PA.com Approach to FERPA and Student Privacy in Research Data Sets

For more information, please see the following documents:

US Department of Education (2015) "Responsibilities of Third-Party Service Providers under FERPA" , Privacy Technical Assistance Center, August 2015, https://studentprivacy.ed.gov/resources/responsibilities-third-party-service-providers-under-ferpa

Overview

At HRC, our job is to make high quality software that protects the privacy of student information stored in PA.com from unauthorized disclosure. There are extensive security tools and methods in place to protect student data from those who wish to break in, and steal or spoil student data. These are described in the document, ***HRC Data Protection and Data Security Policy 2021-06-13***, and other information that can be supplied to your institutional Information Systems office.

The purpose of this document is to help you understand--as an instructor or researcher using PA.com--how we prevent the unauthorized disclosure of your students' data related to FERPA regulations.

> ***What is FERPA?*** *FERPA is a Federal law that protects personally identifiable information* (PII) *in students' education records from unauthorized disclosure. It affords parents the right to access their child's education records, the right to seek to have the records amended, and the right to have some control over the disclosure of personally identifiable information from the education records. When a student turns 18 or enters a postsecondary institution at any age, the rights under FERPA transfer from the parents to the student ("eligible student").*

We will start with PII data, as its protection is central to the FERPA regulations. Also, because HRC is committed to facilitating research in the field of collaborative learning and communications, we have extended this discussion to cover the provision of research data sets.

Personnally Identifiable Information (PII)

PII is a key concept to understand when discussing privacy. We want to protect the sensitive information about an individual, such as their grades, date of birth, or criminal record--in PA.com, that is the peer assessment information. But the best way to protect it is through the PII. The PII is the linking data that ties this specific assessment to that specific student. If a third party does now know who the assessment is about, the student's privacy is not violated.

Removing or obscuring the PII data in storage and/or research data sets allows us to conduct important research using the sensitive data while at the same time protecting individual privacy.

> ***What is PII?*** *FERPA defines the term personally identifiable information (PII) to include direct identifiers (such as a student's or other family member's name) and indirect identifiers (such as a*

student's date of birth, place of birth, or mother's maiden name). Indirect identifiers, metadata about students' interaction with an app or service, and even aggregate information can be considered PII under FERPA if a reasonable person in the school community could identify individual students based on the indirect identifiers together with other reasonably available information, including other public information.

We take two main steps to limit PII.

1. **Collect as little PII as possible.**

   The only student PII collected for everyone is Name and Email, while the institutional student ID could be included by an instructor if they needed it to match grade, etc. But these are the only information fields that could directly identify one individual student. Other fields could be used together to indirectly narrow down and potentially identify a student. These include the institution, course, term, instructor, and team. This can get close to an individual if knowledge about teams exists; this is generally only understood by the instructor and students on the team.

2. **Data Retention Period**

   The second step is to erase even these few pieces of PII data when the "data retention period" ends. Every institution has a data retention period set in PA.com--the default is 36 months. This may be determined by state or institutional standards for the period of time that student data is required to remain on file. However, to err on the side of privacy, it could be set to as little as 1 month, or executed for a given class upon request.

   The period starts for a student when they finish a class assessed in PA.com; if they take subsequent classes, the retention period re-starts when the last class ends. When the period ends, the student's name, email and student id (if it is entered) are erased and the email is "hashed" into a code** that can be stored with the UID (an internally generated User ID--a simple number that has no meaning outside the system). The hash code and UID are saved so that, if a student takes another class later on, their historical data can be retrieved and full "record" with classes, assessments, teams, responses, etc. is tied back together.

   *** Hash codes are a bit technical but the principle is not so bad. Basically, we pass the email address through an algorithm that calculates the unique hash code for that email. But we use a special "one-way hash algorithm" from cryptography. "one-way" ,means it is easy to compute the hash code from the email, but very computationally difficult to go backwards and figure out the email address if only a hash code is known. Hence when PA.com loads new student emails it can quickly determine if the students are in the system and, if so, link them back up. But if anyone were to steal this data, they could not figure out who were the actual students from the hash codes.*

Policies for Providing Data Sets for Research Purposes

HRC is committed to facilitating research in the field of collaborative learning and communications. For this reason we have put in place data protection procedures to securely facilitate the work of researchers while protecting student privacy. In no instance do we release student PII information in any research data sets.

There are two general types of research that are relevant, and which involve slightly different procedures:

1. **Academic Accreditation**

   Accreditation efforts are one type of research application that may make use of PA.com data.* These involve one institution or one academic college/department/program which is using internally generated data to demonstrate fulfillment of its stated learning outcomes. As such this, analysis data files would be provided by PA.com to an officer of the academic unit under review, who is subject to the standards and norms for managing and reporting this data. The data set would include only data relevant to that unit would be included in the data files.

   Accreditation analyses may need to segment groups of data based on courses, sections, terms, or instructors within the institution. Thus, descriptive (meaningful) data values would be included in these fields to support this analysis. No student PII data would be needed, nor would team descriptors, reducing or eliminating the possibility of connecting assessment responses to individual students.

   ** Learning outcomes related to teams--usually task work, interpersonal, leadership, and communication skills--are commonly required in professional programs, such as business, health care and engineering, but these can be found in any discipline.*

2. **Scholarship of Teaching and Learning (SOTL)**

   SOTL research efforts may be carried out by any researcher. Usually these are researchers in a Higher Ed institution, subject to the Human Subjects protocols and FERPA data use agreements of that institution. PA.com strongly requests that SOTL researchers respect and comply with these standards. However, because such standards may be enforced differently by different institutions, PA.com goes byond those standards.

   Our policy for SOTL data sets is to only include internally generated numeric identifiers for institutions, courses, sections, terms, instructors and students. In other words, no descriptive, meaningful values would be included that could be used to narrow down or specific identify one individual human. We will include the internally generated identifiers which enable the researcher to link anonymous individual student assessment data across teams, assessment events, and courses.