Testimony of Derek Willis, Lecturer, University of Maryland, Before the House Appropriations Committee Subcommittee on the Legislative Branch for Fiscal Year 2023

Chairman Ryan, Ranking Member Herrera Beutler, and Members of the Committee, thank you for the opportunity to submit testimony. My name is Derek Willis and I am a lecturer in data and computational journalism at the Philip Merrill College of Journalism at the University of Maryland. I have been a professional journalist at a number of news organizations since 1995, specializing in legislative and electoral data analysis, and I am here today speaking on my own behalf. Today I recommend that the Committee direct the Office of the Chief Administrative Officer be authorized to publish unique identifiers in the electronic version of the "Statement of Disbursements of the House" reports published four times a year. Doing so requires little additional effort and would result in more accurate and useful information for the public.

I applaud the House for its steps to improve the availability and utility of the disbursement reports since first publishing them online in 2009. In particular, since January 1, 2016, the reports have been published as structured data in a spreadsheet, which makes it possible for members of the public, including journalists and researchers, to search and analyze the spending of the House and its offices. The publication of this information online as structured data was in part an accountability measure undertaken by the House in the wake of a scandal. Publishing expenditure information online as data is crucial for providing both context and accountability to the operations of the House.

When the House began publishing disbursement reports in PDF files in 2009, I wrote software to transform those files into data. Because PDF files are mainly used to display information, not to make it usable, that was a difficult and time-consuming process that made it harder to generate timely and accurate information. Still, it was a step in the right direction that, for the first time, made it possible to analyze the spending of House offices in a comprehensive manner.

The publication of structured data significantly reduced the barriers of access to this information and made the process of analyzing it much easier: now users can download the data and use spreadsheet or database software to ask and answer questions.

The Committee should make that structured data more accurate and easier to use by publishing a unique identifier for each entity that appears in the data. For example, each

committee has a unique code that refers only to that committee, much like a drivers license ID. In fact, every entity that does business with the House also has a unique identifier. The CAO already has those unique identifiers in its database and has indicated its readiness to publish that information. However, the CAO needs the Committee's authorization to do so, and that is precisely what I am requesting of you.

Publishing these unique identifiers would not change the nature of the information available to the public; the House already publishes summaries of expenditure information online for public use. The reason to add them is that doing so would make the analysis of this data more accurate, reliable, and easier to use. What this means is that it will be easier for journalists to be accurate when they explain where congressional funds go; for watchdogs to accurately account for every penny; and for academics to properly study the operations of Congress.

Specifically, I am requesting that the Committee direct that:

> To the extent feasible, the Office of the Chief Administrative Officer is requested to publish a unique identifier for each organization and entity listed in the CSV version of the Statement of Disbursements of the House starting with the forthcoming publication of the Statement of Disbursements. Further, the CAO is requested to update prior Statement of Disbursements published as CSVs with this data field to the extent practicable.

This step would improve the accuracy of the data and increase the public's confidence in and understanding of how the House spends the public's money.

Thank you for the opportunity to submit this testimony.