

Variables of concern

- Date variables: 'Signature', 'Force', 'End'
- url
- MainText, AppendixText, AnnexText
- Checked_HUGGO
- Confirmed_HUGGO
- Changes

Cleaning and verifying treaty texts

To verify the treaty texts, the original texts (refer to url) and/or the text stored in TreatyText can serve as references.

Where the current text is incomplete or not clean, new text should be split and added to 'MainText', 'AppendixText', and 'AnnexText' variables.

When adding the new text, preface, preamble, articles, appendices and annexes should be retained. Tables such as the table of contents, schedules, tables of subsidies and other technical details may be removed. Signatures of state parties to the treaties should also be removed. For texts scraped from websites, extra strings of characters or symbols (such as weblinks or random symbols eg. @, #, \$, %, < >) should be removed. Also remove any extra spacing between lines and paragraphs. Note that you might need to insert ' \' before punctuation such as inverted commas " " and dollar signs \$ before pasting in R.

The text may be downloaded from websites in pdf format and converted to text, or copied directly from web pages. Copy and paste the main text including the preamble and preface into 'MainText', the appendices in 'AppendixText', and the annexes in 'AnnexText'. For texts that are not in English, translate using Google Translate before pasting in R, and make a note in the 'Changes' log that the text has been translated.

'Parties' variable: list signatories found in the treaty text (separate states' names with a comma), if available.

Remember to keep the correct url in the 'url' variable.

Managing and verifying data in HUGGO datasets

Three variables, 'Checked_HUGGO', 'Confirmed_HUGGO', and 'Changes' have been added to track progress on manually correcting entries.

'Checked_HUGGO': code 1 when the row's Text and Date (Signature, Force, End) observations have been verified and updated

'Confirmed_HUGGO': list variables for which the observation could be verified and confirmed.

- Eg. List 'Signature' in 'Confirmed_HUGGO' if the Signature date was found and verified in the treaty text or in a manual online search.

- If TreatyText, MainText, AppendixText, or AnnexText are NAs, you can list them in 'Confirmed_HUGGO' if a manual search online does not return any texts.

'Changes': Log changes that have been made to the date variables ('Signature', 'Force', 'End'), url, and text variables ('MainText', 'AppendixText', 'AnnexText')

- Eg. if the url was added or replaced, note 'New url.'
- Eg. if new main agreement text was added, note 'New MainText translated to English from German using Google Translate.'
- Eg. if the Signature date was added or replaced, note 'New Signature date.'