RegData U.K. Technical Documentation

QuantGov

December 10, 2021

1 Purpose

Statutes are the primary tool for addressing public policy challenges, but empirical analysis of the actual effects of statutes has historically been hampered by a paucity of data and format of documents. RegData products are designed to help solve this problem- facilitating third-party usage and independent research.

We describe RegData U.K. 1.0 as the first iteration of the United Kingdom Legal dataset in the RegData Projects. The dataset organizes United Kingdom statutes, enacted and revised, into "Years in effect." In this context, "Years in effect" means that the file has metadata stating that it is in enacted by the first day of the new calendar year and does not have metadata stating that it has been repealed. The dataset provides the following data series: the total number of restrictions, including the types of restrictive terms (*shall, must, may not, required, prohibited*), word counts, and the complexity of the text (sentence length, conditionals, Flesh reading ease, Shannon entropy).

2 Content

RegData U.K. 1.0 quantifies the United Kingdom legislation in force between 1992 and 2020. This excludes acts lacking the metadata necessary to determine whether they are in force, dating back 200 years or so. Additionally, it excludes or includes any legislation that is missing a date in effect or a date expired, respectively.

3   Technical Notes

- RegData U.K. 1.0 is generated from source files collected from the Legislation.gov.uk bulk download website. The files are stored in default XHTML, Crown Legislation Mark-up Language (CLML) XML, Akoma Ntoso (AKN) XML, an HTML 5 serialization of the AKN XML, PDF, and plaintext. For RegData U.K. 1.0, we used XHTML files as the metadata provided within the files hewed closest to our standard metadata collection practices.

- While the bulk downloader has an immense wealth of documents dating back to 1801, the dataset is limited to the years 1992 through 2020. In 1991, such and such took happened which gives us confidence that we have a complete picture of legislation in force starting in 1992. Prior to such and such complete consolidations of legislation do not exist on the website and, therefore, we cannot produce comparable point-in-time data. RegData U.K. 1.0 defines a document, or a single point of data, as an entire statute. This is due to the file structure, as statutes are saved as a single PDF. It is possible to parse each file further, as seen on the Legislation.gov.uk website, so that analyses can be run on the section level. In a future edition, RekData UK may be parsed in a similar way.

- Metadata and data series were produced using text scraping of the HTML documents. As such, the accuracy of the metadata, such as years in effect, are contingent on the accuracy of the source files.

4   Citation

If you use this data, please cite:

McLaughlin, Patrick A., Thurston Powers, and Ethan Greist. RegData U.K. (dataset). QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2021.

Current Version: 1.0