EDPS OPINION ON THE USE OF DATA FROM GISAID PLATFORM BY ECDC

(Case 2021-0650)

1. INTRODUCTION

- On 11 June 2021, the EDPS received a request for consultation from the European Centre for Disease Prevention and Control (ECDC) regarding the processing of data available on GISAID platform. ECDC provided additional information upon EDPS request on 9 July, 23 August and 5 October 2021.

- The EDPS issues this Opinion in accordance with Article 58(3)(c) of Regulation (EU) 2018/1725[1] ('the Regulation').

2. BACKGROUND INFORMATION

2.1. ECDC mandate and fight against COVID-19

ECDC, as part of its activities and under its mandate[2], performs sequence analysis to track variants of SARS-CoV-2 (hereinafter 'COVID-19') in order to assess their impact on the pandemic, such as on the transmissibility of the virus, any reduction in vaccine effectiveness and changes in disease severity.

According to Article 3(1) of the ECDC Founding Regulation, ECDC is mandated to **identify, assess and communicate current and emerging threats to human health from communicable diseases**. In particular, Articles 3(2)(a), Article 5(1), Article 10(1) and Article 11(1) authorise ECDC to collect, evaluate and disseminate relevant scientific and technical data, including to identify emerging health threats and to support the networking activities

[1] Regulation (EU) 2018/1725 of the European Parliament and the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ, L 295, 21.11.2018, pp. 39-98.

[2] Under Regulation (EC) No 851/2004 of the European Parliament and of the Council of 21 April 2004 establishing a European Centre for disease prevention and control, OJ, L 142, 30.4.2004, pp. 1-11. (the 'ECDC Founding Regulation')

EUROPEAN DATA PROTECTION SUPERVISOR
Postal address: rue Wiertz 60 - B-1047 Brussels
Offices: rue Montoyer 30 - B-1000 Brussels
E-mail: edps@edps.europa.eu
Website: edps.europa.eu
Tel.: 32 2-283 19 00 - Fax: 32 2-283 19 50

of the competent bodies recognised by the Member States, through the operation of the dedicated surveillance networks and the provision of technical and scientific expertise. ECDC complements its epidemiological surveillance by scaling up the use whole genome sequencing to better detect and investigate multinational outbreaks[3].

When it comes to COVID-19, in particular, the Council has explicitly commissioned ECDC, in point 2, paragraphs 1, 3 and 4, of Recommendation (EU) 2020/912 on the temporary restriction on non-essential travel into the EU[4]:

- to assess the variants of interest and variants of concern, based on key properties of the virus [COVID-19] such as transmission, severity, and ability to escape immune response;
- to process the data concerning "testing rate", "test positivity rate" and "variant of concern and variant of interest", on the basis of information made available to ECDC; and
- to publish and regularly update a map presenting the situation with regard to variants of concern and variants of interest in third countries.

In order to monitor the prevalence of the variants of the SARS-CoV-2 and to assess their epidemiological impact, ECDC needs information about the genetic sequences of the variants of COVID-19 reported by the Member States (the **sequencing data**).

Furthermore, ECDC's scientific input is the basis for the assessment of the epidemiologic situation in third countries and for policy decisions on lifting the current restrictions on non-essential travel into the EU[5].

To perform this work, the ECDC is using notably[6] the data made publicly available by the **Global Initiative for Sharing Influenza Virus Data** (the so-called GISAID Initiative).

The ECDC processing of the COVID-19 related data made available by GISAID (see details below) is the subject matter of this Opinion.

2.2. GISAID

2.2.1. The GISAID Initiative

According to its website[7] and the additional information subsequently provided by ECDC[8]:

- The **GISAID Plaform** was launched on the 61st World Health Assembly in 2008. Created as an alternative to the public domain sharing model, GISAID's sharing mechanism took into account the concerns of Member States by providing a **publicly accessible**

[3] See ECDC Single Programming Document 2021-2023, pages 45 f.
[4] Council Recommendation (EU) 2020/912 of 30 June 2020 on the temporary restriction on non-essential travel into the EU and the possible lifting of such restriction.
[5] Point 1, paragraph 2, of Recommendation (EU) 2020/912.
[6] ECDC and Member States are also using the European Surveillance System (TESSy) for the exchange of information for the prevention of "communicable diseases" which are relevant for the EU, such as tuberculosis, measles, SARS, H1N1 and others. See EDPS Opinion of 3 September 2010: https://edps.europa.eu/sites/default/files/publication/10-09-03_tessy_ecdc_en.pdf
[7] https://www.gisaid.org/
[8] Following the request for consultation, the EDPS and ECDC had a videoconference at staff level on 1 July 2021. Subsequently, the EDPS asked additional questions (email of 2 July 2021) and ECDC provided additional information and documents (email of 8 July 2021 and call of 23 July 2021).

database designed by scientist for scientist, to improve the sharing of influenza data. In 2013, the European Commission recognised GISAID as a research organisation.

- The Initiative promotes the rapid **sharing of data** from all influenza viruses and the coronavirus causing COVID-19. This includes **genetic sequence** and **related clinical and epidemiological data associated with human viruses**, and geographical as well as species-specific data associated with avian and other animal viruses, to help researchers understand how viruses evolve and spread during epidemics and pandemics.

- As to its **governance**, the GISAID Initiative receives administrative support from a German non-profit association (Freunde von GISAID e.V.), organised and operated exclusively for charitable, scientific and educational purposes. GISAID's activities are governed by several organisational bodies that operate independently of each other. The Executive Board of Freunde von GISAID e.V. charged with overseeing the business affairs, is expected to minimise potential conflicts of interest concerning the funding sources. The combination of public and private funding from a number of entities is expected to give GISAID a degree of independence from individual and national interests.

- GISAID maintains a global database for influenza gene sequences and (the EpiFlu™ Database, hereinafter 'GISAID database'). **Open access** to data in GISAID Database is provided **free-of-charge** to all individuals that agreed to identify themselves and uphold the GISAID sharing mechanism governed through its Database Access Agreement[9]. All bonafide users with GISAID access credentials agree to the basic premise of upholding a scientific etiquette, by acknowledging the Originating laboratories providing the specimens, and the Submitting laboratories generating sequence and other metadata, ensuring fair exploitation of results derived from the data, and that all users agree that no restrictions will be attached to data submitted to GISAID, to promote collaboration among researchers on the basis of open sharing of data and respect for all rights and interests.

2.2.2. The EpiCoV database

In January 2020, at the outset of the COVID-19 outbreak, GISAID created a new database called **EpiCoV™** (hereinafter 'EpiCoV')[10], on the same model as EpiFlu[11].

GISAID users (including Member States' authorities) of EpiCov fill in the following data fields[12]:

Mandatory data fields:

- The genetic sequence of a sample of the SARS-CoV-2 virus
- Date of collection of the sample
- Date of submission of the sample to GISAID
- Country of sample
- Laboratory collecting the sample

[9] https://www.gisaid.org/registration/terms-of-use/

[10] cf. note submitted by ECDC together with its consultation ('GISAID EpiCoV - Note on safeguards for patient privacy and confidentiality').

[11] To ECDC's understanding, it is GISAID's intention, that, in default of specific general terms for EpiCoV, the published general terms regarding the use of EpiFlu also apply to the use of EpiCoV, at least by analogy (cf. ECDC email of 23 August 2021).

[12] cf. explanation provided by ECDC in their email to EDPS of 8 July 2021.

- Laboratory submitting the data to GISAID
- Sequencing technology used to analyse the sample

Voluntary data fields:

- Age of the person from whom the sample was isolated (hereinafter 'the patient')
- Gender of the patient
- Clinical status of the patient (diseased, not diseased)
- Generic information on the location of the patient (e.g. city or town of residence, but no work or residential address)
- Vaccination status of the patient.

EpiCoV functions as a **curated database**, which means that GISAID Curation Team reviews each record submitted by a user becoming publicly available in EpiCoV. This database permits the **collection and sharing of genomic (sequence) data on coronaviruses and associated information** (metadata) such as details on the specimen from which the genome was isolated, and basic characteristics of the originating hosts (animal, environment, human). Metadata are reviewed in order to **avoid that data published in EpiCoV can be retraced to the patients** from whom the clinical sample was collected.

To this end, the curation process includes the following checks:

• The name of the entry (virus name) cannot contain information such as person's name, surname, address, ID, internal protocol number of the hospital/laboratory, etc.

• Location information cannot include detailed information about a person's home or work address. Only general information is allowed, such as city or town. Thus, the place where the patient resides or works cannot be reconstructed.

• **Only general epidemiological information** is allowed i.e. age and gender of the patient from whom the sample was collected.

• Verification that, among all the fields in the metadata, data that could be a source of patient identification has not been reported by mistake, ensuring the confidentiality and anonymity of each patient.

• To prevent the identification of the source patient or of a group of patients, the linking of data entries with other information sources, e.g news articles or other websites, is not permitted.

Data Submitters have the ability to directly communicate - in near real-time - with the GISAID Curation Team. These communications allow for timely solutions to submission-related issues. When questionable data are detected, the Curation Team will alert a submitter to address the concerns raised, to permit an update of the entry if necessary. This exchange of messages between the Curation Team and submitter occurs within hours after the entry was uploaded to EpiCoV™. After obtaining the necessary confirmations and updates from the submitter, the entry is publicly released in EpiCoV.

2.3. Processing of the COVID-19 related data retrieved from EpiCov by ECDC

2.3.1 Added value of GISAID EpiCoV for ECDC's work

To perform sequence analysis to track variants of COVID- 19, ECDC highlights that it has an **operational need** to use the data made available by on EpiCoV. According to ECDC, GISAID is the most important entity worldwide providing open access to genomic data of influenza viruses as well as COVID-19. The genetic sequence and related clinical and epidemiological data associated with the virus thus made available from all over the world help ECDC to better understand how COVID-19 evolves and spreads.

ECDC does not upload any data on EpiCoV. It only **collects and further processes data uploaded by other entities** and in particular by Member States.

Member States are not obliged to provide any information to GISAID but ECDC strongly encourages them to do so (see below)[13].

Indeed, ECDC considers that the optimal method is for Member States to:
- report the sequences of the virus to GISAID;
- report the cases to ECDC using the case-based reporting system (TESSy) and include the GISAID identifier in the data thus reported.

According to ECDC, this reporting method allows for the most in-depth analysis of the data, including detection of individual cross-border transmission events, detecting novel variants, and following trends in geography, investigations of vaccine breakthrough events and reinfections, changes in severity of the disease, etc.

ECDC is, however, aware that this may be too resource intensive or otherwise not possible in the Member States. This is why other options are available, such as reporting the sequences to GISAID and reporting aggregated information on variants to ECDC. This model allows ECDC to follow trends in the countries and to detect novel variants but it does not allow for detailed epidemiological analysis that is enabled by case-based reporting.

Therefore, ECDC issued a **'Guidance for representative and targeted genomic SARS-CoV-2 monitoring'** for EU/EEA Member States in May 2021[14]. As regards data sharing and reporting, the guidance provides the following (p. 11):

> *'Data sharing and reporting are described in the report 'Methods for the detection and identification of SARS-CoV-2 variants' published by ECDC and WHO's Regional Office for Europe on 3 March 2021 and ECDC's Technical Guidance 'Sequencing of SARS-CoV-2: first update' published on 18 January 2021.*
>
> *Detections of SARS-CoV-2 should be reported on a weekly basis to The European Surveillance System (TESSy). Detection of novel VOCs or outbreaks of currently circulating VOCs should be reported immediately through the Early Warning and Response System (EWRS), while VOC detections should be reported to TESSy on a weekly basis.*

[13] See ECDC replies to EDPS questions by email of 8 July 2021.
[14] The guidance is publicly available on ECDC website: technical report of 3 May 2021 on 'Guidance for representative and targeted genomic SARS-CoV-2 monitoring

In addition, ***it is recommended that SARS-CoV-2 sequences are submitted to GISAID (https://www.gisaid.org), the European COVID-19 Data Platform (https://www.covid19dataportal.org)***[15] *or other public databases in a timely manner (ideally within one or two weeks of sample collection).* ***Raw data***[16] ***can also be deposited in the COVID-19 Data Platform. Member States should not upload any information or data that can link back the submitted virus sequence to an individual patient in a national database****,*[17] *such as personal identifiers (e.g. name, date of birth, unique identification code) or granular demographic data (e.g. the patient's age) etc. It is the Member State's responsibility to ensure that sequencing data are submitted in accordance with all pertinent rules and regulations.*

It is also the responsibility of reporting Member States to assess whether the virus is a variant included in the variant list, irrespective of the method used for detection/identification. Variables for reporting VOCs have been implemented within the aggregated (NCOVAGGR) and case-based (NCOV) TESSy record types, where sequence ID ECDC TECHNICAL REPORT Guidance for representative and targeted genomic SARS-CoV-2 monitoring 12 numbers (e.g. ***GISAID identifiers***[18]*) should be reported as well. ENA/SRA accession numbers pointing to raw sequence data, can also be submitted to TESSy. It is also important to share raw sequencing reads because this allows uniform bioinformatic approaches for consensus genome generation, making it possible to avoid potential errors in the generation of consensus sequences while also enabling further analyses (e.g. minority variants). Any epidemiological data available, including severity and probable country of infection, should be reported if data are submitted using the case-based record type (NCOV).* (...)'.

2.3.2. National record ID attributed by Member States and secondary code ID attributed by GISAID

According to ECDC[19], Member States process case-based health data (data pertaining to a particular case of the COVID-19 disease) under a **national record ID**. This national record ID is awarded by the public health authorities of the Member States in accordance with their national law. Member States use the case-based health data for their own analysis. In that context they are subject to the GDPR[20]. The main purposes of the national record ID are to make sure that there are no duplicate entries of the same virus, to be able to update existing records with new information, and for users to be able to request further information from the data provider in case of investigations involving more than one laboratory. As indicated above, this national record ID is not the passport or national ID card number.

With regard to ECDC's analysis of sequencing data in particular, the national record ID, which is included in the case-based reporting made available to ECDC via **TESSy**, aims to ensure that the analysed sequencing data are accurate. This is of paramount importance in

---

15 We underline.

16 The ECDC considers raw data as a raw readout generated by the sequencing laboratory equipment and primary data analysis pipeline and comprising of thousands to millions of non-curated short sequences, as mentioned in the reply of 23 August 2021.

17 We underline.

18 We underline.

19 See ECDC replies to EDPS questions by email of 8 July 2021.

20 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

order to enable the follow-up of any concerns with the Member State and also to ensure that ECDC's scientific outputs comply with established scientific methods and standards.

ECDC is not involved in the attribution of the national record ID. **ECDC** does not possess the key to the national record ID and **has** therefore **not the means to re-identify** the patient from whom the sample has been isolated.

The **secondary code ID** (or GISAID identifier), which pertains to a particular sample of COVID-19 sequences, is randomly created by GISAID when data are uploaded by data submitters (Member States and others). The code ID of the sample does not contain any element that is related to the patient from whom the sample was isolated (see above the description of the curation process).

The ECDC also stated that TESSy has a non-mandatory field where countries can report the associated GISAID identifier for each case. This field has been used by a few Member States. There may be also cases where Member States consistently use the same exact identifier for TESSy and GISAID reporting.

2.3.3. Categories of data collected on EpiCoV and further processing by ECDC

All authorised users of EpiCoV have access to all data, both **sequencing data** (the genome sequence of the virus) and the **metadata** (e.g. age range, gender, clinical status as listed above) stored in the database. They do not have access to the national record ID attributed by the Member States, as this cannot be uploaded to EpiCoV (see the GISAID's Note on safeguards for patient privacy and confidentiality). The random secondary code ID of the sample is, however, available to all authorised users of EpiCoV.

In addition, GISAID provides ECDC with **aggregated data** on the prevalence of COVID-19. These aggregated data are based on the sequencing data stored in EpiCoV. These data are used for epidemiological surveillance purposes. In particular, GISAID provides ECDC with instant, real-time updates of the aggregated data on the prevalence of variants of COVID-19 in the EU/EEA. These data are designed for use in ECDC's weekly outputs and they are shown in ECDC's variant tracker dashboard on ECDC's webpage. As the aggregated data are fully anonymised, no specific security measures apply; they are transferred to ECDC using a secured Application Programming Interface (API).

A limited amount of ECDC's scientific staff are authorised users. These staff members may sign in to EpiCoV with their credentials. They access EpiCoV from their work computer. This processing is subject to the ECDC Information Security Policy (ECDC/IP/63), and the standard ICT security measures implemented to protect ECDC ICT infrastructure, including firewalls, antivirus protection, data encryption and passwords where deemed necessary. Data retrieved from EpiCov and further procccessed by ECDC cannot be accessed by unauthorised users and access from outside ECDC is only possible by using an ECDC secure VPN connection. Data is backed-up regularly and security software is kept up-to-date. ECDC offices are not publicly accessible and are secured in accordance with EU safety/security standards.

ECDC has no retention policy in place as regards the data retrieved from EpiCov and further processed.

2.3.4. ECDC's questions

Considering the information mentioned above, the ECDC wishes to know:

- if the data available on EpiCov is considered personal data;
- if so, what is the nature of the relationship between ECDC and Member States as regards the use of GISAID; and
- if so, how should the ECDC process the data in-house in order to comply with the Regulation.

3. LEGAL ANALYSIS AND RECOMMENDATIONS

This EDPS opinion relates to the data processing operations performed by the ECDC regarding the **collection and further use of data from EpiCoV**. Within this scope, the EDPS highlights below his analysis and recommendations.

3.1. Is the data processed by the ECDC personal data?

'Personal data' is any information that is relating to an identified or identifiable natural person (the data subject). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person[21].

'Pseudonymisation' means the processing of personal data in such a manner that they can no longer be attributed to a specific data subject without the use of additional information, provided that:

- such additional information is kept separately and
- is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

According to Recital 16 of the Regulation, 'Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information, should be considered to be information on an identifiable natural person. **To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person, to identify the natural person directly or indirectly[22].** To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.'

[21] Article 3(1) of the Regulation.
[22] We underline.

Thus, personal data which have undergone pseudonymisation should be considered to be information on an identifiable natural person, because the additional information – such as a code or a number – makes it possible to identify the person[23]. It is also important to note that, in view of the definition of pseudonymous data, it does not matter that those who process the pseudonymous data and those who hold the additional information are not the same. The fact that they are different entities does not render pseudonymous data anonymous. Pseudonymisation is considered to be an additional safeguard which should be employed in the context of scientific research to ensure respect for the principle of data minimisation[24].

At the same time, personal data are such only when the data subject is identifiable taking into account all means reasonably likely to be used either by the controller or by another person, to identify the natural person directly or indirectly (recital 16 to Regulation 2018/1725). Also, in the judgment Breyer (C-582/14), the Court of Justice states that personal data can be such in relation to a specific entity where such entity has the legal means that enables it to identify a data subject by combining additional data from another source. The Breyer-judgment does not imply that the classification as anonymous information solely depends on the fact that additional information can be obtained 'legally' by the controller or a third party. This judgment seems to only use as an example that when the law itself provides for the possibility to gain access to additional information for the purpose of re-identifying a data subject, it is objectively clear that a 'means reasonably likely to be used' is available to re-identify the data subject. Therefore, whenever it is possible to re-identify the data subject, it cannot be said that the information is anonymous.

In the present case, where data are pseudonymised, the only entities that would have the key to potentially link the case-ID assigned by GISAID to a data subject are the uploaders of that data (Members States, which are joint controllers and fall under the GDPR).

In view of the Breyer ruling, ECDC points out that the data uploaded by the Member States (joint controllers with ECDC), might be personal data only with regard to the uploading entities (i.e. the Member States themselves, which are the only entities that can re-identify the data subjects). The ECDC stated that GISAID and any other users that have access to the data on GISAID do not have any legal mean to re-identify the data subjects, and there are no means reasonably likely to be used that would allow such re-identification.

In this respect, the EDPS underlines that the Breyer ruling was rendered before the entry into force of the Regulation, which now includes the definition of 'pseudonymisation', as further outlined by Recital 16, which encompasses a broad interpretation of pseudonymised data.

In the present case, TESSy has a non-mandatory field where Member States can include the GISAID identifier[25]. Moreover, some Member States consistently use the same exact identifier for the TESSy and the GISAID reporting in EpiCoV. Therefore, the ECDC is in a

---

[23] Recital 16 of the Regulation.

[24] Article 13 of the Regulation. See also Para 44, page 11 of the EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research (2 February 2021) : https://edpb.europa.eu/sites/default/files/files/file1/edpb_replyec_questionnaireresearch_final.pdf

[25] The EDPS issued a prior-checking Opinion on TESSY in 2010, which considers that the data processed viaTESSy are personal data (https://edps.europa.eu/sites/default/files/publication/10-09-03_tessy_ecdc_en.pdf)

position that allows it, at least in some instances, to make a link between the data obtained from GISAID (EpiCoV) to data obtained via TESSy. The latter includes the national record ID, which is pseudonymised data. Thus, despite the ECDC not having access to the key to the national record ID, the data collected from EpiCoV by ECDC should be considered as pseudonymised personal data, as data subjects remain indirectly identifiable[26], at least in certain cases (i.e. when it is possible to make a link between data stored on TESSy and those stored on EpiCoV). The EDPS understands that the ECDC does not seem interested in re-identifying the data subjects through the combination of data in TESSy and GISAID. Nonetheless, we cannot deny that the ECDC, given its close cooperation with the Member States and also the fact that some Member States use the exact same identifier when reporting both in GISAID and TESSy, is in abstract in a position to perform such re-identification.

In view of the above; the particular samples of SARS-CoV-2 sequences uploaded to EpiCoV™ and related information are therefore personal data which have undergone pseudonymisation by GISAID and are processed by the ECDC under Recital 16 and Article 3 of the Regulation.

3.2. Relationship between ECDC and the Member States

ECDC considers it acts as joint controller with the Member States regarding the processing of personal data from GISAID, while pointing out that there is no alternative to GISAID and that it is in a 'take it or leave it' situation vis-à-vis GISAID .

First, it is important to identify and explain the concept of a controller of a processing operation within the meaning of the Regulation. A controller is the EU institution or body, or the directorate-general or any other organisational entity which alone or jointly with others determines the purposes and means of the processing of personal data[27].

The controller decides why and how data will be processed.

The controller has actual control, decisive influence, and predominant role over a processing. Hence a controller may cause a processing, can end a processing, can benefit from a processing, may have an interest in a processing[28].

In order to assess whether an entity is a controller or not within the meaning of the Regulation, a list of specific elements should be considered[29].

In particular, only the controller decides exclusively about the essential elements of a processing, namely,

- its legal basis,
- for which purpose,
- which personal data are processed,
- from which individuals will be collected and processed,

---

[26] See recital 16 of the Regulation.
[27] Article 3(8) of the Regulation.
[28] More on the concept of controller, pages 7- 12 of the EDPS Guidelines on the concepts of controller, processor and joint controllership under the Regulation: https://edps.europa.eu/sites/default/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf
[29] Idem, page 13 of the EDPS Guidelines.

- whether the processed data will be disclosed and to whom,
- for how long the personal data will be stored (retention period),
- how data subjects can exercise their rights and how the controller will respond to their requests
- has an autonomy and independence within their assigned tasks
- selects a processor to carry out the processing or part of it on their behalf even if the processor chosen implements specific technical and organisational means (non-essential elements of the processing).

If the entity replies in the affirmative to the majority of the above elements, then the entity qualifies as a controller within the meaning of the Regulation.

In the case at hand, ECDC extracts specific categories of data from GISAID in order to accomplish its own tasks as exposed above in page 2 in line with its Founding Regulation. ECDC determines the purpose of processing the data from GISAID as well as most of the essential means of the processing, namely which data will be processed for its specific purpose, from which categories of data subjects, the recipients, the retention policy and it is autonomous to accomplish its tasks. In light of the above elements, ECDC has an actual control and decisive influence over the data it processes extracting them from GISAID, hence it is a **controller** under Article 3(8) of the Regulation.

The question raised is whether ECDC and Member States can qualify as joint controllers. Joint controllership occurs when two or more controllers together with one or more controllers other than Union institutions and bodies jointly determine the purposes and means of the processing, they are joint controllers[30].

Member States voluntarily upload data on GISAID, which they collect and use for their own analysis under their legal framework including GDPR. ECDC extracts data from GISAID for its own assessments and publications. ECDC published on its website a technical report offering guidance to the Member States. Although both ECDC and Member States process data for similar purposes, they both benefit from GISAID in a separate way. ECDC's report is only guidance and recommendations for best cooperation, but it is not a binding instrument on the Member States and thus cannot be considered as exerting a decisive influence on the purpose and essential elements of the means.

These elements lead to the conclusion that both entities are autonomous in accomplishing their respective analysis of the data and they do not jointly determine any essential means of the processing.

Hence, ECDC and Member States cannot therefore qualify joint controllers within the meaning of Article 28 of the Regulation, but rather as **separate controllers**.

[30] *Article 28(1) of the Regulation.* Examples of already existing joint controllership relationships within the EUIs are the shared EU HR online tool, the EU database for experts, the Portals for research projects of the Executive Agencies, the Platforms for rare diseases with Member States etc .. See also pages 22-26 of the EDPS Guidelines: https://edps.europa.eu/sites/default/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf

3.3. Relationship between ECDC and GISAID

ECDC points out that it does not need to provide specific instructions to GISAID as processor of personal data, nor should ECDC be concerned about who has access to the data. ECDC is currently discussing an agreement to formalise its relation with GISAID.

It is important to identify and explain the concept of processor within the meaning of the Regulation. A processor is a natural or legal person, a public authority, or an agency or other body that processes personal data on behalf of and under the instructions of a controller[31].

In order to assess whether an entity is a processor or not, it is important to go through the following checklist:

A processor should not decide on the essential elements of a processing, namely

- its legal basis,
- its purpose,
- which personal data should be collected and from which individuals
- whether the processed data should be disclosed and to whom,
- the retention period of the data collected.

The processor should only serve the controller's interest in carrying out the assigned task(s) and may have a degree of independence and can make suggestions on matters within their expertise and how they would carry out their tasks in order to comply with the controller's instructions, (for example which equipment to use, which security measures to take).

However, the final decision is up to the controller after having been fully informed on the elements, risks and measures proposed by the processor. The controller and the processor should implement such decisions under a contract or another legal act or binding arrangement.

In the case at hand, ECDC does not give any instructions to GISAID, but it benefits from its open access free of charge service. The fact that ECDC has accepted specific terms of GISAID in order to be able to use the data, does not automatically qualify ECDC as a processor. In light of the checklist above, a processor should not decide on the essential means of a processing, namely the legal basis of the processing of the data used, its purpose, from which data subjects are collected etc. ECDC on the contrary, has a decisive influence on under which legal basis, why and how the data will be processed after extracting them from GISAID and to whom they will be disclosed.

In light of the above, ECDC and GISAID are **separate controllers** determining different purposes and means on the processing of the data, but with obligations under the Regulation and the GDPR respectively.

[31] Article 3(12) of the Regulation. See pages 15-18 of the EDPS Guidelines: https://edps.europa.eu/sites/default/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf

3.4. ECDC's obligations as a controller of the data collected from EpiCoV

The ECDC referred that "EpiCoV is available to accredited members of the scientific community (incl. scholars and researchers from the EU/EEA and from third countries). As authorised users in their personal capacity, ECDC staff can access all the data stored on EpiCoV and harness it for the ECDC's scientific outputs (e.g., risk assessments)." Therefore, the ECDC should give clear instructions to their staff regarding how to process personal data for this processing operation.

It is also stated by the ECDC that "..., where other entities besides Member States upload personal data outside of the recommendations of ECDC, we do not consider that ECDC is in a situation of controllership, as ECDC cannot fully determine the purpose and means of such further processing operations."

These circumstances refer to the processing of data on EpiCov by other entities than ECDC for their own purposes and do not exempt ECDC from any responsibility under the Regulation. Indeed, having regard to the above, the ECDC is a controller in their own data processing operation, i.e. in the data ECDC collects from EpiCoV and further processes within its mandate.

However, considering that the ECDC is not in a position to always identify data subjects, Article 12 of the Regulation is applicable. If the ECDC is in a position to demonstrate that it cannot identify the data subject, for example, if there is a right to access request, the ECDC does not need to perform the re-identification of the data subjects, in order to reply to that request. Therefore, in line with this Article, the ECDC may be exempted of certain controller's obligations, such as the ones described in Articles 17 to 22 of the Regulation.

In any event, within its role as a controller, the ECDC has the obligation to comply with the Regulation and to demonstrate that compliance, even when is processing personal data which does not require identification of data subjects.

Recommendation 1: The ECDC should inform data subjects about this processing operation through a general data protection notice published on their website that would indicate the processing of pseudonymised data only, without the identification of data subjects.

3.5. Recommendations regarding the processing of data by ECDC

3.5.1. Retention period

The ECDC has not issued a retention policy with regard to the processing of sequencing data on SARS-CoV-2 viruses and related personal data.

In light of the storage limitation principle[32] the controller is responsible for adopting a maximum retention period for the personal data that are collected which is necessary for the purpose for which they are processed. Personal data may be stored for longer periods when the personal data are processed for scientific purposes in accordance with Article 13 subject to implementation of the appropriate technical and organisational measures.[33] ECDC

[32] Article 4(2) of the Regulation.
[33] Article 4(1)(e) of the Regulation.

processes pseudonymised data, which is already an appropriate safeguard for the rights and freedoms of the patients.

Recommendation 2: ECDC should set up a maximum retention period for the data it processes assessing the purpose for which they were used and the necessity for keeping them for (x) period under Article 4(1)(e) of the Regulation.

3.5.2. Technical and Organisational security measures

Personal data must be processed in accordance with the data protection principles set up in Article 4 of the Regulation to guarantee data quality. In particular, the accuracy principle (Article 4(d) of the Regulation) is key for these processing operations in order to ensure with reasonable certainty that the data are accurate and up to date.

In accordance with Article 33 of the Regulation, the ECDC as the controller should implement appropriate technical and organisational measures to ensure a level of security appropriate to the risks while having in mind how to minimise those risks to the rights and freedoms of the patients.

As stated in Article 12(1) of the Regulation, if the purposes for which a controller processes personal data do not or no longer require the identification of a data subject by the controller, the controller is not obliged to maintain, acquire or process additional information in order to identify the data subject.

Having into consideration that the data is pseudonymised and that the ECDC does not have a priori access to the pseudonymisation key, the EDPS recognises that the ECDC is free to adopt the appropriate technical and security measures, which can be lowered according to the risks to the data subjects.

Nonetheless, as controller in this processing operation, the ECDC has to implement security measures to ensure that the personal data are processed in accordance with the Regulation, including implementing data protection by design and by default.

The ECDC issued guidelines to the Member States on how to upload information on GISAID with a view to ECDC further optimal use of this information for its own purposes.

Recommendation 3: The ECDC should inform its staff who are users of GISAID of the purpose of these processing operations, as well as their terms and conditions. Furthermore, the ECDC should recommend that staff do not take action to obtain additional information, which might re-identify the data subjects.

Recommendation 4: Considering the constant evolution of technologies, including the possibility to re-identify data subjects based on the information available on EpiCov, the ECDC should regularly assess the level of security of this processing operation, taking into account its particular risks.

4. CONCLUSION

The EDPS has analysed the personal data processing regarding the collection and further processing of the COVID-19 related data made publicly available by GISAID, namely on the concept of personal data, the relationship between ECDC and GISAID, the relationship between ECDC and the Member States and the role of ECDC as controller.

The EDPS has made several recommendations to ensure compliance of the processing with the Regulation, in particular regarding the retention period, the organisational and technical security measures.

In view of the accountability principle, the EDPS expects that the ECDC **implements the above-mentioned recommendations and has therefore decided to close the case**.

Done at Brussels on 30 November 2021

[e-signed]

Wojciech Rafał WIEWIÓROWSKI