Federal Register Description and Codebook

Emily H. Moore
Department of Political Science
Washington University in Saint Louis.

February 25, 2018

To date, most studies of rulemaking utilize data taken from the Unified Agenda. The Unified Agenda is a semi-annual accounting of federal regulatory activity. The Regulatory Information Service Center (a part of GSA) compiles the Unified Agenda in conjunction with the Office of Information and Regulatory Affairs (OIRA) within OMB.[1] The Unified Agenda data has several upsides. Apart from being widely used and accepted, the Unified Agenda notes the significance level of the rules and includes other information about various tests the rules underwent (such as environmental impact assessments). The main problem with the Unified Agenda is its incompleteness. Rules that are started and completed (many direct final rules) between reporting periods may not appear, for example. Moreover, agencies self-report their activities and face no penalty for excluding information (O'Connell 2008). Because all rules must be printed in the Federal Register and because it is the official means of notifying the public of regulatory activity, The Federal Register is the most complete source of regulatory activity. As O'Connell (2011) notes: "In an ideal world, empirical studies of rulemaking would use individual agency rule submissions to the Federal Register. Since virtually all rules are published in the Federal Register, there would be no selection issues as to what rules were included." She goes on to note that using Federal Register data has been prohibitive to collect leading to more limited studies of short time periods and fewer agencies. Finally, using a dataset based on the Federal Register, Nou and Stiglitz

[1] More information from this website: `https://www.reginfo.gov/public/jsp/eAgenda/StaticContent/UA_About.jsp`

(2015) find that the Unified Agenda has severe issues with underreporting of proposed rules and may even exhibit politically strategic underreporting.

1 Collection Method

As mentioned above, Nou and Stiglitz (2015) also collected a dataset of rulemaking activity based on the Federal Register. Their data set is based on "machine-readable Federal Register entries from a variety of sources" (747). They search headings for proposed rules and exclude entries that seem to be for proposed hearings or other activities. My data set covers fewer years but from what I can tell, uses a different collection method and includes more information. My dataset also focuses on final rules rather than proposed rules.

To collect the data, I utilized the Federal Register .xml files held by the Government Printing Office.[2] There is one .xml for every business day which includes everything published in the Federal Register for that day. Helpfully, .xml files contain tags of each type of document, so I could extract individual document types. I used this to obtain final rules (including direct final rules and other rule changes), but the same process could be used to obtain Notices of Proposed Rulemaking and other agency documents published in the Federal Register. Within each document type, there are tags for various pieces of information about the rule itself–the date it was published, the effective date of the rule, the signatories on the rule and more (as described in the next section). Using Python, I extracted the information in each relevant tag and produced a .csv file of these data for each final rule published in the Federal Register from 2000-2015. There are a total of 61,216 final rule documents. This collection could easily be extended to the present as new Federal Register files are quickly placed in the repository. Though I collected data about final rules, the Federal Register also contains the text of final rules as well. The script I used to collect the data is available in the dataverse. It should be noted, however, that the data that came directly from this script have been partially cleaned, so if the script were to be used to collect new data, these would

[2]Available here: `https://www.gpo.gov/fdsys/browse/collection.action?collectionCode=FR`

need to be cleaned as well. (Tags removed, for example).

2 Variable Descriptions

One future challenge of these data is that, while they possess a lot of information, there are also occasionally errors in inputs (information input into the wrong tag, for example) or differences in the way a particular agency was input. For example, the Federal Aviation Administration might be listed as "Department of Transportation (FAA), Federal Aviation Administration, Federal Aviation Administration (DOT)," etc.

Below is a listing of the variables in the dataset and their descriptions, including potential work that will need to be done to use the variable in analysis.

Date: The date of publication in the Federal Register in YYYY-MM-DD format.

Agency: The agency that produced the rule. If the agency is within a department, this is usually the department, though sometimes no department is listed. If an independent agency, it will be listed here. See cleaned agency for a better version of this and the next variable combined. I left in this variable information for completeness and in case department is the relevant level of analysis rather than subunit.

Subagency: The subagency that produced the rule. This is the unit that actually produced the rule if it is within a department. For example, the Farm Service Agency is within the Department of Agriculture. It will be NA in about a third of cases because independent agencies are listed under the agency section instead. It should be noted that the Federal Register did not input these data perfectly (e.g. differences in how to list the same agency, a spelling error or two) and there are some strange instances of multiple agencies sponsoring the same rule that appear oddly. See Cleaned Agency for a cleaner, combined version of the two variables above.

CFR: This is the parts of the Code of Federal Regulations affected by the rule. Sometimes this includes multiple sections as well as multiple parts within the same section. There are a handful of mistakes from the way the data were originally input that remain in this variable and will need to be cleaned.

Signature Date: The date the final rule was signed.

Signatory Name: The name of the person or people who signed the rule.

Signatory Title: The job title of the person or people who signed the rule.

RIN: The rule identification number. Not all rules have one. If it does, it should be trackable with sources like Regulations.gov, the Unified Agenda, etc.

FRDOC: The Federal Register's document identification.

Effective Date: The date the rule is supposed to go into effect. Unfortunately, this is in a text format because agencies input this in a variety of ways. Dates would need to be extracted from text. For example, some rules simply list the date while others have some text before the date. Some rules may have multiple relevant dates. Agencies sometimes switched between using this tag and the following tag. About one third of rules do not have anything listed.

Important Dates: This includes important dates. This sometimes includes effective dates although less frequently than the previous variable. About two thirds of rules do not have anything listed here. Could be combined with Effective Date. Only about 1400 are NA on both, which suggests agencies switched between the two in most cases.

Addresses: Has address for questions or for where paper comments may be sent.

Subject List: This is a very useful list of subjects contained in the rule. This is either coded by the agency or the Federal Register, I'm not sure which. The difficulty with the

subject list is that it is difficult to know how many subjects are contained in each rule. It would be necessary to identify situations where the subject is the same but described in a slightly different way. It may be possible to identify all possible subjects and dummy code each subject for each rule. It will take some work to a

AUTH: This is a list of authorizing material for the rule. This includes areas of the U.S. Code, references to specific public laws, and executive orders.

Agency2: This is another listing of agency provided in the listing. This one has more issues with language differences than most of the others.

Action: The type of rule or action being completed. Includes final rules, direct final rules, and much more. Could probably be consolidated into fewer categories than the many given.

Summary: This is a provided summary of the rule. About 4,700 do not have summaries.

First Page: The page of the Federal Register on which the rule starts.

Last Page: The page of the Federal Register on which the rule ends. May be the same as the first page for a short rule.

Rule Title: The title of the rule.

Cleaned Agency: This is my take on cleaning the agency data. I utilized the subagency header when possible and grabbed the agency variable when the subagency variable was not defined. For example, if the agency was listed as "Department of Transportation" and the subagency as "Federal Aviation Administration," this variable will be "Federal Aviation Administration." If agency is listed as "Environmental Protection Agency" and subagency is blank, this variable will be "Environmental Protection Agency." This creates a more unified listing of the rule producer(s) at the lowest possible level. A small number of rules have multiple creators. Often these are procurement rules from DOD/NASA/GSA. I would

recommend excluding rules with more than two or three sponsors from analysis or find a more appropriate way to list them.

3 Simple Statistics and Plots

Figure 1 shows the number of all final actions and final rules (those tagged as final rules in the action category rather than direct final rules, interim final rules, or other such actions) over time.

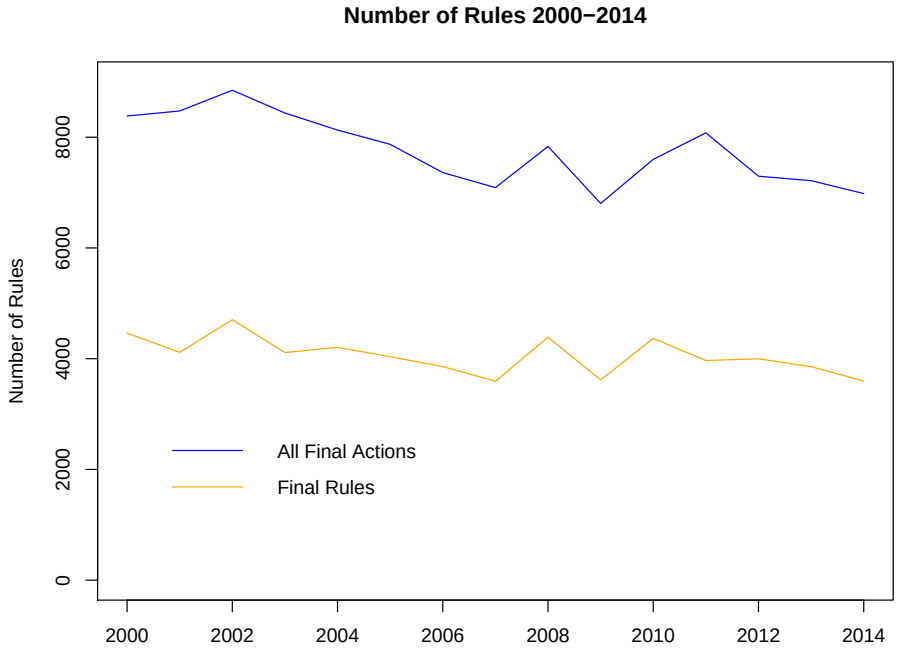

Figure 1: Number of All Final Actions and Final Rules Over Time.

Table 1 shows the top 20 individual rule producers (where a rule is all rules tagged as final in the .xml including direct final rules, temporary rules, and corrections) from 2000-2015. This excludes team producers. The Department of Defense and NASA often co-produce rules and sometimes involve other agencies such as GSA. These seem to be procurement regulations.

Table 1: Table of the top 20 individual rule producers 2000-2015.

Agency	Number of Rules
Federal Aviation Administration	12,275
Environmental Protection Agency	8,681
Coast Guard	6,051
National Oceanic and Atmospheric Administration	4,665
Federal Communications Commission	3,061
Food and Drug Administration	1,886
Internal Revenue Service	1,346
Agricultural Marketing Service	1,241
Federal Emergency Management Agency	1,137
Animal and Plant Health Inspection Service	939
Fish and Wildlife Service	801
Department of Defense (no subagency listed)	612
Defense Acquisition Regulation System	542
Securities And Exchange Commission	508
Department of Veterans Affairs	485
Nuclear Regulatory Commission	474
National Highway Traffic Safety Administration	466
Federal Energy Regulatory Commission	461
Postal Service	415
Office of Personnel Management	407

The Federal Aviation Administration is the largest producers of final actions each year. This is consistent with my previous research on the Regulations.gov data and consistent with Nou and Stiglitz's (2015) Proposed Rules dataset. The second largest producer is the EPA followed by the Coast Guard. The Coast Guard produces a lot of temporary deviations other smaller regulations. It is also worth noting that independent regulatory commissions like the FCC and SEC make a large number of regulations that are often excluded from other sources (including my own Regulations.gov data) because they are typically not subject to OIRA review.

As can be seen in Table 2, there are a variety of different action types listed. Many of these categories can probably be combined but are left in tact for completeness. (e.g. "Final Rule; Correction" is likely similar to "Correcting Amendment".) The top ten categories make up about 75 percent of the rules in the data. Over half of the rules are listed as final rules. Before using, it would be prudent to check for typographical errors in the rule action

listing.

Table 2: Top ten rule action types. These categories make up about 75 percent of all rules in the data.

Action Type	Number of Actions in Category
Final Rule	31,574
Direct Final Rule	2,560
Temporary Final Rule	2,493
Final Rule; Request for Comments	1,864
Interim Rule; Request for Comments	1,541
Final Rule; Correction	1,377
Notice of Temporary Deviation from Regulations	935
Interim Rule	923
Temporary Rule; Closure	916
Correcting Amendment	736

Table 3 lists the 25 most common subjects tagged in rules. As mentioned above, the subjects lists are somewhat messy and will require cleaning to use most effectively. For this table, I cleaned up the subjects somewhat and split them by commas (as this is how multiple subjects in individual rules are divided). I then counted the most common uses. AS is visible, reporting and recordkeeping is the most common subject listed on a rule followed by incorporation by reference. Many of the top subjects list airworthiness, aircraft, or aviation. Given that the FAA is the top rule producer, this is not surprising. An investigator may wish to create a single subject that lists any of these categories, especially since a single rule may list all of these subjects. Air pollution, ozone, pesticides, and environmental protection are not surprising either, considering the proliferation of EPA regulations.

Table 3: List of 25 Most Common Subjects Tagged in Rules

Subject	Count
Reporting and recordkeeping requirements	17,828
Incorporation by reference	12,099
Aircraft	7,520
Aviation safety	7,481
Environmental protection	7,337
Safety	6,820
Administrative practice and procedure	6,759
Air transportation	6,687
Air pollution control	4,708
Intergovernmental relations	4,287
Navigation (water)	3,072
Marine safety	2,842
Waterways	2,594
Security measures	2,471
Ozone	2,436
Airspace	2,411
Harbors	2,407
Agricultural commodities	2,134
Navigation (air)	2,131
Volatile organic compounds	1,933
Imports	1798
Pesticides and pests	1778
Nitrogen dioxide	1762
Government procurement	1663
Particulate matter	1639

4 References

O'Connell, Ann Joseph. 2008. "Political Cycles of Rulemaking: An Empirical Portrait of the Modern Administrative State." *Virginia Law Review* 94(4):889-986.

O'Connell, Ann Joseph. 2011. "Agency Rulemaking and Political Transitions." *Northwestern University Law Review* 105(2):471-534.

Stigliz, Edward H. and Jennifer Nou. 2015. "Strategic Rulemaking Disclosure." *Southern California Law Review* 89: 733-792.