

# 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 (2024): Abwassersurveillance AMELAG, Berlin: Zenodo. DOI: [10.5281/zenodo.13144137](https://doi.org/10.5281/zenodo.13144137)

---

--- see English version below ---

## 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](#) - Einen 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.

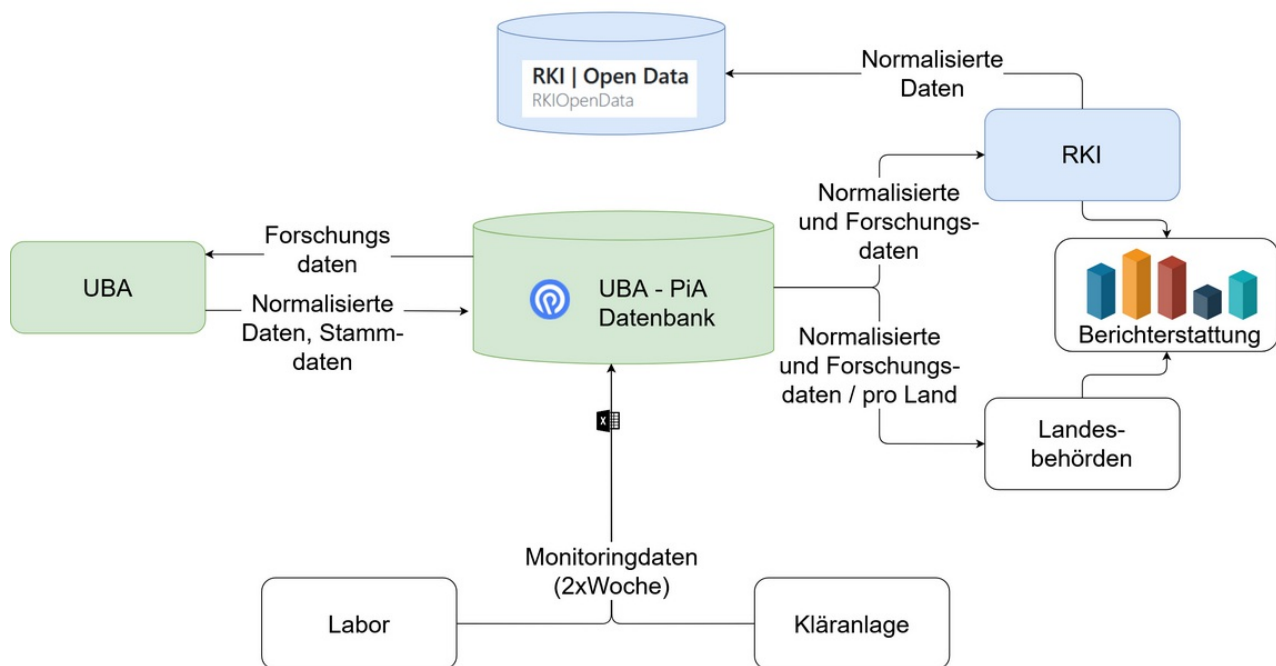
Die Verarbeitung, Aufbereitung und Veröffentlichung der Daten erfolgen durch das Fachgebiet MF 4 | Fach- und Forschungsdatenmanagement. Fragen zum Datenmanagement und zur Publikationsinfrastruktur können an das Open Data-Team des Fachgebiets MF4 unter [OpenData@rki.de](mailto:OpenData@rki.de) gerichtet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

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

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.

- **fallend** : die geglättete Viruslast ist um mehr als 15% zur Vorwoche gesunken
- **ansteigend** : die geglättete Viruslast ist um mehr als 15% zur Vorwoche gestiegen
- **gleichbleibend** : die geglättete Viruslast hat sich nicht mehr als 15% zur Vorwoche verändert
- **keine Daten vorhanden** : für den Mittwoch dieser oder der vergangenen Woche liegt kein geglätteter LOESS-Wert vor
- **NA** : ist für alle Tage außer Mittwoch eingetragen.

## 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 oder NA	Datum, an dem die 24-Stunden-Mischprobe in der Kläranlage begonnen hat.
viruslast	Gleitkommazahl	≥0 oder NA	Gemessene SARS-CoV-2-Viruslast in Genkopien pro Liter.
loess_vorhersage	Gleitkommazahl	≥0 oder NA	Die mittels einer LOESS-Regression (optimiert mittels GCV-Kriterium für die 10er-logarithmierten Viruslasten) vorhergesagten Viruslasten.
loess_obere_schranke	Gleitkommazahl	≥0 oder NA	Obere Grenze des punktwisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
loess_untere_schranke	Gleitkommazahl	≥0 oder NA	Untere Grenze des punktwisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
loess_aenderung	Gleitkommazahl	≥0 oder NA	Änderung des LOESS-Vorhersagewerts zur Vorwoche, d.h. Quotient aus aktuellem Wert und dem Vorwochenwert.
einwohner	Natürliche Zahl	≥0 oder NA	Einwohner, die an das Klärwerk des Standortes angeschlossen sind.
trend	Text	Ansteigend , Fallend , Unverändert , keine Daten vorhanden , NA	Kategorisierte Veränderung des geglätteten LOESS-Wertes von einem Mittwoch zum Mittwoch der Vorwoche (siehe <a href="#">Datenauswertung</a> )
laborwechsel	Text	ja , nein oder NA	Laborwechsel bzw. Änderungen in den Labormethoden.

## 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_aggregierte_kurve.tsv` enthält die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Variable	Typ	Ausprägung	Beschreibung
datum	Datum	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	$\geq 0$ oder NA	Anteil der Gesamtbevölkerung in Deutschland, der an die übermittelnden Klärwerke angeschlossen ist.
viruslast	Gleitkommazahl	$\geq 0$ oder NA	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	$\geq 0$ oder NA	Die mittels einer LOESS-Regression vorhergesagten Viruslasten, zurücktransformiert auf die Originalskala.
loess_obere_schranke	Gleitkommazahl	$\geq 0$	Obere Grenze des punktwisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
loess_untere_schranke	Gleitkommazahl	$\geq 0$	Untere Grenze des punktwisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.

## Kontextmaterialien

Zur Reproduktion der Ergebnisse des [AMELAG Wochenberichts](#) werden die zur Erstellung der Analyse verwendeten R-Skripte 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 Metadaten-datei, diese sind im Metadatenordner hinterlegt:



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.

In der zenodo.json ist neben der 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": "Collected",
    "description": "Date when the Dataset was created"
  }
],
```

## Hinweise zur Nachnutzung der Daten

---

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

## Wastewater surveillance AMELAG

---

[Robert Koch Institute](#) | RKI  
Nordufer 20  
13353 Berlin

### Cite

Fachgebiet 32, Robert Koch-Institut (2024): Abwassersurveillance AMELAG, Berlin: Zenodo. DOI: [10.5281/zenodo.13144137](https://doi.org/10.5281/zenodo.13144137)

---

## Information on the dataset and context of origin

---

The project "Wastewater monitoring for epidemiological situation assessment" (AMELAG) runs from 22.11.2022 to 31.12.2024. Local authorities, wastewater treatment plants (WWTP) and laboratories are working together to take, analyze and evaluate wastewater samples. The project aims at establishing SARS-CoV-2 detection in wastewater as an additional indicator for the epidemiological situation assessment at state and federal level. Further aims of the project include further development of structures and processes for a nationwide network for wastewater surveillance, to develop concepts for continuity and to research the possibilities for monitoring other pathogens in wastewater. Wastewater surveillance is a technique for detecting pathogens in wastewater to better control health protection measures. Wastewater data do not allow for an accurate assessment of disease severity or the burden on the healthcare system. In epidemiological assessments, the data should be combined with other indicators, e.g. from syndromic surveillance.

### Administrative and organizational information

AMELAG is a project funded by the [Federal Ministry of Health \(BMG\)](#) and is being conducted in cooperation with the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and [Consumer Protection \(BMUV\)](#).

The project is being carried out jointly by the Robert Koch Institute (RKI) and the [Federal Environment Agency \(UBA\)](#). Further information on AMELAG can be found on the [project website](#).

The participating WWTPs are responsible for taking samples, which are analyzed by the participating laboratories. In addition to commercial laboratories, state laboratories and the Federal Environment Agency, the Central Medical Service of the German Armed Forces also carries out part of the analysis.

Some of the WWTPs and laboratories are also involved in wastewater surveillance projects in the federal states (Baden-Württemberg, Bavaria, Berlin, Brandenburg, Hamburg, Hesse, Rhineland-Palatinate, Saxony-Anhalt).

Other WWTPs and laboratories are part of the following research projects:

- [WBEready](#) - A research consortium consisting of Emschergenossenschaft and LippeverbandEGLV, Research Institute for Water Management and Climate Future at RWTH Aachen FiW, University Hospital Frankfurt, Goethe University Frankfurt am Main, University Medicine Essen (Institute for Artificial Intelligence, Institute for Urban Public Health), RWTH Aachen, Institute for Urban Water Management.
- Establishment of methods for the detection of viruses in wastewater to assess the infection situation in the population (University of Dresden)
- Development of a state-wide wastewater surveillance system in Thuringia using mobility data and artificial intelligence (research consortium of the University of Weimar, University of Jena, University of Hamburg, Hamm-Lippstadt University of Applied Sciences, SMA Development GmbH, KOWUG Kommunale Wasser- und Umwelttechnik GmbH, Analytik Jena GmbH)
- Establishment of a multiplex PCR from wastewater and for detection and characterization of RSV in

the context of SARS-CoV-2 wastewater monitoring (AMELAG) (University of Bonn and Düsseldorf).

The company [ENDA](#) was commissioned with data management. The data collected are stored and processed in a database (PiA-Monitor).

The data are processed, edited and published by the Department MF 4 | Subject and Research Data Management. Questions about data management and the publication infrastructure can be directed to the Open Data team of the MF4 department at [OpenData@rki.de](mailto:OpenData@rki.de).

## Data collection

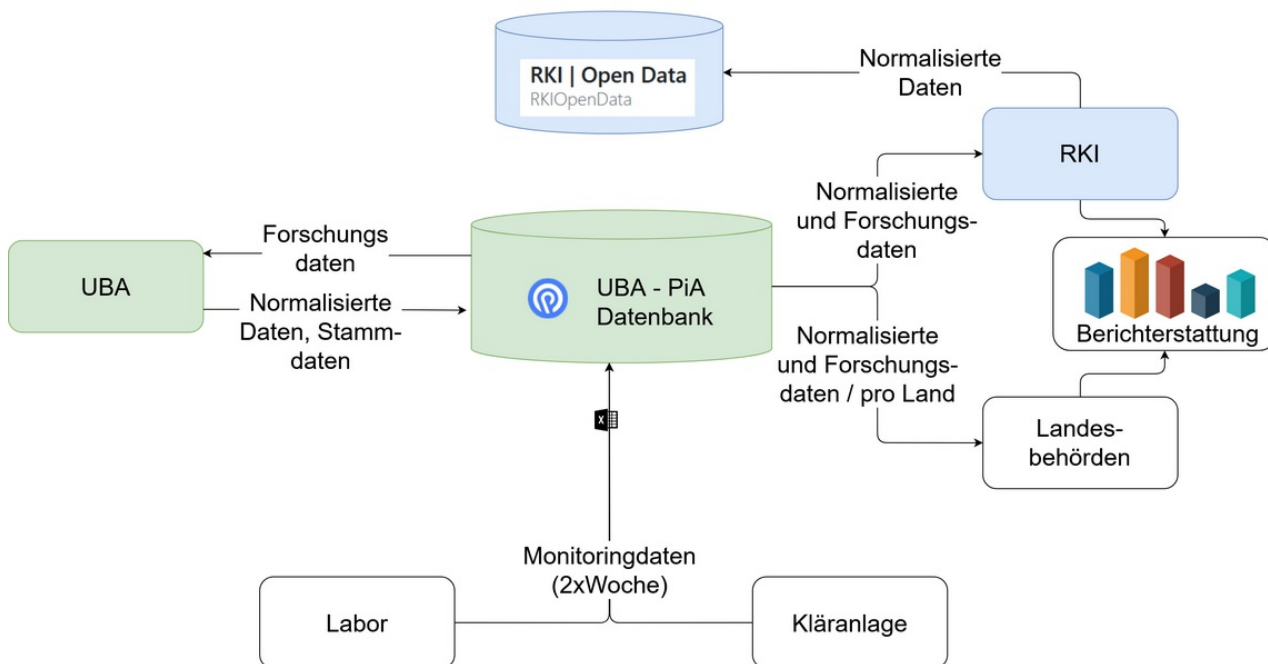
In AMELAG, [technical guidelines](#) were developed based on the handouts for sampling and laboratory analysis created as part of the [ESI-CorA project](#). The raw data of the samples analyzed in the ESI-CorA project are reused in AMELAG and included in the evaluated data.

Raw wastewater samples are generally collected twice a week at each participating WWTP, along with essential parameters such as volume flow, pH value, and temperature. These parameters are necessary for normalization and quality assurance. Where possible, the raw sewage samples should be taken after the grit chamber of the WWTP. A 24-hour composite sample is collected using an automatic sampler. The 24-hour samples are usually taken from Mondays to Tuesdays, and from Wednesdays to Thursdays. As a rule, one liter of the sample is filled into sample bottles and sent to the analysis laboratory.

In the laboratory, the viral nucleic acid is concentrated, extracted and the viral gene sequences are quantified by digital PCR (dPCR) or quantitative real-time PCR (qRT-PCR). At least two representative SARS-CoV-2 gene fragments (preferably N1, N2, E, ORF or RdRp) are determined.

Robert Koch Institute, Department 32 (2024): "ESI-CorA: SARS-CoV-2 wastewater surveillance" [Dataset]. Zenodo. DOI: [10.5281/zenodo.10781653](https://doi.org/10.5281/zenodo.10781653)

## Data flow



At the UBA, metadata on the WWTPs and the laboratories as well as the regularly collected monitoring data are centrally stored and processed further in a web application, the PiA-Monitor (Pathogens in Wastewater). The monitoring data to be collected regularly from the WWTP and the data of the laboratories are merged and imported into the database by the data providers via the web application. The UBA, the RKI and the federal states can access the data within the scope of their respective rights.

## Plausibility check and further processing of the data

A plausibility check is run on the data as they are imported. The formats, completeness of the information (mandatory fields), value ranges of the monitoring data, plausibility of the dates and compliance with stored metadata are checked. Only data records that successfully pass the quality check are imported into the database. The geometric mean of the viral load (gene copies/L) is then determined from the two or more measured target genes.

### Normalization procedure

A varying wastewater composition, e.g. due to irregular industrial influences or heavy rainfall events, can lead to changing concentrations of SARS-CoV-2. To take these external influences into account, the measured viral load can be normalized.

In AMELAG, normalization is performed according to flow rate. The dry weather inflow of the WWTP is the reference. The following formula was used:

$$Gene_{normalized} = Q_{KA\_current} / Q_{KA\_median} \cdot Gene_{averaged}$$

where:

- $Q_{KA\_aktuell}$  : Volume flow of the wastewater treatment plant in the sampling period and
- $Q_{KA\_median}$  : Median of the volume flow of the wastewater treatment plant

Normalization is automated with the data import.

### Data evaluation

The data are evaluated at the RKI using R scripts. The scripts are contained in the [context materials](#). A detailed description of the methodology is provided in the [technical guidelines](#). The results are published in the RKI's [weekly report](#).

For each site, the measured values are reported in gene copies per liter (gene copies/L). In addition, the measured values of the logarithmized normalized gene copies are smoothed using a locally weighted regression (LOESS) and associated confidence intervals are calculated. The trend for a location results from the change in the value estimated by the LOESS method on a Wednesday of a week compared to the value predicted for the previous Wednesday, whereby the values were previously transformed back to the original scale.

- **falling** : The smoothed viral load has fallen by more than 15% compared to the previous week
- **increasing** : The smoothed viral load has increased by more than 15% compared to the previous week
- **stable** : The smoothed viral load has not changed by more than 15% compared to the previous week
- **No data available** : No smoothed LOESS value is available for the Wednesday of this or the previous week
- **NA** : Is entered for all days except Wednesday.

### Aggregation of the location values

The individual time series of the locations are aggregated in order to depict a nationwide course of the SARS-CoV-2 viral load in wastewater. For each week in which measured values are available for at least 10 sites, the average of the logarithmized measured values of the individual sites averaged over one week is calculated. This value is then weighted by the number of inhabitants connected to the WWTPs.

## Notes on data evaluation

Some things to take into account when evaluating the data:

- Different target genes were measured at the different sites (a combination of preferably N1, N2, E, ORF or RdRp).
- The Hamburg site is represented by two inflows: "Hamburg North" and "Hamburg South".
- In summer 2023, the viral load on individual days / locations was sometimes below the limit of quantification (BG). In these cases,  $0.5 \cdot BG$  was entered as the value. If there was no limit of quantification in some rare cases, 4000 gene copies/L was taken as the BG.

## Limitations

Wastewater data do not allow conclusions to be drawn about disease severity or the burden on the healthcare system. As things stand at present, it is not possible to draw precise conclusions about incidence/prevalence or underreporting from wastewater data. When assessing a situation epidemiologically, the data should always be considered in conjunction with other indicators, such as those from syndromic surveillance.

Absolute viral loads cannot be compared directly to the number of infected persons, especially over longer periods of time, as, for example, the amount of virus excreted per infected person can differ between different virus variants.

The values determined are influenced by a variety of factors (e.g. changes in the wastewater supply, heavy rainfall events, or tourist events), which can only be partially compensated for by normalization. The time delay from sampling to transmission and further publication by the RKI can take up to two weeks.

## Content and structure of the dataset

---

The AMELAG dataset provides data and contextual material on SARS-CoV-2 detections in wastewater. The data collected in the project are available for [individual sites](#) and as [aggregated time series](#).

The dataset also contains:

- License file with the license to use the dataset in German and English
- Dataset documentation in German
- Metadata for automated further processing
- Context materials for data analysis

## Normalized data on SARS-CoV-2 viral load

In the file `amelag_einzelstandorte.tsv` contains the normalized SARS-CoV-2 viral load data for the individual sites.

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

## Variables and variable values

The file `amelag_einzelstandorte.tsv` contains the variables and their characteristics shown in the following table:

Variable	Type	Characteristic	Description
standort	Text		Location where the wastewater treatment plant is located.
bundesland	Text	BB , BE , BW , BY , HB , HE , HH , MV , NI , NW , RP , SH , SL , SN , ST , TH	Federal state (abbreviated) in which the wastewater treatment plant is located.
datum	Date	yyyy-mm-dd or NA	Date on which the 24-hour composite sample started in the wastewater treatment plant.
viruslast	Floating point number	$\geq 0$ or NA	Measured SARS-CoV-2 viral load in gene copies per liter.
loess_vorhersage	Floating point number	$\geq 0$ or NA	The viral loads predicted using a LOESS regression (optimized using GCV criterion for the 10s logarithmized viral loads).
loess_obere_schranke	Floating point number	$\geq 0$ or NA	Upper bound of the pointwise 95% confidence interval of the LOESS predicted value.
loess_untere_schranke	Floating point number	$\geq 0$ or NA	Lower bound of the point-wise 95% confidence interval of the LOESS predicted value.
loess_aenderung	Floating point number	$\mathbb{Z}$ or NA	Change in the LOESS predicted value compared to the previous week, i.e. quotient of the current value and the previous week's value.
einwohner	Natural number	$\geq 0$ or NA	Inhabitants connected to the site's sewage treatment plant.
trend	Text	increasing , decreasing , unchanged , no data available , NA	Categorized change in the smoothed LOESS value from a Wednesday to the Wednesday of the previous week (see <a href="#">data evaluation</a> )
laborwechsel	Text	ja , nein or NA	Changes in laboratory or changes in the methods.

## Time series of the SARS-CoV-2 viral load

In the file `amelag_aggregated_curve.tsv` contains the time series of the SARS-CoV-2 viral load on an aggregated or nationwide level.

| [amelag\\_aggregierte\\_kurve.tsv](#)

## Variables and variable characteristics

The file `amelag_aggregierte_kurve.tsv` contains the variables and their values shown in the following table:

Variable	Type	Characteristic	Description
datum	Date	yyyy-mm-dd	date of the Wednesday of a week
n	Natural number	$\geq 0$	Number of locations that have transmitted at least one measured value in the period defined by "date".
anteil_bev	Floating point number	$\geq 0$ or NA	Proportion of the total population in Germany that is connected to the transmitting sewage treatment plants.
viruslast	Floating point number	$\geq 0$ or NA	SARS-CoV-2 viral load in gene copies per liter averaged over all sites and weighted by connected inhabitants of the wastewater treatment plants. Before averaging across the sites, all measured values of the sites in the last 7 days were transformed using the logarithm of 10 and averaged across the individual sites. The indicated viral load is the mean value transformed back to the original scale.
loess_vorhersage	Floating point number	$\geq 0$ or NA	The viral loads predicted using a LOESS regression, transformed back to the original scale.
loess_obere_schranke	Floating point number	$\geq 0$	Upper bound of the 95% confidence interval of the LOESS predicted value.
loess_untere_schranke	Floating point number	$\geq 0$	Lower bound of the 95% confidence interval of the LOESS predicted value.

## Context materials

To reproduce the results of the [AMELAG weekly report](#), the R scripts used to create the analysis are provided. The scripts can be found in the "Contextual materials" folder of the dataset.

### Context Materials

## Structure of the scripts

The R script `main.R` generates all graphics that are displayed in the weekly report. Set `show_log_data = FALSE` at the beginning of `main.R` to generate plots on the original scale (instead of the log scale). The file `main.R` calls all R scripts stored in the subfolder `Scripts` and stores all results in the folder `Results` and its subfolders. The following R scripts are available in the `Scripts` folder:

- `functions_packages.R` : Installs (if required) and loads necessary packages, defines self-written functions and sets parameters and variables used in other scripts.
- `loess_calculation.R` : Deletes LOESS calculations, corresponding confidence intervals and calculated trends from `amelag_einzelstandorte.tsv` in the `Data` folder and shows how to calculate these quantities.
- `aggregation_calculation.R` : Starting with the data `amelag_einzelstandorte.tsv` in the `Data` folder, this script shows how to aggregate the data and calculate the LOESS curve and its respective confidence intervals for the aggregated data. Essentially, this script shows how to obtain the dataset `amelag_aggregated_curve.tsv` from `amelag_single_location.tsv`.
- `plot_single_places.R` : Generates a time series graph with a LOESS curve for each location that has provided sufficient data. Also stores observed and LOESS-estimated wastewater data for each location that provided sufficient data. LOESS estimates are not calculated and stored for sites without sufficient data.
- `plot_aggregated_curve.R` : Generates a time series plot with a LOESS curve for the data aggregated across all sites.
- `plot_heatmap.R` : Creates a heatmap showing trends for all locations that have provided sufficient data.

## Results

After running `main.R` the folder `Results` contains the heatmap and the aggregated curve in its root directory and the curves and data for the individual sites in its subfolder `Single_Sites`.

## 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](https://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": "Collected",
    "description": "Date when the Dataset was created"
  }
],
```

## Guidelines for Reuse of the Data

---

Open data from the RKI are available on [Zenodo.org](https://zenodo.org), [GitHub.com](https://github.com), [OpenCoDE](https://gitlab.opencode.de), and [Edoc.rki.de](https://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 "Abwassersurveillance AMELAG" dataset is licensed under the [Creative Commons Attribution 4.0 International Public License | CC-BY](https://creativecommons.org/licenses/by/4.0/).

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.