

# Internal Procedure on Automated Monitoring of Social Media for Epidemic Intelligence

| Step          | Name       | Function/Section/Unit                                                                             |
|---------------|------------|---------------------------------------------------------------------------------------------------|
| Prepared by:  | [REDACTED] | [REDACTED]                                                                                        |
| Verified by:  | [REDACTED] | Central Coordinator, Planning and Performance Section,<br>Resource Management Services (RMS) Unit |
| Reviewed by:  | [REDACTED] | [REDACTED]                                                                                        |
| Validated by: | [REDACTED] | [REDACTED]                                                                                        |
| Reviewed by:  | [REDACTED] | [REDACTED]                                                                                        |
| Approved by:  | [REDACTED] | [REDACTED]                                                                                        |

**Document Control Sheet****Log of Issuance**

| <b>Issue</b> | <b>Approval Date</b> | <b>Change Description</b>                                                         |
|--------------|----------------------|-----------------------------------------------------------------------------------|
| 000          | 24/03/2026           | First version of the document; transposed to a new IP template and content review |
| 001          |                      |                                                                                   |
| 002          |                      |                                                                                   |
| 003          |                      |                                                                                   |
| 004          |                      |                                                                                   |
| 005          |                      |                                                                                   |
| 006          |                      |                                                                                   |
| 007          |                      |                                                                                   |
| 008          |                      |                                                                                   |
| 009          |                      |                                                                                   |
| 010          |                      |                                                                                   |

**Next Review Deadline**

Three years following approval of this document

# Table of Contents

|                                                        |           |
|--------------------------------------------------------|-----------|
| <b>1 INTRODUCTION</b>                                  | <b>5</b>  |
| <b>1.1 Purpose</b>                                     | 5         |
| <b>1.2 Scope</b>                                       | 5         |
| <b>1.3 Guiding principles</b>                          | 5         |
| <b>1.4 Roles and responsibilities</b>                  | 6         |
| <b>2 PROCEDURE WORKFLOW, ROLES AND RELATED RECORDS</b> | <b>9</b>  |
| <b>3 SEGREGATION OF DUTIES AND CONTROL ACTIVITIES</b>  | <b>13</b> |

| Related Documents                          |                                                                                                                                                                                                                                                                                                                                                                                                     |
|--------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>List of relevant external documents</b> | <ul style="list-style-type: none"> <li>• Regulation (EC) No. 851/2004 establishing a European centre for disease prevention and control</li> <li>• Regulation (EU) 2022/2371 of the European Parliament and of the Council of 23 November 2022 on serious cross-border threats to health and repealing Decision No. 1082/2013/EU (Text with EEA relevance), in particular art 3(6)</li> </ul>       |
| <b>List of relevant ECDC documents</b>     | <ul style="list-style-type: none"> <li>• ECDC Internal Control Framework (Chapter III. Control Activities (Article 10.3)</li> <li>• ECDC Financial Regulation (Article 45.7)</li> <li>• Internal Procedure on Manual Monitoring of Social Media for Epidemic Intelligence (ECDC/IP/146)</li> <li>• Amended ECDC Strategy 2021-2027</li> <li>• ECDC Single Programming Document 2026-2028</li> </ul> |

| Abbreviations/<br>Terms | Descriptions/Definitions                                                                                              |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------|
| AI                      | Artificial Intelligence                                                                                               |
| API                     | Application Programming Interface                                                                                     |
| EBS                     | Event-based surveillance                                                                                              |
| ECDC                    | European Centre for Disease Prevention and Control                                                                    |
| EEA                     | European Economic Area                                                                                                |
| EI                      | Epidemic Intelligence                                                                                                 |
| EU                      | European Union                                                                                                        |
| IP                      | Internal Procedure                                                                                                    |
| Lucene indexes          | Data structure that maps unique terms to the documents containing them, enabling extremely fast search and retrieval. |
| SPR                     | Surveillance, Preparedness and Response                                                                               |
| RT                      | Round Table                                                                                                           |
| VM                      | Virtual Machine                                                                                                       |
| VPN                     | Virtual Private Network                                                                                               |

# 1 Introduction

## 1.1 Purpose

The purpose of this internal procedure (IP) is to define and document the workflow of automated monitoring of publicly available social media posts from BlueSky and any future public micro-blogging services that expose comparable open Application Programming Interface (APIs), performed by the European Centre for Disease Prevention and Control (ECDC) using an R-based package named episomer for epidemic intelligence (EI) activities. This R-based package and pipeline support ECDC's statutory mandate to identify, assess and communicate current and emerging threats to human health from communicable diseases (Regulation 851/2004).

The two main objectives of this automated monitoring of publicly available social media posts using episomer are:

- 1) Automated early warning of potential public health threats: Unusual increases in disease-related conversations in social media has been shown to provide early warning of potential public health threats, often faster than traditional surveillance. Episomer enables the automated monitoring of conversations across social media platforms based on predefined keywords, allowing the detection of unusual increases that may trigger more in-depth assessment and investigation of additional information sources, including official sources, by EI analysts.
- 2) Improvement of model for text geolocation and alerts: Unsupervised machine learning is used to extract geolocation information from text and to classify alerts (i.e., unusual increases in the number of posts per topic, time and location). As unsupervised models cannot be expected to achieve 100% accuracy, manual review of model outputs is required in order to validate results and iteratively improve model performance.

## 1.2 Scope

This IP applies to staff from the Global Epidemic Intelligence and Health Security (GEIHS) and to episomer users involved in the activities described below.

EI activities include the following steps: screening, filtering, validation, analysis, documentation, and communication. This IP focuses on the first three steps.

BlueSky is the sole social media platform used for episomer at this point and this IP remains valid for additional social media platforms provided that (i) only public posts are collected, (ii) the personal data categories do not change, and (iii) the processing steps and safeguards remain the same. An internal assessment of these three conditions will be performed for every new social media platform added and a full Data Protection Impact Assessment (DPIA) update will be required only if any of the three conditions above is no longer met.

It does not cover: (i) the manual monitoring of social media for EI, (ii) the dissemination of information and data by ECDC on social media nor the dissemination of the latter for any other means, (iii) the search and collection of information and data by other sources such as the use of established surveillance networks and databases (e.g., EpiPulse), restricted platforms or from official public sources (e.g., reports from WHO or national public health agencies).

## 1.3 Guiding principles

EI aims to rapidly detect and assess health events related to communicable diseases that threaten EU/EEA Member States and their citizens. The information is collected systematically and collated from a variety of sources, including social media. This ensures a timely response proportionate to risk, and provision of recommendations on appropriate public health measures.

EI experts screen and process a virtually unlimited amount of information from multiple sources to detect any relevant information such as, for example, communicable disease events of public health importance or with high public, media, or political interest, or reports or news of serious unusual or unexpected outbreaks or events of unknown origin. Screening must be sensitive and timely ensuring an early detection of any signal of a potential public health threat, aiming to have a sufficient specificity to be able to handle this huge amount of information in a timely matter.

After screening and processing the different sources, EI experts will filter the information by selecting and deciding which ones constitute items of interest that will require further investigation and followup, and which ones should be discarded early in the EI process. This is done based on pre-defined criteria related to mandate, legislation and context. Validation is the process of confirming the accuracy and credibility of information received from non-official sources. Once the information is validated and deemed accurate, data are further analysed and communicated to relevant stakeholders. The whole process is documented.

Episomer is an R-based package with Scala for the machine learning model to extract text geolocation. This R-based package does not come with data. Users must provide their own credentials to collect data from BlueSky and other social media open APIs.

Episomer is executed in an ECDC-restricted environment. All scripts run on an ECDC virtual machine (VM) in the server in premises which requires VPN and its access is limited to the 24/7 duty officer and ECDC epidemic intelligence team. Data are stored also in this ECDC VM. Raw posts from social media are stored as Lucene indexes which require specific skills and processes to visualise these data.

Social media is one of the sources that ECDC monitors for epidemiological surveillance purposes. The processing operations and the amount of data processed are strictly those that are necessary to achieve the purposes described in the introduction and procedure workflow. The data being processed have been made public by the data subjects. The processing operations are designed to avoid targeting aspects of the life of the individual data subjects, aiming instead at information to be aggregated and that is relevant for early warning of public health threats.

## 1.4 Roles and responsibilities

| Roles                             | Responsibilities                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Epidemic Intelligence (EI) expert | <ul style="list-style-type: none"> <li>• Contributes to the collection, validation, analysis, interpretation and dissemination of communicable disease surveillance data.</li> <li>• Participates in 24/7 duty system (24/7 duty officer and Threat Detection Officer).</li> <li>• Collects data on long-term monitoring of selected diseases, presents the results at the Round Table meeting and includes these in the CDTRs.</li> </ul> |
| Episomer maintainers at ECDC      | <ul style="list-style-type: none"> <li>• Ensuring the functioning of episomer for the purposes mentioned in this IP.</li> <li>• Upgrading episomer when relevant.</li> <li>• Fixing major issues jeopardising the usability of main features of episomer for ECDC's use.</li> </ul>                                                                                                                                                        |

|                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                        | <ul style="list-style-type: none"> <li>• Communicating the upgrades and status of episomer to users.</li> <li>• Manually annotating a selected list of posts or alerts for improving geolocation and alert models in episomer.</li> </ul>                                                                                                                                                                                                                                       |
| Episomer users at ECDC                                                                                                 | <ul style="list-style-type: none"> <li>• Validating the signals detected by episomer.</li> <li>• Understanding the use and procedures of episomer data.</li> <li>• Informing episomer maintainers of any issue.</li> <li>• Manually annotating a selected list of posts or alerts for improving geolocation and alert models in episomer.</li> </ul>                                                                                                                            |
| Head of Section Global Epidemic Intelligence and Health Security                                                       | <ul style="list-style-type: none"> <li>• Plans and supervises the work and projects of the EI experts, including episomer.</li> <li>• Ensures smooth operations of the EI team and 24/7 duty system.</li> <li>• Liaises with internal and external stakeholders in relation to EI activities.</li> <li>• Manages the staff members of the EI experts.</li> <li>• Contact social media users when they are being followed up by ECDC for EI purposes.</li> </ul>                 |
| Head of Unit Surveillance, Preparedness and Response                                                                   | <ul style="list-style-type: none"> <li>• Manages staff within the section, including the preparation and implementation of the section's work plan and budget.</li> <li>• Decides whether RRA should be undertaken in consultation with experts from public health functions and diseases.</li> </ul>                                                                                                                                                                           |
| RT presenter (i.e., 24/7 duty officer and other relevant ECDC experts, EI experts involved in preparation of the item) | <ul style="list-style-type: none"> <li>• Provides an oral report of the respective disease or event.</li> <li>• Prepares the text to be included in the CDTR for the respective disease or event.</li> </ul>                                                                                                                                                                                                                                                                    |
| Threat Detection Officer (TDO) or Epidemic Intelligence Support Officer (EISO)                                         | <ul style="list-style-type: none"> <li>• Supports 24/7 Duty officer during working hours with activities related to EBS, including daily screening.</li> <li>• Oversees coordination and production of the daily and weekly communicable disease threat report (CDTR).</li> <li>• Disseminates EI findings to relevant internal and external stakeholders.</li> <li>• Responsible for using the ECDC tool Epipulse to disseminate the EI findings.</li> </ul>                   |
| 24/7 duty officer                                                                                                      | <ul style="list-style-type: none"> <li>• Ensures 24 hours a day, 7 days a week continuity of EI activities, including threat detection, all year round.</li> <li>• Oversees EI routine procedures</li> <li>• Responsible for work and outputs done to the highest standards.</li> <li>• Presents the daily findings from EI activities at the ECDC Round Table meeting which aims to discuss these with experts from different public health functions and diseases.</li> </ul> |

|  |                                                                                                                                                                                                         |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <ul style="list-style-type: none"><li>• Reviews text to be included in the daily and weekly CDTRs before being presented internally at ECDC and cleared by the chair of the ECDC Round Table.</li></ul> |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## 2 Procedure Workflow, Roles and Related Records

| Description of the Process Steps and their Sub-steps                                                                                                                                       | Resp. Role(s)                                     | Template and/or Form if applicable | Record Owner/ Storage Location                                                                                                                                             | Security Level <sup>1</sup>               |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| <b>Process Step 1: Data collection and preprocessing</b>                                                                                                                                   |                                                   |                                    |                                                                                                                                                                            |                                           |
| 1.1 Public posts containing ECDC-defined keywords of interest are collected using public social media APIs. Approximately, 150 000 BlueSky posts are expected to be collected every month. | R functions from episomer<br>Episomer maintainers | N/A                                | Record: Episomer raw data<br><br>Record owner: Epidemic Intelligence<br><br>Storage location: ECDC-controlled environment (e.g., VM with two-factor authentication access) | √ ECDC NORMAL<br>ECDC SENSITIVE<br>PUBLIC |

<sup>1</sup> Please consult the table in Chapter 5 of the Policy on Information Classification (ECDC/POL/05) to identify the security level for your records. For any questions, contact the Front Office or DTS IT Security ([front.office@ecdc.europa.eu](mailto:front.office@ecdc.europa.eu)) or [cert@ecdc.europa.eu](mailto:cert@ecdc.europa.eu).

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                   |     |                                                                                                                                                                                           |                                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| <p>1.2 New field is added with the location extracted from the text of the posts, fields of the collected data are renamed to match the standardised format for all social media platforms, and data format is modified to Lucene indexes.</p> <p>These Lucene indexes are stored in an ECDC-restricted environment. Specific skills and processes are required to visualise the raw files stored as Lucene indexes.</p> <p>Raw data are kept for ca. 3 months in the ECDC-controlled environment. This retention period is based on potential downtime of the tool. Data is available only in the ECDC-controlled environment, which requires VPN and is only available to 24/7 duty officer and ECDC EI team</p> | Episomer maintainers                              | N/A |                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>✓ ECDC NORMAL</li> <li>ECDC</li> <li>SENSITIVE</li> <li>PUBLIC</li> </ul> |
| <b>Process Step 2: Data aggregation</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                   |     |                                                                                                                                                                                           |                                                                                                                  |
| <p>2.1 Raw posts stored as Lucene indexes are aggregated per topic, day and text location up to country level before being processed. These aggregated data are also stored in the ECDC-controlled environment.</p> <p>No personal data is available in these aggregated data.</p> <p>Aggregated data is kept for ca. 6 months. This retention period is based on epidemiological seasonality</p>                                                                                                                                                                                                                                                                                                                  | R functions from episomer<br>Episomer maintainers | N/A | <p>Record: Episomer aggregated data</p> <p>Record owner: Epidemic Intelligence/</p> <p>Storage location: ECDC-controlled environment (e.g., VM with two-factor authentication access)</p> | <ul style="list-style-type: none"> <li>✓ ECDC NORMAL</li> <li>ECDC</li> <li>SENSITIVE</li> <li>PUBLIC</li> </ul> |
| <b>Process Step 3: Data analysis</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                   |     |                                                                                                                                                                                           |                                                                                                                  |
| <p>3.1 Aggregated data is used for producing visualisations and generating alerts based on the anomalies detected (i.e., number of posts per topic, day and location higher than the expected based on previous X days).</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | R functions from episomer<br>Episomer maintainers | N/A | <p>Records:</p> <ul style="list-style-type: none"> <li>• Episomer alerts and alert emails</li> </ul>                                                                                      | <ul style="list-style-type: none"> <li>✓ ECDC NORMAL</li> <li>ECDC</li> <li>SENSITIVE</li> <li>PUBLIC</li> </ul> |

|                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                        |     |                                                                                                                                                                                                                                                                                                                                                                               |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <p>The outputs of this step are visualisations (trendlines; bar charts with hashtags, most frequent words and URLs, and maps at the country level), alerts and emails containing alerts.</p> <p>The visualisations and alerts are available in the dashboard using VPN, and the alert emails are available in the 24/7 duty email and emails of ECDC EI team members.</p> |                                                                                                        |     | <ul style="list-style-type: none"> <li>• Episomer visualisations</li> </ul> <p>Record owner:<br/>Epidemic Intelligence</p> <p>Storage location:</p> <ul style="list-style-type: none"> <li>• ECDC-controlled environment (e.g., VM with two-factor authentication access)</li> <li>• dashboard available in ECDC browser via VPN</li> <li>• 24/7 and EI mailboxes.</li> </ul> |  |
| <b>Process Step 4: Operational validation</b>                                                                                                                                                                                                                                                                                                                             |                                                                                                        |     |                                                                                                                                                                                                                                                                                                                                                                               |  |
| <p>4.1 Anomalies detected by episomer are manually reviewed to validate the information. This is done checking the alerts sent by email or included in the dashboard and validating the accuracy and veracity of the information with official sources or ECDC stakeholders.</p>                                                                                          | <ul style="list-style-type: none"> <li>• 24/7 duty officer</li> <li>• TDO and/or EI experts</li> </ul> | N/A | N/A                                                                                                                                                                                                                                                                                                                                                                           |  |
| <b>Process Step 5: Ad hoc actions to improve models and data privacy</b>                                                                                                                                                                                                                                                                                                  |                                                                                                        |     |                                                                                                                                                                                                                                                                                                                                                                               |  |

|                                                                                                                                                                                                                                                                                               |                                                                                                                      |                                     |                                                                                                                                                 |                                                                                                                    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| There are three sub-steps which are performed on ad hoc basis to improve models and data privacy in episomer. These sub-steps are independent from each other and do not have to be performed in a specific order:                                                                            |                                                                                                                      |                                     |                                                                                                                                                 |                                                                                                                    |
| 5.1 User-specific deletion or anonymisation. This sub-step deletes or anonymises data from specific users on request. It uses episomer functions and applies these to the data stored in the ECDC-controlled environment                                                                      | <ul style="list-style-type: none"> <li>• EI experts,</li> <li>• Episomer maintainers</li> <li>• HoS GEIHS</li> </ul> | N/A                                 | N/A                                                                                                                                             |                                                                                                                    |
| 5.2 Improvement of geolocation model in episomer. In this sub-step, a selected list of posts text is annotated to review the text geolocation assigned by the episomer model. It uses an Excel file downloaded from episomer in the ECDC-controlled environment, manually annotated.          | Episomer maintainers                                                                                                 | Excel file downloaded from episomer | Record:<br>Excel files with a selected list of posts text (5.2) or detected alerts (5.3)                                                        | <input type="checkbox"/> ECDC NORMAL<br><input type="checkbox"/> ECDC SENSITIVE<br><input type="checkbox"/> PUBLIC |
| 5.3 Improvement of alert model in episomer. In this sub-step, a selected list of alerts detected by episomer are annotated to categorise them (e.g., signal, event, threat or discard). It uses an Excel file downloaded from episomer in the ECDC-controlled environment, manually annotated | Episomer maintainers                                                                                                 | Excel file downloaded from episomer | Record owner:<br>Epidemic Intelligence<br><br>Storage location:<br>ECDC-controlled environment (e.g., VM with two-factor authentication access) |                                                                                                                    |

## 3 Segregation of Duties and Control Activities

### **Segregation of duties**

The following measures are in place to ensure segregation of duties and control activities:

- 1) The 24/7 duty team is composed of two experts, at least one being from the epidemic intelligence group. Both experts keep a close contact and frequent discussions during each of these steps
- 2) There are at least yearly revisions of the episomer package to ensure its main functionalities. In addition, being a public R repository, other users can post any issues they may find.
- 3) All EI experts have access to the ECDC-controlled environment and to the repository to improve or solve issues in the code and tool.

### **Control Activities**

All information collated from social media platforms is validated and confirmed with other sources, especially when it is from unofficial social media accounts, following the EI steps. No episomer data is directly used or included in any ECDC output.