Larry Hogan
Governor

Boyd K. Rutherford
Lt. Governor

Mary Pat Seurkamp, Ph.D.
Chair

James D. Fielder, Jr., Ph. D.
Secretary

Data Suppression Policy for Student Information
Effective: September 3, 2019
Revised: August 16, 2022

Purpose

The purpose of this policy is to help guide the Office of Research and Policy Analysis (the Office) at the Maryland Higher Education Commission (MHEC) in balancing

a) the need to protect student privacy when releasing summary data about Maryland's institutions and students, and

b) the need to ensure the data reported by the Office and MHEC are useful and informative to stakeholders.[1]

This policy applies specifically to *data reporting and publications* by the Office; this policy does not affect the data collected from institutions by MHEC, as outlined in Maryland Statute.[2] This policy applies to all publicly available reports generated by the Office as well as data sharing agreements with contractors and third parties working on behalf of the Office.[3] Last, this policy also applies to Freedom of Information Act requests.

Background

Increased attention to education data has led to an expansion in the amount of information on students and institutions reported by MHEC. Such reports offer a challenge of meeting reporting requirements while also meeting legal requirements to protect each student's personally identifiable information (Family Educational Rights and Privacy Act [FERPA]) (20 U.S.C. § 1232g; 34 CFR Part 99).

It is important to note that much of the material provided by the federal government regarding FERPA compliance and student privacy is in the form of guidance and best practices versus hard and fast rules. In addition, the federal guidance is much clearer on practices around the reporting of K-12 data than reporting the data of students in the postsecondary sector. Helpful entities like the Privacy Technical Assistance Center (PTAC) and the National Center for Educational Statistics (NCES) provide reports and white papers in an attempt to guide state

---

[1] This formal policy went into effect September 3, 2019 and was revised August 2022. The suppression policy was originally implemented in 2019; earlier publications produced by the Research and Policy Analysis unit may not have suppressed data nor has the unit revised past reports to comply with this policy.

[2] Code of Maryland, Education Article (§11-105)
http://mgaleg.maryland.gov/webmga/frmStatutes.aspx?pid=statpage&tab=subject5

[3] The Office may establish data sharing agreements with third party contractors that would allow unsuppressed data to be provided by the Office.

agencies and other organizations. While helpful, these materials have their limitations; they may not have been revised in recent years and therefore may not acknowledge the rapidly changing role that data plays in helping guide policy and practice decisions.

Terms and Definitions

- **Directory Information**: Directory information is information contained in the education records of a student that would not generally be considered harmful or an invasion of privacy if disclosed. FERPA generally permits disclosure of directory information (see 34 CFR § 99.3 and 34 CFR § 99.37). MHEC considers items such as enrollment status, major or program of study, county and state of origin, and degree awarded (in isolation from other information) to be directory information, as these would appear in a campus "directory" or other publicly available documentation.
- **Disclosure:** Disclosure" means to permit access to or the release, transfer, or other communication of PII by any means. Disclosure can be authorized, such as when a parent or an eligible student gives written consent to share education records with an authorized party (e.g., a researcher). Disclosure can also be unauthorized or accidental. An unauthorized disclosure can happen due to a data breach or a loss. An accidental disclosure can occur when data released in public aggregate reports are unintentionally presented in a manner that allows individual students to be identified.
- **Disclosure Avoidance**: Disclosure is avoidance refers to the efforts made to de-identify the data in order to reduce the risk of disclosure of PII. A choice of the appropriate de-identification strategy depends on the nature of the data reporting, the risks inherent in the data, and the usefulness of the resulting data product.
- **Education Records**: Education records are defined by federal regulation to mean those records that are (1) directly related to a student; and (2) maintained by an educational agency or institution or by a party acting for the agency or institution. The full definition can be found in 34 CFR § 99.3 and includes a non-exhaustive list of records that are not included within the definition.
- **FERPA:** The Family Educational Rights and Privacy Act (FERPA) is a federal law that affords a postsecondary student the right to have access to their education records and the right to have some control over the disclosure of personally identifiable information from the education records. When a student is younger than 18 years old, the rights under FERPA is afforded to the student's parent(s). The FERPA statute is found at 20 U.S.C. § 1232g and the FERPA regulations are found at 34 CFR Part 99.
- **Other information**: Other information is defined as information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the institution or community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty. Other information may include, but is not limited to, gender, race/ethnicity, and citizenship. Other information also includes aid awarded for some financial aid programs, that by their nature, risk disclosure (e.g., Tuition Waiver for Unaccompanied Homeless Youth, Tuition Waiver for Students with Disabilities).

- **Outcome Measures**: Outcome measures refer to the student's educational experiences that are recorded in the student's educational records. For example, student grades, courses completed, scores on standardized assessments, time to degree, participation in extracurricular activities, and disciplinary actions are commonly reported measures of student outcomes.
- **Personally identifiable information (PII):** Personally identifiable information for education records is a FERPA term referring to information that is maintained in education records and includes both direct identifiers (such as a student's name or identification number) and indirect identifiers (such as a student's date of birth),[4] or other information which can be used to distinguish or trace an individual's identity either directly or indirectly through linkages with other information. [5] See Family Educational Rights and Privacy Act Regulations, 34 CFR §99.3, for a complete definition of PII specific to education records and for examples of other data elements that are defined to constitute PII. Additional information is available in the PTAC publication *Frequently Asked Questions - Disclosure Avoidance*.[6]
- **Reasonable person standard:** The FERPA standard for de-identifying data is tied to a "reasonable person standard." This standard assesses whether "a reasonable person in the school community who does not have personal knowledge of the relevant circumstances" could identify individual students based on reasonably available information, including other public information released by an agency, such as a report presenting detailed data in tables with small size cells (34 CFR §99.3 and §99.31(b)(1)). This standard can be used to help determine whether statistical information or records are such that a "reasonable person" (e.g., a rational average individual; a fellow student) in the institution or community should not be able to identify specific students. Coaches, instructors, administrators and other insiders are not considered in measuring this standard, as they are presumed to have insider knowledge that would allow them to identify the student.
- **Summary Data:** Summary data include statistical records and reports aggregated from data on individuals in a way that individuals are not identified and from which neither their identities nor any other characteristic that could uniquely identify an individual is easily ascertainable.
- **Suppression**: Suppression is a disclosure limitation method which involves removing data (e.g., from a cell or a row in a table) to prevent the identification of individuals in small groups or those with unique characteristics. *Complementary suppression*, or secondary suppression, is a disclosure limitation where one or more non-sensitive cells

---

[4] Indirect identifiers include information that can be combined with other information to identify specific individuals, including, for example, a combination of gender, birth date, geographic indicator and other descriptors. Other examples of indirect identifiers include place of birth, race, religion, weight, activities, employment information, medical information, education information, and financial information.

[5] MHEC receives the following PII in unit record research files: student first name, middle name, and last name, generational suffix, personal ID (e.g., social security number), birth date, zip code, campus identifier. All but zip code and birthdate are stripped from files for Research purposes as a means to de-identify the data. The personal ID is scrambled.

[6] Frequently Asked Questions - Disclosure Avoidance

or rows is also suppressed so that the values of the suppressed cells cannot be calculated.

- **Suppression Alternatives**: Other mechanisms to help protect student privacy include: 1) perturbation - numeric values are distorted to mask exact numbers; 2) rounding – small values are rounded to larger values; 3) mathematical statements – using less than notations (e.g., <10) in small cells. All of these may be employed but their use introduces other issues that affect data quality and utility (e.g., subtotals and totals must be distorted as well to ensure actual values cannot be calculated) and preparation (these actions must be done manually, which is time-consuming and prone to error).

Scope and Framework

This statement from a document provided by the national Privacy Technical Assistance Center (PTAC) provides the framework for this policy.[7]

Since the release of any data carries at least some element of risk, it may not [be] possible to entirely eliminate the risk of accidental data disclosure. However, organizations disclosing the data in the form of public aggregate reports are responsible for minimizing any such risk while still meeting the disclosure requirements and providing as much useful and transparent information to the public as possible. Before each planned release of student data, an organization must determine the acceptable level of risk of disclosure. This means that in each specific case, the entity disclosing the data should evaluate the risk of PII disclosure within the context that the data will be used, and choose a safeguard strategy that is the most appropriate for that particular context.

Therefore, this policy is framed around two basic tenets: 1) risk reduction and 2) data utility. Using these two tenets, the Office will, when preparing data for reporting and use, perform a risk assessment to weigh whether suppression is necessary and, if so, the methods of data suppression/suppression alternatives to use. When necessary, the Office staff will consult agency counsel for guidance.

Tenet 1 - Risk reduction:

As stated above, there is inherent risk in publishing any student level data, even in summary or aggregated form, with or without suppression. In the publication of data there is a risk of disclosure. It is MHEC's policy to allow the Office to consider such factors as the nature of the data (data content, variables represented), the nature of the characteristics regarding the data (e.g., enrollment intensity of part- or full-time versus unique financial aid categories such as tuition waivers) and the use of the data in making decisions. In addition, it is MHEC's policy to allow the Office to consider which mechanisms to employ to reduce risk (suppression, perturbation), either alone or in combination with each other.

---

[7] From PTAC Frequently Asked Questions – Disclosure Avoidance (October 2012; updated May 2013). https://studentprivacy.ed.gov/sites/default/files/resource_document/file/FAQs_disclosure_avoidance.pdf

August 2022 Data Suppression Policy for Student Information

Tenet 2 - Data utility:
MHEC has a responsibility to provide useable data to its multiple stakeholders. The Office is also committed to employing practices that mitigate the risk of disclosure of PII.

Small numbers in reporting <u>do</u> hold value for users and can help answer pressing policy questions, so employing a standard rule for all reporting is not feasible, practical or useful. It is important to note that reporting unrounded frequency counts in small cells, such as an exact number of students in a small group, does not by itself constitute a disclosure. However, the smaller the cell size, the greater the likelihood that a reasonable person in the school community who does not have personal knowledge of the relevant circumstances might be able to identify an individual within that cell, and thus the greater the risk of disclosure.

A one-size-fits-all method[8] of suppressing or otherwise manipulating small cells sizes can reduce data privacy risk but has some limitations. For example, a common method of rolling small groups of underrepresented populations such as Hispanic, American Indian/Alaska Native, and Hawaiian/Pacific Islander into one category called "underrepresented minority" may work in protecting the privacy of students but risks muting and masking the differences in representation of those groups of students in postsecondary education. In such circumstances, the Office will consider the tenets of this policy – risk reduction and data utility/representation - and weigh the benefits of data transparency with the risks of disclosure. The Office will also consider data reporting methods (e.g., rates but not counts) such that the data can be used and the risk of disclosure is reduced.

Risk Assessment and Practice
<u>Assessment</u> - The Office's data management and privacy practices center on performing risk assessments in preparation of the release of data in reports, dashboards, and other products. When needed, consultation with agency counsel will occur to help with assessment. Specific suppression and suppression alternatives may be employed and may vary depending on the sensitivity level of the data, the risks of/from disclosure and the data usage and utility.

<u>Practice</u> – The Office will employ varying methods of data management and reporting depending on the request or project. Outlined below are practices to be employed.

- The Office will not suppress *Directory Information*. This means the Office will not suppress small numbers of less than 10 when directory information is presented. Directory Information is considered publicly available and the risk of disclosure is low when publishing the data.

---

[8] The policy released in 2019 aligned more with a one-size-fits-all framework wherein the office practiced suppression and complementary suppression without necessarily weighing disclosure risk and data utility. As the Office implemented the policy, it found that this approach often required cumbersome alterations of the data that made it much less useful to the public, even when the data brought very low risk of disclosure and/or very low risk of meaningful harm from any potential disclosure.

- The Office will not suppress summary data if the aggregation does not pose a high risk of disclosure or if the nature of the disclosed data is not considered highly sensitive (e.g., enrollment intensity, state residency). If the summary data poses a risk of disclosure, suppression of cells greater than 1 and less than 10 will be employed. Complementary suppression or other suppression methods may be considered as well.

- The Office will perform a risk assessment utilizing the "reasonable person standard" when preparing data that contains "other information". *Other information* is defined as information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the institution or community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty. Other information may include, but is not limited to: gender, race/ethnicity, citizenship, and financial aid receipt. The Office will report other information with care and will employ disclosure avoidance practices when the risk of disclosure outweighs the data utility. It is possible that the Office will not take specific disclosure avoidance steps should disclosure risks be considered low.

Sharing Unit-Record Data – The Office will employ the following practices when requests are received for unpublished, unit record data.

- The Office will not provide unit record data (i.e., a deidentified record per student) or publish unit record data without specific agreements in place, such as data sharing agreements, or in response to Maryland Public Information Act (PIA) Freedom of Information Act (FOIA) requests.

- Individuals and organizations to which MHEC's Office discloses information will be directed that its re-disclosure to anyone who is not authorized to receive that information under state and/or federal law is prohibited. Disclosure of data by the Maryland Higher Education Commission is subject to the Family Educational Rights and Privacy Act (34 CFR Part 99.31).

- Additionally, any use of education records by another state agency, its employees, agents, or contractors is subject to and shall be consistent with applicable provisions of the Family Educational Rights and Privacy Act (FERPA) including, but not limited to, FERPA regulations at 34 C.F.R. § 99.32 through 99.35, regarding recordkeeping, re-disclosure, and destruction of education records.

It is the Office's aim to apply consistent methods year to year and to use the same disclosure avoidance strategies for similar types of data releases.

Best Practices in Reporting
The practices described below mirror much of what has been presented in this policy. The best practices provided below are taken in part from the U.S. Department of Education's State

Longitudinal Data System Technical Brief (NCES-2011-603), Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting.[9] The Office may leverage these practices when reporting and sharing data.

1. Release summary data, including aggregate student counts of Directory Information on enrollment, major, and completion, for all groups including those greater than zero and less than 10.
2. Report enrollment, degree, some residence data (e.g., county or state of origin) and completion data without suppression, recognizing it as directory information.
3. When suppressing, use a minimum of 10 students for the reporting subgroup size limitation.
4. Suppress results for reporting subgroups (e.g., defined by gender, race and ethnicity, age) with 1 to 9 students and suppress each of the related reporting subgroups regardless of the number of students in the subgroup (i.e., suppress the other subgroup(s) of the set of subgroups that sum to the overall group).
5. Suppress rates or proportions derived from suppressed counts.[10]
6. Use complementary (secondary) suppression, as needed, to assure that suppressed cells cannot be recalculated through subtraction.
7. When possible, aggregate data to segment or statewide level to minimize the need for suppression.
8. Use only whole numbers (versus decimals) when reporting the percentage of students for categories of outcome measures (e.g., the percentage assessed to need remediation, rates of completion by gender).
9. Eliminate totals or subtotals, when needed, as a means to reduce or remove the need for suppression; this action aids in assuring suppressed cells cannot be recalculated through subtraction.

---

[9] Brief can be found here: https://nces.ed.gov/pubs2011/2011603.pdf

[10] This can include such things as average financial aid award per aid program (since average is $ value of total aid awarded in a given category divided by the # of students awarded aid in that aid category) to ensure someone cannot determine the number of award recipients from that calculation.