

Data and Methods for Analyzing Special Interest Influence in Rulemaking

Daniel Carpenter
Harvard University

Devin Judge-Lord
University of Wisconsin

Brian Libgober
Yale University

Steven Rashin*
New York University

January 30, 2020

Abstract

The United States government creates astonishingly complete records relevant to policy creation in executive agencies. In this article, we describe the major kinds of data that have proven useful to scholars studying interest group behavior in bureaucratic politics, how to obtain them, and also the major challenges that we as users have found in working with these data. We discuss established databases such as [regulations.gov](https://www.regulations.gov), which contains comments on regulations issued by executive branch agencies, and new sources of data, such as ex-parte meeting logs, which describe the interest groups and individual lobbyists that bureaucrats are meeting face-to-face about rules, and individually-identified personnel records of nearly all federal employees since 1973. One of the challenges in working with the newer datasets is that the whole databases are often not fully machine-readable. We argue that a productive way forward is to invest in making all the datasets machine-readable and to create consistent way to link them to each other and to outside databases.

Keywords: Interest groups, administrative lobbying, bureaucratic politics, data sources

Submission ID: IGAD-D-19-00044

*Corresponding Author: Steven.Rashin@NYU.edu, 19 West 4th St, 2nd Floor, New York, NY 10012

Submission ID: IGAD-D-19-00044

If the U.S. federal government is unquestionably good at one thing, it is pushing out paper. In theory, governmental records relevant to the creation of policy in executive agencies have long been available to researchers. Practically speaking, obtaining such data has been a difficult and costly undertaking. A decade or more ago, researchers were limited to acquiring data on just a few rules (e.g., [Golden, 1998](#)) or using surveys (e.g., [Furlong, 2004](#)). Beginning in 1994, when the government first released the *Federal Register* online, scholars now have access to astonishingly detailed data on the notice-and-comment rulemaking process and, hence, regulatory policy, regulatory policymakers, and interest group advocacy.

In this article, we describe the major kinds of data that have proven useful to scholars studying interest group behavior in bureaucratic politics, how to obtain them, and also the major challenges that we as users have found in working with these data. Three examples of the data sources that fit this description are machine-readable records of all agency rules published since 1994, comments available on [regulations.gov](#), and metadata about rules contained in the Unified Agenda. We also describe sources that have become more available in recent years, such as ex-parte meeting logs, which describe the interest groups and individual lobbyists that individual bureaucrats are meeting face-to-face about rules, and individually-identified personnel records of nearly all federal employees since 1973.

These data sources do not exhaust the kinds of records relevant to researchers, although, at present, they reflect what is available. Yet data accessibility increases all the time; for this reason, we also highlight sources of data that we believe are already or may soon become comprehensively available, perhaps after enterprising researchers submit the necessary FOIA requests. Examples include agency press releases (see e.g., [Libgober, N.d.](#); [Libgober and Carpenter, 2018](#)) and the Foreign Agents Registration Act (FARA) reports

(Shepherd and You, 2019).¹ To orient potential students of agency policymaking to available sources of data, we identify four units of analysis: (1) participants, (2) policymakers, (3) policy texts, and (4) policy timing. We describe where to find data on each in roughly the order they appear in the process of developing a rule.

Background: Rulemaking and the Administrative Procedures Act

For decades, the volume of legal requirements emerging from executive agencies has dwarfed the law-making activity of Congress, the Supreme Court, and the Presidency. Each year agencies publish four thousand or more regulations. In doing so, agencies are generally required to follow a process prescribed in Section 553 of the Administrative Procedures Act (APA). For each rule, agencies create a collection of documents that mark the stages of this policymaking process. These documents are organized in a “docket” folder with a unique name and ID number. Once kept in a literal folder, these documents are now online. For example, the Centers for Medicare Medicaid Services rulemaking docket 0163 “Transparency in Coverage” can be found at [regulations.gov/docket?D=CMS-2019-0163](https://www.regulations.gov/docket?D=CMS-2019-0163). Section 553 requires an agency to issue a Notice of Proposed Rulemaking (NPRM) in the *Federal Register* to notify potentially affected parties that the regulatory environment might change. Following the NPRM, agencies solicit comments on these regulations through [regulations.gov](https://www.regulations.gov) (for executive agencies) or the agency websites themselves (for independent agencies). Agencies are required to consider the comments but are not required to alter the rules based on them. The fact that the APA requires agencies to assemble a comprehensive record of who sought influence over any part of the policy-making process, and that such records are fairly comprehensive, is highly unusual in the American federal system, and makes the area particularly exciting for interest group scholars.

¹We have links to all the relevant data sources on our github page: <https://github.com/libgober/regdata/blob/master/README.md>.

1 Who participates in the rulemaking process?

Notice and comment rulemaking creates unique feature in American democracy; the right for anyone to participate in, perhaps even influence, most policymaking processes at federal agencies. Participation in rulemaking activities is highly skewed, with a few rules, such as the Federal Communication Commission’s rules on net neutrality receive millions of comments while half of proposed rules open for comment on regulations.gov receive no comments at all ([Libgober, N.d.](#)). Even the median rule designated as ‘economically significant’ – rules projected to have an annual economic impact of over \$100 million – receives fewer than ten comments ([Judge-Lord, 2019](#)). Participants include businesses, public interest groups, trade associations, unions, law firms, and academics ([Cuéllar, 2005](#); [Yackee and Yackee, 2006](#)).

Scholarly interest in patterns of participation in the policymaking process relies on the notion that the participants in the process seek to influence the development of the rule. Participation in this process can begin even before a proposed rule is issued. For example, [You \(2017\)](#) finds that half of all spending on lobbying legislation occurs *after* a bill becomes law. De [Figureido and Kim \(2004\)](#) show that meetings between agency officials and firms spike before an agency issues a policy order. [Libgober \(n.d.\)](#) shows that firms meeting with federal regulators before a rule is issued may receive abnormally high stock market returns upon its release, a finding consistent with the analysis of qualitative researchers that commenters who participate early in the rulemaking process can shape the content of the rule ([Naughton et al., 2009](#)). In some cases public participation can even initiate a rulemaking process via a petition for rulemaking.

To analyze patterns of participation, scholars use data from a variety of sources. Sources for data on participation are agency rulemaking dockets ([Golden, 1998](#); [Yackee, 2006](#); [Young, Marple, and Heilman, 2017](#); [Ban and You, 2019](#)), the *Federal Register* ([Balla, 1998](#), [West \(2004\)](#)), and regulations.gov ([Gordon and Rashin, nd](#)). Though commenters are

not generally required to disclose their names and affiliations, many do. The best current data sources for obtaining the names of the organizations that submit comments on federal regulations are regulations.gov and the websites of the independent agencies themselves.

Obtaining and working with data on comment participants

Scholars wishing to obtain data on comments and commenters from executive agencies are required to use the website regulations.gov. Actually obtaining the data requires overcoming a number of technical and bureaucratic hurdles. First, bulk data from the website can only be accessed via an Application Programming Interface (API) which requires an API key.² Even after signing up for an API key,³ we caution prospective users of the site that these keys can be deactivated without warning or acknowledgment of why they were deactivated. Second, obtaining an accurate count of comments is not straightforward as agencies have different policies regarding duplicated comments⁴ and confidential business information.⁵ Third, obtaining the data is time consuming.⁶

Scholars studying participation outside of executive agencies can often obtain these data from the websites of the agencies themselves. Unlike regulations.gov, websites do not have an API and thus require scholars to use bespoke web scrapers. We have examples of scrapers on <https://github.com/libgober/regdata>. While these websites are not rate limited or gated with API keys, we recommend exercising caution with the rate at which a

²An API is a set of procedures that allow a user to access data from a website in a structured way. Some websites limit API usage by requiring users to get an API key; regulations.gov is one of those websites.

³By emailing regulations@erulemakinghelpdesk.com with ‘your name, email address, organization, and intended use of the API.’ See <https://regulations.gov.github.io/developers/> for more details. The process to approve keys takes a few days, though some do not get approved at all.

⁴We note that sometimes the difference between the total number of reported comments and the number of comments on regulations.gov is due to mass comment campaigns being grouped together.

⁵See e.g., [Lubbers \(2012\)](#) and <https://www.regulations.gov/userNotice>.

⁶Regulations.gov is subject to a rate limit of 1,000 queries per hour which is especially problematic for scholars seeking to analyze the full text of comments, many of which are included only as attachments. As of October 4, 2019 Regulations.gov has 11,238,958 public submission documents representing over 70 million public comments and almost 1.5 million more rules and other documents. Downloading an individual document requires calling the API twice, once for the docket information and then a second time to download the linked file.

scholar solicits information from the servers, as going too fast can result in a temporary ban.

In addition to the challenges of obtaining the data, scholars also must choose how they want to preprocess the data before analyzing it. For example, to answer questions about the types of organizations that participate in notice and comment rulemaking, scholars first need to accurately identify these organizations so that they can be assigned the correct covariate data. This process, however, is not trivial as the same firm can be identified in numerous ways. For example, Goldman Sachs submitted comments to the Securities and Exchange Commission (SEC) and Federal Reserve Board (FRB) as The Goldman Sachs Group, Inc, Goldman Sachs Co. LLC, Goldman Sachs Bank USA, Goldman, Sachs & Co., Goldman Sachs Execution & Clearing, LP, among other identifiers. All of these entities need to be accurately matched to Goldman Sachs, including the Goldman Sachs Execution & Clearing comment which is on Goldman Sachs letterhead.⁷ The problem can be more acute with associations such as the American Bar Association (ABA) sometimes commenting as itself and other times as individual sections of the ABA. One way scholars have been able to get around this problem is to use fuzzy matching. Fuzzy matching is a process that compares two strings of text and gives a distance between them. This procedure will assign a small distance between Goldman Sachs and Goldman, Sachs & Co and a larger distance between Goldman Sachs and Merrill Lynch. The distances and optimal thresholds depend on the algorithm used. Fuzzy matching, however, is not a panacea as the same procedure will also show a small distance between the JP Morgan and Morgan Stanley, which are two separate entities. It will also not reveal whether an ABA letter is from the American Banker's Association or the American Bar Association, and many trade-associations active on similar rulemaking issues do unfortunately often share acronyms or have similar looking names. We note, however, that in expectation, additional noise generated from incorrect matching is likely to bias a scholar against finding any relationships within the data.

⁷See <https://www.sec.gov/comments/s7-03-10/s70310-43.pdf>.

2 Who writes rules?

In the last decade scholars have begun to study how the identities of the policymakers and the networks in which they are embedded affect the policymaking process. This recent scholarly attention is partially a function of the release of new data sources such as the U.S. Office of Personnel Management’s (OPM) data on government employees (e.g., [Bolton, Figueiredo, and Lewis, 2018](#)),⁸ Open Secrets’ lobbying (e.g., [Baumgartner et al., 2009](#)) and revolving door databases (e.g., [i Vidal, Draca, and Fons-Rosen, 2012](#); [Bertrand, Bombardini, and Trebbi, 2014](#)), machine readable lobbying disclosure act reports (e.g., [Boehmke, Gailmard, and Patty, 2013](#); [You, 2017](#)), meeting logs ([Libgober, n.d.](#)), and datasets of corporate board membership such as Boardex (e.g., [Shive and Forster, 2016](#)). Scholars have also been able to study the identities of policymakers through creative use of longstanding data sources such as the *Federal Register* (e.g. [Carrigan and Mills, 2019](#)).

Work on the networks between policymakers and private interests has reshaped our understanding of the policymaking space. Through exploiting meeting logs [Libgober \(n.d.\)](#) finds that comments and meetings with policymakers are associated with abnormal returns in the billions. [Carrigan and Mills \(2019\)](#) find that the number of job functions of the bureaucrats who write the rules are associated with both decreases in the time an agency takes to promulgate a rule and increases in the probability that the rule will be overturned in court. [i Vidal, Draca, and Fons-Rosen \(2012\)](#) find evidence that, among lobbyists, connections to politicians are more valuable than issue expertise. In fact, when a former Senate aide becomes a lobbyist the Senator’s retirement results in a substantial loss of income for that lobbyist.

⁸Note that the dataset [Bolton, Figueiredo, and Lewis \(2018\)](#) used is not public but there is a BuzzFeed version that is substantially similar that we discuss below.

Obtaining and Working with personnel and lobbying data

Scholars wishing to analyze the extent to which personnel influence the development of the rulemaking process can obtain this data from the BuzzFeed personnel data release. The BuzzFeed personnel data⁹ contains information on federal employees such as their salaries, job titles, and basic demographic data from 1973 through 2016.¹⁰ These are the most complete personnel records publicly available but they are subject to several important limitations. First, not all agencies and occupations are a part of this release.¹¹ Second, some of the employees in these data do not have a unique identification number. This means that common names such as ‘John Smith’ match multiple employees. The Veterans Health Administration (VHA) employed 24 John Smiths in 2014. Adding a middle initial still does not lead to unique identifiers as there are five John Smiths at the VHA with the same middle initial.

To obtain machine readable data on the identities of domestic lobbyists, scholars use two databases from Open Secrets – a non profit focused on tracking spending on politics in the US – on administrative and Congressional lobbying. Downloading the lobbying data is straightforward as it only requires an account to access the ‘bulk data’ page. The lobbying data comes from the required disclosures under the Lobbying Disclosure Act of 1995 (LDA). Note that the reporting threshold varies by type of firm (in-house have a higher minimum reporting threshold than lobbying firms) and over time.¹² The lobbying data covers 1999 through 2018 and is broken up into seven tables that can easily be loaded into R or Python.¹³

⁹Available at <https://archive.org/details/opm-federal-employment-data/page/n1>. Note that the complete dataset is massive (>\$30GB).

¹⁰Updated personnel files are available through 2018 from the OPM itself here: <https://www.fedscope.opm.gov/datadefn/index.asp>

¹¹The list includes at least 16 agencies and, within the covered agencies, law enforcement officers, nuclear engineers, and certain investigators (Buzzfeed, 2017). See <https://www.buzzfeednews.com/article/jsvine/sharing-hundreds-of-millions-of-federal-payroll-records>

¹²See <https://lobbyingdisclosure.house.gov/ldaguidance.pdf> for details.

¹³Note that the raw data can be downloaded from the Secretary of the Senate’s Office of Public Records here: https://www.senate.gov/legislative/Public_Disclosure/LDA_reports.htm

We note, however, that the data is not complete as some lobbyists do not disclose required contacts and the data does not contain exact monetary amounts.¹⁴ We recommend Open Secret's database as it is easier to use than the raw Senate data. Open Secrets also has a database on revolving door employees that shows the career paths of federal government workers that went to the private sector and are employed in some capacity where their work depends on interacting with the federal government.¹⁵ Unlike their lobbying database, this one must be scraped as there is no option to download it using bulk data.¹⁶

Lobbyists advocating for foreign clients are required to disclose these contacts under the Foreign Agent Registration Act. Foreign agents often lobby bureaucratic agencies; for example, the state Israel retained law firm Arnold & Porter for advice on, among other issues, registering securities with the SEC.¹⁷ The reports contain a multitude of data including the names of foreign entities, the firms representing them, the nature of the contact, and the specific officials contacted. The data are astonishingly complete - the website contains all 6264 FARA registrants and their foreign contacts since July 3, 1942. The FARA data can be accessed online through the Department of Justice.¹⁸ Unlike other data discussed in this essay, the FARA data is *not* all in machine readable form;¹⁹ converting from PDF to a machine readable format (e.g. csv) is neither easy nor error free. As a result, scholars using the data only focus on a subset such as You (2019) who focuses on lobbying activities by the governments of Colombia, Panama, and South Korea on their free trade agreements from 2003 through 2012. In the absence of a completely digitized archive, scholars can exploit the search functions to search by lobbying registrant (e.g., lobbying firm Squire Patton Boggs) or foreign principal (e.g., the government of Afghanistan). Similar to the discussion of the

¹⁴See <https://www.gao.gov/assets/700/698103.pdf>

¹⁵See <https://www.opensecrets.org/revolving/methodology.php> for details.

¹⁶Note that they promote academic work using their database so they do not actively discourage scraping their database.

¹⁷See <https://efile.fara.gov/docs/1750-Supplemental-Statement-20110729-13.pdf>

¹⁸<https://efile.fara.gov/ords/f?p=1381:1:13132679194789:::>

¹⁹The metadata - e.g. dates, registrants, clients - are available in machine readable form through the API or bulk data downloads page.

diversity of names above, the same actor can participate under different names. For example, the Embassy of the Islamic Republic of Afghanistan and the Embassy of Afghanistan are designated as separate entities under FARA.

Corporate executives, lawyers, and lobbyists often meet with agency officials to discuss rules. Records of these meetings are often recorded by agency personnel. Obtaining these data requires writing a web scraper for each agency which holds the meetings data, since the data are held in different places in each website. Since there are no uniform standards for reporting meeting data, the data differ substantially from agency to agency in both content and organization. The Federal Reserve, for example, groups meetings by subject but not by rule,²⁰ the SEC groups meetings by rule, and the CFTC and FCC has all their meetings in one place.²¹

3 What do the rules say?

All of the work discussed above relies on the notion that public participation and the rulewriters influence the content of rules; numerous studies have found that commenters influence rules (see e.g., Cuéllar, 2005; Yackee and Yackee, 2006; Naughton et al., 2009; Haeder and Yackee, 2015, 2018) at agencies such as the Department of Labor (Yackee and Yackee, 2006), the Department of Treasury (Cuéllar, 2005), the Environmental Protection Agency (Wagner and Peters, 2011), the Securities and Exchange Commission (Rashin, 2019), and numerous other agencies. This finding is consistent across interviews (e.g., Furlong, 1998), qualitative analysis (e.g., Cuéllar, 2005), and quantitative analysis (e.g., Yackee and Yackee, 2006). Firms and groups representing businesses' interests are influential in securing policy concessions from the rulemaking process (Yackee, 2006; Yackee and Yackee, 2006; Carpenter and Moss, 2013), particularly when they are unopposed. Other scholars have found that agencies

²⁰<https://www.federalreserve.gov/regreform/communications-with-public.htm>

²¹See <https://www.cftc.gov/LawRegulation/DoddFrankAct/ExternalMeetings?page=6> and <https://www.fcc.gov/proceedings-actions/ex-parte/archive-of-filings>

respond to citizens (Balla et al., 2019) and medical professionals (Balla, 1998; Gordon and Rashin, nd).

Obtaining and Working with Data on Rule Text

One of the most basic functions of the bureaucracy is to issue regulations that set specific, enforceable requirements implementing a law. The federal government publishes all rules in the *Federal Register*, accessible via [FederalRegister.gov](https://www.federalregister.gov). This database is comprehensive and has machine readable records of all regulations published after 1994.²² Scholars can access search for regulations by the *Federal Register* citation e.g. 47 FR 3431, *Federal Register* number, or, more usefully, through their API; unlike other APIs mentioned in this essay, the *Federal Register* API does not have any rate limits.²³

While the raw text of rules is relatively straightforward to obtain, there are several thorny theoretical and methodological issues scholars must overcome.²⁴ First, the standard path from notice of proposed rulemaking (NPRM) to public comments to final rule is not always straightforward. Some rules are withdrawn before a final rule. Other agencies issue interim final rules subject to comments. For studies that seek to compare the proposed and final rules, the most problematic rules are the ones where one proposed rule gets broken up into a few smaller final rules or the reverse, where small rules become bundled into one final rule. These rules pose challenges for inference as the processes that lead to amalgamation or separation are not well understood. Second, scholars must decide whether, and to what extent, they should preprocess the rulemaking data before feeding the data to a text analysis algorithm.

²²For issues from March 1939 through 1993, see <https://www.govinfo.gov/app/collection/fr/>, although note that they might require optical character recognition to make them in a readable format.

²³However we still recommend small delays to prevent flooding the server with requests.

²⁴We note that raw text of rules is not available from [regulations.gov](https://www.regulations.gov) or the independent agencies themselves. As these rules are often in PDF form, converting them to .txt for analysis introduces errors into the text.

4 Rule metadata

Rulemaking metadata, such as the time rules are released, allow scholars to answer questions about factors that affect the rulemaking environment. Scholars have exploited this data to study questions about regulatory delay, agenda setting, and the financial impact of rules. The often lengthy delays between statutory promulgation and the issuance of regulations is the subject of recent empirical scholarly interest. Scholars attribute various causes to regulatory delay including auditing ([Acs and Cameron, 2013](#)); outside contacts ([Balla, Deets, and Maltzman, 2011](#)); the political climate ([Potter, 2017](#); [Thrower, 2018](#)); personnel types ([Carrigan and Mills, 2019](#)); ossification ([Yackee and Yackee, 2012](#)); deadlines ([Lavertu and Yackee, 2012](#); [Carpenter et al., 2011](#); [Bertelli and Doherty, 2019](#)); staffing levels ([Bolton, Potter, and Thrower, 2015](#)). These scholars relied on data from OIRA ([Acs and Cameron, 2013](#), [Balla, Deets, and Maltzman \(2011\)](#); [Bolton, Potter, and Thrower, 2015](#); [Carrigan and Mills, 2019](#)), the Unified Agenda ([Bertelli and Doherty, 2019](#); [Potter, 2017](#); [Lavertu and Yackee, 2012](#); [Potter, 2019](#)), the *Federal Register* ([Yackee and Yackee, 2012](#); [Thrower, 2018](#)), and FDA’s drug approval and postmarket experience ([Carpenter et al., 2011](#)). Scholars have also exploited rule metadata to show that publicly-traded banks that submit comments account for \$7 billion in excess returns ([Libgober and Carpenter, 2018](#)).

Obtaining and working with the Unified Agenda and Press releases

Obtaining data from the Unified Agenda is relatively straightforward as all issues since 1995 are available online in machine readable form.²⁵ The Unified Agenda contains all the proposed regulations an agency plans to issue in the near future; making it an extremely useful datasource for studying questions about agenda setting and timing. There are, however, important limitations to this data. First, agencies report their early stage rulemaking to the Unified Agenda strategically ([Nou and Stiglitz, 2016](#)). Second, agencies do not list

²⁵The Unified Agenda from 1983 through 1994 is available in the *Federal Register*.

‘failed’ rules that did not become final rules in the Unified Agenda (Yackee and Yackee, 2012). Third, Coglianese and Walters (2016) note that the Unified Agenda misses most of regulatory agency’s work.

In addition to the disclosures mandated by law, agencies often issue press releases to notify the public of important agency actions. Much like the meetings data discussed above, the policies regarding their storage and dissemination differ from agency to agency. The Federal Reserve, for example, lists all press releases since 1996 on their website, while the SEC only has them from 2012.²⁶ When working with the press releases scholars often need data on the exact time documents were made available to the general public (see e.g., Libgober, N.d.). Obtaining the press release metadata requires exploiting Really Simple Syndication (RSS) feeds to extract the exact time a press release becomes public.

5 Discussion: Assembling Complete Databases

The United States government releases troves of data on rulemaking but in forms that require substantial effort from scholars to be useful for research. Scholars working on bureaucratic politics face two primary data challenges going forward: assembling complete, machine-readable datasets of agency rulemaking activity and linking these datasets to one another. There are four fruitful data projects which could increase researcher efficiency by preventing duplicated efforts to download and clean data. First, assembling a complete database of comment activity throughout the Federal government in one location that can be downloaded in bulk, including comment text and metadata. Second, creating a comprehensive database of revolving door rulemakers from the OPM personnel records, LDA disclosure forms, and FARA data. This database would be greatly helped by the addition of data from LinkedIn but that data is not currently public. Third, creating a database of all meeting activities throughout the federal government. Finally, creating unique identifiers for each

²⁶See <https://www.federalreserve.gov/newsevents/pressreleases.htm> and <https://www.sec.gov/news/pressreleases>

commenter that allow researchers to link commenting behavior to firm characteristics and other datasets. These projects, once complete, will allow scholars to pursue novel research on political participation, influence, and public management.

6 Conflicts of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Notes

¹We have links to all the relevant data sources on our github page: <https://github.com/libgober/regdata/blob/master/README.md>.

²An API is a set of procedures that allow a user to access data from a website in a structured way. Some websites limit API usage by requiring users to get an API key; regulations.gov is one of those websites.

³By emailing regulations@erulemakinghelpdesk.com with ‘your name, email address, organization, and intended use of the API.’ See <https://regulationsgov.github.io/developers/> for more details. The process to approve keys takes a few days, though some do not get approved at all.

⁴We note that sometimes the difference between the total number of reported comments and the number of comments on regulations.gov is due to mass comment campaigns being grouped together.

⁵See e.g., [Lubbers \(2012\)](#) and <https://www.regulations.gov/userNotice>.

⁶Regulations.gov is subject to a rate limit of 1,000 queries per hour which is especially problematic for scholars seeking to analyze the full text of comments, many of which are included only as attachments. As of October 4, 2019 Regulations.gov has 11,238,958 public submission documents representing over 70 million public comments and almost 1.5 million more rules and other documents. Downloading an individual document requires calling the API twice, once for the docket information and then a second time to download the linked file.

⁷See <https://www.sec.gov/comments/s7-03-10/s70310-43.pdf>.

⁸Note that the dataset [Bolton, Figueiredo, and Lewis \(2018\)](#) used is not public but there is a BuzzFeed version that is substantially similar that we discuss below.

⁹Available at <https://archive.org/details/opm-federal-employment-data/page/n1>. Note that the complete dataset is massive (>\$30GB).

¹⁰Updated personnel files are available through 2018 from the OPM itself here: <https://www.fedscope.opm.gov/datadefn/index.asp>

¹¹The list includes at least 16 agencies and, within the covered agencies, law enforcement officers, nu-

clear engineers, and certain investigators (Buzzfeed, 2017). See <https://www.buzzfeednews.com/article/jsvine/sharing-hundreds-of-millions-of-federal-payroll-records>

¹²See <https://lobbyingdisclosure.house.gov/ldaguidance.pdf> for details.

¹³Note that the raw data can be downloaded from the Secretary of the Senate's Office of Public Records here: https://www.senate.gov/legislative/Public_Disclosure/LDA_reports.htm

¹⁴See <https://www.gao.gov/assets/700/698103.pdf>

¹⁵See <https://www.opensecrets.org/revolving/methodology.php> for details.

¹⁶Note that they promote academic work using their database so they do not actively discourage scraping their database.

¹⁷See <https://efile.fara.gov/docs/1750-Supplemental-Statement-20110729-13.pdf>

¹⁸<https://efile.fara.gov/ords/f?p=1381:1:13132679194789:::>

¹⁹The metadata - e.g. dates, registrants, clients - are available in machine readable form through the API or bulk data downloads page.

²⁰<https://www.federalreserve.gov/regreform/communications-with-public.htm>

²¹See <https://www.cftc.gov/LawRegulation/DoddFrankAct/ExternalMeetings?page=6> and <https://www.fcc.gov/proceedings-actions/ex-parte/archive-of-filings>

²²For issues from March 1939 through 1993, see <https://www.govinfo.gov/app/collection/fr/>, although note that they might require optical character recognition to make them in a readable format.

²³However we still recommend small delays to prevent flooding the server with requests.

²⁴We note that raw text of rules is not available from regulations.gov or the independent agencies themselves. As these rules are often in PDF form, converting them to .txt for analysis introduces errors into the text.

²⁵The Unified Agenda from 1983 through 1994 is available in the *Federal Register*.

²⁶See <https://www.federalreserve.gov/newsevents/pressreleases.htm> and <https://www.sec.gov/>

[news/pressreleases](#)

References

- Acs, Alex, and Charles M. Cameron. 2013. "Does White House Regulatory Review Produce a Chilling Effect and "OIRA Avoidance" in the Agencies?" *Presidential Studies Quarterly* 43 (3): 443–467.
- Balla, Steven J. 1998. "Administrative Procedures and Political Control of the Bureaucracy." *American Political Science Review* 92: 663–673.
- Balla, Steven J., Alexander R. Beck, William C. Cubbison, and Aryamala Prasad. 2019. "Where's the Spam? Interest Groups and Mass Comment Campaigns in Agency Rule-making." *Policy & Internet*.
- Balla, Steven J., Jennifer M. Deets, and Forrest Maltzman. 2011. "Outside Communications and OIRA Review of Agency Regulations." *Administrative Law Review* 63: 149–177.
- Ban, Pamela, and Hye Young You. 2019. "Presence and influence in lobbying: Evidence from Dodd-Frank." *Business and Politics* 21 (2): 267–295.
- Baumgartner, Frank R., Jeffrey M. Berry, Marie Hojnacki, David C. Kimball, and Beth L. Leech. 2009. *Lobbying and Policy Changes: Who Wins, Who Loses, and Why*. University of Chicago Press.
- Bertelli, Anthony M., and Kathleen M. Doherty. 2019. "Setting the Regulatory Agenda: Statutory Deadlines, Delay, and Responsiveness." *Public Administration Review* 79 (5): 710–720.
- Bertrand, Marianne, Matilde Bombardini, and Francesco Trebbi. 2014. "Is It Whom You Know or What You Know? An Empirical Assessment of the Lobbying Process." *American Economic Review* 104 (12): 3885–3920.
- Boehmke, Frederick J., Sean Gailmard, and John W. Patty. 2013. "Business as usual: interest group access and representation across policy-making venues." *Journal of Public Policy* 33 (1): 3–33.
- Bolton, Alexander, John M. De Figueiredo, and David Lewis. 2018. "Elections, Ideology, and Turnover in the U.S. Federal Government." *Academy of Management Proceedings* 2018 (1): 12717.
- Bolton, Alexander, Rachel Augustine Potter, and Sharece Thrower. 2015. "Organizational Capacity, Regulatory Review, and the Limits of Political Control." *Journal of Law, Economics, and Organization* 32 (2): 242–271.
- Carpenter, Daniel, and David A Moss. 2013. *Preventing regulatory capture: Special interest influence and how to limit it*. Cambridge University Press.
- Carpenter, Daniel, Jacqueline Chattopadhyay, Susan Moffitt, and Clayton Nall. 2011. "The Complications of Controlling Agency Time Discretion: FDA Review Deadlines and Post-market Drug Safety." *American Journal of Political Science* 56 (1): 98–114.

- Carrigan, Christopher, and Russell W. Mills. 2019. "Organizational Process, Rulemaking Pace, and the Shadow of Judicial Review." *Public Administration Review* 79 (5): 721–736.
- Coglianesi, Cary, and Daniel E Walters. 2016. "Agenda-Setting in the Regulatory State: Theory and Evidence." *Administrative Law Review* 68: 93.
- Cuéllar, Mariano-Florentino. 2005. "Rethinking Regulatory Democracy." *Administrative Law Review* 57 (2): 411–499.
- Figureido, J., and J. K. Kim. 2004. "When do Firms Hire Lobbyists? The Organization of Lobbying at the Federal Communications Commission." *Industrial and Corporate Change* 13 (6): 883–900.
- Furlong, S. R. 2004. "Interest Group Participation in Rule Making: A Decade of Change." *Journal of Public Administration Research and Theory* 15 (3): 353–370.
- Furlong, Scott R. 1998. "Political Influence on the Bureaucracy: The Bureaucracy Speaks." *Journal of Public Administration Research and Theory* 8 (1): 39–65.
- Golden, Marissa Martino. 1998. "Interest groups in the rule-making process: Who participates? Whose voices get heard?" *Journal of Public Administration Research and Theory* 8 (2): 245–270.
- Gordon, Sanford, and Steven Rashin. nd. "Citizen Participation in Policymaking: Evidence from Medicare Fee Schedule Revisions." *Journal of Politics* .
- Haeder, Simon F., and Susan Webb Yackee. 2015. "Influence and the Administrative Process: Lobbying the U.S. President's Office of Management and Budget." *American Political Science Review* 109 (3): 507–522.
- Haeder, Simon F, and Susan Webb Yackee. 2018. "Presidentially Directed Policy Change: The Office of Information and Regulatory Affairs as Partisan or Moderator?" *Journal of Public Administration Research and Theory* 28 (4): 475–488.
- i Vidal, Jordi Blanes, Mirko Draca, and Christian Fons-Rosen. 2012. "Revolving Door Lobbyists." *American Economic Review* 102 (7): 3731–3748.
- Judge-Lord, Devin. 2019. Why Do Agencies (sometimes) Get So Much Mail? Lobbying coalitions, mass comments, and political information in bureaucratic policymaking. Southern Political Science Association Annual Conference.
- Lavertu, Stéphane, and Susan Webb Yackee. 2012. "Regulatory Delay and Rulemaking Deadlines." *Journal of Public Administration Research and Theory* 24 (1): 185–207.
- Libgober, Brian. n.d. "Meetings, Comments, and the Distributive Politics of Administrative Policymaking." *Quarterly Journal of Political Science* .
- Libgober, Brian. N.d. "Strategic Proposals, Endogenous Comments, and Bias in Rulemaking." *Journal of Politics*. Forthcoming.

- Libgober, Brian, and Dan Carpenter. 2018. "What's at Stake in Rulemaking? Financial Market Evidence for Banks' Influence on Administrative Agencies."
- Lubbers, Jeffrey S. 2012. *A guide to federal agency rulemaking*. American Bar Association.
- Naughton, Keith, Celeste Schmid, Susan Webb Yackee, and Xueyong Zhan. 2009. "Understanding commenter influence during agency rule development." *Journal of Policy Analysis and Management* 28 (2): 258–277.
- Nou, Jennifer, and Edward H. Stiglitz. 2016. "Strategic Rulemaking Disclosure." *Southern California Law Review* 89: 733–786.
- Potter, Rachel Augustine. 2017. "Slow-Rolling, Fast-Tracking, and the Pace of Bureaucratic Decisions in Rulemaking." *The Journal of Politics* 79 (3): 841–855.
- Potter, Rachel Augustine. 2019. *Bending the Rules*. The University of Chicago Press.
- Rashin, Steven. 2019. "Private Influence over the Policymaking Process."
- Shepherd, Michael, and Hye Young You. 2019. "Exit Strategy: Career Concerns and Revolving Doors in Congress." *American Political Science Review* pp. 1–15.
- Shive, Sophie A., and Margaret M. Forster. 2016. "The Revolving Door for Financial Regulators." *Review of Finance* 21 (4): 1445–1484.
- Thrower, Sharece. 2018. "Policy Disruption Through Regulatory Delay in the Trump Administration." *Presidential Studies Quarterly* 48 (3): 517–536.
- Wagner, Wendy, Katherine Barnes, and Lisa Peters. 2011. "Rulemaking in the Shade: An Empirical Study of EPA's Air Toxic Emissions Standards." *Administrative Law Review* 63 (1): 99–158.
- West, William F. 2004. "Formal Procedures, Informal Processes, Accountability, and Responsiveness in Bureaucratic Policy Making: An Institutional Policy Analysis." *Public Administration Review* 64 (1): 66–80.
- Yackee, Jason Webb, and Susan Webb Yackee. 2006. "A Bias Towards Business? Assessing Interest Group Influence on the U.S. Bureaucracy." *The Journal of Politics* 68 (1): 128–139.
- Yackee, Jason Webb, and Susan Webb Yackee. 2012. "An empirical examination of federal regulatory volume and speed, 1950–1990." *George Washington Law Review* 80: 1414–92.
- Yackee, Susan Webb. 2006. "Sweet-Talking the Fourth Branch: The Influence of Interest Group Comments on Federal Agency Rulemaking." *Journal of Public Administration Research and Theory* 16 (1): 103–124.
- You, Hye Young. 2017. "Ex Post Lobbying." *The Journal of Politics* 79 (4): 1162–1176.
- You, Hye Young. 2019. "Dynamic Lobbying: Evidence from Foreign Lobbying in the US Congress." Working Paper.

Young, Kevin L., Tim Marple, and James Heilman. 2017. "Beyond the revolving door: Advocacy behavior and social distance to financial regulators." *Business and Politics* 19 (2): 327–364.