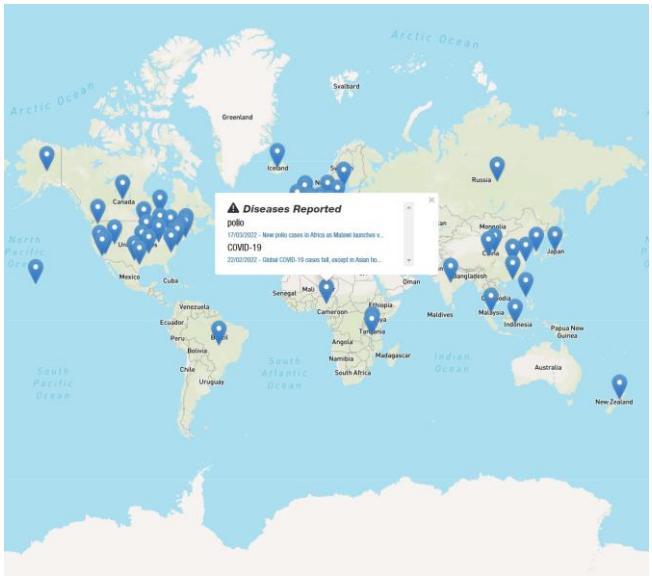

BUSINESS IDEA AND REQUIREMENTS

Deliverable 4

Team QQ

SENG3011

1 CONTENTS

2 INTRODUCTION

The software system which our team created is a user-friendly epidemic information centre called Viralog. This project consisted of 2 separate stages, each of which specified a set of requirements for the system to meet. The first of these stages was to develop a public API to access disease articles and reports. Our team was given the CIDRAP data source to scrape from. The second stage was to utilise this API (and APIs created by other teams using other sources) to develop a platform which presents information about the articles/reports in a meaningful way. For the platform, our team decided to create a website which uses visualisations such as maps and graphs to inform users about current diseases around the world.

Our inspiration for Viralog came from the website which we were introduced to at the very beginning of the course; Epiwatch. We used Epiwatch as our inspiration because we as a group saw massive potential in it, how it uses open-source data such as news reports and social media to quickly detect and identify disease outbreaks, often before they are even known about by health authorities.

However, we have also identified some fundamental problems with Epiwatch which we believe will stop it from becoming successful. Epiwatch is aimed at scientists, however for the average user it is unintuitive and poorly designed. For example, in order to find the website's main dashboard, the user must navigate halfway down the landing page. Additionally, this dashboard takes an extremely long time to populate (sometimes up to 3 minutes). Once the map loads, the user gets a view of a map of the world with numbers scattered all over it, but there is no key explaining what the numbers mean, which for many users is very confusing. The countries on the map aren't labelled which makes it harder to navigate. There is also no functionality to learn more about any particular disease, and since the Epiwatch diseases use their scientific names, many non-technical users won't know what the diseases mean. All of these factors contribute to a poor user experience for users who aren't scientists or topic experts.

So, while we did identify the enormous potential that Epiwatch has to save millions of lives, we just didn't think it could be successful in achieving that goal because it's not user friendly enough to gain traction with the general public. For that reason, our service, Viralog, aims to be an improved version of Epiwatch, differentiating itself from Epiwatch through being much more user friendly and easier to use. The average user without a technical background can go to our site and get simple and digestible information about diseases around the world. However, as will be discussed throughout this report, Viralog is also so much more than just an improved version of Epiwatch, we built our platform with several other use cases and requirements in mind. Also discussed in this report will be the requirements and use cases for the backend public API that we created earlier in the term, which serves information to the Viralog web application.

3 VIRALOG

3.1 REQUIREMENTS

3.1.1 Implemented

3.1.1.1 Interactive Map of Disease Reports

The interactive map (see *Figure 1*) is the first page which users will see when visiting Viralog. This map shows recent disease reports around the world. Users are able to scroll around on this map, zoom into certain locations and click on the blue pins to get the report information. We made use of the react-leaflet library to create this interactive map, and the Google Maps Geocoding API to get the pin locations. Each pin on the map corresponds to a location where at least one article exists which contains a disease report in that given location. As visible in *Figure 1*, many countries such as The United States, China or New Zealand have a corresponding blue pin which means from the sources we are scraping from (CIDRAP, Team Viral API), there exists articles reporting on diseases within those countries. On the other hand, we have not yet come across any articles which report on diseases in South America.

Clicking on these pins will provide users with an information box (see *Figure 2*), containing various pieces of information:

- The precise location of the pin (e.g., *New Zealand*)
- A list of diseases which have been reported within the given location the pin is situated at. Each disease is hyperlinked and will navigate users to the corresponding disease page (see 3.1.1.2 for more information)
- Under each disease is a list of all articles we have scraped for that given disease in location the pin is situated at. Each article is hyperlinked and clicking an article will navigate the user to the corresponding article page (see 3.1.1.3 for more information)

We decided to use an interactive map as the home page of the site as it hooks users with a visually pleasing and familiar representation of disease reports. The interactive pins allow users to access a list of reports for their location of interest easily and quickly, simply navigating to an area and clicking on a pin. Having the map page as the home page grabs the user's attention and encourages them to explore the rest of the site. In contrast, EpiWatch greets users with a visually unpleasing horde of text with no clear access to the site's functionality. This occlusion of features and lack of user friendliness makes users significantly more likely to dismiss the site and exit.

Ultimately, the interactive map adds to the overall value of our project as it enhances the data visualisation and user experience in navigating through large numbers of disease reports over multiple locations and time periods.

3.1.1.2 Article Information Page

An article information page exists for every article which we have scraped from our sources. An example article page can be seen in *Figure 3,* which is about Lassa Fever in the United Kingdom and Northern Ireland. The headline on the article page is hyperlinked to the originally published article we have based our page on, so interested users can find out more information about the given article (such as reading the body of the article). As pointed out in *Figure 3*, our page displays the publication date, category and source of the article, and the list of reports which were found within the article.

Instead of navigating to an article page through clicking a pin, users can also be displayed with a list of articles sorted by publication date through clicking the 'Articles' button on the taskbar (see *Figure 4*). This offers users an alternative method for finding articles; perhaps they didn't want to view an article about a specific disease but more-so recently published articles.

The purpose of having article pages is so users can find out more information about a reported disease by clicking on the hyperlinked headline. The other information displayed on this page provides users with a brief overview of the article, which may be sufficient to users who aren't interested in the specifics of the article. It is important to note that this is not a core feature of our project, however we decided to have dedicated article pages to make the website more complete.

3.1.1.3 Disease Information Page

A disease information page exists for every disease listed on the initial project specification. This page contains important and useful information about each disease such as:

Figure 5 Features

- The aliases (which can also be used for searching, see section 2.1.6 for more information)
- The symptoms
- The ability to 'watch' a given disease (see section 2.1.4 for more information)
- A comprehensive risk analysis, which provides users with the number of reports about this disease in the past 90 days, how many total reports have been made about this disease and the suspected risk of this disease based on the recent report count.

Figure 6 Features

- A report frequency graph which displays the number of reports for a disease over a period of time. Users can modify the period which will determine if reports about this disease are considered in the frequency graph or not.

Figure 7 Features

- A report map which only shows pins in locations where articles about the given disease were written about.

Figure 8 Features

- A list of recent reports about this disease, sorted in reverse chronological order.

Instead of navigating to a disease page through clicking a pin, users can also be displayed with a list of diseases sorted by report count through clicking the 'Diseases' button on the taskbar (*See Figure 9*). In addition to this, we added some advanced filters such as disease names, symptoms and risk level to provide the users with a sense of control over how the disease list is displayed; perhaps they only care about diseases with a given symptom having a risk level of high.

The overall purpose of the disease page is to give users an insight into the activity of each disease and provide them with some important information such as aliases, symptoms, and report frequency.

3.1.1.4 Dashboard & Watching Diseases

As a group, we believed that it would be useful to have a single page which acts as a summary of important information which is why we created the Dashboard (*See Figure 10-13*). The purpose of this dashboard is to allow users to easily view disease related information such as the case frequency and

risk analysis of given diseases, without having to thoroughly navigate the website. In addition to this, we made the dashboard customisable so users can modify it to fit their needs.

The first thing users see when navigating to the dashboard page is a graph of active diseases over the past 90 days (*See Figure 10*). The purpose of providing users with this knowledge is to give them a general idea of which diseases have been active in the current time, and which ones may increase in the near future. There is also an option to change the scale type from Logarithmic Scale to Linear Scale so different perspectives of the disease frequency can be seen allowing for more detailed comparison.

In *Figure 11*, users have the ability to watch diseases, which is a fundamental feature in making this dashboard customisable. Selecting any disease such as 'Ebola' or 'Polio' will provide the user with a risk analysis and report frequency graph directly underneath the list of diseases, which is ordered so the most active diseases appear first. An example of this can be seen in *Figure 12*. The idea behind watching diseases isn't to provide the user with new information unavailable elsewhere, but to concentrate important information in a single location so users aren't required to navigate between multiple pages to access the information they are looking for. Another thing to mention is that user preferences (such as diseases watched) are saved using cookies, removing the need to re-customise the dashboard upon every visit.

The final feature of the dashboard is a list of active diseases (*See Figure 13*), where users can specify the minimum number of reports over a given period from the present day.

3.1.1.5 Disease data visualisations

As previously mentioned in the disease page and dashboard page (*See Figure 6 and 13)*, data visualisations in the form of report frequency graphs are included to provide users with an easily digestible form of data. These visualisations are also modifiable in changing the time scale in which reports are shown for, to give users a better idea of the reported activity of diseases over different periods of time. We used the Chart.js library to render these graphs.

3.1.1.6 Fuzzy Search

The search bar (*See Figure 14*) located in the middle of the taskbar can be used to search for disease names, aliases, or symptoms. We employed a fuzzy searching library called Fuse.js so that if anything searched for isn't an exact match but is closely related, it will still appear in the results if it similar enough to any disease name or alias. This allows for any incorrect spellings of diseases or symptoms to be accounted for. Searching for symptoms is especially effective as users who have a certain symptom can simply use the search bar to learn about possible diseases they may have.

3.1.2 Not yet Implemented

These are features which we would aim to implement given more time.

3.1.2.1 Filter disease information by location

This feature would involve adding another filtering option for each report frequency graph to not only filter by time period, but also by location. This would be effective in allowing users to see the report frequency for a given country or location, which is often in their best interests.

3.1.2.2 Logging searches (anonymously)

This feature would involve logging all searches made using the search bar and storing the frequency of terms used. This data can then be used to understand trending symptoms or diseases which can ultimately be compared to the real data which we have scraped. Furthermore, knowledge of trending search terms can lead to future additions to the site such as personalised suggestions for articles to read.

3.1.2.3 Comparing data

This would involve adding functionality to compare data we have gathered on conditions such as:

- The report frequency between given diseases over a period of time. Users would have the option to specify a start and end date.
- The overall report frequency between two locations, with the option of specifying a disease to use for comparison.
- Comparing the most searched diseases with the most reported diseases.
- Comparing the most searched symptoms with currently active diseases.

Comparing data is beneficial as it will allow users to gain a better understanding of disease activity relative to one another, or to encourage users to reason about a correlation between search terms and disease activity.

3.1.2.4 Trend analysis for epidemic prediction

This feature would involve analysing report frequencies over recent periods in attempt to predict outbreaks or epidemics. One possible way to implement this could be to train machine learning algorithms with pre-existing data (such as a COVID-19 activity status dataset available on Kaggle) and using those algorithms with the data we have available on our website.

3.1.2.5 A designated country page

This feature would involve creating a designated page for each country, containing a summary of all disease information for that given country. Some example features include a report frequency graph for that country filterable by disease, time period or location within that country, a map visualisation of hot-spots where disease activity is highest and a list of the most frequently interacted-with articles for that country.

3.2 USE CASES

3.2.1 Use Case 1: Business owners - Visualising disease activity around the world through the map page

The map page is the first thing which appears when visiting Viralog, therefore it is important that this feature has a purpose for a variety of users. One group of users who can make use of the site is Business Owners, for example, the persona Mary. Mary is a café owner in the United Kingdom who is aware of how COVID may impact her business. She wants to keep up to date with new pandemics and diseases reported in the local area so she can be proactive in the case of an outbreak.

To address her problem, Mary will visit Viralog and navigate to the United Kingdom on the map page by scrolling around and zooming into her area of interest. Clicking on a respective pin will display a list of diseases and articles which have been reported for that given location, providing Mary with a vast amount of useful information about disease activity in her local area. Being aware of disease activity in the local area is important to Mary as she can enforce certain precautions early for business staff and customers if there appears to be an outbreak or large number of reports for the local area – allowing her to be proactive. In similar fashion, individuals can use the map to learn about disease activity in their local area, which will give useful knowledge about which areas to avoid visiting or provide them with information to predict areas which are likely to see an increase in reports. The whole interface was strategically made so users can access articles of interest in a short period of time, removing the need to learn the ins-and-outs of the website.

Mary or any other individual user is now aware that there may be diseases in the local area, but they want to find out additional information about the report made to ensure there is nothing serious to be concerned about. Referring to *Figure 2*, after clicking on the United Kingdom pin Mary is aware that COVID has been reported on in her local area but this is nothing of surprise. Scrolling down, she sees that Lassa Fever, a disease she has never heard of before has also been reported. Mary clicks on one of the articles under the Lassa Fever heading and gets taken to the Article Page (*See Figure 3*). The headline of the article is hyperlinked to the originally posted article, which is exactly what Mary can use to research further into the reports and make a judgement about Lassa Fever being a threat to her business operating – which was the main reason she visited Viralog to begin with.

3.2.2 Use Case 2: Power Users - Using the dashboard to gain deeper insight and keep track of disease activity

The inclusion of a customisable dashboard is aimed at power users, who we define to be more knowledgeable individuals about diseases such as scientists or university students, or people who want to stay updated on the reported frequency of certain diseases such as government officials. The dashboard provides users with a central hub of carefully selected information allowing power users to gain deeper insight into and keep track of disease activity all within a single page.

There are two sections on the dashboard which highlight active disease frequency; the active diseases graph (*See Figure 10*) and the active diseases list (*See Figure 13*). A disease expert would be interested in knowing the most reported-on diseases in the past 90 days, which is shown as a graph at the top of the dashboard page. Alternatively, if they want to modify the minimum report count over a specific time and be provided with a list of diseases which fit the criteria, they can use the active diseases list. The information obtained from these features is useful for the disease expert in staying up to date with active diseases, potentially influencing which diseases they focus their attention to. If on the active disease list, they set the minimum report count to 150 over the past 30 days and found that Lassa Fever or Cholera appears on the returned list, they will become aware of these diseases' activity and be able to find out more information by navigating to those disease pages and exploring the related articles or frequency graphs. In addition to disease experts, users such as government officials will learn about the most actively reported-on diseases, find out if it exists in their local area and be able to raise early awareness of diseases in the case it is active in the community.

The other prominent feature on the dashboard is the ability to watch diseases (*See Figure 11*). We will explore this through the persona of Mike. Mike is a university student studying medicine, has taken interest and wants to keep up to date with some of the less covered diseases despite the recent coverage of COVID-19. Mike first looks at the graph of active diseases and see's that 'polio' is the 6th most reported-on disease in the past 90 days, therefore deciding to 'watch' that disease by expanding the list of diseases and selecting 'polio'. Furthermore, he has heard of Malaria and Cholera, so he decides to watch those too. Note that watching diseases provides users with a risk analysis and report frequency graph for each disease on the dashboard page (*See Figure 12*), so Mike in this case wouldn't have to navigate to each respective disease page to view this information. This feature is useful to Mike as he can now regularly visit the site, navigate to the dashboard, and have access to the disease activity for these watched diseases. The graphs for watched diseases appearing underneath one another allows Mike to compare the frequencies and risk analysis of each disease in a suitable manner. The overall availability of this information will encourage motivated users like Mike to find trends or predict activity for each disease. If Mike decides he wants to watch any new diseases, in a situation where one of his university subjects focuses on Malaria, he can simply add Malaria to the watched diseases and stay up to date with its activity.

3.2.3 Use Case 3: Individuals - Using the universal search bar to quickly find information about diseases

All users, but especially individuals will find the search bar useful for quickly finding out information about diseases. We will explore this through the persona of Simon. Simon, an average Joe in his 40s who doesn't know too much about diseases has heard the word 'Rubella' on the radio. The topic of conversation was to do with diseases, so naturally he became interested and decided to go onto Viralog and use the search bar feature. He didn't use the correct spelling in his search (*See Figure 14 for an example*), but the fuzzy search will solve this problem and return the most relevant diseases. He is surprised that Measles is the first result, but quickly learns that Rubella is an alias for Measles. The functionality of being able to search by disease alias and support incorrect spelling caters to users like Simon who haven't got much knowledge about diseases, allowing them to use the website without worrying it is only for experienced users.

Simon's mother has been feeling unwell lately, coughing a lot and visibly losing weight. Concerned, Simon uses the search bar once again to search for these symptoms. Naturally, he searches for 'cough' first (*See Figure 14, Image 2*) and is provided with several diseases sorted by report count. Scanning through this list he sees that Tuberculosis also has a symptom for 'weight loss', which raises his suspicion that his mother may have this disease. Simon is also able to view more symptoms for Tuberculosis or continue to search for other symptoms using the search bar. Users in general will find the ability to search by symptom very useful as it is often the first indicator of carrying a disease and will be the first thing people search for on the internet. The ability to search by symptom has helped Simon learn about which diseases have 'cough' as a symptom, and hence has gained a stronger understanding of diseases to be aware of relating to his mother's health.

All searches return a list of diseases which best match the search. In the above examples, Simon already learned some basic information about Measles and Tuberculosis by simply looking at the list of diseases. However, he wants to learn more than just the symptoms and clicks on the diseases to navigate to the respective disease pages. Briefly returning to the story of Mary, she is interested in learning more about a Lassa Fever, so she searches 'Lassa Fever' and navigates to the disease page (*See Figures 5-8 for a visual representation of the Lassa Fever disease page*). For someone who knows nothing about Lassa Fever like Mary, she is provided with the aliases and list of symptoms which will prove useful in helping her know which symptoms to monitor for herself and business staff. Additionally, Mary uses the risk analysis as a quick judgement for Lassa Fever being a threat to her business in the local area, and also looks at the report frequency graph to check if there have been a spike in case numbers in the recent days by modifying the time scale. Now, Simon will use the disease page to find out the report frequency of Tuberculosis for recent days – if the number of reports is large and/or increasing, it acts as an indicator that he should look further into this disease as a potential cause for his mother's sickness. For any user, a list of reports about a given disease is shown at the bottom of the disease page sorted by publication date, which will provide them with further reading material to learn about the disease.

Ultimately, the search bar provides individuals with a way to quickly search for information about diseases as users specify disease names or symptoms as a search term, and then are returned with a list of relevant diseases all within a few seconds. Diseases returned will be clicked by the users, where they will learn useful and relevant information about the nature and activity of that disease.

4 PUBLIC API

The public API serves not only as the data source of our website but also a repository of information available to the public. This open-source public dataset advances the world of epidemic prediction by providing an easy way for scientists, businesses, researchers, or hobbyists to obtain disease reports from a variety of sources and use them for good. For example, a scientist or researcher could analyse report frequencies in given locations to predict epidemics, or a business could use the public data to educate the local community in becoming more aware of diseases in their area.

4.1 REQUIREMENTS

There are multiple requirements that have been implemented by our API.

4.1.1 Reports from CIDRAP

The API serves report information attained from CIDRAP (*cidrap.umn.edu*) articles. These articles are scanned for reports and relevant information such as article text, topic, and article publishing date is extracted for entry into the database.

4.1.2 Integration of other teams' APIs

Our API also serves information from other team's APIs, particularly data from the 'WHO' (World Health Organisation), from Team Viral. Rather than connecting the other APIs to Viralog independently of the CIDRAP scraper, they are treated as another independent data source and included in our database. This is to ensure that our API maintains control over the information that is being served and provides consistency between inconsistent services. It also provides additional information to users of the API itself, who would not receive this information if the other APIs were connected independently to Viralog.

4.1.3 Hosted Publicly

The API is not only available to our own services, but also available for use by the public. The database itself, however, is accessible only by the scraper and the API server. As such, the raw data can be hidden from the public and the interface to the API can be carefully controlled.

4.1.4 Users able to access and filter Articles

Article information is accessible via the API, and provides information such as the article headline, URL, main text, disease reports, authors, and source of the article. The article endpoint requires some parameters that allow filtering of articles. These parameters are:

- Two timestamps representing the start and end dates between which the article was posted
- Names of diseases relevant to the articles' reports
- The locations relevant to the articles' reports

Filtering is also optionally possible by data source, where results can be subset by the origin of the article.

4.1.5 Users able to access and filter Reports

Report information is accessible via the API and provides information such as the name of the disease, the symptoms associated with the disease, location (including longitude and latitude), as well as the corresponding article information, such as its name, headline, and source. The report endpoint similarly requires date, name, and location parameters, with an optional parameter available for the source of the data.

4.1.6 Well Documented

Documentation for the APIs endpoints are available for each route, providing information on parameter requirements, structure of results, and example responses. An interactive version of the documentation is available on Swagger, which provides a way for a user to sample use of the API via a web interface.

The documentation also defines conditions in which the API will fail. These conditions are also returned by the API response (with the error code) in order for the user to fix their erroneous query.

4.2 Use Cases

This section will discuss the potential uses for our API, as well as how our endpoints can be used to satisfy these use cases.

These use cases are divided into two categories: those that were designed with academic research in mind, and those designed with businesses or hobbyists in mind.

The first use case is targeted to academic researchers who may want to access a dataset for analysis or comparison with other data sources. The second category is aimed at businesses or hobbyists that may want to utilise our data in their own applications.

These use cases revolve around the idea that the API provides a standard format for queries and responses from our API.

4.2.1 Use Case 1 – Scientists or Researchers using the API to Advance their Research

Analysis of data is central to the role of scientists and researchers. As such, it is common for these users to attain their data from external sources rather than the generation of their own data. This could be useful as supplementary data for a report, or as a review of current technologies pertinent to disease and worldwide infection rates.

The analysis of data by Viralog is limited, largely due to our aim of creating a website that is simple as possible. There is potential for further processing of the data to generate new conclusions, create insightful data visualisations, or for further comparison with other data sources.

The articles endpoint could be used by researchers to determine from which sources reports are coming from. One possible application of this information is to determine if any news source is biased to a particular pattern or method of disease reporting.

The reports endpoint could be used for analysis of the data to determine which parts of the world are being affected by which diseases. This information could be used to target areas for aid, or to determine if a disease is becoming more or less prevalent.

The predictions endpoint could be used to monitor the current rate of disease predictions being made. If the rate is decreasing, this may be an indication that the disease is evolving and becoming more difficult to predict. This can especially be seen in ubiquitous disease such as COVID-19, where a high infection rate leads to the rapid evolution of the disease into new strains with differing symptoms.

4.2.2 Use Case 2 – Businesses or Hobbyists using the API to Create their own Application

As is exemplified by the nature of the second stage of this assignment, our API can be used by people who are developing their own applications for business purposes, or simply for the development of hobby applications that might utilise disease report data. It can be used to supplement existing services, or for the creating of new applications.

Viralog serves as an example of what is possible with the API, however there are a vast number of potential applications that could be developed using it.

The articles endpoint could be used to create a news application that collects data from many different sources to provide a single source for all the latest news on a particular disease.

The reports endpoint could be used to create an application that allows people to track the spread of a disease in real time, or to monitor the global situation of multiple diseases. One kind of application that was extensively explored by other groups was the use of disease data to inform users to make educated decisions when making travel plans.

The predictions endpoint could be used to create an application that monitors the spread of the virus and predicts the areas that will be most affected in the future. This could be used to warn people of potential dangers, or to target areas for aid.

5 APPENDIX

Figure 1: Map containing interactive pins for all articles we have scraped.

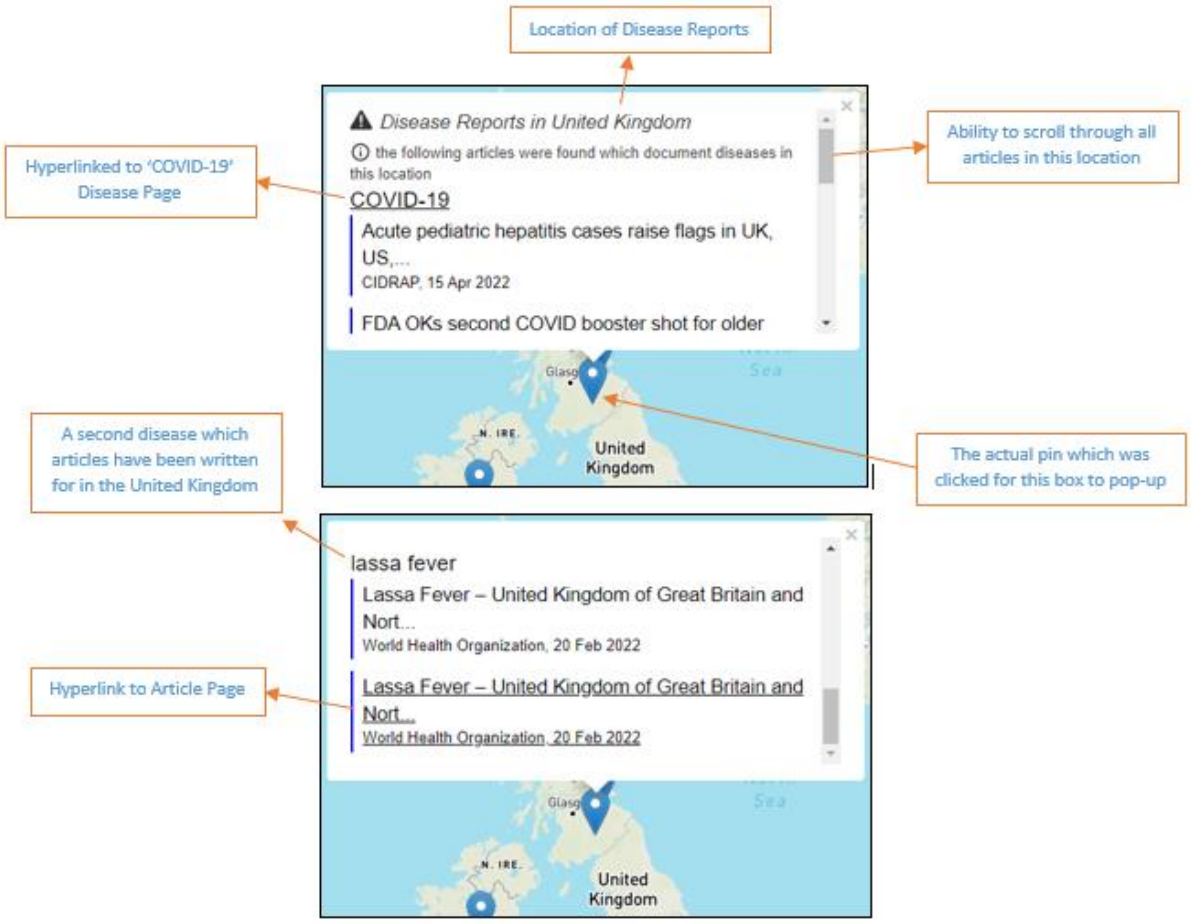

Figure 2: Pop-Up box when clicking a pin on the map.

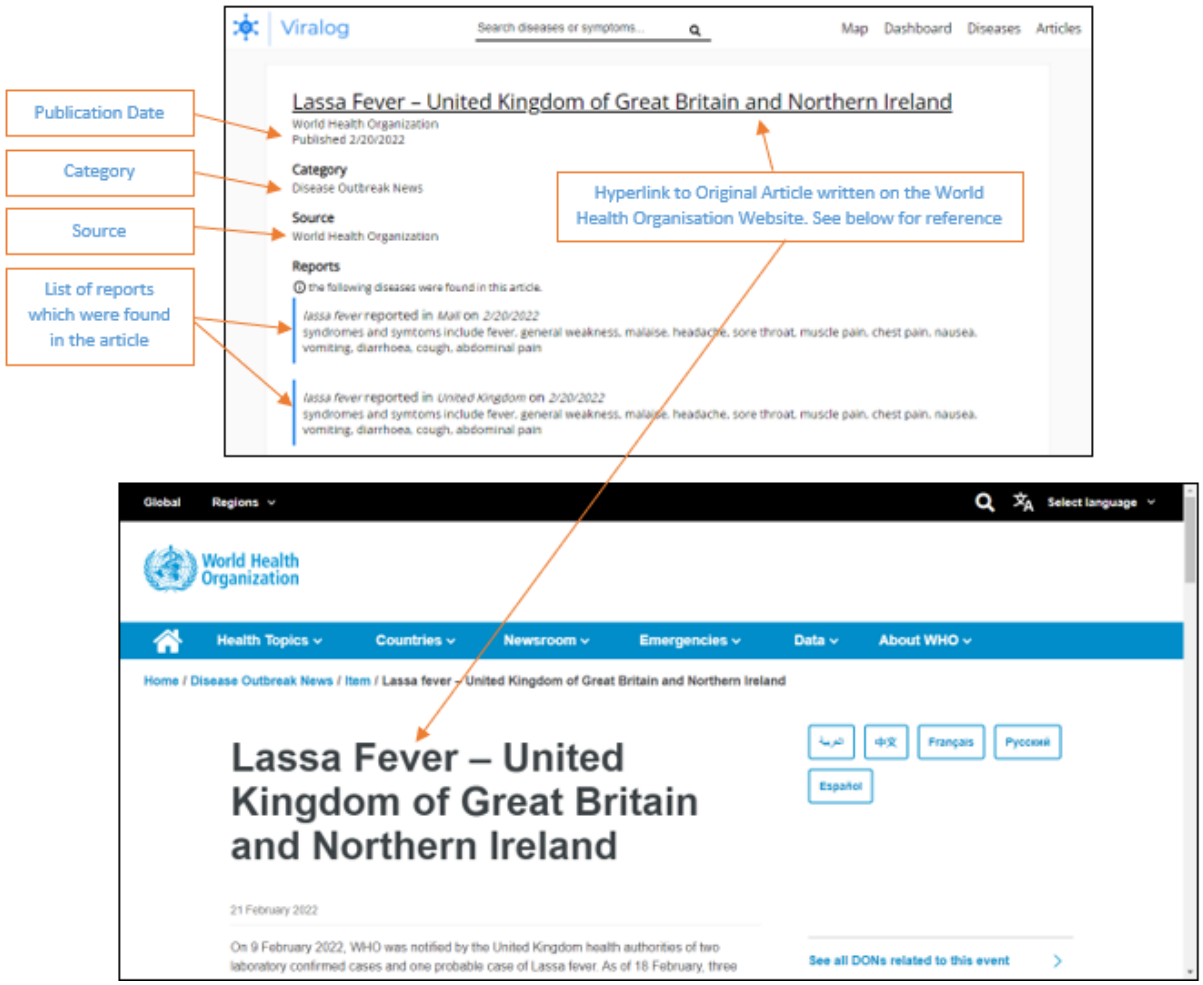

Figure 3: Example article page and the original article this page is based on.

Figure 4: The article list, navigated to by clicking the 'Articles' button on the taskbar.

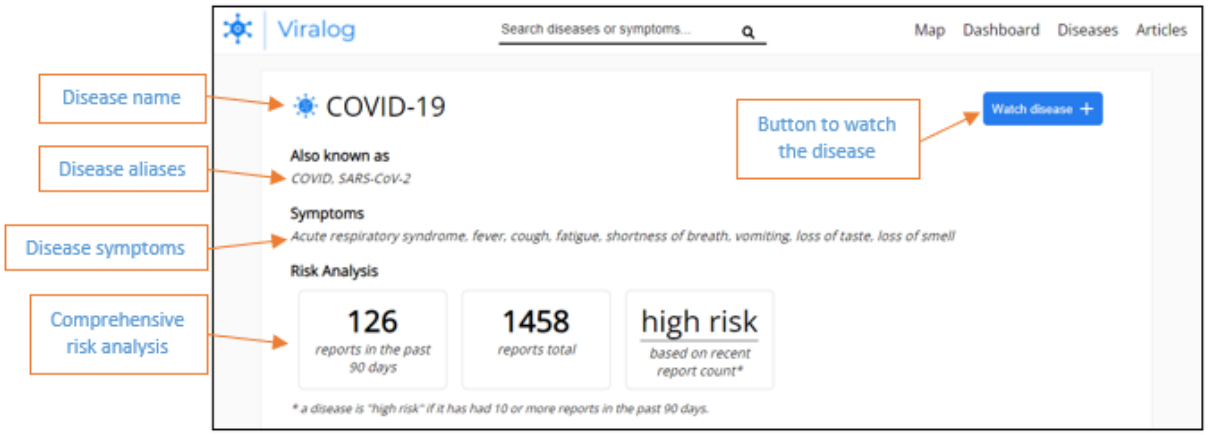

Figure 5: Example disease page (COVID-19)

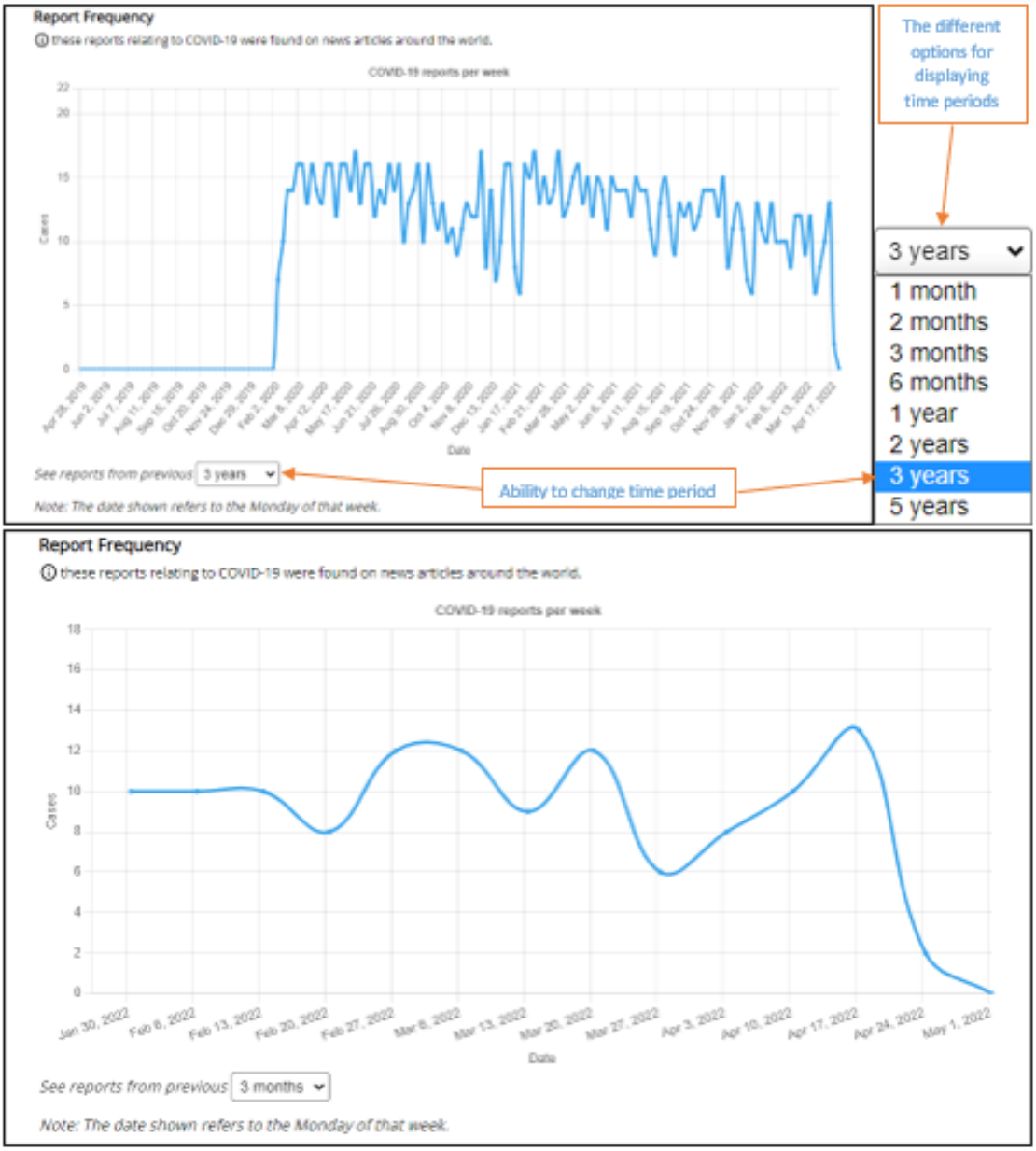

Figure 6: Report frequency graph for COVID-19. The top and bottom images are a representation for all reports made in the previous 3 years and 3 months respectively.

Figure 7: An interactive map on the COVID disease page, containing pins only for COVID reports.

Recent Reports

China tracks rising COVID activity in Beijing as US cases climb
COVID-19 reported in *China* on *4/25/2022*

Omicron less severe than Delta but more easily evades boosters
COVID-19 reported in *Scotland Denmark* on *4/25/2022*

WHO backs Paxlovid for high-risk COVID patients
COVID-19 reported in *China* on *4/22/2022*

COVID-19 third leading cause of death in US in 2021
COVID-19 reported in *United States* on *4/22/2022*

Figure 8: A list of recent reports on the COVID disease page.

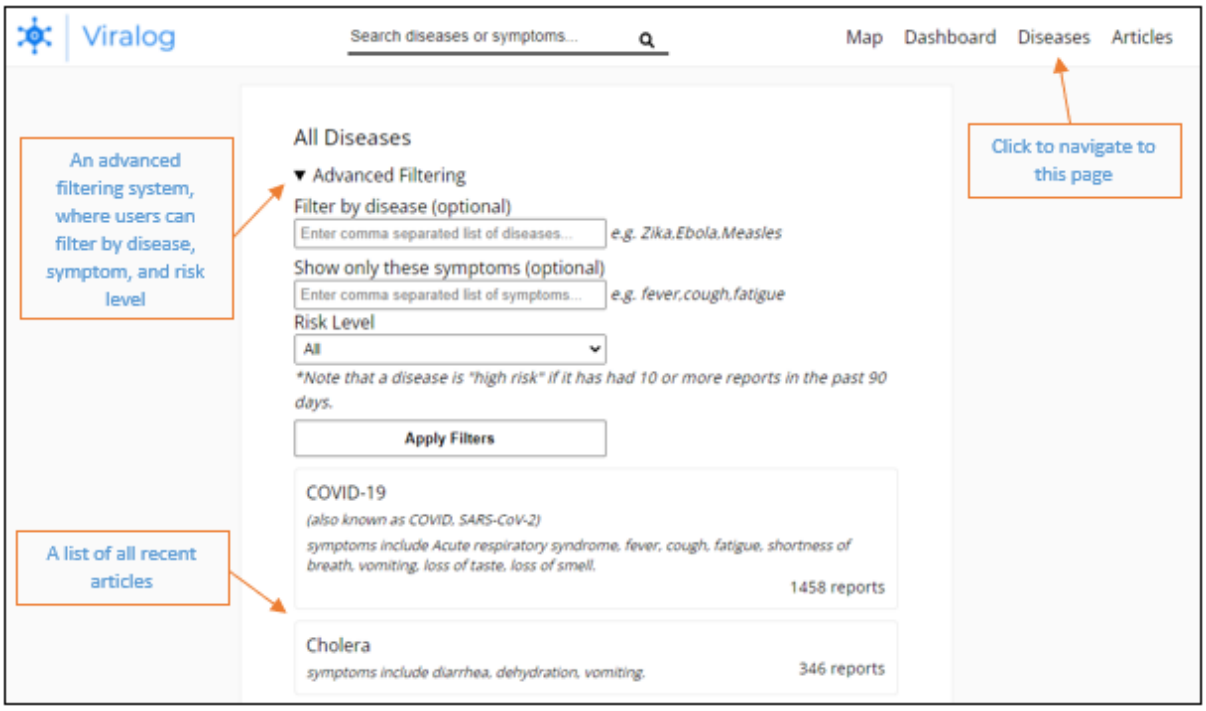

Figure 9: The disease list, navigated to by clicking the 'Diseases' button on the taskbar.

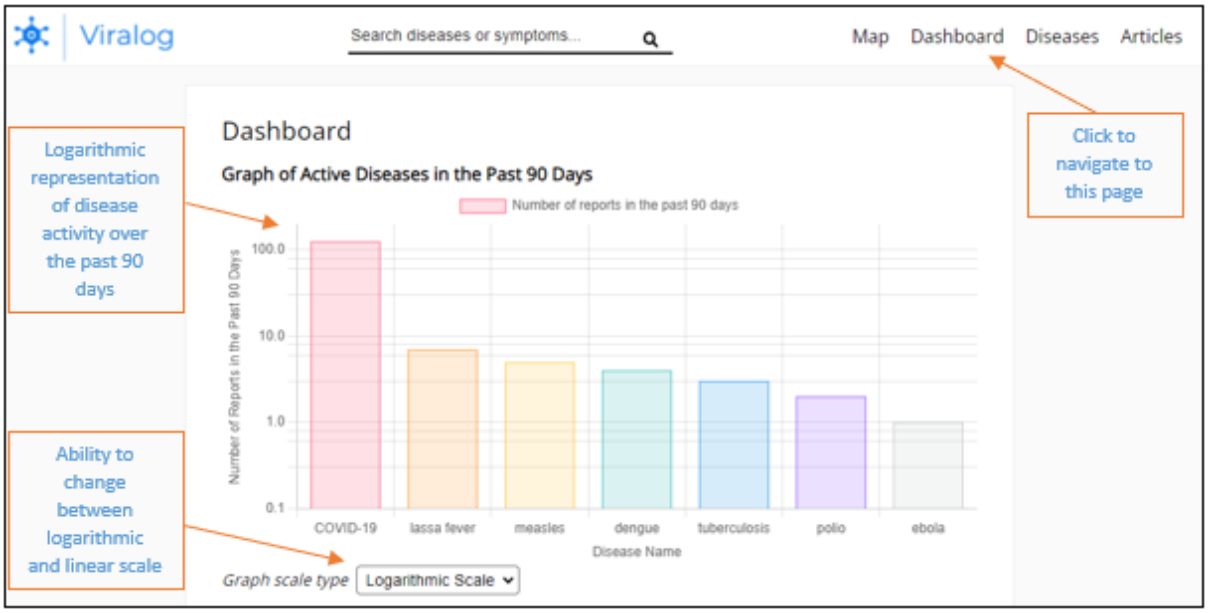

Figure 10: The top of the dashboard, navigated to by clicking the 'Dashboard' button on the taskbar.

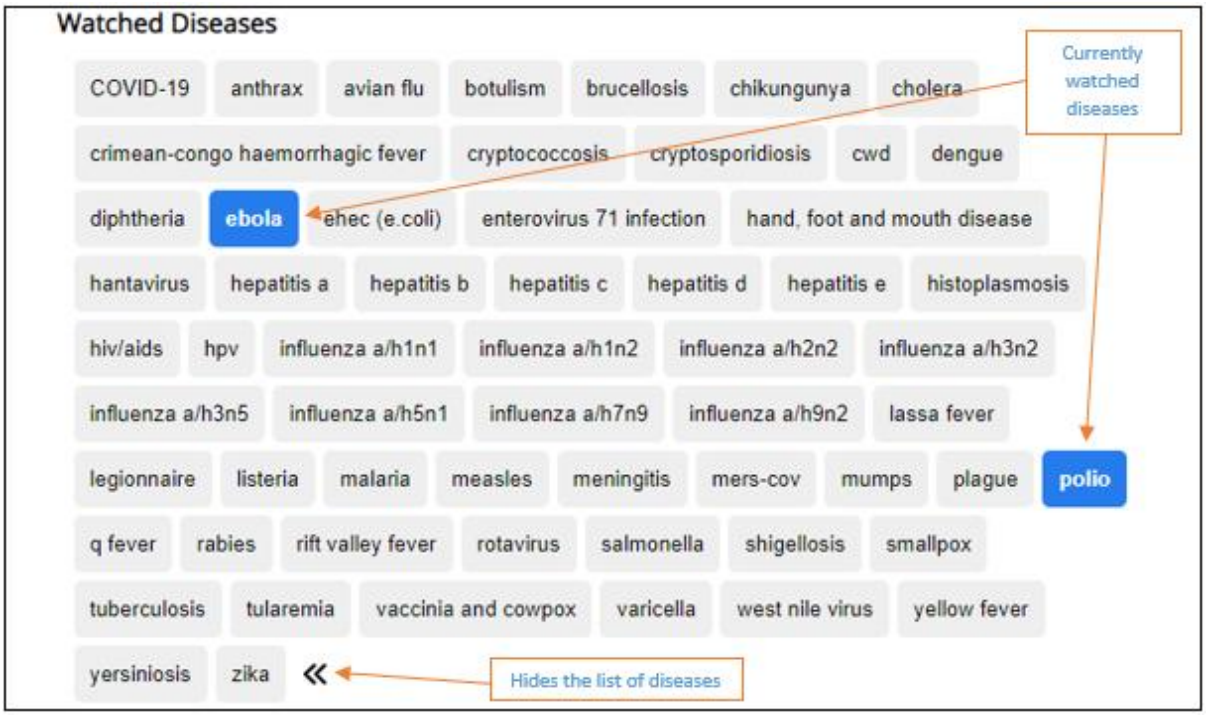

Figure 11: The list of diseases on the dashboard which can be watched. Currently watched diseases are highlighted in blue, namely 'ebola' and 'polio'.

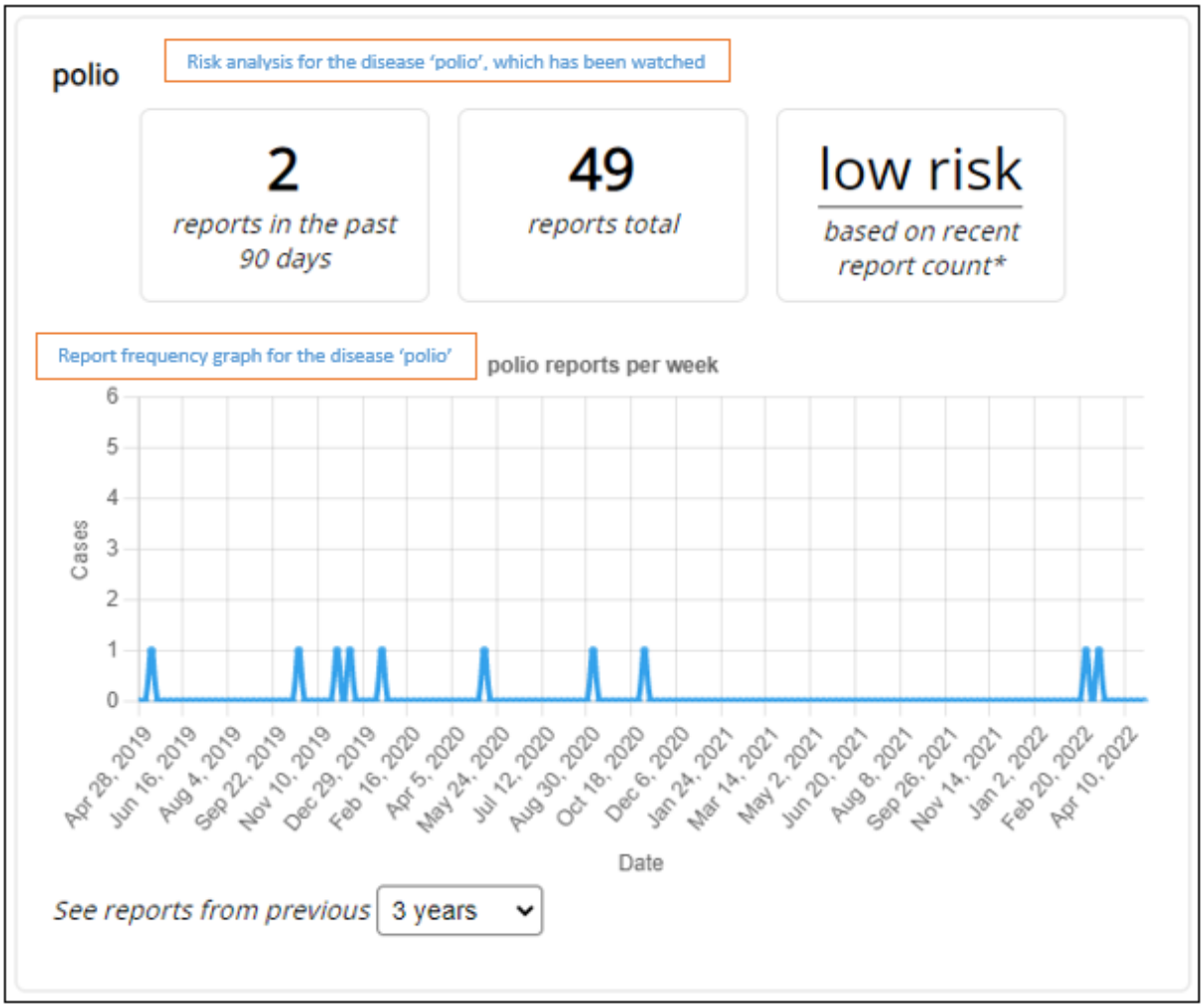

Figure 12: The summary displayed on the dashboard for a watched disease, in this case 'polio'.

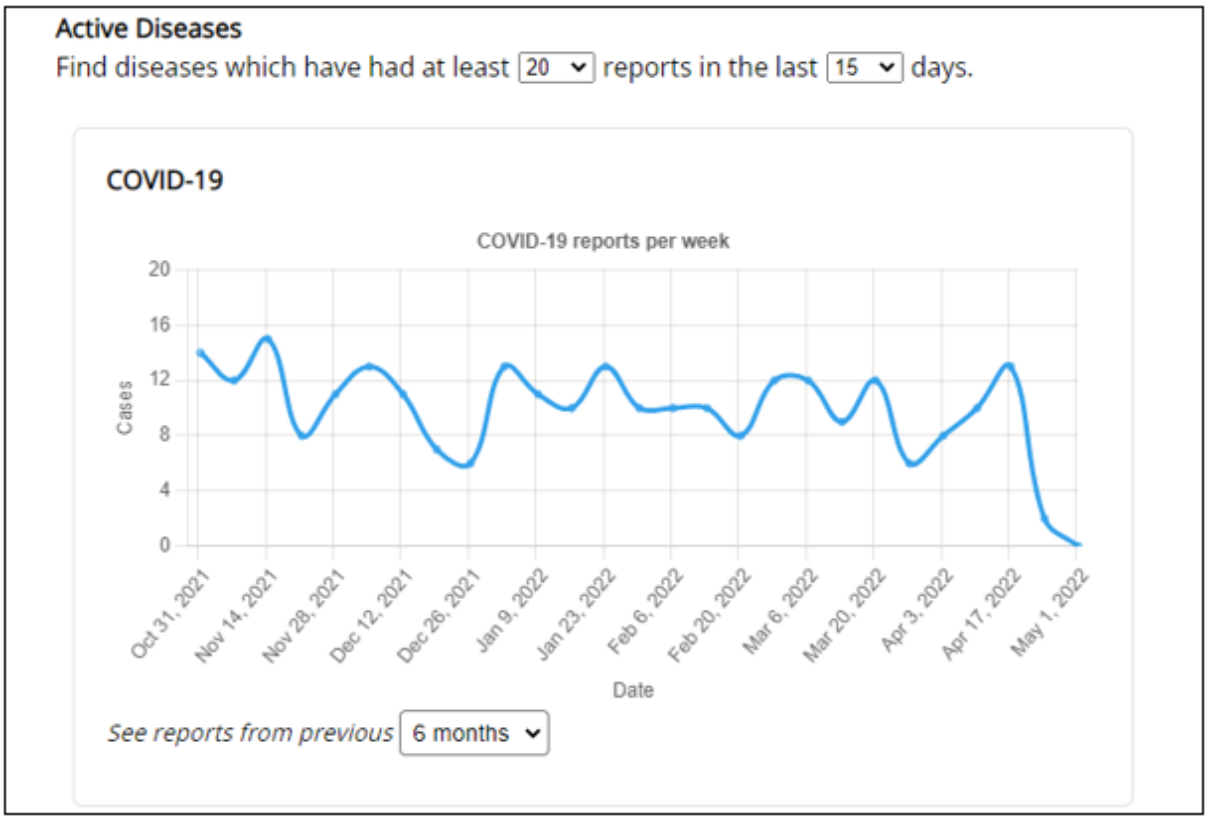

Figure 13: A list of active diseases, filterable by minimum report count in a given number of days.

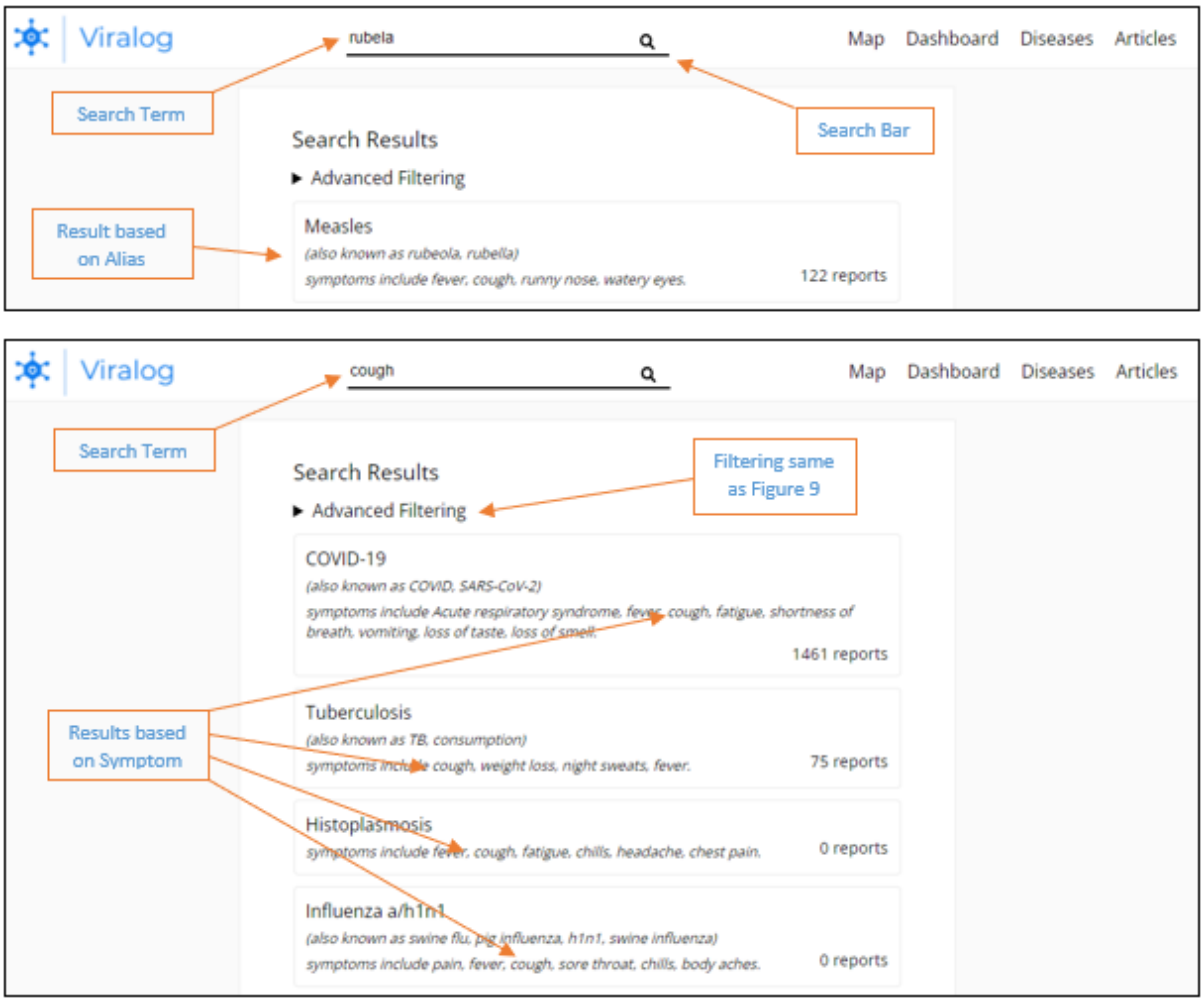

Figure 14: Results of using the search bar, firstly searching by disease alias and then by symptom.