REDUCING ADMINISTRATIVE INCOMPREHENSIBILITY WITH DATA TOOLS AND STANDARDIZATION

Oliver Sherouse
Policy Analytics Lead, Program for Economic Research on Regulation, Mercatus Center at George Mason University

House Committee on Oversight and Government Reform, Subcommittee on Intergovernmental Affairs
Regulatory Divergence: Failure of the Administrative State

July 18, 2018

Chairman Palmer, Ranking Member Raskin, and members of the Subcommittee on Intergovernmental Affairs:

Thank you for the chance to speak to you today about the important and often overlooked problem of duplicative regulations and regulatory standards. My name is Oliver Sherouse and I am the policy analytics lead for the Program for Economic Research on Regulation at the Mercatus Center, a 501(c)(3) academic research center at George Mason University.

My testimony today will focus on one cause of regulatory duplication: the incomprehensible scale of the administrative state. I will also present two ways my colleagues and I are working to reduce that problem: first, through the application of text analysis and machine learning in our QuantGov project; and second, by developing an open, machine-readable, and data-first standard rulemaking format called XRRL.

THE INCOMPREHENSIBILITY OF THE ADMINISTRATIVE STATE

Since "policy analytics" is not a very common phrase, I will explain what it is that I do more simply: I teach computers to read policy documents, especially regulation. We have to use computers because the administrative state has grown to an incomprehensible size. I mean that quite literally: there are simply too many rules for any one person to understand, whether that person is trying to follow those rules or write new ones. So using text analysis and machine learning, my colleagues and I have created a dataset called RegData to quantify how much regulation there is, who writes it, and whom it affects.

RegData tells us that today there are more than 103 million words in the *Code of Federal Regulations* (CFR), including 1.08 million individual regulatory restrictions—words and phrases such as *shall* and *must* that indicate a particular mandated or prohibited activity.[1] To put that number in context, if you were to read the CFR as your full-time job, at 250 words a minute for 40 hours a week, it would take you three years, 111 days, and a bit over 5 hours.

---

[1] Patrick A. McLaughlin and Oliver Sherouse, RegData US 3.0 Daily (dataset), QuantGov, Mercatus Center at George Mason University, 2018.

By the time you had finished, of course, you would need to immediately start figuring out what had been added in the interim. That's no easy task, since according to RegData, from 1970 to 2017 the CFR increased by an average of more than 1.4 million words and 14,000 regulatory restrictions every year.[2]

HOW QUANTGOV DATA TOOLS CAN HELP REDUCE INCOMPREHENSIBILITY

While reading, let alone understanding, the entire CFR is impossible, data tools like those we have produced for the QuantGov project at the Mercatus Center can help us begin to make better sense of the administrative state.

RegData, in fact, does more than count total words and restrictions. It attributes them to the individual agencies and departments that create those words and restrictions, and it predicts which industries will be affected by them. All of our data is freely available, and our website now features a daily updated interactive tracker with which users can break down federal regulation by industry and by agency.

We can use the same kind of text analysis to understand regulation currently being developed. To create our RegPulse dataset, our system examines rules as they are published in the *Federal Register* and, as with RegData, quantifies those rules, tracks the agencies promulgating them, and predicts which industries are likely to be affected by them. And as with RegData, we have built a daily updated interactive tool that allows users to see which industries have more or fewer relevant rules coming into effect over the next several years, and what those rules are.

With QuantGov we are producing not only these kinds of data, but also these kinds of interactive tools for states, for other countries, and for a broader spectrum of policy documents. The software we use to produce QuantGov is also open source and freely available for anyone to use, modify, and build on.

XRRL: RULEMAKING AS DATA

A more comprehensive understanding of the large mass of federal regulation, however, could be achieved by going one level deeper and reexamining the medium by which regulations are made. The current regulatory process is made for paper: paper rules and analyses published in a paper *Federal Register* and compiled into a paper *Code of Federal Regulations*. While there are now electronic versions of these documents, they essentially mimic the paper-based system in use since the Administrative Procedure Act of 1946.

Seventy years later, it is time for an upgrade. A modern approach to rulemaking should insist on the use of an open, machine-readable, and data-first standard format for regulatory documents. A standard format could liberate the information about whom regulations will affect and how they will be affected—information that is currently trapped in dense prose—and transform it into discoverable, machine-readable data.

That data can be used by Congress to ensure effective oversight. It can be used by regulators to avoid duplication within or across federal agencies and potentially even across jurisdictions. A modern regulatory standard could also facilitate the review of regulatory programs so that those that are broken can be fixed and those that are successful can be recognized. And it can be used by businesses to ensure that they know what the law is and what they need to do to follow it.

---

[2] Patrick A. McLaughlin and Oliver Sherouse, RegData US 3.1 Annual (dataset), QuantGov, Mercatus Center at George Mason University, 2018.

My colleagues and I are currently developing such a standard, the eXtensible Regulatory Reporting Language, or XRRL. Our goal with this project is to build an open and nonproprietary standard incorporating insights from the academy, government, and industry that can be adapted to any level of government.

CONCLUSION

Duplication in regulation is a side effect of an administrative state grown too large to manage effectively. Tools like the ones we have built with QuantGov are a step toward making an incomprehensible collection of rules somewhat less so, and we will continue to produce them. But the implementation of an open, data-first standard format such as XRRL for rulemaking would be an even more powerful way to render the administrative state more manageable, while also providing benefits to both those writing rules and those subject to them.

I thank you again for the opportunity to testify, and I look forward to answering your questions.

Sincerely,

Oliver Sherouse

Policy Analytics Lead, Program for Economic Research on Regulation
Mercatus Center at George Mason University