Leveraging social media data for public health threat detection: from manual monitoring to prototyping and escalating the automated monitoring

Laura Espinosa, DVM MPH MSc MScRes(Agr) PMP

Expert Epidemic Intelligence, Surveillance section, ECDC

Rencontres de Santé publique France, 14 February 2024

Declaration of interest

I declare no conflict of interest

Outline

Background: ECDC and epidemic intelligence

Social media and epidemic intelligence

Roadmap of activities

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

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

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

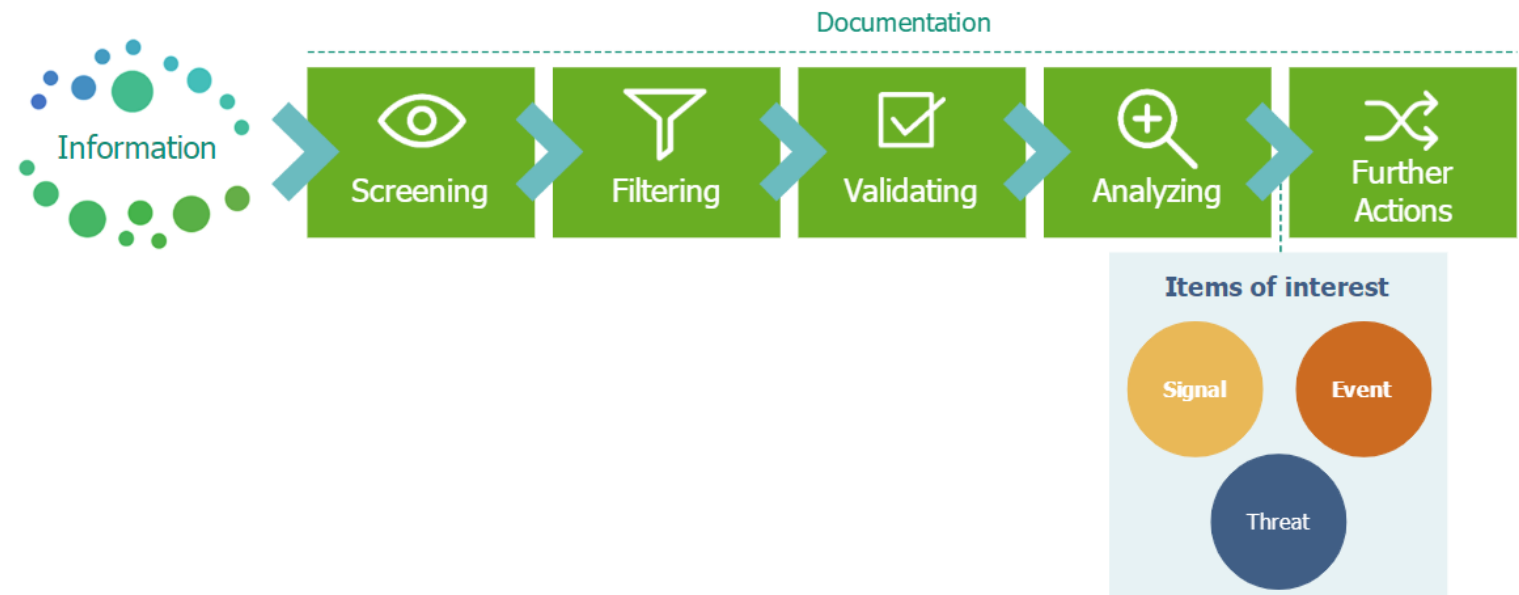4. Changes in data accessibility and how to more forward

Lessons learned

Background

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.

Documentation

Information → Screening → Filtering → Validating → Analyzing → Further Actions

Items of interest

Signal   Event

Threat

Social media and epidemic intelligence

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

ECDC developed a project in 2019 to improve the **timeliness and effectiveness** of using social media, more specifically **Twitter**, for early detection of **public health threats**.

Roadmap of activities

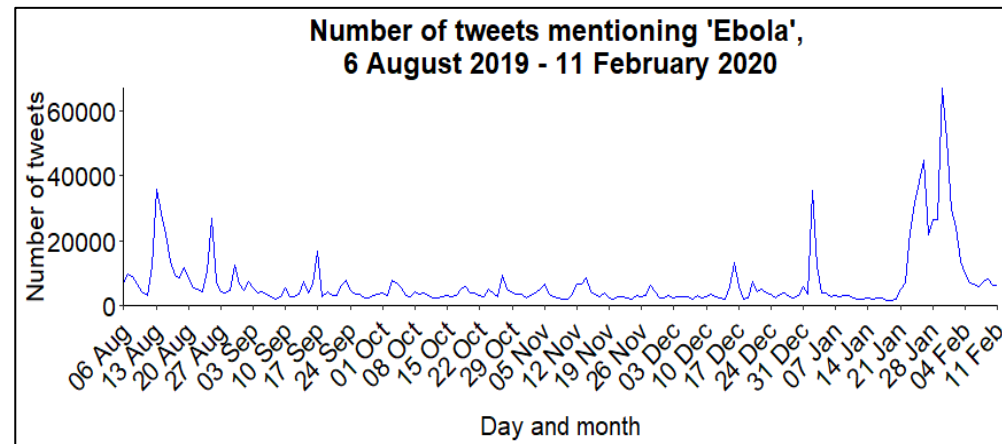| 2019	2020	2021-2022	2023	2024-2025
Prototype of R package	Development and launch of epitweetr	Use of epitweetr for baseline, ad hoc threats and mass gatherings monitoring	Retirement of epitweetr in April 2023 due to changes in data access levels	Review of data accessibility to several social media platforms and possible new tool

1. Proof-of-concept study: initial non-automatised prototype (July 2019 – February 2020)
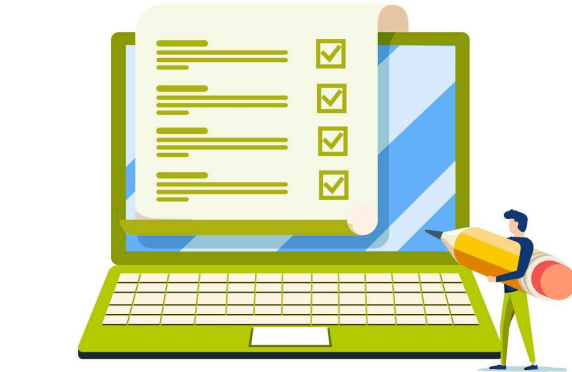
List of ECDC requirements

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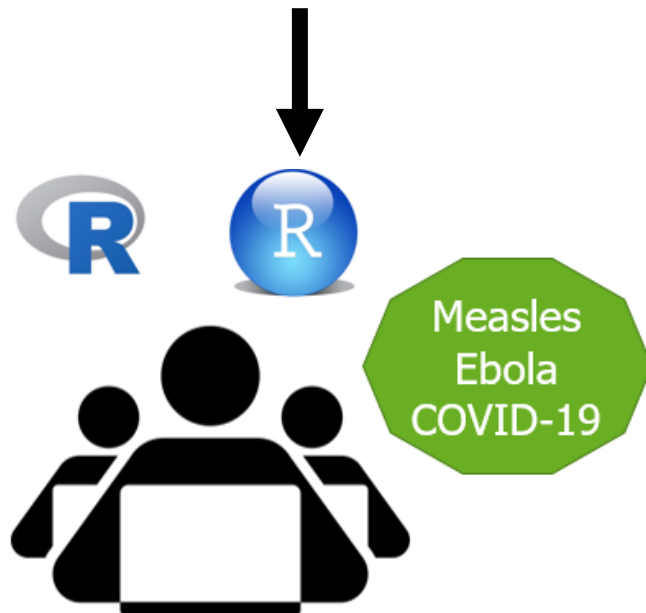
☑ Data collection, aggregation, and visualisation

☑ 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

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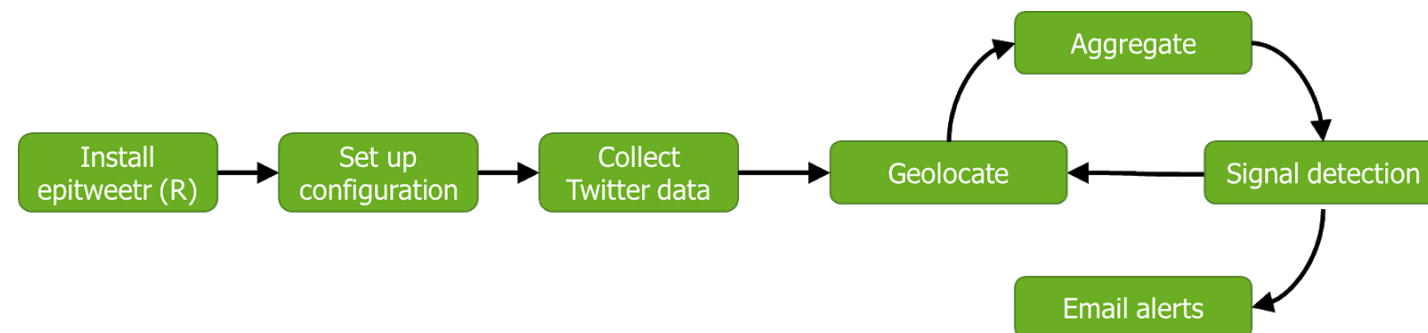

Multidisciplinary team

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)

2. 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

Sources: Video: The 24/7 responsibility - ECDC Epidemic Intelligence (europa.eu); Searching for infectious diseases with open source | Joinup (europa.eu); Usage of social media in epidemic intelligence activities in the WHO, Regional Office for the Eastern Mediterranean | BMJ Global Health

3. 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.

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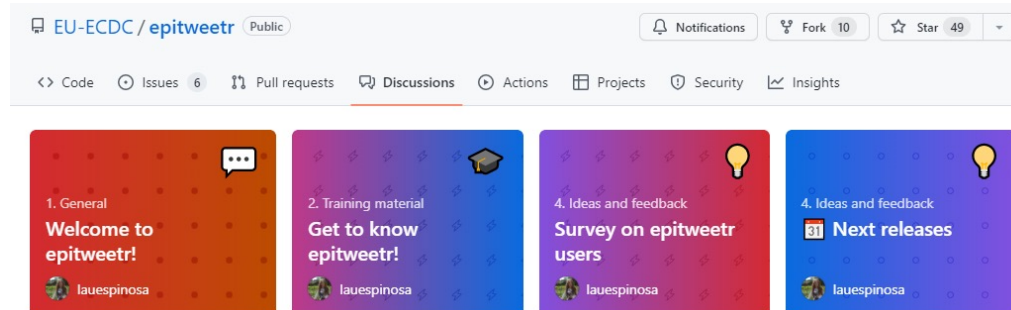
`epitweetr: Early Detection of Public Health Threats from 'Twitter' Data`

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

Epitweetr community

downloads 16K

🖥 EU-ECDC / **epitweetr** (Public)   🔔 Notifications   ⑂ Fork 10   ☆ Star 49 ⌄

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

1. General	2. Training material	4. Ideas and feedback	4. Ideas and feedback
💬	🎓	💡	💡
Welcome to epitweetr!	**Get to know epitweetr!**	**Survey on epitweetr users**	📅 **Next releases**
🧑 lauespinosa	🧑 lauespinosa	🧑 lauespinosa	🧑 lauespinosa

4. Changes in data accessibility and how to move forward

	Free	Basic	Pro	Enterprise
Getting access	Get Started	Get Started	Get Started	Get Started
Price	Free	$100/month	$5000/month	
Access to X API v2	✔ (Only Post creation)	✔	✔	
Access to standard v1.1	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	
Project limits	1 Project	1 Project	1 Project	
App limits	1 App per Project	2 Apps per Project	3 Apps per Project	
Post caps - Post	1,500	3,000	300,000	
Post caps - Pull	✖	10,000	1,000,000	
Filtered stream API	✖	✖	✔	
Access to full-archive search	✖	✖	✔	
Access to Ads API	✔	✔	✔	

Data accessibility of other social media platforms
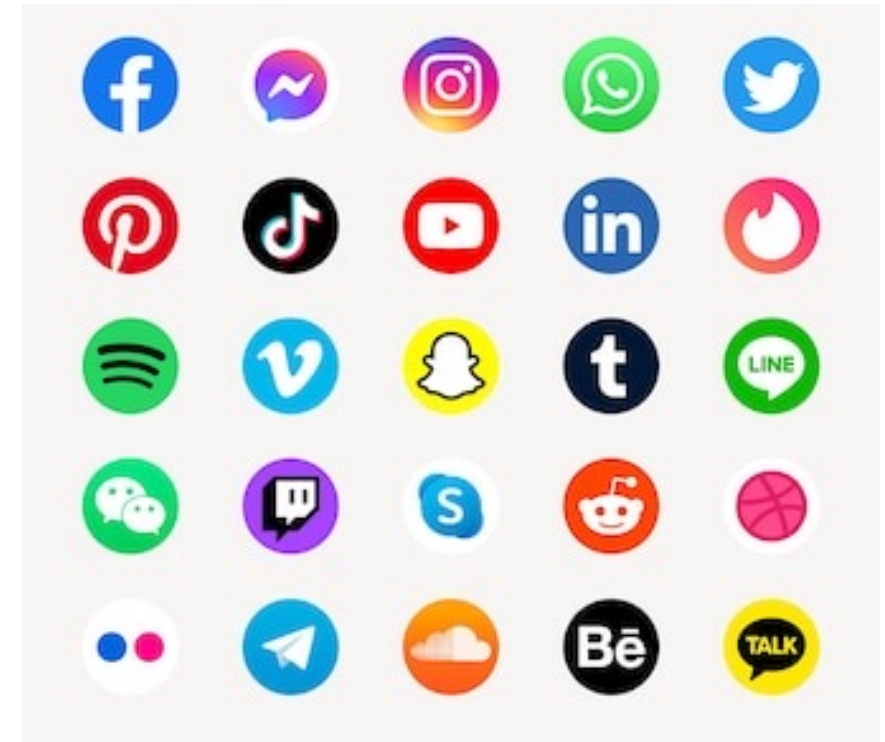
Lessons learned

❑ 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 increased the demands for support and further development (**sustainability**)

❑ Unexpected changes in these platforms may happen, be ready for finding other solutions, not the only solution

Thank you for your attention

Laura.Espinosa@ecdc.europa.eu