

# Open Data Sandbox

## Vorname Name<sup>1</sup>

<sup>1</sup> Robert Koch-Institut

## Zitieren

Name, V. (2025). Open Data Sandbox [Data set]. Zenodo. <https://doi.org/10.5072/zenodo.324600>

## Zusammenfassung

## Inhaltsverzeichnis

- Beispieltext
  - Hinweise zur Nachnutzung der Daten
  - Appendix

## Inhaltsverzeichnis

## Example text

*Etiam velit utrumque. Ut enim ad minimus quisque deus est, et non nisi deus. Quod si deus est, non potest non esse bene. Non potest non esse bene, quod si deus est, non potest non esse bene.*

Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a,

## Variables and values

Die Datei [Sandbox\\_Data.tsv](#) enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen. Ein maschinenlesbares Datenschema ist im [Data Package Standard](#) in [tableschema\\_Sandbox\\_Data.json](#) hinterlegt:

## tableschema\_Sandbox\_Data.json

Variable	Typ	Ausprägungen	Beschreibung
date_and_time	date	Format: YYYY-MM-DDTHH:MM:SS	Datum der Probeentnahme im ISO 8601 Format ohne Zeitzone
date	date	Format: YYYY-MM-DD	Datum der Probeentnahme im ISO 8601 Format ohne Zeitzone
week_as_date	date	Beispiel: 2021-04 Format: YYYY-ww	Berichtswoche im ISO-8601 Format ohne Zeitzone
season	string	Beispiel: 2012/13	Saison jeweils von Kalenderwoche 40 bis Kalenderwoche 39 des Folgejahres (z.B. Saison 2012/13 = 2012W40 bis 2013W39)
text_with_fixed_set_of_possible_values	string	Werte: random, requested, clinical, unknown, other	Mögliche Werte: random, requested, clinical, unknown, other
		Beispiel: 873a7cc28d29e3f17b0544ea6e9e84	Eine Variable

text_with_very_long_example	string	36defe32f6d60649159ee8ac78d414 7ac9	die eine lange ID enthält
number_with_minimum	number	Werte: ≥2.5	Gleitkommazahl mit inklusivem Minimum
integer_with_range	integer	Werte: 0 - 99999 Beispiel: 1095	Ganze Zahl (Zählwerte)
integer_with_missing_values	integer	Werte: ≥-1 Fehlende Werte: NA	Ganze Zahl mit fehlenden Werten
text_with_json_example	string	Beispiele: [{"method": "PANGOLIN_LATEST", "classification_version": "PUSHER-v1.28.1", "tool_version": "4.3", "lineage": "BA.2", "@qc_notes": "Ambiguous_content:0.02", "@is_designated": False, "@qc_status": "pass", "@conflict": 0.0, "@note": "Usher placements: BA.2(1/1)"}]	Text mit Beispiel im JSON-Format
unique_variable	integer		Jeder Wert kommt nur einmal im Datensatz vor.
url	string	Beispiel: <a href="https://www.gbe.rki.de/DE/Themen/EinflussfaktorenAufDieGesundheit/GesundheitsUndRisikoverhalten/Alkoholkonsum/Rauschtrinken/rauschtrinken_node.html">https://www.gbe.rki.de/DE/Themen/EinflussfaktorenAufDieGesundheit/GesundheitsUndRisikoverhalten/Alkoholkonsum/Rauschtrinken/rauschtrinken_node.html</a>	Link zu Ressource

Die Datei [Sandbox\\_Data\\_lfs.tsv](#) enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen. Ein maschinenlesbares Datenschema ist im [Data Package Standard](#) in [tableschema\\_Sandbox\\_Data\\_lfs.json](#) hinterlegt:

#### [tableschema\\_Sandbox\\_Data\\_lfs.json](#)

Variable	Typ	Ausprägungen	Beschreibung
date_and_time	date	Format: YYYY-MM-DDTHH:MM:SS	Datum der Probeentnahme im ISO 8601 Format ohne Zeitzone
date	date	Format: YYYY-MM-DD	Datum der Probeentnahme im ISO 8601 Format ohne Zeitzone

week_as_date	date	Beispiel: 2021-04 Format: YYYY-ww	Berichtswoche im ISO-8601 Format ohne Zeitzone
season	string	Beispiel: 2012/13	Saison jeweils von Kalenderwoche 40 bis Kalenderwoche 39 des Folgejahres (z.B. Saison 2012/13 = 2012W40 bis 2013W39)
text_with_fixed_set_of_possible_values	string	Werte: random , requested , clinical , unknown, other	Mögliche Werte: random , requested , clinical , unknown, other
text_with_very_long_example	string	Beispiel: 873a7cc28d29e3f17b0544ea6e9e84 36defe32f6d60649159ee8ac78d414 7ac9	Eine Variable die eine lange ID enthält
number_with_minimum	number	Werte: ≥2.5	Gleitkommazahl mit inklusivem Minimum
integer_with_range	integer	Werte: 0 - 99999 Beispiel: 1095	Ganze Zahl (Zählwerte)
integer_with_missing_values	integer	Werte: ≥-1 Fehlende Werte: NA	Ganze Zahl mit fehlenden Werten
text_with_json_example	string	Beispiele: [ { 'method': 'PANGOLIN_LATEST' , 'classification_version': 'PUSHER-v1.28.1' , 'tool_version': '4.3' , 'lineage': 'BA.2' , '@qc_notes': 'Ambiguous_content:0.02' , '@is_designated': False , '@qc_status': 'pass' , '@conflict': 0.0 , '@note': 'Usher placements: BA.2(1/1)' } ]	Text mit Beispiel im JSON-Format
unique_variable	integer		Jeder Wert kommt nur einmal im Datensatz vor.
		Beispiel: <a href="https://www.gbe.rki.de/DE/Themen/EinflussfaktorenAufDieGesundheit.html">https://www.gbe.rki.de/DE/Themen/EinflussfaktorenAufDieGesundheit.html</a>	Link zu

url	string	dheit/GesundheitsUndRisikoverh alten/Alkoholkonsum/Rauschtrin ken/rauschtrinken_node.html	Ressource
-----	--------	---	-----------

## Metadaten

Zur Erhöhung der Auffindbarkeit sind die bereitgestellten Daten mit Metadaten beschrieben. Über GitHub Actions werden Metadaten an die entsprechenden Plattformen verteilt. Für jede Plattform existiert eine spezifische Metadatendatei, diese sind im Metadatenordner hinterlegt:

### Metadaten/

Versionierung und DOI-Vergabe erfolgt über [Zenodo.org](#). Die für den Import in Zenodo bereitgestellten Metadaten sind in der [zenodo.json](#) hinterlegt. Die Dokumentation der einzelnen Metadatenvariablen ist unter <https://developers.zenodo.org/#representation> nachlesbar.

### Metadaten/zenodo.json

In der zenodo.json ist neben dem Publikationsdatum (`"publication_date"`) auch der Datenstand in folgendem Format enthalten (Beispiel):

```
"dates": [
  {
    "start": "2023-09-11T15:00:21+02:00",
    "end": "2023-09-11T15:00:21+02:00",
    "type": "Created",
    "description": "Date when the published data was created"
  }
],
```

Zusätzlich beschreiben wir tabellarische Daten mithilfe des [Data Package Standards](#).

Ein Data Package ist eine strukturierte Sammlung von Daten und zugehörigen Metadaten, die den Austausch und die Wiederverwendung von Daten erleichtert. Es besteht aus einer datapackage.json-Datei, die zentrale Informationen wie die enthaltenen Ressourcen, ihre Formate und Schema-Definitionen beschreibt.

Der Data Package Standard wird von der [Open Knowledge Foundation](#) bereitgestellt und ist ein offenes Format, das eine einfache, maschinenlesbare Beschreibung von Datensätzen ermöglicht.

Die Liste der in diesem Repository enthaltenen Daten ist in folgender Datei hinterlegt:

### datapackage.json

Für tabellarische Daten definieren wir zusätzlich ein [Table Schema](#), das die Struktur der Tabellen beschreibt, einschließlich Spaltennamen, Datentypen und Validierungsregeln. Diese Schema-Dateien finden sich unter:

### Metadaten/schemas/

## Hinweise zur Nachnutzung der Daten

Offene Forschungsdaten des RKI werden auf [Zenodo.org](#), [GitHub.com](#), [OpenCoDE](#) und [Edoc.rki.de](#) bereitgestellt:

- <https://zenodo.org/communities/robertkochinstitut>
- <https://github.com/robert-koch-institut>

- <https://gitlab.opencode.de/robert-koch-institut>
- <https://edoc.rki.de/>

## Lizenz

Der Datensatz "Open Data Sandbox" ist lizenziert unter der [Creative Commons Namensnennung 4.0 International Public License | CC-BY 4.0 International](#).

Die im Datensatz bereitgestellten Daten sind, unter Bedingung der Namensnennung des Robert Koch-Instituts als Quelle, frei verfügbar. Das bedeutet, jede Person hat das Recht die Daten zu verarbeiten und zu verändern, Derivate des Datensatzes zu erstellen und sie für kommerzielle und nicht kommerzielle Zwecke zu nutzen. Weitere Informationen zur Lizenz finden sich in der [LICENSE](#) bzw. [LIZENZ](#) Datei des Datensatzes.

## Appendix

---

This is an example appendix .