



# ENDIG: Interactive Geovisualization of Surveillance Systems for Notifiable Diseases in Europe

Nils B. Tjaden <sup>1</sup> and Justine I. Blanford <sup>1</sup>

<sup>1</sup> Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation–ITC, University of Twente, Enschede, Netherlands

Correspondence: Nils Tjaden ([n.b.tjaden@utwente.nl](mailto:n.b.tjaden@utwente.nl))

**Abstract.** National disease surveillance systems are among the most important tools for the observation and management of communicable diseases. They help to detect outbreaks and monitor the success of public health responses. The European Union aims to harmonize the European landscape of disease surveillance systems for better interoperability across the member states. However, the progress of these efforts is difficult to assess, as information is available in different formats, written reports or numerous tables for individual diseases. Here we present ENDIG, an interactive geovisualization tool, that makes the existing information easily accessible to end users and researchers in public health and adjoining disciplines. ENDIG allows for convenient exploration about the development of national disease surveillance systems for more than 60 diseases in the EU member states, as well as Iceland, Liechtenstein and Norway, over the course of the past 7 years (2015–2021).

**Keywords.** Disease surveillance, notifiable diseases, geovisualization, shiny

## 1 Introduction

### 1.1 Disease Surveillance Systems

Infectious diseases are a major threat to human and animal health. In particular, communicable diseases (i.e. infectious diseases that can be transmitted between hosts) such as bacterial or viral infections have the ability to spread beyond the geographical areas where they originally occurred. In contrast to many non-communicable diseases such as strokes, most cancers, or autoimmune diseases, communicable infectious diseases tend to show more dynamical patterns in space and time. In order to keep track of relevant communicable diseases within their territory, many countries define a set of

notifiable diseases that must be reported whenever they occur. Typically, a disease surveillance agency is in charge of gathering and curating these data.

Despite advances in digital surveillance based on internet search behaviour and social media (Aiello et al., 2020), one of the most important tools for observing communicable disease activity is still the classical disease surveillance system (Murray and Cohen, 2017). These systems provide the necessary infrastructure that enables health care providers and laboratories to report case records to the surveillance agency.

Depending on the disease they are monitoring, disease surveillance systems are typically used for *early warning* or *programme monitoring* purposes (WHO, 2006): For emerging or re-emerging communicable diseases, early detection of occurring cases may be crucial for preventing isolated cases from turning into local outbreaks or large scale epidemics. For endemic diseases, disease surveillance systems can be used to monitor the progress of control programs or the success of vaccination campaigns. For researchers, disease surveillance systems can also supply valuable data that can be used to train or validate models.

### 1.2 The Need for a Pan-European View

The COVID-19 pandemic has demonstrated the potential for communicable diseases to overcome political boundaries. Long before, the European Union (EU) already realized the need for a pan-European approach to disease surveillance. In 1998, the European Commission decided to set up a “network for the epidemiological surveillance and control of communicable diseases” (Decision No 2119/98/EC of the European Parliament and of the Council, 1998). A list of diseases to be covered was established in 2000 (Commission Decision 2000/96/EC, 1999) and updated in 2018 (Commission Implementing

Decision (EU) 2018/945, 2018). In 2022, a new regulation on “serious cross-border threats to health” was put in place (Regulation (EU) 2022/2371 of the European Parliament and of the Council, 2022), that picks up upon the lessons learnt from the COVID-19 pandemic and aims to improve data flows within the EU. The European Centre for Disease Prevention and Control (ECDC) serves as the central surveillance agency the national agencies report their data to. Anonymized disease surveillance data at the country level is shared with the public through ECDC’s Surveillance Atlas of Infectious Diseases (2023) and *Annual Epidemiological Reports* are being published for individual diseases.

What is missing, however, is an accessible way to assess the degree to which the individual member countries already fulfill the requirements set up by the EU. To fill that gap, we present the European Notifiable Diseases Interactive Geovisualization (ENDIG). ENDIG synthesizes information that is publicly available but cumbersome to handle into a series of maps and figures. Embedded in an interactive web application, these enable public health professionals, scientists as well as the general public to easily explore the spatial and temporal developments in the diverse landscape of European surveillance systems.

## 2 Data sources and processing

### 2.1 Data and Software Availability

Research data supporting this publication was accessed on January 25<sup>th</sup>, 2023. The computational workflow supporting this publication, including data download from external sources, is executed via a set of numbered R-scripts. Together with ENDIG itself, these are published under GNU General Public License v3.0 at <https://github.com/nbtjaden/ENDIG>.

### 2.2 Data on National Surveillance Systems

Alongside their written *Annual Epidemiological Reports* for individual diseases, ECDC publishes yearly *surveillance system overviews* – lists of national disease surveillance systems that report cases of certain diseases to ECDC (Introduction to the Annual Epidemiological Report, 2023). The selection of diseases covered in the *surveillance system overviews* is based on the official EU list of “communicable diseases and related special health issues to be covered by the epidemiological surveillance network” (hereafter: “diseases”) published in Annex 1 to Commission Implementing Decision (EU) 2018/945 (2018). At the time of download, this data described the development of 4 surveillance system characteristics for

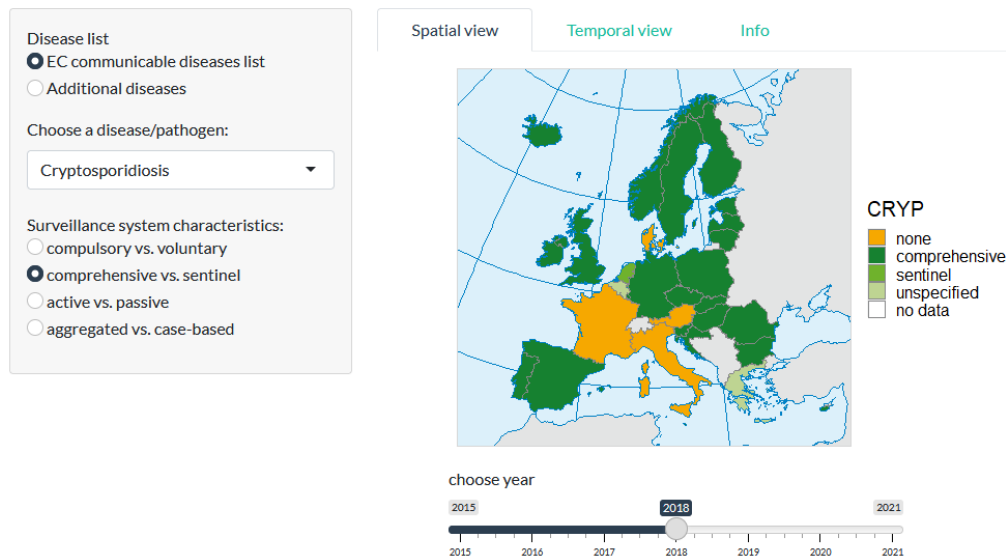
>60 diseases over 7 years, for the EU member states as well as Iceland, Liechtenstein and Norway.

*Surveillance system overview* files for all available years (2015–2021) were downloaded from their respective pages (e.g. <https://www.ecdc.europa.eu/en/publications-data/surveillance-systems-overview-2020> for the 2020 data; see data processing script for a full list). These semi-standardized Office Open XML Workbook files (\*.xlsx) contain one table sheet per disease. Each sheet then lists those countries that report surveillance information for the corresponding disease to ECDC. The lists contain information about the types of surveillance systems each country uses to gather the information, and give a reference to the formal case definition each country uses for the respective disease. Surveillance systems are classified according to four pre-defined characteristics (compare Murray and Cohen, 2017):

- **Compulsory vs. voluntary** reporting;
- **Active vs. passive** data collection, where in an active system the national surveillance agency is responsible for collecting the data from healthcare providers, and in a passive one the healthcare providers are responsible for reporting to the surveillance agency;
- **Comprehensive vs. sentinel-based** surveillance, where either all or only a representative sample of healthcare providers supply data, respectively;
- **Case-based vs. aggregated** reporting, where either full (anonymized) case data or total number of cases are reported, respectively.

The data was cleaned and restructured in R 4.2.2 (R Core Team, 2022) using the packages *xlsx* version 0.6.5 (Dragulescu and Arendt, 2020) for reading \*.xlsx files, *stringr* version 1.4.1 (Wickham, 2022) for processing character strings as well as *dplyr* version 1.0.10 (Wickham et al., 2022) and *tidyr* version 1.2.1 (Wickham and Girlich, 2022) for general data processing. Cleaning involved routine tasks such as harmonizing file, disease, class, and country names, as well as scrubbing footnotes and other non-data content from the tables. The classes “Other” and “Not specified/unknown” (often encoded as “.”) were merged into a single “unspecified” class for easier visualization.

# Disease surveillance systems in Europe



**Figure 1.** The interactive web application in *Spatial view*, showing a map for comprehensive vs. sentinel-based disease surveillance systems for Cryptosporidiosis in 2019.

Internal inconsistencies in the 2015 and 2016 files additionally required separate processing for Hepatitis B and C (which have an additional column for national coverage), as well as HIV/AIDS (which follows a completely different layout).

There were several instances where a disease's table sheet contained multiple entries for a single reporting country. These were automatically merged into a single entry per country, keeping classifications intact if all entries contained the same value. Otherwise, the classification value was set to "unspecified", as it was impossible to determine a single "correct" value.

Existing data for diseases not currently on the official EU list of relevant diseases (Commission Implementing Decision (EU) 2018/945, 2018) was kept separately (n=10). This includes discontinued systems for influenza but also additional surveillance of diseases such as hantavirus infections or Ebola.

## 2.3 Spatial Data

Spatial polygon data of country outlines was acquired from the Natural Earth data set. We extracted data supplied through the *rnaturalearthdata* R-package version 0.1.0 (South, 2017), using *rnaturalearth* version 0.3.2 (Massicotte and South, 2023). Spatial data was handled and joined with disease surveillance system data for visualization purposes using *sf* version 1.0-9 (Pebesma, 2018).

## 3 Interactive Visualization

### 3.1 Technical Implementation

As plotting a map of each possible combination of diseases and surveillance system characteristics for 7 years of data would have resulted in more than 1600 individual figures, traditional visualization techniques were clearly inappropriate. To make the data more meaningful, an interactive approach using *shiny* was chosen. Shiny is a popular web application framework that runs inside R – either locally on the user's device, on a cloud service, or on a dedicated server (Fay et al., 2021). We used *shiny* version 1.7.3 (Chang et al., 2022) in combination with *ggplot2* version 3.3.6 (Wickham, 2016) and *sf* version 1.0-9 (Pebesma, 2018) to develop an interactive web application that enables the user to explore the diseases that are being monitored across Europe.

A working implementation of ENDIG is available at <https://nbtjaden.shinyapps.io/ENDIG>. The source code for the ENDIG, along with the R scripts necessary to download and process the underlying data is available at <https://github.com/nbtjaden/ENDIG>.

### 3.2 User Interface

The ENDIG interactive web application utilizes a simple layout (Fig. 1). In the sidebar, the user can choose a disease from a dropdown list. A set of radio buttons allows to switch between the four different surveillance system

## Disease surveillance systems in Europe



**Figure 2.** The interactive web application in *temporal view*, showing the establishment of surveillance systems for Zika virus disease after the 2015–2016 epidemic in the Americas (Ikejezie et al., 2017).

characteristics. The main panel allows to switch between three tabs for a spatial view, a temporal view, and additional information. The *Spatial view* tab (Fig. 1) shows a map of the reporting states, illustrating the chosen surveillance system characteristic for the selected disease. This allows for easy identification of individual countries and spatial patterns. Below, a slider allows to select which year of data should be shown. In the *Temporal view* tab (Fig. 2), a heatmap shows the temporal development throughout 2015–2021 for each reporting country. The *Info* tab holds information about the underlying data, links to the source code repository, and gives a short explanation of the four characteristics used to describe the surveillance systems.

## 4 Discussion

The ENDIG fills a gap between existing geovisualizations currently provided by ECDC. While interactive maps for the occurrence of diseases (Surveillance Atlas of Infectious Diseases, 2023) and static maps for the occurrence of certain disease vectors (Surveillance and disease data for disease vectors, 2023) in the EU exist, an easily accessible overview for disease surveillance systems is still missing.

ENDIG makes important information on European disease surveillance systems easily accessible in a way that neither the original tables nor the written Annual

Epidemiological Reports can provide. At a glance, public health professionals and policymakers as well as the general public can gain valuable insights about a number of health-related questions: Which diseases are being monitored and reported to ECDC by the different EU member countries? Which diseases are not yet being monitored/reported even though they should? Which countries report more/less diseases than others? How has the current state evolved since 2015? The maps showing compulsory reporting give an indication of which diseases are considered notifiable by law in the different member states. The remaining three characteristics may hint at how thoroughly individual surveillance systems can be expected to work.

It also shows that different kinds of diseases are being treated with different urgency. (Commission Implementing Decision (EU) 2018/945, 2018) adds three mosquito-borne diseases (chikungunya, dengue, and Zika) as well as the tick-borne Lyme neuroborreliosis. At that point, most member countries were already reporting the former three diseases, but even in 2020 less than half of them reported Lyme neuroborreliosis.

### 4.1 Challenges and Limitations

Quality and completeness of the underlying *surveillance system overview* files were the most significant hurdles to overcome during the creation of ENDIG. The original data files' layout and structure prioritize an aesthetically

pleasing appearance over data usage, and thus required considerable amounts of re-formatting. This was additionally hampered by structural inconsistencies both between and within the individual yearly files.

After data cleaning, the temporal view reveals a series of data gaps that otherwise might have gone unnoticed. For instance, only the 2017 *surveillance system overview* file contains data about measles and rubella surveillance, even though surveillance was carried out and reported to ECDC before and afterwards (Annual Epidemiological Reports for measles, 2023). Similarly, influenza surveillance has not been recorded in the files since 2016, despite existing surveillance for both seasonal and zoonotic influenza in later years (ECDC, 2021a, b). The temporal view for yellow fever reveals that in the 2015 file the bottom part of the alphabetically sorted table is missing. This demonstrates the ENDIG's usefulness beyond its originally intended scope – as a simple tool for data quality assessment.

It must be noted, that the underlying data only covers diseases that are being reported to ECDC. Additional surveillance systems for other diseases may exist in the individual EU member states, that are not linked to the EU system and thus not covered by ENDIG. For example, while the United Kingdom stopped reporting most diseases after leaving the EU, that does not mean that existing surveillance systems were shut down. Liechtenstein is a special case in that it shares its surveillance with neighbor Switzerland (both non-EU members), but still reports tuberculosis data to ECDC.

## 4.2 Outlook

Future iterations of ENDIG may feature a zoomable map for better visibility of smaller countries such as Liechtenstein, Luxembourg, or Malta. As the ECDC Copyright and Limited Reproduction Notices (2023) allow for their contents to be modified and re-distributed, adding a convenient download option for the cleaned data or subsets thereof would benefit users who do not want to run the cleaning scripts themselves. A more streamlined and automated data processing queue is currently under development. This will allow for easy updates once a year, as long as future *surveillance system overview* files follow the established data structure.

In principle, ENDIG is not limited to the data supplied by ECDC. Data from neighboring countries or subnational regions could be integrated if available. Given the EU's plans towards a OneHealth approach (Regulation (EU) 2022/2371 of the European Parliament and of the Council, 2022), it can be expected that the list of disease surveillance systems will soon be extended to animal diseases. Ideally, data of actual disease occurrence would be integrated as well, although that would require the

European authorities to re-think their open-access strategy and make such data available in a public database.

In addition to these technical considerations, we intend to conduct a usability survey to assess the needs of different target groups. For better accessibility, the central points of this publication will be summarized in the *Info* tab of the application.

## Acknowledgements

Author contributions: NT & JB conceived the paper; NT performed the analysis and developed the application, JB advised. Both authors wrote the paper.

Funding was provided by the Ingenuity grant “Geospatial for health education and research”.

## References

- Aiello, A. E., Renson, A., and Zivich, P.: Social media- and internet-based disease surveillance for public health, *Annu. Rev. Public Health*, 41, 101–118, <https://doi.org/10.1146/annurev-publhealth-040119-094402>, 2020.
- Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B.: shiny: Web Application Framework for R, 2022.
- Commission Decision 2000/96/EC: of 22 December 1999 on the communicable diseases to be progressively covered by the Community network under Decision No 2119/98/EC of the European Parliament and of the Council (notified under document number C(1999) 4015), *Official Journal L028*, 1999.
- Commission Implementing Decision (EU) 2018/945: of 22 June 2018 on the communicable diseases and related special health issues to be covered by epidemiological surveillance as well as relevant case definitions (Text with EEA relevance.), *Official Journal L170*, 2018.
- Decision No 2119/98/EC of the European Parliament and of the Council: of 24 September 1998 setting up a network for the epidemiological surveillance and control of communicable diseases in the Community, *Official Journal L268*, 1998.
- Dragulescu, A. and Arendt, C.: xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files, 2020.
- ECDC Copyright and Limited Reproduction Notices: <https://www.ecdc.europa.eu/en/copyright>, last access: 18 April 2023.

- Introduction to the Annual Epidemiological Report: <https://www.ecdc.europa.eu/en/surveillance-and-disease-data/annual-epidemiological-reports/introduction-annual>, last access: 13 February 2023.
- ECDC: Seasonal Influenza, in: Annual Epidemiological Report for 2020, European Centre for Disease Prevention and Control, Stockholm, 2021a.
- ECDC: Zoonotic Influenza, in: Annual Epidemiological Report for 2020, European Centre for Disease Prevention and Control, Stockholm, 2021b.
- Annual Epidemiological Reports for measles: <https://www.ecdc.europa.eu/en/measles/surveillance-and-disease-data/aer-measles>, last access: 14 February 2023.
- Surveillance and disease data for disease vectors: <https://www.ecdc.europa.eu/en/disease-vectors/surveillance-and-disease-data>, last access: 17 February 2023.
- Surveillance Atlas of Infectious Diseases: <https://atlas.ecdc.europa.eu/public/index.aspx?Dataset=27&HealthTopic=11>, last access: 17 February 2023.
- Fay, C., Rochette, S., Guyader, V., and Cervan, G.: Deploy Your Application, in: Engineering Production-Grade Shiny Apps: A Workflow, CRC Press, Boca Raton, 2021.
- Ikejezie, J., Shapiro, C. N., Kim, J., Chiu, M., Almiron, M., Ugarte, C., Espinal, M. A., and Aldighieri, S.: Zika Virus Transmission—Region of the Americas, May 15, 2015–December 15, 2016, *Am. J. Transplant.*, 17, 1681–1686, <https://doi.org/10.1111/ajt.14333>, 2017.
- Massicotte, P. and South, A.: *naturalearth*: World Map Data from Natural Earth, 2023.
- Murray, J. and Cohen, A. L.: Infectious Disease Surveillance, in: *International Encyclopedia of Public Health*, vol. 4, Elsevier, Kidlington, Oxford, 222–229, 2017.
- Pebesma, E.: Simple Features for R: Standardized Support for Spatial Vector Data, *R J.*, 10, 439–446, <https://doi.org/10.32614/RJ-2018-009>, 2018.
- R Core Team: *R: A Language and Environment for Statistical Computing*, 2022.
- 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), *Official Journal L314*, 2022.
- South, A.: *naturalearthdata*: World Vector Map Data from Natural Earth Used in “*naturalearth*,” 2017.
- WHO: *Communicable disease surveillance and response systems*, 2006.
- Wickham, H.: *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York, 2016.
- Wickham, H.: *stringr: Simple, Consistent Wrappers for Common String Operations*, 2022.
- Wickham, H. and Girlich, M.: *tidyr: Tidy Messy Data*, 2022.
- Wickham, H., François, R., Henry, L., and Müller, K.: *dplyr: A Grammar of Data Manipulation*, 2022.