

# Abwassersurveillance AMELAG

---

Robert Koch-Institut | RKI

Nordufer 20  
13353 Berlin

**Fachgebiet 32 | Surveillance und elektronisches Melde- und Informationssystem (DEMIS) | ÖGD-Kontaktstelle**

## Zitieren

Fachgebiet 32, Robert Koch-Institut (2023): Abwassersurveillance AMELAG, Berlin: Zenodo. DOI: [10.5281/zenodo.10782702](https://doi.org/10.5281/zenodo.10782702)

---

## Informationen zum Datensatz und Entstehungskontext

---

Das Vorhaben „Abwassermonitoring für die epidemiologische Lagebewertung“ (AMELAG) läuft vom 22.11.2022 bis zum 31.12.2024. Behörden, Kläranlagen und Labore arbeiten zusammen, um Proben zu nehmen, zu analysieren und zu bewerten. Das Ziel dieses Vorhabens ist es, SARS-CoV-2-Nachweise aus dem Abwasser als zusätzlichen Indikator zur epidemiologischen Lagebewertung auf Länder- und Bundesebene zu etablieren. Ebenso ist es das Ziel, Strukturen und Prozesse für ein bundesweites Netzwerk für die Abwassersurveillance weiter auszubauen, Konzepte für eine Verstetigung zu erstellen und die Möglichkeiten für ein Monitoring von weiteren Krankheitserregern im Abwasser zu erforschen.

Abwassersurveillance ist eine Technik, um Erreger im Abwasser nachzuweisen, um Gesundheitsschutzmaßnahmen besser steuern zu können. Abwasserdaten erlauben keine genaue Einschätzung von Krankheitsschwere oder der Belastung des Gesundheitssystems. Bei der epidemiologischen Bewertung sollten die Daten mit anderen Indikatoren, z.B. aus der syndromischen Surveillance, kombiniert werden.

## Administrative und organisatorische Angaben

AMELAG ist ein vom [Bundesministerium für Gesundheit \(BMG\)](#) gefördertes Vorhaben und wird in Kooperation mit dem Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und [Verbraucherschutz \(BMUV\)](#) durchgeführt.

Das Vorhaben wird vom Robert Koch-Institut (RKI) und [Umweltbundesamt \(UBA\)](#) gemeinsam durchgeführt. Weitere Informationen zu AMELAG finden Sie auf der [Projektwebseite](#).

Die Durchführung der Probenahme erfolgt durch die teilnehmenden Kläranlagen. Die Analyse der Proben erfolgt durch die teilnehmenden Labore. Neben kommerziellen Laboren, Landeslaboren und dem Umweltbundesamt führt der zentrale Sanitätsdienst der Bundeswehr einen Teil der Analytik durch.

Ein Teil der Kläranlagen und Labore sind gleichzeitig in Projekten der Bundesländer zur Abwassersurveillance beteiligt (Baden-Württemberg, Bayern, Berlin, Brandenburg, Hamburg, Hessen, Rheinland-Pfalz, Sachsen-Anhalt).

Weitere Kläranlagen und Labore sind Teil der folgenden Forschungsprojekte:

- [WBeready](#) - Ein Forschungskonsortium bestehend aus Emschergenossenschaft und LippeverbandEGLV, Forschungsinstitut für Wasserwirtschaft und Klimazukunft an der RWTH Aachen FiW, Universitätsklinikum Frankfurt, Goethe-Universität Frankfurt am Main, Universitätsmedizin Essen (Institut für künstliche Intelligenz, Institut für Urban Public Health), RWTH Aachen, Institut für Siedlungswasserwirtschaft.
- Etablierung von Verfahren für den Nachweis von Viren im Abwasser zur Bewertung der Infektionslage in der Bevölkerung (Universität Dresden)
- Entwicklung einer landesweiten Abwassersurveillance in Thüringen mittels Mobilitätsdaten und künstlicher Intelligenz (Forschungskonsortium der Universität Weimar, Universität Jena, Universität Hamburg, Hochschule Hamm-Lippstadt, SMA Development GmbH, KOWUG Kommunale Wasser- und Umwelttechnik GmbH, Analytik Jena GmbH)
- Etablierung einer Multiplex-PCR aus Abwasser und für Detektion und Charakterisierung von RSV im Rahmen des SARS-CoV-2-Abwasser-Monitoring (AMELAG) (Universität Bonn und Düsseldorf).

Die Firma [ENDa](#) wurde mit der Datenhaltung beauftragt. Die erhobenen Daten werden dort in einer Datenbank (PiA-Monitor ) gespeichert und weiterverarbeitet.

Die Verarbeitung, Aufbereitung und Veröffentlichung der Daten erfolgen durch das Fachgebiet MF 4 | Fach- und Forschungsdatenmanagement. Fragen zum Datenmanagement und zur Publikationsinfrastruktur können an das Open Data-Team des Fachgebiets MF4 unter [OpenData@rki.de](mailto:OpenData@rki.de) gerichtet werden.

## **Datenerhebung**

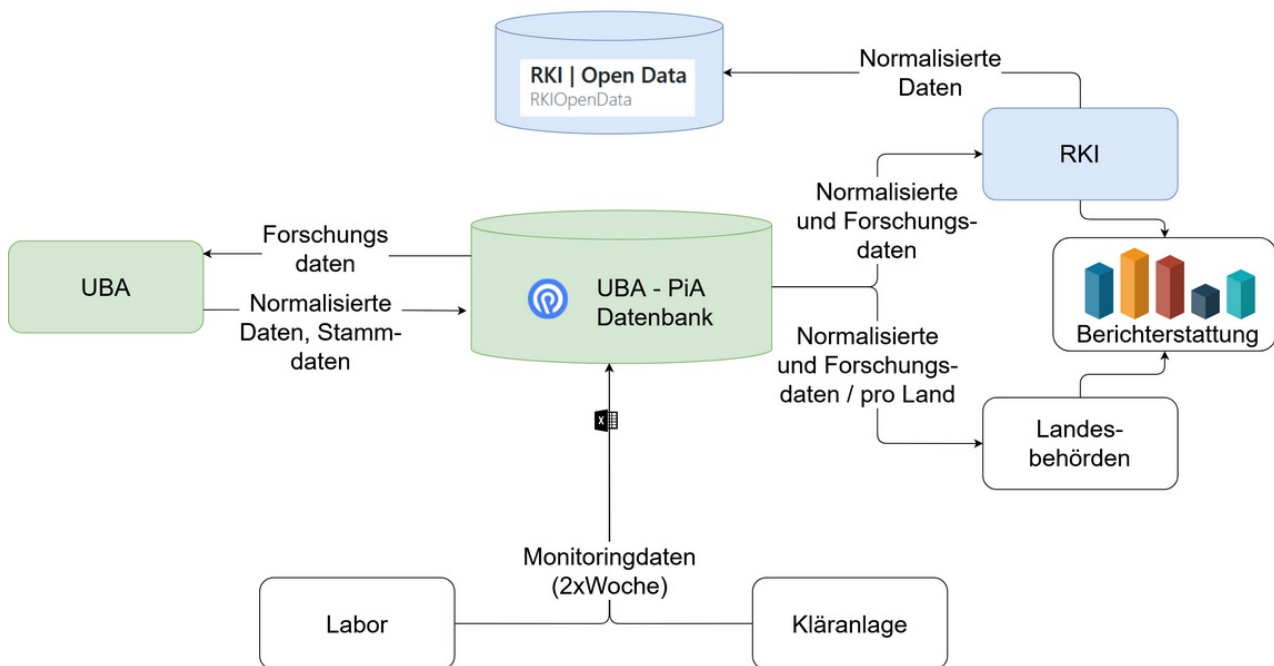
In AMELAG wurden aufbauend auf die im Rahmen des [ESI-CorA-Projekts](#) erstellten Handreichungen zur Probennahme und Laboranalytik [technische Leitfäden](#) entwickelt. Die Rohdaten der im ESI-CorA-Projekt analysierten Proben sind in AMELAG nachgenutzt und in den ausgewerteten Daten enthalten.

An jeder beteiligten Kläranlage werden in aller Regel zwei Mal pro Woche Rohabwasserproben entnommen und zusammen mit den Begleitparametern (z.B. Volumenstrom, pH-Wert, Temperatur), die für die Normalisierung und Qualitätssicherung nötig sind, erhoben. Die Rohabwasserproben sollen, sofern möglich, nach dem Sandfang der Kläranlage entnommen werden. Es wird eine 24-Stunden-Mischprobe entnommen, welche mit einem automatischen Probennehmer durchgeführt wird. Die 24-Stunden-Probennahmen erfolgen in der Regel jeweils montags auf dienstags und mittwochs auf donnerstags. Im Regelfall wird ein Liter der Probe in Probenflaschen abgefüllt und an das Analyselabor versendet.

Im Labor erfolgt die Aufkonzentrierung, Extraktion der viralen Nukleinsäure und Quantifizierung der viralen Gensequenzen durch digitale PCR (dPCR) oder quantitative real time PCR (qRT-PCR). Mindestens zwei repräsentative SARS-CoV-2 Genfragmente (Vorzugsweise N1, N2, E, ORF oder RdRp) werden bestimmt.

Robert Koch-Institut, Fachgebiet 32 (2024): "ESI-CorA: SARS-CoV-2-Abwassersurveillance" [Data set]. Zenodo. DOI: [10.5281/zenodo.10781653](https://doi.org/10.5281/zenodo.10781653)

## Datenfluss



Beim UBA laufen die Metadaten zu den Kläranlagenstandorten und den Laboren sowie die regelmäßig erhobenen Monitoringdaten zentral in einer Webanwendung, dem PiA-Monitor (Pathogene im Abwasser), zusammen, werden dort gespeichert und weiterverarbeitet. Die regelmäßig zu erfassenden Monitoringdaten der Kläranlagen und die Analysedaten der Labore werden zusammengeführt und von den datenliefernden Stellen über die Web-Anwendung der Datenbank importiert. Das Umweltbundesamt, das RKI und die Bundesländer können auf die Daten im Rahmen ihrer jeweiligen Rechte zugreifen.

## Plausibilitätsprüfung und Weiterverarbeitung der Daten

Mit dem Datenimport werden die Daten auf Plausibilität geprüft. Dabei werden die Formate, Vollständigkeit der Angaben (Pflichtfeldangaben), Wertebereiche der Monitoringdaten, Plausibilität der Datumsangaben und die Übereinstimmung mit hinterlegten Metadaten geprüft. Nur Datensätze, welche die Qualitätsprüfung erfolgreich durchlaufen, werden auch in die Datenbank importiert. Es wird der geometrische Mittelwert der Viruslast (Genkopien/Liter) aus den zwei oder mehr gemessenen Zielgenen ermittelt.

## Normalisierungsverfahren

Eine variierende Abwasserzusammensetzung, z. B. aufgrund von unregelmäßigen industriellen Einflüssen oder Starkregenereignissen, kann zu veränderten Konzentrationen von SARS-CoV-2 führen. Um diese externen Einflüsse zu berücksichtigen, kann die gemessene Viruslast normalisiert werden.

In AMELAG wird nach Durchfluss normalisiert. Dabei ist der Trockenwetterzufluss der Kläranlage die Referenz. Folgende Formel wurde hierbei verwendet:

$$Gene_{normalisiert} = \frac{Q_{KA\_aktuell}}{Q_{KA\_median}} \cdot Gene_{gemittelt}$$

wo:

- $Q_{KA\_aktuell}$  : Volumenstrom der Kläranlage im Probenahmezeitraum und
- $Q_{KA\_median}$  : Median des Volumenstrom der Kläranlage

Die Normalisierung erfolgt automatisiert mit dem Datenimport.

## Datenauswertung

Die Auswertung der Daten erfolgt am RKI über R-Skripte. Die Skripte sind in den [Kontextmaterialien](#) enthalten. Eine genaue Beschreibung der Methodik ist in den [technische Leitfäden](#) hinterlegt. Die Ergebnisse werden in einem wöchentlichen Bericht des RKI [Wochenbericht](#) veröffentlicht.

Für jeden Standort werden die Messwerte in Genkopien pro Liter (Genkopien/L) angegeben. Zusätzlich werden die Messwerte der logarithmierten normalisierten Genkopien mittels einer lokal gewichteten Regression (LOESS) geglättet und zugehörige Konfidenzintervalle berechnet. Der Trend für einen Standort ergibt sich aus der Veränderung des von der LOESS-Methode geschätzten Werts an einem Mittwoch einer Woche gegenüber dem für den vorherigen Mittwoch vorhergesagten Wert, wobei die Werte vorher zurück auf die Originalskala transformiert wurden. Diese Trends des jeweiligen Standorts werden kategorisiert in „ansteigend“ (definiert als Anstieg um mehr als 15 % zur Vorwoche), „unverändert“ (Veränderung zwischen -15 % und 15 % zur Vorwoche) und „fallend“ (Rückgang um mehr als 15 % zur Vorwoche).

## Aggregation der Standortwerte

Es werden die einzelnen Zeitreihen der Standorte aggregiert, um einen bundesweiten Verlauf der SARS-CoV-2-Viruslast im Abwasser abzubilden. Dafür werden in jeder Woche, in der für mindestens 10 Standorte Messwerte vorliegen, der Mittelwert über die über eine Woche gemittelten logarithmierten Messwerte der einzelnen Standorte berechnet. Dabei wird nach den angeschlossenen Einwohnern der Kläranlage gewichtet.

## Hinweise zur Datenauswertung

Bei der Datenbewertung sind einige Besonderheiten zu beachten:

- Es wurden an den unterschiedlichen Standorten verschiedene Zielgene gemessen (eine Kombination aus vorzugsweise N1, N2, E, ORF oder RdRp).
- Der Standort Hamburg ist mit zwei Zuläufen vertreten: "Hamburg Nord" und "Hamburg Süd".
- Im Sommer 2023 lag die Viruslast an einzelnen Tagen / Standorten teilweise unter der Bestimmungsgrenze (BG). In diesen Fällen wurde  $0.5 \cdot BG$  als Wert eingetragen. Lag in einigen, seltenen Fällen keine Bestimmungsgrenze vor, wurde 4000 Genkopien/L als BG genommen.

## Limitationen

Abwasserdaten erlauben keinen Rückschluss auf die Krankheitsschwere oder die Belastung des Gesundheitssystems. Aus Abwasserdaten kann nach aktuellem Stand nicht präzise auf Inzidenz/Prävalenz oder die Untererfassung (die sog. „Dunkelziffer“) geschlossen werden. Für die epidemiologische Lagebewertung sollten die Daten immer in Zusammenschau mit anderen Indikatoren, z.B. aus der syndromischen Surveillance, betrachtet werden.

Absolute Viruslasten können insbesondere über längere Zeiträume nicht direkt im Hinblick auf die Anzahl an Infizierten verglichen werden, da sich die ausgeschiedene Virusmenge pro infizierter Person beispielsweise zwischen verschiedenen Virusvarianten unterscheiden kann. Die ermittelten Werte werden durch eine Vielzahl von Faktoren (z.B. Veränderungen der Abwasserzuleitung, Starkregenereignisse oder touristische Ereignisse) beeinflusst, was durch die Normalisierung nur teilweise ausgeglichen werden kann.

Der Zeitverzug von der Probenahme bis zur Übermittlung und weitere Veröffentlichung vom RKI kann bis zu zwei Wochen dauern.

## Inhalt und Aufbau des Datensatzes

---

Im AMELAG-Datensatz werden Daten und Kontextmaterialien zu SARS-CoV-2-Nachweisen im Abwasser bereitgestellt. Die im Projekt erhobenen Daten, liegen für [einzelne Standorte](#) und als [aggregierte Zeitreihe](#) vor.

Im Datensatz zusätzlich enthalten sind:

- Lizenz-Datei mit der Nutzungslizenz des Datensatzes in Deutsch und Englisch
- Datensatzdokumentation in deutscher Sprache
- Metadaten zur automatisierten Weiterverarbeitung
- Kontextmaterialien zur Datenanalyse

## Normalisierten Daten zur SARS-CoV-2-Viruslast

In der Datei [amelag\\_einzelstandorte.tsv](#) sind die normalisierten Daten zur SARS-CoV-2-Viruslast für die einzelnen Standorte angegeben.

| [amelag\\_einzelstandorte.tsv](#)

## Variablen und Variablenausprägungen

Die Datei `amelag_einzelstandorte.tsv` enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Variable	Typ	Ausprägung	Beschreibung
standort	Text		Standort, an dem sich die Kläranlage befindet.
bundesland	Text	BB , BE , BW , BY , HB , HE , HH , MV , NI , NW , RP , SH , SL , SN , ST , TH	Bundesland (abgekürzt), in dem sich die Kläranlage befindet.
datum	Datum	jjjj-mm-tt	Datum, an dem die 24-Stunden-Mischprobe in der Kläranlage begonnen hat.
viruslast	Gleitkommazahl	$\geq 0$	Gemessene SARS-CoV-2-Viruslast in Genkopien pro Liter.
loess_vorhersage	Gleitkommazahl	$\geq 0$	Die mittels einer LOESS-Regression (optimiert mittels GCV-Kriterium für die 10er-logarithmierten Viruslasten) vorhergesagten Viruslasten.
loess_obere_schranke	Gleitkommazahl	$\geq 0$	Obere Grenze des punktweisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
loess_untere_schranke	Gleitkommazahl	$\geq 0$	Untere Grenze des punktweisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.

loess_aenderung	Gleitkommazahl	$\geq 0$	Einwohner, die an das Klärwerk des Standortes angeschlossen sind.
einwohner	Natürliche Zahl	$\geq 0$	Einwohner, die an das Klärwerk des Standortes angeschlossen sind.
trend	Text	Ansteigend , Fallend , Unverändert , keine Daten vorhanden	Kategorisierte Veränderung des geglätteten LOESS-Wertes zur Vorwoche: "fallend" bedeutet, dass die geglättete Viruslast um mehr als 15% zur Vorwoche gesunken ist. "ansteigend" bedeutet, dass die geglättete Viruslast um mehr als 15% zur Vorwoche gestiegen ist. "gleichbleibend" bedeutet, dass die geglättete Viruslast sich nicht mehr als 15% zur Vorwoche verändert hat.

## Zeitreihe der SARS-CoV-2-Viruslast

In der Datei `amelag_aggregierte_kurve.tsv` ist die Zeitreihe der SARS-CoV-2-Viruslast auf aggregierter bzw. bundesweiter Ebene enthalten.

`amelag_aggregierte_kurve.tsv`

### Variablen und Variablenausprägungen

Die Datei `amelag_einzelstandorte.tsv` enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Variable	Typ	Ausprägung	Beschreibung
----------	-----	------------	--------------

datum	Datu,	jjjj-mm-tt	Datum des Mittwochs einer Woche.
n	Natürliche Zahl	$\geq 0$	Anzahl der Standorte, die mindestens einen Messwert im durch "datum" definierten Zeitraum übermittelt haben.
anteil_bev	Gleitkommazahl	$\geq 0$	Anteil der Gesamtbevölkerung in Deutschland, der an die übermittelnden Klärwerke angeschlossen ist.
viruslast	Gleitkommazahl	$\geq 0$	SARS-CoV-2-Viruslast in Genkopien pro Liter gemittelt über alle Standorte und gewichtet nach angeschlossenen Einwohnern der Kläranlagen. Vor der Mittelung über die Standorte wurden alle Messwerte der Standorte in den letzten 7 Tagen jeweils mittels 10er-Logarithmus transformiert und über die einzelnen Standorte gemittelt. Die angegebene Viruslast ist der auf die Originalskala zurücktransformierte Mittelwert.
loess_vorhersage	Gleitkommazahl	$\geq 0$	Die mittels einer LOESS-Regression vorhergesagten Viruslasten,



			zurücktransformiert auf die Originalskala.
loess_obere_schranke	Gleitkommazahl	$\geq 0$	Obere Grenze des punkweisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
loess_untere_schranke	Gleitkommazahl	$\geq 0$	Untere Grenze des punkweisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.

## Kontextmaterialien

Zur Reproduktion der Ergebnisse des [AMELAG Wochenberichts](#) werden die zur Erstellung der Analyse verwendeten R-Skripte zur bereitgestellt. Die Skripte befinden sich im Ordner "[Kontextmaterialien](#)" des Datensatzes.

### Kontextmaterialien

### Struktur der Skripte

Das R-Skript `main.R` erzeugt alle Grafiken, die im Wochenbericht angezeigt werden. Setzen Sie `show_log_data = FALSE` am Anfang von `main.R`, um Plots auf der Originalskala (statt auf der Logskala) zu erzeugen. Die Datei `main.R` ruft alle R-Skripte auf, die im Unterordner `Scripts` gespeichert sind und speichert alle Ergebnisse im Ordner `Results` und seinen Unterordnern. Die folgenden R-Skripte sind im Ordner `Scripts` verfügbar:

- `functions_packages.R`: Installiert (falls erforderlich) und lädt notwendige Pakete, definiert selbst geschriebene Funktionen und setzt Parameter und Variablen, die in anderen Skripten verwendet werden.
- `loess_calculation.R`: Löscht LOESS-Berechnungen, entsprechende Konfidenzintervalle und berechnete Trends aus `amelag_einzelstandorte.tsv` im Ordner `Data` und zeigt, wie man diese Größen berechnet.
- `aggregation_calculation.R`: Ausgehend von den Daten `amelag_einzelstandorte.tsv` im Ordner `Data` zeigt dieses Skript, wie man die Daten aggregiert und die LOESS-Kurve und ihre jeweiligen Konfidenzintervalle für die aggregierten Daten berechnet. Im Wesentlichen zeigt dieses Skript, wie man aus `amelag_einzelstandorte.tsv` den Datensatz `amelag_aggregierte_kurve.tsv` erhält.

- `plot_single_places.R`: Erzeugt eine Zeitreihengrafik mit einer LOESS-Kurve für jeden Standort, der genügend Daten geliefert hat. Speichert auch beobachtete und mittels LOESS geschätzte Abwasserdaten für jeden Standort, der genügend Daten geliefert hat. Für Standorte ohne ausreichende Daten werden keine LOESS-Schätzungen berechnet und gespeichert.
- `plot_aggregated_curve.R`: Erzeugt eine Zeitreihendarstellung mit einer LOESS-Kurve für die über alle Standorte aggregierten Daten.
- `plot_heatmap.R`: Erzeugt eine Heatmap, die Trends für alle Standorte zeigt, die genügend Daten geliefert haben.

## Ergebnisse

Nach dem Ausführen von `main.R` enthält der Ordner `Results` die Heatmap und die aggregierte Kurve in seinem Hauptverzeichnis und die Kurven und Daten für die einzelnen Standorte in seinem Unterordner `Single_Sites`.

## 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](https://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](#)

## Hinweise zur Nachnutzung der Daten

---

Offene Forschungsdaten des RKI werden auf [GitHub.com](https://github.com), [Zenodo.org](https://zenodo.org) und [Edoc.rki.de](https://edoc.rki.de) bereitgestellt:

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

## Lizenz

Der Datensatz " Abwassersurveillance AMELAG" 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 und des Umweltbundesamtes 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.