Corpus of Resolutions

————————

UN Security Council

(CR-UNSC)

CODEBOOK

Version 2024-05-19

Title	Corpus of Resolutions: UN Security Council
Abbreviation	CR-UNSC
Author	Seán Fobbe, Lorenzo Gasbarri and Niccoló Ridi
Version	2024-05-19
Download	https://doi.org/10.5281/zenodo.11212056
License	CC0 1.0 Universal

Citation

Seán Fobbe, Lorenzo Gasbarri and Niccoló Ridi (2024). Corpus of Resolutions: UN Security Council (CR-UNSC). Version 2024-05-19. Zenodo. DOI: 10.5281/zenodo.11212056.

Digital Object Identifiers: Concept DOI and Version DOI

This data set is uniquely identified via the Digital Object Identifier (DOI) system. DOIs are persistent identifiers that are globally unique and can be resolved as a link by entering a DOI into the web service at www.doi.org. The DOI given in this document is a 'Version DOI', which uniquely identifies version 2024-05-19. Academics and others who wish to enable replication analyses are strongly advised to cite the *version DOI* and the precise version of the data used. A 'Concept DOI' is available from the page of the Zenodo record under the heading 'Cite all versions?' and will always resolve to the latest version.

Public Domain Status

The full data set and this document are distributed under a **Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication** license. The person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. In no way are the patent or trademark rights of any person affected by CC0, nor are the rights that other persons may have in the work or in how the work is used, such as publicity or privacy rights. Unless expressly stated otherwise, the person who associated a work with this deed makes no warranties about the work, and disclaims liability for all uses of the work, to the fullest extent permitted by applicable law.

Please see <https://creativecommons.org/publicdomain/zero/1.0/legalcode> for the full terms of the license.

Disclaimer

This data set is a personal academic initiative and is not associated with or endorsed by the the United Nations or any of its constituent organs and organizations.

Contents

1 Introduction

The **United Nations Security Council (UNSC)** is the most influential of the principal UN organs. Composed of five permanent and ten non-permanent members, its functioning is constrained by the political context in which it operates. During the Cold War, the complex political relationships between the permanent members and their veto powers significantly affected the capacity of the UNSC to address violations of international peace and security, with only 646 resolutions passed from 1946 to 1989.

Since the 1990s, the activity of the UN Security Council has increased dramatically and produced 2721 resolutions up to the end of 2023. The length, complexity and thematic breadth of the resolutions has also increased, prompting calls to redefine it as a quasi-legislative body.

Under Articles 24 and 25 of the UN Charter, member states have conferred upon the UNSC the 'primary responsibility for the maintenance of international peace and security' and have agreed 'to accept and carry out' its decisions. The discharge of this function is carried out through the powers bestowed upon it under Chapter VI of the UN Charter, 'Pacific Settlement of Disputes', Chapter VII, 'Action with Respect to Threats to the Peace, Breaches of the Peace, and Acts of Aggression', Chapter VIII, 'Regional Arrangements', and Chapter XII, 'International Trusteeship System'.

Under the peace and security mandate, its areas of activity cover disarmament, pacific settlement of disputes, enforcement, and, until 1994, strategic areas in a trusteeship agreement. Its functions also pertain to the correct working of the United Nations, covering issues of membership, the appointment of the Secretary General, the elections of judges of the International Court of Justice (ICJ), the calling of special and emergency sessions of the General Assembly, the amendment of the Charter and of the ICJ Statute.

The **Corpus of Resolution: UN Security Council (CR-UNSC)** collects and presents for the first time all resolutions, drafts, and meeting records of the UN Security Council with their associated metadata, as published by the UN Digital Library.[1]

The quantitative analysis of international legal data is still in its infancy, a situation exacerbated by the lack of high-quality empirical data. Most advanced data sets are held in commercial databases and are therefore not easily available to academic researchers, journalists and the general public. With this data set we hope to contribute to a more systematic and empirical view of the international legal system. In an international community founded on the rule of law the activities of the United Nations must be public, transparent and defensible. In the 21st century this requires quantitative scientific review of decisions and actions.

Design, construction and compilation of this data set are based on the principles of general availability through freedom from copyright (public domain status), strict transparency and full scientific reproducibility. The *FAIR Guiding Principles for Scientific Data Management and Stewardship* (Findable, Accessible, Interoperable and Reusable) inspire both the design and the manner of publication.[2]

---

[1]   https://digitallibrary.un.org/.
[2]   Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.

2   Reading Files

The data are published in open, interoperable and widely used formats (CSV, TXT, PDF, GraphML). They can be used with all modern programming languages (e.g. Python or R) and graphical interfaces. The PDF collections are intended to facilitate traditional legal research.

Important: Missing values are always coded as 'NA'.

2.1   CSV Files

Working with the CSV files is recommended. CSV[3] is an open and simple machine-readable tabular data format. In this data set values are separated by commas. Each column is a variable and each row is a document. Variables are explained in detail in section 5.

To read **CSV** files into R we strongly recommend using the fast file reader **fread()** from the **data.table** package (available on CRAN). The file can be read into **R** like so:

```
library(data.table)

unsc <- fread("CR-UNSC_2023-08-23_EN_CSV_Main-Dataset.csv")
```

2.2   TXT Files

The **TXT** files, including metadata, can be read into **R** with the package **readtext** (available on CRAN):

```
library(readtext)

unsc <- readtext("txt_res_en/*.txt",
                 docvarsfrom = "filenames",
                 docvarnames = c("body",
                                 "doctype",
                                 "res_no",
                                 "year",
                                 "language"),
                 dvsep = "_",
                 encoding = "UTF-8")
```

---

[3]   The CSV format is defined in RFC 4180: <https://tools.ietf.org/html/rfc4180>.

3 Data Set Design

3.1 Overview

The **Corpus of Resolutions: UN Security Council (CR-UNSC)** collects and structures in human- and machine-readable form all published resolutions of the United Nations Security Council (UNSC) in all six official UN languages (Arabic, Chinese, English, French, Spanish and Russian). Specialized variants of the data set provide citation data and a bibliopgrahic database compatible with a wide range of end-user applications.

This structure and content of this section build on the suggestions offered in: Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

3.2 Table of Sources

Data Source	Citation
Primary Data Source	https://digitallibrary.un.org/
Source Code	https://doi.org/10.5281/zenodo.11212057

3.3 Motivation

Our motivation in creating this dataset is to enhance international legal and political research in general, increase understanding of the work of the UN Security Council and facilitate creative digital work in the area of international peace and security.

The CR-UNSC is a general-purpose dataset intended to provide high-quality legal and political data that enables researchers, practitioners, civil society, UN officials, governments and any other interested parties to address a wide range research questions related to the UN Security Council.

The use of the main data set is not limited to a particular line of inquiry or specific discipline. Specialized variants of the data set offer citation data and a bibliographic database compatible with a wide range of end-user applications.

3.4 Funding

The authors did not receive any specific funding to work on this dataset. All work on the dataset was carried out in addition to their regular academic activities.

3.5 Composition

The dataset is primarily composed of the texts of all resolutions in all official UN languages published by the UN Security Council. We aim for a complete accounting of all resolutions, up to the most recent resolution that was available through the UN Digital Library on the date of data set creation. We further provide the English texts of the draft resolutions and meeting records associated with each resolutons.

Each resolution is accompanied by as much high-quality metadata as could be acquired and validated, as well as additional metadata extracted from the text of the resolutions. The exact composition and scope of the data set is documented and visualized with frequency tables at the end of this Codebook.

The data set is self-contained, but it contains the original URLs to UN Digital Library record pages and URLs to individual PDF files for purposes of reproducibility, data validation and creating digital applications that would benefit from integrating linked open data. We assume these URLs will remain stable over time, but their maintenance depends on the actions of the UN Digital Library. In case of link failure, newer version of the data set will contain updated links.

There are no recommended data splits. However, all resolutions are consecutively numbered and timestamped with the date of adoption. Time series analyses should take this into account.

3.6   Sensitive Data

Resolutions of the UN Security Council are publicly available official documents of an international organization. The metadata strictly relates to these public documents. While the text of the resolutions contains references to individuals and organizations, this information has always been published in unredacted form by the UN Security Council. We have not redacted any of this information to ensure the integrity of these official documents. In any case, the maintenance of international peace and security at the highest international level is a public affair of such magnitude that privacy rights of individuals and organizations cannot prevent their mention and discussion of their activities in official texts of the UN Security Council.

3.7   Data Collection

Data were collected with the explicit consent of the United Nations. All documents were downloaded via TLS-encrypted connections and cryptographically signed after data processing was complete. The data set collects all resolutions issued by the UN Security Council that were published by the UN Digital Library up to resolution number 2722.

3.8   Processing and Quality Assurance

Dozens of automated tests were conducted to ensure the quality of the data and metadata. These include:

- Validation of variable types
- Plausibility checks on expected values for many variables (e.g. minima, maxima, plausible values)
- Completeness checks on download and full dataset
- Text quality and language purity checks
- Creation of visualizations for many common descriptive analyses

For results of each test and more information on the construction of the data set please refer to the Compilation Report and the Quality Assurance Report.

Figure 1: Dataset Compilation Process

The raw data for all stages of the data pipeline ('targets') was stored and is published together with the source code. The original files acquired from the UN Digital Library are part of the dataset itself, insofar as we consider them useful to other users.

3.9 Uses

The dataset has been designed to be useful for any and all purposes related to the quantitative analysis of international law and to traditional researchers who seek a complete collection of UN Security Councilresolutions. The dataset is identified by a unique Digital Object Identifier (DOI) and, if correctly cited, all public applications should be searchable on the internet.

The dataset can potentially be used in the following applications

- Information retrieval (discovery of relevant and related resolutions)
- Automatic summarization (full text and brief summaries and titles available)
- Topic analysis (full-text available)
- Citation analysis (UNSC and UNGA citation data is included)
- Machine translation (all six UN languages available, but note quality and completeness constraints on non-English texts)
- Argument mining

The dataset should not be used for the following tasks:

- Prediction of adoption success or voting behavior (dependent variable) based on the text of a resolution or draft (input variable). The data set contains only successfully adopted resolutions, so there is no control group of unsuccessful drafts.

3.10 Distribution

The datast is cleared for immediate use by the general public and professional audiences. It is distributed free of charge.

Authentic copies of the data set and source code will be distributed on the internet via Zenodo, a non-profit scientific repository managed by CERN.

The concept of the data set is uniquely identified by a digital object identififer (DOI), which always resolves to the latest version. In addition to the 'Concept DOI' each version of the data set is further uniquely identified with its own DOI, the 'Version DOI'.

The dataset and source code are distributed under standardized and widely recognized open licenses. We are not aware of any regulatory restrictions that would limit distribution or re-distribution of the dataset.

Development versions of the source code are available on GitHub but carry no promise or expectation of correctness or functionality.

3.11 Copyright Status: Public Domain

The copyright status of UN documents, resolutions, drafts and meeting records is governed by UN Administrative Instruction ST/AI/189/Add.9/Rev.2 of 17 September 1987.[4] UN Administrative Instruction ST/AI/189/Add.9/Rev.2/Add.2 extended the temporal scope

---

[4]   https://en.wikisource.org/wiki/Administrative_Instruction_ST/AI/189/Add.9/Rev.2

of the former until 'further revision to that instruction'.[5] As no revision has been published to this day, the rules remain in force. The most recent Index to Administrative Issuances ST/IC/2019/1 lists Administrative Instruction ST/AI/189/Add.9/Rev.2 as active.

UN Administrative Instruction ST/AI/189/Add.9/Rev.2 of 17 September 1987 clearly states:

> 2. The following categories of material will, as at present, be left in the public domain, i.e., the United Nations will not seek copyright therefor unless, prior to issue and in exceptional circumstances, the Publications Board decides otherwise, in consultation with the Office of Legal Affairs.
>
>   (a) Official Records: a series of printed publications relating to the proceedings of organs or conferences of the United Nations. They include verbatim or summary records, documents or check-lists of documents, issued in the form of annexes to those records, including periodic supplements, such as the quarterly ones of the Security Council; and reports to those organs of their subordinate or affiliated bodies, compilations of resolutions, certain reports of the Secretary-General and other selected publications, which are issued in the form of supplements;
>   (b) United Nations documents: written material officially issued under a United Nations document symbol, regardless of the form of production, although, in practice, the term is applied mainly to material offset from typescript and issued under a masthead. The term also applies to written material issued simultaneously or sequentially in the form of documents and publications;

We wish to honor the letter and spirit of this UN policy. To ensure the widest possible distribution of official UN documents and to promote the international rule of law we waive any copyright that might have accrued by creating the dataset under a **Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication**.

For details of the waiver please refer to the CC0 copyright notice at the beginning of this Codebook or visit the Creative Commons website for the full terms.[6]

3.12   Maintenance and Updates

Authentic versions of the dataset and source are always published and hosted on Zenodo.[7] Zenodo guarantees the professional storage of published records in CERN's Data Centre and public availability for the lifetime of CERN, which has a defined research program for at least 20 years.

Maintenance and further development of the dataset and source code will be carried out by the author(s) for the foreseeable future. Contact details are available on the Zenodo page of the data set.

We intend to update the dataset at least once per year. All updates will be published on Zenodo and uniquely identified by date stamps and Version DOIs. Older versions will continue to be available on Zenodo under their original Version DOI.

---

[5]   https://en.wikisource.org/wiki/Administrative_Instruction_ST/AI/189/Add.9/Rev.2/Add.2
[6]   https://creativecommons.org/publicdomain/zero/1.0/legalcode
[7]   https://www.zenodo.org

Contributions to the dataset can be made by reporting issues and suggesting enhancements to the authors, as well as proposing code changes and additions on GitHub.

3.13   Source Code and Compilation Report

The full Source Code for the creation of this data set, the resulting Compilation Report, a Quality Assurance Report and this Codebook are published open access and permanently archived with Zenodo, the scientific repository of CERN.

With every compilation of the full data set an extensive **Compilation Report** is created in a professionally layouted PDF format. The Compilation Report includes the full data pipeline, comments and explanations of design decisions, relevant computational results, exact timestamps and a table of contents with clickable internal hyperlinks to each section. The Compilation Report is published under the same DOI as the Source Code.

For details of the construction and validation of the data set please refer to the Compilation Report and Quality Assurance Report in addition to this Codebook.

3.14   Limitations

Users should bear in mind certain important limitations:

- The data set contains only those documents which were published by the UN Security Council and have been made available by the UN Digital Library on its official website (*publication bias*). This mainly affects languages other than English
- While Tesseract yields high-quality OCR results, current OCR technology is not perfect and minor errors must be expected (*OCR bias*). This affects resolutions and associated drafts and meeting records up to resolution 899.
- The OCR texts for all languages other than English were *not* manually post-processed, so, in addition to OCR errors, may contain superfluous texts from other documents printed on the same page. This affects resolutions and associated drafts and meeting records up to resolution 899.

4 Variants and Target Audiences

4.1 Overview of Variants

The published data set contains all six official UN languages (English, French, Spanish, Arabic, Chinese and Russian) and is published in several different variants intended for specific target audiences. In particular, a tabular CSV variant for general data analysis, a GraphML network representation for citation analysis, a PDF variant for traditional research and a BIBTEX variant to assist with efficient citation of UNSC resolutions.

4.1.1 Recommended for Quantitative Research

- ALL_CSV_FULL
- ALL_CSV_META
- ALL_TXT
- ANALYSIS
- CITATIONS_GRAPHML
- EN_TXT_BEST

4.1.2 Recommended for Traditional Research

- ALL_PDF_RESOLUTIONS
- ALL_TXT
- ANALYSIS
- BIBTEX_OSCOLA
- EN_PDF_Drafts
- EN_PDF_Meeting-Records
- EN_TXT_BEST

4.2 ALL_CSV_FULL

A structured tabular representation of the full data set, stored within a single comma-delimited file (CSV). Includes the full texts of resolutions in all languages, draft texts in English, meeting records in English and the full complement of metadata described in the Codebook.

4.3 ALL_CSV_META

A structured tabular representation of the full metadata, contained within a single comma-delimited file (CSV). The *META* CSV variant is equivalent to the *FULL* CSV variant, except that all text variables have been removed. This is particularly useful where disk space is limited or where an analysis requires only metadata.

4.4 ALL_PDF_Resolutions

All resolutions issued by the UNSC in all six official UN languages in PDF format as published by the UN Digital Library. Includes resolutions 1 through 899 in English with a new and enhanced OCR text layer created with an advanced LSTM neural network machine learning engine.

OCR versions of other languages are not included in this PDF archive to keep file size reasonable. However, the data pipeline produces PDF files with enhanced OCR text for all languages.

Please note that the gold-standard text for the English versions of resolutions 1 through 899 is only available in the *TXT* variant, the text layer in the English PDF files is OCR output.

4.5   ALL_TXT

All resolutions issued by the UNSC in all six official UN languages, as well as English drafts and meeting records, in TXT format. The TXT representation was extracted from the original PDF files with simple text layer extraction and alternatively OCR recognition with Tesseract, an advanced LSTM neural network machine learning engine. This variant also includes the revised gold standard texts for resolutions 1 through 899 in English.

Researchers with slow internet connections or who wish to save disk space should consider the *TXT* variant, as this still provides a reasonable visual approximation of the original documents, but offers the advantage of drastically reduced file size.

4.6   ANALYSIS

This archive contains almost all of the machine-readable analysis output generated during the data set creation process to facilitate further analysis, in CSV format for tables, PDF and PNG format for diagrams. Minor analysis results are documented only in the Compilation Report and the Quality Assurance Report.

4.7   BIBTEX_OSCOLA

A selection of variables from the metadata converted to `.bibtex` format compliant with the **bl-oscola** 1.7 standard. This file should be importable in all standard reference managers such as Zotero, EndNote, Citavi and Jabref.

Note that the content of the variables 'title' and 'other_title' are switched in the BIBTEX variant to make for a better experience in bibliography software. The 'other_title' variable is more informative, but the main data set tries to respect the UN Digital Library naming convention.

4.8   CITATIONS_GRAPHML

A graph (network) representation of all citations from UNSC resolutions to other UNSC and UNGA resolutions in GraphML format. The format is widely supported, in particular by GUI software such as Gephi.

The UNSC resolution nodes are associated with as much metadata as makes sense from a network perspective. The file does not include full texts of resolutions or URLs to PDF files. If you wish to select nodes based on the contents of the full text, please run your analysis on the CSV file, export the resolution numbers and use these to select the appropriate nodes in your network analysis software.

4.9   EN_PDF_Drafts

All drafts of resolutions texts in English as published by the UN Digital Library, in PDF format. We only include English draft resolutions, as the draft text is primarily negotiated in English. Enhanced OCR versions of the PDF files are produced by the data pipeline, but not included in the published data set to keep file size reasonable.

Please note that the files are named according to the associated final resolution (and not the official document symbol of the draft) to speed up findability.

4.10   EN_PDF_Meeting-Records

All meeting records associated with resolutions in English as published by the UN Digital Library, in PDF format. We only include English draft resolutions, as the draft text is primarily negotiated in English. Enhanced OCR versions of the PDF files are produced by the data pipeline, but not included in the published data set to keep file size reasonable.

Please note that the files are named according to the associated final resolution (and not the official document symbol of the meeting record) to speed up findability.

4.11   EN_TXT_BEST

All resolutions issued by the UNSC in English, in the best quality available to the dataset authors. This is a merged set of gold-standard, expert-revised OCR text for resolutions 1 through 899 (and some others) and extracted born-digital text for resolutions 900 and higher.

This TXT version is equivalent to the "text" variable in the main CSV dataset. We strongly recommend using this variant for all serious work on the English texts, if use of the CSV file is not feasible.

5 Variable Descriptions

5.1 General Remarks

- Missing values are always coded as 'NA'.
- All Strings are encoded in UTF-8.
- A significant part of the metadata has been drawn from the UN Digital Library metadata records and was originally curated by the UN Digital Library.

5.2 ID Variables

The group ID variables identifies the resolution, its draft, meeting record and various associated documents with symbols and dates.

Variable	Type	Description
body	String	The body issuing the document. In the main data set this is always the UN Security Council ('S'), in the network data this may also be the UN General Assembly ('A').
call_number	String	The call number of the resolution. This is either 'UNS(01)/R3' or NA. Curated by the UN Digital Library.
committee_report	String	The official UN symbol of the committee report. Example: 'S/2006/58'. Curated by the UN Digital Library.
date	Date	The adoption date of the resolution in the format YYYY-MM-DD (ISO 8601). Example: '2016-11-09'. Extracted with REGEX from the 'other_titles' variable and corrected by the authors.
doc_id	String	The filename of the extracted English language TXT file.
doctype	String	The type of the main document. In this data set always 'RES'.
draft	String	The official UN symbol of the draft. Example: 'S/2023/102'. Curated by the UN Digital Library.
meeting_no	Integer	The number of the UN Security Council meeting in which the resolution was adopted.
meeting_record	String	The official UN symbol of the meeting record. Example: 'S/PV.8891'. Curated by the UN Digital Library.

Variable	Type	Description
related_resource	String	The official UN symbol of related resources. Example: 'S/1234'. Curated by the UN Digital Library.
res_no	Integer	The number of the resolution. Example: '156'. Extracted with REGEX from the 'other_titles' variable.
symbol	String	The official UN symbol of the resolution. Example: 'S/RES/2316' (2016)
year	Integer	The adoption year of the resolution in the format YYYY (ISO 8601). Example: '2016'

5.3   Text Variables

Text variables contain the full text of all six language versions of the UN Security Council's resolutions. Also provided are pre-computed statistical measures of the length of each resolution (English text only).

Variable	Type	Description
nchars	Integer	The number of characters in the English text of the resolution. Calculated with the nchar() function in R.
npages	Integer	Number of pages of the English PDF version of the document. Curated by the UN Digital Library.
nsentences	Integer	The number of sentences in a given document. Computed with the quanteda package for R. This count was generated based on plain tokenization with no further pre-processing (e.g. stopword removal, removal of numbers, lowercasing) applied. The rules for detecting sentence boundaries are very complex and are described in 'Unicode Standard Annex No 29'.
ntokens	Integer	The number of tokens (an arbitrary character sequence bounded by whitespace) in a given document. Computed with the quanteda package for R. This count was generated based on plain tokenization with no further pre-processing (e.g. stopword removal, removal of numbers, lowercasing) applied.

Variable	Type	Description
ntypes	Integer	The number of types (unique tokens) in a given document. Computed with the quanteda package for R. This count was generated based on plain tokenization with no further pre-processing (e.g. stopword removal, removal of numbers, lowercasing) applied.
text	String	The English full text of the resolution. Up to resolution 899 this is OCR text revised by international lawyers (gold standard), later resolutions are assumed to be born-digital and the text is extracted from the PDF file as-is.
text_ar	String	The Arabic full text of the resolution. Up to resolution 899 this is non-revised OCR text, later texts are assumed to be born-digital and the text is extracted from the PDF file as-is. WARNING: non-revised OCR text may contain errors and superfluous text from other documents if they were printed on the same page.
text_draft	String	The English full text of the draft resolution. Up to resolution 899 this is non-revised OCR text, later texts are assumed to be born-digital and the text is extracted from the PDF file as-is. WARNING: non-revised OCR text may contain errors and superfluous text from other documents if they were printed on the same page.
text_es	String	The Spanish full text of the resolution. Up to resolution 899 this is non-revised OCR text, later texts are assumed to be born-digital and the text is extracted from the PDF file as-is. WARNING: non-revised OCR text may contain errors and superfluous text from other documents if they were printed on the same page.
text_fr	String	The French full text of the resolution. Up to resolution 899 this is non-revised OCR text, later texts are assumed to be born-digital and the text is extracted from the PDF file as-is. WARNING: non-revised OCR text may contain errors and superfluous text from other documents if they were printed on the same page.

Variable	Type	Description
text_meeting	String	The English full text of the meeting record. Up to resolution 899 this is non-revised OCR text, later texts are assumed to be born-digital and the text is extracted from the PDF file as-is. WARNING: non-revised OCR text may contain errors and superfluous text from other documents if they were printed on the same page.
text_ru	String	The Russian full text of the resolution. Up to resolution 899 this is non-revised OCR text, later texts are assumed to be born-digital and the text is extracted from the PDF file as-is. WARNING: non-revised OCR text may contain errors and superfluous text from other documents if they were printed on the same page.
text_zh	String	The Chinese full text of the resolution. Up to resolution 899 this is non-revised OCR text, later texts are assumed to be born-digital and the text is extracted from the PDF file as-is. WARNING: non-revised OCR text may contain errors and superfluous text from other documents if they were printed on the same page.

5.4 Thematic Variables

Thematic variables contain a number of different classifications of UN Security Council texts and will prove especially useful for refining the data to fit a thematic research question.

Variable	Type	Description
agenda_information	String	Information on the agenda. Curated by the UN Digital Library.
aggression	Logical	Whether the string 'aggression' is present in the English text of the resolution. Either TRUE or FALSE.
chapter6	Logical	Whether the string 'Chapter VI' is present in the English text of the resolution. Either TRUE or FALSE.
chapter7	Logical	Whether the string 'Chapter VII' is present in the English text of the resolution. Either TRUE or FALSE.

Variable	Type	Description
chapter8	Logical	Whether the string 'Chapter VIII' is present in the English text of the resolution. Either TRUE or FALSE.
contains	String	The full name of an instrument annexed to a resolution, otherwise 'NA'. Example: Statute of the International Tribunal for Rwanda (1994). Curated by the UN Digital Library.
human_rights	Logical	Whether the string 'human rights' is present in the English text of the resolution. Either TRUE or FALSE.
notes	String	Additional notes on the resolution. Curated by the UN Digital Library.
other_titles	String	The number, year, body, meeting number and date of the resolution in narrative text. Curated by the UN Digital Library.
peace_breach	Logical	Whether the strings 'breach of the peace' or 'breach of international peace' is present in the English text of the resolution. Either TRUE or FALSE.
peace_threat	Logical	Whether the strings 'threat to peace', 'threat to international peace' or 'threat to the peace' are present in the English text of the resolution. Either TRUE or FALSE.
self_defence	Logical	Whether the strings 'self-defense' or 'self-defence' are present in the English text of the resolution. Either TRUE or FALSE.
series	String	Indicates whether an Advisory Opinion was requested from the International Court of Justice. Curated by the UN Digital Library.
subjects	String	One or more thematic labels. Multiple labels are separated by vertical bars. Curated by the UN Digital Library.
summary	String	A summary of the resolution, presumably human-created. Curated by the UN Digital Library.
title	String	The number, year and topic of the resolution. Curated by the UN Digital Library.

Variable	Type	Description
topic	String	A short description of the resolution's specific topic. Example: 'on withdrawal of USSR troops from Iran'. Extracted with REGEX from the 'other_titles' variable.

5.5   Vote Variables

Vote variables document the voting results and voting patterns of UNSC resolutions. Includes counts, total and disaggregated, as well as individual votes cast by each State.

Note that the UNSC was enlarged from 11 to 15 members (from 6 to 10 non-permanent members) in 1965, so comparing absolute vote counts is only appropriate before and after those years. All other analyses should use proportions.

When attempting to build prediction models, please bear in mind that the data set only documents successfuly adopted resolutions and their voting data. Many draft resolutions fail during the negotiation stage or because they cannot secure the necessary number of votes. Any prediction model should inclues these failed drafts, but they are not included as part of this dataset.

Variable	Type	Description
vote_abstention	Integer	The number of abstentions. Based on the variable 'vote_summary'.
vote_date	Date	The date of the vote on the resolution in the format YYYY-MM-DD (ISO 8601). Example: 1948-08-19. Curated by the UN Digital Library.
vote_detail	String	A breakdown of all votes by State. Curated by the UN Digital Library.
vote_no	Integer	The number of 'no' votes. Extracted with REGEX from the variable 'vote_summary'.
vote_nonvote	Integer	The number of non-voting members. Extracted with REGEX from the variable 'vote_summary'.
vote_summary	String	A narrative summary of the voting counts (yes, no, abstention, non-voting, total voting membership). Curated by the UN Digital Library.
vote_total	Integer	The total voting membership of the UN Security Council at the time of the vote. Extracted with REGEX from the variable 'vote_summary'.
vote_yes	Integer	The number of 'yes' votes. Extracted with REGEX from the variable 'vote_summary'.

5.6 Geographic Variables

Geographic variables provide insight into the geographic distribution of the UNSC's attention. Note that the voting variables contain the geographic distribution of votes.

Variable	Type	Description
iso_alpha3	String	The ISO 3166-1 Alpha-3 codes of all countries mentioned in the text of the resolution. Example: GBR. If more than one, separated by a pipe character. Extracted with REGEX from the English resolution text by searching for short names listed in the UN M49 standard (slightly modified for better accuracy, for example by adding 'DPRK' for North Korea). Warning: a hit for the Congo will always show up when the Democratic Republic of the Congo (DRC) is mentioned. Further disambiguation is not possible because both the short name (Congo) and the full name (Republic of the Congo) are perfect subsets of the official name of the DRC.
iso_name	String	The ISO 3166-1 Alpha-3 names of all countries mentioned in the resolution. Example: Afghanistan. If more than one, separated by a pipe character. Transformation of the variable 'iso_alpha3'.
m49_countrycode	String	The UN M49 country codes of all countries mentioned in the resolution. Example: 818. If more than one, separated by a pipe character. Transformation of the variable 'iso_alpha3'.
m49_region	String	The UN M49 regions of all countries mentioned in the resolution. Example: Americas. If more than one, separated by a pipe character. Transformation of the variable 'iso_alpha3'.
m49_region_intermediate	String	The UN M49 intermediate regions of all countries mentioned in the resolution. Example: South America. If more than one, separated by a pipe character. Note that not all countries have intermediate regions assigned. Transformation of the variable 'iso_alpha3'.
m49_region_sub	String	The UN M49 sub-regions of all countries mentioned in the resolution. Example: Latin America and the Caribbean. If more than one, separated by a pipe character. Transformation of the variable 'iso_alpha3'.

5.7 Source Variables

Source variables provide explicit references for the data sources via URLs to UN Digital Library record pages and documents.

Variable	Type	Description
url_record	String	The URL of the UN Digital Library record for the main resolution.
url_record_draft	String	The URL of the UN Digital Library record for the draft resolution.
url_record_meeting	String	The URL of the UN Digital Library record for the meeting transcript.
url_res_ar	String	The URL of the PDF file containing the Arabic resolution text. Hosted by the UN Digital Library.
url_res_en	String	The URL of the PDF file containing the English resolution text. Hosted by the UN Digital Library.
url_res_es	String	The URL of the PDF file containing the Spanish resolution text. Hosted by the UN Digital Library.
url_res_fr	String	The URL of the PDF file containing the French resolution text. Hosted by the UN Digital Library.
url_res_ru	String	The URL of the PDF file containing the Russian resolution text. Hosted by the UN Digital Library.
url_res_zh	String	The URL of the PDF file containing the Chinese resolution text. Hosted by the UN Digital Library.
url_draft_ar	String	The URL of the PDF file containing the Arabic draft text. Hosted by the UN Digital Library.
url_draft_en	String	The URL of the PDF file containing the English draft text. Hosted by the UN Digital Library.
url_draft_es	String	The URL of the PDF file containing the Spanish draft text. Hosted by the UN Digital Library.
url_draft_fr	String	The URL of the PDF file containing the French draft text. Hosted by the UN Digital Library.
url_draft_ru	String	The URL of the PDF file containing the Russian draft text. Hosted by the UN Digital Library.

(continued)

Variable	Type	Description
url_draft_zh	String	The URL of the PDF file containing the Chinese draft text. Hosted by the UN Digital Library.
url_meeting_ar	String	The URL of the PDF file containing the Arabic meeting record text. Hosted by the UN Digital Library.
url_meeting_en	String	The URL of the PDF file containing the English meeting record text. Hosted by the UN Digital Library.
url_meeting_es	String	The URL of the PDF file containing the Spanish meeting record text. Hosted by the UN Digital Library.
url_meeting_fr	String	The URL of the PDF file containing the French meeting record text. Hosted by the UN Digital Library.
url_meeting_ru	String	The URL of the PDF file containing the Russian meeting record text. Hosted by the UN Digital Library.
url_meeting_zh	String	The URL of the PDF file containing the Chinese meeting record text. Hosted by the UN Digital Library.

5.8 Meta Variables

Meta variables provide further information on the data set, including version, DOIs and licensing information. These document provenance and ensure compliance with FAIR Data Principles F1, F3 and R1.1.

Variable	Type	Description
doi_concept	String	The Digital Object Identifier (DOI) for the *concept* of the data set. Resolving this DOI via www.doi.org allows researchers to always acquire the *latest version* of the data set. The DOI is a persistent identifier suitable for stable long-term citation. Principle F1 of the FAIR Data Principles ('data are assigned globally unique and persistent identifiers') recommends the documentation of each data set with a persistent identifier and Principle F3 its inclusion with the metadata. Even if the CSV data set is transmitted without the accompanying Codebook this allows researchers to establish provenance of the data. Added by the dataset authors.
doi_version	String	The Digital Object Identifier (DOI) for the *specific version* of the data set. Resolving this DOI via www.doi.org allows researchers to always acquire this *specific version* of the data set. The DOI is a persistent identifier suitable for stable long-term citation. Principle F1 of the FAIR Data Principles ('data are assigned globally unique and persistent identifiers') recommends the documentation of each data set with a persistent identifier and Principle F3 its inclusion with the metadata. Even if the CSV data set is transmitted without the accompanying Codebook this allows researchers to establish provenance of the data. Added by the dataset authors.
license	String	The license for the data set. In this data set it is always Creative Commons Zero 1.0 Universal. Useful when merging individual rows with other data sets. Added by the dataset authors.
version	Date	The version of the data set as a date in the form YYYY-MM-DD (ISO-8601). The version represents the date on which the data set creation process was begun. Added by the dataset authors.

6   Citation Network Data

The CR-UNSC comes with a specialized variant of the dataset that contains all citations from UNSC resolutions to other UNSC resolutions and UNGA resolutions. This weighted, directed graph (network) representation of citations is available in GraphML format.

The number of nodes gives a count of resolutions with in- or outgoing citations. The number of edges gives a count of resolution-pairs that are connected by at least one citation. The individual weights for directed edges give the number of citations for the bilateral links. The out-strength gives the total count of citations between all resolutions (excluding self-citations).

Metric	Value	Metric	Value
Number of Nodes	2,516.00	Min In-Degree	0.00
Number of Edges	15,507.00	Mean Out-Degree	6.16
Strength (Out)	26,866.00	Max Out-Degree	49.00
Mean Degree	12.33	Min Out-Degree	0.00
Max Degree	336.00	Diameter	40.00
Min Degree	1.00	Radius	1.00
Mean In-Degree	6.16	Assortativity (Degree)	0.10
Max In-Degree	331.00	Mean Distance	6.68

The UNSC resolution nodes are associated with as much metadata as makes sense from a network perspective. The following meta data variables are included:

body, call_number, committee_report, date, doc_id, doctype, draft, meeting_no, meeting_record, related_resource, res_no, symbol, year, nchars, npages, nsentences, ntokens, ntypes, agenda_information, aggression, chapter6, chapter7, chapter8, contains, human_rights, notes, other_titles, peace_breach, peace_threat, self_defence, series, subjects, summary, title, topic, vote_abstention, vote_date, vote_detail, vote_no, vote_nonvote, vote_summary, vote_total, vote_yes, iso_alpha3, iso_name, m49_countrycode, m49_region, m49_region_intermediate, m49_region_sub, doi_concept, doi_version, license, version, name

The GraphML file does not include full texts of resolutions. If you wish to select nodes based on the contents of the full text, please run your analysis on the CSV file, export the resolution numbers and use these to select the appropriate nodes in your network analysis software.

CR–UNSC | Version 2024–05–19 | Degree Distribution of Citations

Fobbe/Gasbarri/Ridi | DOI: 10.5281/zenodo.11212056
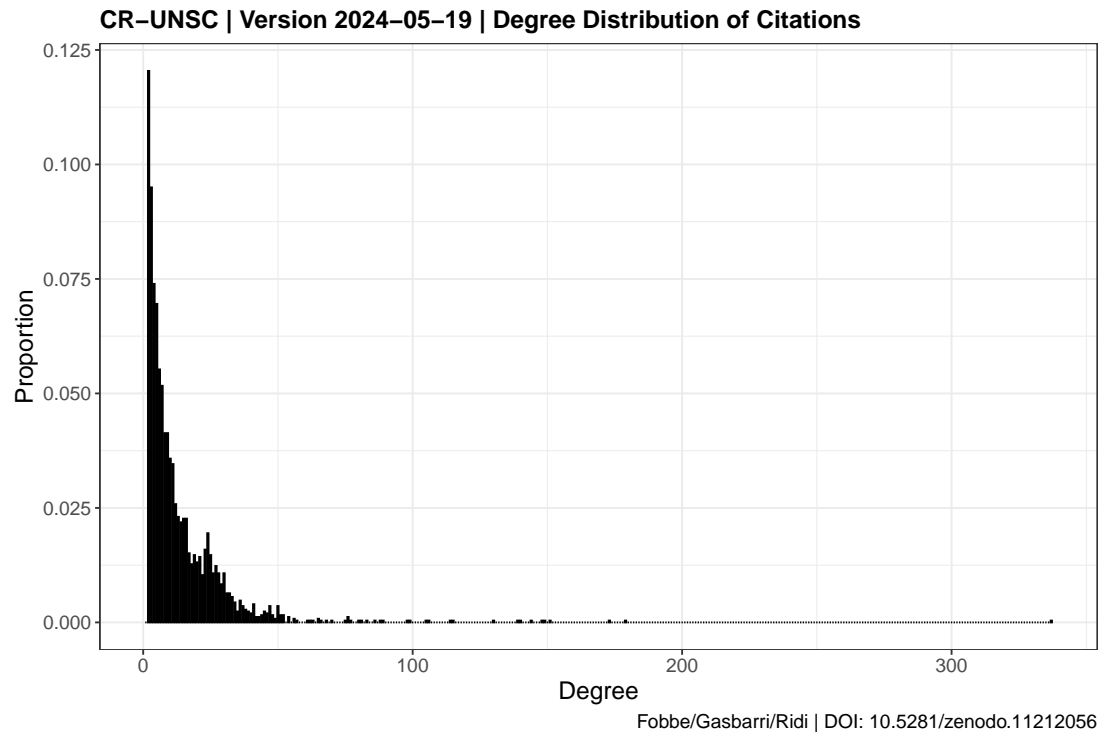
Figure 2: Degree distribution of the UNSC citation network.

CR–UNSC | Version 2024–05–19 | In–Degree Distribution of Citations

Fobbe/Gasbarri/Ridi | DOI: 10.5281/zenodo.11212056
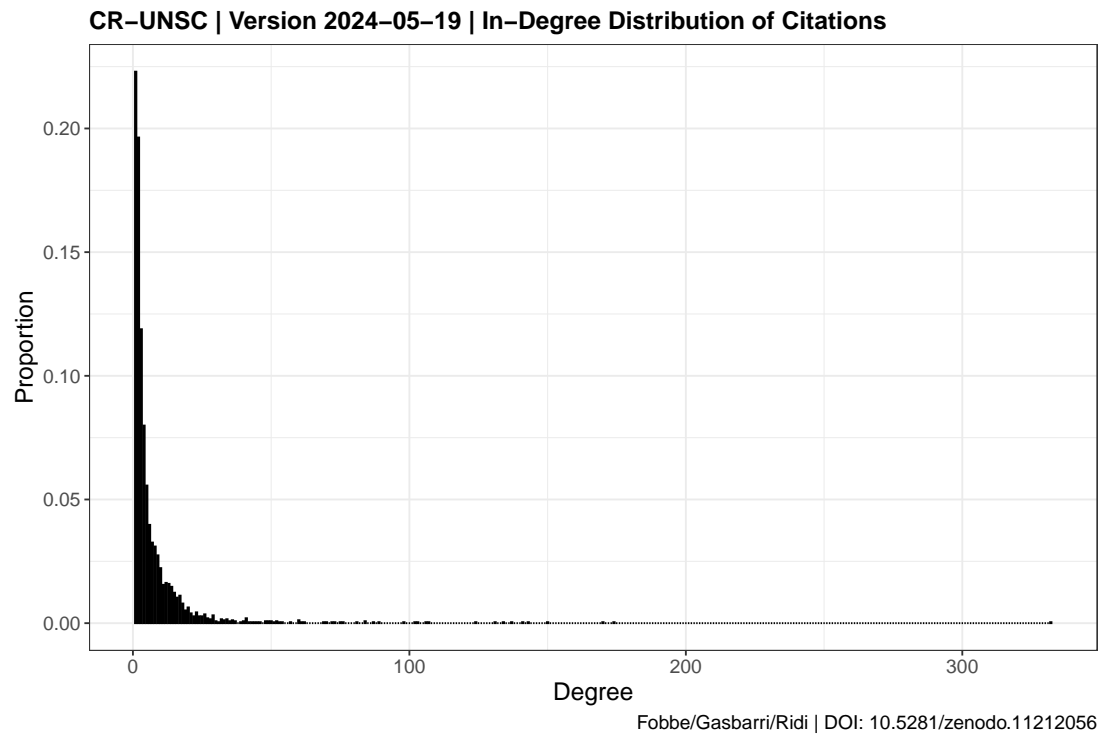
Figure 3: In-Degree distribution of the UNSC citation network.
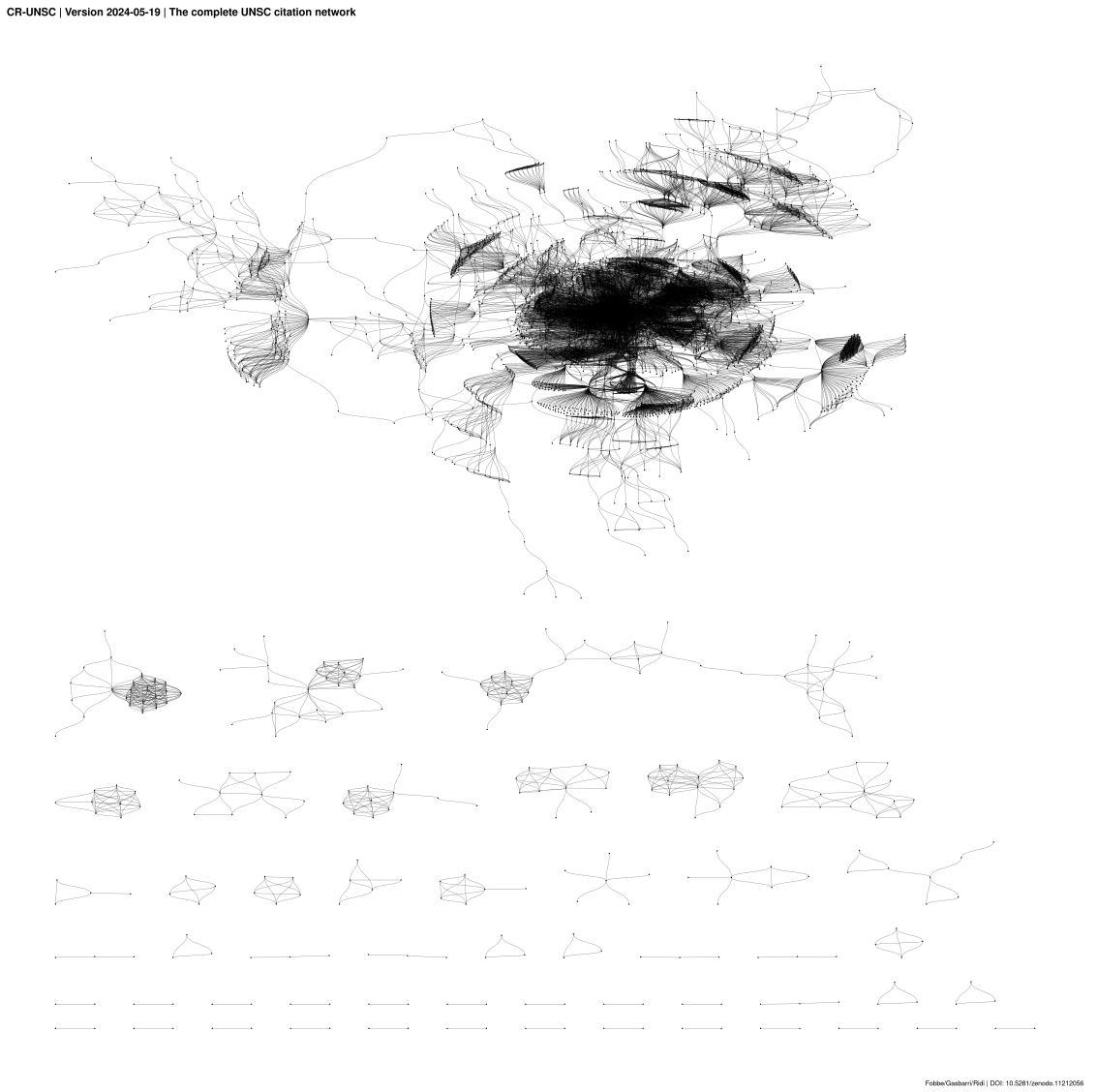
Figure 4: The citation network of the UNSC

7 Linguistic Metrics

7.1 Explanation of Metrics

We provide a number of classic linguistic metrics and visualize their distributions to better communicate the scope of the corpus and its constituent documents:

| Metric	Definition
Characters	Characters roughly correspond to graphemes, the smallest functional unit in a writing system. The word 'judge' is composed of 5 characters, for example.
Tokens	An arbitrary character sequence delimited by whitespace on both sides, e.g. it roughly corresponds to the notion of a 'word'. However, due to its strictly syntactical definition it might also include arbitrary sequences of numbers or special characters.
Types	Unique tokens. If, for example, the token 'human' appeared one hundred times in a given document, it would be counted as only one type.
Sentences	Corresponds approximately to the colloquial definition of a sentence. The exact rules for determining sentence boundaries are very complex and may be reviewed in 'Unicode Standard: Annex No 29'.

7.2 Summary Statistics

Variable	Sum	Min	Quart1	Median	Mean	Quart3	Max
nchars	26,071,742	189	1,914.75	4,669.5	9,578.16	11,515.00	343,887
ntokens	3,704,016	32	310.00	717.0	1,360.77	1,643.75	45,030
ntypes	32,316	28	142.00	265.0	367.16	485.00	5,179
nsentences	43,302	1	6.00	10.0	15.91	19.00	1,313

7.3 Distributions of Document Length

The diagrams in this section are combined violin and box plots. They are especially useful in visualizing distributions of quantitative variables. Their interpretation is fairly straightforward: the greater the area under the curve for a given range, the more frequent the values are in this range. The thick center line of the box indicates the median, the outer lines of the box the first and third quartiles. Whiskers extend outwards to 1.5 times the inter-quartile range (IQR). Outliers beyond 1.5 times IQR are shown as individual points.

Please note that the x-axis is logarithmically scaled, i.e. in powers of 10. It therefore increases in a non-linear fashion. Additional sub-markings are included to assist with interpretation.

CR−UNSC | Version 2024−05−19 | Distributions of Document Length (EN)
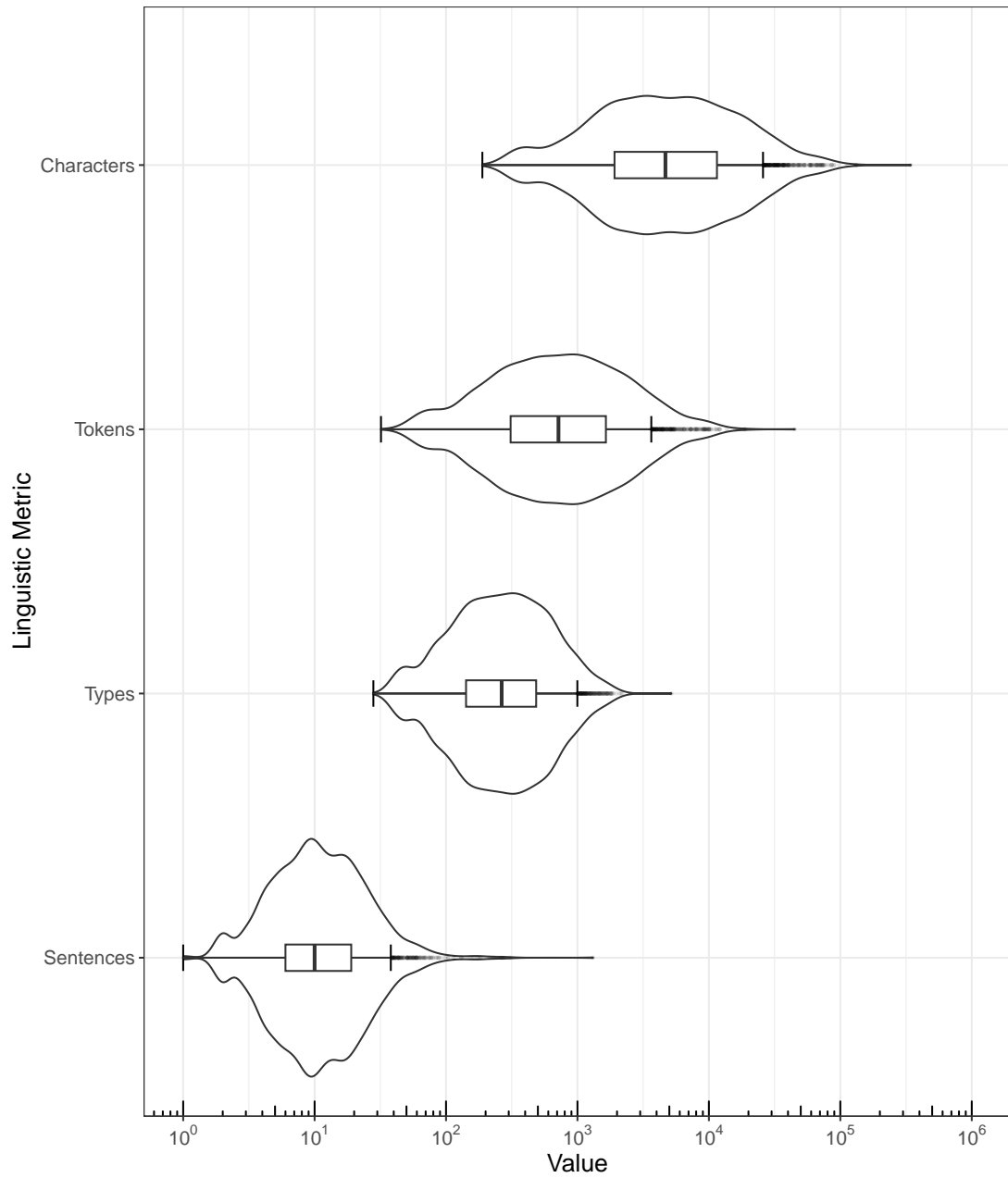
8 Metadata Frequency Tables

8.1 By Year

CR–UNSC | Version 2024–05–19 | Resolutions by Year
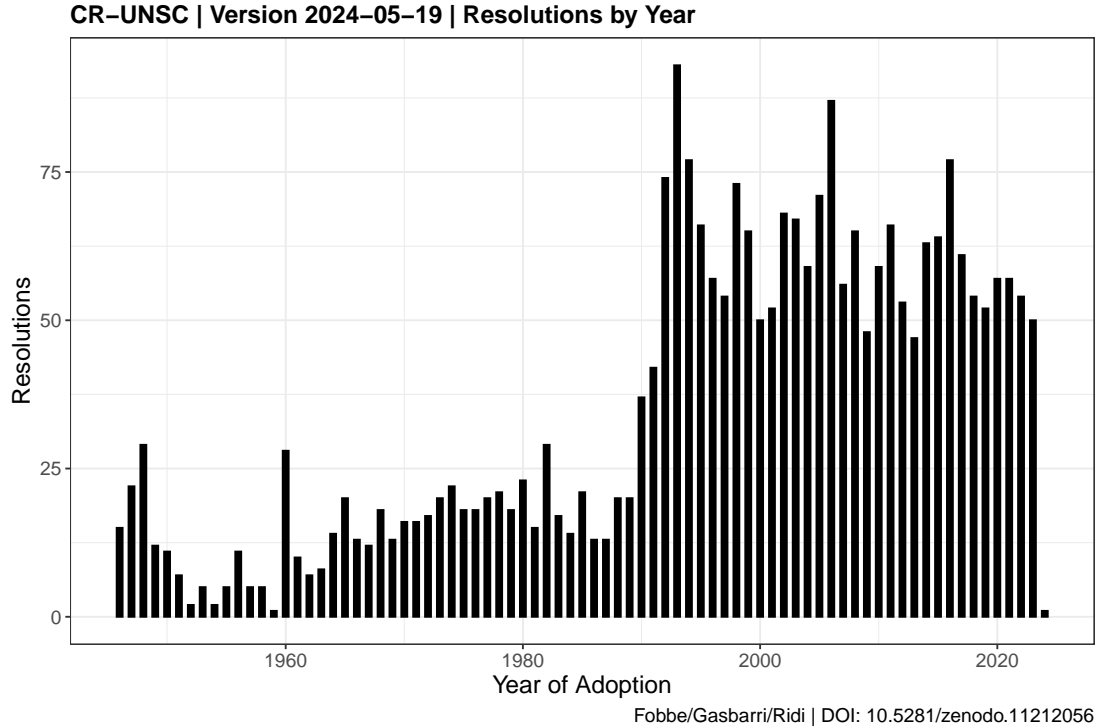
Year	Resolutions	% Total	% Cumulative
1946	15	0.55	0.55
1947	22	0.81	1.36
1948	29	1.07	2.42
1949	12	0.44	2.87
1950	11	0.40	3.27
1951	7	0.26	3.53
1952	2	0.07	3.60
1953	5	0.18	3.78
1954	2	0.07	3.86
1955	5	0.18	4.04
1956	11	0.40	4.45
1957	5	0.18	4.63
1958	5	0.18	4.81
1959	1	0.04	4.85
1960	28	1.03	5.88

Year	Resolutions	% Total	% Cumulative
1961	10	0.37	6.25
1962	7	0.26	6.50
1963	8	0.29	6.80
1964	14	0.51	7.31
1965	20	0.73	8.05
1966	13	0.48	8.52
1967	12	0.44	8.96
1968	18	0.66	9.63
1969	13	0.48	10.10
1970	16	0.59	10.69
1971	16	0.59	11.28
1972	17	0.62	11.90
1973	20	0.73	12.64
1974	22	0.81	13.45
1975	18	0.66	14.11
1976	18	0.66	14.77
1977	20	0.73	15.50
1978	21	0.77	16.27
1979	18	0.66	16.94
1980	23	0.84	17.78
1981	15	0.55	18.33
1982	29	1.07	19.40
1983	17	0.62	20.02
1984	14	0.51	20.54
1985	21	0.77	21.31
1986	13	0.48	21.79
1987	13	0.48	22.26
1988	20	0.73	23.00
1989	20	0.73	23.73
1990	37	1.36	25.09
1991	42	1.54	26.63
1992	74	2.72	29.35
1993	93	3.42	32.77

(continued)

Year	Resolutions	% Total	% Cumulative
1994	77	2.83	35.60
1995	66	2.42	38.02
1996	57	2.09	40.12
1997	54	1.98	42.10
1998	73	2.68	44.78
1999	65	2.39	47.17
2000	50	1.84	49.01
2001	52	1.91	50.92
2002	68	2.50	53.42
2003	67	2.46	55.88
2004	59	2.17	58.05
2005	71	2.61	60.65
2006	87	3.20	63.85
2007	56	2.06	65.91
2008	65	2.39	68.30
2009	48	1.76	70.06
2010	59	2.17	72.23
2011	66	2.42	74.65
2012	53	1.95	76.60
2013	47	1.73	78.32
2014	63	2.31	80.64
2015	64	2.35	82.99
2016	77	2.83	85.82
2017	61	2.24	88.06
2018	54	1.98	90.04
2019	52	1.91	91.95
2020	57	2.09	94.05
2021	57	2.09	96.14
2022	54	1.98	98.13
2023	50	1.84	99.96
2024	1	0.04	100.00
Total	2722	100.00	100.00

8.2 By Country Mentioned

CR–UNSC | Version 2024–05–19 | Top 50 Countries mentioned
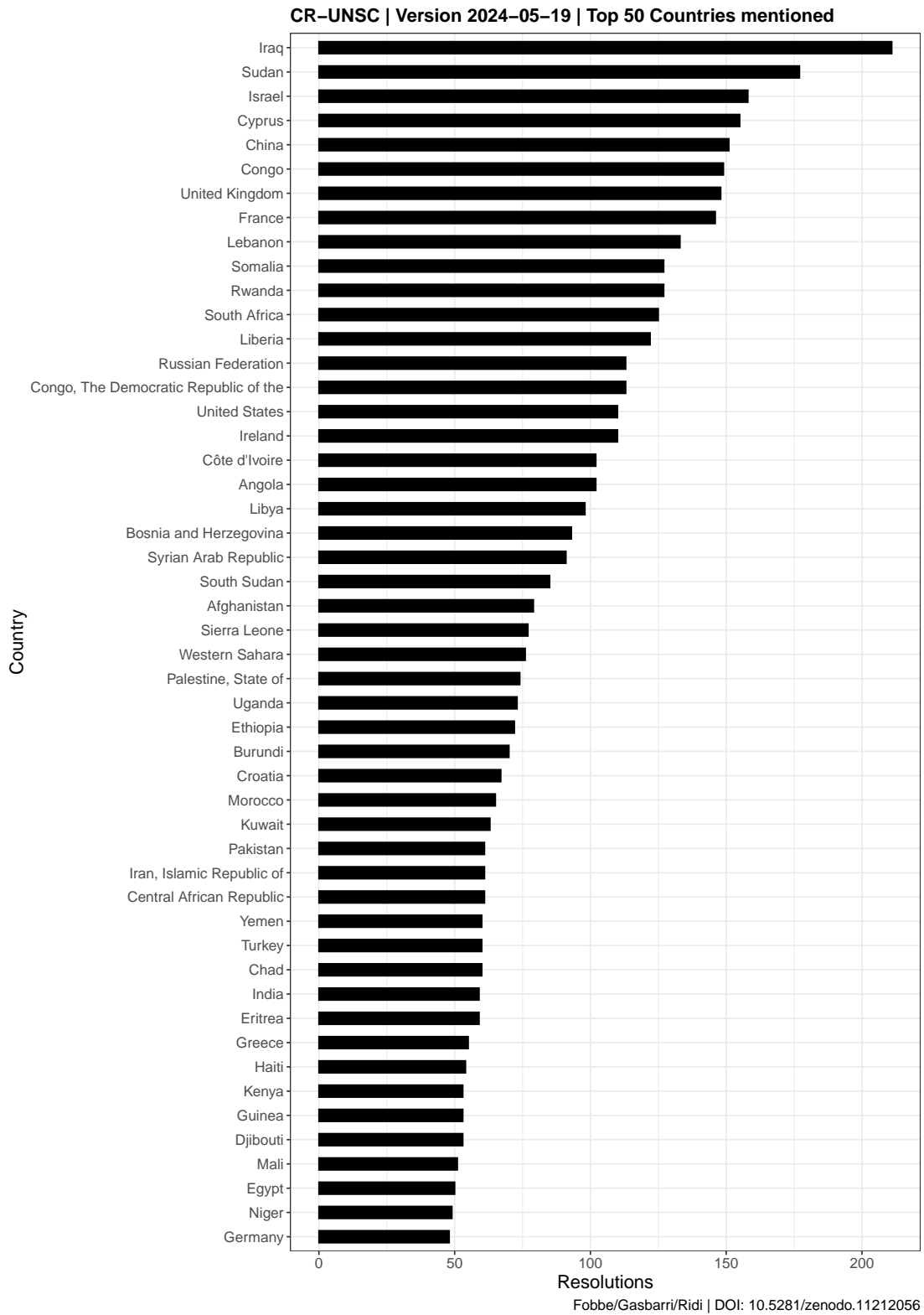
Country	Resolutions	% Total	% Cumulative
Afghanistan	79	1.26	1.26
Albania	23	0.37	1.63
Algeria	27	0.43	2.06
Andorra	1	0.02	2.08
Angola	102	1.63	3.71
Antigua and Barbuda	1	0.02	3.73
Argentina	19	0.30	4.03
Armenia	5	0.08	4.11
Australia	21	0.34	4.45
Austria	8	0.13	4.57
Azerbaijan	6	0.10	4.67
Bahamas	10	0.16	4.83
Bahrain	4	0.06	4.89
Bangladesh	4	0.06	4.96
Barbados	1	0.02	4.97
Belgium	30	0.48	5.45
Belize	1	0.02	5.47
Benin	17	0.27	5.74
Bhutan	1	0.02	5.76
Bonaire, Sint Eustatius and Saba	2	0.03	5.79
Bosnia and Herzegovina	93	1.49	7.28
Botswana	11	0.18	7.45
Brazil	3	0.05	7.50
Brunei Darussalam	1	0.02	7.52
Bulgaria	10	0.16	7.68
Burkina Faso	19	0.30	7.98
Burundi	70	1.12	9.10
Cabo Verde	1	0.02	9.11
Cambodia	17	0.27	9.39
Cameroon	17	0.27	9.66

Country	Resolutions	% Total	% Cumulative
Canada	21	0.34	9.99
Central African Republic	61	0.98	10.97
Chad	60	0.96	11.93
Chile	4	0.06	11.99
China	151	2.41	14.41
Colombia	32	0.51	14.92
Comoros	2	0.03	14.95
Congo	149	2.38	17.33
Congo, The Democratic Republic of the	113	1.81	19.14
Costa Rica	3	0.05	19.19
Croatia	67	1.07	20.26
Cuba	26	0.42	20.67
Cyprus	155	2.48	23.15
Côte d'Ivoire	102	1.63	24.78
Denmark	21	0.34	25.12
Djibouti	53	0.85	25.97
Dominica	6	0.10	26.06
Dominican Republic	4	0.06	26.13
Ecuador	3	0.05	26.18
Egypt	50	0.80	26.97
El Salvador	12	0.19	27.17
Equatorial Guinea	2	0.03	27.20
Eritrea	59	0.94	28.14
Estonia	2	0.03	28.17
Ethiopia	72	1.15	29.33
Falkland Islands (Malvinas)	2	0.03	29.36
Fiji	3	0.05	29.41
Finland	8	0.13	29.53
France	146	2.33	31.87

Country	Resolutions	% Total	% Cumulative
Gabon	17	0.27	32.14
Gambia	3	0.05	32.19
Georgia	42	0.67	32.86
Germany	48	0.77	33.63
Ghana	24	0.38	34.01
Greece	55	0.88	34.89
Grenada	1	0.02	34.91
Guatemala	3	0.05	34.95
Guinea	53	0.85	35.80
Guinea-Bissau	18	0.29	36.09
Guyana	7	0.11	36.20
Haiti	54	0.86	37.06
Honduras	3	0.05	37.11
Hungary	13	0.21	37.32
Iceland	1	0.02	37.34
India	59	0.94	38.28
Indonesia	30	0.48	38.76
Iran, Islamic Republic of	61	0.98	39.73
Iraq	211	3.37	43.11
Ireland	110	1.76	44.87
Israel	158	2.53	47.39
Italy	42	0.67	48.07
Jamaica	15	0.24	48.31
Japan	38	0.61	48.91
Jordan	32	0.51	49.42
Kazakhstan	4	0.06	49.49
Kenya	53	0.85	50.34
Kiribati	1	0.02	50.35
Korea, Democratic People's Republic of	12	0.19	50.54
Korea, Republic of	40	0.64	51.18

Country	Resolutions	% Total	% Cumulative
Kuwait	63	1.01	52.19
Kyrgyzstan	3	0.05	52.24
Latvia	4	0.06	52.30
Lebanon	133	2.13	54.43
Lesotho	9	0.14	54.57
Liberia	122	1.95	56.52
Libya	98	1.57	58.09
Liechtenstein	2	0.03	58.12
Lithuania	1	0.02	58.14
Luxembourg	8	0.13	58.27
Madagascar	27	0.43	58.70
Malawi	4	0.06	58.76
Malaysia	19	0.30	59.07
Mali	51	0.82	59.88
Malta	13	0.21	60.09
Marshall Islands	2	0.03	60.12
Mauritania	15	0.24	60.36
Mauritius	13	0.21	60.57
Mexico	8	0.13	60.70
Micronesia, Federated States of	2	0.03	60.73
Moldova, Republic of	2	0.03	60.76
Monaco	1	0.02	60.78
Mongolia	4	0.06	60.84
Montenegro	34	0.54	61.38
Morocco	65	1.04	62.42
Mozambique	26	0.42	62.84
Myanmar	3	0.05	62.89
Namibia	48	0.77	63.66
Nauru	2	0.03	63.69
Nepal	11	0.18	63.86
Netherlands	33	0.53	64.39

Country	Resolutions	% Total	% Cumulative
New Zealand	2	0.03	64.42
Nicaragua	12	0.19	64.61
Niger	49	0.78	65.40
Nigeria	41	0.66	66.05
Northern Mariana Islands	1	0.02	66.07
Norway	18	0.29	66.36
Oman	3	0.05	66.41
Pakistan	61	0.98	67.38
Palau	3	0.05	67.43
Palestine, State of	74	1.18	68.61
Panama	8	0.13	68.74
Paraguay	2	0.03	68.77
Peru	3	0.05	68.82
Philippines	5	0.08	68.90
Poland	30	0.48	69.38
Portugal	40	0.64	70.02
Qatar	20	0.32	70.34
Romania	4	0.06	70.40
Russian Federation	113	1.81	72.21
Rwanda	127	2.03	74.24
Saint Kitts and Nevis	7	0.11	74.35
Saint Lucia	1	0.02	74.37
Samoa	1	0.02	74.38
San Marino	2	0.03	74.42
Sao Tome and Principe	1	0.02	74.43
Saudi Arabia	6	0.10	74.53
Senegal	23	0.37	74.90
Serbia	34	0.54	75.44
Seychelles	20	0.32	75.76
Sierra Leone	77	1.23	76.99

Country	Resolutions	% Total	% Cumulative
Singapore	3	0.05	77.04
Slovakia	1	0.02	77.05
Slovenia	4	0.06	77.12
Solomon Islands	1	0.02	77.13
Somalia	127	2.03	79.17
South Africa	125	2.00	81.16
South Sudan	85	1.36	82.52
Spain	23	0.37	82.89
Sri Lanka	8	0.13	83.02
Sudan	177	2.83	85.85
Sweden	22	0.35	86.20
Switzerland	14	0.22	86.42
Syrian Arab Republic	91	1.46	87.88
Tajikistan	24	0.38	88.26
Tanzania, United Republic of	46	0.74	89.00
Thailand	3	0.05	89.05
Timor-Leste	15	0.24	89.29
Togo	12	0.19	89.48
Tonga	1	0.02	89.49
Trinidad and Tobago	10	0.16	89.65
Tunisia	11	0.18	89.83
Turkey	60	0.96	90.79
Turkmenistan	4	0.06	90.85
Tuvalu	1	0.02	90.87
Uganda	73	1.17	92.04
Ukraine	9	0.14	92.18
United Arab Emirates	5	0.08	92.26
United Kingdom	148	2.37	94.63
United States	110	1.76	96.39
Uruguay	2	0.03	96.42

(continued)

Country	Resolutions	% Total	% Cumulative
Uzbekistan	6	0.10	96.51
Vanuatu	1	0.02	96.53
Venezuela, Bolivarian Republic of	4	0.06	96.59
Viet Nam	1	0.02	96.61
Virgin Islands, British	1	0.02	96.63
Western Sahara	76	1.22	97.84
Yemen	60	0.96	98.80
Zambia	38	0.61	99.41
Zimbabwe	37	0.59	100.00
Total	6254	100.00	100.00

8.3 By Region of Countries Mentioned

CR–UNSC | Version 2024–05–19 | Top Regions of Countries Mentioned
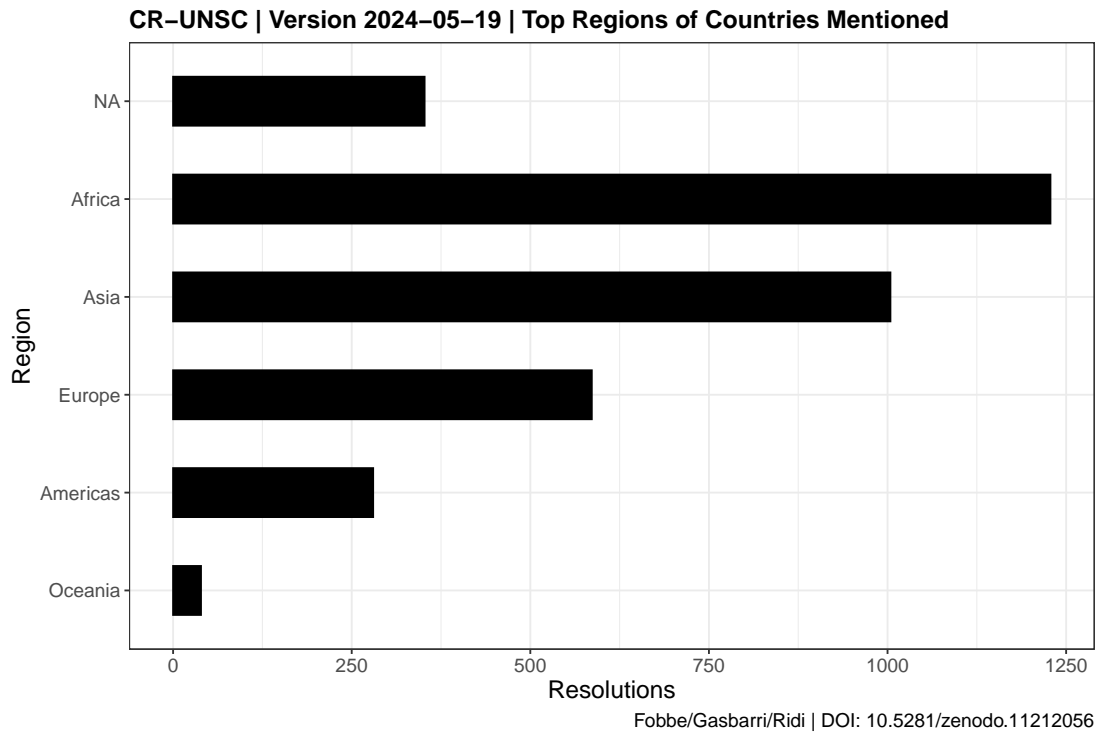
Region	Resolutions	% Total	% Cumulative
NA	352	10.09	10.09
Africa	1228	35.20	45.29
Americas	280	8.03	53.31
Asia	1004	28.78	82.09
Europe	586	16.80	98.88
Oceania	39	1.12	100.00
Total	3489	100.00	100.00

8.4 By Sub-Region of Countries Mentioned

CR–UNSC | Version 2024–05–19 | Top Sub–Regions Mentioned
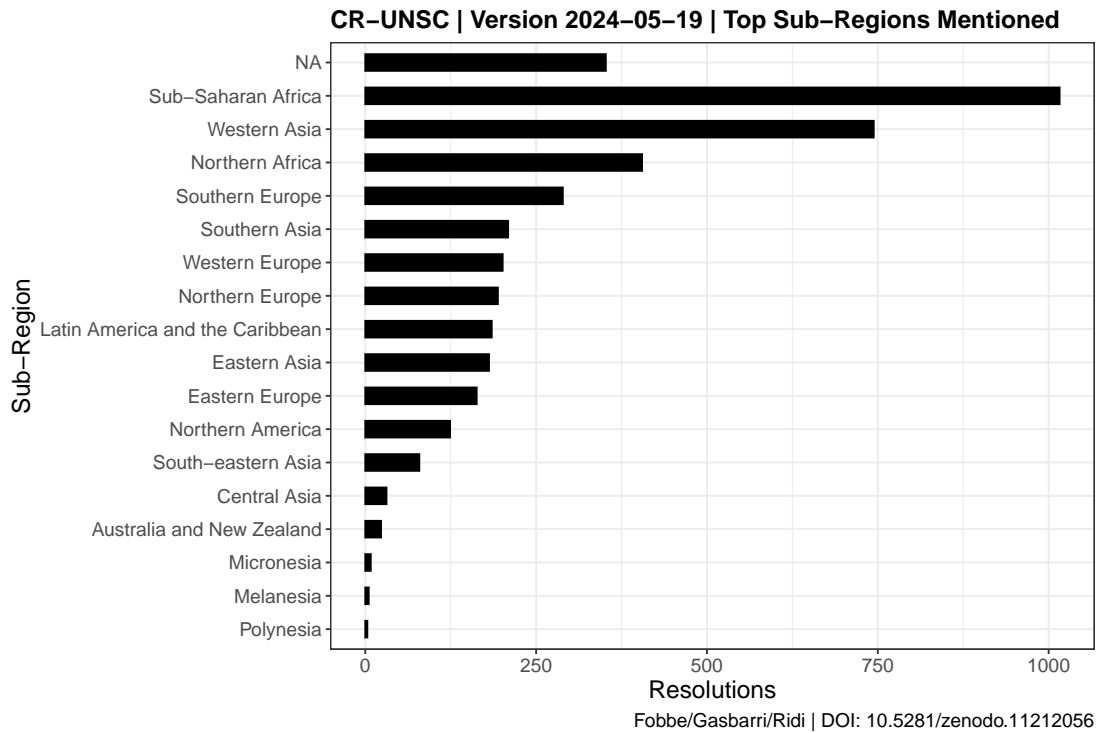
Sub-Region	Resolutions	% Total	% Cumulative
NA	352	8.36	8.36
Australia and New Zealand	23	0.55	8.90
Central Asia	31	0.74	9.64
Eastern Asia	181	4.30	13.94
Eastern Europe	163	3.87	17.81
Latin America and the Caribbean	185	4.39	22.20
Melanesia	5	0.12	22.32
Micronesia	8	0.19	22.51
Northern Africa	405	9.62	32.12
Northern America	124	2.94	35.07
Northern Europe	194	4.61	39.67
Polynesia	3	0.07	39.74
South-eastern Asia	79	1.88	41.62

Sub-Region	Resolutions	% Total	% Cumulative
Southern Asia	209	4.96	46.58
Southern Europe	289	6.86	53.44
Sub-Saharan Africa	1016	24.12	77.56
Western Asia	744	17.66	95.23
Western Europe	201	4.77	100.00
Total	4212	100.00	100.00

8.5 By Subject

CR–UNSC | Version 2024–05–19 | 50 Most Frequent Subjects
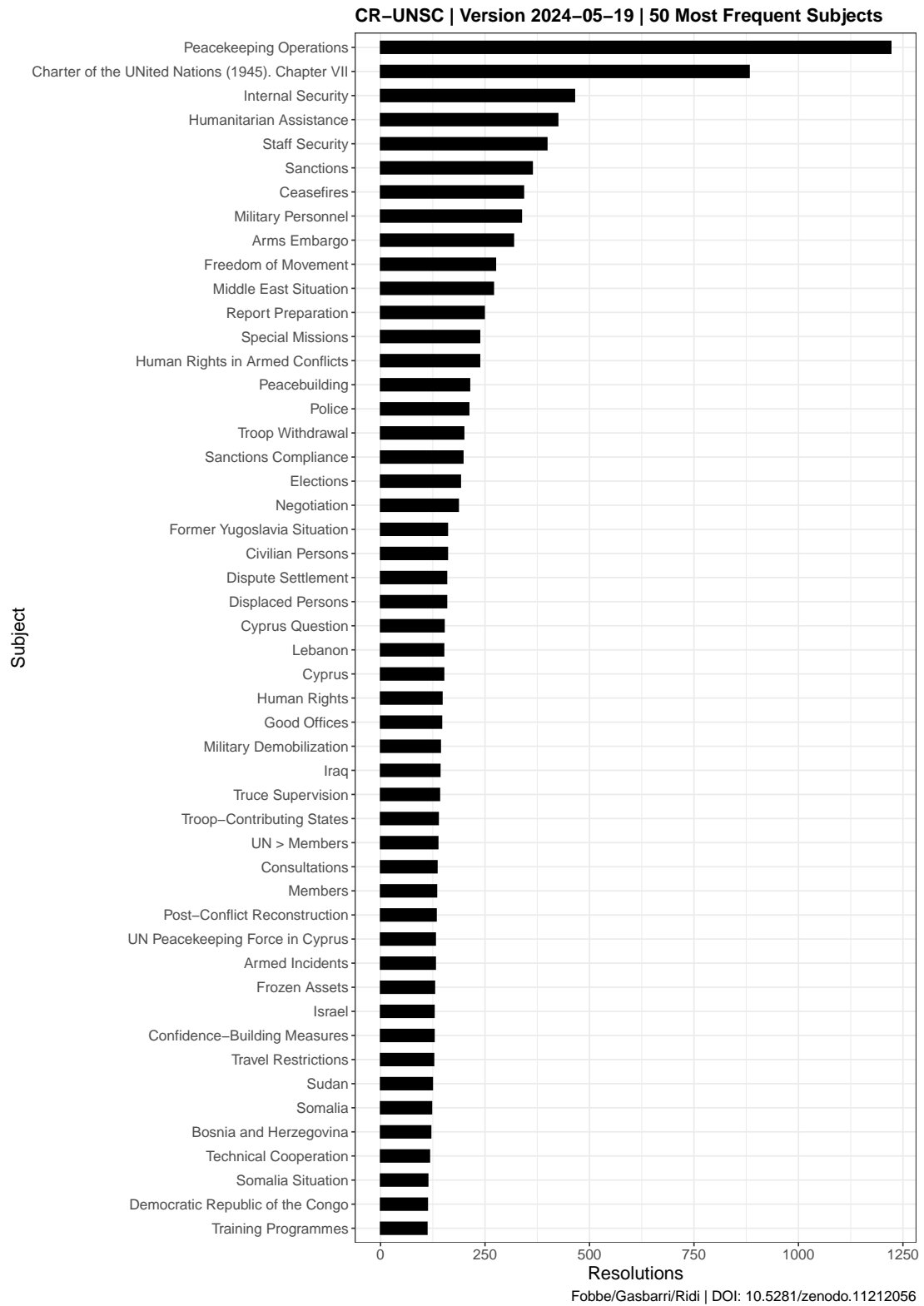
Subject	Resolutions	% Total	% Cumulative
Peacekeeping Operations	1221	3.68	62.16
Charter of the UNited Nations (1945). Chapter VII	882	2.66	19.87
Internal Security	464	1.40	41.27
Humanitarian Assistance	424	1.28	37.54
Staff Security	398	1.20	78.19
Sanctions	363	1.10	72.23
Ceasefires	342	1.03	10.52
Military Personnel	337	1.02	53.32
Arms Embargo	318	0.96	4.48
Freedom of Movement	275	0.83	30.62
Middle East Situation	270	0.81	51.28
Report Preparation	248	0.75	69.68
Human Rights in Armed Conflicts	237	0.72	35.83
Special Missions	237	0.72	76.88
Peacebuilding	213	0.64	58.47
Police	211	0.64	63.67
Troop Withdrawal	199	0.60	83.57
Sanctions Compliance	197	0.59	72.82
Elections	191	0.58	26.17
Negotiation	186	0.56	55.31
Civilian Persons	160	0.48	12.02
Former Yugoslavia Situation	160	0.48	29.64
Displaced Persons	158	0.48	23.52

Subject	Resolutions	% Total	% Cumulative
Dispute Settlement	158	0.48	23.99
Cyprus Question	152	0.46	16.94
Cyprus	151	0.46	16.48
Lebanon	151	0.46	47.61
Human Rights	147	0.44	35.10
Good Offices	146	0.44	32.92
Military Demobilization	143	0.43	52.11
Iraq	142	0.43	43.13
Truce Supervision	141	0.43	84.41
Troop-Contributing States	138	0.42	83.99
UN > Members	137	0.41	85.84
Consultations	135	0.41	13.70
Members	134	0.40	50.11
Post-Conflict Reconstruction	133	0.40	65.05
Armed Incidents	131	0.40	3.45
UN Peacekeeping Force in Cyprus	131	0.40	90.74
Frozen Assets	129	0.39	31.06
Confidence-Building Measures	128	0.39	13.03
Israel	128	0.39	43.80
Travel Restrictions	127	0.38	82.48
Sudan	124	0.37	78.81
Somalia	122	0.37	74.82
Bosnia and Herzegovina	120	0.36	7.75
Technical Cooperation	117	0.35	80.48
Somalia Situation	113	0.34	75.16

Subject	Resolutions	% Total	% Cumulative
Democratic Republic of the Congo	112	0.34	21.81
Training Programmes	111	0.33	81.88

8.6 Proportions of Votes Cast

CR−UNSC | Version 2024−05−19 | Voting Data

Fobbe/Gasbarri/Ridi | DOI: 10.5281/zenodo.11212056
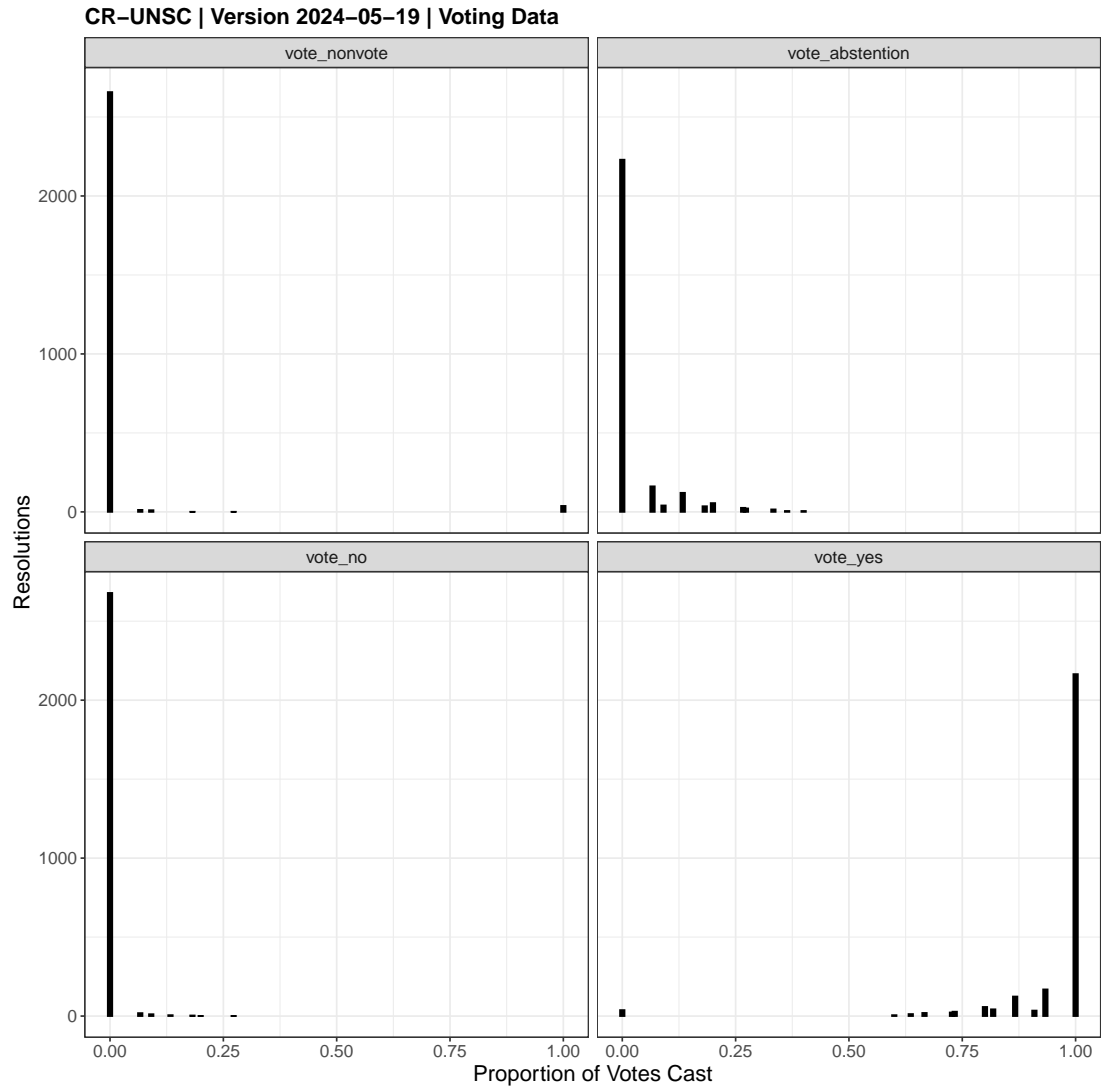
Variable	Proportion	N
vote_nonvote	0.0000000	2658
vote_nonvote	0.0666667	13
vote_nonvote	0.0909091	11
vote_nonvote	0.1818182	1
vote_nonvote	0.2727273	1
vote_nonvote	1.0000000	38
vote_abstention	0.0000000	2230
vote_abstention	0.0666667	162
vote_abstention	0.0909091	41

49

(continued)

Variable	Proportion	N
vote_abstention	0.1333333	121
vote_abstention	0.1818182	36
vote_abstention	0.2000000	56
vote_abstention	0.2666667	26
vote_abstention	0.2727273	22
vote_abstention	0.3333333	16
vote_abstention	0.3636364	6
vote_abstention	0.4000000	6
vote_no	0.0000000	2679
vote_no	0.0666667	19
vote_no	0.0909091	12
vote_no	0.1333333	6
vote_no	0.1818182	4
vote_no	0.2000000	1
vote_no	0.2727273	1
vote_yes	0.0000000	38
vote_yes	0.6000000	6
vote_yes	0.6363636	13
vote_yes	0.6666667	20
vote_yes	0.7272727	23
vote_yes	0.7333333	28
vote_yes	0.8000000	58
vote_yes	0.8181818	43
vote_yes	0.8666667	124
vote_yes	0.9090909	35
vote_yes	0.9333333	169
vote_yes	1.0000000	2165

9   FAIR Data Conformity Declaration

The FAIR Principles lay out an influential set of guiding standards for the management of scientific data. Four principles define the FAIR standard: [F]indability, [A]ccessibility [I]nteroperability and [R]eusability.

It is the intention of the author(s) to be fully compliant with the FAIR standard. This conformity declaration is intended to provide further information on what measures have been taken to achieve this compliance.

For the original publication proposing the FAIR standard, please see: Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

9.1   [F] Findable

[F1] (Meta)data are assigned a globally unique and persistent identifier

- The dataset and source code are published with Zenodo, a major scientific repository operated by CERN.
- Zenodo issues a DOI for the dataset as a whole and a DOI for each individual version of the dataset and source code.

[F2] Data are described with rich metadata (defined by R1 below)

- Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms
- The author(s) have provided significant metadata to augment the Zenodo record, such as a lengthy description, keywords, publication details, license, related works and contact details for the maintainer(s)

[F3] Metadata clearly and explicitly include the identifier of the data they describe

- Zenodo requires the DOI as a top-level and mandatory field in the metadata of each record

[F4] (Meta)data are registered or indexed in a searchable resource

- Metadata is indexed in and searchable via Zenodo
- Metadata is indexed in and searchable via DataCite

9.2   [A] Accessible

[A1] (Meta)data are retrievable by their identifier using a standardised communications protocol

- Zenodo provides a public OAI-PMH interface to retrieve metadata of individual records and whole collections
- Zenodo provides a public REST API to retrieve metadata of individual records

[A1.1] The protocol is open, free, and universally implementable

- The OAI-PMH and REST APIs provided by Zenodo are based on open, free and universally implementable protocols

[A1.2] The protocol allows for an authentication and authorisation procedure, where necessary

- Zenodo provides all metadata publicly and without authentication under a public domain license

[A2] Metadata are accessible, even when the data are no longer available

- Zenodo guarantees that metadata remain available for the lifetime of the parent organization CERN, which has a defined experimental program for the next 20 years
- Metadata are stored in different servers than the data at Zenodo
- Zenodo commits to providing a "data tombstone" of the record that contains all metadata even if data becomes unavailable for legal or other reasons

9.3 [I] Interoperable

[I1] (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

- Zenodo metadata is internally stored as JSON and can be exported to MARCXML, Dublin Core XML, CSL, DataCite JSON, DataCite XML, BibTeX, GeoJSON and DCAT

[I2] (Meta)data use vocabularies that follow FAIR principles

- Zenodo refers to external vocabularies for license (Open Definition), funders (FundRef) and grants (OpenAIRE)

[I3] (Meta)data include qualified references to other (meta)data

- The CR-UNSC embeds the Concept DOI and Version DOI for the relevant Zenodo record in the CSV dataset and the Codebook
- The CR-UNSC includes the unique URL for the resolution text in all six UN languages, the URL for the draft and meeting record in English and the URL for the UN Digital Library record pages of the resolution, draft and meeting record. UN Digital Library record pages provide machine-readable metadata in a wide range of export formats.

9.4 [R] Reusable

[R1] (Meta)data are richly described with a plurality of accurate and relevant attributes

[R1.1] (Meta)data are released with a clear and accessible data usage license

- Zenodo metadata includes standardized representations of each license compliant with the OpenDefinition standard
- The CR-UNSC data is published under a standardized Creative Commons CC0 1.0 waiver
- The CR-UNSC source code is published under a standardized GNU GPL v3 license
- The CR-UNSC embeds the data license in the CSV file and in the Codebook

[R1.2] (Meta)data are associated with detailed provenance

- The CR-UNSC embeds the Concept DOI and Version DOI for the relevant Zenodo record in the CSV dataset and the Codebook
- The CR-UNSC includes the unique URL for the resolution text in all six UN languages, the URL for the draft and meeting record in English and the URL for the UN Digital Library record pages of the resolution, draft and meeting record

[R1.3] (Meta)data meet domain-relevant community standards

- The CR-UNSC was crafted by specialized international lawyers, complies with international law terminology and respects international legal usage wherever possible
- The CR-UNSC respects United Nations terminology and United Nations document symbols
- The CR-UNSC adapts all metadata conventions from the UN Digital Library
- The CR-UNSC complies with ISO and UN M49 standards for country and area codes

10 Cryptographic Signatures

10.1 Two Phase Signature

The integrity and veracity of the ZIP archives and other output of the data set are guaranteed with two-phase *hash-and-sign* signature process.

In **Phase I** the data pipeline automatically computes digital fingerprints (hashes) with two different secure algorithms (SHA2-256 und SHA3-512) and stores these in a CSV file.

In **Phase II** the CSV files with hashes is manually signed with a secret GPG key.

This process ensures that the compilation of the data set can be performed by anyone, such as in scientific replication studies, but provides the same guarantees as signing the full data set would.

10.2 Personal GPG Key

The CSV file is signed with the personal GPG key of the first author. The public key is included with the data set and source code publication on Zenodo, as well as the project GitHub repository. The key is identified as follows:

Name: Sean Fobbe (fobbe-data@posteo.de)

Fingerprint: FE6F B888 F0E5 656C 1D25 3B9A 50C4 1384 F44A 4E42

11   Changelog

11.1   Version 2024-05-19

- New variant: EN_TXT_BEST containing a write-out of the English resolution texts equivalent to the CSV file text variable
- New diagrams: bar charts of top M49 regions and sub-regions of countries mentioned in resolution texts
- Fixed naming mix-up of BIBTEX and GRAPHML zip archives
- Fixed whitespace character detection in citation extraction (adds ca. 10% more citations)
- Fixed improper merging of weights in citation network
- Fixed "cannot xtfrm data frames" warning
- Improve REGEX detection for certain geographic entities
- Improve Codebook (headings, citation network docs)

11.2   Version 2024-05-03

- Initial Release

12 Replication Parameters

```
## [1] "OpenSSL 3.0.2 15 Mar 2022 (Library: OpenSSL 3.0.2 15 Mar 2022)"
```

```
## R version 4.2.2 (2022-10-31)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 22.04.2 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.20.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] data.table_1.14.8        future.apply_1.10.0
##  [3] future_1.32.0            textcat_1.0-8
##  [5] quanteda.textplots_0.94.2 quanteda.textstats_0.96.1
##  [7] quanteda_3.2.4           stringi_1.7.12
##  [9] readtext_0.81            pdftools_3.3.3
## [11] RColorBrewer_1.1-3       viridis_0.6.2
## [13] viridisLite_0.4.1        magick_2.7.4
## [15] scales_1.2.1             igraph_1.4.1
## [17] ggraph_2.1.0             ggplot2_3.4.1
## [19] kableExtra_1.3.4         knitr_1.42
## [21] rvest_1.0.3              testthat_3.1.7
## [23] zip_2.2.2                RcppTOML_0.2.2
## [25] tarchetypes_0.7.5        targets_0.14.3
##
## loaded via a namespace (and not attached):
##  [1] fs_1.6.1          bit64_4.0.5       webshot_0.5.4
##  [4] httr_1.4.5        rprojroot_2.0.3   future.callr_0.8.1
##  [7] tools_4.2.2       backports_1.4.1   utf8_1.2.3
## [10] R6_2.5.1          colorspace_2.1-0  withr_2.5.0
## [13] tidyselect_1.2.0  gridExtra_2.3     processx_3.8.0
## [16] bit_4.0.5         compiler_4.2.2    cli_3.6.0
## [19] humaniformat_0.6.0 xml2_1.3.3       desc_1.4.2
## [22] labeling_0.4.2    stringfish_0.15.7 slam_0.1-50
## [25] callr_3.7.3       askpass_1.1       systemfonts_1.0.4
## [28] stringr_1.5.0     digest_0.6.31     rmarkdown_2.20
## [31] svglite_2.1.1     tau_0.0-24        pkgconfig_2.0.3
## [34] htmltools_0.5.4   parallelly_1.34.0 fastmap_1.1.1
## [37] rlang_1.0.6       rstudioapi_0.14   RApiSerialize_0.1.2
## [40] farver_2.1.1      generics_0.1.3    dplyr_1.1.0
```

```
## [43] magrittr_2.0.3      Matrix_1.5-1         waldo_0.4.0
## [46] Rcpp_1.0.10         munsell_0.5.0       fansi_1.0.4
## [49] lifecycle_1.0.3     furrr_0.3.1         yaml_2.3.7
## [52] MASS_7.3-58.1       brio_1.1.3          grid_4.2.2
## [55] parallel_4.2.2      listenv_0.9.0       ggrepel_0.9.3
## [58] lattice_0.20-45     graphlayouts_0.8.4  ps_1.7.2
## [61] pillar_1.8.1        base64url_1.4       pkgload_1.3.2
## [64] codetools_0.2-18    stopwords_2.3       fastmatch_1.1-3
## [67] glue_1.6.2          evaluate_0.20       qpdf_1.3.0
## [70] RcppParallel_5.1.7  vctrs_0.5.2         tweenr_2.0.2
## [73] gtable_0.3.1        purrr_1.0.1         polyclip_1.10-4
## [76] tidyr_1.3.0         bib2df_1.1.1        qs_0.25.5
## [79] xfun_0.37           ggforce_0.4.1       tidygraph_1.2.3
## [82] nsyllable_1.0.1     tibble_3.2.0        globals_0.16.2
```

Bibliography

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2023. *Rmarkdown: Dynamic Documents for R.*

Arel-Bundock, Vincent. 2022. *Countrycode: Convert Country Names and Country Codes.* https://vincentarelbundock.github.io/countrycode/.

Arel-Bundock, Vincent, Nils Enevoldsen, and CJ Yetman. 2018. "Countrycode: An R Package to Convert Country Names and Country Codes." *Journal of Open Source Software* 3 (28): 848. https://doi.org/10.21105/joss.00848.

Bengtsson, Henrik. 2021. "A Unifying Framework for Parallel and Distributed Processing in R Using Futures." *The R Journal* 13 (2): 208–27. https://doi.org/10.32614/RJ-2021-048.

———. 2022. *Future.apply: Apply Function to Elements in Parallel Using Futures.*

———. 2023. *Future: Unified Parallel and Distributed Processing in R for Everyone.*

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3 (30): 774. https://doi.org/10.21105/joss.00774.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, Akitaka Matsuo, and William Lowe. 2022. *Quanteda: Quantitative Analysis of Textual Data.* https://quanteda.io.

Buchta, Christian, and Kurt Hornik. 2022. *ISOcodes: Selected Iso Codes.*

Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal* Complex Systems: 1695. https://igraph.org.

Csárdi, Gábor, Kuba Podgórski, and Rich Geldreich. 2022. *Zip: Cross-Platform Zip Compression.* https://github.com/r-lib/zip#readme.

Dowle, Matt, and Arun Srinivasan. 2023. *Data.table: Extension of 'Data.frame'.*

Eddelbuettel, Dirk. 2023. *RcppTOML: Rcpp Bindings to Parser for "Tom's Obvious Markup Language".* http://dirk.eddelbuettel.com/code/rcpp.toml.html.

file., See AUTHORS. 2023. *Igraph: Network Analysis and Visualization.*

Gagolewski, Marek. 2022. "stringi: Fast and Portable Character String Processing in R." *Journal of Statistical Software* 103 (2): 1–59. https://doi.org/10.18637/jss.v103.i02.

Gagolewski, Marek, Bartek Tartanus, others; Unicode, Inc., and others. 2023. *Stringi: Fast and Portable Character String Processing Facilities.*

Landau, William Michael. 2021a. *Tarchetypes: Archetypes for Targets.*

———. 2021b. "The Targets R Package: A Dynamic Make-Like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing." *Journal of Open Source Software* 6 (57): 2959. https://doi.org/10.21105/joss.02959.

———. 2023a. *Tarchetypes: Archetypes for Targets.*

———. 2023b. *Targets: Dynamic Function-Oriented Make-Like Declarative Pipelines.*

Ooms, Jeroen. 2023. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents.*

Ottolinger, Philipp. 2019. *Bib2df: Parse a Bibtex File to a Data Frame.* https://github.com/ropensci/bib2df.

Pedersen, Thomas Lin. 2022. *Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks.*

Ushey, Kevin. 2023. *Renv: Project Environments.* https://rstudio.github.io/renv/.

Wickham, Hadley. 2022. *Rvest: Easily Harvest (Scrape) Web Pages.*

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

———. 2015. *Dynamic Documents with R and Knitr.* 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. https://yihui.org/knitr/.

———. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in R.* https://yihui.org/knitr/.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide.* Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook.* Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook.

Zhu, Hao. 2021. *KableExtra: Construct Complex Table with Kable and Pipe Syntax.*