Kevin Shotwell

Professor John Wilkerson

8 August 2013

Whitepaper on the "Auth" Dataset

Introduction

This whitepaper attempts to give a brief description of the "Auth" dataset. It is anticipated that this dataset will be useful primary for researchers attempting to examine the manner by which laws passed by Congress effect administrative regulations of the federal government. It starts by describing the publications from which the dataset has been created, followed by a more technical description of the dataset itself. The paper concludes by discussing some ways in which the dataset can be used and built upon.

Publications

The path from a bill to an administrative regulation involves numerous government publications. The process is not typically linear from document to document—that is, it is difficult or impossible to compare any single publication to any single succeeding publication. The following table provides a summary of the involved publications, which are later discussed in more detail. The table is organized roughly in order of the "chain" that must be followed to track sources of authority.

Publication Name	**Best Format**	**Electronic Availability**	**Historical Versions**
Legislation	Text	GPO starting 1993	N/A
Public/Private Laws	Text	GPO starting 1995	N/A
Statutes-at-Large	Text	GPO starting 1951	N/A
Table III	HTML	OLRC	No
United States Code	HTML	GPO starting 1994	Annually
Parallel Table of Authorities	Text	GPO	No
Code of Federal Regulations	XML	GPO starting 1996	Annually
Federal Register	XML	GPO starting 2000	N/A

Legislation

Only legislation passed by both chambers of Congress and approved by the President constitute laws of the United States. *U.S. Const. art. I, § 7*. Legislation that has the potential to become a law includes all bills, both public and private, and most joint resolutions. The final form of a piece of legislation that has passed both chambers is known as the "enrolled" version.

Legislation typically consists of one or more sections, numbered in increasing (although not necessarily sequential) order, each containing "a single proposition of enactment." 1 U.S.C. §

104. Sections also have headings that briefly describe the intent of the section. These headings are written by members of Congress, and are sometimes politicized[1].

It should be noted that legislation can contain a single section that amends existing law to insert or change multiple sections. This most commonly happens when amending major pieces of non-positive law (defined later); for example, the Public Health Service Act or Social Security Act.

Public/Private Laws

Upon passage by both chambers of Congress and approval by the President, bills and joint resolutions are given a either a "Public Law number" or a "Private Law number." This number is assigned by the Office of the Federal Register (OFR) by taking the number of the current Congress, following it with a hyphen, and adding a number one greater than the previous public or private law, numbered separately and resetting to one at the start of each new Congress. For example, the Patient Protection and Affordable Care Act is Public Law 111-148, because it is the 148th public law enacted by the 111th Congress.

Statutes-at-Large

At the end of each Congressional session, all laws—public and private—passed during the session are concatenated and published in a volume of the Statutes-at-Large. The OFR handles this as well, in coordination with the Government Printing Office. Concurrent resolutions, Presidential proclamations, and Constitutional amendments proposed to or ratified by the states are also included in the statutes-at-large, but are less commonly a source of authority and are laws in a Constitutional sense. 1 U.S.C. § 112. The content of the statutes-at-large is legal evidence of the law in a court of law. *Id.*

United States Code (USC)

As a legal reference and research aid, the United States House of Representatives Office of the Law Revision Council (OLRC) annually publishes compilations of all statutes currently in effect that are of a "general and permanent" nature. 2 U.S.C. § 285b. "General" excludes private laws and laws with a very narrow effect (for example, naming a postal office), while "permanent" excludes temporary public laws, such as appropriation bills. This compilation is a significant undertaking, which involves significant discretion on behalf of the professional staff of the House.

The USC consists of 51 titles and 5 appendixes. 26 of these titles constitute "positive law," which means that the entire title was enacted in identical form by Congress—including the organization, section numbers, and statutory text. Bills that amend a positive law title reference the USC title and section directly. The other 25 titles are "non-positive law," which are compiled editorially by the OLRC from statutes. Bills that amend a non-positive law title must reference the specific statute being amended, repeal the prior statute and reenact it with the amendments, or create an entirely new section in the title. For example, the Internal Revenue Code of 1986, the last major tax restructure law, is directly amended by nearly all newly introduced taxation

[1] For an example, see Section 1003 of the Affordable Care Act, which has the header, "Ensuring that consumers get value for their dollars."

bills. The 5 appendixes to the USC have the full force of law, but are separated for organizational purposes—for example, the Federal Rules of Civil Procedure are an appendix to Title 28.

Titles are inconsistently subdivided into a combination of any of the following: subtitles, chapters, subchapters, parts, and subparts. Regardless of the method of subdivision, each section number is unique within the title—making only the title and section number necessary to locate a specific section.

Each section in the USC consists of five distinct parts: the section number, subject, body, source credit, and notes. Because the USC is highly cited in legal publications, the OLRC avoids renumbering sections. Because the USC is also highly categorized by topic, this means that new sections sometimes must be inserted between existing sections. To do so, the OLRC adds additional parts to the section number. So far, the most complex USC section numbers take the following form:

- One or more numbers
- One or more letters
- A single dash
- One or more numbers
- One or more letters
- One or more "version" numbers

For example, Title 42 has been so heavily amended that section 3201 of the Affordable Care Act was partially codified as 42 U.S.C. § 1395w–27a.

The title and body of a USC section are largely as passed by Congress, with minor editorial changes in non-positive law titles. A section's source credit consists of the information necessary to locate statutes that the code section was derived from.

Finally, the notes portion of a section consists of statutory and editorial notes. Statutory notes are directly taken from a statute, and have the full force of law. The decision to place statutory language in section notes rather than a section body is an editorial decision made by the OLRC. Typically, notes are "boilerplate" language, such as short title names and effective dates, as well as statutory text that is not "general and permanent" but still necessary to fully understand a section. Additionally, new public laws that do not amend a positive law title, but create new provisions that best fit under an existing positive law title, *must* be placed in the positive law title as a statutory note rather than a section body. Editorial notes are written by the OLRC, and detail their classification and codification work with regard to a section. Editorial notes are purely research tools, and have no legal effect.

Positive law titles of the USC are legal evidence of the law, while non-positive law titles are only *prima facie* evidence of the law. 1 U.S.C. § 204.

Table III

Table III is a resource maintained by the OLRC. It maps statutory sections, or portions thereof, to their ultimate destination in the USC. It is typically used by researchers interested in how a particular statute has been classified and codified into the USC. Table III contains all statutes that

have ever had provisions placed into the USC, even if such provisions have been later repealed. This information exists for the entire history of the nation.

Federal Register

Where authorized by Congress, federal agencies are permitted to promulgate administrative regulations that have the force of law. Starting in 1936, Congress began requiring that federal agencies submit their regulations for publication in the newly created Federal Register. Eventually, the Federal Register became the "newspaper of the federal government," containing notices and proposed rules in addition to final agency regulations.

A new edition of the federal register is published every government business day. Each edition contains multiple publications; each with a specific agency's recently submitted proposed or final rules in a subject area. Typically, proposed rules are published for a public comment period, after which they are revised and become final rules. Regardless, both types of publications contain the name of the responsible agency and subagency, a unique Federal Register Document number, a subject, and an "action." Actions are not standardized, but are most commonly something along the lines of "final rule," "proposed rule," "temporary final rule," "interim final rule with request for comments," "final rule, correction," etc. Some agencies choose to use more specific language, for example, "final rule; notification of U.S. fish quotas."

Each rule publication, at some point or another, contains "instructions" on how to amend existing regulations or create new regulations. These instructions start with an amendatory line, which specifies the target CFR location (see below) and the operation to be applied (insertion, replacement, etc.) Usually, each new regulatory section is then set forth at full length. However, in the case of minor amendments or corrections, the amendatory line may simply prescribe words or sentences to be changed.

It is also typical for rule publications to include an "authority update." This usually takes the form of a sentence either ratifying the existing authority or amending it in response to changes in the publication. This authority update applies to the entire CFR part (see below) in which the regulation is to be codified.

Publications in the Federal Register create a rebuttable presumption of proper rule promulgation and text. 44 U.S.C. § 1507.

Code of Federal Regulations (CFR)

The Code of Federal Regulations is an annual publication of the Office of the Federal Register (OFR). The CFR is to the Federal Register what the USC is to the Statutes-at-Large. That is, the "official" source of federal regulations is the Federal Register, and the CFR is codified source. It is a codification of the general and permanent administrative regulations that have been established by executive departments and agencies of the federal government and published in the Register, divided into 50 categorical titles. These titles are then divided into chapters, which are then assigned by OFR to agencies that issue regulations in that title's subject area.

Each chapter is then divided into one or more parts, which agencies themselves create. Parts have a unique number within each title, and are therefore able to be located without regard to

chapters. Parts may be further subdivided into subparts or other divisions, but their essential unit is the section. Each section has a number which is unique within each part.

Generally, each part of the CFR contains a single "authority clause" that cites the statutes or other documents that allow the agency to enforce the regulations contained within that part. Each part also cites the Federal Register publications that the part was derived from. On occasion, specific sections will have their own source citations to the Register.

The CFR constitutes *prima facie* evidence of the text of federal regulations. 44 U.S.C. § 1510.

Parallel Table of Authorities

The Parallel Table of Authorities, technically an index to the CFR, is a compilation of the authority citations for each part of the CFR. These authority citations are provided by the promulgating agency, and are nearly always references to a USC section. Other citation classes currently include portions of the Statutes at Large, public laws, presidential notices and proclamations, Executive Orders, determinations, directives, memorandums, and reorganization plans. In theory, the Parallel Table contains only the information available by parsing the authority clause from each part of the latest CFR publication. Because federal agencies themselves are responsible for providing and updating these authorities, and because failure to publish a proper authority clause is not necessarily a bar to enforcing a regulation, the Parallel Table contains errors and inconsistencies.

As noted by the Cornell Legal Information Institute[2], the Parallel Table typically omits references to 5 U.S.C. § 301, which provides broad authority to agency heads with regard to internal management.

Dataset Description

The dataset that this whitepaper describes is in the form of a MySQL relational database containing 12 tables. The following table provides a summary of the database tables, which are later discussed in more detail.

Table Name	Description	Source	Historical	Rows	Size
cfr_key	One row for each unique CFR title and part number. Provides a global key for each CFR part.	GPO XML	Annually from 1996	9,105	1 MB
cfr_part	One row for each CFR part for each year.			117,701	33 MB
cfr_section	One row for each CFR section for each year.			3,539,645	11 GB
parallel	One row for each authority to CFR part mapping. Official listing of authority for each CFR part's rules.	GPO Text	No	80,620	11 MB
register_entries	One row for each distinct publication in the Federal	GPO XML	Daily from 2000	86,381	25 MB

[2] http://www.law.cornell.edu/ptoa

	Register.				
register_rules	One row for each section found in a proposed or final rule Federal Register publication.			305,847	861 MB
statutes	One row for each section found in a public law.	GPO PLaw Text	Annually from 1995	37,409	251 MB
table3	One row for each USC section mapping in Table III.	OLRC HTML	No	288,871	44 MB
usc	One row for each USC section for each year.	GPO HTML	Annually from 1994	761,260	6 GB
usc_key	One row for each unique USC section citation. Provides a global key for each Code section.			53,800	10 MB

CFR Tables

The three CFR tables ("cfr_key", "cfr_part", and "cfr_section") are derived by parsing XML files provided by the GPO[3]. In general, these XML files are well structured and relatively easy to parse. However, because of the vast range of data contained in the CFR, some notable issues do arise. In addition, the amount of data means that manual review of the entire process is not possible, so additional issues may be undiscovered. The schema for the tables is as follows:

Table	**Column**	**Type**	**Nullable**	**Description**
cfr_key	id	Unsigned integer	No	Unique CFR part ID
	title	Unsigned integer	No	CFR title
	part_number	Unsigned integer	No	Concatenated together forms the CFR part number
	part_letter	1 character	Yes	
cfr_parts	id	Unsigned integer	No	Unique CFR part & year ID
	year	Unsigned integer	No	Publication year that part is from
	cfr_key_id	Unsigned integer	No	Reference to the cfr_key "id" column
	subject	Text	Yes	Subject of the entire part
	auth	Text	Yes	Authorization line of the entire part
	source	Text	Yes	Source line of the entire part
cfr_sections	id	Unsigned integer	No	Unique row ID
	part_id	Unsigned	No	Reference to the cfr_parts "id" column

[3] See http://www.gpo.gov/fdsys/bulkdata/CFR.

		integer		
	section	Text	No	Section number
	subject	Text	No	Subject of the section
	body	Text	No	Body of the section
	source	Text	Yes	Source line of the section

The first table created is the "cfr_key" table. This table exists so that each CFR part referenced in the entire dataset can be given a unique integer as an identification number. The "id" column from this table is used in the cfr_parts and parallel tables, and could also be used with the "register_rules" table.

The second table created is the "cfr_parts" table. This table exists to provide information that is applicable to an entire part of the CFR for a specific year. Typically, the "parallel" table should be used rather than the "auth" column of this table—unless there is a specific need for the historical information that the "parallel" table lacks. The "source" column most commonly refers to one or more publications in the Federal Register and applies to the entire part unless otherwise noted by a specific section.

The third table created is the "cfr_sections" table. This table exists to provide information on each section that is contained in the "cfr_parts" table above. The "section" column can be quite messy, but it typically consists of the part number, followed by a period, followed by the section number (which may sometimes contain non-numeric characters). The "body" column contains the full text of each section, with some caveats. Most notably, tables, graphics, and other non-textual data may not be present or usable. In addition, the text contains formatting tags used by the GPO that may not be useful for all applications. Finally, the "source" column, if provided, supersedes the part's source line to provide a source for the specific section.

Parallel Table

The "parallel" table is created by parsing, line by line, the Parallel Table of Authorities and Rules provided by the GPO[4]. This file is formatted as plain text, and was not designed for machine processing—indeed it contains a number of "bugs" that required manual patching to allow for machine processing. The table's schema is as follows:

Table	**Column**	**Type**	**Nullable**	**Description**
parallel	id	Unsigned integer	No	Unique row ID
	class	Enumeration	No	Type of entry (USC, Executive Order, etc.)
	auth_fail	Text	Yes	Text of authorization citation that could not be assigned a "usc_id"
	usc_id	Unsigned integer	Yes	Identification number from "usc_key" table that corresponds to the authorization citation
	usc_suffix	Text	Yes	Suffix to a USC citation (such as "note" or "et seq")
	cfr_part_id	Unsigned	Yes	Identification number from "cfr_key" table

[4] http://www.gpo.gov/help/parallel_table.txt

		integer		that corresponds to the target CFR part
	cfr_part_fail	Text	Yes	Text of target CFR part that could not be assigned a "cfr_part_id"

This table is one of the most important in the dataset, and also one of the tables most reliant on references to other tables. In rows where they are present, these references make use of the table much easier to perform automatically. This table consists of 80,620, of which 72,045 make full use of the "usc_id" and "cfr_part_id" reference columns. The "usc_id" column points to the USC section that the row references, and the "cfr_part_fail" column points to the CFR part that the row contains the authorization for. Where a reference could not automatically be created, the text that was attempted to be parsed is placed in the corresponding "fail" column. At this time, all rows that have a class that is not the USC will fail to have an authorization reference. This includes rows that reference an appendix of the USC for authorization.

Notably, the creation of this table involves some amount of "range expansion." For example, the authorization citation in the Parallel Table might read, "458aa--458gg." This is expanded into 42 separate sections (going as deep as "458ddd-2") by cross-referencing the "usc_key" table. This may present research design issues, because it is not necessarily the case that all sections within a range are intended to serve as authority for a regulation.

Federal Register Tables

The two Federal Register tables ("register_entries" and "register_rules") are generated from the daily XML files of the Federal Register provided by the GPO[5]. There are some challenges associated with processing these files, largely because the Federal Register is intended for reading by interested parties rather than machine processing. The Federal Register also consists of an enormous amount of information, some of which is very minor or irrelevant for most research purposes. The schema for the tables is as follows:

Table	**Column**	**Type**	**Nullable**	**Description**
register_entries	id	Unsigned integer	No	Unique Register publication ID
	year	Unsigned integer	No	Year of publication
	month	Unsigned integer	No	Month of publication
	day	Unsigned integer	No	Day of publication
	agency	Text	No	Agency responsible for the publication
	subagency	Text	No	Subagency or office responsible for the publication
	cfr	Text	No	Affected CFR part or parts
	subject	Text	No	Subject of the publication

[5] http://www.gpo.gov/fdsys/bulkdata/FR

	action	Text	Yes	Publication's action
	frdoc	Text	No	Register document number assigned by OFR
	proposed	Boolean	No	True if entry refers to a proposed rule
register_rule	id	Unsigned integer	No	Unique row ID
	entry_id	Unsigned integer	No	Reference to "register_entries" "id" column
	title	Text	Yes	CFR title that the rule section will be placed in
	part	Text	Yes	CFR part that the rule section will be placed in
	auth	Text	Yes	Nearest authorization line for the section
	amend_par	Text	Yes	Directive for the amendment in the section
	section_num	Text	Yes	Section number that will be assigned upon CFR placement
	section_subject	Text	Yes	Subject of the section
	section_body	Text	Yes	Body of the section

The Federal Register is published every government business day, and contains various publications submitted by federal agencies. Each publication generally deals with a single topic.

The first table created is the "register_entries" table. It contains information that relates to a single publication or "entry" in the Federal Register. Each publication that is tagged by the Office of the Federal Register as a "Rule" or "Proposed Rule" constitutes a single row. The "proposed" column can be used to differentiate between publications that are tagged as rules and those that are tagged as proposed rules. The "action" column should also be considered before using these tables for research purposes—there are many types of "actions" that an agency can take with regard to a rule (notice of proposed rulemaking, temporary final rule, interim rule, etc.).

The second table created is the "register_rule" table. This table contains one row for each section of regulatory text that is found within each entry (see above). Caution should be used with the "title," "part", "auth," and "amend_par" columns. It is not always the case that a section of the Register has internal information to populate these columns. In such a case, the columns will contain information from the nearest ancestor to the section (for example, a part or subpart), or, failing that, the nearest preceding section. This approach is used because it generally results in reliable results—the Register often organizes information using units larger than the section, with the intent that attributes of the containing unit apply to its sections.

Statutes Table

The statutes table is generated from text files of public laws provided by the GPO[6]. Each public law is broken up by section, with each row in the table constituting one section. Although the separation of a statute into its sections is generally reliable, there are likely cases where it has failed in one or more way. The schema for the table is as follows:

Table	Column	Type	Nullable	Description
statutes	id	Unsigned integer	No	Unique row ID
	congress	Unsigned integer	No	Congress of statute passage
	number	Unsigned integer	No	Public Law number of statute
	sec_number	15 characters	No	Section number in statute
	sec_subject	Text	No	Subject of the section
	sec_body	Text	No	Body of the section

The "sec_number" column is usually a number, but occasionally contains letters as a suffix. Note that, despite the name of this table, it does not include all statutes—private laws are excluded. The table *does* include public laws that are created by joint resolutions having the force of law, which are not always parsed correctly due to their irregular structure.

Table III Table

The Table III table ("table3") is generated from HTML files provided by the OLRC[7]. These files are parsed fairly easily, despite not being designed for machine processing. Because of the cumulative nature of Table III, historical versions are not necessary. The schema for the table is as follows:

Table	Column	Type	Nullable	Description
table3	id	Unsigned integer	No	Unique row ID
	volume	Unsigned integer	No	Volume that the statute section was published in
	act_number	20 characters	No	Number of the statute within the volume
	act_section	100 characters	No	Number of the section within the statute
	statute_pg	50 characters	No	Page of the volume that the section starts on
	usc_status	20 characters	No	Status of codification (eliminated, repealed, etc.)
	prefix	10 characters	No	Codification prefix (note preceding, see, paragraph, etc.) to the codification target section
	usc_id	Unsigned integer	Yes	Identification number from "usc_key" table that corresponds to the codification target section

[6] http://www.gpo.gov/fdsys/browse/collection.action?collectionCode=PLAW
[7] http://uscodebeta.house.gov/table3/table3statutesatlargevolumes.htm

	suffix	10 characters	No	Codification suffix (notes, notes, etc.) to the codification target section
	usc_fail	Text	Yes	Text of target CFR part that could not be assigned a "cfr_part_id"

Each row in the table corresponds to a single row of Table III, after applying range expansion rules. Range expansion relies on the "usc_keys" table in a manner similar to the range expansion used for the "parallel" table, although the potential research design issues are less because ranges tend to be much smaller and carefully drawn.

The "volume" column uses two numbering systems. Statutes passed in or after the 85th Congress are contained in a volume that matches the Congress number (from 85 to 112). Statutes passed prior to the 85th Congress are contained in a volume that matches the year of passage (from 1789 to 1956). The "act_section" column can sometimes be complex and difficult to parse, containing multiple sections, partial sections, and ranges consisting thereof.

References to the USC are created as in the "parallel" table. This table consists of 288,871 rows, of which 213,035 make use of the "usc_id" reference column. The "usc_id" column points to the USC section that the row references. Where a reference could not automatically be created, the text that was attempted to be parsed is placed in the "usc_fail" column. It should be noted that some reference failures are normal—the "usc_key" table only contains sections in the USC after 1994, while Table III contains USC sections since the creation of the code system in 1926. The "usc_status" column can be used to locate these rows.

U.S. Code Tables

The USC tables ("usc" and "usc_key") are generated from HTML files provided by the GPO and originally created by the OLRC[8]. These files contain some markers that make machine processing easier, but are still designed primary for human use.

Table	Column	Type	Nullable	Description
usc	id	Unsigned integer	No	Unique row ID
	year	Unsigned integer	No	Publication year that section is from
	usc_key_id	Unsigned integer	No	Reference to the "usc_key" "id" column
	subject	Unsigned integer	Yes	Subject of the section
	body	Text	Yes	Body of the section
	citation	Text	Yes	Statutory sources of the section
	notes	Text	Yes	Notes of the section, both editorial and statutory
usc_key	id	Unsigned integer	No	Unique USC section ID

[8] http://www.gpo.gov/fdsys/browse/collectionUScode.action?collectionCode=USCODE

	title	Unsigned integer	No	USC title
	number	Text	No	Concatenated together form the USC section number
	letter	Text	Yes	
	number2	Text	Yes	
	letter2	Text	Yes	
	version	Text	Yes	

The first table that is created is the "usc_key" table. This table, similar to the "cfr_key" table, contains a single row for each discrete title and section number found in the USC. This allows each USC section to be referenced throughout the entire dataset using a unique integer as an identification number. The "id" column from this table is used in the "usc" and "table3" tables.

The second table that is created is the "usc" table. This table contains a single row for every section of the USC for each year that is available. Most sections that are designated as "repealed," "omitted," or "reserved" are automatically skipped.

Discussion of Use Scenarios

This dataset was initially created with the intent to be used as a tool for measuring the impact of Congressional enactments on administrative regulations. The dataset remains well suited for that purpose, although further development may be necessary to achieve a satisfactory level of confidence in results. The general process for tracing a new law to the regulations that it has affected would be something like this:

- Locate the statute (or one or more sections thereof) in "table3" and obtain the resulting "usc_ids"
- Look each "usc_id" up in "parallel" and obtain the resulting "cfr_part_id"
- Use each "cfr_part_id" with "cfr_parts" and "cfr_sections" to obtain the full text of each affected section for the years immediately prior to and succeeding the statute's enactment

A significantly more precise way of locating specific sections would be to further develop the Federal Register tables to add cross-references to "usc_id" and "cfr_part_id." It could then be reasonably inferred that publications in the Federal Register *adding* the "usc_id" as an authority for a specific CFR part are also making regulatory changes in the same publication in response to the new law. By examining the sections contained in the publication, it may even be possible to determine individual sections that the new law has affected.

The ability to link a statutory section to sections of the CFR would be valuable for investigating a number of political science questions. The bulk of day-to-day government operation is guided by regulations, not statutes, but these regulations are typically adopted by Presidential appointees—not the democratically elected members of Congress. It might be valuable to compare the public policy goals of a statutory section (and that statute's sponsors in Congress) to the impact of the resulting regulations. Also interesting would be an examination of the "unintended consequences" of delegating rule making authority to an agency. This is particularly true for the broad grants of general authority that are practically boilerplate in the creation of new federal agencies.

Another potential use scenario is an investigation of the process of classification and codification by the OLRC. By comparing the "statutes" table with the "usc" table, it would be possible—albeit somewhat cumbersome—to identify specific editorial decisions of the OLRC. However, this wouldn't necessarily be a valuable use of time considering that the OLRC is only authorized to make purely technical changes, and is not generally accused of straying from this mandate.

Another interesting use would be to compare the text of proposed rules to the text of final rules using the two Federal Register tables. In theory, the process of promulgating a new administrative regulation is guided by public comment and stakeholder participation. The extent to which text changes between versions of the same rule proposal may be indicative of the receptiveness of a particular agency to input, the policy goals of a particular presidential administration, or the persuasiveness of stakeholder lobbies.

Finally, it may be interesting to compare the historical development of administrative regulations. Administrative regulations evolved as a response to the need for flexible and detailed regulation in areas that Congress was ill-equipped to handle. The three CFR tables and two Federal Register tables might be used to measure the frequency of major updates to regulations, correlations between new presidential administrations and regulatory activity, the speed with which regulations were updated in response to major events or scientific discoveries, or the extent to which specific agencies or administrations contributed to the regulatory scheme.

One final thing worth noting is the importance of using effective SQL JOIN operations with this dataset. Although it is possible to use the dataset without them, it is intended that tables be joined together using foreign keys when necessary.

Conclusion

This dataset remains in its infancy. Particularly with regard to the Federal Register, there is a significant amount of additional pre-processing and parsing that should be completed. Potential uses of machine learning technology should also be considered, especially for text comparison and similarity analysis. Nevertheless, the current dataset should provide—with some manual intervention—the tools necessary to partially trace the regulatory impact of a statute.