

## Checkliste zur Steigerung der Datenqualität von CSV-Dateien

Die nachfolgende Checkliste soll Ihnen dabei helfen, die Qualität Ihrer CSV-Dateien zu überprüfen und auf einfache Weise mögliche Qualitätsprobleme zu identifizieren. Die Checkliste basiert auf dem NQDM-Leitfaden zur Steigerung der Qualität von Daten und Metadaten. Ausführliche Erläuterungen zu untenstehenden Empfehlungen finden Sie im NQDM-Leitfaden jeweils unter der angegebenen Kapitelnummer. Der Leitfaden kann unter folgendem Link heruntergeladen werden:

<https://www.nqdm-projekt.de/de/downloads/leitfaden>

Weitere Informationen über das NQDM-Projekt finden Sie unter:

<https://www.nqdm-projekt.de/>

### Allgemein

- Der Titel des Datensatzes ist prägnant und verständlich (Kapitel 8.2).
- Der Datensatz liegt in einem offenen und maschinenlesbaren Format vor (Kapitel 4.1).
- Für Schlagworte werden bestehende Vokabulare verwendet (Kapitel 8.2).
- Die Daten sind sowohl für Fachexperten als auch für Laien verständlich beschrieben (Kapitel 8.2).
- Personenbezüge in den Daten sind entsprechend dem Verwendungszweck (mögliche Veröffentlichung) anonymisiert (Kapitel 7.1.1).
- Alle in den Daten enthaltenen Datumsangaben entsprechen folgendem Format: *JJJ-MM-TT* (Kapitel 7.1.3).
- Alle in den Daten enthaltenen Zeitangaben entsprechen folgendem Format: *HH:MM:SS* (Kapitel 7.1.3).

### Qualitätsdimensionen (Kapitel 5)

- Aktualität:** Die Daten sind auf dem neuesten Stand.
- Fehlerfreiheit:** Die Daten sind korrekt und fehlerfrei.
- Genauigkeit:** Die Daten sind so präzise wie möglich angegeben.
- Konformität:** Die Daten entsprechen den domänenspezifischen Standards.
- Konsistenz:** Daten und Metadaten sind widerspruchsfrei.
- Vertrauenswürdigkeit:** Der Ursprung der Daten ist kenntlich gemacht.
- Transparenz:** Veränderungen an den Daten ist für Dritte ersichtlich (beispielsweise durch die Angabe einer Versionsnummer).
- Verständlichkeit:** Die Daten sind so strukturiert und bezeichnet, dass Außenstehende dieses leicht verstehen können.
- Vollständigkeit:** Die Daten sind vollständig (beispielsweise sind alle Datenfelder befüllt). Falls die Daten unvollständig sind, wird auf die Unvollständigkeit hingewiesen.
- Zugänglichkeit:** Die Daten können auf einfache Weise abgerufen werden.

Gefördert durch:



aufgrund eines Beschlusses  
des Deutschen Bundestages

## CSV-Dateien (Kapitel 7.2)

- Als Trennzeichen zwischen den einzelnen Werten wird das Semikolon verwendet.
- Die CSV-Datei besteht lediglich aus einem Tabellenblatt.
- Die CSV-Datei enthält neben der Kopfzeile und der Inhaltsdaten keine weiteren Informationen.
- Die CSV-Datei enthält keine visuellen Optimierungen wie zusätzlichen Überschriften oder Leerzeilen enthalten.
- Es gibt eine Kopfzeile mit verständlichen Spaltenüberschriften, die lediglich aus den Buchstaben A-Z und Ziffern 0-9 bestehen.
- Der Datensatz enthält keine Umlaute oder Sonderzeichen.
- Die Anzahl der Werte in jeder Zeile ist identisch mit der Anzahl der Spaltenüberschriften.
- Bei der Angabe von Zahlen wird auf die Darstellung von Tausender-Stellen verzichtet.
- Dezimalzahlen werden mit einem Punkt anstelle eines Kommas dargestellt.
- Zahlenwerte sind pro Spalte einheitlich dargestellt (einheitlich als Ganzzahlen oder einheitlich als Dezimalzahlen).
- Einheiten werden in der Spaltenüberschrift angegeben (z.B. „Ausgaben in Mrd. EUR“).
- Differieren die Einheiten pro Spalte, wird eine extra Spalte zur Angabe der Einheit genutzt.
- Alle Null-Werte sind kenntlich gemacht (z.B. durch „kein Wert“ oder „null“).
- Alle Werte in der CSV-Datei entsprechen der UTF-8-Zeichenkodierung.
- Werte, die ein Semikolon oder andere Zeichen enthalten, die interpretiert werden könnten, sind entsprechend maskiert.

## Hilfreiche Tools und Quellen

### Beschreibung

### Link

- |  |   |
|--|---|
| <ul style="list-style-type: none"><li>▪ CSV-Spezifikation</li><li>▪ Sammlung hilfreicher Tools für CSV</li><li>▪ Webseite zur Prüfung von CSV-Dateien auf Konsistenz und Konformität</li></ul> | <ul style="list-style-type: none"><li><a href="https://tools.ietf.org/html/rfc4180">https://tools.ietf.org/html/rfc4180</a></li><li><a href="https://onlinecsvtools.com/">https://onlinecsvtools.com/</a></li><li><a href="https://csvlint.io/">https://csvlint.io/</a></li></ul> |
|--|---|

Gefördert durch:



aufgrund eines Beschlusses  
des Deutschen Bundestages