Chair Young, Ranking Member Miller, and all honorable members of the House Higher Education Committee, thank you for hearing my testimony today. In the process of earning my Ph.D. in Cognition, Brain, and Behavior from the University of Notre Dame, as well as in writing dozens of scholarly peer reviewed publications, I have worked extensively with data and the processes to convert, rearrange, or otherwise shape data into forms necessary to execute tasks. I was asked to imagine a few scenarios in which I would be provided data inputs of university and major and would need to match those inputs to a database that included that information and the 25th percentile and 75th percentile incomes at 1 and 5 years out. I was asked to consider a university provided database and to consider publicly available census data. After matching, I would return the income values in a form that could be broken up into individual letters for students. I spent a few hours generating examples and writing code that would deliver the request.

I believe this task to be relatively straightforward, affordable, simple, and secure, and I can demonstrate how one could accomplish this task using a few common tools such as R, a free open source software with the tidyverse package, and Microsoft word with a mail merge function. I use these tools for convenience but believe other software could be used to accomplish similar tasks, such as Microsoft excel.

First, please see Supplemental1, which is an example of publicly available census data.

Next, please see Supplemental2, which is an example of student outcome data that a university might provide.

Either Supplemental1 or Supplemental2 could serve as the source of information for this form.

Please see Supplemental3 for an example of a few imaginary students who have been accepted and would require this information in the acceptance letter.

Please see Supplemental4, for how a university could send out this information replacing first and last names with numbers so that any matching would be ‘blinded’, or lacking in personal or identifiable information.

Please see Supplemental5, for what I could use R, or other programs like excel, to use to match the data and return relevant information.

Please see Supplemental6 for what the university could do after it had the file it sent and the file it received back.

Please see Supplemental7 for an example of a mail merge template that is combined with 6.

Please see Supplemental8 for the outcome of the mail merge template: individual PDFs that have associated a specific person's name with their major and the expected income 1 and 5 years out of the 25th and 75th percentile.

Based on these examples, I believe a few hours of work with an appropriately skilled individual could satisfactorily complete them. If I were to be asked to do this task in actuality, I would probably ask for a $1,000 initially to set it up, $200 an hour for any consultations required or calculations performed (e.g. Training in the use of mail merge, if different logic was required, screening data for misspellings, calculating averages rather than returning matched values, etc.), and charge $1 a completed record match. I imagine the first year would cost a university about $20,000 and each subsequent year $10,000. Thank you for the opportunity to provide testimony here today, and I am happy to answer any questions the committee may have.

Supplemental files:

1) Example census data
    a) https://lehd.ces.census.gov/data/pseo_experimental.html

2) Example Database
    a) Generated by me

3) Example University Acceptance and Major List
    a) Generated by me

4) Example University blinded Major List
    a) Generated by me

5) Example Matching Output
    a) Generated by me

6) Example University combination file
    a) Generated by me

7) Example Mail Merge word doc
    a) Generated by me

8) Example Individualized PDFs
    a) Generated by me