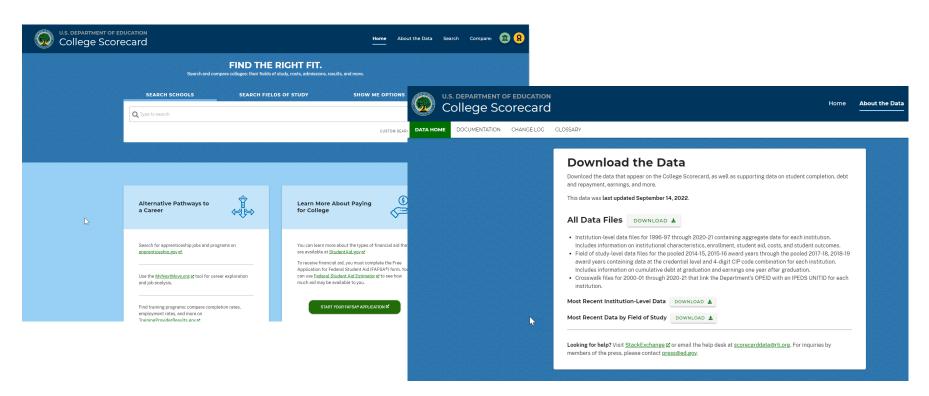


# Statistics of Income: College Scorecard Project

The College Scorecard is an online tool owned by the Department of Education (ED) and available to the public to search and compare colleges by field of study, costs, admissions, economic outcomes, and more.





# **IRS** Internal Revenue Code Section 6108(b)

Statistics of Income (SOI) is part of the Research, Applied Analytics and Statistics arm of the IRS

SOI is authorized under the Internal Revenue Code (IRC) Section 6108(b) to release aggregated level statistics to the Department of Education. The section authorizes the Secretary of Treasury to make special statistical studies and compilations involving return information (as defined in section IRC 6103(b)(2)).



### **College Scorecard Earnings Data**

SOI matches data on recipients of federal student aid (provided by ED) to IRS administrative tax records from two databases that hold W2 and 1040-SE data.

SOI has produced the earnings metrics portion of the College Scorecard for six years. The last three years we've used the SafeTables software to produce statistics that satisfy differential privacy.



# College Scorecard Earnings Data

#### Original disclosure method: Data suppression and rounding

- Data were aggregated at the institution or institution and program level.
- Earnings were capped at a max dollar amount and rounded.
- Suppressed rows and cells based on predetermined value.

#### Current disclosure method: Differential Privacy and suppression

- Tumult Labs, Inc. SafeTables software application executes differentially private computations of the statistics to be released.
- Initial suppressions are based on a minimum threshold, but additional suppressions are made based on relative error.



### **College Scorecard Workflow**

- SafeTables application specifications are updated to reflect potential changes in inputs and outputs
- SOI provides Tumult with synthetic microdata (no PII) to determine the privacy algorithm that will satisfy differential privacy.
- Tumult designs the evaluation tool (CSExplore) that allows SOI to adjust the level of privacy loss utilized.
- A limited version is shared with ED, which allows them to explore alternative suppression thresholds and workload specifications.



## **College Scorecard Workflow**

- Once SOI and ED have settled on final details of the deliverables, the microdata (no PII) are submitted to the SafeTables application to produce aggregated statistics.
- SOI applies suppressions based on, noisy and true, count threshold.
- SOI assess SafeTables output by comparing with confidential data. Additional suppressions are applied based on relative error threshold.
- In addition to delivering aggregated earnings data, summary statistics are provided to ED.