

Using Comments as Data for Research

By: Zhoudan Xie | April 13, 2020

The current COVID-19 pandemic is disrupting many activities in the economy. Thanks to eRulemaking, people can still [track](#) federal regulatory activities and [submit](#) comments on proposed regulations online. For government agencies, public comments are [considered](#) an important source of information during the rulemaking process that can improve the effectiveness of regulations and enhance democratic accountability. For researchers, comments have been a valuable source of data to study public participation and bureaucratic behavior.

A recent GW Regulatory Studies Center [report](#) examines how public comments could inform retrospective review of existing regulations. It analyzes public comments from an unconventional perspective and reveals an opportunity for using comments as data that has not been fully realized in the existing research.

Public comments submitted over the Internet

Historically, public comments were submitted via postal mail and in-person delivery. Although those options are still retained in some cases, most agencies accept comments over the Internet today. To make it easier for the public to participate in the rulemaking process, the [eRulemaking Program](#) launched [Regulations.gov](#) in January 2003. Since then, the website has become a central portal for public users to access federal regulatory materials and submit comments on regulations.

Today, nearly 300 [federal agencies](#) post an [average](#) of 8,000 regulations per year, among which the majority receive comments and share them on Regulations.gov, while others accept and post comments via other online platforms (e.g., [Surface Transportation Board](#)). In 2019, the documents posted to the website [include](#) 1,372 proposed rules, 3,176 rules, 17,151 notices, and 2,205,631 public submissions. The huge number of public submissions not only reflects active public participation in the rulemaking process, but also represents a source of rich data that enables academic and practical research.

In brief...

Public comments have been a valuable source of data in research studying public participation and bureaucratic behavior. Our recent report analyzes public comments from an unconventional perspective and reveals a new way for researchers to use comments as data.

Research using public comments

An extensive body of research has used public comments to study the nature of public participation in the regulatory process and its implications for participatory democracy. The research generally focuses on addressing three questions: who comments, what they say, and how the government responds. Researchers typically answer these questions by qualitatively [coding](#) the content of comments for a select set of regulations. For example, Cuéllar [studied](#) the complexities of participation under existing legal structures by analyzing thousands of comments received on three regulations along multiple dimensions, including commenter identity, the level of sophistication, issues of concern, and recommendations. GW Regulatory Studies Center senior scholar Professor Steve Balla and his coauthors [examined](#) the sponsorship and content of mass comment campaigns (MCCs)¹ based on information collected on more than one thousand MCCs submitted on Environmental Protection Agency (EPA) rulemakings between 2012 and 2016.

Studies analyzing government responsiveness to public input involve comparisons between changes in content from proposed to final rules and preferences expressed in comments. This literature includes [several studies](#) by Susan Yackee and her coauthors evaluating the influence of interest group comments on agency rulemaking. In subsequent analysis on MCCs, Balla et al. [addressed](#) the association between MCCs and the content of final rules, finding that agencies were more likely to respond to substantive comments than to campaigns.

As such, public comments have served as valuable data for researchers to answer a set of interesting questions. However, our recent [report](#) suggests that comments may reveal more information than the prior research has recognized. In this study, we examined a special set of comments submitted to the U.S. Department of Agriculture, EPA, and the Food and Drug Administration in response to recent deregulatory initiatives. Unlike comments received on proposed regulations, these comments were solicited to help agencies identify existing regulations for repeal, replacement, and modification. We found that, although the content of the comments exhibited significant variation, a substantial number of comments [identified](#) specific regulations (such as references to CFR parts or sections) as candidates for review and [provided](#) relevant feedback on regulations (such as [forms of regulation](#)) that could inform the directions of agency regulatory reform.

The report also [shows](#) that these comments can be used as a source of data to identify regulations affecting specific industries. Through text mining of the comments, we identified a set of regulations that are likely to affect the crop production sector and confirmed with empirical analysis that these regulations appear to slow down the productivity growth in the relevant industries. The analysis implies that public comments could provide information that helps answer questions that have not traditionally been addressed in research, such as the economic impact of regulations.

¹ [MCCs](#) “consist of identical and near-duplicate comments sponsored by organizations and submitted by group members and supporters to government agencies in response to proposed rules.”

Comments as text data

The fact that comments are available in an unprecedented volume and format today offers numerous research opportunities in the future. The existing research using comments mostly relied on qualitative coding through human reading, which largely limited the scope of comments analyzed, as opposed to vast quantities of comments available on the Internet. Moreover, advanced text mining and analysis [techniques](#) developed during the recent years have made it possible to convert the information encoded in text into more structured data, which can enable empirical research using much larger data sets of comments.

As an effort to bring new resources and technologies into the existing scholarship, we applied several new techniques in our study of public comments, including retrieving tens of thousands of comments via the [Regulations.gov API](#) and mining the text of comments to extract relevant information. Regulations.gov offers all its public data in machine readable format via an API (Application Programming Interface), which allows users to search and retrieve data on public submissions and other regulatory materials in an automated way. To be more explicit, all the content that can be obtained using the regular search function on Regulations.gov is available in json or xml format if an equivalent [API query](#) is used. This includes the full text of comments and rules, as well as their metadata such as agency name, commenter name, publication date, etc.

Considering that programming is not a conventional specialty of social science scholars (while many scholars are also good programmers), we created a [Github repository](#) to share the Python code we developed in our research. The goal of the repository is to provide code that can be easily modified for use in research using public comments. The initial content available in the repository includes the code to retrieve public submissions via the Regulations.gov API, including comments submitted as PDF attachments, and convert them into text data. As we continue uploading new code for other parts of our analysis, we hope other researchers and programmers can contribute to the repository or suggest improvements to make it a more useful tool.

No description, website, or topics provided.

Edit

Manage topics

28 commits 1 branch 0 packages 0 releases 1 contributor MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

Table with 3 columns: File Name, Action, Time Ago. Rows include Retrieve Comments, LICENSE, and README.md.

README.md

Background

Under a cooperative agreement with the U.S. Department of Agriculture (USDA), the George Washington University Regulatory Studies Center conducted research analyzing public comments to inform agency regulatory reform efforts in the United States. A group of faculty members and researchers affiliated with the GW Regulatory Studies Center contributed to the research project, and subject matter experts at the USDA Office of the Chief Economist provided technical advice. The research report is available at: https://regulatorystudies.columbian.gwu.edu/putting-food-table-agriculture-and-regulation.