

# Abwassersurveillance AMELAG

---

**Robert Koch-Institut, & Umweltbundesamt**

## Beitragende

Fachgebiet 32<sup>1</sup>

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

## Zitieren

Robert Koch-Institut, & Umweltbundesamt. (2025). Abwassersurveillance AMELAG [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.17045957>

---

--- please find the English version here ---

## Informationen zum Datensatz und Entstehungskontext

---

Das Vorhaben „Abwassermonitoring für die epidemiologische Lagebewertung“ (AMELAG) läuft vom 22.11.2022 bis zum 31.12.2025. 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. Aktuell werden Abwasserproben von ausgewählten Kläranlagen auf SARS-CoV-2, Influenzaviren und Respiratorische Synzytial-Viren (RSV) untersucht.

Bei der Abwassersurveillance werden Erreger im Abwasser gemessen um Gesundheitsschutzmaßnahmen besser steuern zu können. Abwassersurveillance kann einen Beitrag für eine Reihe von [Anwendungsfällen](#) liefern. Abwasserdaten unterliegen speziellen Limitationen, beispielsweise erlauben sie keine genaue Einschätzung von Krankheitsschwere oder 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.

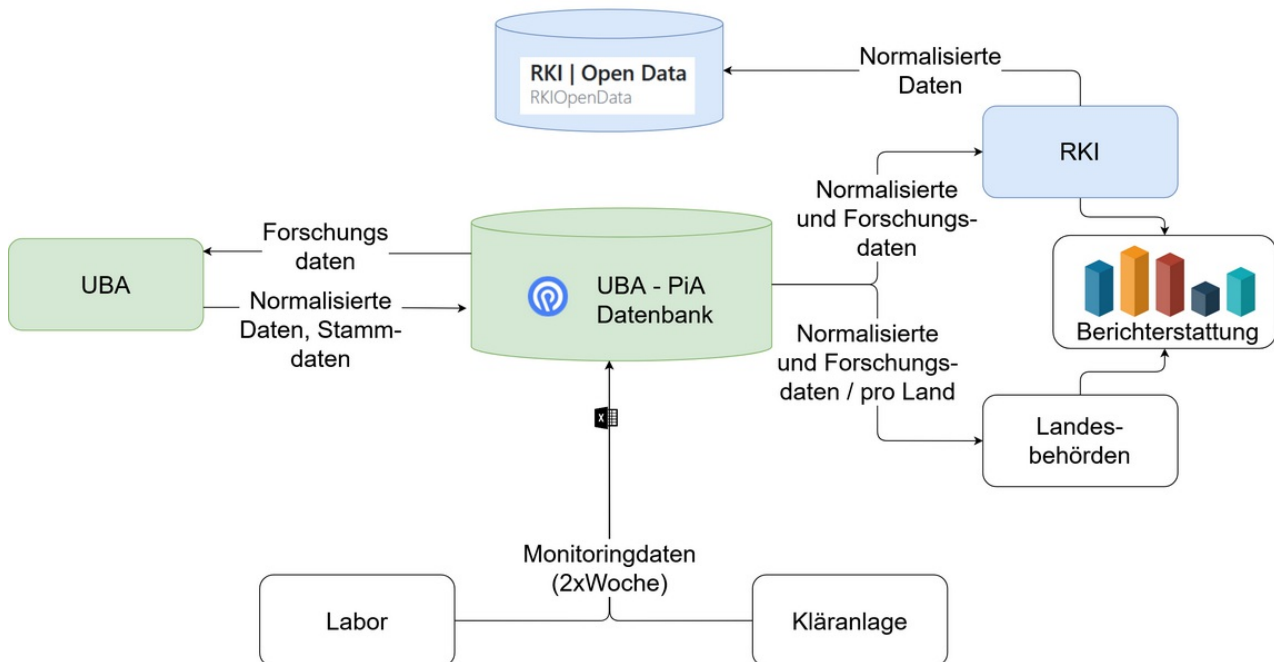
Die Verarbeitung, Aufbereitung und Veröffentlichung der Daten erfolgen durch das Fachgebiet MF 4 | Fach- und Forschungsdatenmanagement. Fragen zum Datenmanagement und zur Publikationsinfrastruktur können an das Open Data-Team des Fachgebiets MF4 unter [OpenData@rki.de](mailto:OpenData@rki.de) gerichtet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

## Datenerhebung

In AMELAG wurden aufbauend auf die im Rahmen des [ESI-CorA-Projekts](#) erstellten Handreichungen zur Probennahme und Laboranalytik [technische Leitfäden](#) entwickelt. Die SARS-CoV-2-Rohdaten der im ESI-CorA-Projekt analysierten Abwasserproben wurden in AMELAG nachgenutzt und sind 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). Bei SARS-CoV-2 werden mindestens zwei Genfragmente (vorzugsweise N1, N2, E, ORF oder RdRp) bestimmt, bei den Influenzaviren ein Genfragment (vorzugsweise M1 für Influenza-A-Virus und M1, NS1, NS2 oder HA für Influenza-B-Virus) und bei RSV ebenfalls ein Genfragment (vorzugsweise N für RSV A und RSV B, M oder N für das gemeinsam gemessene RSV A/B).

## Datenfluss



Beim UBA laufen die Metadaten zu den Kläranlagen 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 UBA, das RKI und die Landesbehörden 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. Für SARS-CoV-2 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 der Viren führen. Um diese externen Einflüsse zu berücksichtigen, kann die gemessene Viruslast normalisiert werden.

In AMELAG ist die auf den Durchfluss der Kläranlage normalisierte Viruslast unter der Variable `viruslast_normalisiert` verfügbar. 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. Die gemessenen Viruslasten im Wochenbericht sowie in der Variable `viruslast` sind nicht normalisiert, da sich keine verbesserte Datenqualität durch die Normalisierung feststellen lässt.

## 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 [technischen Leitfäden](#) hinterlegt. Die Ergebnisse werden wöchentlich im AMELAG [Wochenbericht](#) auf der RKI-Webseite veröffentlicht.

Für jeden Standort werden die nicht normalisierten Messwerte (optional können auch die normalisierten Messwerte verwendet werden) für SARS-CoV-2, Influenza A und B-Virus sowie RSV A, RSV B und das gemeinsam gemessene RSV A/B in Genkopien pro Liter (Genkopien/L) angegeben. Zusätzlich werden die Messwerte der logarithmierten Genkopien mittels einer lokal gewichteten Regression (LOESS) geglättet und zugehörige Konfidenzintervalle berechnet.

## Aggregation der Standortwerte

Es werden die einzelnen Zeitreihen der Standorte aggregiert, um einen bundesweiten Verlauf der jeweiligen Viruslast im Abwasser abzubilden. Dafür wird zunächst der Mittelwert über die über eine Woche gemittelten Messwerte der einzelnen Standorte berechnet. Dann wird für jeden Standort für jede Woche die Differenz vom Wochenmittelwert über alle Standorte dieser Woche berechnet. Für jede Standort-Labor-Kombination wird der Mittelwert über diese Differenzen über alle Wochen gebildet um diesen Mittelwert danach von den ursprünglich gemessenen Werten abzuziehen. Dadurch wird für mittlere Unterschiede in den Viruslasten zwischen unterschiedlichen Standort-Labor-Kombinationen, adjustiert. Abschließend wird in jeder Woche, in der für mindestens 20 Standorte Messwerte vorliegen, der Mittelwert über diese adjustierten Werte berechnet. Dabei wird nach den angeschlossenen Einwohnern der Kläranlage gewichtet.

Da mit verschiedenen Kläranlagen und Laboren viele Akteure in die Datenerhebung involviert sind, kann es zu Unregelmäßigkeiten in den Daten der Einzelstandorte kommen, die einen großen Einfluss auf die über alle Standorte aggregierten Werte haben können. Sobald solche Auffälligkeiten festgestellt werden, werden diese Werte bis zur vollständigen Abklärung der Ursachen aus den aggregierten Kurven (`amelag_aggregierte_kurve.tsv`) ausgeschlossen. In den Daten der Einzelstandorte (`amelag_aggregierte_kurve.tsv`) bleiben die Werte weiterhin enthalten.

## Hinweise zur Datenauswertung

Bei der Datenbewertung sind einige Besonderheiten zu beachten:

- Es wurden an den unterschiedlichen Kläranlagen und für die unterschiedlichen Viren verschiedene Zielgene gemessen
  - SARS-CoV-2: eine Kombination aus vorzugsweise N1, N2, E, ORF oder RdRp
  - Influenza A-Virus: vorzugsweise M1
  - Influenza B-Virus: vorzugsweise M1, NS1, NS2, HA
  - RSV A: vorzugsweise N
  - RSV B: vorzugsweise N
  - RSV A/B: M, N
- Einige Städte sind mit mehr als einer Kläranlage bzw. mehr als einem Zulauf vertreten.
- Bei Werten unterhalb der Bestimmungsgrenze (BG) wird mit der Hälfte der Bestimmungsgrenze als

Wert gerechnet ( $0,5 * BG$ ).

## 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 Normalisierungsmethoden bisher nicht ausgeglichen werden kann.

Von der Probenahme bis zur Datenübermittlung und Veröffentlichung vom RKI kann es zu einem Zeitverzug von bis zu zwei Wochen kommen.

## 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

## Daten für die einzelnen Standorte

In der Datei [amelag\\_einzelstandorte.tsv](#) sind die normalisierten und nicht normalisierten Daten zur Viruslast für SARS-CoV-2, Influenza A und B (einzeln und als Summe) sowie zu RSV A, RSV B, deren Summe (RSV A+B) und gemeinsamen Messungen von RSV A und B (RSV A/B) 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. Ein maschinenlesbares Datenschema ist im [Data Package Standard](#) in [tableschema\\_amelag\\_einzelstandorte.json](#) hinterlegt:

[tableschema\\_amelag\\_einzelstandorte.json](#)

Variable	Typ	Ausprägungen	Beschreibung
standort	string	Beispiele: Aachen , Ratzeburg , Weil am Rhein	Standort der Kläranlage.

bundesland	string	Werte: BB, BE, BW, BY, HB, HE, HH, ...	Kürzel des Bundeslandes, in dem sich die Kläranlage befindet.
datum	date	Format: YYYY-MM-DD Fehlende Werte: NA	Datum, an dem die 24-Stunden-Mischprobenahme in der Kläranlage begonnen hat.
viruslast	number	Werte: $\geq 0$ Fehlende Werte: NA	Gemessene Viruslast (in Genkopien pro Liter).
viruslast_normalisiert	number	Werte: $\geq 0$ Fehlende Werte: NA	Nach dem Durchfluss normalisierte Viruslast (wie in der Variable "viruslast" beschrieben).
loess_vorhersage	number	Werte: $\geq 0$ Fehlende Werte: NA	Vorhergesagte Viruslast (mittels einer LOESS-Regression auf Basis der nicht-normalisierten Viruslasten, optimiert mittels AICc-Kriterium für die 10er-logarithmierten Viruslasten).
loess_obere_schranke	number	Werte: $\geq 0$ Fehlende Werte: NA	Obere Grenze des punktwisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
loess_untere_schranke	number	Werte: $\geq 0$ Fehlende Werte: NA	Untere Grenze des punktwisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
einwohner	integer	Werte: $\geq 0$ Fehlende Werte: NA	Anzahl an Einwohnern, die an die Kläranlage angeschlossen sind.
laborwechsel	string	Werte: ja, nein Fehlende Werte: NA	Gibt an, ob es Laborwechsel bzw. Änderungen in den Labormethoden gab.
typ	string	Werte: SARS-CoV-2, Influenza A, Influenza B, Influenza A+B, RSV A, RSV B, RSV A+B, ...	Virustyp.
unter_bg	string	Werte: ja, nein Fehlende Werte: NA	Gibt an, ob mindestens die Hälfte der gemessenen Gene unterhalb der Bestimmungsgrenze liegen.

## Daten über alle Standorte aggregiert

In der Datei [amelag\\_aggregierte\\_kurve.tsv](#) ist die Zeitreihe der SARS-CoV-2-, Influenza A- und Influenza B-Viruslast einzeln und als Summe (Influenza A+B) sowie von RSV A, RSV B einzeln und als Summe (RSV A+B) und der gemeinsamen Messungen von RSV A und RSV B (RSV A/B) 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. Ein maschinenlesbares Datenschema ist im [Data Package Standard](#) in [tableschema\\_amelag\\_aggregierte\\_kurve.json](#) hinterlegt:

[tableschema\\_amelag\\_aggregierte\\_kurve.json](#)

Variable	Typ	Ausprägungen	Beschreibung
datum	date	Format: YYYY-MM-DD	Datum des Mittwochs einer Woche. Die Daten der unterliegenden Einzelzeitreihen werden innerhalb des Zeitraums vom vorherigen Donnerstag bis zum angegebenen Mittwoch gemittelt.
n	integer	Werte: $\geq 0$ Fehlende Werte: NA	Anzahl der Standorte, die mindestens einen Messwert im durch "datum" beschriebenen Zeitraum übermittelt haben.

anteil_bev	number	Werte: $\geq 0$ Fehlende Werte: NA	Anteil der Gesamtbevölkerung in Deutschland, der an die übermittelnden Kläranlagen angeschlossen ist.
viruslast	number	Werte: $\geq 0$ Fehlende Werte: NA	Gemessene 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.
viruslast_normalisiert	number	Werte: $\geq 0$ Fehlende Werte: NA	Nach dem Durchfluss normalisierte Viruslast (wie in der Variable "viruslast" beschrieben).
loess_vorhersage	number	Werte: $\geq 0$ Fehlende Werte: NA	Vorhergesagte Viruslast (mittels einer LOESS-Regression auf Basis der nicht-normalisierten Viruslasten, zurücktransformiert auf die Originalskala).
loess_obere_schranke	number	Werte: $\geq 0$	Obere Grenze des punkweisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
loess_untere_schranke	number	Werte: $\geq 0$	Untere Grenze des punkweisen 95%-Konfidenzintervalls des LOESS-Vorhersagewerts.
typ	string	Werte: SARS-CoV-2 , Influenza A , Influenza B , Influenza A+B , RSV A , RSV B , RSV A+B , ...	Virustyp.

## 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](#)

## 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": "Collected",  
    "description": "Date when the Dataset 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 "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.



---

Documentation

## Wastewater Surveillance AMELAG

---

**Robert Koch Institute**, & **Federal Environment Agency**

### Contributors

Unit 32<sup>1</sup>

<sup>1</sup> Robert Koch Institute

### Cite

Robert Koch Institute, & Federal Environment Agency. (2025). Wastewater Surveillance AMELAG [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.17045957>

---

## Information on the dataset and context of origin

---

In AMELAG ("Abwassermonitoring für die epidemiologische Lagebewertung", German for wastewater monitoring for epidemiological situation assessment), running from 22.11.2022 to 31.12.2025, local authorities, wastewater treatment plants (WWTP) and laboratories are working together to take, analyze and evaluate wastewater samples. The project aims at testing wastewater samples for selected pathogens and to establish it 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 wastewater surveillance network, to develop concepts for continuity and to research the possibilities for monitoring other pathogens in wastewater. Currently, wastewater samples from selected treatment plants are being tested for SARS-CoV-2, influenza viruses and eespiratory syncytial viruses (RSV).

Wastewater surveillance is a technique for detecting pathogens in wastewater to better control health protection measures. Wastewater surveillance has a range of [applications](#). Wastewater data, however,

underlie several limitations. For example, they 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 SARS-CoV-2 samples analyzed in the ESI-CorA project are reused in AMELAG and included in the evaluated data.

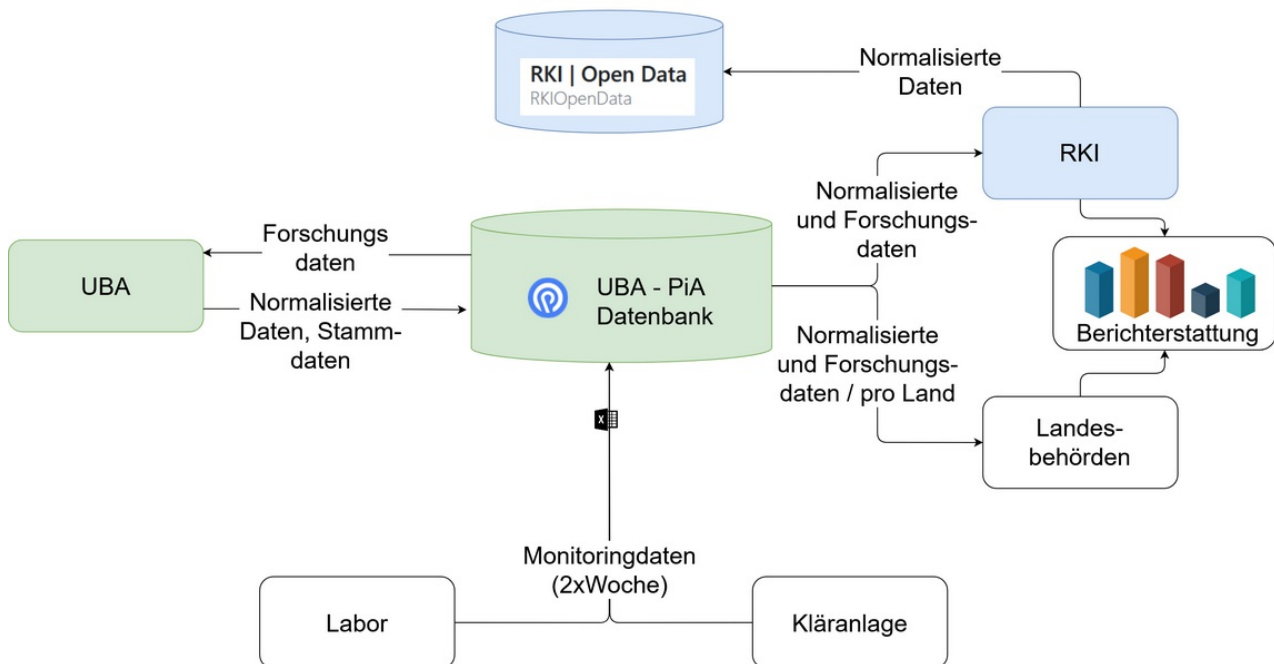
Raw wastewater samples are generally collected twice a week at each participating WWTP, along with essential parameters such as volume flow, pH value, and temperature. These parameters are necessary for normalization and quality assurance. Where possible, the raw sewage samples should be taken after 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). For SARS-CoV-2, at least two

representative gene fragments (preferably N1, N2, E, ORF or RdRp) are determined, for the Influenza virus only one gene fragment (preferably M1 for Influenza A Virus and M1, NS1, NS2 or HA for Influenza B Virus), for RSV also only one gene fragment (preferably N for RSV A and RSV B, M or N for RSV A/B).

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. For SARS-CoV-2, 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 the pathogens. To take these external influences into account, the measured viral load can be normalized.

In AMELAG, the viral load normalized to the flow rate of the wastewater treatment plant is available as `viruslast_normalisiert`. 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. The measured virus loads provided in the weekly report and in the variable `virusload` are not normalized, as normalization does not improve data quality.

## 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 WWTP, the non-normalized measured values (optional: the normalized values) for SARS-CoV-2, Influenza A and B viruses and RSV A, RSV B and RSV A/B are reported in gene copies per liter (gene copies/L). In addition, the measured values of the logarithmized gene copies are smoothed using a locally weighted regression (LOESS) and associated confidence intervals are calculated.

## Aggregation of the WWTP values

The individual time series of the WWTPs are aggregated in order to depict a nationwide course for each pathogen. To do this, the mean value is first calculated over the logarithmized measured values of the individual WWTPs averaged over one week. Then, for each location and for each week, the deviation from the weekly mean value over all locations is calculated. For each WWTP-laboratory combination, the mean of these differences over all weeks is determined and then subtracted from the originally measured values. This procedure adjusts for mean differences in viral loads between different WWTP-laboratory combinations. Finally, the weighted mean (weighted according to the number of inhabitants connected to the respective WWTP) of these adjusted values is calculated for each week in which measured values are available for at least 20 locations.

As different WWTPs and laboratories are involved in the data collection, inconsistencies in the data from individual sites may arise, potentially having a significant impact on the values aggregated across all sites. As soon as such inconsistencies are detected, these values are excluded from the aggregated curves (`amelag_aggregierte_kurve.tsv`) until the underlying causes have been fully clarified. The values remain in the data for the individual sites (`amelag_aggregierte_kurve.tsv`).

## Notes on data evaluation

Some things to take into account when evaluating the data:

- Different target genes were measured at the different sites
  - SARS-CoV-2: a combination of preferably N1, N2, E, ORF oder RdRp
  - Influenza A-Virus: preferably M1
  - Influenza B-Virus: preferably M1, NS1, NS2, HA
  - RSV A: preferably N
  - RSV B: preferably N
  - RSV A/B: M, N
- Some cities have more than one sewage treatment plant or more than one inflow.
- For values below the limit of quantification (LOQ), half of the LOQ is used as the value ( $0.5 * LOQ$ ).

## Limitations

Wastewater data do not allow conclusions to be drawn about disease severity or the burden on the healthcare system. 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 combination 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 determined values are influenced by a variety of factors (e.g., changes in the wastewater inflow, heavy rainfall events, or tourist events), which cannot be compensated by normalization methods thus far.

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

### Data for individual WWTP

The file [amelag\\_einzelstandorte.tsv](#) contains the normalized and non-normalized data on viral load of SARS-CoV-2, influenza and RSV data for the individual sites.

[amelag\\_einzelstandorte.tsv](#)

### Variables and variable values

The file [amelag\\_einzelstandorte.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\\_amelag\\_einzelstandorte.en.json](#):

[tableschema\\_amelag\\_einzelstandorte.en.json](#)

Variable	Type	Characteristic	Description
standort	string	Examples: Aachen, Ratzeburg, Weil am Rhein	Location of the wastewater treatment.
bundesland	string	Values: BB, BE, BW, BY, HB, HE, HH, ...	Federal state (abbreviated) in which the wastewater treatment plant is located.
datum	date	Format: YYYY-MM-DD Missing values: NA	Date on which the 24-hour composite sample started in the wastewater treatment plant.
viruslast	number	Values: $\geq 0$ Missing values: NA	Measured viral load in gene copies per liter.

viruslast_normalisiert	number	Values: $\geq 0$ Missing values: NA	Flow-normalized viral load (as described in the variable "viruslast").
loess_vorhersage	number	Values: $\geq 0$ Missing values: NA	Predicted viral load (predicted using a LOESS regression and the non-normalized viral loads, transformed back to the original scale).
loess_obere_schranke	number	Values: $\geq 0$ Missing values: NA	Upper bound of the pointwise 95% confidence interval of the LOESS predicted value.
loess_untere_schranke	number	Values: $\geq 0$ Missing values: NA	Lower bound of the pointwise 95% confidence interval of the LOESS predicted value.
einwohner	integer	Values: $\geq 0$ Missing values: NA	Number of inhabitants connected to the site's sewage treatment plant.
laborwechsel	string	Values: ja, nein Missing values: NA	Indicates whether change in laboratory or change in laboratory methods occurred.
typ	string	Values: SARS-CoV-2, Influenza A, Influenza B, Influenza A+B, RSV A, RSV B, RSV A+B, ...	Virus type.
unter_bg	string	Values: ja, nein Missing values: NA	Indicates if at least half of the measured genes are under the limit of quantification ( ja =yes, nein =no).

## Data aggregated across all WWTP

In the file [amelag\\_aggregated\\_curve.tsv](#) contains the time series of the SARS-CoV-2, influenza virus and RSV loads 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. A machine-readable data schema is stored in [Data Package Format](#) in [tableschemamamelag\\_aggregierte\\_kurve.en.json](#):

[tableschemamamelag\\_aggregierte\\_kurve.en.json](#)

Variable	Type	Characteristic	Description
datum	date	Format: YYYY-MM-DD	Date of Wednesdays of a week. The data of the underlying individual time series are averaged within the period from the previous Thursday to the specified Wednesday.
n	integer	Values: $\geq 0$ Missing values: NA	Number of locations that have transmitted at least one measured value in the period defined by "date".
anteil_be	number	Values: $\geq 0$ Missing values: NA	Proportion of the total population in Germany that is connected to the transmitting sewage treatment plants.
viruslast	number	Values: $\geq 0$ Missing values: NA	Measured 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.

viruslast_normalisiert	number	Values: $\geq 0$ Missing values: NA	Flow-normalized viral load (as described in the variable "viruslast").
loess_vorhersage	number	Values: $\geq 0$ Missing values: NA	Predicted viral load (predicted using a LOESS regression and the non-normalized viral loads, transformed back to the original scale).
loess_obere_schranke	number	Values: $\geq 0$	Upper bound of the 95% confidence interval of the LOESS predicted value.
loess_untere_schranke	number	Values: $\geq 0$	Lower bound of the 95% confidence interval of the LOESS predicted value.
typ	string	Values: SARS-CoV-2 , Influenza A , Influenza B , Influenza A+B , RSV A , RSV B , RSV A+B , ...	Virus type.

## 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

## 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": "Collected",
    "description": "Date when the dataset 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 "Wastewater Surveillance AMELAG" 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.