

Revision Memo for, “Government Websites As Data: A methodological pipeline with application to the websites of municipalities in the United States”

June 4, 2021

Dear Profs. Copeland, Gainous, and Towner,

Thank you for your comments and for giving us the opportunity to revise and resubmit our paper. Thanks also to the reviewers for their constructive comments. As we hope you will see both in the revised manuscript and in this response letter, we have taken each comment seriously and have endeavored to address each as completely as possible. Each of the comments from reviewers is addressed in the order presented in the review. We should note that we are now approximately 300 words over the 5,000 word limit. We trimmed the text substantially as we revised the paper, but we think the current version provides the most effective presentation of our core contributions and the revisions suggested by the reviewers. Of course, we are happy to trim further if necessary. We do believe that addressing reviewers' comments has significantly improved the paper.

Reviewer 1

1. *My main recommendation is to add some figures/plots that show the substantive findings about the differences between Democratic and Republican mayors. You do a good job describing the findings in the text, but I think figures would also help make the findings clear to readers.*

Response: We thank the reviewer for this recommendation. We recognize that the STM results presented in Table 2 conveyed the content, direction of partisan bias, and statistical significance, but did not convey magnitude of differences/effects in any meaningful way. We have added Figure 3, in which we illustrate the magnitudes of the differences between the prevalence of topics based on the mayor's partisanship. We also added discussion of the differences illustrated in this figure. We see this as a major enhancement of our results interpretation.

2. *Also, in the conclusion can you give some more examples of other research areas and questions where this approach will be useful? I think that would increase the usage and citations.*

Response: We have added a paragraph to the conclusion in which we recap the uses of website content analysis that we reviewed in the front end of the paper, and also note and cite potential applications outside of the analysis of government websites and candidates. We give examples of politics and policy research in which researchers have analyzed (1) company websites, and (2) the websites of interest groups.

3. *Along similar lines, I hope you plan to produce step-by-step replication code and instructions about how readers can use your workflow. I think that too will increase the usage and citations of your approach.*

Response: We will make three categories of online materials available to help readers/users effectively apply the methods workflow we develop. First, we will post a conventional replication archive to Dataverse. The replication archive will include all of the code necessary to replicate the text analysis results we present, as well as the

raw text corpora that we extracted from the websites. Second, the `gov2text` package is available on GitHub. Third, we have posted an interactive tutorial on Google Colab that demonstrates the key steps in the application of our pipeline using the `gov2text` package.

4. *One minor point – there’s a typo in the abstract – “can can”*

Response: Fixed.

Reviewer 2

1. *Research question objective: From the title, it seems that the main objective of this study is to develop a new methodology that aid in collecting and analyzing data from government websites and examining the association between website content and mayoral partisanship is a way to validate. However, the author(s) did not develop the discussion about the challenges and issues that they found in the existing methodology after indicating the problem statement in the Introduction as “The conventional approach to data collection in projects focused on government websites involves manual content extraction from each website in the dataset (p. 1, line 42-44).” This leaves me confused about the objective and the research question of this study. While the objective is to develop a new methodology, the research question is to examine the mayoral partisanship in the website contents? The author(s) should address this gap by clearly explaining the differences in a set of research objectives throughout the manuscript.*

Response: In the initial draft we did too little to walk the reader through the challenges posed by manual collection of website data, and also did not clearly describe the organization and relationships between our primary and secondary research objectives. We have substantially revised the front end of the paper to further justify and clarify our research objectives. To better situate our contribution in the existing literature and methodology, we discuss an example of the manual data collection process used in a recent study in which manual data collection was used to gather data on the contents of the websites of U.S. municipalities. This example anchors our discussion of the cost and scale limitations associated with manual data collection.

2. *Locating within the existing literature: While the main objective of this study is to develop a new methodology, it lacked discussion of the existing methodologies. Instead, the author(s) discussed the current literature about the partisanship and government website contents in section 2 and then jumped right into the method part in sections 3-5. Without locating the objective of the study in the existing literature, the conclusion is a mere summary of this study and it did not develop into a deeper discussion of theoretical and practical implications of this study.*

Response: We agree with this critique of our literature discussion, and we have completely revised Section 2 to focus primarily on the methods that have been applied to collecting data from government websites, and to specifically focus on the challenges posed by manual data collection. We also incorporated a footnote about the literature on campaign websites, which uses similar methods and thus faces the same issues.

The discussion of mayoral partisanship and government websites has now been shifted down to the point at which we introduce our application. Finally, we revised the conclusion to point specifically to the implications of our contributions. In response to R1 (point 2) we extended the conclusion to discuss more of the literature in which our methodological pipeline would be applicable, which helps to further unpack the implications.

3. *The abstract is not well framed to effectively introduce the goal and the question of this study. The last sentence is particularly confusing to grasp the main contribution of this study.*

Response: We revised the abstract to more effectively present and justify our main objective of introducing a methodological pipeline for extracting content from government website. We also contextualized discussion of the application by noting that it serves to illustrate and validate the methodological contribution.

4. *The last paragraph that explains the structure of this study is clear, which should be used as the reference to organize the overall narratives of the manuscript.*

Response: We unpacked and expanded upon this paragraph in the revised presentation of our research objectives and approach.

5. *While it is essential to provide the background of the association between government website content and mayoral partisanship, it did not well developed. As the findings are related to the types of contents of government websites depending on mayors' partisanship, while it explained the ways government websites are used to reflect policy priorities, it lacked the discussion of the overall political objectives of each party.*

Response: We now recognize that our initial submission jumped right into the estimation and interpretation of differences based on mayoral partisanship, without first discussing why we would expect the priorities of mayors to differ according to party. We added discussion of findings from two recent articles that address how Republican and Democratic mayors differ in terms of policy priorities.

6. *As this paper's contribution is mainly methodological, there should be a separate, in-depth discussion of the problems and limitations in the existing methodologies. This discussion is more important than the reflection of partisanship on government website contents because it explains the need for a new methodology, supporting the contribution of this study.*

Response: Our revised and focused discussion of the scale limitations, and costs, associated with manual web content collection in Section 2 (with the former contents of Section 2 being moved down to Section 3) address this point. See also our responses to R2's comments 1–3.

7. *The conclusion should be improved by adding its theoretical and practical implications by aligning with the limitations of the existing studies that you should discuss earlier*

(see my comment #6.)

Response: We have revised the conclusion in accordance with this comment (as well as earlier comments from R2 and R1).

8. *The author(s) should check typos after revising. Abstract: "... from large samples of government websites can can compliment ... (p. 1, line 22)." Section 2: "... mobile accessibility, and verall function... (p. 3, line 42)"*

Response: We have closely proof read the revised manuscript.