Legal requirements are important considerations when informing an education institution's deletion and retention activities. Typically, legal requirements should inform but not be the main driver to such activities as additional actions beyond what is legally required are needed to strike the right balance between data retention and deletion. These legal requirements can be both federal and state in nature as well as education-specific, child-focused, or more general.

At the federal level, three laws primarily inform a student data deletion and retention strategy:

| *Federal Laws* | Deletion and Retention Requirements |
|---|---|
| *Family Educational Rights and Privacy Act (FERPA)*[4] | • Parents can request that an education agency amend any information in the student record that they deem incorrect or misleading.<br>• An educational agency cannot destroy any education records if there is an outstanding request from a parent to inspect and review the records.<br>• Third parties, like vendors and researchers, are mandated to destroy all personally identifiable information when it's no longer needed for the purpose for which it was disclosed. |
| *Children's Online Privacy Protection Act (COPPA)*[5] | • A parent has the right to refuse to permit the operator's future online collection of personal information from a child who is under 13 years old and direct the operator to delete the child's personal information.<br>• An operator of a website shall delete all personal information once it is no longer necessary to fulfill the purpose for which the information was initially collected. |
| *Individuals with Disabilities Education Act (IDEA)*[6] | • Public agencies must disclose to parents when the personally identifiable information collected, maintained, or used is no longer needed to provide educational services to the child, and the information must be destroyed if the parent requests so.<br>• A permanent record of a student's name, address, phone number, grades, attendance record, classes attended, grade level completed, and year completed may be maintained without time limitation. |

[4] 34 C.F.R. §§ 99.10, 99.20, 99.31 (2019). Retrieved from: https://www.ecfr.gov/cgi-bin/text-idx?rgn=div5&node=34:1.1.1.1.33.
[5] 16 C.F.R. §§ 312.6, 312.10 (2019). Retrieved from: https://www.ecfr.gov/cgi-bin/text-idx?SID=50aa19d46a91816536da1cd3c6ba5c6c&mc=true&node=pt16.1.312&rgn=div5.
[6] 34 C.F.R. § 300.624 (2019). Retrieved from: https://www.ecfr.gov/cgi-bin/text-idx?SID=7e53d34ff60b04cde95cdc59e4ebf85c&node=pt34.2.300&rgn=div5#se34.2.300_1624.

At the state level, laws that should inform an education institution's data deletion and retention strategy can be education-specific, child-focused, and/or address data disposal more generally. California, for example, has at least nine distinct laws that should be considered when formulating a student data deletion and retention strategy. As described in more detail in Appendix D, five education-specific laws in California impose requirements related to empowering parents and students with data deletion rights, adhering to data deletion standards for information that is maintained by schools and districts, and holding third parties accountable for deleting student data that is shared with them. In addition to these laws, it has a child-focused law regarding companies that collect information from minors and gives users rights to request the deletion of such information. Lastly, it has three general data disposal laws that would inform student data retention and deletion in that they require companies to meet certain standards when destroying data, empower consumers with rights to ask that companies delete data, and post a privacy policy that provides details about the process for users to delete their data. Although these laws are insufficient to generate a comprehensive data retention and deletion strategy, understanding how these disparate requirements should inform a data retention and deletion approach is a complicated undertaking.

Lastly, current trends in consumer privacy legislation are focused on empowering the consumer to make decisions about their personal data. California, as mentioned above, has several laws that inform data deletion and retention in education, one of which is the California Consumer Privacy Act, which provides more general deletion rights to all consumers. In the European Union, the General Data Protection Regulation (GDPR) provides rights to consumers regarding the deletion and retention of their data. Similar efforts are underway to empower consumers in the U.S., which could affect an education institution's data deletion and retention requirements and are beginning to appear in state student privacy laws like Utah's recent bill (see below).

| *Representative Laws* | **Deletion and Retention Requirements** |
|---|---|
| GDPR Article 17 [Right to deletion][7] | • In response to a request from a data subject, the controller is obligated to delete all personal data concerning the subject where one of the following applies: (a) the data is no longer necessary for the purposes for which they were collected; (b) the data subject withdraws consent on which the processing is based, and there is no legal ground for the processing; or (c) the personal data was unlawfully acquired or processed.<br>• Where the controller has made the personal data public, and one of the above conditions applies, the controller is obligated to take reasonable steps, taking account of available technology and the cost of implementation, including technical measures, to inform other controllers (third parties) that are processing the data that the data subject has requested their personal data to be deleted. |

[7] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, 2016 O.J. (L 119) 1 (EU), art. 17. Retrieved from: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

| *Representative Laws* | Deletion and Retention Requirements |
|---|---|
| Utah 53A-1-1407[8] | • Requires that an education entity expunge a student's data if the student is at least 23 years old and the student requests that this data is expunged. |

## *What are Best Practices for Student Data Deletion and Retention?*

Education leaders and the third parties they work with can effectively address student data deletion and retention by focusing on three key actions:

1. Conduct comprehensive inventory of student data;
2. Create an organizational student data retention policy; and
3. Implement technical best practices when deleting student data.

### 1. Conduct Comprehensive Inventory of Student Data

Before building a data management plan, it's important to know what information an organization has. Inventorying data can feel like a herculean task, but it's an important first step in an overall data management program. A data inventory should document what data an organization has, what format the data is in, where that data is stored and what system it is in, and how it is used and by whom.[9] The inventory should include both digital and paper records that are maintained about students. So, where to start?

A primary goal of a data inventory in education is to compile a list of what data an organization maintains about students. This list might include, but is not limited to, student demographics, attendance, assessment information, special education information, disciplinary records, and health-related data. To provide an example of what a data inventory might produce, Appendix B excerpts the Wisconsin Department of Public Instruction's data inventory.

Inventorying and documenting student data will be most thorough and effective with the following steps:

- Establish a representative group to participate in data inventory process;

[8] Utah Code Ann. § 53A-1-1407 (2016). Retrieved from: https://le.utah.gov/~2016/bills/static/HB0358.html#53a-1-1407.
[9] GovEx Labs, *Data Inventory Guide.* Retrieved from: https://labs.centerforgov.org/data-governance/data-inventory/. Accessed 2019, Feb.

- Consider how to engage the entire organization;
- Document consistently key components of data and systems applications; and
- Continue to reassess data and systems applications on an ongoing basis.

### *Establish a Representative Group to Participate in Data Inventory*

Inventorying and documenting student data should be an iterative and inclusive process. The inclusiveness component is key because people in different roles collect, produce, and maintain different kinds of data, and it can be difficult to assess from the outside what all that data is. Establishing a representative group of stakeholders who collect and maintain student data will provide good coverage of roles (for instance, data may be maintained by multiple departments and at different levels of the organization). This team will be responsible for working with their colleagues to ensure all data is known and documented and can act as liaisons to broader groups of employees. This group should remain actively involved throughout the duration of the data inventory. A sample student data deletion and retention initiative kick-off letter is provided in Appendix D.

### *Consider How to Engage the Entire Organization*

In addition to establishing a representative group to participate in the data inventory, consider how to engage all members of the organization. Sending out a survey to all employees asking about how they use and collect data in their work can help to ensure that the inventory process does not miss any data (such as a teacher storing student grades and outcomes in a spreadsheet to track his own progress as a teacher) and will ensure awareness that the data inventory is happening, why it is important, and how it will help the organization improve its management of student data.

### *Document Consistently Key Components of Data and Systems Applications*

After the initial survey, examine the results and compile a list of data sources, types, and systems that store data. The stakeholder group should review the list to identify any places where data or systems have been omitted, and ensure those are incorporated into the inventory.

### *Continue to Reassess Data and Systems Applications on an Ongoing Basis*

Data inventorying should not be a one-time event. Data should be reviewed on a regular schedule, and as situations arise that may affect the data map. Think of it like health insurance: An individual re-enrolls every year, which offers a regular opportunity to assess a health plan to make sure it still makes sense for the individual. In the data inventory world, an organization

should reevaluate whether the data map still matches the actual state of the data. Again like health insurance, in addition to regularly scheduled re-evaluations, there are also qualifying events that should trigger a reevaluation. For data, this might be adoption of new systems or phase-outs of old databases. Make sure any data mapping and inventorying documents include any new data collected by the new system, or any data retired along with the old database.

### *Approach the Inventory Process in a Manageable and Comprehensive Way*

In addition to getting organization-wide engagement, another tactic to organize a data inventory is to start high level by collecting an inventory of all systems and repositories that contain, produce, or collect data. Next, inventory each system and repository in more detail. Starting high level helps ensure you do not miss systems that contain data, and allows you to run your data inventory in stages. After inventorying each new system, perform a reconciliation step to determine if you are collecting the same data in multiple places. Duplicated data often indicates unnecessary risk. See if it is possible to store that data set in a single place and reference the central copy, rather than duplicating it.

## 2. Create an Organizational Student Data Retention Policy

The potential of a data-rich environment is better insights about students and educational practices that lead to improved outcomes and better educated citizens. However, in order to achieve those goals, data has to be well managed and potential risks mitigated. Developing a data retention and deletion schedule helps ensure the data hygiene of your organization and that you are not introducing risks to students by maintaining information longer than is necessary. To see an example of a retention schedule, Appendix A provides an excerpt of Colorado's retention policy.

A data retention schedule should include information on:

- How long each type of data is to be retained;
- Deletion practices required for each type of data;
- How permanently retained data will be archived;
- Any legal obligations that inform the retention rules for a particular type of data (for instance, if a law mandates retention or deletion of a particular type of data); and
- Any situations that would override the retention schedule (for instance, ongoing litigation may require certain data to be retained for the duration of the litigation, even if the schedule would mandate deletion).

***Archiving Data for Long-Term Storage:***

For data that must be retained for long periods of time (e.g., ten years or longer) and is accessed infrequently, archiving that data can mitigate some of the risks associated with maintaining it. Archiving is the practice of moving data into offline storage media, so it can be more easily maintained for long periods of time. Generally, this means moving it out of any databases or systems that hold frequently accessed data. Because the archived data is out of the day-to-day flow, it is less likely to be accidentally lost or modified. Additionally, because archived data is usually offline (i.e., not stored as part of a system that is connected to a network), it is at far lower risk of being hacked or accidentally disclosed (however, the data does still require physical security).

There are a number of different media options when archiving data, such as traditional hard drives, solid state drives, Linear Tape-Open (LTO), Blu-ray Recordable (BD-R) discs, etc. These media carry a range of tradeoffs that organizations will have to weigh when determining if and how to archive data. Chief among these factors is cost. The primary factor driving cost will be dollars per terabyte of the media itself, but any other equipment required to archive data will contribute to the cost as well. For example, LTO tapes are generally cheaper per terabyte than hard drives, but the tape drive required to write them can cost several thousand dollars, whereas hard drives could be written by computers the organization already has. Another factor to consider is how long the media will last and what sort of maintenance it needs. Hard drives can last a decade or two and require that the data on them be "refreshed" (rewritten to the drive) every few years. Some types of BD-R discs can last for several decades, if stored properly. Ultimately, an organization has to weigh these factors to determine which archival solution (if any) makes sense for them.

When designing a data retention schedule, the following practices support a realistic policy that is grounded in the needs of the organization while ensuring student privacy:

- Solicit opinions from a range of stakeholders: legal, technical, parents, students, teachers, etc.
- Have a single point of responsibility for deleting data. Even if the task is delegated, there must be a single person who is ultimately responsible for the enforcement of the retention schedule. This person should be a high-level employee, such as the Chief Information Security Officer (CISO) or Chief Information Officer (CIO).

- While there should be a single point of responsibility in general, each type of data should have a steward who is responsible for the maintenance of that data, and for deleting that data in accordance with the schedule.
- There should be an auditable "deletion trail" that shows what data was deleted, when and how it was deleted, and by whom. (In the case of digital records, this information is logged ideally by default, and is unalterable. For paper records, this may mean sign-off sheets indicating the date of deletion and the person who carried out the deletion.)
- If a class of data is due to be deleted, but incorporated into an aggregate statistic, ensure that this is clearly noted in the retention policy. Where possible, ensure that a personally identifiable record cannot be reconstructed from the aggregate statistic.[10]
- Consider making the retention and deletion schedule a public document, or creating a public-facing version. In addition to providing guidelines for the organization and its employees to follow around data management, a retention schedule can also help parents and students understand how their data is being used and managed.

A retention schedule is most effective as part of broader data management plan. Other aspects of that plan might include:

- Usage guidelines for different types of data;
- Limited and appropriate controls for data access as well as employee training on protecting information to which they have access;
- Protocols for students and parents to access their data (and to request updates, amendments, or corrections of that data as needed); and
- Clear documentation about what data is collected and how it is used; if an organization is collecting data that does not have a clear use, consider halting that collection so resources can be dedicated to better maintenance of the more valuable data (in general, consider the principle of data minimization).

## 3. Implement Technical Best Practices When Deleting Student Data

Organizational policies around data are a key component of keeping student information safe, but to be effective those policies have to be implemented in a technically sound and secure

[10] The National Institute of Standards and Technology (NIST) and the National Center for Education Statistics (NCES) both provide guidance on the risks and best practices around de-identifying data.
*Deidentification@NIST.GOV.* Retrieved from: https://www.nist.gov/itl/iad/deidentificationnistgov.
*Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting.* (2010, Dec) Retrieved from: https://nces.ed.gov/pubs2011/2011603.pdf.

way. The following recommendations are intended to provide actionable technical guidance for data management with a focus on disposal and responsible storage of data.

- Encryption can protect information, whether it is used as a deletion substitute (see the table below for some of the advantages and disadvantages of this approach), or as a way to keep still-in-use data safe from unauthorized access. Use encryption for all data holdings at rest and in transit, regardless of sensitivity.[11]
- When media (computers, flash drives, etc.) is being retired (whether to be thrown away, sold off, or otherwise repurposed), sanitize or destroy the media before it leaves your control.
    - If the media is going to be thrown away, follow National Institute of Standards and Technology (NIST) guidelines for making the media permanently unreadable. This might mean shredding documents with a crosscut shredder, or pulverising a disk drive, or whatever makes sense for the particular type of media of which you are disposing.
    - If the media is going to be donated, sold, or otherwise repurposed, follow best practices for clearing the media (overwriting for disk-based drives, deletion-via-encryption, and factory-wipe for flash-based media).
- For data stored by a third party, contracts and data-sharing agreements should require that the third party provide deletion certificates (Appendix C provides an example certificate used by Washington D.C.'s Office of the State Superintendent of Education). This certificate should include:
    - What data was destroyed;
    - What method was used to destroy the data;
    - The date the data was destroyed; and
    - The individual that deleted the data.

Another approach to data "destruction" is de-identifying data so that it can still be used for research and analysis purposes, but cannot be linked to any specific student. De-identification can be an effective tool for preserving student privacy while still gaining the benefit of the data. However, the de-identification must be done correctly to limit the risk of the data being re-identified (i.e., a particular record could be re-associated with the student to whom it belongs). Re-identification is a particular risk when data sets are combined, so these techniques

[11] NIST offers a guide on encrypted storage technologies that includes several useful standards for different contexts, including full disk encryption, virtual disk encryption, as well as individual file or folder encryption and includes guidance on steps to design and deploy cryptographic solutions.
Scarfone, K., Souppaya, M., and Sexton, M. (2007, Nov) *Guide to Storage Encryption Technologies for End User Devices*. Retrieved from: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublica-tion800-111.pdf.

should only be used with caution, and it is important to place limits on the sharing and reuse of de-identified data. (Please see the box for more detail.)

### *De-identifying Data*

One approach to deletion is to remove students' personally identifiable information so that the remaining information cannot be linked to an individual student. To meet the definition of de-identification in FERPA, education entities must remove enough student information such that, "a student's identity is not personally identifiable, whether through single or multiple releases, and taking into account other reasonably available information."[12] However, this is more complicated than it might seem. For example, approaches to de-identification can range from simply deleting direct identifiers like student name or ID number (which is typically not sufficient to prevent the data from being re-identified)[13] up to more sophisticated techniques like shuffling or adding noise to the data which make recovery more difficult (these more complex approaches are generally referred to as "anonymization" in computer science). Whichever of these methods is used, it is important to understand the techniques and business rules that are being applied when taking steps to remove personally identifiable information as a form of deletion because, depending on the approach, data may still be recoverable and thus not actually deleted. As a result, this approach should be taken with extreme caution and de-identified data sets should carry re-use limitations when shared.[14]

There are several different technical approaches to deletion. The following table *(next page)* lays out a spectrum of technical approaches, along with some of the advantages and drawbacks of each approach. Green-light approaches are strong forms of deletion that will provide sufficient deletion for any context (though they may not be feasible in every case). Yellow-light approaches may be sufficient depending on the context and the data in question, but still carry some risk of data recovery, so they may not be suitable for highly sensitive data. Red light approaches carry a high risk of data recovery, and are not recommended.

---

[12] 34 C.F.R. § 99.31 (2019). Retrieved from: https://www.ecfr.gov/cgi-bin/text-idx?rgn=div5&node=34:1.1.1.1.33.

[13] Ochoa, S., Rasmussen, J., Robson, C., and Salib, M. (2002, Aug). *Reidentification of Individuals in Chicago's Homicide Database: A Technical and Legal Study. Massachusetts Institute of Technology.* Retrieved from: https://www.researchgate.net/publication/2838440_Reidentification_of_Individuals_in_Chicago%27s_Homicide_Database_A_Technical_and_Legal_Study.

[14] Guidance on the risks and best practices around de-identifying data can be found from NIST at https://www.nist.gov/itl/iad/deidentificationnistgov and from NCES at https://nces.ed.gov/pubs2011/2011603.pdf.

| Strength | Deletion Method | Description | Use Cases |
|---|---|---|---|
|  | Data overwrite | Overwriting data can be time consuming for large data sets, but it is a much more effective method of deletion than a soft delete (see below) and can be used for individual files, rather than wiping the entire drive. | Use this approach to delete data from a disk drive, either before it is reused for another purpose, or before destroying the drive. |
|  | Solid state drive factory wipe | Overwriting is not a viable method for solid state drives, because they handle memory allocation differently than traditional disk drives. Most solid state drives have a factory wipe option to erase data from the drive. However, this method typically erases the entire drive, and is not available for individual files. | Use this approach to clear a solid state drive (flash-based hard drives), either before it is reused for another purpose, or before destroying the drive. |
|  | Media destruction | Physical destruction of media is the most extreme method of deletion. It is expensive, because the media obviously cannot be reused or sold. Methods that will destroy a disk-based drive (such as demagnetizing the drive) may be ineffective on a solid-state or flash-based drive, so it is important to know what type of drive is being destroyed. | Use this approach when the media is due to be thrown away. Overwrite or factory-wipe the drive first. |
|  | Soft delete | Basic “deletion” operations, such as dragging a file to the trash and emptying the trash, do not remove information, but rather signal to the operating system that block of memory is free for reuse. Thus, the information contained there can still be recovered up until the OS repurposes that memory. | Use this when deleting low-sensitivity data elements. If, however, you are erasing an entire drive, use overwriting or factory-wipe, even for low-sensitivity data. |
|  | Deletion via encryption | Once data is encrypted and the encryption keys are erased, the information is rendered irretrievable, the same goal of | Use this method for protecting information in cloud environments |

| | | | |
|---|---|---|---|
| | | deleting information. However, if the encryption is broken or the key is guessed, the data can be recovered, making this approach less safe than overwriting, factory wiping drives, or media destruction. | where access to the physical media is restricted. |
|  | De-identification with limits on reuse and sharing | When data is retained for research purposes, but no longer needs to be associated with a particular student, de-identifying the data protects the student while allowing the data to be used for research. Limiting sharing and reuse helps protect against re-identification. Effectively de-identifying data requires statistical expertise to minimize risk of re-identification. | Use this approach when sharing data with external researchers, or preparing data for internal research. |
|  | De-identification with no limits on reuse and sharing | Allowing de-identified data to be re-shared and repurposed without limitations fails to manage the risk of re-identification. Therefore, the data cannot be considered deleted. | We do not recommend this approach. Without specifying limits on reuse or sharing, it is difficult to ensure the data will not be re-identified. |

## Conclusion

Although student data deletion and retention are technical concepts, they are also critical strategies to protecting and utilizing student data and should be led by the highest levels of an organization. These concepts are also much more difficult than they might initially seem. Crafting sound data deletion and retention policies, complying with relevant federal and state laws, and deploying technical best practices requires involvement from diverse stakeholders within and outside the education system, from parents to policymakers. They also require a dynamic approach which recognizes that data has a lifecycle and must be constantly maintained, safeguarded, and deleted. Recognizing this reality will allow education leaders and the companies they work with to fulfill their responsibilities to protect student data while realizing the benefits of data and technology to improve student outcomes.

# APPENDIX

A

B

C

D

E

CENTER FOR
DEMOCRACY
& TECHNOLOGY

**Appendix A: Sample student data retention policy**
Excerpted here is a portion of Colorado's retention policy, the record schedule for student records. The full record schedule can be found at
https://www.colorado.gov/pacific/sites/default/files/SchoolsRMManual.pdf

# RECORDS MANAGEMENT MANUAL - SCHOOL DISTRICTS

## SCHEDULE 3

## STUDENT SERVICES RECORDS

General Description: Records generally relating to academic records of children within the school district. The specified retention period applies to the information contained within the record, regardless of the physical format of the record (paper, microfilm, computer disk or tape, optical disk, etc.).

Duplicate Copies: Provided that no retention period is specified for duplicate copies, retain those that are created for administrative purposes for 1 year, and retain those created for convenience or reference purposes until no longer needed or for 1 year, whichever is first. Duplicate copies should not be retained longer than the record copy.

NOTE: RECORDS OF THE STUDENT FROM ELEMENTARY, MIDDLE AND JUNIOR HIGH SCHOOL SHOULD BE MERGED INTO THE STUDENT PERMANENT RECORD WHEN HE OR SHE REACHES HIGH SCHOOL.

1. **Student Permanent Record** These records are divided into three categories: personal information, enrollment history and academic performance. Each Colorado school district keeps information about students in different ways and on different forms. Therefore, the retention schedule presents the kinds of information or data elements that are maintained in files, rather than the names of the forms on which information may be found.

    a. Personal Information -- This information, except for the immunization record, is usually found with the student's permanent record.
        i. Student's identification number - A number used for recordkeeping purposes. It might be one assigned by the district or a Social Security number
        ii. Legal name of student
        iii. Legal name of parent or guardian
        iv. Date of birth
        v. Address
        vi. Sex
        vii. Telephone number
        viii. Immunization record for withdrawals

b. Enrollment History -- This information may be with the transcript or it may be on a different form, depending upon the district. It consists of the following:
   i. Exact date the student enrolled in the district
   ii. Name, city and state of the previous school(s) attended outside the district
   iii. The schools attended within the district
   iv. The dates and grade levels of the student
   v. Date the student withdrew or graduated from the district
   vi. Name, city and state of the school to which the student is withdrawing

c. Academic Performance -- usually found on the transcript or on report cards.
   i. Classes and/or grade level taken
   ii. Semester grades
   iii. Postsecondary courses/semester grades
   iv. Standardized test scores
   v. Advanced placement (AP) test scores
   vi. Grade point average (GPA)
   vii. Class rank
   viii. College placement test scores (i.e., ACT/SAT)

Retention: Permanent

2. **Student Fall Enrollment Report (October Count)** (Report to the Colorado Department of Education of the number of students enrolled.)

   Retention: Permanent

3. **Student End of School Year Enrollment** Report to the Colorado Department of Education that reports the number of students in school at the close of the academic year.

   Retention: Permanent

4. **Student Cumulative Records** that contain optional information on students attending school in the district. The record may contain but is not limited to:
   a. Other such information as shall enable school officials to counsel with students and plan appropriate activities.
   b. Immunization record for graduates
   c. Ethnic code - This code is of use only to the district in which the student is enrolled.
   d. Withdrawal Grades (sometimes called grades in progress) - Withdrawal grades are not official grades, i.e., no credits are earned. Their purpose is to facilitate enrollment at the student's next school.

e. Supplementary programs - Examples of such programs are gifted and talented, bi¬lingual, English as a Second Language (ESL), Chapter 1, etc.
f. Health records - Hearing and vision screenings, visits to the school clinics, or similar records are not required information.
g. Signed releases of records - The purpose of this record is to document whether or not student record information was released, as requested by the parent or student.
h. Progress reports - Mid-semester grades which inform parents and students of how the student is doing. These are not official grades and do not have long-term value.
i. Out-of-district records
j. School fines
k. Emergency information
l. Marriage licenses - Students may obtain a copy from the state or country in which they were married. It is not the responsibility of the school district to maintain these records permanently.
m. Birth certificates - Students may obtain a copy from the state or country in which they were born. It is not the responsibility of the school district to maintain these records permanently.
n. Court orders denying access to records
o. Adoptions - The child's legal name should be changed on the transcript, although the previous name should also remain part of the transcript. It is not the responsibility of the school district to maintain permanent adoption records.
p. Guardianships - It is not the responsibility of the school district to maintain guardianship records.
q. GED records - This information is retained permanently at the Colorado State Department of Education.

Retention: (1) For graduates: purge immediately after graduation. (2) For withdrawals: destroy after the student leaves the district unless there is a compelling need to keep it longer.

5. **Student Drop Out Records** are distinct from the student cumulative record and are maintained as a separate file.

   Retention: 10 years

6. **Student Transfer In-Transfer Out Records** are distinct from the student cumulative record and are maintained as a separate file.

   Retention: 10 years

7. **New Student Orientation Schedules**

   Retention: 1 year

8. **Report Cards** that document the periodic report by a school about a student's academic, social, emotional, and physical progress. Information includes but is not limited to full legal name of student; teacher's name; name and address of school; indication of attendance during reporting period; grades; and other related information.)

   Retention: 1 year after school year in which records were created provided semester grade is recorded in the student permanent record

9. **Student Schedules File** of forms completed by school personnel for student scheduling into class. Information includes printouts of student schedules, class lists, student class assignments and requests for change of schedule.

   Retention: Until no longer needed for administrative purposes, then destroy

10. **Student Discipline, Suspension, and Expulsion Records** documenting inappropriate student behavior and corrective actions taken. Information includes referral and action form, notes, letters to parents, suspension documentation, detention documents, hearing notices, bus driver referrals, statements and conference notes.

    Retention: (1) When suspended and subsequently expelled permanently: Transfer to Student Permanent Record File and retain until student reaches the age of 21. (2) When disciplined or temporarily suspended and returned to school with no further rules infractions: 3 years Return to Colorado School District Records

11. **Student Truancy Records** - Records created to document student's excessive absences and action taken to correct the problem by school personnel. Information includes referral and action forms, letters to parents, attendance profile sheets, correspondence, release forms, copies of initial court petitions, copies of court orders, hearing notes, affidavits and visitation documentation.

    Retention: 3 years after school year in which records were created

CENTER FOR DEMOCRACY & TECHNOLOGY

## Appendix B: Sample student data inventory template

The Wisconsin Department of Public Instruction provides an example of what a data inventory looks like and can serve as a useful template. For instance, their description of a disciplinary information (reproduced here) clearly describes each data element and explains the purpose for collecting it (including any regulatory requirements). The full inventory is at https://dpi.wi.gov/wise/data-dpi.

| **Student Data Elements (Discipline Data Non-Special Education; Safe and Drug Free Schools)** https://dpi.wi.gov/sites/default/files/imce/wise/pdf/Inventory%20-%20Discipline%20Data%20Non-Special%20Education%3B%20%20Safe%20and%20Drug%20Free%20Schools.pdf | | | | | | |
|---|---|---|---|---|---|---|
| Data Element | Definition | Notes | Data Category | Date Element Implemented | Specific Use/Relevance | Matching Element? |
| Property | Definition | Rules & Notes - SIS | e.g. Student, School, etc. | Initial School Year of Data Implementation | Requirement - State/Federal/ AdminRule | Required for Identity Matching? |
| Incident Date | Date the incident resulting in removal occurred. | An incident is reported only if it is associated with an expulsion or a removal of half a day or more. | student | 2006-07 | Used to uniquely identify incident and associate it with a school year | No |
| Incident Type - Primary | Primary reason for removal. Most serious infraction or offense committed. | | student | 2006-07 | Required by the Wisconsin School Performance Report law, identification of persistently dangerous schools under ESEA Safe School Transfer, ESEA Consolidated Performance Report and ESEA Gun-Free Schools Act Report to the US Department of Education, and (for removals to an Interim Alternative Educational Settings - School Personnel) public reporting and reports to the US Department of Education under IDEA. | No |
| Incident Type - Secondary | Any reason for removal other than the primary reason. | | student | 2006-07 | Used for ESEA persistently dangerous schools and gun free and safe schools reporting | No |
| Early Reinstatement Condition | Condition that a student is required to meet before he or she may be granted early reinstatement or conditions that a student is required to meet after his or her early reinstatement but before the expiration of the term of expulsion specified in the student's expulsion order under s.120.13(1)(c)3 or s. 120.13(1)(e)3, or s.119.25, Wis Stats. See s.120.13(1)(h)1. Wis Stats. | Gathered for expulsions only. | student | 2006-07 | Required by Wisconsin School Performance Report law and is required if Removal Type is Expulsion. | No |
| Expulsion Period Return Year | The school year during which an expelled student is expected to return to school grounds | Gathered for expulsions only. | student | 2006-07 | Used to translate WI definition of expulsion into the IDEA definition. The IDEA definition is used for | No |

| | | | | | | |
|---|---|---|---|---|---|---|
| | based on the expulsion expiration date specified in the expulsion order issued under s.120.13(1)(c) 3 or (e)3, Wis Stats | | | | reporting SwD and SwoD. | |
| Modified Term Firearms | Reduction in the minimum one-year (12 month) expulsion period for firearms. The expulsion period begins on the removal period start date. This reduction is specified in the student's expulsion order under s.120.13(1)(c)3 or (e)3, Wis Stats | Gathered for expulsions resulting from a firearms incident only. | student | 2006-07 | Required for ESEA Gun-Free Schools Act Report and federal EDFacts reporting. | No |
| Removal Type | Type of disciplinary removal for student involved in incident. | Out-of-school Suspension and Expulsion are the only removal types gathered for SwoD. | student | 2006-07 | Required by Wisconsin School Performance Report law, identification of persistently dangerous schools under ESEA Safe School Transfer, ESEA Consolidated Performance Report to the US Department of Education, and public reporting and reports to the US Department of Education under IDEA. | No |
| Removal Period Start Date | Start date of the student's removal type. | | student | 2006-07 | Used to uniquely identify a removal and associate it with a school year. One incident may result in multiple removals | No |
| Return to School after Expulsion | Return to school on or before the expulsion expiration date specified in the expulsion order. | | student | 2006-07 | Required by Wisconsin School Performance Report law. | No |
| School Days Removed This Term | Number of school days during which student is removed for this school term, incident, and removal type | | student | 2006-07 | Required by Wisconsin School Performance Report law and for public reporting and reports to the US Department of Education under IDEA | No |
| Services Provided During Expulsion | The student who was expelled from School A received educational services during the expulsion period during this school term. | Gathered for expulsions only. | student | 2006-07 | Required by Wisconsin School Performance Report law, public reporting and reports to the US Department of Education under IDEA, and the ESEA Gun-Free Schools Report to the US Department of Education | No |
| WSN / WISEid | Wisconsin Student Identifier | | student | 2004-05 | Identifies student. | Yes |
| n/a | | | student - school association | 2004-05 | Every discipline record must be associated to enrollment data submitted separately. | No |
| n/a | | | student, student - special education program association | 2004-05 | Every discipline record must be associated to demographic data submitted separately. | No |

CENTER FOR DEMOCRACY & TECHNOLOGY

**Appendix C: Sample Deletion Certificate**

Presented here is the deletion certificate used by the Office of the State Superintendent of Education of the District of Columbia to certify the personally identifiable information of their students and families has been appropriately deleted.

**CERTIFICATE OF DATA DESTRUCTION**

*Required of All Projects Receiving Access to*
*Personally Identifiable Information from OSSE*

In accordance with the provisions of the data sharing Memorandum of Agreement (MOA) executed on [insert date] and modified on [insert date] between the Office of the State Superintendent of Education and [insert organization], the data files and all related information described below were destroyed as required in Section [insert section] of the Agreement.

_____
Date destruction completed

Description of records destroyed

| File name | Media type | File description | Inclusive dates covered | Comment |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

CENTER FOR DEMOCRACY & TECHNOLOGY

Method of destruction
*Check all that apply*

| X | Type of destruction | Provide details on methods |
|---|---|---|
| | Secure file deletion | |
| | Cross cut paper shredding | |
| | Hard disk physical destruction | |
| | Other (such as external drives) | |

We certify that all copies of the files listed and described, in all media, and by all individuals with access have been destroyed in the manner indicated.

_____ _____

[Insert name of Principal Investigator] Date

_____

[Insert title of Principal Investigator]

Signed:

_____ _____

[Insert name of Witness] Date

_____

[Insert title of Witness]

When signed, please submit this certificate to [insert contact information].

**Appendix D: Sample student data deletion and retention initiative kickoff letter**

Dear colleagues,

I am writing to ask for your participation in an effort to protect student data while ensuring we can support its use to improve educational outcomes. I am launching an effort for [Insert organization name] to balance how long we keep data with getting rid of data that no longer serves students. This is important because we believe deeply in the power of data to improve outcomes for students, but it puts our organization and the students we serve at risk if we keep data beyond its usefulness. It's expensive, makes us a target for hackers, and most importantly, any data that is used out of context in ways that harm students damages public trust in us. Rest assured, we are committed to retaining data that we know is useful, especially when it comes to data we believe individual students may need in the future.

To that end, we are going to take three steps to improve our management of the student data with which we are entrusted:

1. Conduct comprehensive inventory of student data (approximately XX months),
2. Create an organizational student data retention policy (approximately XX months), and
3. Implement technical best practices when deleting student data (approximately XX months).

We want to ensure diverse perspectives are included in this process, which is why we plan to establish a working group that will oversee this process, conduct an organization-wide survey, and provide forums for external engagement with important stakeholders like parents and students. As described above, we estimate that this process could take up to [insert time frame]. We are committed to a transparent, thorough process that serves the needs of [Insert organization name] and the students and families that we serve.

Our initial meeting is on [Insert date, time, location], so I ask that you and/or your designee attend this kick-off meeting as your perspective is critical in this process.

Please do not hesitate to contact me with any questions, and I appreciate your input in shaping policies and practices that are student-centered and protect their privacy while supporting data use that improves their outcomes.

Sincerely,

[Insert name and contact information]

CC: [Superintendent name]

CENTER FOR DEMOCRACY & TECHNOLOGY

**Appendix E: California's Laws Related to Data Deletion**

| | State Law | Deletion and Retention Requirements |
|---|---|---|
| **Education -specific** | California 49062 [General Provisions] | School districts shall establish, maintain, and destroy pupil records according to regulations adopted by the State Board of Education. Pupil records shall include a pupil's health record. No pupil records shall be destroyed except pursuant to such regulations or as provided in Section 49070. |
| | California 49070 [Rights of Parents] | The parent/guardian of a pupil may file a written request to correct/remove any information recorded in the written records concerning his/her child which the parent/guardian alleges to be inaccurate, misleading, or in violation of another student's privacy. |
| | California 49073.1 [Privacy of Pupil Records] | A local education agency that enters into a contract with a third party shall ensure the contract contains:<br>• A description of the procedures by which a parent, legal guardian, or eligible pupil may review personally identifiable information in the pupil's records and correct erroneous information.<br>• A certification that a pupil's records shall not be retained or available to the third party upon completion of the terms of the contract and a description of how it will be enforced. |
| | California 49073.6 [Privacy of Pupil Records] | A school district or charter school that considers a program to gather or maintain in its records any information obtained from social media of any enrolled pupil shall: (a) Provide a pupil an opportunity to correct/delete any information about them obtained from social media (b) Destroy social media information maintained in its records within one year after a pupil turns 18 or within one year after the pupil is no longer enrolled in the school district, whichever occurs first. |
| | California 22584 [Student Online Personal Information Protection Act] | • An operator has to delete a student's covered information if the school/district requests deletion of data under their control. |

| | | |
|---|---|---|
| | California 437 [Retention and Destruction of Pupil Records] | • Mandatory permanent pupil records shall be preserved in perpetuity by all California schools.<br>• Mandatory Interim Pupil Records are those records which schools are required to compile and maintain for stipulated periods of time and are then destroyed as per California statute or regulation.<br>• Permitted pupil records may be destroyed when their usefulness ceases. |
| **Child-focused** | California 22581 [Privacy rights for California minors in the digital world] | An operator of an Internet Web site, online service, online application, or mobile application directed to minors, or one that has actual knowledge that a minor is using its service/application shall permit a minor who is a registered user of the operator's service/application to remove or, if the operator prefers, to request and obtain removal of, content or information posted on the operator's service/application by the user. |
| **General data disposal** | California 1798.81 [Customer Records] | A business shall take all reasonable steps to dispose, or arrange for the disposal, of customer records within its custody or control containing personal information when the records are no longer to be retained by the business by (a) shredding, (b) erasing, or (c) otherwise modifying the personal information in those records to make it unreadable or indecipherable through any means. |
| | California SB 1121 [California Consumer Privacy Act] | A business that receives a verifiable request from a consumer to delete any personal information which the business has collected from them, shall delete the consumer's personal information from its records and direct any service providers to delete the consumer's personal information from their records. |
| | California 22575 [Internet Privacy Requirements] | An operator of a commercial website or online service that collects PII through the internet about individual consumers residing in California who use or visit its website or online service shall conspicuously post its privacy policy on its website or online service and if the operator maintains a process for an individual consumer to review and request changes to any of his/her PII that is collected through the website or online service, provide a description of that process. |