Background:

The Center for Education Policy Research at Harvard is seeking a tool for scraping school board websites. Starting in early February, our research team is planning to send information to more than 15,000 school board members in 40+ states. We will vary the content of the messages and then seek to measure whether which information is subsequently mentioned in the minutes of local school board meetings. (We will also invite school board members to sign up for later updates.) Eventually, we are looking for a tool which would allow us to download all documents from a school board's website mentioning a specific key word combination. *Over time, we will use the results to optimize our success in communicating research findings to local school district leadership.*

Many school districts in the U.S. use one of two products to post their meeting minutes online: BoardDocs and simbli by eboardsolutions. Unfortunately, we have been told by Board Docs that there is no single database which contains all the board documents. (With the appropriate sign-on, school board members themselves <u>do have</u> the ability to search the documents in those school districts which have not yet transitioned to the new version of the Board Docs software. We can connect the student with a school board member with the appropriate credentials.) However, because they are public entities, each district has a website which their residents can use to search. Fortunately, it seems that the school boards using Board Docs use the *same http address format*:

e.g.

https://go.boarddocs.com/mabe/pgcps/Board.nsf/Public# for Prince Georges County in Maryland

https://go.boarddocs.com/oh/jtmo/Board.nsf/Public   for Johnstown County in Ohio

https://go.boarddocs.com/fla/orcpsfl/Board.nsf/Public   for Orange County Florida

Deliverable #1: Find all the websites using the https://go.boarddocs.com/ stem (even if they are not school boards), capture the name and address of the organization (typically in the top left corner of the webpage as in the above).

Deliverable #2: Merge that list of sites to the National Center for Education Statistics' Common Core of Data, using the name and addresses of the organizations. The CCD provides an id for each local education agency which we can use to match. (Many organizations other than school boards use Board Docs.)

Deliverable #3: Repeat #1 and #2 for simbli by eboardsolutions.

Deliverable #4:   Develop a tool which would take a key word, such as "Education Recovery Scorecard," visit each of the sites, search the "meetings" tab within the Board Docs site and download all documents containing the key words.

Our research team will then be able to go through the documents from Deliverable #4 to determine which were true and false positives.

Timing: We would very much like to have **Deliverables 1 and 2** by mid January 2025.   (Deliverables 3 and 4 are less time sensitive.)   We plan use that list of school district names to form a sampling frame for our messaging in February.

Pay: $23 per hour

Supervision: The student will meet weekly with Professor Tom Kane and CEPR research analyst, Dan Dewey to problem-solve and report progress.

Contact: tom_kane@gse.harvard.edu AND Ashley_snowdon@gse.harvard.edu.