De-Identification of Student IDs

Brandt Redd, CIO, 30 Sept 2015

Each year members submit to Smarter Balanced their student test results and student responses to items. The minimum requirement is for de-identified data; members may elect to send identified data. Smarter Balanced uses these data to maintain and improve the assessments including calibrating new items and producing technical reports.

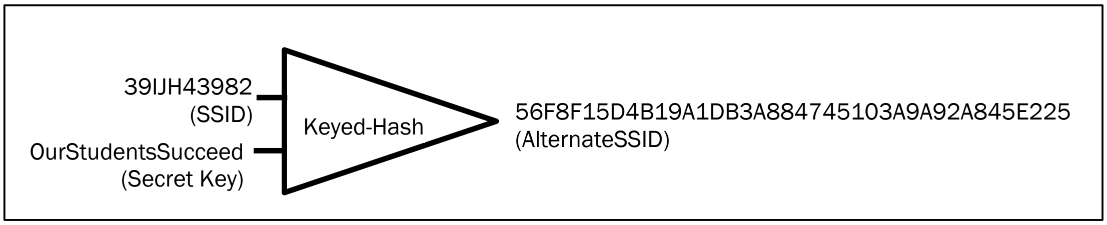
When submitting de-identified data, members should include in the data an Alternate State Student ID (AlternateSSID) that uniquely identifies each student and is consistent from year to year. This alternate ID will be used in future years to compare student performance with previous years' data and to measure growth.

To preserve student privacy, it should not be possible for anyone to derive the State Student ID (SSID) from the AlternateSSID. Smarter Balanced has recommended the use of the HMAC-SHA1 cryptographic hashing algorithm for this purpose. The technical recommendation and sample source code are here: http://www.smarterapp.org/deployment/HashStudentIdSample.html.

While other methods are acceptable, many service providers have chosen to follow the Smarter Balanced recommendation. This memo is to help members understand what to expect from that practice.

Keyed Cryptographic Hash

A keyed cryptographic hash function accepts a string of data (the State Student ID) and a secret key. From these inputs it produces a hash code typically in the form of an alphanumeric string. This can be used for the AlternateSSID in de-identified data sets.



The same SSID and secret key will always generate the same AlternateSSID. However, it is computationally impractical to reverse the hash and gain the SSID back, even if the secret key is known. Thus, an entity would have to have both the secret key *and* the entire roster of all unencrypted student IDs to be able to match a de-identified record to a student ID.

What Members Should Expect

Even when the summative student data supplied to Smarter Balanced are de-identified, Smarter Balanced applies the same security procedures that would be required for identified data and complies with all data security requirements of member MOUs and data sharing agreements.

In most cases, service providers will produce student test results and send those to the member agency. The member then forwards the data on to Smarter Balanced. Members should verify that the data meet their specifications including whether de-identified data are devoid of student identity information such as name, birthdate, and SSID. If the service provider is using a secure hash, then the AlternateSSID will be a 40-character string of letters and numbers similar to the one in the figure above. Depending on the data format, the AlternateSSID may also appear in the SSID field.

If a secure hash is used, the service provider should also provide each member with the secret key they used when preparing that member's data. The key might be a password or passphrase like the one above, or it might be a long string of arbitrary letters and numbers. *The secret key should not be shared with Smarter Balanced.* Members and service providers should treat the secret key with the same sensitivity as identified student data. In future years, they should use the same secret key for the same member. Doing so will allow Smarter Balanced to match student records across years while continuing to protect student privacy.