

# CLEANING CORPORATE GOVERNANCE

## CODEBOOK CCG CORPUS

### ***cik***

Central Index Key assigned by the SEC to the company. Note that CIKs functioned as the main identifying variable in our project. Accordingly, should there be a mismatch between CIKs and company names / tickers, treat CIKs as authoritative.

### ***charter\_id***

Unique identifier for the charter.

### ***text***

The text of the charter. Most text available in HTML format was converted in plain text using the package `html2text` in python. Examples apply for charters harvested early in the project, for which we used custom-written code in python. The algorithm used to convert html to text does not affect the content of the text. By contrast, the code used affects the location of line breaks, white spaces, and similar markers of formatting. Text available in image format (e.g., jpg, png) was converted to text by means of OCR using `tesseract`.

### ***OCR***

Dummy variable indicating whether the charter text was available in image format only, which required a conversion to text format by means of OCR. Note that OCR is likely to lead to imperfect results, especially for older charters and charters for which the quality of the image file is poor. '1' indicates a document that was converted via OCR, '0' a document that was available either in text or HTML format.

### ***date\_coded***

Effective date of the charter (or alternatively, the date of adoption). This variable was hand-coded and is not available for all charters in the dataset.

### ***date\_filing***

Filing date of the filing from which the charter was obtained. This variable was automatically coded from meta-information available on EDGAR.

***fullrestatement***

Variable indicating whether a charter document contains the full text of the charter ('Y'), or whether it merely restated parts of the charter text ('N'). This variable was manually coded.

***coname***

Name of the company. This variable is provided for informational purposes only and should not be treated as authoritative (see above ***cik***).

***ticker***

Ticker associated with the company. This variable is provided for informational purposes only and should not be treated as authoritative (see above ***cik***).