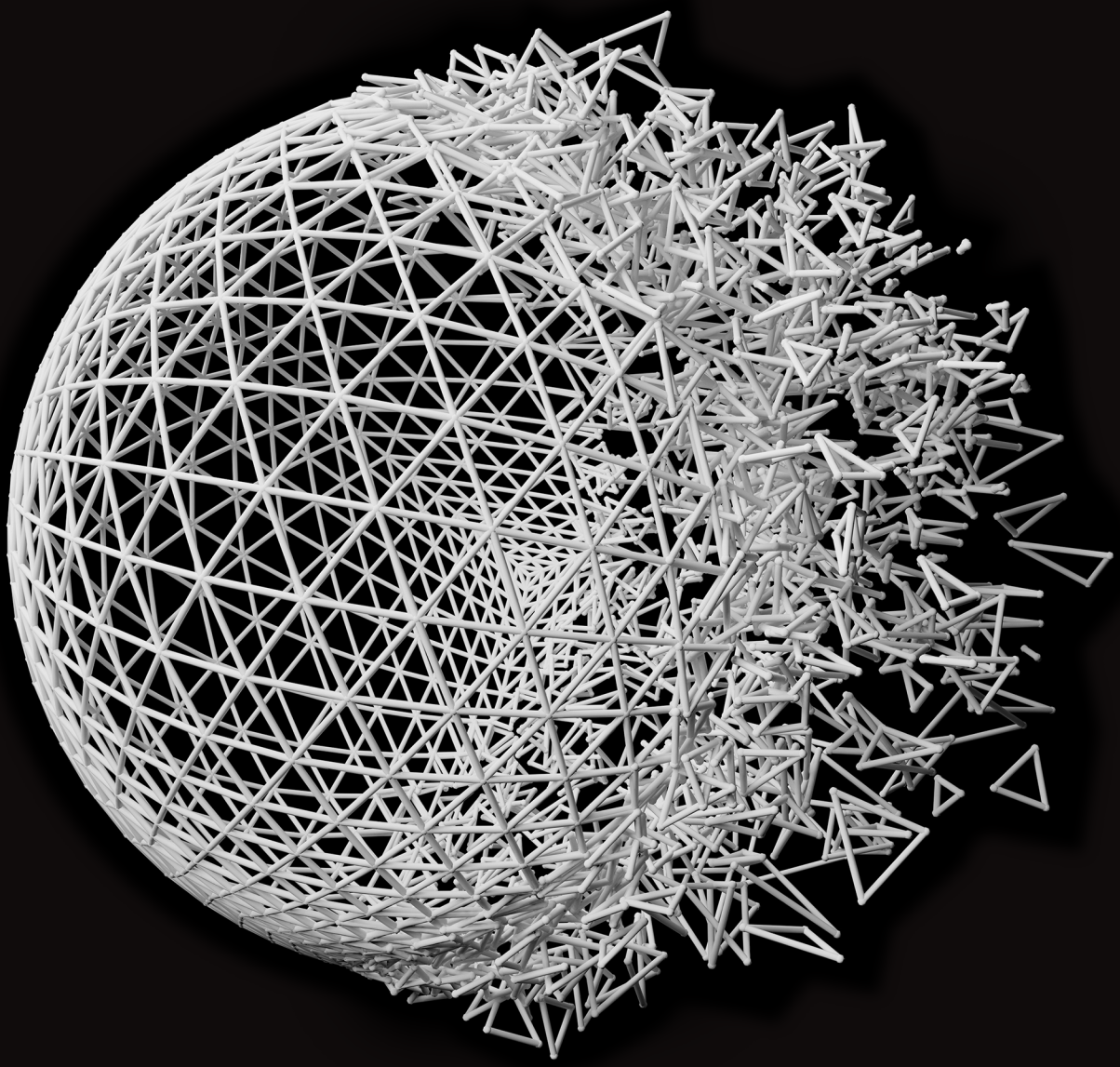
Deloitte.



Using advanced analytics to drive regulatory reform

Understanding presidential orders on regulation reform

Deloitte Center *for* Government Insights

About the authors

Daniel Byler is the technical lead on the Research and Insights Data Science Team for Deloitte Services LP. Within this role, he is responsible for creating research projects and software tools that produce actionable quantitative insights. He and his team span the full gamut of technical methods, ranging from neural networks to classic econometrics. He also has experience serving clients in the federal government to help them take action based on insights derived from data. Daniel received an MS in Statistics from Texas A&M University and a BA in Economics from The College of William and Mary.

Beth Flores is a senior leader for Deloitte Consulting LLP in the Federal Strategy practice with expertise in strategy, organizational change, policy development, regulatory reform, customer experience, and human-centered design. She focuses on helping mission-driven organizations adapt to the future and achieve business results through the power of data and design. Prior to joining Deloitte, she was a policy advisor in the Department of Defense and led an organizational transformation initiative focused on leadership and innovation. She received an MBA and a MS in Foreign Service from Georgetown University and a BA in American Studies from Yale University.

Jason Lewris is a data scientist in Research and Insights with Deloitte Services LP. He primarily focuses on developing advanced analytics solutions and using modern data science techniques to distill meaningful insights for research engagements. Prior to joining Deloitte, Jason received a Master of Data Science from the University of Virginia, where he specialized in text analytics. His current research focus is using neural networks to model complex relationships buried within text.

About the Deloitte Center for Government Insights

The Deloitte Center for Government Insights shares inspiring stories of government innovation, looking at what's behind the adoption of new technologies and management practices. We produce cutting-edge research that guides public officials without burying them in jargon and minutiae, crystalizing essential insights in an easy-to-absorb format. Through research, forums, and immersive workshops, our goal is to provide public officials, policy professionals, and members of the media with fresh insights that advance an understanding of what is possible in government transformation.

Using advanced analytics to drive regulatory reform

Advanced text analytics can help agency leaders look more specifically into regulations with a higher potential opportunity for improving efficiencies—without reducing protections.

Imagine you are a federal leader who is now tasked with finding regulations, which may contain duplications, contradictions, or simply be outdated. In light of Executive Order 13771 "Reducing Regulation and Controlling Regulatory Costs," this challenging task has taken on a new urgency. With 186,000+ pages of the Code of Federal Regulations (CFR), how would you even begin?

For some, achieving regulatory reform while still executing their mission can seem impossible. However, due in part to advances in text analytics, regulatory reform is more achievable than ever. In this study, we show how advanced analytics can support regulation reform by identifying three kinds of regulations:

1. Regulations written more than 20 years ago
2. CFR sections that are exact copies of other regulations in the CFR
3. Regulations that touch on the same topic in multiple (possibly contradictory) ways

The last time the average CFR section was edited was nearly 20 years ago

To identify regulations that are candidates for reform, Deloitte analyzed all 217,714[1] sections of the 2017 CFR.[2] Our results suggest that analytics can provide practical assistance in identifying targets of opportunity for regulatory reform. Throughout this study, unless otherwise noted, all results are from Deloitte's analysis of the CFR.

Key findings

- It has been 19.4 years[3] since the average section of the CFR has been edited or updated in any way. Twelve percent of all CFR sections haven't been updated since the 1970s or before. Sixty–seven percent of all CFR sections currently on the books have never been edited since they were originally created. This makes them likely candidates for having become outdated and/or superseded by other more modern regulations.
- 17,800 sections are "extremely close matches" that differ only by a few words from other sections in the CFR.
- There is substantial overlap in what these rules cover with 47 percent of all sections falling into conceptual clusters that have a gap between the oldest and youngest section of 41 years or more.

What is the Code of Federal Regulations?

The Code of Federal Regulations (CFR) combines all of the federal regulations published in the Federal Register into a single document that acts as the official record of US regulation. The CFR[4] was first published in 1938. An annual "final" copy of the CFR is published by the Office of the Federal Register in addition to interim quarterly updates. The e-CFR represents the most up–to–date information on rules that have been translated into specific regulations. Organizationally, the CFR is divided into **titles** (e.g., agriculture) which are made up of **parts** (e.g., watershed projects) and then tactical individual **sections** (e.g., defining eligible watershed projects). Our analysis is conducted mostly at the section level because it provides a very fine grain level of detail.

Many federal regulations are quite old

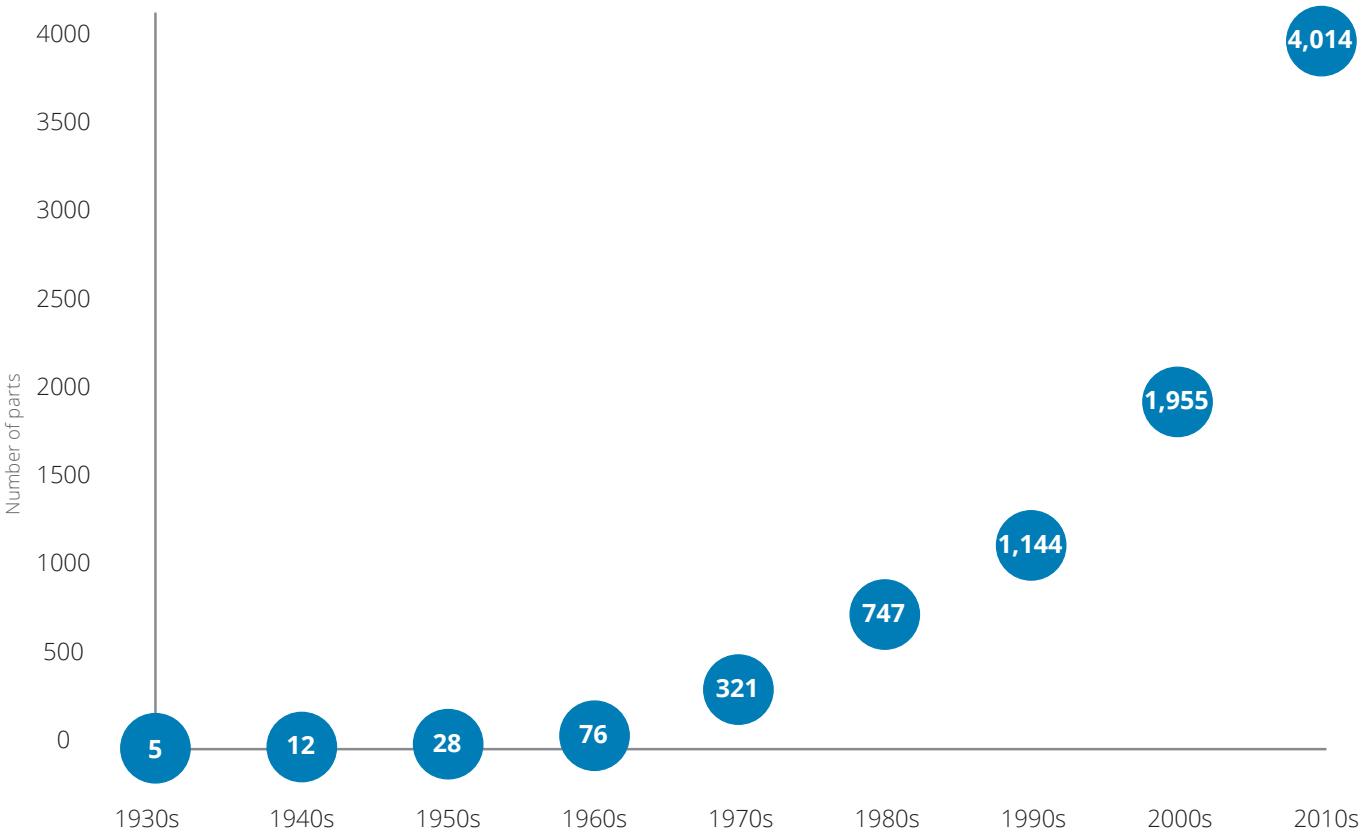
The CFR is divided into broad parts that pertain to a single subject (like railroads). These parts are often made up of many individual sections that build toward the part's overall purpose. In total, there are 8,310[i] of these parts[ii] in the CFR. In most cases, individual sections within these parts are updated as times change. However, a total of 2,878 parts have underpinning sections[5] which have not been updated since the year they were implemented. It might be expected that some parts implemented in the 2010s and even the 2000s would not yet have required updates.

However, there are 1,189 full parts (covering 6.1 percent of all CFR sections[6]) that have not been updated since the 1980s or before. These unchanged parts span almost every title in the CFR with 44 out of 49 titles having at least 1 part that has not been edited[iii] since the 1980s. Notably, there are 45 parts totaling 282 sections which have not been updated since the 1950s or before. The chart below shows the number of parts that have never been changed by decade.

The topics covered by these parts vary widely. As an example, three parts from the 1930s that have not been edited since their creation[7] cover:

- The rules of procedure of the National Railroad Adjustment Board
- Standards for Navajo, Pueblo, and turquoise products
- The use of government marks of genuineness for Alaskan Indian and Alaskan Eskimo handmade products

Decade in which CFR part was last updated



Source: Deloitte Analysis of the CFR

i. We used part numbers presented as they appeared in the part headers tag of the CFR along with the title number to construct part representations for analysis.
ii. A small number of CFR parts did not have date information that was easily machine readable and so were excluded from this analysis.
iii. We did not include reserved titles in our analysis.

Out of date: Many individual sections of the CFR have not been updated

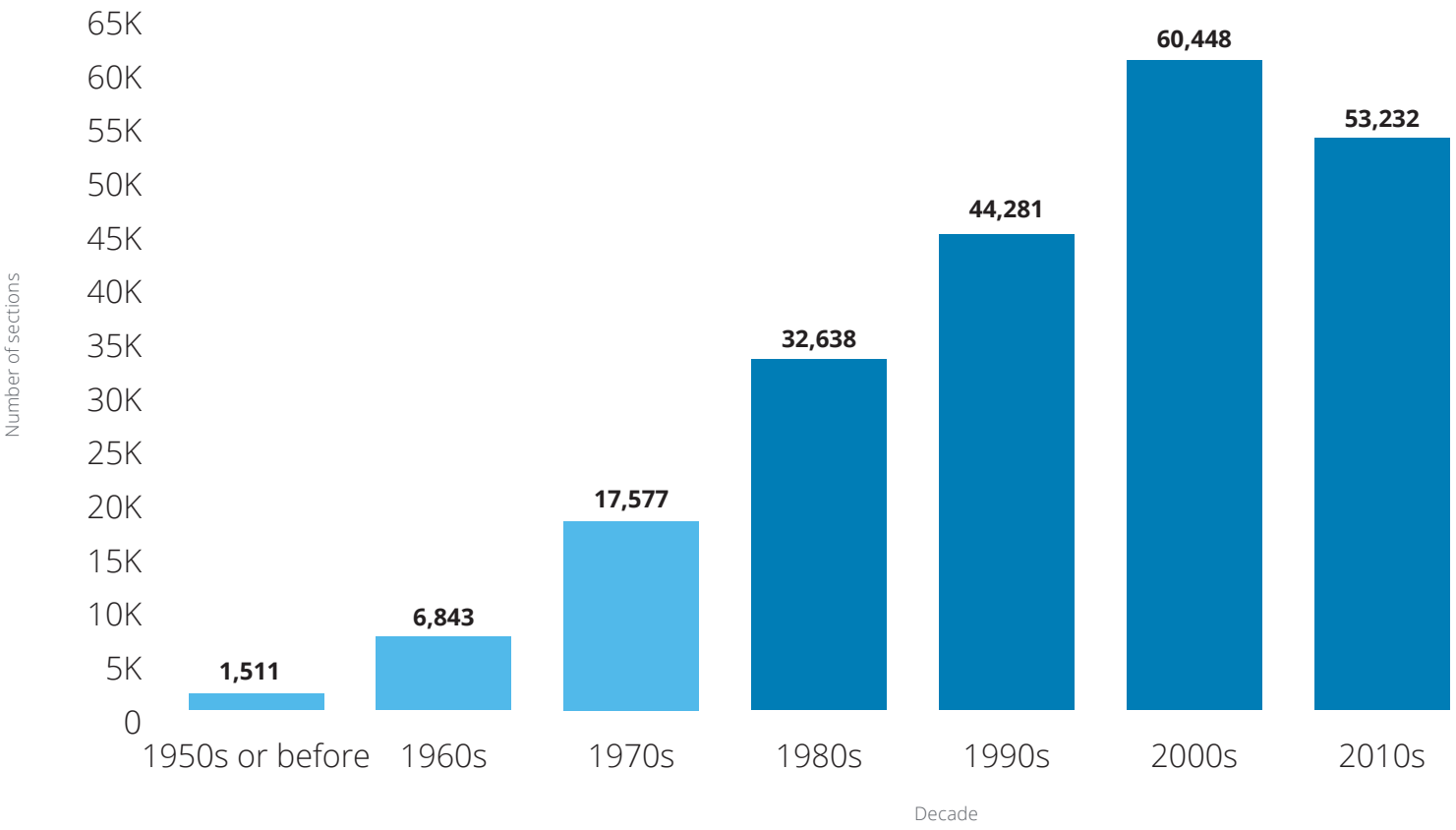
The last time the average CFR section was edited was 19.4 years ago. Age is not a proxy for irrelevance, but if a regulation hasn't been updated for decades, federal leaders may want to examine whether it is still having its intended effect. For example, one of the oldest sections in the CFR (from 1905) pertains to a method of log transportation (log running) in a particular Minnesota river which has not occurred since 1937.[8]

CFR Title 33 Part 207 Section 380 reads in part:

"(a) Parties wishing to run logs on Red Lake River must provide storage booms near the head of the river to take care of said logs. (b) No one will be permitted to turn into the river at any time more logs than he can receive at his storage boom. (c) Tows arriving at the head of the river shall turn their logs into the river successively in the order of their arrival, and such logs shall be at once driven to the owner's storage boom ..." [full text of regulation is here]

This 19.4 year average age of regulations that have not been updated is partially driven by regulations that date back to the earlier part of the last century. In total, 25,931 sections (12 percent of the total) were last modified in the 1970s or before.[9] While potentially appealing, it would be unwise for federal leaders to eliminate all regulations older than a certain age as some portion of them are still relevant. However, the age of these sections does present an opportunity for examination and potential improvement.

Decade in which section was last edited



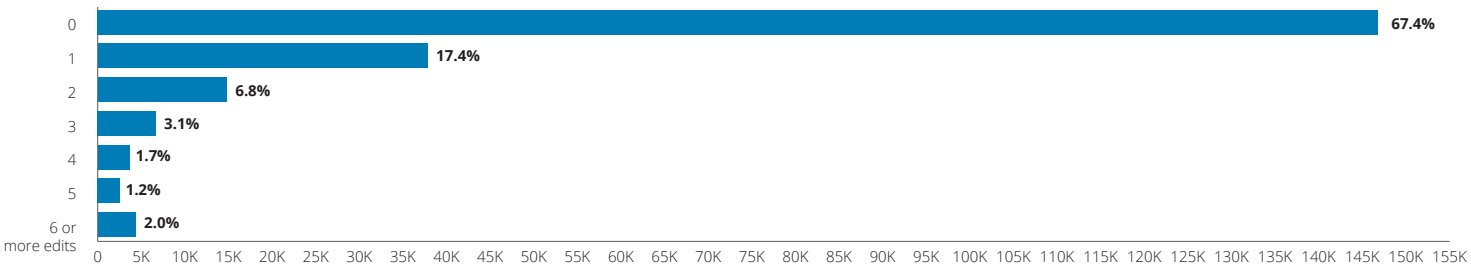
Source: Deloitte Analysis of the CFR

Most regulations in the CFR have never been updated since their passage

After being created and put into the CFR, each section of the CFR can be updated in future revisions to stay up to date. Unfortunately, this is rarely done in practice. Sixty–seven percent of all sections have not been updated since they were initially passed. An additional 17 percent have been edited only once. This total doesn't include deletions but, given the nearly uniform and rapid growth[10] in the size of the CFR and the consistent use of terms which indicate forcing actions like shall and must,[11] we do not believe including deletions would substantively change this finding.

This shows that, to the extent that agencies have chosen to pursue regulatory reform, it has been by deleting sections rather than revising them. This gives hope to the idea that these initial efforts at regulatory reform could bear good fruit by revising thousands of sections of the CFR that have never been thoughtfully revisited.

Number of times a section is ever revised



Source: Deloitte Analysis of the CFR

Massive duplication of regulations

Within the CFR, 17,800 sections have extremely similar text as another section. Most of these matches are exact matches where the words are precisely the same in both sections (often legal definitions or standard penalties/definitions). Other "very close" matches have exactly the same main words but differ in the order of the words or in other small ways such as the number of spaces between words, missing an inconsequential word like "the," and, most interestingly, differ only in the numbers they refer to (changing how many milligrams are referenced, etc.).

Agency leaders can examine these sections to see if they present opportunities to streamline their portion of the CFR. In addition to reducing the length of the code, it would improve consistency if all sections were pointing to a single common definition with a consistent location. More importantly, streamlining these definitions into common sections could help regulators avoid unintentionally setting different standards simply due to referencing differing newer and older parts of the CFR.

While these may be ripe targets for reform, human intervention and analysis is still needed. In some cases these repeated sections are common legal requirements that genuinely may need to be repeated.

What is text analytics?

Text analytics involves the use of computers/software algorithms to process large volumes of text–based information with the intent of extracting quality insights. It is widely used both within and outside of the federal government to analyze documents (like the CFR), which are too voluminous for humans to efficiently handle. It makes it possible to detect trends and patterns that, due to the size of the underlying text, humans would generally have trouble identifying.

Within the CFR, nearly 18,000 sections have extremely similar text as another section.

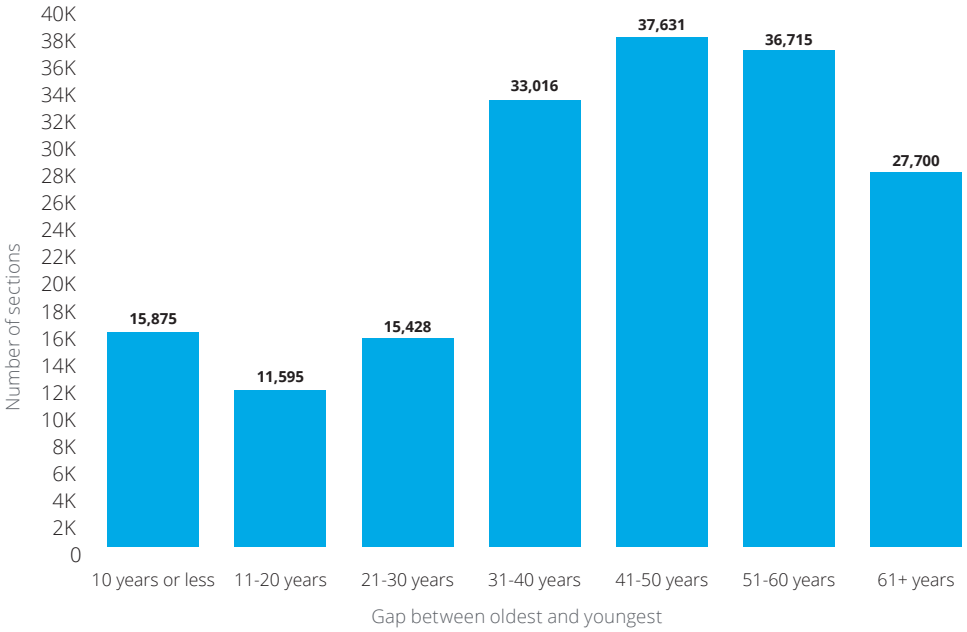
Regulations with overlapping purpose

Federal leaders are often faced with organizations that are made up of components that each individually achieve their aims but fail to coordinate effectively across organizational silos. Federal leaders can use text analytics to identify parts of the CFR that govern similar topics and then use those findings to encourage collaboration and a unified front on key regulatory topics within their own agency.

This process can be scaled up for issues that cut across multiple agencies. These areas of multiagency overlap are crucial for two reasons. First, each agency's expert staff may not know about competing efforts in the same area at another agency. Second, businesses looking to comply with a regulation can be quickly frustrated by incongruent definitions and practices from multiple competing agencies. Combined, this creates a recipe for persistent frustration that agencies may be unaware of. The analytic methods described here can help identify these areas of overlap and encourage rationalization of these regulatory structures.

Beyond spanning across organizational silos, these clusters can be used to identify topics that have been re-litigated over time. In particular, there are 102,046 sections that come from clusters where the oldest section is at least 41 years older than the most recent section. This is particularly troubling as these particular clusters were developed within the walls of each agency—minimizing the chance these sorts of large time gaps would be present. This indicates that the same topic may have been touched on again and again, potentially without due consideration for the removal or revision of previous regulations. The large number of these sections should encourage leaders who are seeking to rationalize their regulatory structure.[13]

Difference between youngest and oldest section within a cluster



Source: Deloitte Analysis of the CFR

Using neural networks to find common clusters

In addition to classic text analytics techniques, we used neural networks (commonly called artificial intelligence) to boost the quality of our analysis. Artificial intelligence allows for computers to understand how concepts in a given piece of text relate to each other—for example, that boat and ship are similar. This is a very substantial improvement over prior methods, which were based solely on keyword style list matching that would have missed the deeper meaning of terms.

We tested this particular application of artificial intelligence and text analytics on the CFR to see whether it yielded intelligible results.[12] Our clusters were validated not only by statistical tests, but also by regulatory specialists. In one example, our clusters flagged medical devices as an opportunity area for improvement without any help or guidance from a specialist. Simultaneously and independently of the tool, we discovered our regulatory specialists had also identified this as a key area of concern. The difference is that the clusters already "knew" what specific sections of the code should be examined in more detail by these specialists. This gives us confidence that this approach can be used by federal leaders to map broad initiatives to tactical sections that can support regulatory reform.

Five steps agency leaders should consider to help identify CFR sections that can be eliminated

Every federal agency differs in the volume and complexity of the regulations they have produced. Some agencies have smaller and more manageable numbers of sections while others have thousands of sections just within a sub part of the agency. Regardless of the size of the regulatory structure, some common elements can be used by all agencies to help improve the quality of the regulations that they do have on the books.

1. Using text analytics to group sections across institutional silos

Federal leaders can use methods similar to what we describe in this study to group similar regulations into manageable "haystacks" for their analysts to search through. Federal executives can search for redundancy both within their organization and across the rest of the federal government. This frees up valuable time that agencies can spend on improving regulations.

2. Using Journey Mapping and input from industry to gain insight into key pain points

Advanced analytics is key to making the task of regulatory reform faster and more effective. However, it doesn't replace the traditional task of finding out what the biggest-felt pain points are for the businesses and citizens that are impacted by them. For additional information on constructing effective regulatory reform, readers should see our prior work, "How To Implement President Trump's One In, Two Out Regulation Initiative."[14]

3. Prioritizing sections that are out of date for early review during regulatory reform

Old parts of the CFR present an opportunity to identify regulations that may need updating or be genuinely irrelevant in today's environment. For agencies with a voluminous amount of old regulations, text analytics can be used to focus in on only old regulations that are key topics of interest.

4. Streamlining sections that are precise, or nearly precise, duplicates of other sections

Some sections that are precisely duplicative of each other may represent an opportunity for a streamlining of definitions or point to unnecessary redundancy. These can be easy wins for leaders looking to make their code clearer, and therefore less confusing, to those they regulate.

5. Implementing an ongoing review process to surface out-of-date regulations

It is tempting to treat regulatory reform as a one-off that can be done and forgotten. Unfortunately, as we have shown, 67 percent of all sections of the CFR have never been updated after publication. This contributes to the current unmanageable size of the CFR and likelihood of needless burdens on businesses. Establishing a clear process to continuously evaluate existing regulations can prevent this build-up from continuing into the future.

Especially as the rate of technological change accelerates, the time it takes for a genuinely useful protection to transition into a burdensome antiquity will likely shorten. This is expected to only increase pressure on regulators to institute new and meaningful protections while weeding out the old.

A note of caution

Federal regulators should have hope from the results in this study that there are opportunities to reduce burdens while maintaining protection. However, it would be a mistake to capriciously remove regulations solely based on their age or other metrics. The evaluation of every section and every rule must be human driven and systematic.

Conclusion

Federal agency leaders can feel frustrated by the massive task of regulatory reform. This discomfort can lead to a focus on examining only those regulations that are familiar. Those are, after all, the regulations that naturally surface in discussions. This strategy may be convenient but makes the classic mistake of "looking for your keys underneath a streetlamp because it's easy to look there." If agency executives only examine recent regulations that they are familiar with, they may miss an opportunity for reform that could help eliminate unnecessary burden. Advanced analytics, as outlined in this study, can help agency leaders both expand their view beyond what they and their staff are familiar with and look more specifically into sets of regulations with a higher potential opportunity for improving efficiencies without reducing protections.

Endnotes

1. We excluded "reserved" sections from all counts as they do not contain any text content.
2. As pulled on June 20, 2017. Not all regulations issued during the last year of the Obama administration have yet been codified into the CFR and so it will likely be meaningfully larger by the end of the year.
3. There were 1,180 sections where we could not extract a citation or a citation year didn't exist.
4. Code of Federal Regulations. Retrieved from https://en.wikipedia.org/wiki/Code_of_Federal_Regulations.
5. It is possible some of these parts once had more sections but were deleted. Deleted sections are not well tracked in a computer readable format and so are difficult to analyze. However, given the immense growth rate of the CFR, it is unlikely many of these parts had dramatic deletions that represent an "invisible edit."
6. These older parts tend to be shorter than the other more modern parts. Part of this may be that portions of these parts were deleted over time, but it is difficult to formally identify whether a portion was deleted or simply never made in the first place.
7. There are two additional parts that had edits that came only from the 1930s. However, they didn't make the criteria for this chart because their surviving sections had been edited at least one time after their initial publication. For example, a part may have been in publication in 1937 but then edited in 1938 in which case it wouldn't make it into the count of "never having been edited."
8. Forest History Center, Log Drives. Retrieved from http://sites.mnhs.org/historic-sites/forest-history-center/log-drives.
9. The totals below sum to 215,627 sections, which is less than 217,886 total sections that exist in the CFR. This is because there are a small portion of sections that do not have a readily machine readable citation mentioning when they were last edited.
10. GW Regulatory Studies Center, Total Pages, Code of Federal Regulations (1975-2015). Retrieved from https://regulatorystudies.columbian.gwu.edu/sites/regulatorystudies.columbian.gwu.edu/files/downloads/Pages_CFR_0.JPG.
11. Al-Ubaydli, O., and McLaughlin, P. A. (2017) RegData: A numerical database on industry-specific regulations for all United States industries and federal regulations, 1997–2012. Regulation & Governance, 11: 109–123.
12. Through a combination of neural networks (also known as AI), information extraction, and multiple clustering techniques, we grouped sections in the CFR together. Crucially, unlike keyword-based systems, this system allows leaders and analysts to analyze clusters of regulations that share a common topic even if the keywords are not precisely the same (such as "boat" and "fishing ship").
13. Many sections are excluded from the chart below because they either lack a citation year or, more commonly, fall into a cluster with one member. This is not a failing of the clustering method but rather it flags that a section is an extreme outlier and should be examined individually to see what drives its oddball status. Additionally, if desired, "oddball clusters" can be forcibly assigned to their nearest neighboring clusters or can be "vacuumed" into clusters made entirely of oddballs. We felt both of these methods would inflate our results for the purpose of this study but could be performed depending on the specific needs of an agency.
14. Kohli, J., Chew, W., (2017) How to implement regulatory reform. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-public-sector-regulatory-reform.pdf.

Acknowledgments

The authors thank Kenton Andersen, Sushmita Biswal, Bruce Chew, Lucy Ferguson, Shanti Jha, Michael Petrillo, Nell Todd, and many others who contributed their ideas and insights to this project.

Contacts

Bill Eggers
Executive Director, Deloitte
Center for Government Insights
Deloitte Services LLP
+ 571 882 6585
weggers@deloitte.com

Daniel Byler
Technical Lead, Research and
Insights Data Science Team
Deloitte Services LP
+1 571 814 7166
dbyler@deloitte.com

Jitinder Kohli
Government Performance Leader
Monitor Deloitte
+1 571 858 0821
jkohli@deloitte.com

Deloitte.