Privacy Technical Assistance Center

Webinar

November 7, 2012

Best Practices for Access Controls and Disclosure Avoidance Techniques

Webinar Introductions

- Michael Hawes, Statistical Privacy Advisor, U.S. Department of Education

- Baron Rodriguez, Privacy Technical Assistance Center Director

Webinar Overview

- One hour
- **Please enter your questions in the chat box in the lower right hand corner of your screen.**
- Recorded webinar will be available on the PTAC website in approximately two weeks: https://studentprivacy.ed.gov

Why is ED providing guidance?

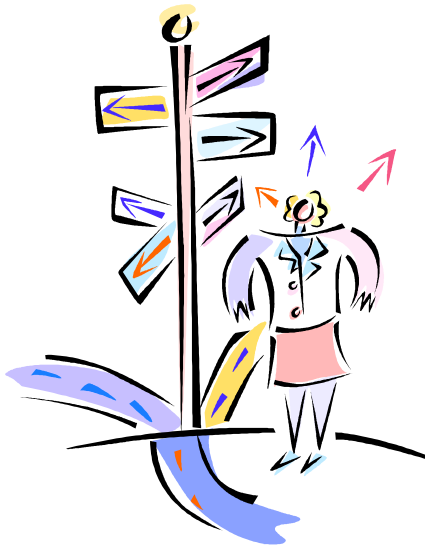
- Privacy of individual student records is protected under FERPA
- Students' data must be adequately protected at all times
- Published reports and/or sharing student data with external researchers increases the likelihood of unauthorized disclosure

What are the guidance documents?

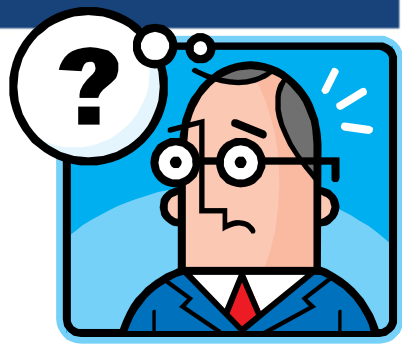
- FERPA Case Study #5 “Minimizing Access to PII: Best Practices for Access Controls and Disclosure Avoidance Techniques”
- “Data De-Identification: An Overview of Basic Terms”
- “Frequently Asked Questions — Disclosure Avoidance”

Basic Terms: Why is ED defining what I already know?

- Many interpretations of common disclosure avoidance terminology
- To clarify the distinction between various disclosure avoidance techniques



Frequently Asked Questions Samples



- *What is the definition of "disclosure" and "disclosure avoidance"?*
- *What legal obligation do educational agencies and institutions have to protect PII in aggregate reports?*
- *Is public reporting of data for small groups ("small cells") the same thing as a disclosure?*

It's all about risk

"The release of any data usually entails at least some element of risk. A decision to eliminate all risk of disclosure would curtail [data] releases drastically, if not completely. Thus, for any proposed release of [data] the acceptability of the level of risk of disclosure must be evaluated."

Federal Committee on Statistical Methodology, "Statistical Working Paper #2"

Frequently Asked Question: Risk

What standard is used to evaluate disclosure risk?

Frequently Asked Question: Risk

Can a reasonable person in the school community who does not have personal knowledge of the relevant circumstances identify an individual in the publicly released data with reasonable certainty?

Paraphrased from 34 CFR § 99.3 and § 99.31(b)(1)

ED Requirements on methodology

Does the U.S. Department of Education require educational agencies and institutions to use specific data disclosure avoidance techniques?

Disclosure Avoidance Techniques

Three Basic Flavors of Disclosure Avoidance:

- Suppression
- Blurring
- Perturbation

Suppression

Definition:	Removing data to prevent the identification of individuals in small cells or with unique characteristics
Examples:	• Cell Suppression • Row Suppression • Sampling
Effect on Data Utility:	• Results in very little data being produced for small populations • Requires suppression of additional, non-sensitive data (e.g., complimentary suppression)
Residual Risk of Disclosure:	• Suppression can be difficult to perform correctly (especially for large multi-dimensional tables) • If additional data is available elsewhere, the suppressed data may be re-calculated.

Blurring

Definition:	Reducing the precision of data that is presented to reduce the certainty of identification
Examples:	• Aggregation • Percents • Ranges • Top/Bottom-Coding • Rounding
Effect on Data Utility:	• Users cannot make inferences about small changes in the data • Reduces the ability to perform time-series or cross-case analysis
Residual Risk of Disclosure:	• Generally low risk, but if row/column totals are published (or available elsewhere) then it may be possible to calculate the actual values of sensitive cells

Perturbation

Definition:	Making small changes to the data to prevent identification of individuals from unique or rare characteristics
Examples:	• Data Swapping • Noise • Synthetic Data
Effect on Data Utility:	• Can minimize loss of utility compared to other methods • Seen as inappropriate for program data because it reduces the transparency and credibility of the data, which can have enforcement and regulatory implications
Residual Risk of Disclosure:	• If someone has access to some (e.g., a single state's) original data, they may be able to reverse-engineer the perturbation rules used to alter the rest of the data

State Methodologies

Does the U.S. Department of Education intend to release more in-depth guidance on data disclosure avoidance techniques?

State Education Agency Guidance

- Select appropriate disclosure avoidance methods
  - Context/Characteristics of data
  - Availability of related data

Best Practices for Access Controls and Disclosure Avoidance Techniques

- Implement in accordance with your State's standards
  - Periodically review your public reports and data tables to ensure that the disclosure avoidance methods you use are implemented correctly
  - Consider checking previously released reports and tables to make sure you applied disclosure avoidance techniques effectively;

Best Practice: Complementary Suppression

- Most states use some form of suppression
- Often additional suppression of non-sensitive data is necessary to ensure protection of PII
- Failure to apply complementary suppression may result in inadvertent disclosure of PII

Example: Complementary Suppression

Subgroup	Number of Students	Percent Proficient
American Indian	***	***
Asian	15	87.7%
Black	12	91.7%
Hispanic	21	81.0%
Two or More Races	13	76.9%
White	24	79.2%
Female	45	84.4%
Male	41	78.0%

Example: Complementary Suppression

Subgroup	Number of Students	Percent Proficient
American Indian	*** (1 student)	***
Asian	15	87.7%
Black	12	91.7%
Hispanic	21	81.0%
Two or More Races	13	76.9%
White	24	79.2%
Female	45	84.4%
Male	41	78.0%

15 + 12 + 21 + 13 + 24 = 85
45 + 41 = 86
86 – 85 = 1

Example: Complementary Suppression

Subgroup	Number of Students	Percent Proficient
American Indian	*** (1 student)	***
Asian	15 (13 proficient)	87.7%
Black	12 (11 proficient)	91.7%
Hispanic	21 (17 proficient)	81.0%
Two or More Races	13 (10 proficient)	76.9%
White	24 (19 proficient)	79.2%
Female	45 (38 proficient)	84.4%
Male	41 (32 proficient)	78.0%

87.7% = 13/15
91.7% = 11/12
etc.

Example: Complementary Suppression

Subgroup	Number of Students	Percent Proficient
American Indian	*** (1 student) (0 proficient)	0.0%
Asian	15 (13 proficient)	87.7%
Black	12 (11 proficient)	91.7%
Hispanic	21 (17 proficient)	81.0%
Two or More Races	13 (10 proficient)	76.9%
White	24 (19 proficient)	79.2%
Female	45 (38 proficient)	84.4%
Male	41 (32 proficient)	78.0%

13 + 11 + 17 + 10 + 19 = 70
38 + 32 = 70
70 – 70 = 0

Example: Complementary Suppression

Subgroup	Number of Students	Percent Proficient
American Indian	***	***
Asian	15	87.7%
Black	***	***
Hispanic	21	81.0%
Two or More Races	13	76.9%
White	24	79.2%
Female	45	84.4%
Male	41	78.0%

By suppressing an additional subgroup (in this case, Black students, who are the next smallest subgroup) you prevent re-identification of the protected values for the American Indian subgroup.

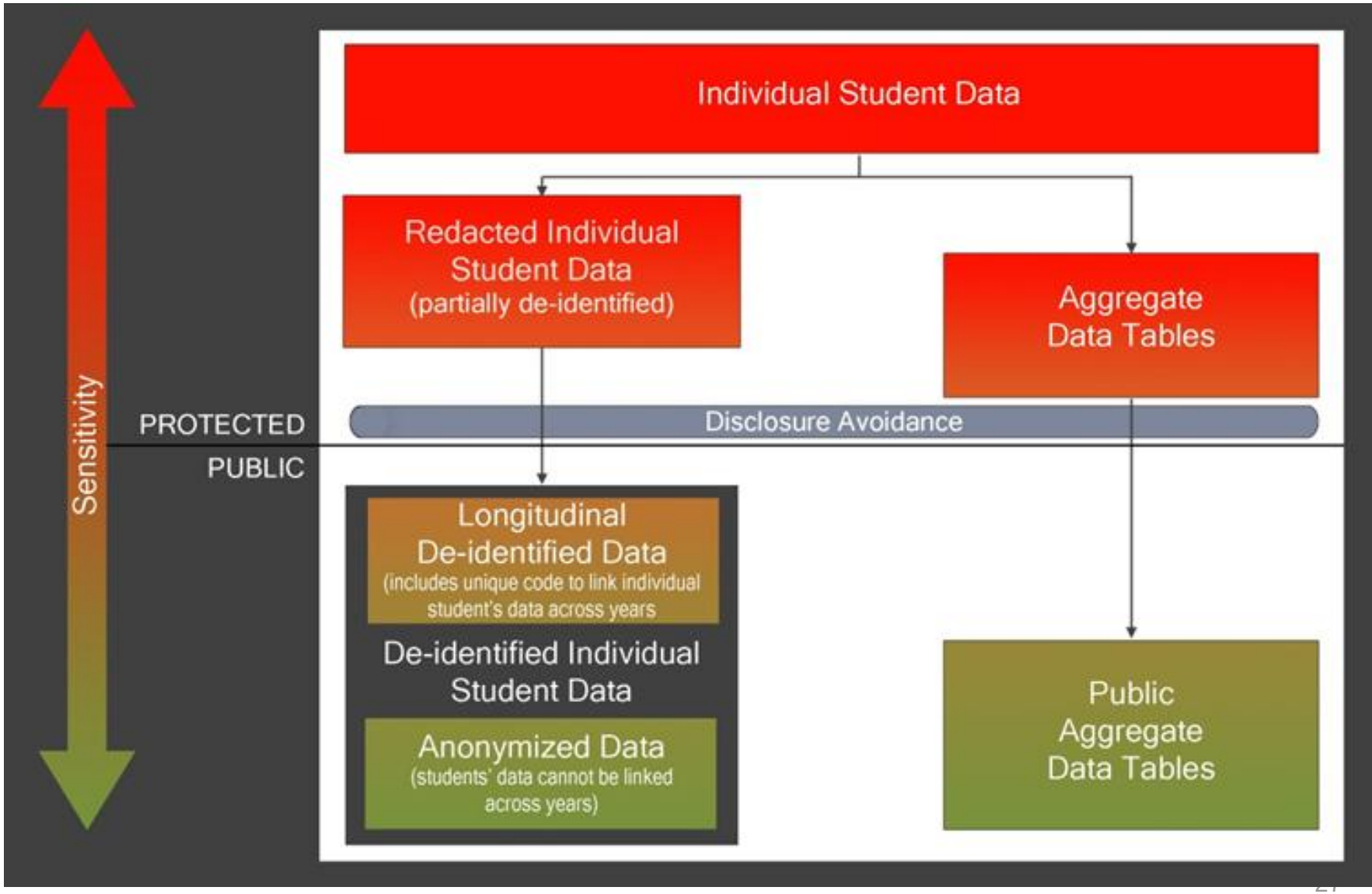
De-Identification

Can't I just de-identify the data by removing the names, SSNs, etc.?

Internal Access Controls

Isn't it just easier from a systems standpoint to provide PII to all my IT staff and researchers? They have a "legitimate" need, right?

Best Practices for Access Controls and Disclosure Avoidance Techniques



Still need help?

- **PTAC Helpdesk** – Review specific questions, reports, or issues jointly with ED officials.

  Write: PrivacyTA@ed.gov or call: **(855) 249-3072**

- **PTAC Site Visits** – PTAC can provide onsite assistance through training (such as FERPA for colleges & universities at yearly registrars meeting), facilitated discussions and can bring experts in the areas of privacy, governance, security, data disclosure avoidance techniques or SLDS expertise.

Questions?

Please enter your questions in the Chat Box in the lower right hand corner of the webinar!