

Important Note: Please read the methodology notes for Legislative Data in Interim Report of the Financial Services Inquiry if you wish to understand the full limitations of the data. The key limitation is the inconsistent quality of the HTML for each piece of legislation that is obtained from Federal Register of Legislation, particularly in older legislation, treaties, or intergovernmental agreements. Variables that rely on analysing HTML may be an undercount. Some older pieces of legislation or treaties and intergovernmental agreements do not use markup at all in the HTML for certain textual data, such as chapter, section, and subsection numbers.

Important concepts – Legislation

The following two definitions are based on the Federal Register of Legislation Glossary. For more information, see <https://www.legislation.gov.au/content/whatisit>. The Federal Register of Legislation also publishes a guide to the typical structure of a Commonwealth Act: <https://www.legislation.gov.au/file/StructureOfAnAct>.

As made legislation is legislation in the form in which it was originally made. The text may not reflect in force as it will not include any later amendments that may have been made.

A *compilation* is a version of a law that shows the text of the law as amended at a particular point in time.

Important concepts – Process of analysis

Stopwords are words that are removed for certain types of text analysis, such as determining the number of unique words in a text (ie its vocabulary). The ALRC uses R’s Quanteda package, which includes 172 stopwords.

Excluded structural elements refers to subsection, paragraph, subparagraph, and sub-subparagraph numbers and lettering that appear at the beginning of lines in legislation (eg (1), (d), (iv), (A)). The ARLC removes these for certain types of text analysis.

Legislation Name	Name of the legislation as scraped.
FRL ID	The unique identification number given to the piece of legislation by the Federal Register of Legislation.
URL	Link to the legislation on the Federal Register of Legislation
Enabling Legislation	As identified in the ‘Enabled by’ tab on the series page of the legislation.
Legislation Number/Year	Information indicating the position of the instrument in a series as described under the ‘Details’ tab of the legislation page. For example, ASIC CO 14/1001 or AASB 1004. Not all instruments are given this information.
Instrument type	The type of legislative instrument as described under the ‘Details’ tab of the legislation page. Not all instruments are given a type.
Summary	The second italicised paragraph description that appears under the ‘Details’ tab of the legislation page.
Administrator	The administrator of the legislation as listed under the ‘Details’ tab of the legislation page – see ‘Administered by:’ under this tab. Legislation can have more than one administrator.

Sunset exemption	The legislation authorising an exemption from the sunset provisions of the <i>Legislation Act 2003</i> (if applicable). Listed under the ‘Details’ page of the legislation
Sunset date	The date at which the legislation will sunset (if applicable). Listed under the ‘Series’ page of the legislation
Date of Registration	Date of registration for the legislation as listed under the ‘Details’ tab of the legislation page – see ‘Registered’ under this tab.
Start Date	Date from which a legislation compilation takes effect as listed under the ‘Details’ tab of the legislation page – see ‘Start Date’ under this tab. Only applicable to compilations.
Date of Gazettal	Date the instrument was Gazetted (generally only applicable to instruments made prior to the creation of the ComLaw website (the predecessor to the Federal Register of Legislation). Listed under the ‘Details’ tab of the legislation
Date of Repeal	Date on which an as made piece of legislation was repealed. Only applicable to legislation as made. Compilations have an ‘End Date’.
End Date	Date on which a legislation compilation ceases to have effect as listed under the ‘Details’ tab of the legislation page – see ‘End Date’ under this tab. Only applicable to compilations.
Repealed by	Legislation repealing the instrument (if applicable).
First Tabling Chamber	The first Chamber of Parliament in which the instrument was tabled as listed under the ‘Details’ tab of the legislation page. Not applicable to compilations.
First Tabling Date	The date at which the instrument was tabled in the First Tabling Chamber as listed under the ‘Details’ tab of the legislation page. Not applicable to compilations.
Second Tabling Chamber	The second Chamber of Parliament in which the instrument was tabled as listed under the ‘Details’ tab of the legislation page. Not applicable to compilations.
Second Tabling Date	The date at which the instrument was tabled in the Second Tabling Chamber as listed under the ‘Details’ tab of the legislation page. Not applicable to compilations.
Legislation Pages	The number of pages of the legislation as appears on the ‘Download’ page for the legislation.
Word count	Counts the number of ‘tokens’ using R’s Quanteda package, splitting the document using ‘fasterword’. Does not include endnotes, table of contents, and excluded structural elements. Does not split hyphens.
Entropy	<p>Calculated using the below equation from Patrick McLaughlin et al, ‘Is Dodd-Frank the Biggest Law Ever?’ (2021) 7(1) <i>Journal of Financial Regulation</i> 149, 170. Does not include endnotes, table of contents, and excluded structural elements. Does not split hyphens. Also excludes stopwords, numbers, and alphanumeric words (eg 601AKC). ‘where D is a document, H(D) is the Shannon entropy of document D, WD is the set of unique words occurring in</p> $H(D) = -\sum_{w \in W_D} p_w \log_2(p w),$ <p>document D, and pw is the probability of encountering one of these words at a random point in the text—that is, the frequency of that word as a percentage of the total word count.’</p>

Number of unique substantive words	Counts the number of unique ‘tokens’ using R’s Quanteda package. Does not include endnotes, table of contents, and excluded structural elements. Does split hyphens. Also excludes stopwords, numbers, and alphanumeric words (eg 601AKC).
Average substantive word length	Average word length in a text. Does not include endnotes, table of contents, and excluded structural elements. Does split hyphens. Also excludes stopwords, numbers, and alphanumeric words (eg 601AKC).
Unique word stems for substantive words	The number of unique stem words that appear in a text. For example, the stemmed word of ‘provider’ and ‘provided’ is ‘provid’. Does not include endnotes, table of contents, and excluded structural elements. Does split hyphens. Also excludes stopwords, numbers, and alphanumeric words (eg 601AKC).
Readability	Flesch-Kinkaid score - the lower the score the less readable the text. Has limitations in legislation because headings do not use end of sentence punctuation. This means the sentence length can be overestimated, which makes the text appear less readable.
Defined terms	Counts the number of references to "means", "has the meaning", "has the same meaning" as a proxy for the number of defined terms introduced in the legislation.
Bold and italicised terms	Number of bold and italicised terms marked-up in the HTML.
Act cross-references	Counts the number of references to ‘^Act\$’ that appear in the text of the legislation. The code then deletes results that are immediately preceded by any of the following terms: ‘[Tt]his’, ‘An’, ‘)’’, ‘[Tt]hat’, ‘[Tt]he’, or ‘[Aa]pplication’. The ALRC identified these terms did not indicate an external cross-reference.
Cross references to other Acts per 100 words	‘Act cross-references’ divided by the ‘Word count’.
Internal cross-references - Sections	Counts the number of references to ‘^section.*’, ‘^subsection.*’, ‘^subparagraph.*’, ‘^paragraph.*’. The code then deletes results that are immediately preceded by ‘this’ or followed by ‘of’. The ALRC identified these terms did not indicate an internal cross-reference.
Cross references to other sections per 100 words	‘Internal cross-references – Sections’ divided by the ‘Word count’.
Cross-references - Regulations	Only applicable to Regulations. Counts the number of references to ‘^regulation\$’, ‘^subregulation.*’. The code then deletes results that are immediately preceded by ‘this’ or followed by ‘of’. ALRC identified these terms did not indicate an internal cross-reference.
Cross references to Regulations per 100 words	‘Cross-references – Regulations’ divided by the ‘Word count’.
Conditional Statements	Counts the number of the following that appear in the text of the legislation: ‘^if\$’, ‘^except\$’, ‘^but\$’, ‘^provided\$’, ‘^when\$’, ‘^where\$’, ‘^whenever\$’, ‘^unless\$’, ‘^notwithstanding\$’
Conditional statements per 100 words	‘Conditional Statements’ divided by the ‘Word count’.
Obligations	Counts the number of the following that appear in the text of the legislation: ‘^must\$’, ‘^shall\$’, ‘^may not’, ‘^prohibited’, ‘^required’, ‘^may only’, ‘^cannot be’

Obligations per 100 words	'Obligations' divided by the 'Word count'.
Offences	Counts the number of the following that appear in the text of the legislation: ^offence\$
Offences per 100 words	'Offences' divided by the 'Word count'.
Reasonableness	Counts the number of the following that appear in the text of the legislation: ^reasonabl.*
Reasonableness per 100 words	'Reasonableness' divided by the 'Word count'.
Modifications	Counts the number of the following that appear in the text of the legislation: ^omit.*, ^insert\$', ^substitute\$'
Modifications per 100 words	'Modifications' divided by the 'Word count'.
Contravene	Counts the number of the following that appear in the text of the legislation: ^contravene.*
Contravene per 100 words	'Contravene' divided by the 'Word count'.
Discretions	Counts the number of the following that appear in the text of the legislation: 'Minister may', 'ASIC may', 'ACCC may', 'RBA may', 'APRA may'
Discretions per 100 words	'Discretions' divided by the 'Word count'.
Legislative instruments	Counts the number of the following that appear in the text of the legislation: 'legislative instrument'
Legislative instruments per 100 words	'Legislative instruments' divided by the 'Word count'.
Regulations	Counts the number of the following that appear in the text of the legislation: ^regulations\$
Regulations per 100 words	'Regulations' divided by the 'Word count'.
Strict liability	Counts the number of the following that appear in the text of the legislation: strict liability
Strict liability per 100 words	'Strict liability' divided by the 'Word count'.
Civil liability	Counts the number of the following that appear in the text of the legislation: civil penalt.*
Civil liability per 100 words	'Civil liability' divided by the 'Word count'.
Misleading	Counts the number of the following that appear in the text of the legislation: mislead.*
Misleading per 100 words	'Misleading' divided by the 'Word count'.
Unconscionable	Counts the number of the following that appear in the text of the legislation: unconsciona.*
Unconscionable per 100 words	'Unconscionable' divided by the 'Word count'.
Dishonesty	Counts the number of the following that appear in the text of the legislation: dishonest.*
Dishonesty per 100 words	'Dishonesty' divided by the 'Word count'.
Honesty	Counts the number of the following that appear in the text of the legislation: ^honest.*
Honesty per 100 words	'Honesty' divided by the 'Word count'.
Good faith	Counts the number of the following that appear in the text of the legislation: good faith

Good faith per 100 words	'Good faith' divided by the 'Word count'.
Unfair	Counts the number of the following that appear in the text of the legislation: ^unfair.*
Unfair per 100 words	'Unfair' divided by the 'Word count'.
Fair	Counts the number of the following that appear in the text of the legislation: ^fair.*
Fair per 100 words	'Fair' divided by the 'Word count'.
Mistake	Counts the number of the following that appear in the text of the legislation: ^mistake.*
Mistake per 100 words	'Mistake' divided by the 'Word count'.
Unjust	Counts the number of the following that appear in the text of the legislation: ^unjust.*
Unjust per 100 words	'Unjust' divided by the 'Word count'.