

Open Data Sandbox

Vorname Name¹

¹ Robert Koch Institute

Cite

Name, V. (2026). Open Data Sandbox [Data set]. Zenodo. <https://doi.org/10.5072/zenodo.318172>

Abstract

[illegible]

Table of Content

- Example text
- Guidelines for reuse of the data
- Appendix

Inhaltsverzeichnis

Example text

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec

quam felis, ultricies nec, pellentesque eu, pretium quis, sem.

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

The file [Sandbox_Data.tsv](#) contains the variables and their values shown in the following table. A machine-readable data schema is stored in [Data Package Format](#) in [tableschema_Sandbox_Data.en.json](#):

[tableschema_Sandbox_Data.en.json](#)

Variable	Type	Characteristic	Description
date_and_time	date	Format: YYYY-MM-DDTHH:MM:SS	
date	date	Format: YYYY-MM-DD	
week_as_date	date	Example: 2021-04 Format: YYYY-ww	
season	string	Example: 2012/13	
text_with_fixed_set_of_possible_values	string	Values: random, requested, clinical, unknown, other	
text_with_very_long_example	string	Example: 873a7cc28d29e3f17b0544ea6e9e8436defe32f6d60649159ee8ac78d4147ac9	
number_with_minimum	number	Values: ≥ 2.5	
integer_with_range	integer	Values: 0 - 99999 Example: 1095	
integer_with_missing_values	integer	Values: ≥ -1 Missing values: NA	
text_with_json_example	string	Examples: [{'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',	

Variable	Type	Characteristic	Description
		'@conflict': 0.0, '@note': 'Usher placements: BA.2(1/1)'}]	
unique_variable	integer		
url	string	Example: https://www.gbe.rki.de/DE /Themen/EinflussfaktorenA ufDieGesundheit/Gesundhei tsUndRisikoverhalten/Alko holkonsum/Rauschtrinken/r auschtrinken_node.html	

The file [Sandbox_Data_lfs.tsv](#) contains the variables and their values shown in the following table. A machine-readable data schema is stored in [Data Package Format](#) in [tableschema_Sandbox_Data_lfs.en.json](#):

[tableschema_Sandbox_Data_lfs.en.json](#)

Variable	Type	Characteristic	Description
date_and_time	date	Format: YYYY-MM-DDTHH:MM:SS	
date	date	Format: YYYY-MM-DD	
week_as_date	date	Example: 2021-04 Format: YYYY-ww	
season	string	Example: 2012/13	
text_with_fixed_set_of_possible_values	string	Values: random, requested, clinical, unknown, other	
text_with_very_long_example	string	Example: 873a7cc28d29e3f17b0544ea6 e9e8436defe32f6d60649159e e8ac78d4147ac9	
number_with_minimum	number	Values: ≥2.5	
integer_with_range	integer	Values: 0 - 99999 Example: 1095	
integer_with_missing_values	integer	Values: ≥-1 Missing values: NA	
text_with_json_example	string	Examples: [{'method': 'PANGOLIN_LATEST', 'classification_version': 'PUSHER- v1.28.1' , 'tool_version': '4.3',	

Variable	Type	Characteristic	Description
		<pre>'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)'}]</pre>	
unique_variable	integer		
url	string	Example: <code>https://www.gbe.rki.de/DE/Themen/EinflussfaktorenAufDieGesundheit/GesundheitsUndRisikoverhalten/Alkoholkonsum/Rauschtrinken/rauschtrinken_node.html</code>	

Metadata

To increase findability, the provided data are described with metadata. The Metadata are distributed to the relevant platforms via GitHub Actions. There is a specific metadata file for each platform; these are stored in the metadata folder:

[Metadaten/](#)

Versioning and DOI assignment are performed via [Zenodo.org](#). The metadata prepared for import into Zenodo are stored in the `zenodo.json`. Documentation of the individual metadata variables can be found at <https://developers.zenodo.org/representation>.

[Metadaten/zenodo.json](#)

The `zenodo.json` includes the publication date and the date of the data status in the following format (example):

```
"publication_date": "2024-06-19",
"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"
  }
],
```

Additionally, we describe tabular data using the [Data Package Standard](#).

A Data Package is a structured collection of data and associated metadata that facilitates data exchange and reuse. It consists of a `datapackage.json` file that contains key information such as the included resources, their formats, and schema definitions.

The Data Package Standard is provided by the [Open Knowledge Foundation](#) and is an open format that enables a simple, machine-readable description of datasets.

The list of data included in this repository can be found in the following file:

[datapackage.json](#)

For tabular data, we additionally define a [Table Schema](#) that describes the structure of the tables, including column names, data types, and validation rules. These schema files can be found in:

[Metadaten/schemas/](#)

Guidelines for reuse of the data

Open data from the RKI are available on [Zenodo.org](#), [GitHub.com](#), [OpenCoDE](#), and [Edoc.rki.de](#):

- <https://zenodo.org/communities/robertkochinstitut>
- <https://github.com/robert-koch-institut>
- <https://gitlab.opencode.de/robert-koch-institut>
- <https://edoc.rki.de/>

License

The "Open Data Sandbox" dataset is licensed under the [Creative Commons Attribution 4.0 International Public License | CC-BY](#).

The data provided in the dataset are freely available, with the condition of attributing the Robert Koch Institute as the source, for anyone to process and modify, create derivatives of the dataset and use them for commercial and non-commercial purposes.

Further information about the license can be found in the [LICENSE](#) or [LIZENZ](#) file of the dataset.

Appendix

This is an example appendix 📁.