Safe Aggregation of Personally Identifiable Information Data in Generally Available Aggregated Reports and Cubes

Purpose

The protection of individuals’ personally identifiable information (PII) in University of Washington (UW) information systems is regulated by federal and state privacy laws; yet, in the interest of transparency, the UW must publish aggregate reports to inform the public about the students it serves and the individuals it employs to fulfill its mission. While the aggregation of individual-level data into department- or university-level data significantly reduces the risk of disclosing PII, an accidental disclosure can nevertheless occur when aggregate reports are presented in a way that allows individuals to be identified, either directly or in combination with other available information. To comply with privacy laws, UW designated officials who aggregate or publish data at the UW must take steps to minimize the risk of disclosing PII.

This guideline summarizes the Best Practices for Aggregate Reporting of Individual Data at the UW.

Note that these guidelines are not all inclusive and do not replace the need for data users to be familiar with any laws and regulations that relate to PII, as well as any department-specific policies.

Overview / Problem Statement

Many reports and cubes published by UW officials first group individuals into various categories, (e.g. by college, by gender, by ethnicity) and then show aggregated counts or averages for those groups. While these reports do not display individual attributes (e.g. name, address), it may be possible to deduce PII in relatively small group sizes. For example, if a report of graduation rates by race and ethnicity reveals that there is one African American in the cohort and that the graduation rate for African Americans is 0%, it would result in disclosure: anyone who knows this student could deduce his or her graduation status from the report.

Observations, Recommendations and Agreements

The following is a summary of the major observations and best-practice recommendations. Data users who publish aggregate reports are required to follow these best practices. If there are questions regarding the interpretation or application of these best practices, UW designated officials should request clarification from appropriate Data Custodians and/or the Data Management Committee (DMC).

- The risk and strategies for avoiding the accidental disclosure of PII should be evaluated on a case-by-case basis.
- Publishing aggregate reports refers to making the reports generally available on the UW Tableau Public server with broad user access, or publishing physical copies. The UW Tableau Public web site does not have a “login” or university-controlled authentication.

Document Location: https://www.washington.edu/provost/oim/dmc
Questions regarding this document should be emailed to: dmc-support@u.washington.edu
• The risk of disclosing PII in aggregate data should be assessed at the lowest available level of aggregation, taking into account the sensitivity of the data, in concert with the relevant Data Custodian.
• Generally, the number of individuals in the smallest group at the lowest available level of aggregation should be at least 10.
• In the presence of cell sizes that are smaller than 10, the following strategies should be employed to reduce the risk of disclosure of PII: 1) suppression or removal of information about small cell sizes; or 2) reduction of the precision of the data by rounding or providing ranges. The choice of strategy will be determined by the sensitivity of the data and the purpose of the aggregate report.
• The data publisher should clearly indicate and document which techniques are being applied to reduce the risk of disclosure of PII.
• Only aggregated data should be stored or extracted in cubes or data sets used by reports. Data at the lowest level of detail should only be accessible through a drill-through action executed against source database using live connection. This will ensure that proper Data Access Control (DAC) permissions are leveraged to access data.
• In the case of any accidental disclosure of PII, the department which built and published the Tableau visualization and corresponding data sets, is responsible for: 1) notifying the Office of the Chief Information Security Officer (CISO) according to Administrative Policy Statement 2.5 Information Security and Privacy Incident Management Policy; and 2) paying all expenses related to the accidental disclosure.

Advisory Opinions Sought

These best practices and guidelines were developed with input from the Institutional Analysis department in Planning & Management, the Graduate School, Enrollment Management, CISO, Human Resrouces, Academic Human Resources and UW Information Technology.

Concurrence / Responsibility

The undersigned concur with observations, recommendations and Agreement in this data usage guideline document. They understand the implied risks and endorse summary of the best practices into this guideline.

Signature
Vice Provost, Planning and Budgeting, Gary Quarfoth
(Co-Chair of Data Management Committee)

Date
6/9/15

Signature
Associate Dean, Arts and Sciences, Steve Majeski
(Co-Chair of Data Management Committee)

Date
6/15/15

Document Location: https://www.washington.edu/provost/oim/dmc
Questions regarding this document should be emailed to: dmc-support@u.washington.edu