Final Report: Website Data Hub for Public Health Data

Madyson Bradford, Ethan Reed, Zach Watson, Jackson Stilwell, Marc Walters
Texas A&M University, College Station

Website Data Hub for Public Health Data

Introduction

In the last two years, the availability and accessibility of public health related data has become vital to informing local policy decisions and influencing the behavior of communities. According to the CDC Foundation, public health is "the science of protecting and improving the health of people and their communities". For this reason, we will create a Google website to host a multitude of public health data for easy access by the public. This will eliminate the need for searching multiple databases from different organizational sources to collect substantial public health data. Although there are multiple factors considered determinants of public health, the following outlines public health determinants as defined for the scope of this project:

Public Health Topic	Determinants
Community Health	Access to health services, acute disease, chronic disease, outbreaks, endemics, pandemics, etc.
Environmental Health	Air pollution, water pollution, natural hazards, etc.

Figure 1. Public Health Topics addressed in the project, as well as determinants commonly considered aspects of those topics.

To truly create an accurate picture of the state of public health in any community, factors from all of these topics must be addressed in some form or fashion.

As public health continues to remain on the forefront of national security concerns, access to data associated with public health research will become more important (Frist, 2002). So will answering questions about how this data will be stored, what this data can provide, and the best use-cases for public health data. Many of these questions remained unanswered, even as the Covid-19 pandemic continued worldwide. Although Covid-19 is at the forefront of public health efforts right now, the project will give equal weight to other outbreaks and health conditions. This ensures that the goals of this project remain novel compared to trending current research, since many resources are already available for Covid-19 response.

Data

There is a diverse selection of public health data available to the public. This project will include any data from the Centers for Disease Control (CDC), the World Health Organization (WHO), and the U.S. Census Bureau that pertains to public health. This includes but is not limited to: instances of [certain] acute and chronic diseases, at-risk populations, and public health

geographies, and environmental quality data. Below is a table that provides the most likely data sources for this project, as well as a hyperlink to the data hub most likely used for data collection. This list is subject to change as the project progresses, since there will likely be a need for a more diverse set of data sources as data collection for environmental quality data begins.

Data Source	Link
Center for Disease Control (CDC)	https://www.cdc.gov/datastatistics/index.html
World Health Organization (WHO)	https://www.who.int/data/collections
U.S. Census Bureau	https://www.census.gov/data.html
Environmental Protection Agency (EPA)	https://www.epa.gov/data

Figure 2. Data sources for this project as well as links to data likely to be hosted on the public health data website. This list is subject to change as data collection progresses and evolves to include more topics.

Environmental quality data is an important factor in public health, so the scope of data included on this subject may grow to include more topics within this field as the project progresses. All data will come from well renowned and reliable sources that have been used in other scientific research as reputable sources. Part of data collection will be small amounts of data validation to ensure that the provided data is free of errors and from reliable sources only.

WebGIS methods

I. Website Creation

This project will utilize Google's free Web Hosting services. Google enables hosting of both static and dynamic websites on the cloud, in most cases free of charge. This project will use the Google Sites provided by Texas A&M University as part of the suite of software. Group members have experience with the platform, and it will also ensure that the hosted data and the website are included in the same service (Google Drive). This enables both efficient communication within the group as well as a collaborative environment to work on the site.

HTML

HyperText Markup Language (HTML) is the standard markup language for documents designed to be displayed on a web browser. With the assistance of other programming languages like Cascading Style Sheets, more detailed web pages can be created. In this project, HTML will

be used as the basis for certain sections of the website such as displaying clickable links, showing screenshots, and more. All HTML is embedded into the website.

Figure 3. All HTML was embedded into the Google Site for maximum customization and use of WebGIS learning objectives.

CSS

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is considered a cornerstone technology among the internet, in addition to HTML and JavaScript. In this project, CSS will be used to better chisel the smaller designs throughout the website, changing the presentation of certain pages with colors and fonts.

Results

We created a Public Health Data Hub hosted through the Texas A&M University System using Google Sites. Here is a link to the published site:

https://sites.google.com/tamu.edu/tamu-public-health-datahub/home

The site includes three pages: Homepage, Data, and About (Figures 4-7).

Website Data Hub for Public Health Data

Figure 4. Public Health Data Hub Homepage.

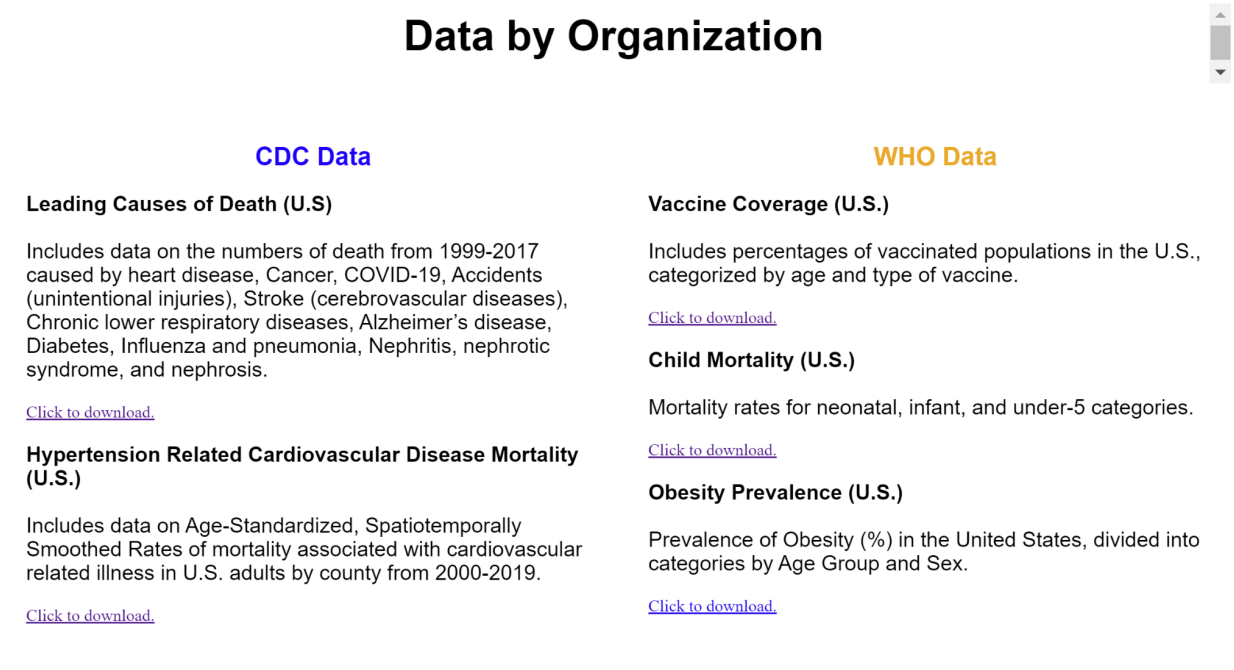

Figure 5. Public Health Data Hub Data page.

Website Data Hub for Public Health Data

Click to view all CDC Data.

Click to view all WHO Data.

Click to view all U.S. Census Data

Click to view all EPA data

Texas A&M University - College Station

Figure 6. At the bottom of the Public Health Data Hub Data page, users can access the entire collection of data by organization.

Madyson

Senior GIST Major

Academic Interests: Open Source GIS & Rare Earth Minerals.

Ethan

Senior GIST Major

Academic Interests: Human/Environment Interactions and Historical GIS.

Jackson

Senior Environmental Geoscience Major

Academic Interest: Environmental Law and Policymaking.

Zach

Senior GIST-CDA Major

Academic Interests: Computer Science and Data Analysis/Presentation.

Figure 7. Public Health Data Hub About Me page.

All data was added to a file in Google Drive and hosted through there. Anyone can access the data through a Texas A&M University gmail account. Data is divided by the organization it was

retrieved from, and each link below each description of each dataset leads to the google drive where the data is stored. All data can be downloaded to a local machine by following the links (Figure 4).

At the bottom of the Data page, there is a section where a user can access the entirety of data hosted for each organization. Not all collected data is listed in the organizational structure— only datasets we chose to highlight are included on the main Data page. For access to a multitude of other datasets, users can explore through the links at the bottom of the page (Figure 5).

Each element was embedded using HTML and inline CSS (Figure 3).

Discussion and Challenges

I. Programming Experience within the Team

The group for this project contained a variety of skill levels for programming languages. This did not end up hindering our ability to work together effectively. Each team member, upon completion of this project, gained basic HTML skills and was able to understand how elements were created on the site.

II. Hosting Large Datasets

Hosting large datasets, although believed to be an issue, ended up not being a significant issue at all. Through the infrastructure provided by Google Drive, all datasets were easily hosted without running into space issues. There was no issue creating the download links either, so basically this issue devolved into a non-issue for our finished project.

III. Broad Topic: Public Health

Initially, we believed it would be difficult to discern what sort of public health topics we should include on the website. This turned out to not be a problem! As a group, we decided which topics would be most relevant to us if we were looking for public health data. Because of this, we have a wide variety of topics from different organizations.

IV. Visualizations

One of our initial goals of the project was to create sample visualizations of the different datasets we provided. This ended up not being in-line without end goals. We did not want to reformat the data to make a visualization or map, since then a user may assume the data was properly formatted as provided through the download links to recreate the graphics. We did not want to reformat only some datasets seemingly at random, so in the end it was decided that visualizations of the data would not be provided.

Conclusion

Texas A&M University students and alumni now have access to a comprehensive data hub of public health data. We hope this not only improves access to public health data, but helps consolidate useful data in an easy-to-download format for analysis.

References

https://www.cdcfoundation.org/

Frist, B. (2002). Public health and national security: The critical role of increased federal support. Health Affairs, 21(6), 117–130. https://doi.org/10.1377/hlthaff.21.6.117