

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

October 1, 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 Reviewer 1 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 is addressed in the order presented in the review. We should note that we are now approximately 500 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. *Abstract: in the last sentence, "The application validates the utility of website content data...", I'd suggest changing the word, validate, to demonstrate/illustrate as validation of a new method requires a much more rigorous approach.*

Response: We agree with this comment, and revised it to be, "illustrates."

2. *In section 2, the author(s) added the discussion of the existing studies that used manual methods. However, it is not clear how the author(s) connect the website evaluation studies to the use of website data. These seem to be two separate branches of study in the e-government website literature. If there is an important connection, the author(s) should explain and make a clearer argument on how the current study can help with the limitations found in the manual website evaluation study. Otherwise, the author(s) should use more relevant examples of the existing studies that are more tightly aligned with the subject of the current study.*

Response: We revised this section to note the connection/relevance. We note that it is now standard practice for researchers to conduct "supervised" or "human-in-the-loop" machine coding exercises in which a relatively small number of human-coded items are used to train a machine classification algorithm such that a large set of items can be machine-coded. Given the ubiquity and effectiveness of human-coding attributes of government websites, there is substantial potential for supervised machine coding. However, in order to train a machine classifier, researchers must have access to website contents in a machine-readable format. Our pipeline can be used to collect such website content from a large set of websites.

3. *Conclusion: In the last paragraph, the author(s) presented two studies by explaining the implications of the study. These sets of existing studies should have been discussed earlier. Also, the author(s) should explain more in detail why these studies show the usefulness of the new methodology that this study provides. Have these studies used*

the manual method? If not, any ideas on how the approaches these studies used can be applied in the context of e-government studies?

Response: We removed the discussion of these studies from the conclusion. We added a brief paragraph to Section 2, in which we discuss the fact that researchers in other areas—beyond the analysis of government websites—have conducted computational content analyses of websites. We point to these studies as evidence that our pipeline will be applicable beyond the study of government websites. The previous paragraph already explains that our methodology would improve the approaches in these studies by making it possible to use the same scraping technique for all websites being studied, rather than having to specify HTML tags for each site individually.