

# Corpus des Deutschen Bundesrechts (C-DBR)

CODEBOOK

Version 2021-01-05



DOI: [10.5281/zenodo.4384771](https://doi.org/10.5281/zenodo.4384771)

<b>Titel</b>	Corpus des Deutschen Bundesrechts
<b>Abkürzung</b>	C-DBR
<b>Autor</b>	Seán Fobbe
<b>Version</b>	2021-01-05
<b>Download</b>	<a href="https://doi.org/10.5281/zenodo.4384771">https://doi.org/10.5281/zenodo.4384771</a>
<b>Lizenz</b>	CC0 1.0 Universal

### Zitiervorschlag

*Seán Fobbe* (2021). Corpus des Deutschen Bundesrechts (C-DBR). Version 2021-01-05. Zenodo. DOI: [10.5281/zenodo.4384771](https://doi.org/10.5281/zenodo.4384771).

### Digital Object Identifier (DOI): Concept DOI und Version DOI

Soweit nicht anders angegeben ist die DOI immer eine »Version DOI« und bezieht sich nur auf eine bestimmte Version des Datensatzes. Sie verweist daher nur auf Version 2021-01-05. Für das Gesamtkonzept dieses Datensatzes steht eine »Concept DOI« zur Verfügung, die auf der Zenodo-Seite jeder Version unter »Cite all versions?« zu finden ist. Sie lautet [10.5281/zenodo.3832111](https://doi.org/10.5281/zenodo.3832111). Die »Concept DOI« verlinkt immer die aktuellste Version.

### Urheberrecht

Der Datensatz und dieses Dokument sind unter einer **Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication Lizenz** veröffentlicht. Ich stelle den Datensatz und das Codebook vollständig gemeinfrei und verzichte weltweit auf alle damit verbundenen Urheberrechte, einschließlich aller ähnlichen Rechte, soweit dies gesetzlich möglich ist.

Sie können die Werke kopieren, modifizieren, verteilen und aufführen ohne um Erlaubnis bitten zu müssen, selbst für kommerzielle Zwecke. Patente und Markenschutzrechte bleiben von CC0 unberührt. CC0 hat auch keine Auswirkungen auf etwaige Datenschutz- oder Persönlichkeitsrechte. Jegliche Haftung für die Benutzung dieses Werkes ist ausgeschlossen, bis zu dem maximalen Umfang in dem dies gesetzlich möglich ist.

Wenn Sie diese Werke nutzen oder zitieren sollten Sie nicht den Eindruck erwecken, der Autor unterstütze ihre Nutzung.

Dies ist nur eine unverbindliche deutsche Zusammenfassung der Lizenz, den vollständigen und rechtsverbindlichen Lizenztext finden Sie hier: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>

### Disclaimer

Dieser Datensatz ist eine private wissenschaftliche Initiative und steht in keiner Verbindung zu Behörden, Gerichten oder anderen amtlichen Stellen der Bundesrepublik Deutschland.

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>5</b>
<b>2</b>	<b>Nutzung</b>	<b>6</b>
2.1	CSV-Dateien . . . . .	6
2.2	XML-Dateien . . . . .	6
2.3	TXT-Dateien . . . . .	6
<b>3</b>	<b>Konstruktion</b>	<b>7</b>
3.1	Datenquellen . . . . .	7
3.2	Beschreibung des Datensatzes . . . . .	7
3.3	Sammlung der Daten . . . . .	7
3.4	Source Code und Compilation Report . . . . .	7
3.5	Grenzen des Datensatzes . . . . .	8
3.6	Urheberrechtsfreiheit von Rohdaten und Datensatz . . . . .	8
3.7	Metadaten . . . . .	8
3.7.1	Schema für die Dateinamen (PDF, TXT, EPUB) . . . . .	8
3.7.2	Beispiel eines Dateinamens . . . . .	9
3.8	Qualitätsprüfung . . . . .	9
<b>4</b>	<b>Varianten und Zielgruppen</b>	<b>10</b>
<b>5</b>	<b>Variablen</b>	<b>13</b>
5.1	Hinweise . . . . .	13
5.2	Erläuterungen der einzelnen Variablen . . . . .	13
5.3	Konkordanztabelle: XML-Struktur und CSV-Variablen . . . . .	18
<b>6</b>	<b>Linguistische Kennzahlen</b>	<b>19</b>
6.1	Erläuterung der Kennzahlen und Diagramme . . . . .	19
6.2	Kennzahlen: Einzelnormen . . . . .	19
6.3	Kennzahlen: Rechtsakte . . . . .	19
6.4	Verteilung Tokens . . . . .	20
6.5	Verteilung Typen . . . . .	21
6.6	Verteilung Sätze . . . . .	22
<b>7</b>	<b>Inhalt</b>	<b>23</b>
7.1	Nach Periodikum . . . . .	23
7.1.1	Einzelnormen . . . . .	23
7.1.2	Rechtsakte mit veröffentlichtem Normtext . . . . .	25
7.1.3	Alle Rechtsakte (mit und ohne Normtext) . . . . .	27
7.2	Nach Ausfertigungsjahr . . . . .	30
7.2.1	Einzelnormen . . . . .	30
7.2.2	Rechtsakte mit veröffentlichtem Normtext . . . . .	35
7.2.3	Alle Rechtsakte (mit und ohne Normtext) . . . . .	40
<b>8</b>	<b>Dateigrößen: Summen und Verteilungen</b>	<b>45</b>
<b>9</b>	<b>Signaturprüfung</b>	<b>48</b>
9.1	Allgemeines . . . . .	48
9.2	Persönliche GPG-Signatur . . . . .	48

9.3	Import: Public Key . . . . .	48
9.4	Prüfung: GPG-Signatur der Hash-Datei . . . . .	49
9.5	Prüfung: SHA3-512 Hashes der ZIP-Archive . . . . .	50
<b>10</b>	<b>Changelog</b>	<b>51</b>
<b>11</b>	<b>Parameter für strenge Replikationen</b>	<b>52</b>
	<b>Literaturverzeichnis</b>	<b>53</b>

# 1 Einführung

Dem **Bundesrecht** kommt im Normengefüge der Bunderepublik Deutschland herausragende Bedeutung zu. Zwar sind die Länder gemäß Art. 30, 70 GG primär für die Gesetzgebung zuständig, im Katalog der Art. 71 ff GG sind aber derart viele Kompetenzen dem Bund zugewiesen, dass das Bundesrecht praktisch jedes rechtliche Problem in der Bundesrepublik dominiert. Ausnahmen sind in der Regel nur die Bereiche innere Sicherheit, Bildung und Kultur, die weitgehend in der Hand der Bundesländer verblieben sind. Aber auch in diesen Bereichen finden sich Regelungen des Bundes. Beispiele dafür sind manche Regelungen des Bundespolizeigesetzes (BPolG) oder das Kulturgutschutzgesetz (KGSG).

Bundesgesetze werden vom Bundestag im Zusammenwirken mit dem Bundesrat erlassen und vom Bundespräsidenten ausgefertigt (Art. 76 ff GG). Das Initiativrecht liegt bei Abgeordneten aus der Mitte des Bundestags, der Bundesregierung und dem Bundesrat (Art. 76 Abs. 1 GG). Der Bundesrat ist je nach Gesetzescharakter mit einem Zustimmungserfordernis oder einem Einspruchsrecht beteiligt (Art. 77, 78 GG).

Verordnungen werden in der Regel von der Exekutive erlassen, in seltenen Fällen vom Bundestag selbst. Durch Bundesgesetz können nur Bundesregierung, Bundesminister oder Landesregierungen hierzu ermächtigt werden (Art. 80 Abs. 1 S.1 GG), eine im Gesetz vorgesehene Sub-Delegation ist aber möglich (Art. 80 Abs. 1 S. 4 GG). Verordnungen müssen einem speziellen Bestimmtheitsgebot genügen und ihre Rechtsgrundlage in der Verordnung angeben (Art. 80 Abs. 1 S. 2 und 3). Der Erlass von Verordnungen erfordert zudem nicht selten die Zustimmung des Bundesrates, entweder aufgrund von Art. 80 Abs. 2 GG oder bedingt durch eine Regelung in einem einfachen Bundesgesetz.

Die quantitative Analyse von juristischen Texten, insbesondere von Gesetzen und Verordnungen, ist in den deutschen Rechtswissenschaften ein noch junges und kaum bearbeitetes Feld.<sup>1</sup> Zu einem nicht unerheblichen Teil liegt dies auch daran, dass die Anzahl an freizugänglichen Datensätzen außerordentlich gering ist.

Die meisten hochwertigen Datensätze lagern (fast) unerschlossen in kommerziellen Datenbanken und sind wissenschaftlich gar nicht oder nur gegen Entgelt zu nutzen. Freizugängliche Datenbanken wie *Opinio Iuris*<sup>2</sup> und *openJur*<sup>3</sup> verbieten ausdrücklich das maschinelle Auslesen der Rohdaten.<sup>4</sup> Wissenschaftliche Initiativen wie der Juristische Referenzkorpus (JuReKo) sind nach jahrelanger Arbeit hinter verschlossenen Türen verschwunden.

In einem funktionierenden Rechtsstaat muss die Rechtsetzung öffentlich, transparent und nachvollziehbar sein. Im 21. Jahrhundert bedeutet dies auch, dass sie quantitative Analysen zugänglich sein muss. Der Erstellung und Aufbereitung des Datensatzes liegen daher die Prinzipien der allgemeinen Verfügbarkeit durch Urheberrechtsfreiheit, strenge Transparenz und vollständige wissenschaftliche Reproduzierbarkeit zugrunde. Die FAIR-Prinzipien (Findable, Accessible, Interoperable and Reusable) für freie wissenschaftliche Daten inspirieren sowohl die Konstruktion, als auch die Art der Publikation.<sup>5</sup>

---

<sup>1</sup> Positive Ausnahmen finden sich vor allem unter: <https://www.quantitative-rechtswissenschaft.de/>

<sup>2</sup> <https://opinioiuris.de/>

<sup>3</sup> <https://openjur.de/>

<sup>4</sup> Openjur beabsichtigt eine API anzubieten, diese war aber im Januar 2021 immernoch nicht verfügbar. Openjur ist seit 2008 in Betrieb.

<sup>5</sup> 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 Nutzung

Die Daten sind in offenen, interoperablen und weit verbreiteten Formaten (CSV, XML, TXT, PDF, EPUB) veröffentlicht. Sie lassen sich grundsätzlich mit allen modernen Programmiersprachen (z.B. Python oder R), sowie mit grafischen Programmen nutzen.

**Wichtig:** Nicht vorhandene Werte sind sowohl in den Dateinamen als auch in der CSV-Datei mit »NA« codiert.

### 2.1 CSV-Dateien

Am einfachsten ist es die **CSV-Dateien** einzulesen. Die Nutzung der CSV-Varianten ist aus Qualitätsgründen und den umfangreicheren Metadaten **dringend empfohlen**. CSV<sup>6</sup> ist ein einfaches und maschinell gut lesbares Tabellen-Format. In diesem Datensatz sind die Werte komma-separiert. Jede Spalte entspricht einer Variable, jede Zeile einer Einzelnorm bzw. einem Rechtsakt (je nach Variante). Die Variablen sind unter Punkt 5 genauer erläutert.

Hier empfehle ich für **R** dringend das package **data.table** (via CRAN verfügbar). Dessen Funktion **fread()** ist etwa zehnmal so schnell wie die normale **read.csv()**-Funktion in Base-R. Sie erkennt auch den Datentyp von Variablen sicherer. Ein Vorschlag:

```
library(data.table)
csv.dbr <- fread("filename.csv")
```

### 2.2 XML-Dateien

Das Einlesen der **XML-Rohdaten** ist komplex und die Entscheidung welche XML-Nodes zu extrahieren sind wird ganz erheblich von der Forschungsfrage beeinflusst. Falls Sie über vertiefte XML-Kenntnisse verfügen, sollten Sie eine eigenständige Extraktion dennoch in Erwägung ziehen, weil sie so die Datenanalyse besser auf Ihre Bedürfnisse anpassen können. Lesen Sie hierfür bitte die Document Type Definition (DTD) genau und greifen Sie ggf. auf den im Source Code zur Verfügung gestellten XML Parser zurück.

### 2.3 TXT-Dateien

Die TXT-Dateien enthalten nur sehr rudimentäre Metadaten! Benutzen Sie daher für statistische Analysen vorzugsweise die CSV- oder XML-Dateien! Die **TXT-Dateien** inklusive Metadaten können zum Beispiel mit **R** und dem package **readtext** (via CRAN verfügbar) eingelesen werden. Ein Vorschlag:

```
library(readtext)
txt.dbr <- readtext("./*.txt",
                   docvarsfrom = "filenames",
                   docvarnames = c("kurztitel",
                                   "langtitel"),
                   dvsep = "_",
                   encoding = "UTF-8")
```

<sup>6</sup> Das CSV-Format ist in RFC 4180 definiert, siehe <https://tools.ietf.org/html/rfc4180>

## 3 Konstruktion

### 3.1 Datenquellen

Datenquelle	Fundstelle
Primäre Datenquelle	<a href="https://www.gesetze-im-internet.de/">https://www.gesetze-im-internet.de/</a>
Source Code	<a href="https://doi.org/10.5281/zenodo.4384767">https://doi.org/10.5281/zenodo.4384767</a>

**Hinweis:** Alle auf Zenodo bereitgestellten Datenquellen wurden vom Autor persönlich für die Zwecke dieses Datensatzes erstellt.

### 3.2 Beschreibung des Datensatzes

Der Datensatz ist eine möglichst vollständige Sammlung der konsolidierten Fassungen aller Gesetze und Verordnungen auf Bundesebene. Änderungsgesetze und -verordnungen sind nicht enthalten. Er enthält alle Rechtsakte, die auf der amtlichen Webseite »Gesetze im Internet« des Bundesministerium des Justiz am jeweiligen Stichtag verfügbar waren. Die Stichtage für jede Version sind in der Versionsnummer festgehalten.

Zusätzlich zu den einfach maschinenlesbaren Formaten (CSV und TXT) sind die XML-, PDF- und EPUB-Rohdaten enthalten, damit Analysten gegebenenfalls ihre eigene Konvertierung vornehmen können. Die Rohdaten wurden inhaltlich nicht verändert. Die PDF- und EPUB-Varianten der Rechtsakte sollen primär traditionelle juristische Forschung und mixed-methods-Ansätze unterstützen.

In diesem Datensatz sind nur Rechtsakte mit Außenwirkung (d.h. das Grundgesetz, Bundesgesetze und Bundesverordnungen) enthalten. Verwaltungsvorschriften sind nicht Teil des Datensatzes.

### 3.3 Sammlung der Daten

Die Daten wurden vollautomatisiert gesammelt und mit Abschluss der Verarbeitung kryptographisch signiert. Die Webseite des Justizministeriums ist laut dem Reiter »Hinweise«<sup>7</sup> ausdrücklich für die vollautomatisierte Datensammlung freigegeben. Der Abruf geschieht ausschließlich über TLS-verschlüsselte Verbindungen.

### 3.4 Source Code und Compilation Report

Der gesamte Source Code — sowohl für die Erstellung des Datensatzes, als auch für dieses Codebook — ist öffentlich einsehbar und dauerhaft erreichbar im wissenschaftlichen Archiv des CERN unter dieser Adresse hinterlegt: <https://doi.org/10.5281/zenodo.4384767>

Mit jeder Kompilierung des vollständigen Datensatzes wird auch ein umfangreicher **Compilation Report** in einem attraktiv designten PDF-Format erstellt (ähnlich diesem Codebook). Der Compilation Report enthält den vollständigen Source Code, dokumentiert

<sup>7</sup> <https://www.gesetze-im-internet.de/hinweise.html>

relevante Rechenergebnisse, gibt sekundengenaue Zeitstempel an und ist mit einem klickbaren Inhaltsverzeichnis versehen. Er ist zusammen mit dem Source Code hinterlegt. Wenn Sie sich für Details des Erstellungs-Prozesses interessieren, lesen Sie diesen bitte zuerst.

### 3.5 Grenzen des Datensatzes

Nutzer sollten folgende wichtige Grenzen beachten:

1. Der Datensatz enthält nur das, was das Ministerium auch tatsächlich veröffentlicht (*publication bias*). Es fehlen insbesondere Änderungsgesetze und -verordnungen. Manche Rechtsakte sind zudem nur mit den Metadaten nachgewiesen — ihr Inhalt fehlt aus technischen Gründen oder weil in der Bereinigten Sammlung Bundesgesetzblatt Teil III nur bibliographische Angaben enthalten sind (§ 3 Abs. 2 BRSG).<sup>8</sup>
2. Es kann aufgrund technischer Grenzen bzw. Fehler sein, dass manche — im Grunde verfügbare — Rechtsakte nicht oder nicht korrekt abgerufen werden (*automation bias*).
3. Es sind nur am Tag des Abrufs veröffentlichte konsolidierte Rechtsakte enthalten, eine diachronische Untersuchung muss somit mehrere verfügbare Versionen auswerten (*temporal bias*).

### 3.6 Urheberrechtsfreiheit von Rohdaten und Datensatz

An den Rechtsakten und Rechtsnormen besteht gem. § 5 Abs. 1 UrhG kein Urheberrecht, da sie amtliche Werke sind. § 5 UrhG ist auf amtliche Datenbanken analog anzuwenden (BGH, Beschluss vom 28.09.2006, I ZR 261/03, »Sächsischer Ausschreibungsdienst«).

Alle eigenen Beiträge (z.B. durch Zusammenstellung und Anpassung der Metadaten) und damit den gesamten Datensatz stelle ich gemäß einer *CC0 1.0 Universal Public Domain Lizenz* vollständig urheberrechtsfrei.

### 3.7 Metadaten

Alle Metadaten wurden aus den XML-Rohdaten zeitgleich mit dem Text der Normen extrahiert. Der volle Satz an Metadaten ist nur in den CSV-Dateien enthalten. Bitte beachten Sie, dass bei weitem nicht alle XML-Nodes ausgewertet wurden. Viele Nodes enthalten nur optische Informationen und wurden deshalb ignoriert. Manche Nodes (z.B. einzelne Absätze, Listen) wurden nicht extrahiert, weil nicht alle Normen in Absätze und Listen unterteilt sind und die Bereitstellung in einem nicht-hierarchischen Format wie CSV keine Vorteile gegenüber dem XML-Format bringen würde.

Die Dateinamen der PDF-, TXT und EPUB-Dateien enthalten nur eine Abkürzung und einen modifizierten Langtitel (auf 200 Zeichen gekürzt und um Sonderzeichen bereinigt). Diese wurden aus den jeweiligen Header-Markierungen der HTML-Seiten extrahiert.

#### 3.7.1 Schema für die Dateinamen (PDF, TXT, EPUB)

[Abkürzung]\_[modifizierter\_Langtitel]

<sup>8</sup> <https://www.gesetze-im-internet.de/hinweise.html>

### **3.7.2 Beispiel eines Dateinamens**

2.WasSV\_ZweiteWassersicherstellungsverordnung.pdf

## **3.8 Qualitätsprüfung**

Insgesamt werden zusammen mit jeder Kompilierung Dutzende Tests zur Qualitätsprüfung durchgeführt. Alle Ergebnisse der Qualitätsprüfungen sind aggregiert im Compilation Report und einzeln im Archiv »ANALYSE« zusammen mit dem Datensatz veröffentlicht.

## 4 Varianten und Zielgruppen

Dieser Datensatz ist in verschiedenen Varianten verfügbar, die sich an unterschiedliche Zielgruppen richten. Zielgruppe sind nicht nur quantitativ forschende RechtswissenschaftlerInnen, sondern auch traditionell arbeitende JuristInnen. Idealerweise müssen quantitative Methoden ohnehin immer durch qualitative Interpretation, Theorie-Bildung und kritische Auseinandersetzung verstärkt werden (*mixed methods*).

Lehrende werden zudem von den vorbereiteten Tabellen und Diagrammen besonders profitieren, die bei der Erläuterung der Charakteristika der Daten hilfreich sein können und Zeit im universitären Alltag sparen. Alle Tabellen und Diagramme liegen auch als separate Dateien vor um sie einfach z.B. in Präsentations-Folien oder Handreichungen zu integrieren.

Variante	Zielgruppe und Beschreibung
CSV_Einzelnormen_Datensatz	<b>Legal Tech/Quantitative Forschung.</b> Diese CSV-Datei ist eine der für statistische Analysen empfohlenen Varianten des Datensatzes. Sie enthält den Volltext aller Rechtsakte, disaggregiert nach Einzelnormen, sowie alle in diesem Codebook beschriebenen Metadaten. Enthält nur Rechtsakte, für die mindestens eine Einzelnorm mit Normtext veröffentlicht wurde!
CSV_Einzelnormen_Metadaten	<b>Legal Tech/Quantitative Forschung.</b> Wie die andere CSV-Datei mit Einzelnormen, nur ohne die Normtexte. Sinnvoll für AnalystInnen, die sich nur für die Metadaten interessieren und Speicherplatz sparen wollen. Enthält nur Rechtsakte, für die mindestens eine Einzelnorm mit Normtext veröffentlicht wurde!
CSV_Rechtsakte_Datensatz	<b>Legal Tech/Quantitative Forschung.</b> Diese CSV-Datei ist eine der für statistische Analysen empfohlenen Varianten des Datensatzes. Sie enthält den Volltext aller Rechtsakte, sowie fast alle in diesem Codebook beschriebenen Metadaten. Die gegenüber den Einzelnormen fehlenden Metadaten betreffen vor allem Gliederungsdaten (z.B. Gliederungsüberschrift), die auf Rechtsakts-Ebene keinen Sinn ergeben. Wurde durch ein Zusammenfügen der Einzelnorm-Variante erstellt. Enthält nur Rechtsakte, für die mindestens eine Einzelnorm mit Normtext veröffentlicht wurde!
CSV_Rechtsakte_Metadaten	<b>Legal Tech/Quantitative Forschung.</b> Wie die andere CSV-Datei mit Rechtsakten, nur ohne die Normtexte. Sinnvoll für AnalystInnen, die sich nur für die Metadaten interessieren und Speicherplatz sparen wollen. Enthält nur Rechtsakte, für die mindestens eine Einzelnorm mit Normtext veröffentlicht wurde!
CSV_MetadatenXML	<b>Legal Tech/Quantitative Forschung.</b> Diese CSV-Datei enthält Metadaten für jeden auf der amtlichen Webseite nachgewiesenen Rechtsakt, unabhängig davon, ob mit oder ohne Normtext veröffentlicht. Die Zahl der Rechtsakte ist daher um etwa 1000 höher als bei den anderen CSV-Dateien, es sind aber keine Normtexte enthalten.

Variante	Zielgruppe und Beschreibung
XML_Datensatz	<b>Legal Tech/Quantitative Forschung.</b> Die XML-Rohdaten. Alle CSV-Dateien wurden aus diesen Rohdaten extrahiert. XML ist ein komplexes Format und daher nur für erfahrene ForscherInnen geeignet.
XML_Anlagen	<b>Legal Tech/Quantitative Forschung.</b> Manche XML-Dateien verweisen auf Anlagen, vorwiegend Bild-Dateien. Diese sind hier zusammengefasst.
PDF_Datensatz	<b>Traditionelle juristische Forschung.</b> Die PDF-Dokumente wie sie vom Bundesjustizministerium auf der amtlichen Webseite bereitgestellt werden, jedoch verbessert durch semantisch hochwertige Dateinamen, die sowohl die Abkürzung, als auch einen modifizierten Langtitel enthalten. Die Dateinamen sind so konzipiert, dass sie auch für traditionelle qualitative juristische Arbeit einen erheblichen Mehrwert bieten. Im Vergleich zu den CSV-Dateien enthalten die Dateinamen nur einen drastisch reduzierten Umfang an Metadaten, um Kompatibilitätsprobleme unter Windows zu vermeiden und die Lesbarkeit zu verbessern. Besonders geeignet für die Arbeit an Desktop PCs.
EPUB_Datensatz	<b>Traditionelle juristische Forschung.</b> Die PDF-Dokumente wie sie vom Bundesjustizministerium auf der amtlichen Webseite bereitgestellt werden, jedoch verbessert durch semantisch hochwertige Dateinamen, die sowohl die Abkürzung, als auch einen modifizierten Langtitel enthalten. Die Dateinamen sind so konzipiert, dass sie auch für traditionelle qualitative juristische Arbeit einen erheblichen Mehrwert bieten. Im Vergleich zu den CSV-Dateien enthalten die Dateinamen nur einen drastisch reduzierten Umfang an Metadaten, um Kompatibilitätsprobleme unter Windows zu vermeiden und die Lesbarkeit zu verbessern. Besonders geeignet für die Arbeit an mobilen Endgeräten, weil sich das Format der Bildschirmgröße anpassen kann.
TXT_Datensatz	<b>Subsidiär.</b> Diese Variante enthält die vollständigen aus den PDF-Dateien extrahierten Normtexte der Rechtsakte, aber nur einen drastisch reduzierten Umfang an Metadaten, der dem der PDF-Dateien entspricht. Die TXT-Dateien sind optisch an das Layout der PDF-Dateien angelehnt. Geeignet für qualitative Forscher, die nur wenig Speicherplatz oder eine langsame Internetverbindung zur Verfügung haben und für quantitative Forscher, die beim Einlesen der CSV-Dateien Probleme haben.

Variante	Zielgruppe und Beschreibung
ANALYSE	<p><b>Alle Lehrenden und Forschenden.</b> Dieses Archiv enthält alle während dem Kompilierungs- und Prüfprozess erstellten Tabellen (CSV) und Diagramme (PDF, PNG) im Original. Sie sind inhaltsgleich mit den in diesem Codebook verwendeten Tabellen und Diagrammen. Das PDF-Format eignet sich besonders für die Verwendung in gedruckten Publikationen, das PNG-Format besonders für die Darstellung im Internet. AnalystInnen mit fortgeschrittenen Kenntnissen in R können auch auf den Source Code zurückgreifen. Empfohlen für Nutzer die einzelne Inhalte aus dem Codebook für andere Zwecke (z.B. eigene Publikationen) weiterverwenden möchten.</p>

## 5 Variablen

### 5.1 Hinweise

- Fehlende Werte sind immer mit »NA« codiert
- Strings können grundsätzlich alle in UTF-8 definierten Zeichen (insbesondere Buchstaben, Zahlen und Sonderzeichen) enthalten.
- Alle Variablen sind in der hier beschriebenen Form nur in der CSV-Datei enthalten. Die meisten davon sind jedoch aus gleichlautenden oder ähnlich lautenden Nodes in den XML-Daten vorhanden.

### 5.2 Erläuterungen der einzelnen Variablen

Variable	Typ	Erläuterung
doc_id	String	Der Name der extrahierten XML-Datei. Einzelnormen enthalten den Namen der XML-Datei und jeweils eine fortlaufende Zahl um die Einzigartigkeit der ID für die Nutzung mit NLP-Tools herzustellen.
text	String	Der vollständige Normtext der Einzelnorm oder des Rechtsaktes, so wie er in den XML-Dateien dokumentiert ist. Nur die Varianten »Einzelnormen« und »Rechtsakte« enthalten Textdaten. Hierzu schreibt das Ministerium in den Hinweisen: »Einzelne Vorschriften sind nur mit der Überschrift aufgenommen. In einigen Fällen hat dies technische Gründe, in anderen Fällen ist dies dadurch bedingt, dass die Vorschrift nur mit ihren bibliographischen Angaben in der Bereinigten Sammlung Bundesgesetzblatt Teil III enthalten ist (§ 3 Abs. 2 BRSG).«
builddate_original	String	Datum und Uhrzeit an dem die XML-Repräsentation der Norm konstruiert wurde, eine Serie von Zahlen ohne Interpunktion. Das genaue Format ist nicht dokumentiert, es ist aber sehr wahrscheinlich so aufgebaut: vierstellige Jahreszahl, zweistellige Monatszahl, zweistellige Tageszahl, zweistellige Stundenzahl, zweistellige Minutenzahl und eine zweistellige Sekundenzahl.
builddate_iso	Datum (ISO)	Eine Interpretation der builddate-Variable im ISO-Format (z.B. 2016-09-12T18:12:16Z). Das genaue Original-Format ist nicht dokumentiert, die Variable wurde aber unter folgenden Annahmen extrahiert: vierstellige Jahreszahl, zweistellige Monatszahl, zweistellige Tageszahl, zweistellige Stundenzahl, zweistellige Minutenzahl und eine zweistellige Sekundenzahl.

Variable	Typ	Erläuterung
fundstellentyp	String	Ob es sich um eine amtliche Fundstelle handelt oder nicht. Mögliche Werte sind »amtlich« oder »nichtamtlich«. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
periodikum	String	Die Abkürzung des Periodikums in dem die amtliche Fassung des Rechtsaktes erschienen ist, beispielsweise »BGBI I« (Bundesgesetzblatt I). Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
zitstelle	String	Die genaue Fundstelle im jeweiligen Periodikum. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
stand	String	Informationen zum aktuellen Stand des Rechtsaktes, als Fließtext. Enthält insbesondere Informationen zur letzten Änderung und dem letzten Änderungsrechtsakt. Falls nicht vorhanden ist der Wert »NA«. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
aufh	String	Informationen zur etwaigen Aufhebung des Rechtsaktes, als Fließtext. Falls nicht vorhanden ist der Wert »NA«. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
neuf	String	Informationen zur letzten Neufassung des Rechtsaktes, als Fließtext. Falls nicht vorhanden ist der Wert »NA«. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
hinweis	String	Hinweise zur dokumentarischen Bearbeitung des Rechtsaktes, als Fließtext. Falls nicht vorhanden ist der Wert »NA«. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
sonst	String	Sonstige Informationen zum Stand des Rechtsaktes, als Fließtext. Falls nicht vorhanden ist der Wert »NA«. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
check_*	String	Ob die Angabe der jeweiligen Stand-Variable geprüft wurde. Es ist unklar, welche Prüfung hier vom Ministerium vorgenommen wurde. Mögliche Werte sind »ja« oder »NA«. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
amtabk	String	Die amtliche Abkürzung des Rechtsaktes. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!

Variable	Typ	Erläuterung
jurabk	String	Unter »Hinweise« schreibt das Ministerium: »Zu einigen Gesetzen und Verordnungen existieren keine amtlichen Abkürzungen. In diesen Fällen sind die Vorschriften in den alphabetischen Listen von ›Gesetze im Internet‹ anhand der von der Dokumentationsstelle im BfJ gebildeten und in der Bundesrechtsdatenbank verwendeten Abkürzungen eingeordnet. Diese nichtamtlichen Abkürzungen können von Abkürzungen, die andere Anbieter verwenden, abweichen.« Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
langue	String	Die Langform des Namens (Langüberschrift) eines Rechtsaktes. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
kurzue	String	Die Kurzform des Namens (Kurzüberschrift) eines Rechtsaktes. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt!
gliederungskennzahl	String	(Nur CSV-Datei mit Einzelnormen) Die Kennzahl der jeweiligen Gliederungsebene. Das zugrundeliegende System ist vom Ministerium nicht dokumentiert und wissenschaftlich noch nicht erschlossen. Beispielsweise »010050030«.
gliederungsbez	String	(Nur CSV-Datei mit Einzelnormen) Die Bezeichnung der Gliederungsebene. Beispielsweise »Titel 3«.
gliederungstitel	String	(Nur CSV-Datei mit Einzelnormen) Der Name der Gliederungsebene. Beispielsweise »Rechtsfolgen der Verjährung«.
enbez	String	(Nur CSV-Datei mit Einzelnormen) Die Bezeichnung der Einzelnorm. Beispielsweise »§ 214«.
ausfertigung_datum	Datum	Das Datum an dem der Rechtsakt ausgefertigt wurde. Das Format ist YYYY-MM-DD (Langform nach ISO-8601). Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt! Daher sind beispielsweise für das BGB alle Normen mit dem Ausfertigungsjahr 1896 versehen, auch wenn die Einzelnorm später erlassen wurde.
ausfertigung_jahr	Natürliche Zahl	Das Jahr in dem der Rechtsakt ausgefertigt wurde. Für Einzelnormen bezieht sich die Angabe auf den gesamten Rechtsakt. Das Format ist eine vierstellige Jahreszahl (YYYY). Wurde durch den Autor aus dem Ausfertigungsdatum berechnet.

Variable	Typ	Erläuterung
tokens	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
typen	Natürliche Zahl	Die Anzahl einzigartiger Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung und Typenzählung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
saetze	Natürliche Zahl	Die Anzahl Sätze. Entsprechen in etwa dem üblichen Verständnis eines Satzes. Die Regeln für die Bestimmung von Satzanfang und Satzende sind im Detail sehr komplex und in »Unicode Standard Annex No 29« beschrieben. Diese Zahl kann je nach Software und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Zählung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
version	Datum	Die Versionsnummer des Datensatzes im Format YYYY-MM-DD (Langform nach ISO-8601). Die Versionsnummer entspricht immer dem Datum an dem der Datensatz erstellt und die Daten von der Webseite des Gerichts abgerufen wurden.

Variable	Typ	Erläuterung
doi_concept	String	Der Digital Object Identifier (DOI) des Gesamtkonzeptes des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via <a href="http://www.doi.org">www.doi.org</a> immer die <b>aktuellste Version</b> des Datensatzes abgerufen werden. Prinzip F1 der FAIR-Data Prinzipien (»data are assigned globally unique and persistent identifiers«) empfiehlt die Dokumentation jeder Messung mit einem persistenten Identifikator. Selbst wenn die CSV-Dateien ohne Kontext weitergegeben werden kann ihre Herkunft so immer zweifelsfrei und maschinenlesbar bestimmt werden.
doi_version	String	Der Digital Object Identifier (DOI) der <b>konkreten Version</b> des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via <a href="http://www.doi.org">www.doi.org</a> immer diese konkrete Version des Datensatzes abgerufen werden. Prinzip F1 der FAIR-Data Prinzipien (»data are assigned globally unique and persistent identifiers«) empfiehlt die Dokumentation jeder Messung mit einem persistenten Identifikator. Selbst wenn die CSV-Dateien ohne Kontext weitergegeben werden kann ihre Herkunft so immer zweifelsfrei und maschinenlesbar bestimmt werden.

### 5.3 Konkordanztabelle: XML-Struktur und CSV-Variablen

CSV-Variablen	XPath	Attribut
text	/norm/textdaten/text/Content	-
builddate_original	/norm	builddate
fundstellentyp	/norm/metadaten/fundstelle	typ
periodikum	/norm/metadaten/fundstelle/periodikum	-
zitstelle	/norm/metadaten/fundstelle/zitstelle	-
stand	/norm/metadaten/standangabe/standtyp	-
	/norm/metadaten/standangabe/standkommentar	-
aufh	/norm/metadaten/standangabe/standtyp	-
	/norm/metadaten/standangabe/standkommentar	-
neuf	/norm/metadaten/standangabe/standtyp	-
	/norm/metadaten/standangabe/standkommentar	-
hinweis	/norm/metadaten/standangabe/standtyp	-
	/norm/metadaten/standangabe/standkommentar	-
sonst	/norm/metadaten/standangabe/standtyp	-
	/norm/metadaten/standangabe/standkommentar	-
check_*	/norm/metadaten/standangabe	checked
amtabk	/norm/metadaten/amtabk	-
jurabk	/norm/metadaten/jurabk	-
langue	/norm/metadaten/langue	-
kurzue	/norm/metadaten/kurzue	-
gliederungskennzahl	/norm/metadaten/gliederungseinheit/gliederungskennzahl	-
gliederungsbez	/norm/metadaten/gliederungseinheit/gliederungsbez	-
gliederungstitel	/norm/metadaten/gliederungseinheit/gliederungstitel	-
enbez	/norm/metadaten/enbez	-
ausfertigung_datum	/norm/metadaten/ausfertigung-datum	-

## 6 Linguistische Kennzahlen

### 6.1 Erläuterung der Kennzahlen und Diagramme

Zur besseren Einschätzung des inhaltlichen Umfangs des Korpus dokumentiere ich an dieser Stelle die Verteilung der Werte für drei verschiedene klassische linguistische Kennzahlen:

Variable	Definition
Tokens	Eine beliebige Zeichenfolge, getrennt durch whitespace-Zeichen, d.h. ein Token entspricht in der Regel einem »Wort«, kann aber gelegentlich auch sinnlose Zeichenfolgen enthalten, weil es rein syntaktisch berechnet wird.
Typen	Einzigartige Tokens. Beispiel: wenn das Token »gewerblich« mehrmals in einer Norm vorhanden ist, wird es als ein Typ gezählt.
Sätze	Entsprechen in etwa dem üblichen Verständnis eines Satzes. Die Regeln für die Bestimmung von Satzanfang und Satzende sind im Detail aber sehr komplex und in »Unicode Standard Annex No 29« beschrieben. Für Rechtsnormen ist diese Zählweise vermutlich nicht robust genug, interpretieren Sie die Ergebnisse mit großer Vorsicht!

Es handelt sich bei den Diagrammen jeweils um »Density Charts«, die sich besonders dafür eignen die Schwerpunkte von Variablen mit stark schwankenden numerischen Werten zu visualisieren. Die Interpretation ist denkbar einfach: je höher die Kurve, desto dichter sind in diesem Bereich die Werte der Variable. Bitte beachten Sie, dass die x-Achse logarithmisch (d.h. nicht-linear!) skaliert ist.

### 6.2 Kennzahlen: Einzelnormen

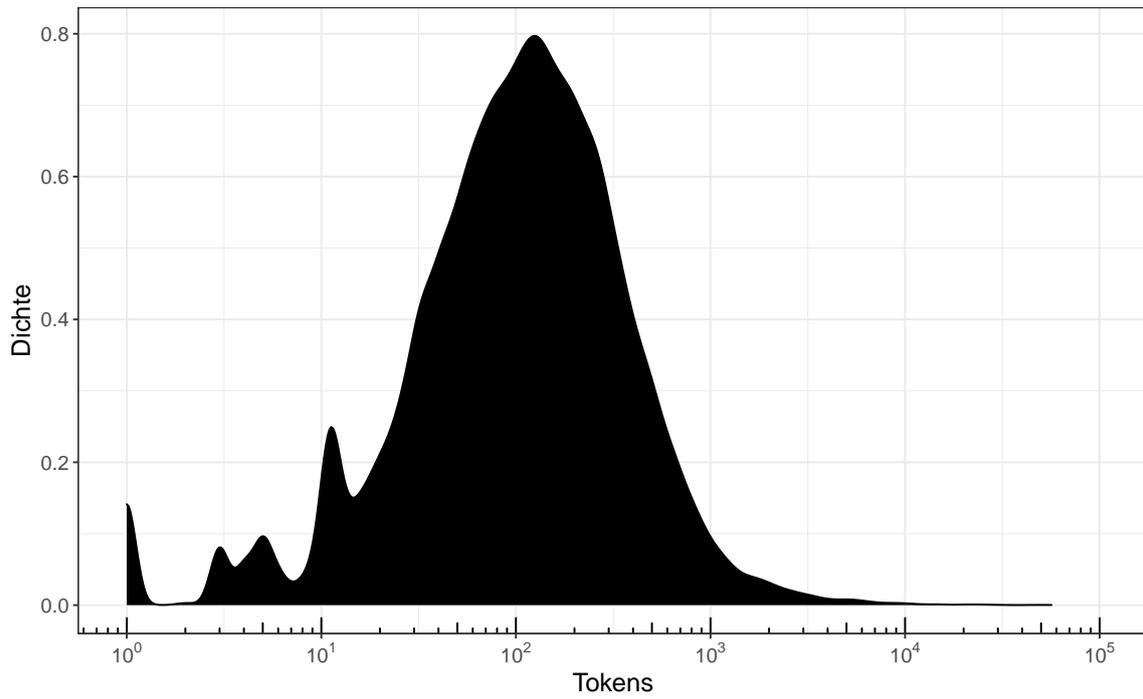
Kennzahl	Summe	Min	1. Quartil	Median	Mittel	3. Quartil	Max.
Tokens	21,446,633	1	46	107	217.454150	230	56,584
Sätze	724,389	1	2	4	7.344808	8	2,381
Typen	NA	1	35	66	91.997932	113	11,096

### 6.3 Kennzahlen: Rechtsakte

Kennzahl	Summe	Min	1. Quartil	Median	Mittel	3. Quartil	Max.
Tokens	21,446,633	17	263	923	3,851.0743	3,378	273,108
Sätze	712,908	1	13	36	128.0136	114	14,306
Typen	NA	15	129	303	633.4236	803	14,622

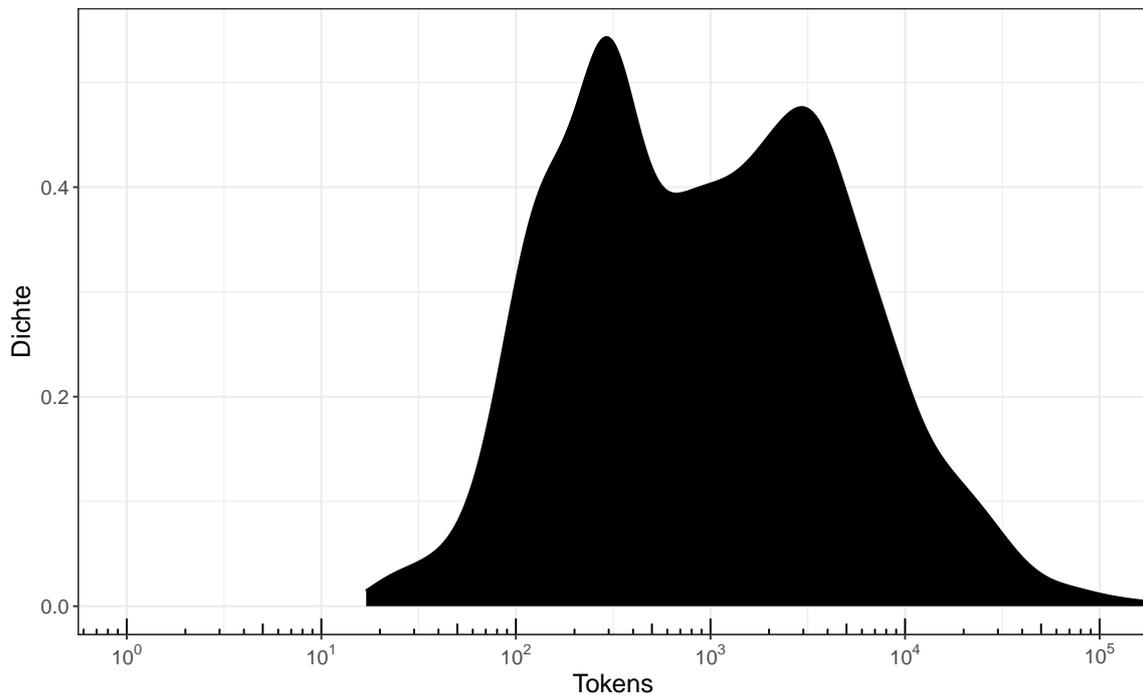
## 6.4 Verteilung Tokens

C-DBR | Version 2021-01-05 | Verteilung der Tokens je Norm



DOI: 10.5281/zenodo.4384771

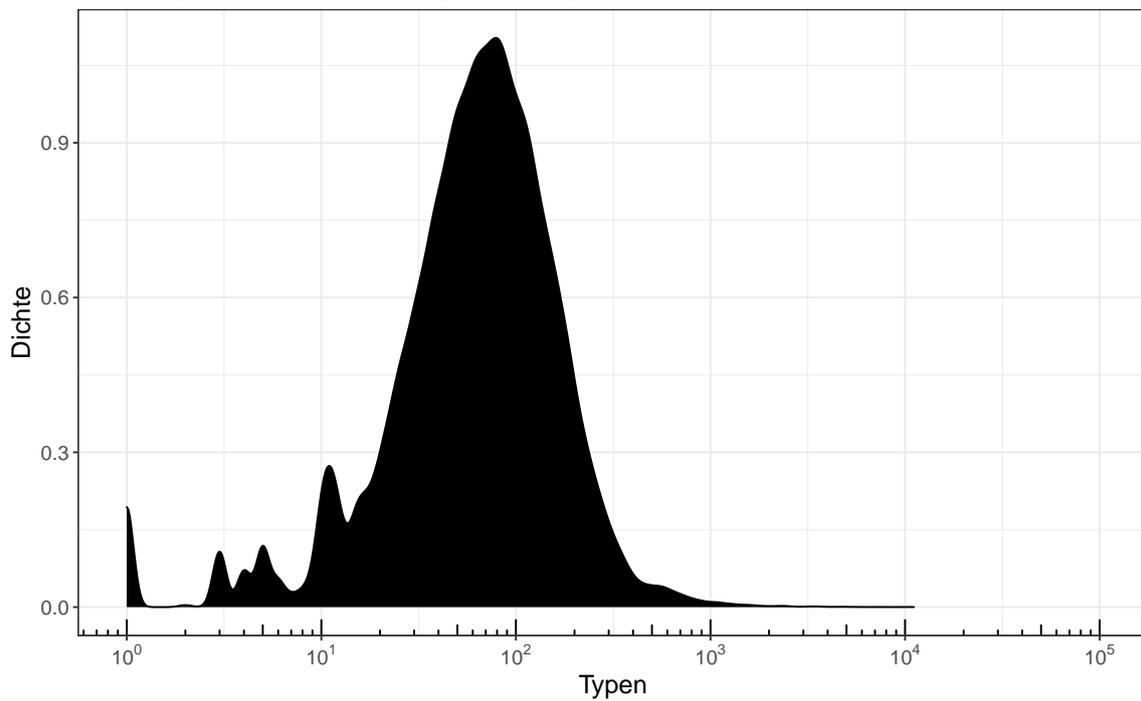
C-DBR | Version 2021-01-05 | Verteilung der Tokens je Rechtsakt



DOI: 10.5281/zenodo.4384771

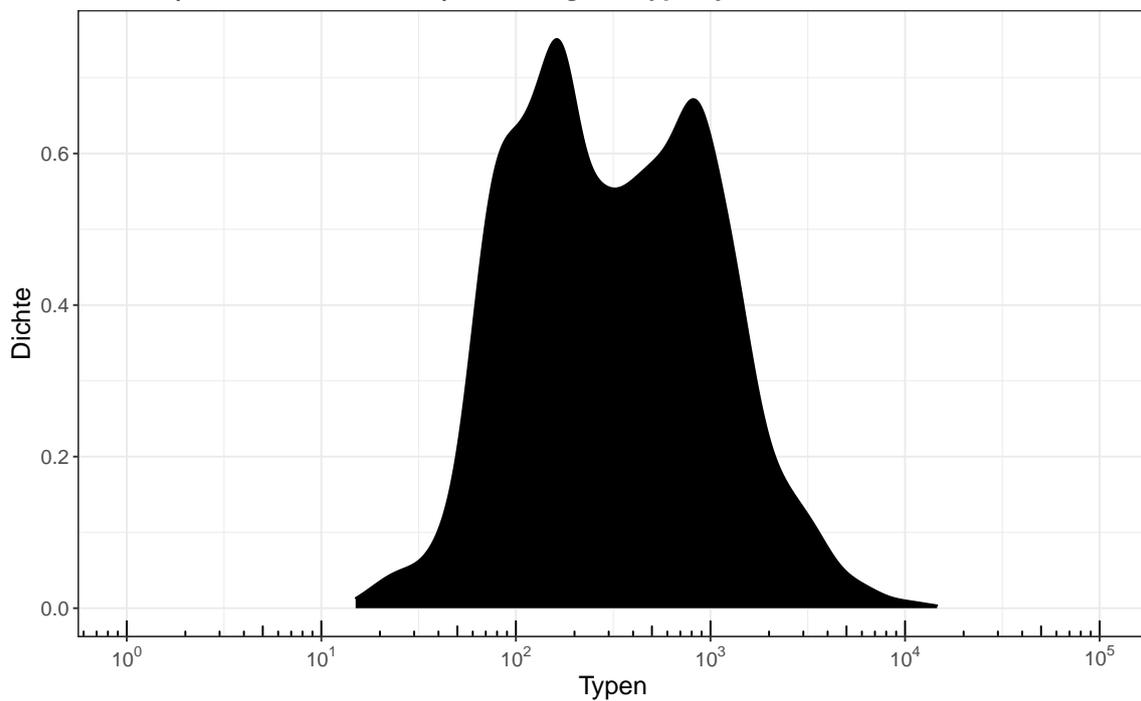
## 6.5 Verteilung Typen

C-DBR | Version 2021-01-05 | Verteilung der Typen je Norm



DOI: 10.5281/zenodo.4384771

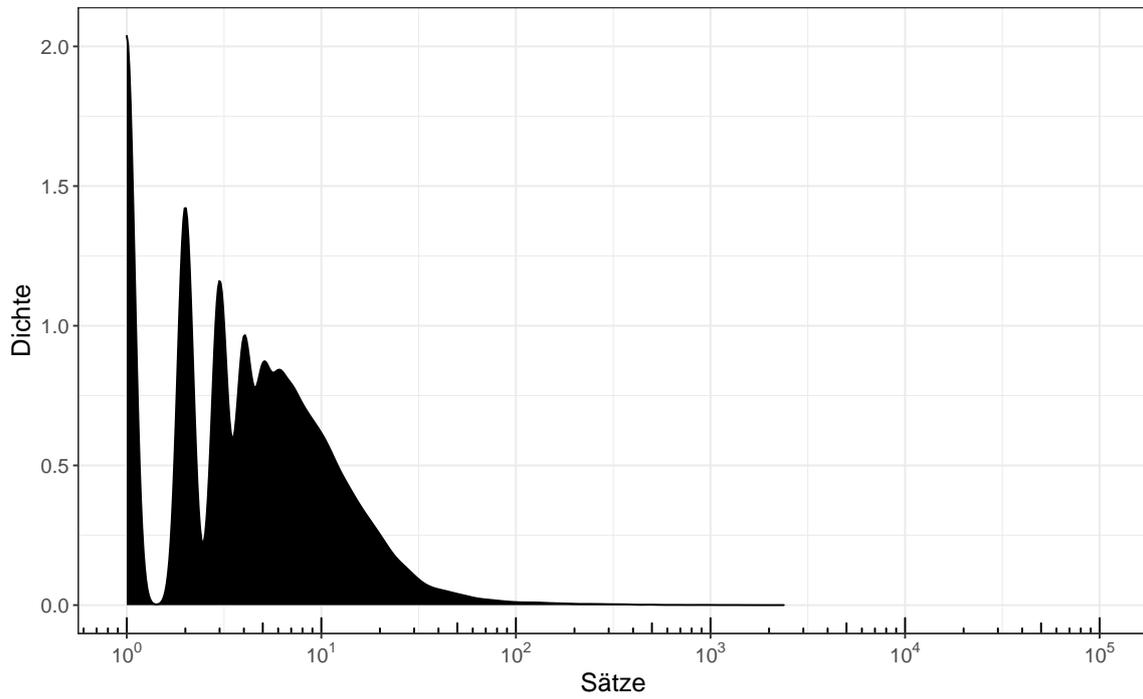
C-DBR | Version 2021-01-05 | Verteilung der Typen je Rechtsakt



DOI: 10.5281/zenodo.4384771

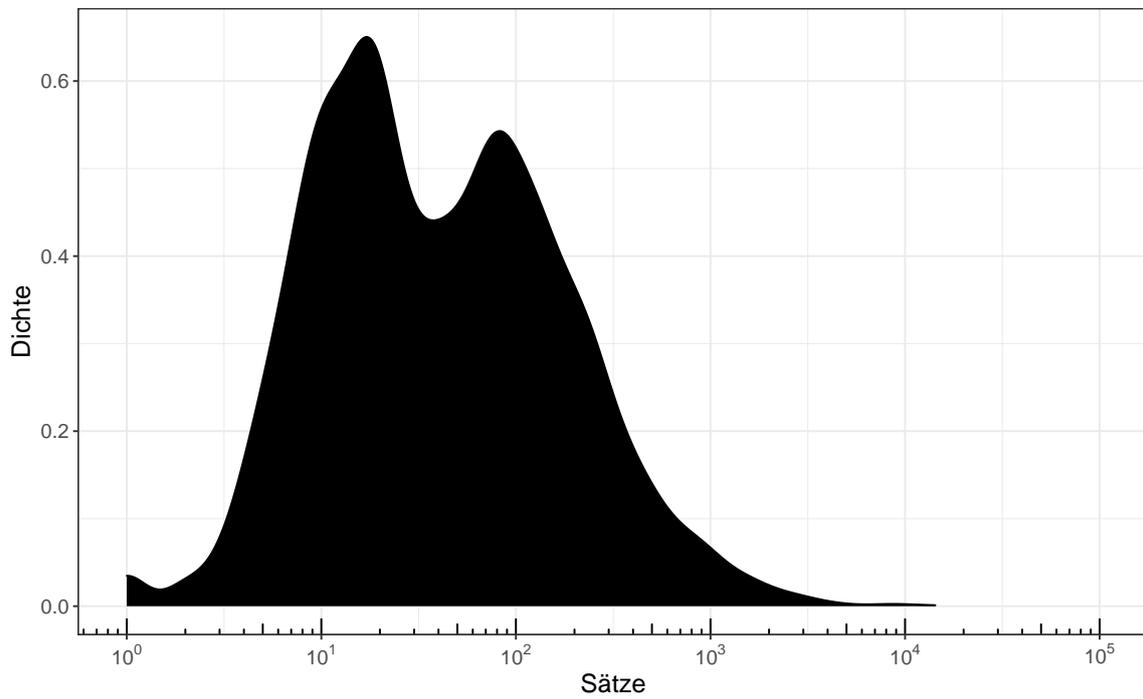
## 6.6 Verteilung Sätze

C-DBR | Version 2021-01-05 | Verteilung der Sätze je Norm



DOI: 10.5281/zenodo.4384771

C-DBR | Version 2021-01-05 | Verteilung der Sätze je Rechtsakt

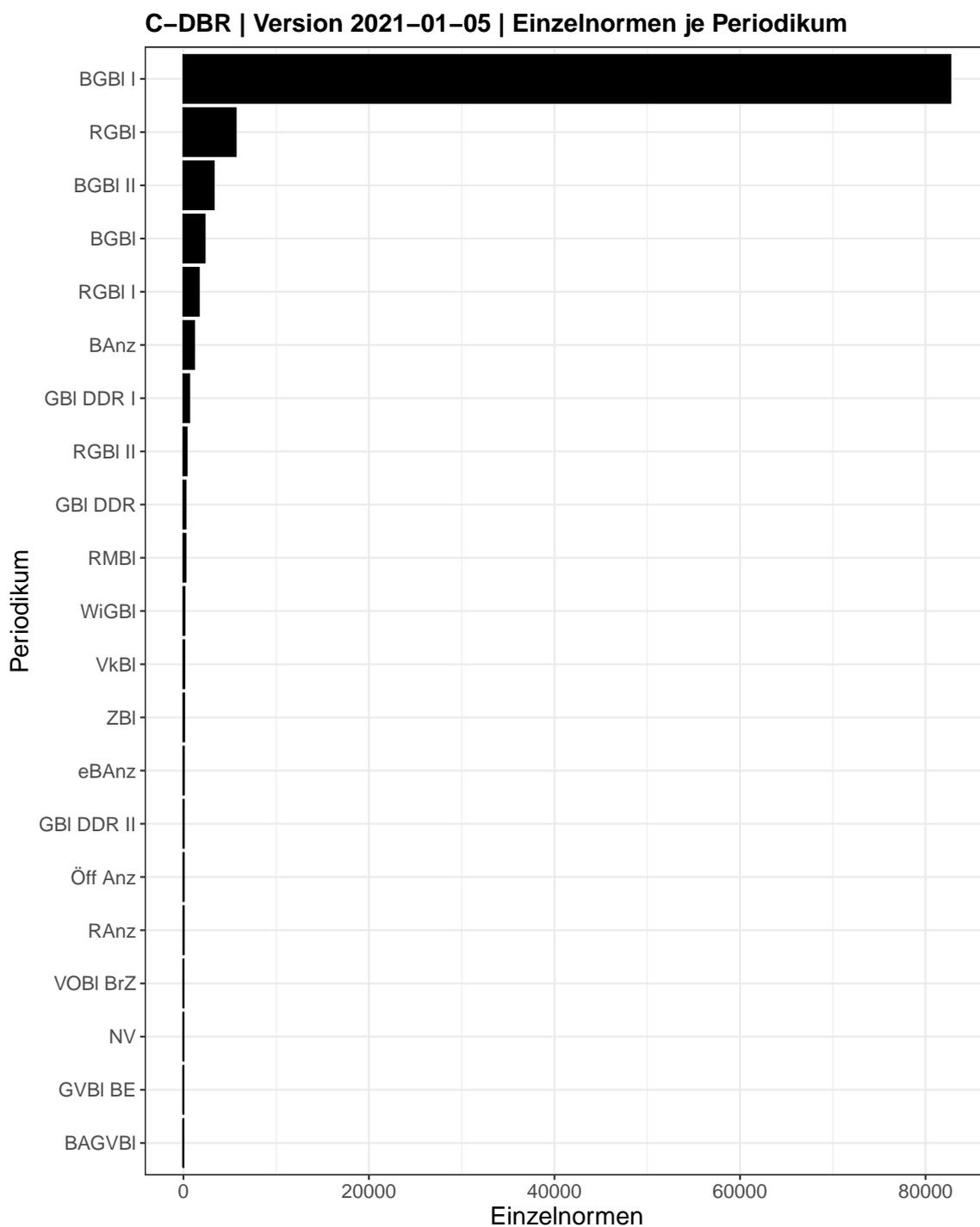


DOI: 10.5281/zenodo.4384771

## 7 Inhalt

### 7.1 Nach Periodikum

#### 7.1.1 Einzelnormen

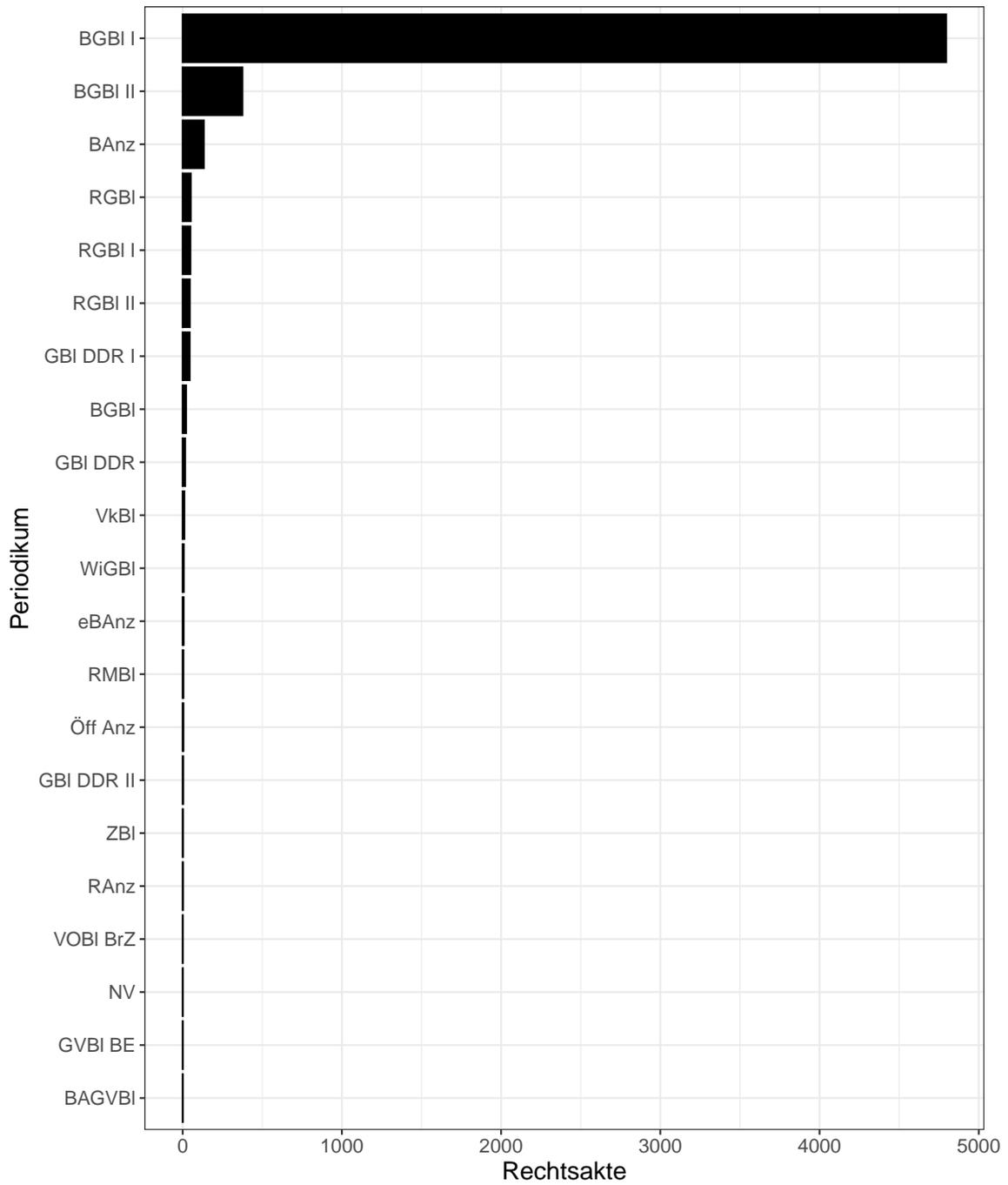


DOI: 10.5281/zenodo.4384771

Periodikum	Einzelnormen	% Gesamt	% Kumulativ
BAGVBl	1	0.00	0.00
BAnz	1156	1.17	1.17
BGBI	2286	2.32	3.49
BGBI I	82671	83.82	87.31
BGBI II	3265	3.31	90.62
GBl DDR	233	0.24	90.86
GBl DDR I	622	0.63	91.49
GBl DDR II	43	0.04	91.53
GVBl BE	9	0.01	91.54
NV	15	0.02	91.56
RAnz	38	0.04	91.60
RGBI	5634	5.71	97.31
RGBI I	1657	1.68	98.99
RGBI II	347	0.35	99.34
RMBI	230	0.23	99.58
VOBl BrZ	19	0.02	99.59
VkBl	105	0.11	99.70
WiGBI	122	0.12	99.82
ZBl	80	0.08	99.91
eBAnz	52	0.05	99.96
Öff Anz	41	0.04	100.00

### 7.1.2 Rechtsakte mit veröffentlichtem Normtext

C-DBR | Version 2021-01-05 | Rechtsakte mit Inhalt je Periodikum

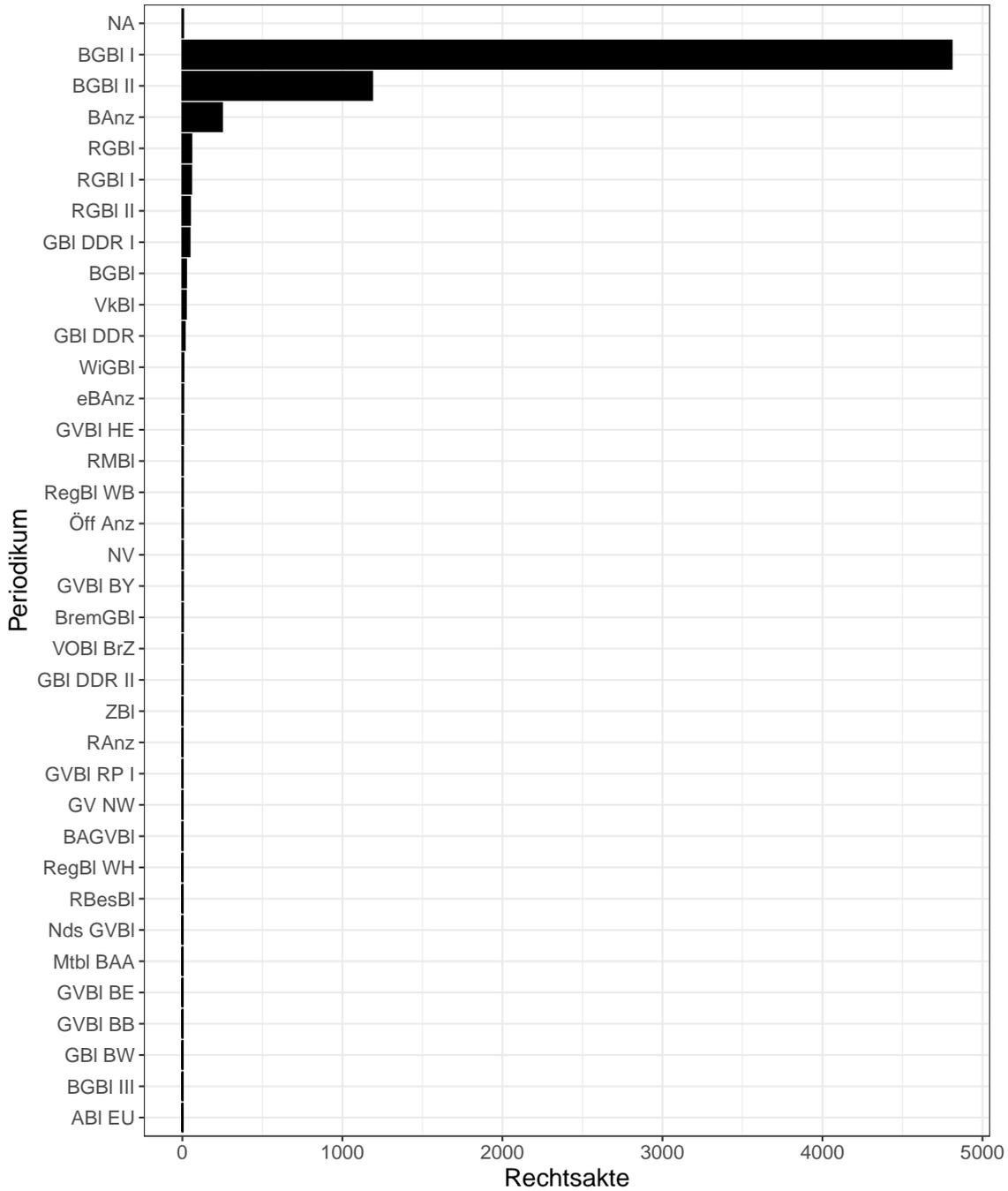


DOI: 10.5281/zenodo.4384771

Periodikum	Rechtsakte	% Gesamt	% Kumulativ
BAGVBl	1	0.02	0.02
BAnz	133	2.39	2.41
BGBI	21	0.38	2.78
BGBI I	4795	86.10	88.88
BGBI II	375	6.73	95.62
GBl DDR	15	0.27	95.89
GBl DDR I	44	0.79	96.68
GBl DDR II	3	0.05	96.73
GVBl BE	1	0.02	96.75
NV	1	0.02	96.77
RAnz	2	0.04	96.80
RGBI	51	0.92	97.72
RGBI I	48	0.86	98.58
RGBI II	45	0.81	99.39
RMBI	4	0.07	99.46
VOBl BrZ	1	0.02	99.48
VkBl	10	0.18	99.66
WiGBI	7	0.13	99.78
ZBl	2	0.04	99.82
eBAnz	6	0.11	99.93
Öff Anz	4	0.07	100.00

### 7.1.3 Alle Rechtsakte (mit und ohne Normtext)

C-DBR | Version 2021-01-05 | Rechtsakte nach Metadaten je Periodikum



DOI: 10.5281/zenodo.4384771

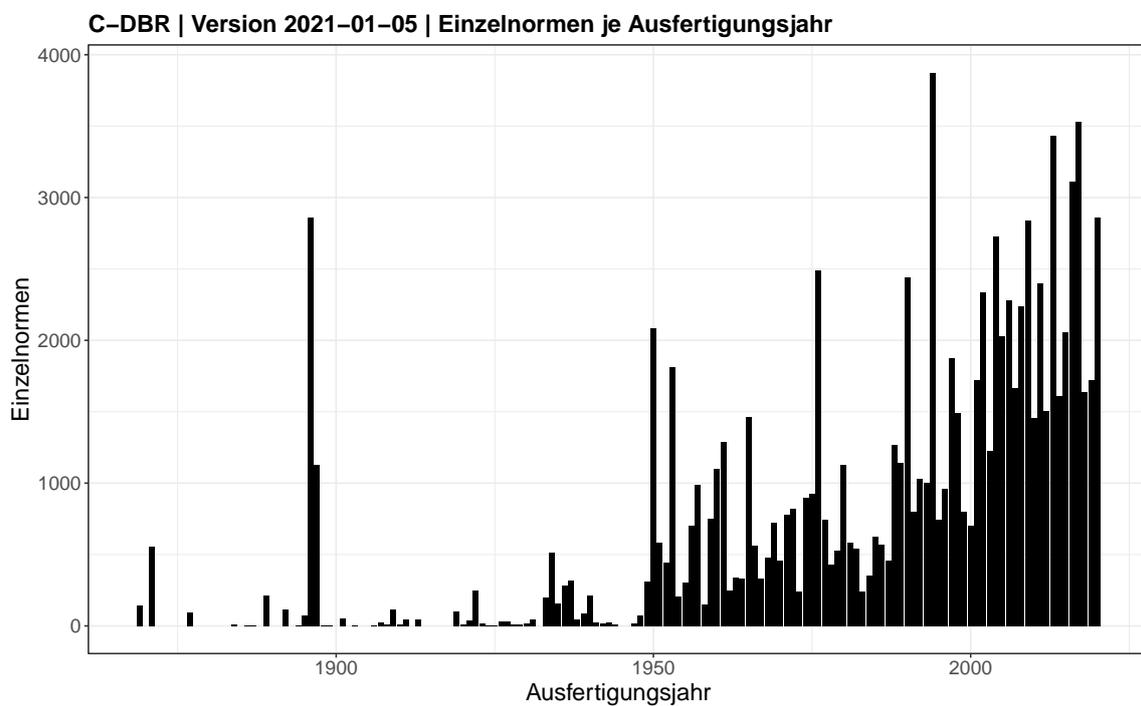
Periodikum	Rechtsakte	% Gesamt	% Kumulativ
NA	5	0.08	0.08
ABl EU	1	0.02	0.09
BAGVBl	2	0.03	0.12
BAnz	248	3.77	3.89
BGBI	23	0.35	4.24
BGBI I	4807	73.09	77.33
BGBI II	1186	18.03	95.36
BGBI III	1	0.02	95.38
BremGBI	4	0.06	95.44
GBI BW	1	0.02	95.45
GBI DDR	15	0.23	95.68
GBI DDR I	45	0.68	96.37
GBI DDR II	3	0.05	96.41
GV NW	2	0.03	96.44
GVBl BB	1	0.02	96.46
GVBl BE	1	0.02	96.47
GVBl BY	4	0.06	96.53
GVBl HE	5	0.08	96.61
GVBl RP I	2	0.03	96.64
Mtbl BAA	1	0.02	96.66
NV	4	0.06	96.72
Nds GVBl	1	0.02	96.73
RAnz	2	0.03	96.76
RBesBl	1	0.02	96.78
RGBI	56	0.85	97.63
RGBI I	55	0.84	98.46
RGBI II	48	0.73	99.19
RMBl	4	0.06	99.25
RegBl WB	4	0.06	99.32
RegBl WH	1	0.02	99.33
VOBl BrZ	3	0.05	99.38
VkBl	22	0.33	99.71

*(continued)*

Periodikum	Rechtsakte	% Gesamt	% Kumulativ
WiGBI	7	0.11	99.82
ZBI	2	0.03	99.85
eBAnz	6	0.09	99.94
Öff Anz	4	0.06	100.00

## 7.2 Nach Ausfertigungsjahr

### 7.2.1 Einzelnormen



DOI: 10.5281/zenodo.4384771

Jahr	Einzelnormen	% Gesamt	% Kumulativ
1869	146	0.15	0.15
1871	558	0.57	0.71
1877	97	0.10	0.81
1884	10	0.01	0.82
1886	3	0.00	0.83
1887	4	0.00	0.83
1889	214	0.22	1.05
1892	115	0.12	1.16
1894	4	0.00	1.17
1895	74	0.08	1.24
1896	2863	2.90	4.14
1897	1131	1.15	5.29
1898	3	0.00	5.29
1899	2	0.00	5.30
1901	51	0.05	5.35
1903	1	0.00	5.35

*(continued)*

Jahr	Einzelnormen	% Gesamt	% Kumulativ
1906	1	0.00	5.35
1907	26	0.03	5.38
1908	7	0.01	5.38
1909	118	0.12	5.50
1910	11	0.01	5.51
1911	47	0.05	5.56
1913	45	0.05	5.61
1919	99	0.10	5.71
1920	7	0.01	5.72
1921	40	0.04	5.76
1922	248	0.25	6.01
1923	15	0.02	6.02
1924	2	0.00	6.02
1925	2	0.00	6.03
1926	31	0.03	6.06
1927	32	0.03	6.09
1928	8	0.01	6.10
1929	8	0.01	6.11
1930	14	0.01	6.12
1931	46	0.05	6.17
1933	196	0.20	6.37
1934	512	0.52	6.89
1935	157	0.16	7.04
1936	284	0.29	7.33
1937	320	0.32	7.66
1938	48	0.05	7.71
1939	90	0.09	7.80
1940	215	0.22	8.02
1941	25	0.03	8.04
1942	20	0.02	8.06
1943	24	0.02	8.09

*(continued)*

Jahr	Einzelnormen	% Gesamt	% Kumulativ
1944	12	0.01	8.10
1947	19	0.02	8.12
1948	73	0.07	8.19
1949	314	0.32	8.51
1950	2087	2.12	10.62
1951	584	0.59	11.22
1952	445	0.45	11.67
1953	1815	1.84	13.51
1954	209	0.21	13.72
1955	303	0.31	14.03
1956	700	0.71	14.74
1957	990	1.00	15.74
1958	149	0.15	15.89
1959	750	0.76	16.65
1960	1097	1.11	17.77
1961	1292	1.31	19.08
1962	245	0.25	19.32
1963	338	0.34	19.67
1964	330	0.33	20.00
1965	1462	1.48	21.48
1966	564	0.57	22.06
1967	330	0.33	22.39
1968	478	0.48	22.87
1969	722	0.73	23.61
1970	457	0.46	24.07
1971	776	0.79	24.86
1972	823	0.83	25.69
1973	239	0.24	25.93
1974	900	0.91	26.85
1975	927	0.94	27.79
1976	2492	2.53	30.31

*(continued)*

Jahr	Einzelnormen	% Gesamt	% Kumulativ
1977	747	0.76	31.07
1978	429	0.43	31.50
1979	528	0.54	32.04
1980	1127	1.14	33.18
1981	580	0.59	33.77
1982	542	0.55	34.32
1983	244	0.25	34.57
1984	356	0.36	34.93
1985	623	0.63	35.56
1986	569	0.58	36.14
1987	456	0.46	36.60
1988	1268	1.29	37.89
1989	1142	1.16	39.04
1990	2438	2.47	41.52
1991	798	0.81	42.32
1992	1030	1.04	43.37
1993	1003	1.02	44.39
1994	3875	3.93	48.31
1995	747	0.76	49.07
1996	958	0.97	50.04
1997	1872	1.90	51.94
1998	1492	1.51	53.45
1999	802	0.81	54.27
2000	705	0.71	54.98
2001	1725	1.75	56.73
2002	2339	2.37	59.10
2003	1226	1.24	60.35
2004	2728	2.77	63.11
2005	2030	2.06	65.17
2006	2282	2.31	67.48
2007	1667	1.69	69.17

*(continued)*

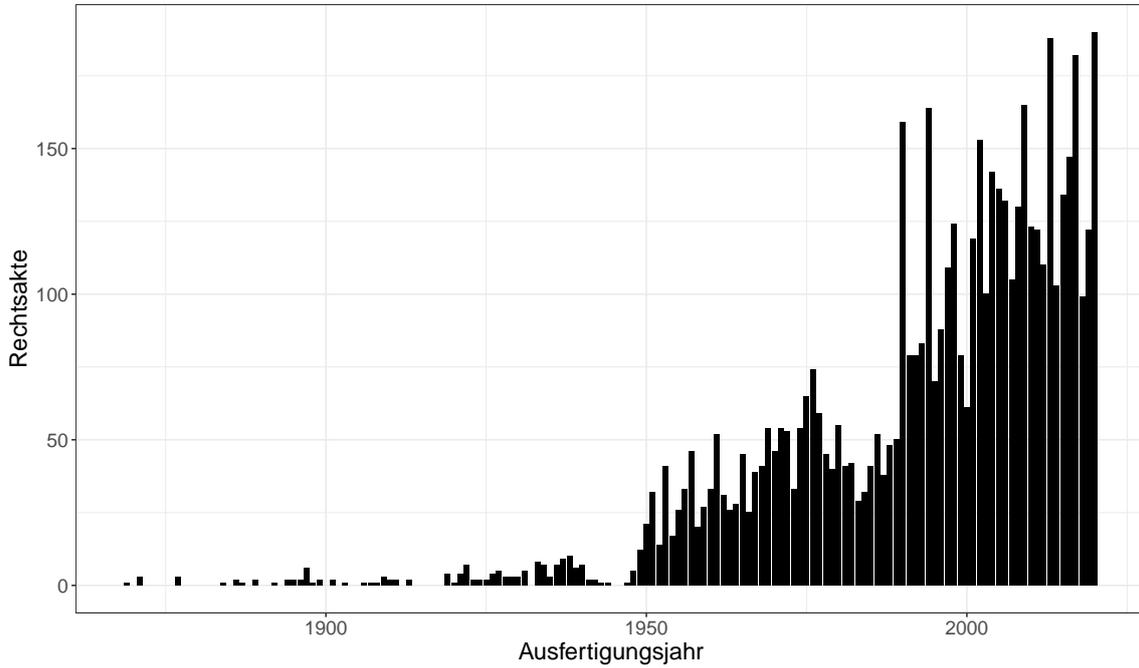
---

Jahr	Einzelnormen	% Gesamt	% Kumulativ
2008	2238	2.27	71.44
2009	2840	2.88	74.32
2010	1455	1.48	75.80
2011	2399	2.43	78.23
2012	1507	1.53	79.76
2013	3435	3.48	83.24
2014	1611	1.63	84.88
2015	2054	2.08	86.96
2016	3109	3.15	90.11
2017	3530	3.58	93.69
2018	1640	1.66	95.35
2019	1721	1.74	97.10
2020	2863	2.90	100.00

---

## 7.2.2 Rechtsakte mit veröffentlichtem Normtext

C-DBR | Version 2021-01-05 | Rechtsakte mit Inhalt je Ausfertigungsjahr



DOI: 10.5281/zenodo.4384771

Jahr	Rechtsakte	% Gesamt	% Kumulativ
1869	1	0.02	0.02
1871	3	0.05	0.07
1877	3	0.05	0.13
1884	1	0.02	0.14
1886	2	0.04	0.18
1887	1	0.02	0.20
1889	2	0.04	0.23
1892	1	0.02	0.25
1894	2	0.04	0.29
1895	2	0.04	0.32
1896	2	0.04	0.36
1897	6	0.11	0.47
1898	1	0.02	0.48
1899	2	0.04	0.52
1901	2	0.04	0.56
1903	1	0.02	0.57
1906	1	0.02	0.59

*(continued)*

Jahr	Rechtsakte	% Gesamt	% Kumulativ
1907	1	0.02	0.61
1908	1	0.02	0.63
1909	3	0.05	0.68
1910	2	0.04	0.72
1911	2	0.04	0.75
1913	2	0.04	0.79
1919	4	0.07	0.86
1920	1	0.02	0.88
1921	4	0.07	0.95
1922	7	0.13	1.08
1923	2	0.04	1.11
1924	2	0.04	1.15
1925	2	0.04	1.19
1926	4	0.07	1.26
1927	5	0.09	1.35
1928	3	0.05	1.40
1929	3	0.05	1.45
1930	3	0.05	1.51
1931	5	0.09	1.60
1933	8	0.14	1.74
1934	7	0.13	1.87
1935	3	0.05	1.92
1936	7	0.13	2.05
1937	9	0.16	2.21
1938	10	0.18	2.39
1939	6	0.11	2.50
1940	7	0.13	2.62
1941	2	0.04	2.66
1942	2	0.04	2.69
1943	1	0.02	2.71
1944	1	0.02	2.73

*(continued)*

Jahr	Rechtsakte	% Gesamt	% Kumulativ
1947	1	0.02	2.75
1948	5	0.09	2.84
1949	12	0.22	3.05
1950	21	0.38	3.43
1951	32	0.57	4.00
1952	14	0.25	4.26
1953	41	0.74	4.99
1954	17	0.31	5.30
1955	26	0.47	5.76
1956	33	0.59	6.36
1957	46	0.83	7.18
1958	20	0.36	7.54
1959	27	0.48	8.03
1960	33	0.59	8.62
1961	52	0.93	9.55
1962	31	0.56	10.11
1963	26	0.47	10.58
1964	28	0.50	11.08
1965	45	0.81	11.89
1966	25	0.45	12.34
1967	39	0.70	13.04
1968	41	0.74	13.77
1969	54	0.97	14.74
1970	46	0.83	15.57
1971	54	0.97	16.54
1972	53	0.95	17.49
1973	33	0.59	18.08
1974	54	0.97	19.05
1975	65	1.17	20.22
1976	74	1.33	21.55
1977	59	1.06	22.61

*(continued)*

Jahr	Rechtsakte	% Gesamt	% Kumulativ
1978	45	0.81	23.42
1979	40	0.72	24.13
1980	55	0.99	25.12
1981	41	0.74	25.86
1982	42	0.75	26.61
1983	29	0.52	27.13
1984	32	0.57	27.71
1985	41	0.74	28.44
1986	52	0.93	29.38
1987	38	0.68	30.06
1988	48	0.86	30.92
1989	50	0.90	31.82
1990	159	2.86	34.67
1991	79	1.42	36.09
1992	79	1.42	37.51
1993	83	1.49	39.00
1994	164	2.94	41.95
1995	70	1.26	43.20
1996	88	1.58	44.78
1997	109	1.96	46.74
1998	124	2.23	48.97
1999	79	1.42	50.39
2000	61	1.10	51.48
2001	119	2.14	53.62
2002	153	2.75	56.37
2003	100	1.80	58.16
2004	142	2.55	60.71
2005	136	2.44	63.15
2006	132	2.37	65.52
2007	105	1.89	67.41
2008	130	2.33	69.74

*(continued)*

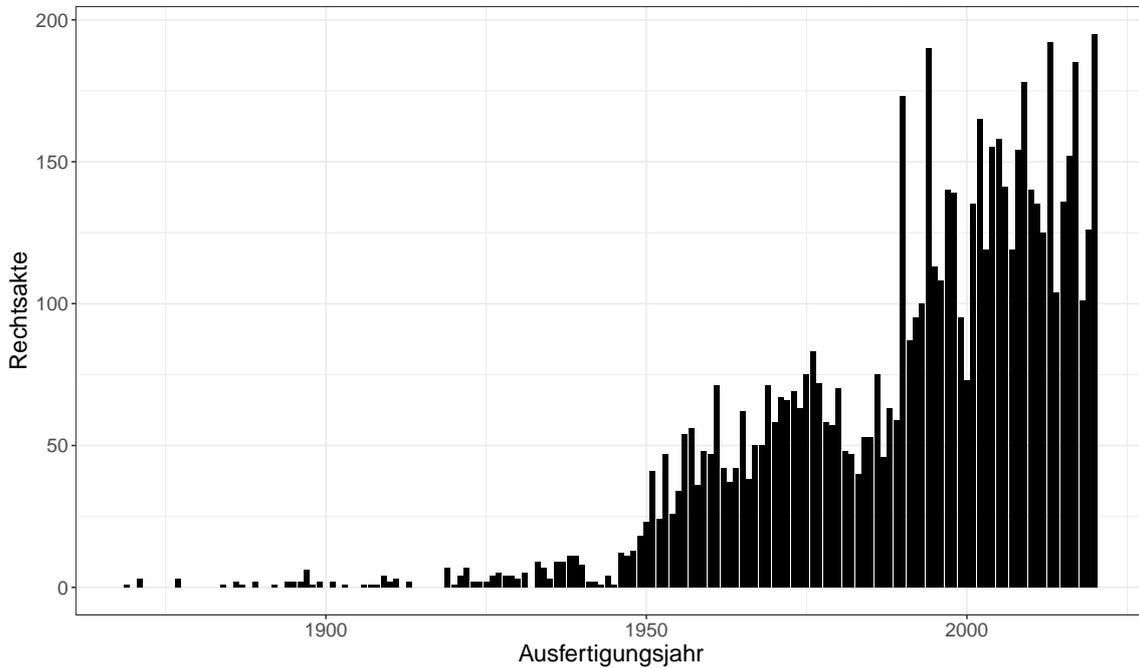
---

Jahr	Rechtsakte	% Gesamt	% Kumulativ
2009	165	2.96	72.71
2010	123	2.21	74.91
2011	122	2.19	77.11
2012	110	1.98	79.08
2013	188	3.38	82.46
2014	103	1.85	84.31
2015	134	2.41	86.71
2016	147	2.64	89.35
2017	182	3.27	92.62
2018	99	1.78	94.40
2019	122	2.19	96.59
2020	190	3.41	100.00

---

### 7.2.3 Alle Rechtsakte (mit und ohne Normtext)

C-DBR | Version 2021-01-05 | Rechtsakte nach Metadaten je Ausfertigungsjahr



DOI: 10.5281/zenodo.4384771

Jahr	Rechtsakte	% Gesamt	% Kumulativ
NA	1	0.02	0.02
1869	1	0.02	0.03
1871	3	0.05	0.08
1877	3	0.05	0.12
1884	1	0.02	0.14
1886	2	0.03	0.17
1887	1	0.02	0.18
1889	2	0.03	0.21
1892	1	0.02	0.23
1894	2	0.03	0.26
1895	2	0.03	0.29
1896	2	0.03	0.32
1897	6	0.09	0.41
1898	1	0.02	0.43
1899	2	0.03	0.46
1901	2	0.03	0.49
1903	1	0.02	0.50

*(continued)*

Jahr	Rechtsakte	% Gesamt	% Kumulativ
1906	1	0.02	0.52
1907	1	0.02	0.53
1908	1	0.02	0.55
1909	4	0.06	0.61
1910	2	0.03	0.64
1911	3	0.05	0.68
1913	2	0.03	0.71
1919	7	0.11	0.82
1920	1	0.02	0.84
1921	4	0.06	0.90
1922	7	0.11	1.00
1923	2	0.03	1.03
1924	2	0.03	1.06
1925	2	0.03	1.09
1926	4	0.06	1.16
1927	5	0.08	1.23
1928	4	0.06	1.29
1929	4	0.06	1.35
1930	3	0.05	1.40
1931	5	0.08	1.47
1933	9	0.14	1.61
1934	7	0.11	1.72
1935	3	0.05	1.76
1936	9	0.14	1.90
1937	9	0.14	2.04
1938	11	0.17	2.20
1939	11	0.17	2.37
1940	8	0.12	2.49
1941	2	0.03	2.52
1942	2	0.03	2.55
1943	1	0.02	2.57

*(continued)*

Jahr	Rechtsakte	% Gesamt	% Kumulativ
1944	4	0.06	2.63
1945	1	0.02	2.65
1946	12	0.18	2.83
1947	11	0.17	3.00
1948	13	0.20	3.19
1949	18	0.27	3.47
1950	23	0.35	3.82
1951	41	0.62	4.44
1952	24	0.36	4.80
1953	47	0.71	5.52
1954	26	0.40	5.91
1955	34	0.52	6.43
1956	54	0.82	7.25
1957	56	0.85	8.10
1958	36	0.55	8.65
1959	48	0.73	9.38
1960	47	0.71	10.10
1961	71	1.08	11.18
1962	42	0.64	11.81
1963	37	0.56	12.38
1964	42	0.64	13.02
1965	62	0.94	13.96
1966	38	0.58	14.54
1967	50	0.76	15.30
1968	50	0.76	16.06
1969	71	1.08	17.14
1970	58	0.88	18.02
1971	67	1.02	19.04
1972	66	1.00	20.04
1973	69	1.05	21.09
1974	63	0.96	22.05

*(continued)*

Jahr	Rechtsakte	% Gesamt	% Kumulativ
1975	75	1.14	23.19
1976	83	1.26	24.45
1977	72	1.09	25.54
1978	58	0.88	26.43
1979	57	0.87	27.29
1980	70	1.06	28.36
1981	48	0.73	29.09
1982	47	0.71	29.80
1983	40	0.61	30.41
1984	53	0.81	31.21
1985	53	0.81	32.02
1986	75	1.14	33.16
1987	46	0.70	33.86
1988	63	0.96	34.82
1989	59	0.90	35.72
1990	173	2.63	38.35
1991	87	1.32	39.67
1992	95	1.44	41.11
1993	100	1.52	42.63
1994	190	2.89	45.52
1995	113	1.72	47.24
1996	108	1.64	48.88
1997	140	2.13	51.01
1998	139	2.11	53.12
1999	95	1.44	54.57
2000	73	1.11	55.68
2001	135	2.05	57.73
2002	165	2.51	60.24
2003	119	1.81	62.05
2004	155	2.36	64.41
2005	158	2.40	66.81

*(continued)*

---

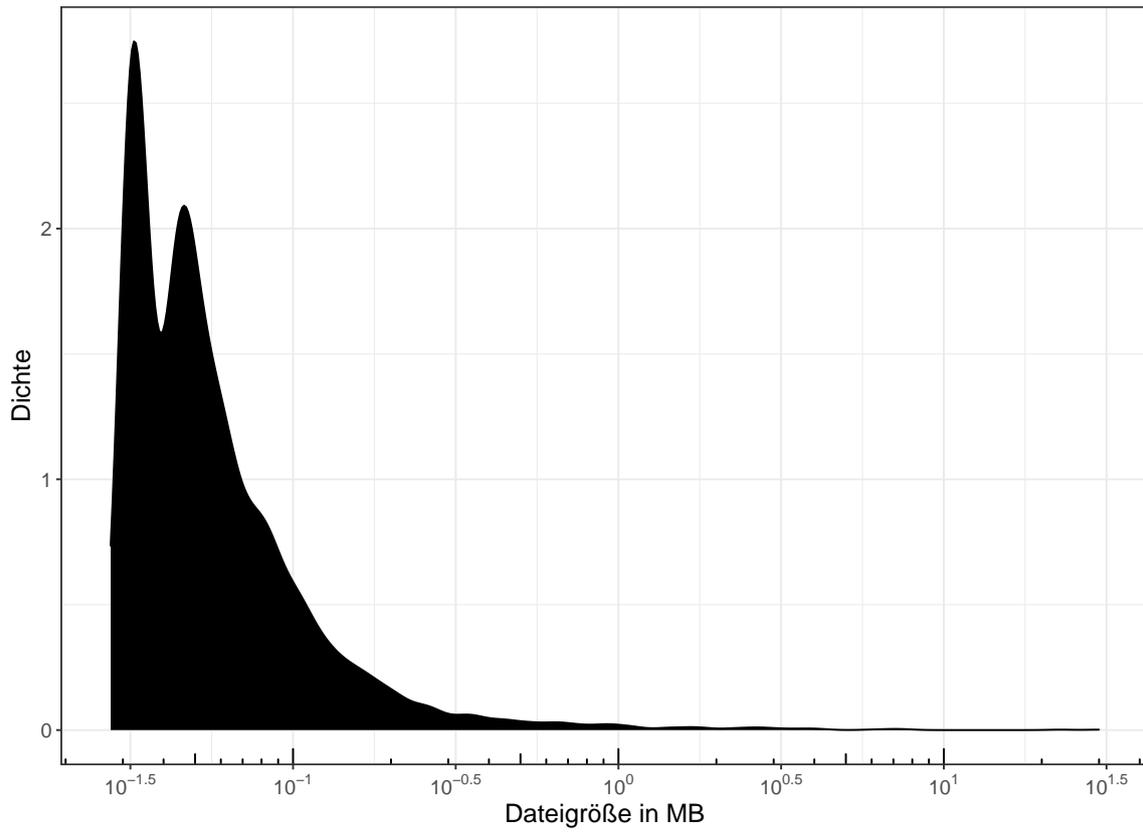
Jahr	Rechtsakte	% Gesamt	% Kumulativ
2006	141	2.14	68.95
2007	119	1.81	70.76
2008	154	2.34	73.10
2009	178	2.71	75.81
2010	140	2.13	77.94
2011	135	2.05	79.99
2012	125	1.90	81.89
2013	192	2.92	84.81
2014	104	1.58	86.39
2015	136	2.07	88.46
2016	152	2.31	90.77
2017	185	2.81	93.58
2018	101	1.54	95.12
2019	126	1.92	97.04
2020	195	2.96	100.00

---

## 8 Dateigrößen: Summen und Verteilungen

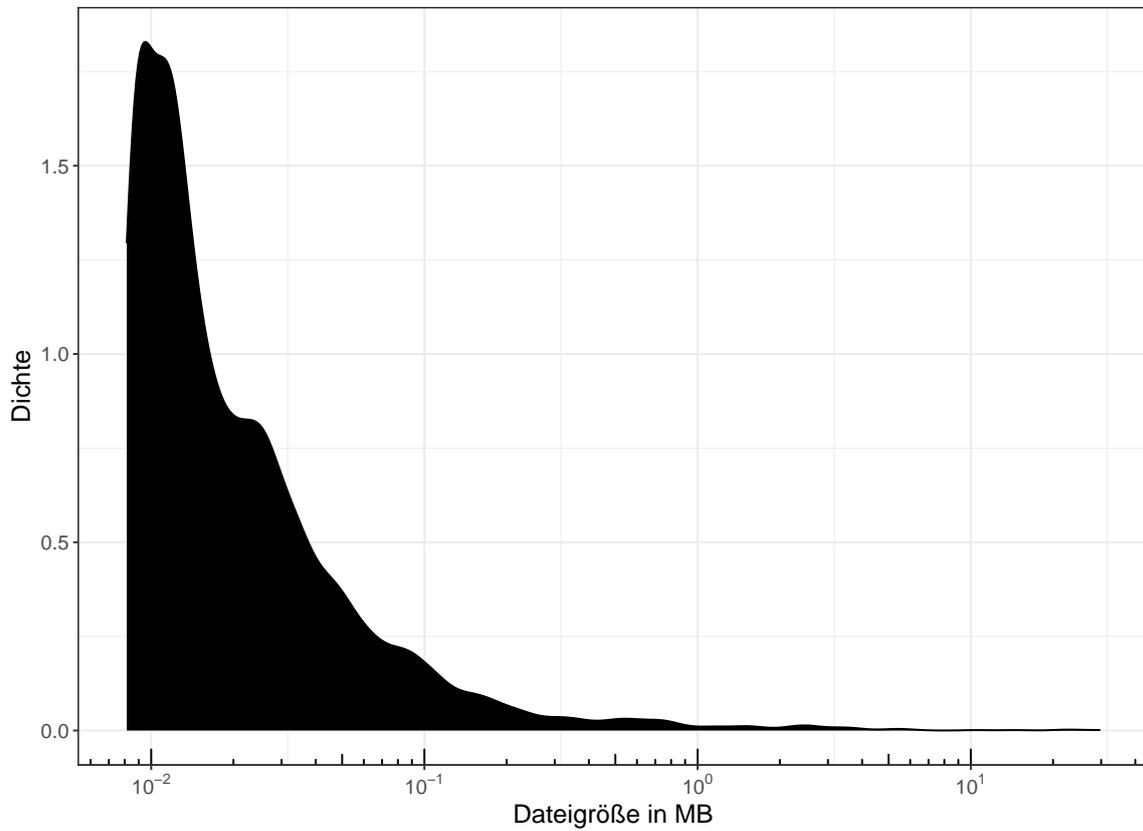
Datei	Größe in MB
C-DBR_2021-01-05_DE_ANALYSE.zip	0.99
C-DBR_2021-01-05_DE_CSV_Einzelnormen_Datensatz.zip	37.16
C-DBR_2021-01-05_DE_CSV_Einzelnormen_Metadaten.zip	2.41
C-DBR_2021-01-05_DE_CSV_MetadatenXML.zip	0.54
C-DBR_2021-01-05_DE_CSV_Rechtsakte_Datensatz.zip	33.72
C-DBR_2021-01-05_DE_CSV_Rechtsakte_Metadaten.zip	0.45
C-DBR_2021-01-05_DE_EPUB_Datensatz.zip	420.94
C-DBR_2021-01-05_DE_PDF_Datensatz.zip	562.13
C-DBR_2021-01-05_DE_TXT_Datensatz.zip	45.61
C-DBR_2021-01-05_DE_XML_Anlagen.zip	254.08
C-DBR_2021-01-05_DE_XML_Datensatz.zip	47.68
C-DBR_2021-01-05_Source_Files.zip	0.05

C-DBR | Version 2021-01-05 | Verteilung der Dateigrößen (PDF)



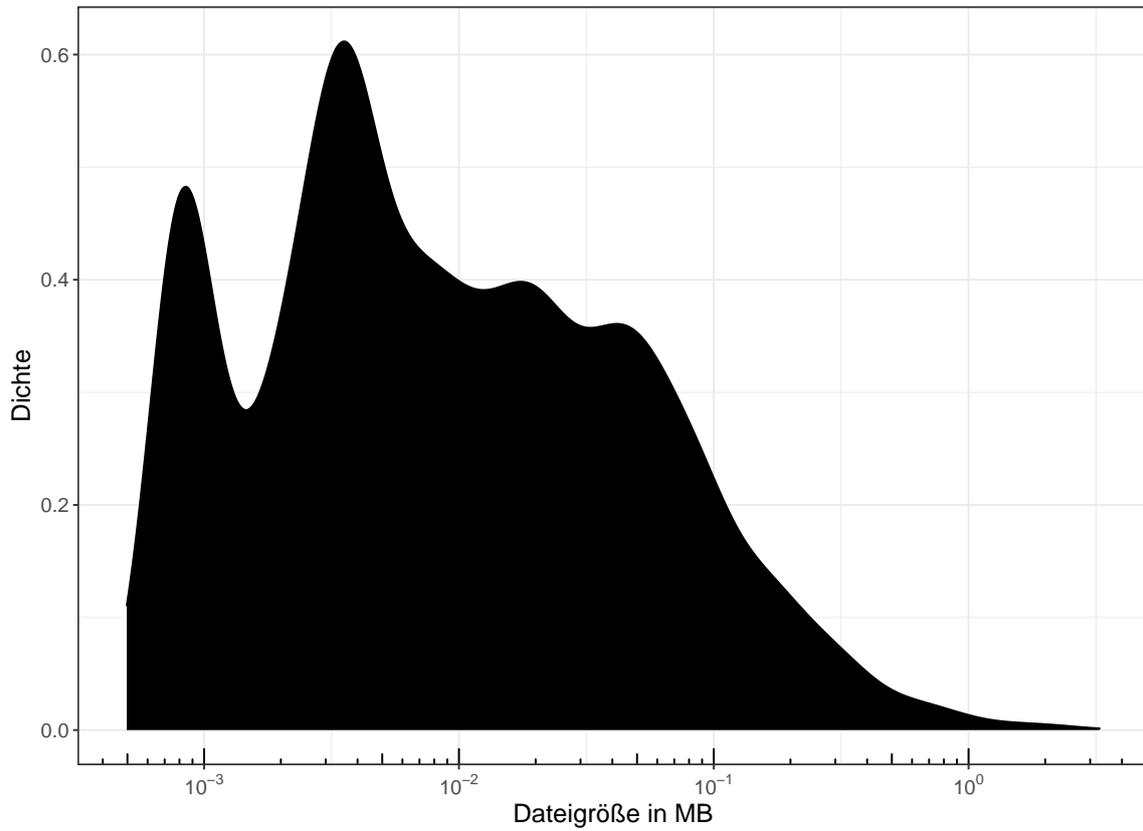
DOI: 10.5281/zenodo.4384771

C-DBR | Version 2021-01-05 | Verteilung der Dateigrößen (EPUB)



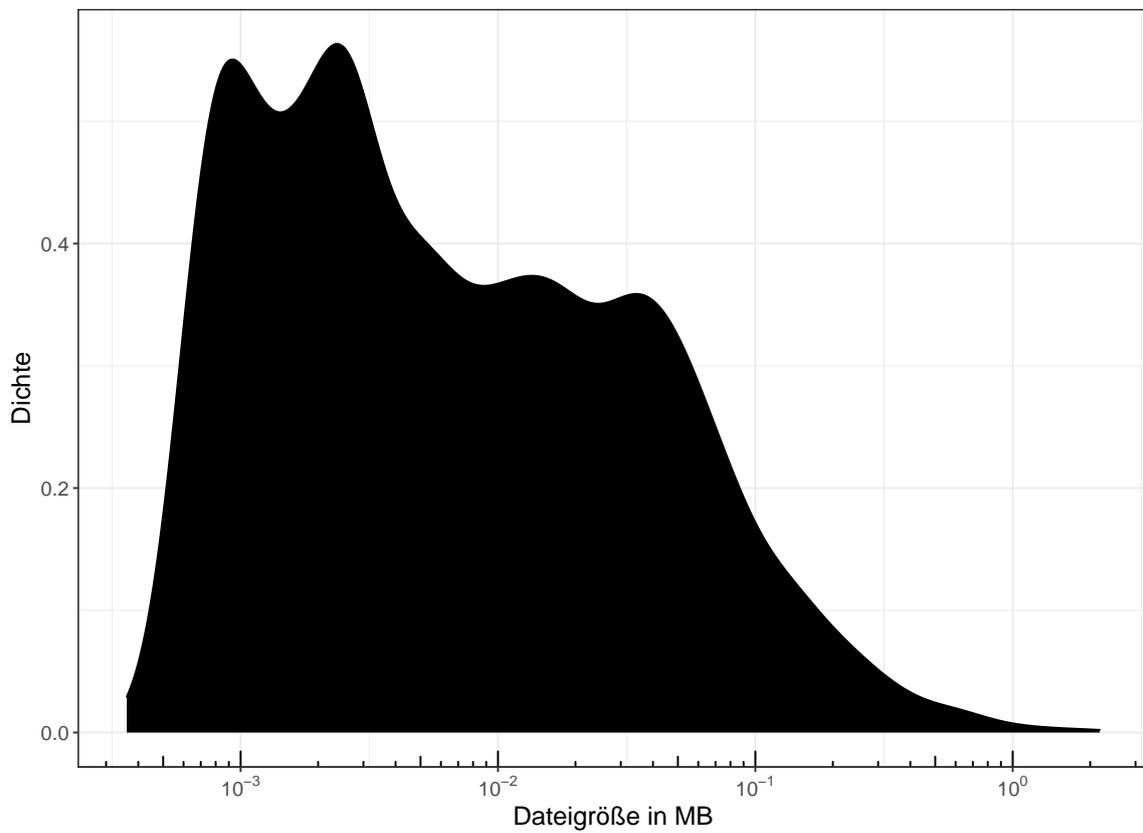
DOI: 10.5281/zenodo.4384771

C-DBR | Version 2021-01-05 | Verteilung der Dateigrößen (XML)



DOI: 10.5281/zenodo.4384771

C-DBR | Version 2021-01-05 | Verteilung der Dateigrößen (TXT)



DOI: 10.5281/zenodo.4384771

## 9 Signaturprüfung

### 9.1 Allgemeines

Die Integrität und Echtheit der einzelnen Archive des Datensatzes sind durch eine Zwei-Phasen-Signatur sichergestellt. In Phase I werden während der Kompilierung für jedes ZIP-Archiv Hash-Werte in zwei verschiedenen Verfahren berechnet und in einer CSV-Datei dokumentiert. In Phase II wird diese CSV-Datei mit meinem persönlichen geheimen GPG-Schlüssel signiert. Dieses Verfahren stellt sicher, dass die Kompilierung von jedermann durchgeführt werden kann, insbesondere im Rahmen von Replikationen, die persönliche Gewähr für Ergebnisse aber dennoch vorhanden bleibt.

Dieses Codebook ist vollautomatisch erstellt und prüft die kryptographisch sicheren SHA3-512 Signaturen (»hashes«) aller ZIP-Archive, sowie die GPG-Signatur der CSV-Datei, welche die SHA3-512 Signaturen enthält. SHA3-512 Signaturen werden durch einen system call zur OpenSSL library auf Linux-Systemen berechnet. Eine erfolgreiche Prüfung meldet »Signatur verifiziert!«. Eine gescheiterte Prüfung meldet »FEHLER!«

### 9.2 Persönliche GPG-Signatur

Die während der Kompilierung des Datensatzes erstellte CSV-Datei mit den Hash-Prüfsummen ist mit meiner persönlichen GPG-Signatur versehen. Der mit dieser Version korrespondierende Public Key ist sowohl mit dem Datensatz als auch mit dem Source Code hinterlegt. Er hat folgende Kenndaten:

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

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

### 9.3 Import: Public Key

```
system2("gpg2", "--import GPG-Public-Key_Fobbe-Data.asc",
        stdout = TRUE,
        stderr = TRUE)
```

```
## [1] "gpg: key 50C41384F44A4E42: \"Sean Fobbe <fobbe-data@posteo.de>\" not
      changed"
## [2] "gpg: Total number processed: 1"
## [3] "gpg:                unchanged: 1"
```

## 9.4 Prüfung: GPG-Signatur der Hash-Datei

```
# CSV-Datei mit Hashes  
print(hashfile)
```

```
## [1] "C-DBR_2021-01-05_KryptographischeHashes.csv"
```

```
# GPG-Signatur  
print(signaturefile)
```

```
## [1] "C-DBR_2021-01-05_FobbeSignatur.gpg"
```

```
# GPG-Signatur prüfen  
testresult <- system2("gpg2",  
                      paste("--verify", signaturefile, hashfile),  
                      stdout = TRUE,  
                      stderr = TRUE)  
  
# Anführungsstriche entfernen um Anzeigefehler zu vermeiden  
testresult <- gsub("'", "", testresult)
```

```
kable(testresult, format = "latex", booktabs=TRUE,  
      longtable=TRUE, col.names = c("Ergebnis"))
```

---

Ergebnis

---

gpg: Signature made Tue 05 Jan 2021 10:52:33 AM CET

gpg: using RSA key FE6FB888F0E5656C1D253B9A50C41384F44A4E42

gpg: Good signature from Sean Fobbe <fobbe-data@posteo.de> [full]

---

## 9.5 Prüfung: SHA3-512 Hashes der ZIP-Archive

```
# Prüf-Funktion definieren
sha3test <- function(filename, sig){
  sig.new <- system2("openssl",
                    paste("sha3-512", filename),
                    stdout = TRUE)
  sig.new <- gsub("^.*\\|= ", "", sig.new)
  if (sig == sig.new){
    return("Signatur verifiziert!")
  }else{
    return("FEHLER!")
  }
}

# Ursprüngliche Signaturen importieren
table.hashes <- fread(hashfile)
filename <- table.hashes$filename
sha3.512 <- table.hashes$sha3.512

# Signaturprüfung durchführen
sha3.512.result <- mcmapply(sha3test, filename, sha3.512, USE.NAMES=FALSE)

# Ergebnis anzeigen
testresult <- data.table(filename, sha3.512.result)
```

```
kable(testresult, format = "latex", booktabs=TRUE,
      longtable=TRUE, col.names = c("Datei", "Ergebnis"))
```

Datei	Ergebnis
C-DBR_2021-01-05_DE_ANALYSE.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_CSV_Einzelnormen_Datensatz.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_CSV_Einzelnormen_Metadaten.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_CSV_MetadatenXML.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_CSV_Rechtsakte_Datensatz.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_CSV_Rechtsakte_Metadaten.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_EPUB_Datensatz.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_PDF_Datensatz.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_TXT_Datensatz.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_XML_Anlagen.zip	Signatur verifiziert!
C-DBR_2021-01-05_DE_XML_Datensatz.zip	Signatur verifiziert!
C-DBR_2021-01-05_Source_Files.zip	Signatur verifiziert!

## 10 Changelog

---

Version	Details
2021-01-05	<ul style="list-style-type: none"><li>• Vollständige Aktualisierung der Daten</li><li>• Komplette Überarbeitung des Source Codes</li><li>• Erstveröffentlichung eines Codebooks</li><li>• Einführung der vollautomatischen Erstellung von Datensatz und Codebook</li><li>• Einführung von Compilation Reports um den Erstellungsprozess exakt zu dokumentieren</li><li>• CSV-Dateien werden durch Parsing der XML-Dateien erstellt</li><li>• Automatisierung und deutliche Erweiterung der Qualitätskontrolle</li><li>• Einführung von Diagrammen zur Visualisierung von Prüfergebnissen</li><li>• Einführung kryptographischer Signaturen</li></ul>
2020-10-09	<ul style="list-style-type: none"><li>• Vollständige Aktualisierung der Daten</li><li>• Erstveröffentlichung des Source Codes</li><li>• XML-Daten nun fehlerfrei. In Version 2020-07-08 waren XML-Dateien mit Anhängen fehlerhaft.</li></ul>
2020-07-08	<ul style="list-style-type: none"><li>• Vollständige Aktualisierung der Daten</li></ul>
2020-05-18	<ul style="list-style-type: none"><li>• Erstveröffentlichung</li></ul>

---

## 11 Parameter für strenge Replikationen

```
## [1] "OpenSSL 1.1.1i FIPS 8 Dec 2020"
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora 32 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib64/libopenblas-r0.3.12.so
##
## locale:
## [1] LC_CTYPE=en_US.utf8      LC_NUMERIC=C
## [3] LC_TIME=en_US.utf8       LC_COLLATE=en_US.utf8
## [5] LC_MONETARY=en_US.utf8   LC_MESSAGES=en_US.utf8
## [7] LC_PAPER=en_US.utf8      LC_NAME=C
## [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.utf8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] parallel stats graphics grDevices utils datasets methods
## [8] base
##
## other attached packages:
## [1] magick_2.5.2 scales_1.1.1 quanteda_2.1.2 data.table_1.13.4
## [5] ggplot2_3.3.2 doParallel_1.0.16 iterators_1.0.13 foreach_1.5.1
## [9] pdftools_2.3.1 kableExtra_1.3.1 knitr_1.30 rvest_0.3.6
## [13] xml2_1.3.2
##
## loaded via a namespace (and not attached):
## [1] tinytex_0.28 qpdf_1.1 tidyselect_1.1.0 xfun_0.19
## [5] purrr_0.3.4 lattice_0.20-41 colorspace_2.0-0 vctrs_0.3.6
## [9] generics_0.1.0 htmltools_0.5.0 usethis_2.0.0 viridisLite
## [13] yaml_2.2.1 rlang_0.4.9 pillar_1.4.7 glue_1.4.2
## [17] withr_2.3.0 selectr_0.4-2 bit64_4.0.5 lifecycle_0.2.0
## [21] stringr_1.4.0 munsell_0.5.0 gtable_0.3.0 codetools_0.2-18
## [25] evaluate_0.14 labeling_0.4.2 curl_4.3 highr_0.8
## [29] Rcpp_1.0.5 RcppParallel_5.0.2 webshot_0.5.2 bit_4.0.4
## [33] farver_2.0.3 fs_1.5.0 fastmatch_1.1-0 stopwords_2.1
## [37] askpass_1.1 digest_0.6.27 stringi_1.5.3 dplyr_1.0.2
## [41] grid_4.0.3 tools_4.0.3 magrittr_2.0.1 tibble_3.0.4
## [45] crayon_1.3.4 pkgconfig_2.0.3 ellipsis_0.3.1 Matrix_1.2-18
## [49] rmarkdown_2.5 httr_1.4.2 rstudioapi_0.13 R6_2.5.0
## [53] compiler_4.0.3
```

## Literaturverzeichnis

- Analytics, Revolution, and Steve Weston. 2020. *Iterators: Provides Iterator Construct*. <https://github.com/RevolutionAnalytics/iterators>.
- 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, Jiong Wei Lua, Jouni Kuha, and William Lowe. 2020. *Quanteda: Quantitative Analysis of Textual Data*. <https://quanteda.io>.
- Corporation, Microsoft, and Steve Weston. 2020. *DoParallel: Foreach Parallel Adaptor for the Parallel Package*. <https://CRAN.R-project.org/package=doParallel>.
- Dowle, Matt, and Arun Srinivasan. 2020. *Data.table: Extension of ‘Data.frame’*. <https://CRAN.R-project.org/package=data.table>.
- Ooms, Jeroen. 2020. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents*. <https://CRAN.R-project.org/package=pdfutils>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Revolution Analytics, and Steve Weston. n.d. *Foreach: Provides Foreach Looping Construct*.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2020. *Rvest: Easily Harvest (Scrape) Web Pages*. <https://CRAN.R-project.org/package=rvest>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2020. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.
- Wickham, Hadley, Jim Hester, and Jeroen Ooms. 2020. *Xml2: Parse Xml*. <https://CRAN.R-project.org/package=xml2>.
- Wickham, Hadley, and Dana Seidel. 2020. *Scales: Scale Functions for Visualization*. <https://CRAN.R-project.org/package=scales>.
- 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. <http://www.crcpress.com/product/isbn/9781466561595>.
- . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- . 2020. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.

Zhu, Hao. 2020. *KableExtra: Construct Complex Table with Kable and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.