

Data Reporting Guidelines
Office of Institutional Research: Princeton University

Purpose

The Data Reporting Guidelines are intended to:

1. Ensure that University information is used in a manner that is consistent with state and federal regulations and with University policies on data security and privacy.
2. Reduce the risk of unintentional, accidental and/or secondary release of University information.
3. Reduce the risk of unintended, direct or indirect, disclosure of personally identifiable information.
4. Encourage the consistent presentation and use of University information across administrative and academic units.
5. Facilitate the appropriate use of University information derived from multiple sources.

General Principles

1. Data products should be designed in a manner that considers intended audience and the associated risk of unintentional or accidental and/or secondary release of the data product.
2. Data products should indicate the sensitivity of the information being delivered/released. At a minimum the data product should indicate the Data Classification of the data product.
3. Data products should indicate any restrictions regarding the redistribution, repurposing, reuse, and/or reformatting of the data product.
4. Data products should indicate the source of information and the date of the production.
5. Data products should include sufficient explanatory content to enable the “user”/“reader” to understand and interpret the data product in the absence of supporting materials. This may include definitions of field values, column headings, row headings, abbreviations, and calculated variables.
6. Data products should be delivered/released in a format that maintains the integrity of the data product and appropriately manages the end users ability to manipulate, copy, clone, and/or replicate the data product.
7. Direct personal identifiers should not be included in data products unless they are required to meet a valid business need.
8. Data products based upon “person-data” (headcounts, FTEs, etc.) should be presented in the aggregate. The smallest cell/record value should be based upon at least five (5) individuals—unless there is a valid business need that requires greater detail.
9. Data products that provide unit-to-unit comparisons should be approved by the appropriate authority prior to release.
10. Data products, as appropriate, should include a cover-sheet that summarizes the acceptable uses of the data product.

Suggested Data Classification of data products by intended audience

1. Public

- a. Information intended to be released outside of the Princeton University
- b. Information published or posted on open web sites
- c. Campus publications (Princeton profile)
- d. Information presented at open meetings (CPUC)
- e. Information presented at meeting accessible to media

2. Public within Princeton University

- a. Reports available to campus community that require authentication to access
- b. Information published or posted on open web sites that require authentication to access
- c. Information presented at open meetings where the audience is limited to members of the Princeton community (AAMG)

3. Confidential

- a. Task forces / Ad hoc committees
- b. Department meetings
- c. Department officials (Chairs, DGS, Director of Studies)
- d. Trustees
- e. Cabinet officers

4. Restricted

- a. Defined audience with specific business need that necessitates access to Restricted Data
- b. Audience members sign data-use agreement tailored to meet specific business need.

Suggested reporting formats

1. Suggested strategies to mitigate the risk of disclosure of personally identifiable information
 - a. Suppress results for all reporting groups comprised of 0 to 4 individuals
 - b. Report whole-number percentages without underlying counts or totals.
 - c. Create and report group variables that contain at least five individuals (salaries ranges)
 - d. Collapse existing categories into broader group variables (Black, Hispanic, Native American, Native Hawaiian into underrepresented minorities)
 - e. Create and report “top-coded” and/or “bottom coded” variables (>90%, <10%)
 - f. Suppress sub-groups and report at the group level (report at divisional level instead of the departmental level)