Epitweetr: development and implementation of an AI tool to monitor Twitter trends for early warning of public health threats

Laura Espinosa, DVM MPH MSc MScRes(Agr)

Expert Epidemic Intelligence, Surveillance section, European CDC

Columbia, 10 February 2023

Outline

Background and what was the issue?

Spiral process of social innovation for developing epitweetr project

1. Challenges and opportunities: Why Twitter?

2. Proof-of-concept study: Initial non-automatised prototype

3. Sustainability: Epitweetr R package and Shiny app first release, and integration in the ECDC Epidemic Intelligence processes

4. Scaling and systemic change: Evaluation of epitweetr, new releases and more users

Lessons learned

Background (I)

ECDC aims at strengthening Europe's defences against **infectious diseases.**

It is a **decentralised agency** of the **European Commission** founded in **2005**.
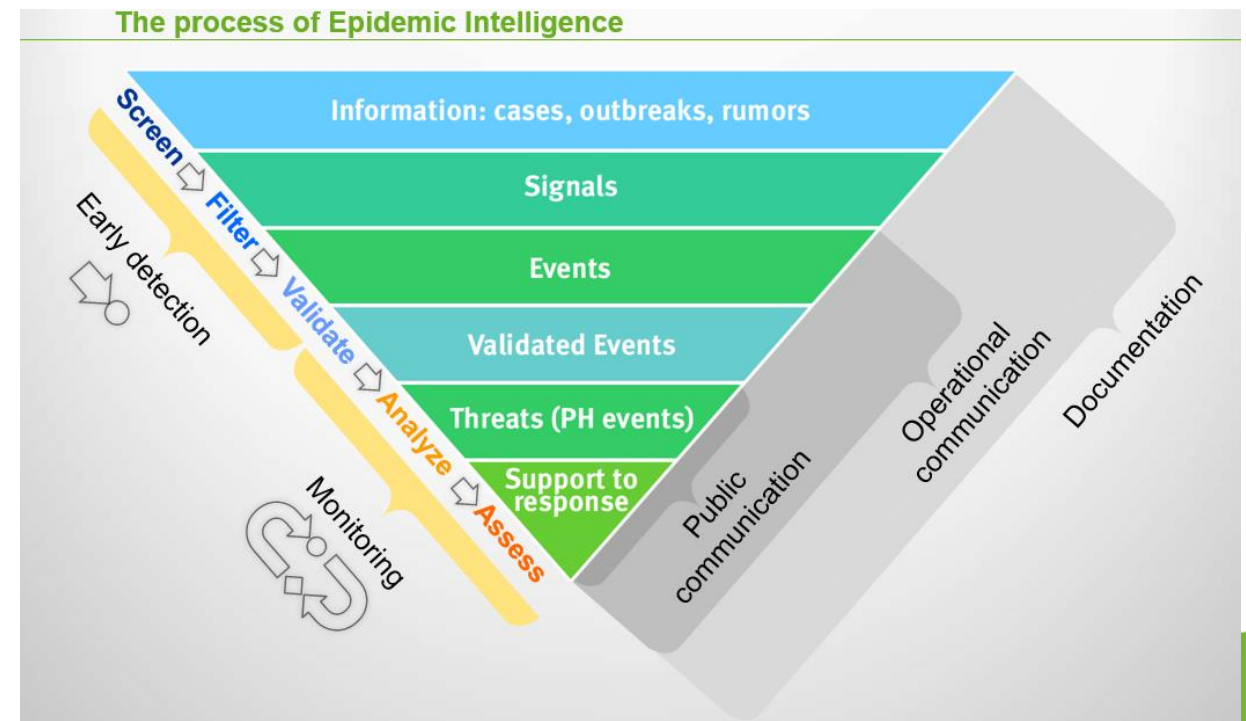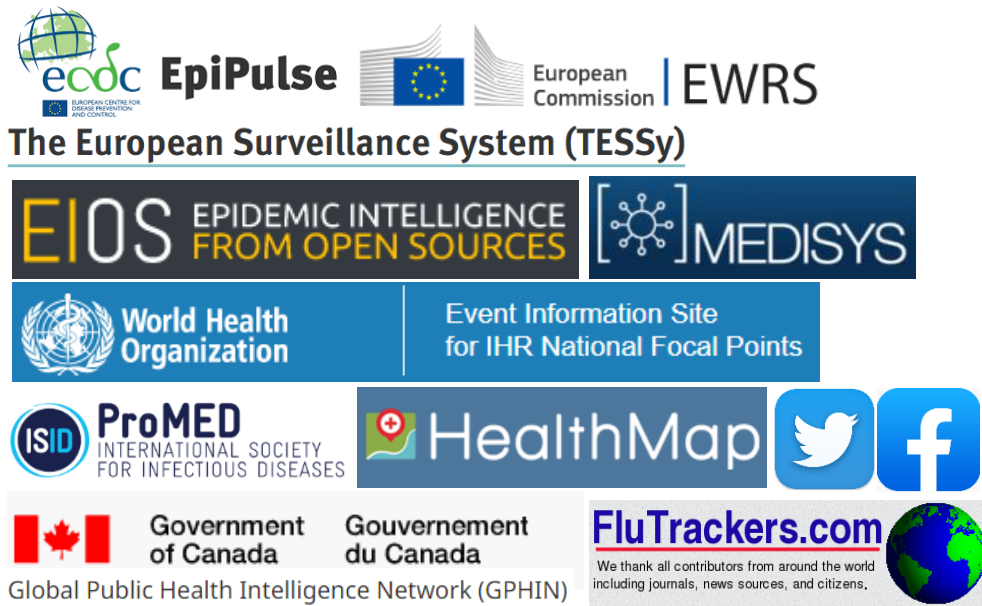
Epidemic intelligence (EI) is one of the **core functions** of the organisation.

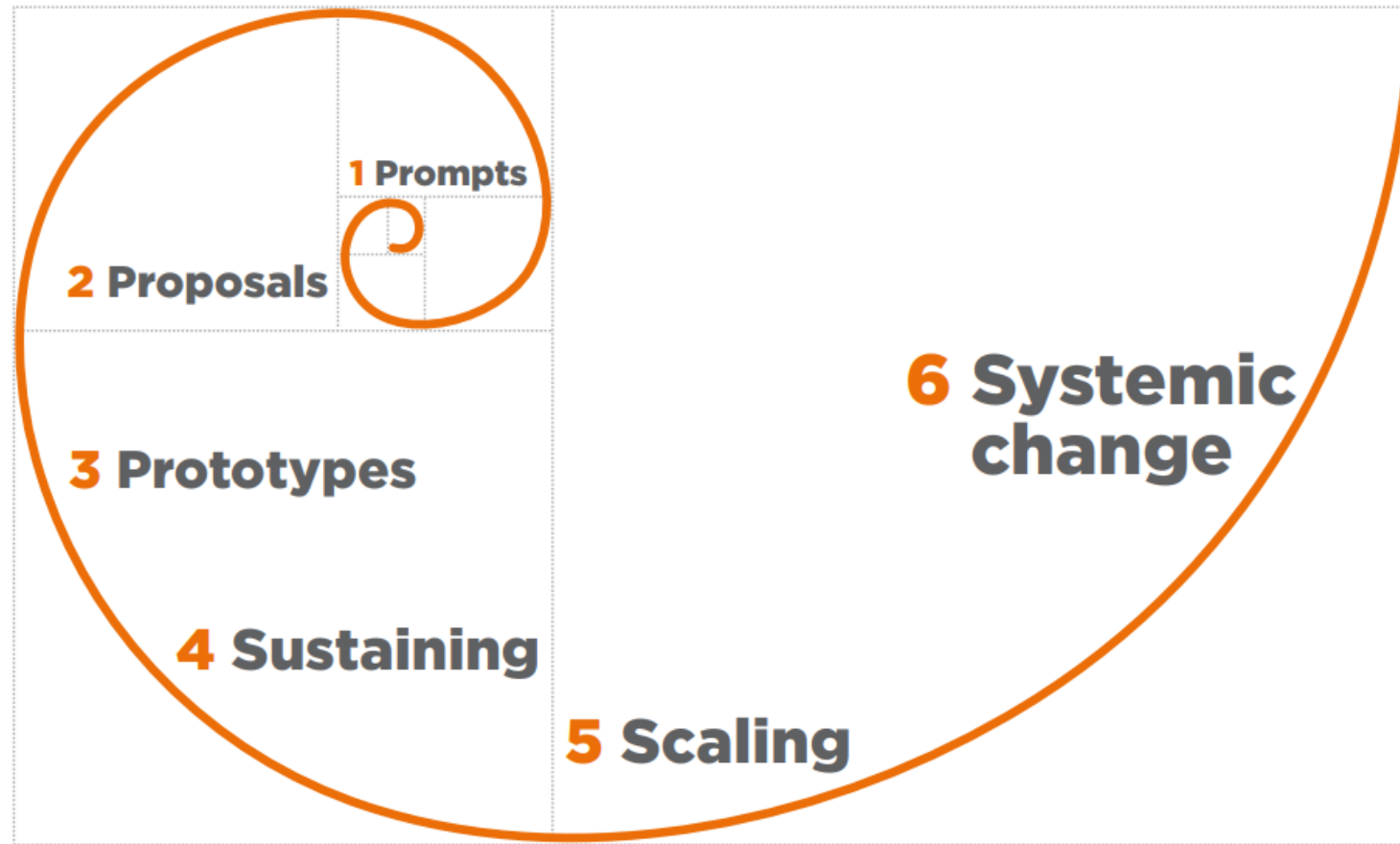At ECDC, there is a **24/7 EI duty team** comprising:

- 24/7 duty officer with weekly rotation
- EI support officer working during working hours with weekly rotation

Background (II)

Epidemic intelligence is the process of screening, filtering, validating, analysing and assessing potential public health threats.

Several sources are screened from **indicator-based surveillance** (e.g., The European Surveillance System) and **event-based surveillance** (official public sources, web aggregators, social media and restricted platforms).

The process of Epidemic Intelligence

4

What was the issue?

Previous studies have shown the **usefulness of social media** platforms for public health **surveillance** and real-time monitoring or rapid detection of **outbreaks**.

The **objective** of this project was to improve the **timeliness and effectiveness** of using social media, more specifically **Twitter**, for early detection of **public health threats**.

Spiral process of social innovation for developing epitweetr project

1 Prompts

2 Proposals

3 Prototypes

4 Sustaining

5 Scaling

6 Systemic change

# 1. Challenges and opportunities: Why Twitter?

Challenges

Lack of configurable tool for social media trend monitoring for early warning

Time and human resources required for manual monitoring of social media
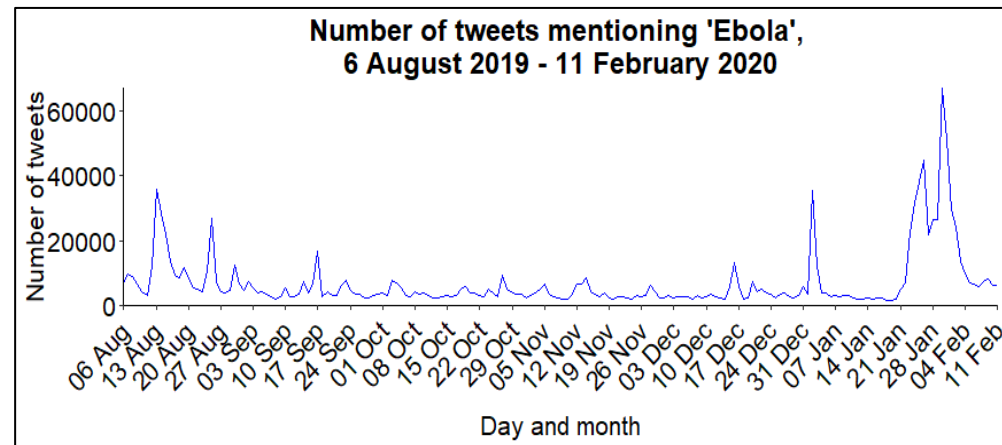
Publicly available data

Well documented API

R packages facilitating Twitter data collection

Opportunities Why Twitter?

2. Proof-of-concept study: initial non-automatised prototype (July 2019 – February 2020)

List of ECDC requirements

Measles
Ebola
COVID-19
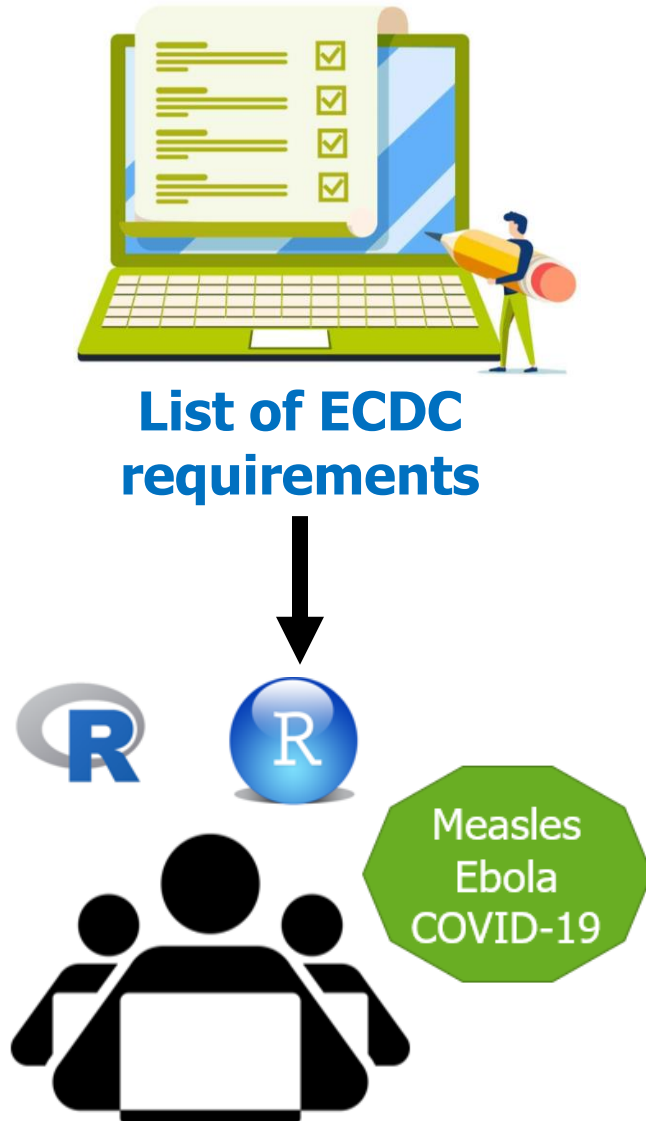
☑ Data collection, aggregation, and visualisation

☑ Email notifications

☒ Automation

☒ Signal detection algorithm (identified as key functionality)

Number of tweets mentioning 'Ebola', 6 August 2019 - 11 February 2020

Number of tweets

60000
40000
20000
0

06 Aug 13 Aug 20 Aug 27 Aug 03 Sep 10 Sep 17 Sep 24 Sep 01 Oct 08 Oct 15 Oct 22 Oct 29 Oct 05 Nov 12 Nov 19 Nov 26 Nov 03 Dec 10 Dec 17 Dec 24 Dec 31 Dec 07 Jan 14 Jan 21 Jan 28 Jan 04 Feb 11 Feb

Day and month

3. Sustainability: Epitweetr R package and Shiny app first release in October 2020
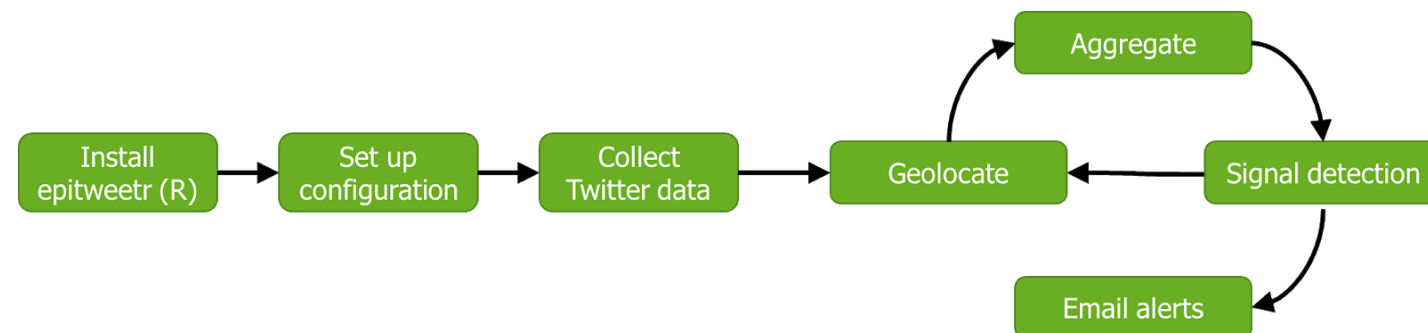

Multidisciplinary team

Expert workshop (June 2020)

With 4 potential users

- ✓ Twitter API 1.1
- ✓ 70 topic and 4 languages (up to 157 fasttext)
- ✓ Signal detection
- ✓ Machine learning models for geolocation (tweet & users)

3. Sustainability: Integration in the ECDC EI processes

- Rapid increase of epitweetr's use

- Integrated in the daily ECDC epidemic intelligence activities

- The ECDC EI team and the ECDC 24/7 duty officers have been trained on its use and maintenance

- Epitweetr has been presented in different fora

- Online trainings have been organised for public health experts

- Developing epitweetr as a free open-source R package enhanced its use outside of ECDC

4. Scaling and systemic change: Evaluation of epitweetr, new releases and more users

Research
🔓 **Open Access**

Epitweetr: Early warning of public health threats using Twitter data 💬 🔴 Check for updates

Laura Espinosa[1,*] iD, Ariana Wijermans[1,*], Francisco Orchard[2] iD, Michael Höhle[3] iD, Thomas Czernichow[2,4] iD, Pietro Coletti[5] iD, Lisa Hermans[5] iD, Christel Faes[5] iD, Esther Kissling[2] iD, Thomas Mollet[1,6] iD

The main objective of our study is to evaluate epitweetr version 1 published in October 2020, a new automated, open-source, R-based tool for early detection of public health threats using Twitter data. The specific objectives are to assess the performance of the geolocation and signal detection algorithms used by epitweetr and to assess the performance of epitweetr in comparison with the manual monitoring of Twitter for early detection of public health threats.

Epitweetr community

downloads 16K

Jan 5, 2022

lauespinosa

v2.0.3

c3185b1 ✓

Compare ▾

v2.0.3

New data architecture/storage, additional functionalities and machine learning layers for geolocation, signal categorisation and data protection.

EU-ECDC / epitweetr Public

<> Code ⊙ Issues 6 ⇄ Pull requests 💬 Discussions ⊙ Actions ⊞ Projects ⊙ Security ⌁ Insights

1. General
Welcome to epitweetr!
lauespinosa

2. Training material
Get to know epitweetr!
lauespinosa

4. Ideas and feedback
Survey on epitweetr users
lauespinosa

4. Ideas and feedback
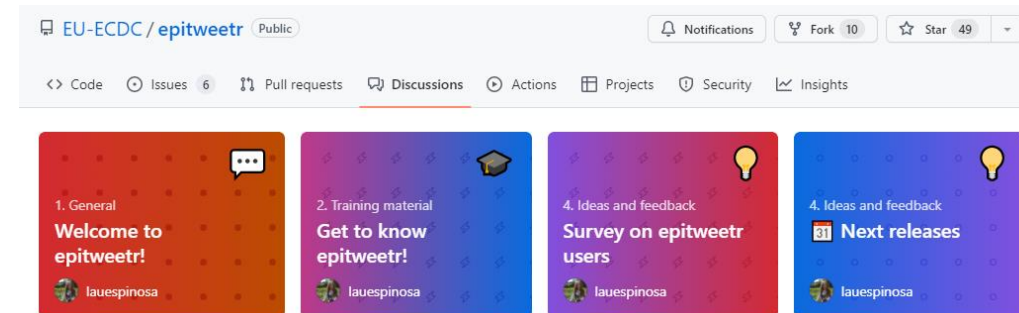**Next releases**
lauespinosa

epitweetr: Early Detection of Public Health Threats from 'Twitter' Data

It allows you to automatically monitor trends of tweets by time, place and topic aiming at detecting public health threats early through the detection of signals (e.g. an unusual increase in the number of tweets). It was designed to focus on infectious diseases, and it can be extended to all hazards or other fields of study by modifying the topics and keywords. More information is available in the 'epitweetr' peer-review publication (<https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2022.27.39.2200177>).

Version: 2.2.13

Lessons learned

❑ This project has shown the **importance** of following the steps in the **innovation spiral**, from developing a prototype through scaling and finally systemic change, to achieve a successful output.

❑ Having a **multidisciplinary team** behind its development and making epitweetr as an **open source tool**, has allowed for:

- Continuous **improvement**

- Increased **usability** of epitweetr in the public health community

❑ The increased use of the tool has increased the demands for support and further development (**sustainability**)

Acknowledgement

Epitweetr's development core team:

- Ariana Wijermans

- Francisco Orchard

- Michael Hoehle

ECDC EI former and current team members, especially:

- Gianfranco Spiteri

- Jon Bilbatua

- Thomas Mollet

- Hasan Tareq

Experts from Epiconcept and Hasselt University involved at some point in the project

Epitweetr beta testers and users

Thank you for your attention

Laura.Espinosa@ecdc.europa.eu