

Documentación del conjunto de datos  
# Infecciones por SARS-CoV-2 en Alemania

Instituto Robert Koch | RKI  
Nordufer 20  
13353 Berlín

FG 32 | Vigilancia y sistema electrónico de notificación e información  
(DEMIS) | Punto de contacto de la ÖGD  
Michaela Diercke (Gestión)

IMF | Desarrollo de métodos, infraestructura de investigación y tecnología de  
la información  
Linus Grabenhenrich (Gestión)

IT 4 | Arquitectura y desarrollo de software  
Herrmann Claus (Director)

MF 4 | Información y gestión de datos de investigación  
Hannes Wuensche (Curación de datos)

---

#### **Cita**

Instituto Robert Koch (2023): Infecciones por SARS-CoV-2 en Alemania, Berlín:  
Zenodo. DOI:10.5281/zenodo.4681153.

### **Información sobre el conjunto de datos y el contexto de origen**

Este conjunto de datos contiene información exhaustiva sobre las infecciones por SRAS-CoV-2 en Alemania, que fueron notificadas al Instituto Robert Koch (RKI) por las autoridades sanitarias de conformidad con la Ley de Protección de las Infecciones (IfSG). Los datos incluyen información sobre el número de casos confirmados, fallecimientos y recuperaciones, a partir de los cuales pueden obtenerse otras cifras clave relacionadas con la pandemia de COVID-19. El conjunto de datos se actualiza diariamente y contiene información detallada a nivel de condado, desglosada por diferentes grupos de edad. El objetivo del conjunto de datos es ayudar a comprender mejor la pandemia de COVID-19 en Alemania y apoyar la elaboración de informes, la investigación y el análisis en este ámbito.

#### **Información administrativa y organizativa**

El conjunto de datos "Infecciones por SRAS-CoV-2 en Alemania" proporciona el número de casos diarios de infecciones positivas por SRAS-CoV-2, muertes y recuperaciones notificadas por las autoridades sanitarias de Alemania de acuerdo con los requisitos de la Ley de Protección de la Infección (IfSG).

Los datos subyacentes se transmiten al Instituto Robert Koch (RKI) a través del sistema de notificación de conformidad con la IfSG. Responsable del

funcionamiento del

El punto de contacto del RKI para el sistema de notificación es el Departamento 32 | Vigilancia y Sistema Electrónico de Notificación e Información (DEMIS) | ÖGD.

El procesamiento y la preparación de los datos brutos disponibles en el sistema de información corren a cargo del departamento de Arquitectura y Desarrollo de Software del RKI.

La publicación de los datos, su conservación y la gestión de la calidad de los (meta)datos corren a cargo del Departamento MF 4 | Gestión de la Información y los Datos de Investigación. Las preguntas sobre la gestión de datos y la infraestructura de publicación pueden dirigirse al Equipo de Datos Abiertos del departamento MF4 en [OpenData@rki.de](mailto:OpenData@rki.de).

### **Contenido y estructura del registro de datos**

El conjunto de datos contiene datos epidemiológicos sobre la evolución de las infecciones por SRAS-CoV-2 en Alemania. El conjunto de datos contiene

- Datos sobre el número de casos con informes actualizados diariamente de infecciones por SRAS-CoV-2
- Archivo con la recopilación de todas las tablas de números de casos anteriores
- Fichero con la licencia de uso del conjunto de datos
- Documentación del conjunto de datos en alemán
- Archivo de metadatos para la importación en Zenodo

### **Datos y preparación de datos**

Los datos sobre el número de casos representan un estado actualizado diariamente (00:00) de todos los casos de infección notificados previamente en Alemania. Esto significa que todas las infecciones por SRAS-CoV-2 transmitidas al sistema de notificación del RKI por las autoridades sanitarias a través de las autoridades estatales responsables antes de las 00:00 horas del día AAAA-MM-DD se incluyen en el estado de los datos. Los datos se generan completamente nuevos cada día y este estado de datos sustituye al estado de datos del día anterior.

Los datos del número de caso contienen el ID del distrito como única geoinformación. Se basa en la clave oficial del municipio (AGS) del trimestre 2 de 2020, obtenida del portal de la Oficina Federal de Estadística. La ID del distrito se deriva del código del estado federal (Land), el distrito regional (RB) y el distrito administrativo (LK). Para una representación más precisa de Berlín, los 12 distritos de la ciudad se desglosan como "Landkreise" separados. Esto se desvía de las especificaciones del AGS. La distribución es la siguiente:

IdLandkreis	Distrito	IdLandkreis	Distrito
11001	Centro de Berlín	11007	Berlín Tempelhof- Schöneberg
11002	Berlín Friedrichshain- Kreuzberg	11008	Berlín Neukölln
11003	Berlín Pankow	11009	Berlín Treptow- Köpenick
11004	Berlín Charlottenburg- Wilmerdorf	11010	Berlín Marzahn- Hellersdorf
11005	Berlín Spandau	11011	Berlín Lichtenberg
11006	Berlín Steglitz- Zehlendorf	11012	Berlín Reinick- endorf

### Datos del número de caso

Archivo/YYYY-MM-DD\_Germany\_SARSCoV2\_Infections.csv

La fecha central del conjunto de datos es el número de caso actual. La carpeta de archivo contiene los datos del número de caso con el nombre de archivo "AAAA-MM-DD\_Alemania\_SARSCoV2\_Infecciones.csv". En el nombre del archivo, la secuencia "AAAA-MM-DD" representa la fecha de creación del archivo y, por tanto, también la fecha del estado de los datos que contiene. "AAAA" representa el año, "MM" el mes y "DD" el día de creación o del estado de datos que contiene.

**Características de los datos del número de caso** En la tabla de números de caso .csv, las columnas diferencian las distintas características de un grupo de casos. Se muestra un único grupo de casos por fila. Un grupo de casos no incluye casos individuales. Sin embargo, es posible que el grupo de casos contenga un solo caso. Un grupo de casos se caracteriza básicamente por las siguientes propiedades (las características de estas propiedades se muestran entre paréntesis):

- Localización de las infecciones (IdLandkreis)
- Grupo de personas (sexo, grupo de edad)
- Hora de notificación de la infección (fecha de notificación)
- Inicio de la enfermedad (fecha de referencia, inicio real de la enfermedad)

- Potencia del grupo (número de casos, número de muertes, número de recuperaciones)
- Estado de la notificación (nuevo caso, nueva defunción, nueva recuperación)

Un grupo de casos asume una característica única con respecto a su número de casos ("NumberCase"), "AgeGroup", "Gender", su distrito ("Idistrict"), "NotificationDate", fecha de enfermedad ("Refdate") y la información sobre si se conoce la fecha de enfermedad ("ActualDiseaseStart"). Además, se especifica el "número de fallecimientos" o el "número de recuperaciones" de cada grupo de casos, por lo que sólo puede asumirse una de las dos características "número de fallecimientos" o "número de recuperaciones". Esto significa que si hay muertes o recuperaciones en un grupo de casos, el número de muertes o el número de casos recuperados se especifica en un nuevo grupo. Si, por ejemplo B. ambos casos están en un grupo de casos, el grupo de casos se divide en otros dos grupos, a saber, un grupo de fallecimientos y un grupo de recuperaciones.

---

### Ejemplo

Se registra un nuevo grupo de casos w (IdPaís, sexo, grupo de edad, fecha de registro, fecha de referencia, inicio real de la enfermedad son constantes). Contiene un grupo de casos al principio:

Grupo de casos w: 5 infectados, 0 muertes y 0 casos recuperados

Si 1 de los casos muere y 2 se recuperan, el grupo de casos w se divide en 3 grupos: Grupo de casos x: 2 infectados, 0 muertes y 0 casos recuperados  
Grupo de casos y: 1 persona infectada, 1 muerte y 0 casos recuperados  
Grupo de casos z: 2 personas infectadas, 0 muertes y 2 casos recuperados

---

Las características del estado de notificación indican si, en relación con el día anterior, se han producido cambios en los casos de infección, muertes y recuperaciones en un grupo de casos. Esto permite hacer un seguimiento de los cambios en comparación con el día anterior. Éstos son el resultado de nuevas notificaciones de infecciones (incluidas las notificaciones tardías), correcciones (por ejemplo, debidas a notificaciones erróneas, pero también correcciones relativas al distrito, la edad, el sexo o el inicio de la enfermedad) y cambios en el estado de salud (recuperado, fallecido). Las características del estado de notificación dividen temporalmente los grupos de casos. La división es temporal, ya que sólo muestra los cambios desde el día de la publicación hasta el día anterior. Los casos nuevos forman un grupo de casos separado para el día de la nueva notificación. Como un caso sólo se notifica, se recupera o fallece o se corrige en un día, la división temporal del grupo de casos el día de la nueva notificación del estado de notificación va seguida de una fusión de los grupos al día siguiente. En la sección siguiente se ofrece una explicación más detallada de este proceso.

**Valores característicos** Los datos del número de caso contienen las características y sus valores que se muestran en la siguiente tabla:

Característica	Caracterización	Explicación
IdLandkreis	1001 a 16077	Número de identificación del distrito basado en el Oficial Clave del municipio (AGS) más los 12 Distritos de Berlín (11001 a 11012);
Sexo	W, M, desconocido	Estatuto territorial: 30 de junio de 2020 (2º trimestre) Sexo del grupo de casos: femenino (W), masculino (M) y (desconocido)
Grupo de edadA00-A04	, A05-A14, A15-A34, A35-A59, A60-A79, A80+, desconocido	Rango de edad de los casos incluidos en el grupo, estratificados por 0-4 años, 5-14 años, 15-34 años, 35-59 años, 60-79 años, 80+ años y desconocido
Fecha de registro	AAAA-MM-DD	Fecha en que el asunto ha llegado a conocimiento de la autoridad sanitaria. AAAA corresponde al año, MM al mes y DD al día.
Fecha de actualización	AAAA-MM-DD	Fecha de inicio de la enfermedad. Si no se conoce, la fecha de notificación.
Inicio real de la enfermedad 0, 1		1: Refdate es la Inicio de la enfermedad 0: Refdate es la fecha de notificación

Característica	Caracterización	Explicación
NúmeroCaso	Número entero	Número de informes Casos en los correspondiente Grupo de casos Para NewCase = -1, el Número negativo: Es es un Corrección de la Grupo de casos indicando, cuántas infecciones se ha informado mucho Número de informes
Número de muertes	Número entero	Muertes en el correspondiente Grupo de casos Para NuevaMuerte = -1, es el número es negativo: Es es un Corrección de la Grupo de casos indicando, cuántas muertes se ha informado mucho Número de recuperados
NúmeroGeneses	Número entero	Casos en los correspondiente Grupo de casos Para NeuGenesen = -1, es el número es negativo: Es es un Corrección de la Grupo de casos indicando, Cuántos casos recuperados se ha informado demasiado son

Característica	Característica	Explicación
NuevoCaso, NuevaMuerte, NuevaGénesis	0, 1, -1	<p>0 : Los casos del grupo figuran en la publicación del día en curso y en la del día anterior. Esto significa que estos casos se conocen desde hace más de un día. 1 : Los casos del grupo figuran por primera vez en la publicación actual. Esto significa que son casos recién transmitidos o reevaluados para la fecha de publicación. - 1: Los casos del grupo se incluyen en la publicación del día anterior, pero se eliminan de la publicación del día siguiente.</p> <p>Datos del número de caso eliminados. Esto significa que los casos se eliminan del día actual.</p> <p>Un grupo de casos de este tipo puede surgir, por ejemplo, debido a informes incorrectos, que se muestran como una corrección.</p>



Característica	Característica	Explicación
NuevaMuerte, NuevaGénesis		-9Casos en el grupo no se notifican como recuperados ("Nuevo recuperado") o fallecidos ("Nueva muerte") ni en la publicación del día actual ni en la del día anterior. Esto significa que no se conoce ninguna información sobre la evolución de la infección para los casos del grupo. Esto suele ocurrir, por ejemplo, cuando se acaba de notificar la infección de un grupo de casos.

La división temporal de los grupos de casos por las características del estado de notificación se ilustra en el siguiente ejemplo. Los grupos temporales se caracterizan por un '. Las nuevas notificaciones quedan claras cuando se analizan las características:

### Ejemplo

Si en el día TT se registra un nuevo grupo de casos (IdCounty, sexo, grupo de edad, fecha de notificación, refdate, inicio real de la enfermedad son constantes), asume el estado de notificación NewCase = [1]. Si no se conocen recuperaciones o muertes, pero se notifican en el grupo de casos, NewDeath y NewRecoveries = [-9]:

Los casos del grupo de casos w' son nuevos en el conjunto de datos del día TT (nuevo caso [1]), los casos del grupo no son casos de muerte o recuperación (nueva muerte [-9], nueva recuperación [-9]).

Grupo de casos w':

Infectados [4], fallecidos [0] y casos recuperados [0]

NuevoCaso [1], NuevaMuerte [-9], NuevaGénesis [-9]

Al día siguiente, DD+1, los casos del grupo de casos w' ya no son nuevos. Por tanto, su estado de notificación pasa de [1] a [0]. El grupo de casos temporal w'

(NewCase [1]) se convierte en el grupo de casos continuos w (NewCase [0]):

Grupo de casos w: infectados [4], muertes [0] y casos recuperados [0]  
Nuevo caso [0], nueva muerte [-9], nueva recuperación [-9]

El día DD+1 se registra un nuevo caso adicional en el grupo de casos w. Como se trata de un nuevo caso, vuelve a formar un grupo w' temporal y separado:

Grupo de casos w':  
Infectados [1], fallecidos [0] y casos recuperados [0]  
NuevoCaso [1], NuevaMuerte [-9], NuevaGénesis [-9]

Al día siguiente, DD+2, los casos del grupo de casos w'(DD+1) ya no son nuevos, su estado de notificación cambia como lo hizo el día anterior para el grupo de casos w'(DD). Al cambiar el estado de notificación a w'(TT+1), w' se fusiona con w. El número de personas infectadas en ambos grupos de casos se suma.

Grupo de casos w: infectados [5], muertes [0] y casos recuperados [0]  
Nuevo caso [0], nueva muerte [-9], nueva recuperación [-9]

Las notificaciones de defunciones o recuperaciones son similares a las notificaciones de nuevas infecciones. Estos forman grupos de casos temporales y' y z' que posteriormente se convierten en grupos de casos permanentes y y z:

Día TT+3

>Grupo de casos w:  
>Infectados [4], muertes [0] y casos recuperados [0]  
>NuevoCaso [0], NuevaMuerte [-9], NuevaRecuperación [-9]

Grupo de casos y':  
Infectados [1], fallecidos [1] y casos recuperados [0].  
NuevoCaso [0], NuevaMuerte [1], NuevaGénesis [-9]

Día TT+4

>Grupo de casos w:  
>Infectados [2], muertes [0] y casos recuperados [0]  
>NuevoCaso [0], NuevaMuerte [-9], NuevaRecuperación [-9]

Grupo de casos y:  
Infectados [1], fallecidos [1] y casos recuperados [0].  
NuevoCaso [0], NuevaMuerte [0], NuevaGénesis [-9]

Grupo de casos z':  
Infectados [2], fallecidos [0] y casos recuperados [2]  
NuevoCaso [0], NuevaMuerte [-9], NuevaGénesis [1]

---

Nota sobre los convalecientes

A partir de la información detallada sobre un caso de enfermedad facilitada al RKI por las autoridades sanitarias, se calcula la duración de la enfermedad para cada caso.

. Para los casos en los que sólo se notifican síntomas que indican un curso leve de la enfermedad, se supone que la duración de la enfermedad es de 14 días. Para los casos hospitalizados o con síntomas que indican un curso grave (por ejemplo, neumonía), se supone que la duración de la enfermedad es de 28 días. Para cada caso se calcula una fecha estimada de recuperación basada en el inicio de la enfermedad o, si no se conoce, en la fecha de notificación. Dado que en algunos casos también es posible que la enfermedad dure mucho más tiempo o que la información utilizada no se transmita al RKI en todos los casos, los datos calculados de este modo son sólo estimaciones aproximadas del número de personas recuperadas y, por tanto, sólo deben utilizarse teniendo en cuenta estas limitaciones.

**Formateo de los datos** Los datos de vigilancia del registro de emergencias se incluyen en el conjunto de datos como un archivo .csv separado por comas. El juego de caracteres utilizado en el archivo .csv es UTF-8. Los valores individuales están separados por una coma ",". Las fechas se formatean según la norma ISO 8601.

- Juego de caracteres: UTF-8
- Formato de fecha: ISO 8601
- Separador .csv: Coma ","
- comprimido en formato .xz

### **Metadatos**

Los datos proporcionados se describen con metadatos para aumentar su accesibilidad. Los metadatos se distribuyen a las plataformas correspondientes a través de GitHub Actions. Existe un archivo de metadatos específico para cada plataforma; estos se almacenan en la carpeta de metadatos:

Metadatos/

El control de versiones y la asignación de DOI se realizan a través de Zenodo.org. Los metadatos proporcionados para la importación en Zenodo se almacenan en zenodo.json. La documentación de cada una de las variables de metadatos puede consultarse en <https://developers.zenodo.org/#representation>.

Metadatos/zenodo.json

### **Notas sobre el uso posterior de los datos**

Los datos de investigación abiertos del RKI están disponibles en GitHub.com, Zenodo.org y Edoc.rki.de:

- <https://github.com/robert-koch-institut>

- <https://zenodo.org/communities/robertkochinstitut>
- <https://edoc.rki.de>

### **Licencia**

El conjunto de datos "Infecciones por SARS-CoV-2 en Alemania" está bajo licencia Creative Commons Attribution 4.0 International Public License | CC-BY 4.0 International.

Los datos proporcionados en el conjunto de datos son de libre acceso, siempre que se mencione la fuente. Esto significa que cualquier persona tiene derecho a procesar y modificar los datos, crear derivados del conjunto de datos y utilizarlos con fines comerciales y no comerciales. Encontrará más información sobre la licencia en el archivo LICENSE del conjunto de datos.