U.S.-Mexico Risk Taskforce to Support the Health Supply Chain Systems for Infrastructure and Workforce Threatened by the COVID19 Pandemic

Name	Role	Position Title	Affiliations
Zenon Medina-Cetina, PhD	PI	Associate Professor	College of Engineering and College of Geosciences Texas A&M University (TAMU)
Matt Cochran, DVM, MIA	Co-PI	Research Director	Cross-Border Threat Screening and Supply Chain Defense (CBTS), AgriLife Research, Texas A&M University System (TAMUS)
Maria Jose Perez-Patron, PhD	Co-PI	Research Assistant Professor, Associate Director	School of Public Health (SPH) Texas A&M University (TAMU), U.S. Census Bureau Research Data Center at Texas A&M University (RDC-TAMU)
Miriam Olivares, GISP	Co-PI	GIS Librarian	Marx Science and Social Science Library, Yale University
Greg Pompelli, PhD	Co-PI	Director	Cross-Border Threat Screening and Supply Chain Defense (CBTS), AgriLife Research, Texas A&M University System (TAMUS)

I. Objective/Purpose:

The purpose of this project was to integrate a binational Taskforce from Mexico and the U.S. that could serve as an advisory group to guide the identification and characterization of key variables and processes addressing the impact of COVID-19 in the state of risk of supply chains of critical value for trade between Mexico and the U.S. This project aimed as well as providing the best technology integrator to assimilate evidence that could facilitate the production of risk-based analytics to better inform stakeholders about the likely implementation of risk mitigating strategies that secure the optimal operation of supply chain systems involving trade between both countries. It aimed as well to produce a risk information system that would produce periodical information about the state of risk of Mexico and U.S. supply chains. The objectives of this project were:

1. To integrate a triple-helix binational *Taskforce* comprised of representatives from academia, industry and government from the U.S. and Mexico. Address the public health impacts of the COVID-19 pandemic on the U.S. – Mexico health supply chain systems for *health infrastructure* and for the *health of the workforce*, considering current and emerging regional social, economic and environmental Risks.
2. To develop a *Data-Lake* platform concentrating near real-time analytics following a risk systems approach that can provide strategic information about the evolution of COVID-19 and related current and emerging *threats*, the state of *vulnerability* of the health supply chain systems, and the likely *impacts* a combination of these may cause to society, the economy and the environment.
3. To publish a monthly *U.S.-Mexico COVID-19 Risk Bulletin* to provide scientific, technological, and strategic cultural support to secure the operation of the U.S.-Mexico health supply chain systems.

II. Research Results:

- **Explanation of any changes from the approved workplan**
- **Methodology**

Taskforce

The original approach from the workplan envisioned the development and integration of a binational triple-helix *Taskforce* comprised of representatives from academia, industry and government, from the U.S. and Mexico (Key Milestone).

The first *Taskforce* meeting was held on October 23rd, 2020. The inaugural Taskforce members are listed in Table 1 below. Since the Taskforce kickoff, biweekly meetings (every other week) were held with participation above 90%.

Organization	Member
Texas A&M University – College of Engineering	**Dr. Zenon Medina-Cetina** Project PI Associate Professor, Civil Engineering, Petroleum Engineering, Ocean Engineering, Geography
CBTS Director	**Gregory Pompelli, PhD** Project Co-PI Director, CBTS-TAMU
CBTS – Texas A&M University	**Dr. Matt Cochran** Project Co-PI Research Director, CBTS-TAMU
Texas A&M University	**Dr. Maria Jose Perez-Patron** Project Co-PI Adjunct Assistant Professor Department of Epidemiology & Biostatistics Texas A&M School of Public Health
GIS Consulting	**Miriam Olivares** Project Co-PI and Consultant Marx Science and Social Science Library, Yale University THEI Consulting
Mexican National Business Advisory Council (Consejo Coordinador Empresarial, CCE)	**Victor Gutierrez Martinez** CEO, Grupo Plenum Chair of CCE's Innovation Commission
Mexican National COVID-19 Scientific Advisory Board (Grupo CONACYT COVID-19)	**Dr. Oscar Sanchez Siordia** Director, Mexican National Geo-Intelligence Laboratory Institutional Member of Grupo CONACYT COVID-19
Mexico Census Bureau	**Dr. Sergio Carrera** Director, Integration, Analysis and Research México Census Bureau
U.S. Census Bureau	**Dr. Bethany DeSalvo** Chief, Small Area Modeling and Development Social, Economic and Housing Statistics Division U.S. Census Bureau

National Institute of Mathematics (CIMAT)	**Dr. Graciela Ma. De los Dolores González Farías** Director, CIMAT Campus Monterrey
U.S. Department of Homeland Security	**Duane C. Caneva, MD, MS** Chief Medical Officer Department of Homeland Security
U.S. Department of Homeland Security	**Tom McGinn DVM** Senior Veterinarian Office of the Chief Medical Officer (OCMO) Countering Weapons of Mass Destruction (CWMD) Department of Homeland Security
U.S. Department of Homeland Security	**Alexander L. Eastman, MD, MPH, FACS, FAEMS** Senior Medical Officer - Operations Office of the Chief Medical Officer Countering Weapons of Mass Destruction Office U.S. Department of Homeland Security

Table 1.- CBTS R13 U.S.-Mexico Taskforce

New *Taskforce* members were invited on the March 5th, 2021 meeting. These are listed in Table 2 below.

Organization	Member
Texas A&M University – Department of Geography	**Julie Loisel, PhD** Associate Professor Department of Geography Texas A&M University
U.S. Department of Health and Human Services (HHS)	**Maria Julia Marinissen, PhD, MSc** Health Attaché U.S. Department of Health and Human Services (HHS) U.S. Embassy | Mexico City
U.S. Department of Homeland Security	**Thomas M. Wilkinson, MD** Chief Medical Information Officer Countering Weapons of Mass Destruction U.S. Department of Homeland Security

Table 2.- CBTS R13 U.S.-Mexico New Taskforce Members

As of June 30th, 2021, a total of thirteen *Taskforce* meetings took place. During these meetings, the *Taskforce* was in charge of facilitating the identification and provisioning of evidence to be integrated into the *Data-Lake (CBTS' Data-Lake System, CBTS-DLS).* R13' contractors and SGL students created a workflow to integrate supporting evidence to produce risk-based analytics, following R7's Risk Assessment and Management model (CBTS-RAM), which served as a reference for discussion with Taskforce members.

The *Taskforce* meetings contributed as well to the development of CBTS-SGL Dashboard by providing databases, and by suggesting the production of Risk Analytics such as the US-Mexico Border States Analysis and Excess Deaths that are a fundamental part of the current status of the dashboard.

Data-Lake

The *Data-Lake System* (CBTS-DLS) focused initially to provide elements to assess the state of risk of U.S.-Mexico health supply chain systems for both *infrastructure* and *workforce.* This platform provides access to risk analytics and to its supporting variables and processes, including datasets and predictive models, needed to produce evidence-based support on the causes and effects posed by COVID-19 on the U.S. health supply chain.

Risk Assessment Framework

In order to assess risks, a risk assessment framework is adopted that is defined as follows:

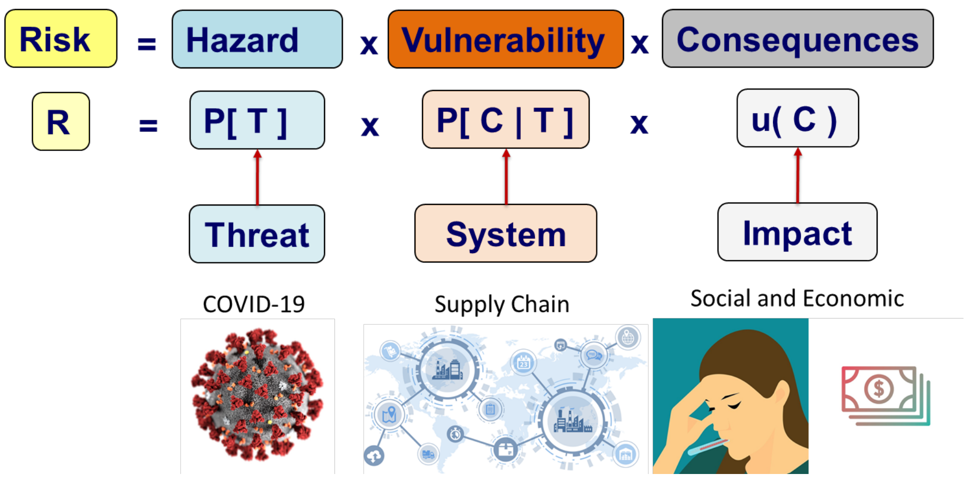


Figure 1.- Risk Assessment Framework.

Where:

Hazard is the probability that a particular Threat T (for example, COVID-19) with a given intensity is exceeded within a given period of time.

Vulnerability is the probability of reaching a Consequence or damage in the element or system of interest (for this project is supply chains systems), conditioned on a given Threat intensity.

And the consequences are the expected Consequence value (for example people getting ill, economic losses) of the element or system of interest exposed to a given Threat intensity.

Data-Lake System (CBTS-DLS)

Due to university security protocols, it was decided to modify our original system's architecture, resulting in a clear division; servers located on-premise are hosted and managed by TAMU-IT with restricted access to public resources outside the internal firewall, while a separate cloud-based server hosted by Amazon's Web Services (AWS) acts as the public-facing element for R13, where the projects' data warehouse and dashboard are hosted.

CBTS-DLS' on-premise servers can be only accessed through internal TAMU's network following its university-wide security protocols. The cloud-based system hosts the public dashboards of synthesized data and risk analytics with exclusive communication with CBTS-DLS' on-premise system to stream the projects' information. Figure 2 describes CBTS-DLS' two main components:

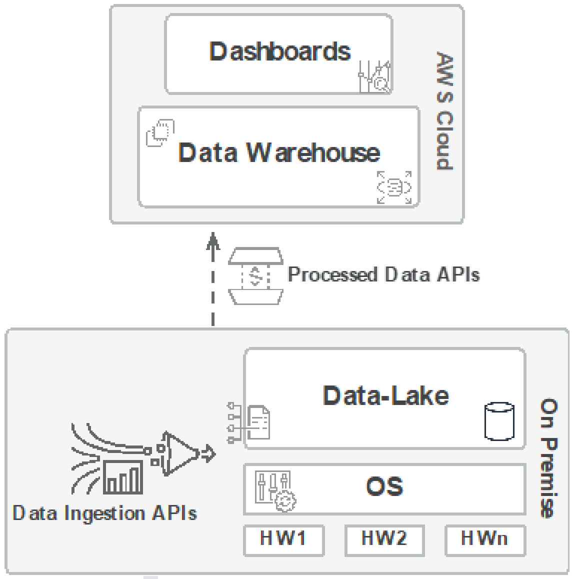


Figure 2.- Data-Lake System Diagram.

On the other hand, CBTS-DLS' implementation and operation process are depicted in Figure 3:

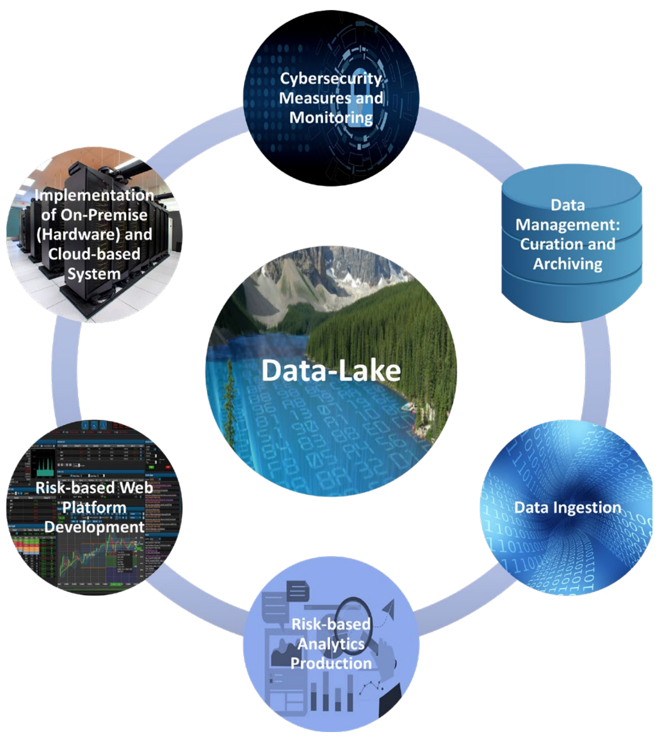


Figure 3.- Data-Lake Design, Development, Implementation and Operation diagram.

This figure illustrates that the CBTS-DLS' operation started with implementation of on-premise (hardware) and the cloud-based component (AWS' hardware instance). It's worth noticing that CBTS-DLS' operation was split for security and management purposes in two collaborative efforts, the on-premise system was allocated to R13 and the cloud system was allocated to R7.

The on-premise architecture is described below. The cloud architecture is described in R13's final report.

On-premise cluster setup

The R13 servers' cluster infrastructure has 3 physical servers, with the following characteristics:

Server Name	Specifications
r13app1.engr.tamu.edu	• Cores: 96 • Hard disks: 1 with approximately 1.2 TB* and 1 with approximately 11 TB* • RAM: Aproximately 370 GB*
r13app2.engr.tamu.edu	• Cores: 96 • Hard disks: 1 with approximately 1.2 TB* and 1with approximately 11TB* • RAM: Aproximately 370 GB*
r13data1.engr.tamu.edu	• Cores: 96 • Hard disks: 1 with approximately 1.75 TB and 1 with approximately 11TB* • RAM: Aproximately 370 GB*

Table 3.- On-premise cluster setup.

Each physical server is divided into 5 virtual machines; 1 VMs for hosting master components and 4 VMs for hosting worker components.

On-premise cluster architecture diagrams

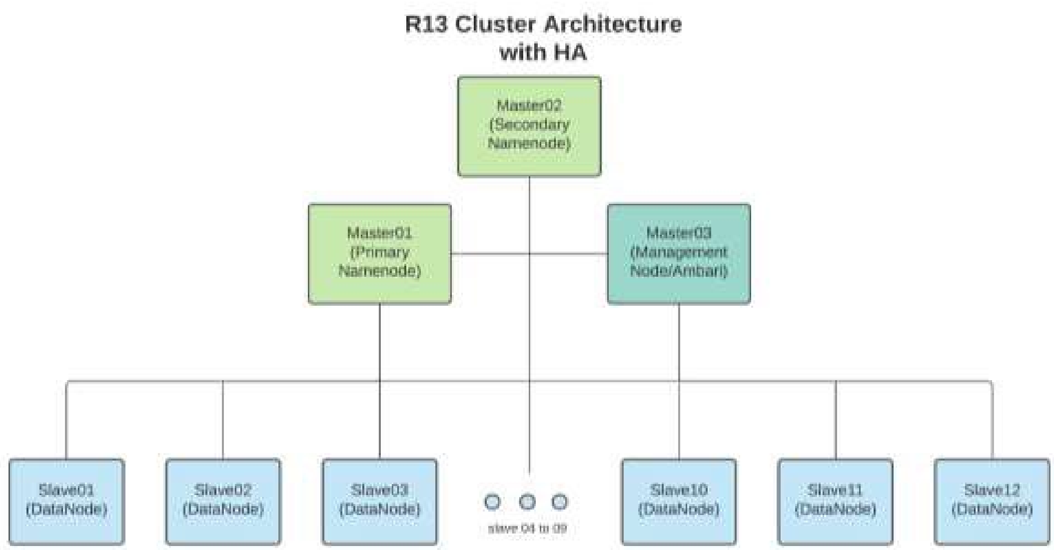


Figure 4.- Cluster Architecture Diagram.

Cybersecurity measures and monitoring

A background check and security protocols are in place for everyone involved in the team according to government and institutional regulations. Also, the project's leads attended more than a dozen meetings with TAMU-IT specialists to follow strict security cybersecurity protocols. Furthermore, a screening of every foreign national involved in the project has been submitted to DHS due to the binational nature of the project.

Data management: curation and archiving

Utilization of the Texas Data Repository (TDR) is in place to publish curated data ingested and produced in CBTS-DLS' dashboard for archiving. The key strategy to manage all data and meta-data is the development of a data management matrix, which serves as a gate-keeper for data clearance (Figure 5). Additionally, the OAKTrust Digital Repository was adopted to generate proper citations for all the products and data produced by the R13 project.

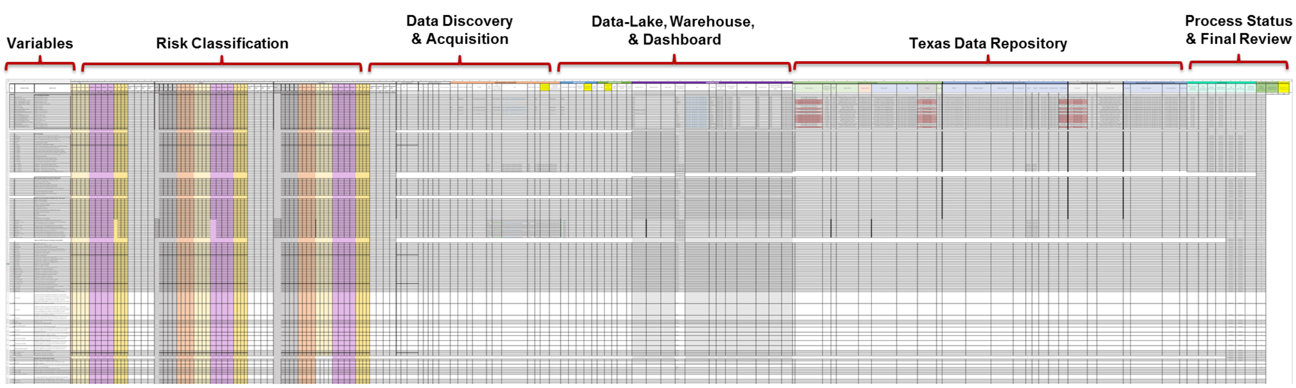


Figure 5- R13 Data Management Matrix

Data ingestion

The data ingestion process (Figure 6) starts with the identification of databases such as John Hopkins and CONACYT from Mexico. The raw data is ingested to the data-lake using apache NIFI tool, and it's curated using Apache Spark and apache Zeppelin tools. Then the curated data is managed using MySQL tool to deploy it to the cloud AWS dashboard, via python and php.

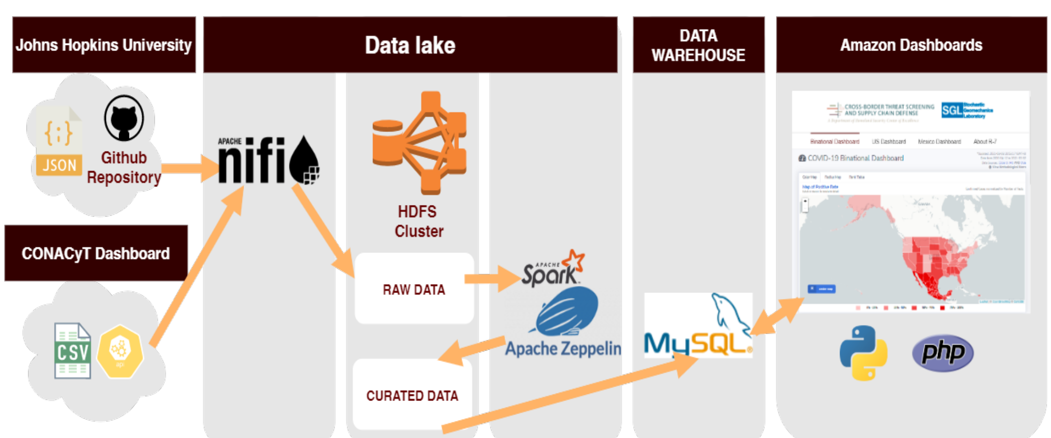


Figure 6.- R13 Data Ingestion Process.

As of June 30th, 2021, the following list of datasets have been ingested:

- John Hopkins University COVID-19 data.
- Secretaria de Salud México – COVID-19 data.
- CDC Social Vulnerability Index (SVI) data.
- Mexico CDC-based SVI data.
- CENSUS' Community Resilience Estimates (CRE) data.
- Mexico's COVID-19 Social Vulnerability Index (MCVI) data.
- Google Mobility data.
- US Excess Deaths data.
- Mexico Excess Deaths data.

Risk-based Analytics

The analytics, maps, and visualizations were produced using the list of datasets presented in "Data ingestion" section. This set of products was presented in Data-Lake Team and Taskforce meetings in order to provide feedback to the Data-Lake Team. Of special interest is a set of products produced based on the results of CIMAT modeling team and SGL risk team.

Risk-based Web Platform Development

A Risk-based web dashboard was developed to make available information and risk analytics to report R13's results. This platform allows the user to explore spatio-temporal information in an interactive and dynamic format through the use of maps, tables, and risk analytics organized in 3 main sections: Binational Dashboard, US Dashboard and Mexico Dashboard (Figure 7).

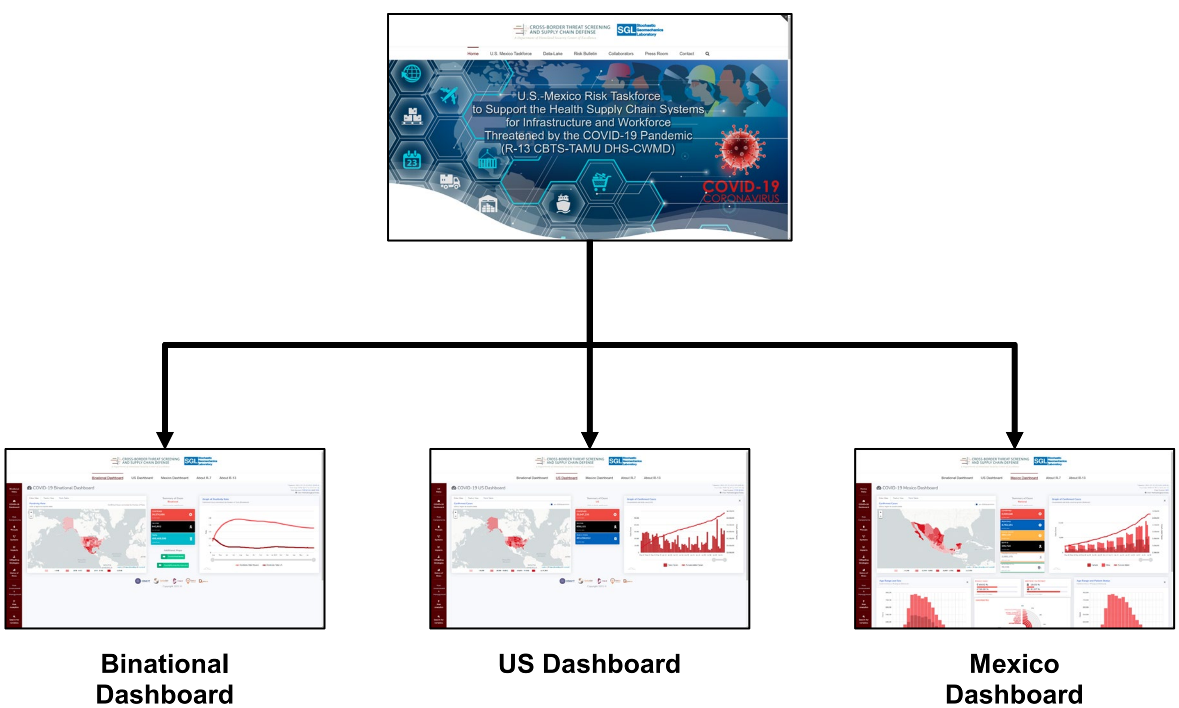


Figure 7.- R13 Situational awareness dashboard diagram.

The platform contains visualization and download features of more than 120 categorized variables in terms of Risk (Figure 8a), according to our Risk Assessment Framework (Figure 1). Additionally, it includes a “Risk Analytics” module with a set of maps and variables that are the result of model predictions (Figure 8b). All the variables are located in the sidebar menu at the left of your screen (Figure 8), and they are organized in 3 main sections (Figure 7):

- Binational Dashboard: variables and maps that have information for both US and Mexico.
- US Dashboard: variables and maps that only have US information.
- Mexico Dashboard: variables and maps that only have Mexico information.

In order to facilitate the search and analysis of variables in the platform, a search module was developed that allows users to filter variables (Figure 8c and Figure 9).

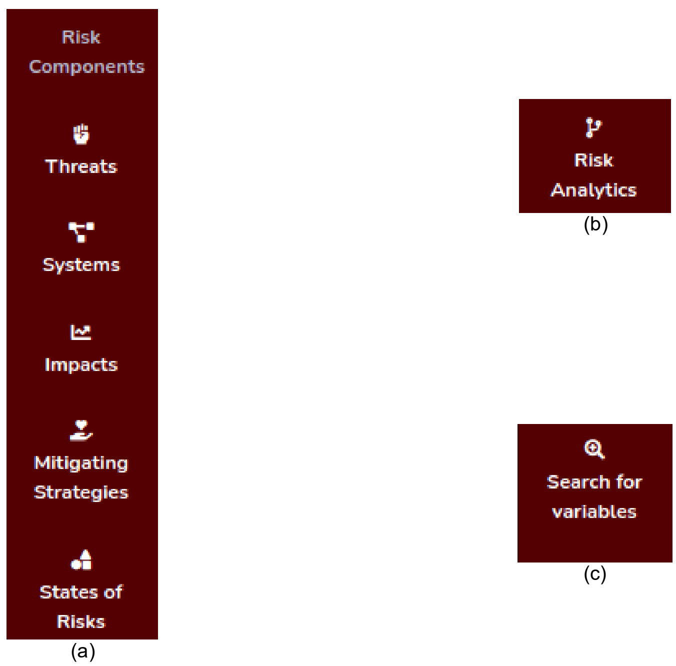


Figure 8.- Menu to access variables and maps (a) according to Risk Components, (b) Risk Analytics and (c) variable search.

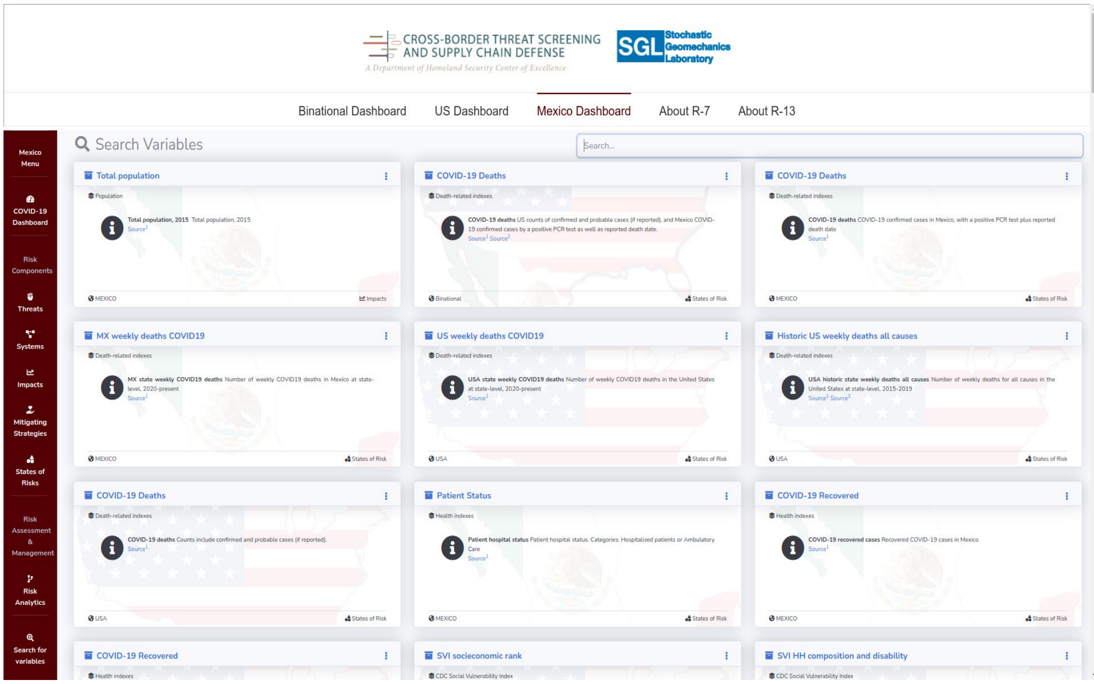
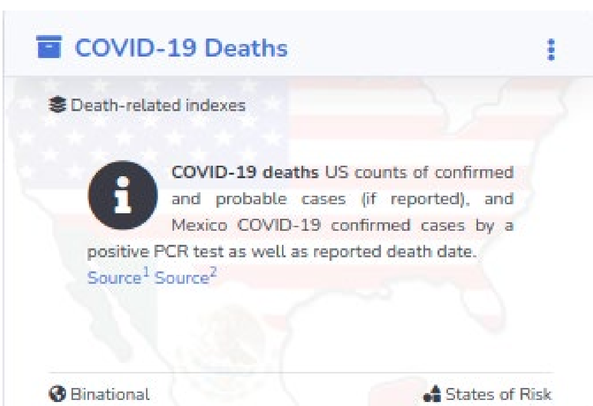
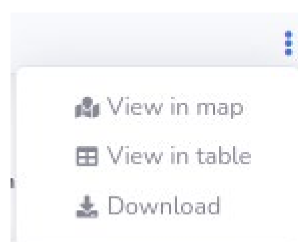


Figure 9.- Search for variables module.

The search for variables module shows information such as name, description, classification (according to the Risk framework), spatial granularity, and sources of data of variables that are filtered according to the user's search (Figure 10a). It also has a menu that allows users to visualize maps, tables, and download feature of CSV files (Figure 10b).



(a) card info body



(b) card info menu

Figure 10.- Information cards for available variables.

The platform has 3 types of visualizations:

- Dashboard view.
- General Map view.
- Table view.

The Dashboard view presents available information about COVID-19 pandemic associated to its geographical location (Binational, US, and Mexico). Each one of the Dashboard views have their original sources of data, maps, informative ‘cards’, and interactive graphs (Figure 11).

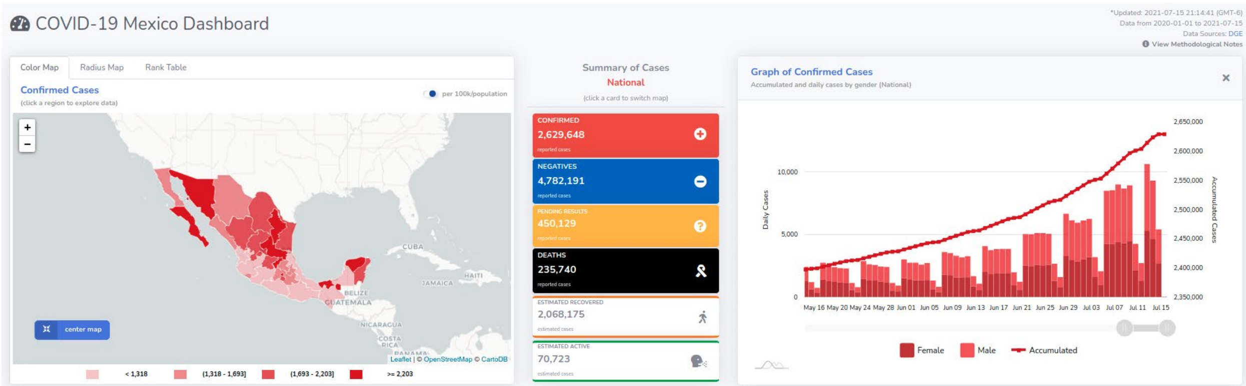


Figure 11.- Dashboard view.

The interactive maps have the ability to display specific information regarding a geographic area when data is available. A ‘click’ on an informative card will generate new map and analytics that represent the variable in the card. The interactive graphs allow users to hide/show elements in the graphs by clicking on the label. They also provide a slider feature in the horizontal axis to narrow down the time step.

The General Map view consists of a set of variables associated to one Risk Component. This view has a drag and drop tool that generates maps and reports (Figure 12).

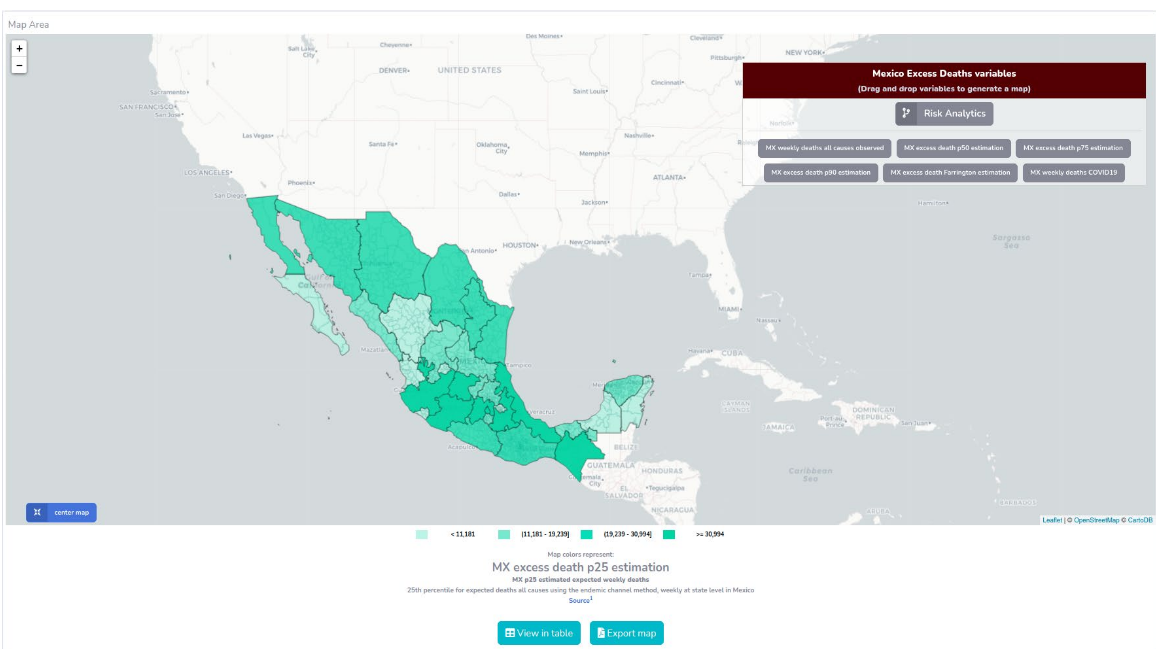
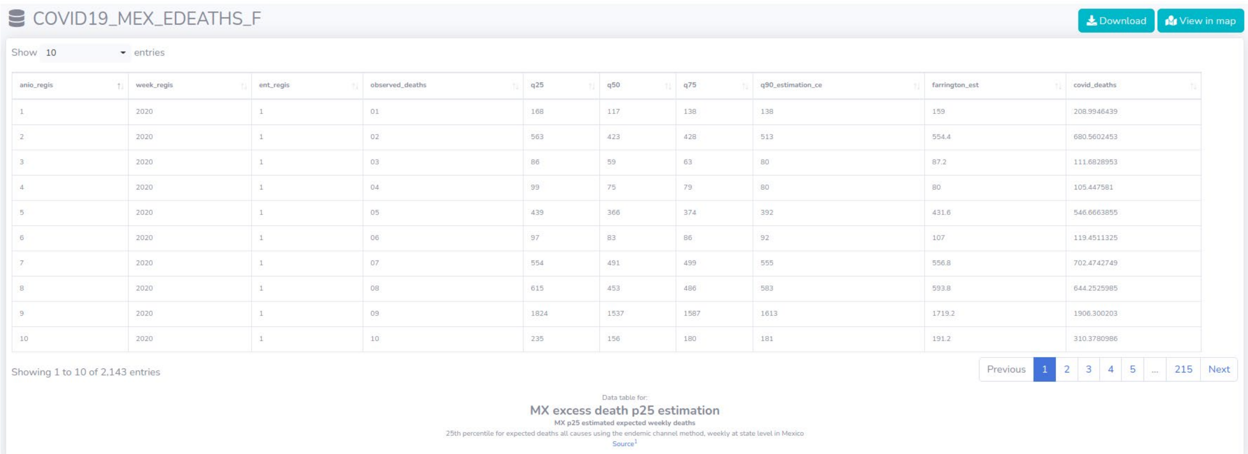


Figure 12.- General Map view.

After dragging and dropping a variable, a map will be generated automatically according to the variable's coverage and scale. Below the map, users can find the label and descriptive information of variables, and their original sources of data. There are also two buttons that are linked to the Table view of the variables, and a PDF report generator.

The Table view contains the original data of a variable. By default, the table view displays groups of 10 elements, and the first 5,000 records (Figure 13).


COVID19_MEX_EDEATHS_F

Download | View in map

Show 10 entries

anio_regis	week_regis	ent_regis	observed_deaths	q25	q50	q75	q90_estimation_ce	farrington_est	covid_deaths
1	2020	1	01	168	117	138	138	159	208.9946439
2	2020	1	02	563	423	428	513	554.4	680.5602453
3	2020	1	03	86	59	63	80	87.2	111.6828953
4	2020	1	04	99	75	79	80	80	105.447581
5	2020	1	05	439	366	374	392	431.6	546.6663855
6	2020	1	06	97	83	86	92	107	119.4511325
7	2020	1	07	554	491	499	555	556.8	702.4742749
8	2020	1	08	615	453	486	583	593.8	644.2525985
9	2020	1	09	1824	1537	1587	1613	1719.2	1906.300203
10	2020	1	10	235	156	180	181	191.2	310.3780986

Showing 1 to 10 of 2,143 entries

Previous 1 2 3 4 5 ... 215 Next

Data table for:
MX excess death p25 estimation
MX p25 estimated expected weekly deaths
25th percentile for expected deaths all causes using the endemic channel method, weekly at state level in Mexico
Source[1]

Figure 13.- General Map view.

Below the table, users can find the label and descriptive information of variables, and their original sources of data. Above the table, two buttons are provided to return to the General Map view or download the CSV file.

U.S.-Mexico COVID-19 Risk Bulletin

The monthly Risk Bulletin is a compilation of research advances and outcomes from R13 contractors and SGL students, *Taskforce* contributions and inputs, and Data-Lake generated Risk Analytics. The main focus of the Bulletin is U.S. & Mexico Health Infrastructure, and Health of the Workforce. The Bulletin aims to improve decision making across international supply chain systems by providing timely and synthesized relevant information to decision-makers on both sides of the border to help them make more informed choices.

Risk bulletins are compiled using the key takeaways and major milestones accomplished by CBTS-SGL-TAMU team from:

- CBTS-SGL internal weekly meetings.
- R13 Data-Lake Team weekly meetings.
- R13 Contractors' weekly meetings.
- Taskforce bi-weekly meetings.
- Meetings minutes.
- Meeting recordings.
- R13 research briefs.

The bulletins are composed in Markdown, and written utilizing Pandoc format, so that it can be run through Pandoc processes which produce a range of file formats without sacrificing design. Once the formats are generated, the files can then be displayed in a number of ways, including webpages, in pdf form or via email. See Figure 14 for a depiction of this process.

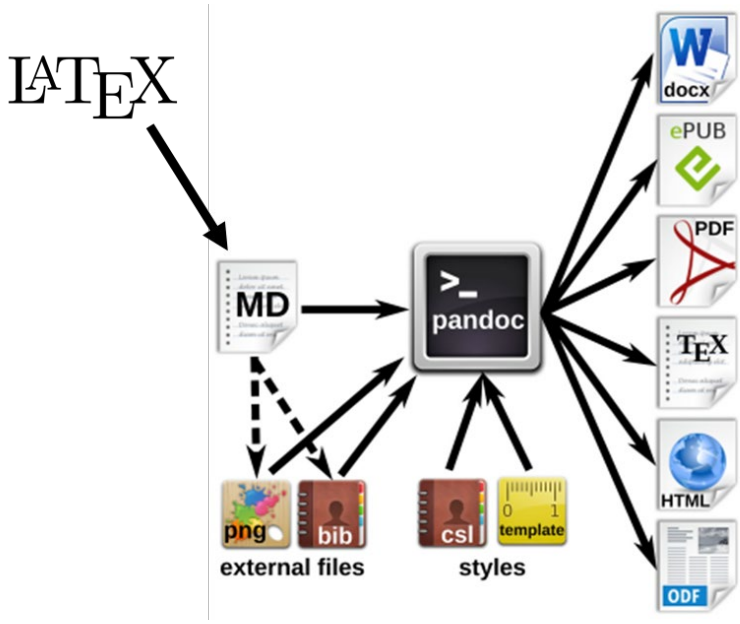


Figure 14.- Risk Bulletin production workflow.

As of June 30th, 2021, a total of seven Risk Bulletins have been published at R13 project website.

Report card grading scheme

A report grading scheme was developed to define the status of threats, system vulnerabilities and impacts, applicable to any U.S. supply chain. Presently, three evidence sources are considered to produce a preliminary risk index (SGL Risk Index). These are the CDC Social Vulnerability Index (SVI)[1] which provides specific socially and spatially pertinent information that is primarily used to aid public health officials and local planners to better prepare communities to emergency events and all of the variables used to compute SVI are extracted from the ACS 2014-2018 5-year estimates[2], the COVID-19 Community Profile Report (CCPR)[3] that issues COVID-19 related statistics on different geographical unit based on the last 7 day trends and is released by the White House COVID-19 team and lastly, the U.S. COVID-19 Community Vulnerability Index (CCVI)[4] , which serves as an extension of the aforementioned CDC SVI to account for COVID-19 specific epidemiological risk factors, public health system capacity and high-risk environment, on top of the socioeconomic and demographics information.

Across all of the data sources, 40 variables are selected and classified according to the SGL risk framework into the category of theme (Figure 1), system and impact (19 variables are from SVI, 3 variables from CCPR and 18 variables from CCVI[5][6][7]). Each variable is classified into one of the risk factors (e.g. threats, systems, and impacts or exposure). A simple and robust methodology is utilized inspired in the CDC SVI's index. After the initial classification of the variables in one of the three risk components, the variables are ranked by the geographical unit of choice (i.e. county). The ranking ultimately normalizes the categories assigned to each county to a value ranging between 0 and 1. The formula used is given below:

$$Rank = \frac{relative\ position\ of\ a\ county\ with\ respect\ to\ the\ variable - 1}{Total\ number\ of\ county - 1}$$

Some variables are normalized to the county's estimate population before ranking them to better represent the state of the county with respect to the variables of interest. The ranked variables are then summed according to their respective category (32 variables are classified to system, 7 variables to threat and 1 variable to impact) which were classified earlier according to the risk factor they represent. Subsequently, the summed categories are each ranked again using the same formula as above and mapped to a choropleth map. The same procedure is applied to the subgroups. The subgroups are divided to 2 types which are static if the subgroups variables are sourced only from the ACS 5-year estimates & CCVI, and dynamic if the subgroups contain variables from CCPR as CCPR report being published almost daily to account for the COVID-19 dynamics. The SGL risk index [8] is the description used for the state of risk at a particular county and is made up of system, threats and impact. The risk is computed as follows:

$$Risk = Hazard * Vulnereability * Consequence$$

The threat analysis leads to Hazard assessment, the system analysis leads to the Vulnerability assessment, and the impact analysis leads to the Consequence assessment. Maps of each risk factor to illustrate the formulation of SGL risk index are shown in Figure 15, 16 and 17 respectively, as computed for an exercise run for January 31st of 2021 (highest number of deaths that was registered due to COVID-19). The product of the risk factors results in the State of Risk map (Figure 18), which is compared against the COVID-19 deaths per county to show qualitatively the potential to adopt this simple methodology to overlay pathways of any given supply chain (Figure 19), and then have a preliminary assessment of its current state of risk.

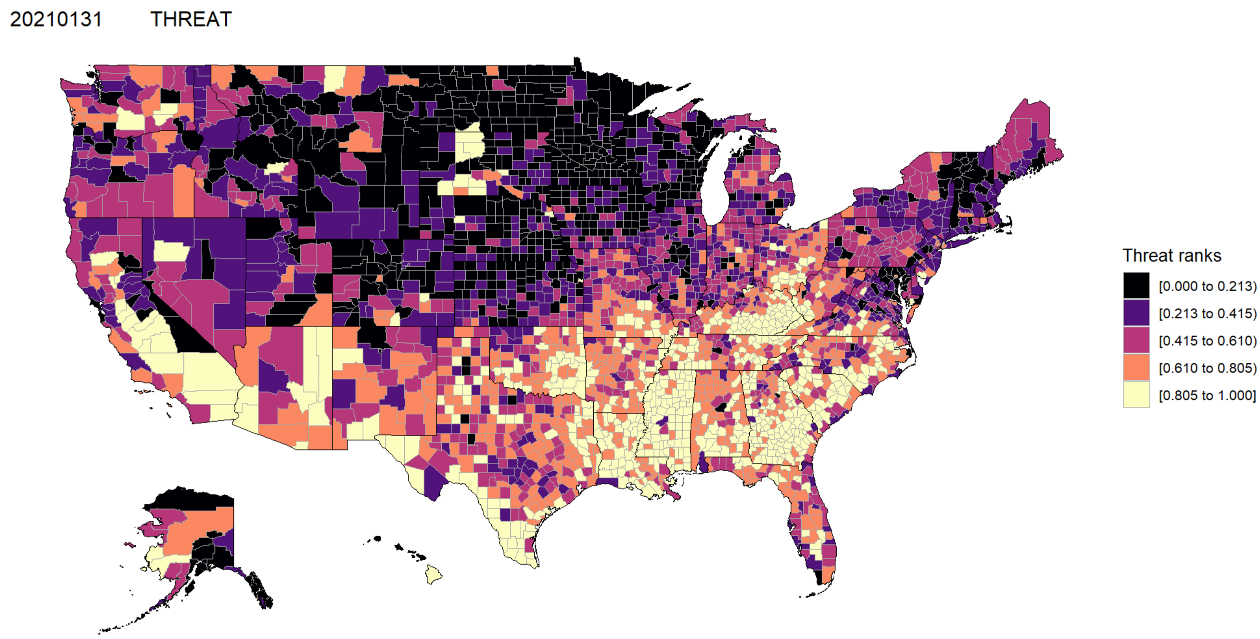


Figure 15.- Threat map as a dynamic category.

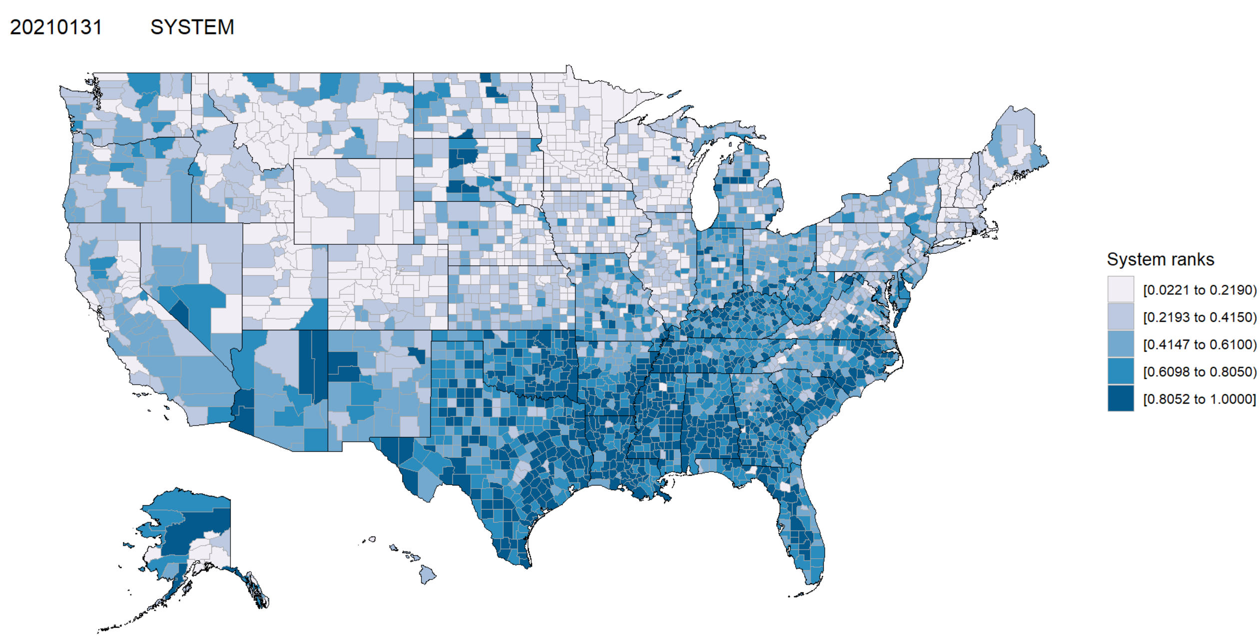


Figure 16.- System vulnerability map as a static category.

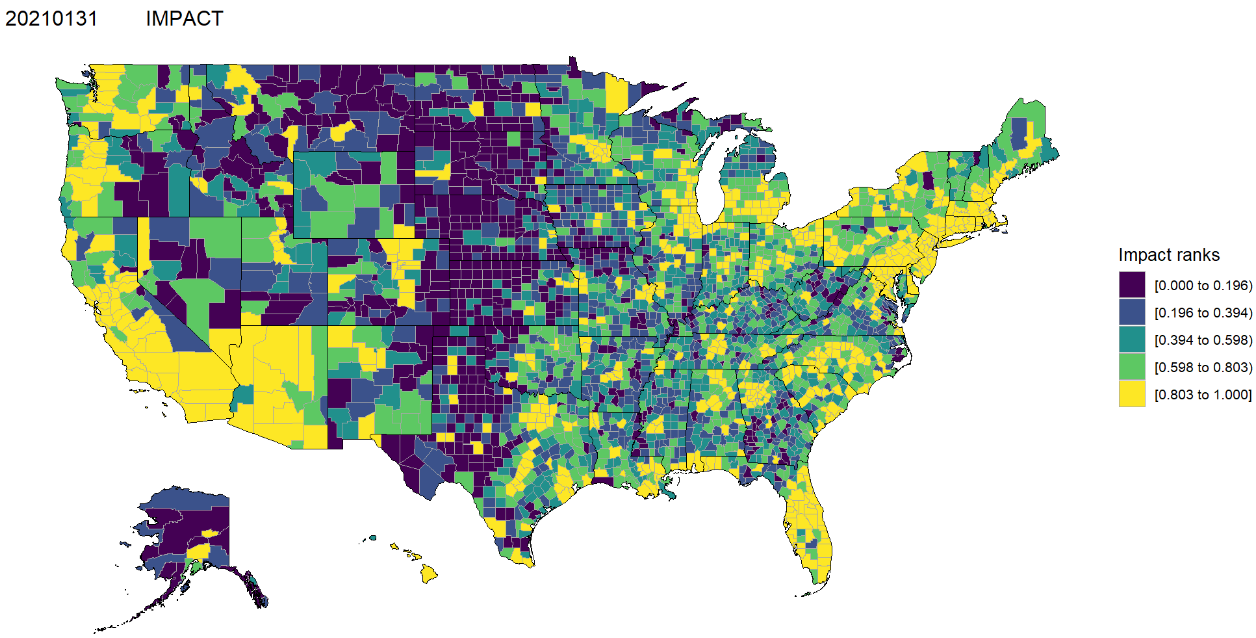


Figure 17.- Impact map as a static category.

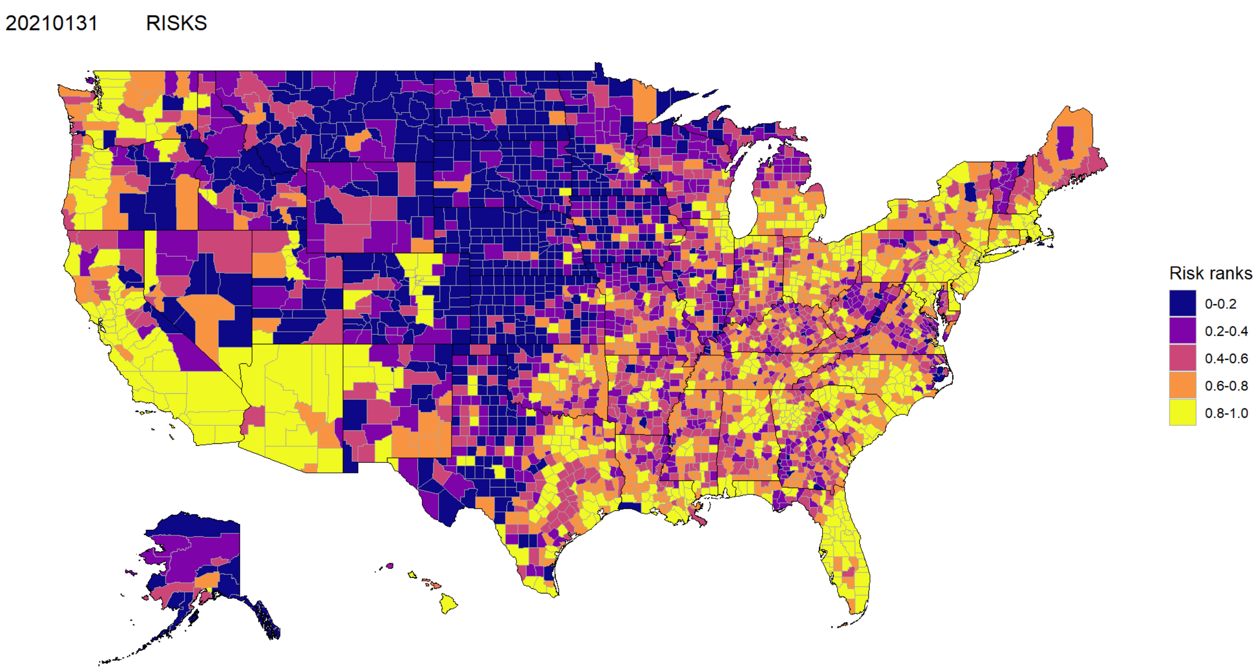


Figure 18.- SGL risk index as a dynamic category.

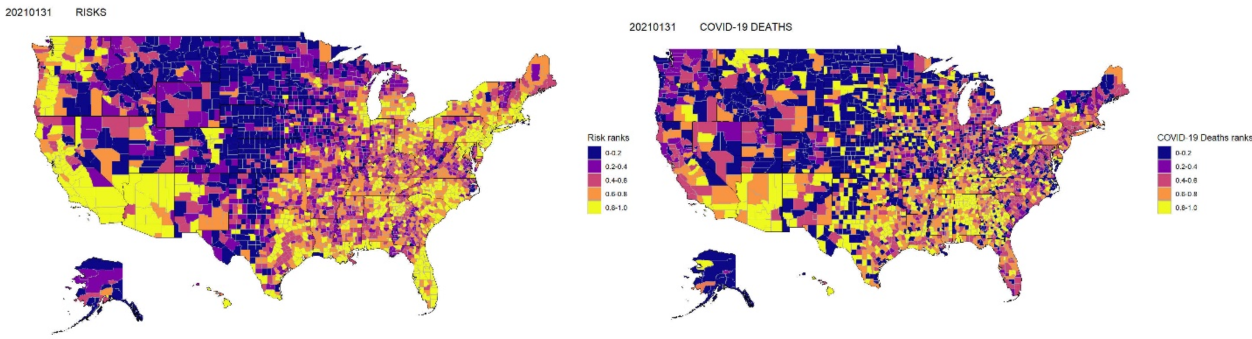


Figure 19.- SGL risk index vs COVID-19 deaths.

Risk communication platform

COVID19 has revealed the impacts of natural disasters by disrupting supply chains globally, causing many businesses to seek alternative means of securing the goods or services they need. The goal of the R13 Risk Communication Platform is to facilitate and grow connections between businesses within supply chains to increase resilience, and provide information on other suppliers, shippers or buyers if an alternative is needed.

A review (Table 4) was conducted of web-based platforms which seek to connect businesses, and this was used to inform design and structure of the platform. A database of businesses searchable by specific criteria which provides contact and location information as well as products and industry was decided upon from this review, wherein a business representative can log on, find a business which can fill the current need they have, and contact the business through the platform.

Platforms reviewed:	**Structure:**
ThomasNet.com	Searchable Database
GovShop.com	Searchable Database
Dnb.com	Products based on data, data not displayed
Nga.org	Unformatted list of businesses
WisconsinSupplierNetwork.com	Searchable Database

Table 4.- Literature Review of web-based platforms.

A few specific components are necessary to facilitate this connection. The platform must have an advanced search feature, a page to display and filter results from the database of businesses, a registration feature for new businesses, and a profile page to display in depth information on each business. In addition, a social or connective element such as a posting feature for advertising sales or purchases, and a messaging feature would be important elements to connect new businesses. This informed our ideas of structure, displayed in Figure 20, of how the platform would be laid out from a technical standpoint.

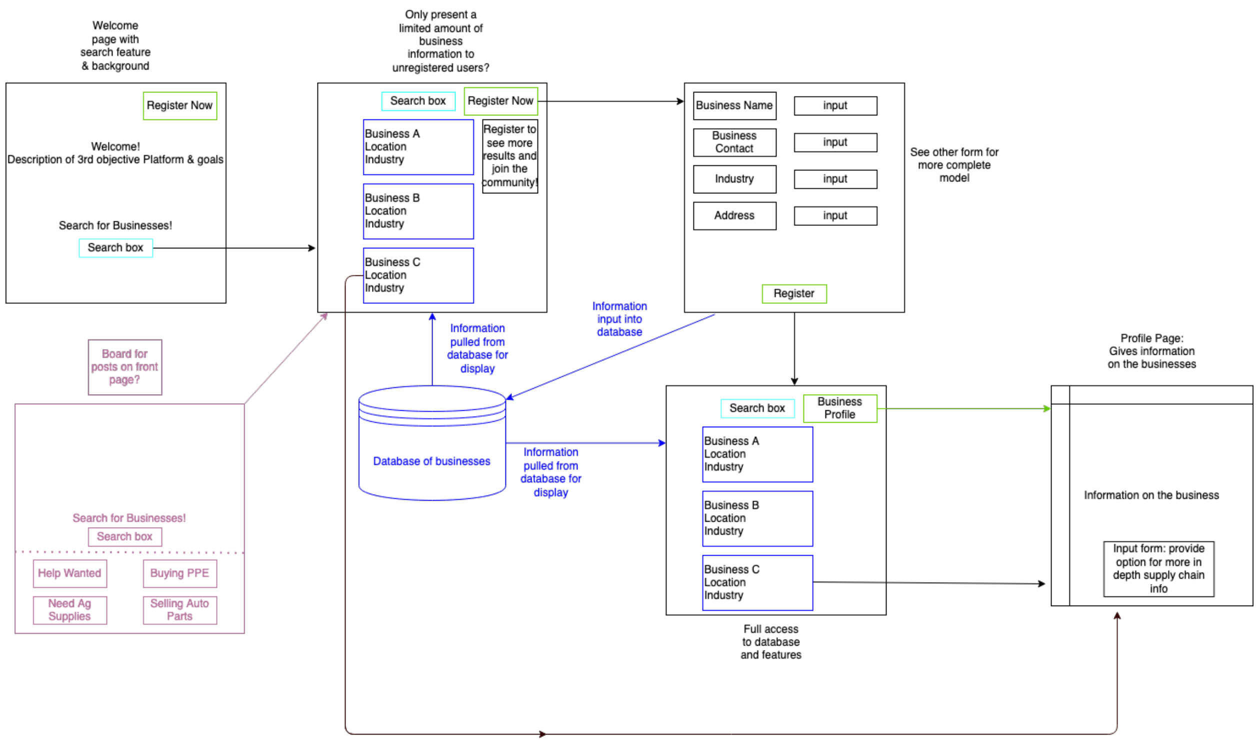


Figure 20.- Technical Structure of Web Platform.

Each number in the structure represents a separate web page in the platform. The flow of a user will start at the landing page (1), which will contain a simple search feature as well as an explanation or mission statement of the project. The draft of the landing page is displayed in Figure 21.

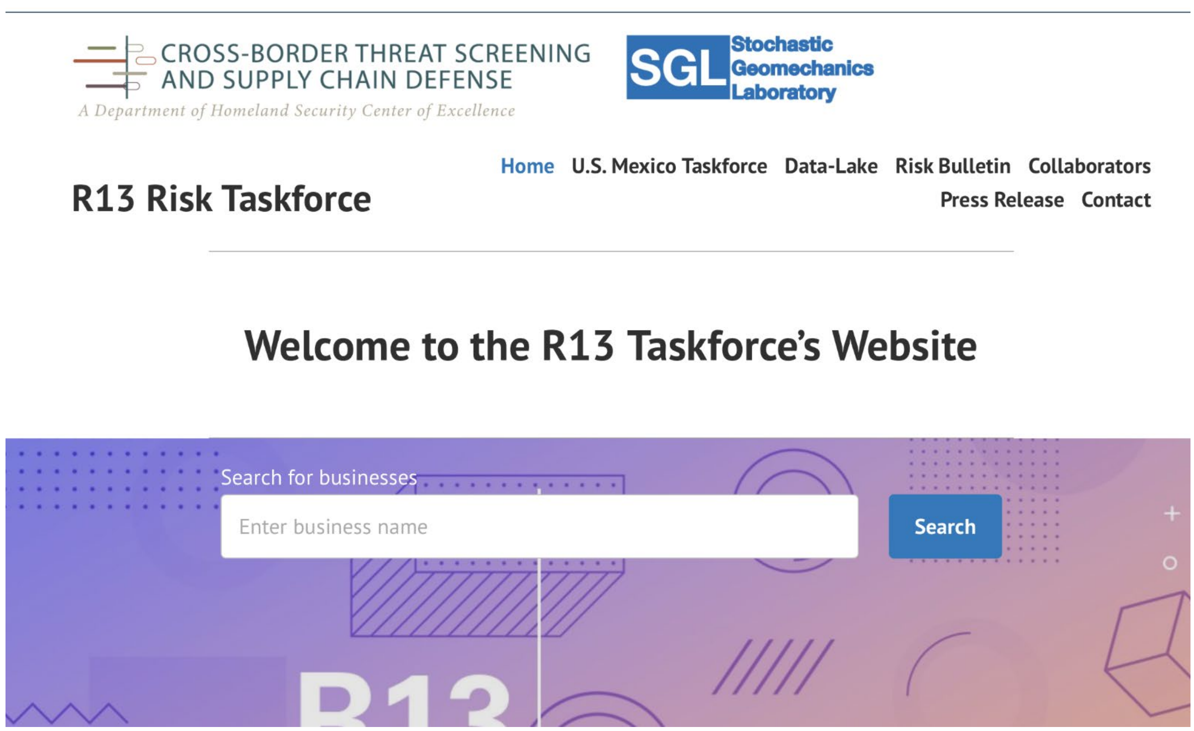


Figure 21.- Landing page.

After a search is initiated, the results will be pulled for display in a list (2) which shows all of the businesses with matching criteria, and each individual card displaying a short description of each business, and which can be selected for viewing more in-depth information about each. This will be the main page of the platform, with each business having its own "card" or description in a small box so that the user can select the most appropriate from the list. Figure 24 is a rough estimation of the structure of the web page.

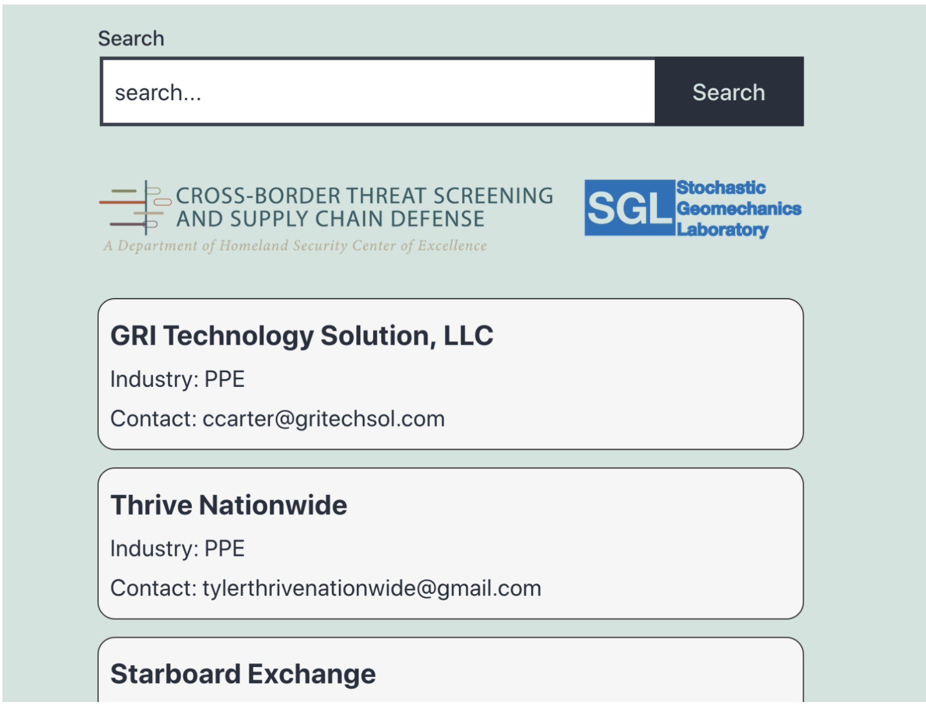


Figure 22.- Database Display Page Structure.

After identifying the structure necessary for such a platform, development started utilizing the tools WordPress, PHP coding, Formidable Forms and XYZPHP to create the tools necessary for each part of the platform. The database was collected and preprocessed from open-source databases of businesses in the medical, agricultural and automotive industries. WordPress was selected for the platform due to the team's familiarity with the tool as well as its customizability. PHP code as well as the Formidable Forms and XYZPHP plugins are used for the functionality of the database display and search features, which make up the backbone of the platform.

The platform is currently in the development phase, with the database display feature finished but the dynamic search feature still being implemented. The PHP code is being tested and developed to deliver the functionality established from the literature review. The next steps are to successfully implement the search feature, create a profile page for each business, and implement a registration feature.

Near-to-real-time Risk Analytics: Semantic Analysis

The goal of this effort is to provide near-to-real-time risk analytics related to information produced in the R13 dashboard. The threats, systems and impacts can be analyzed through the use of Bibliometric Analysis, where the content of published information from various sources reexamined to recognize patterns. This data will assist the team in creating statistics that can be portrayed for users in the dashboard.

The objectives of this effort were to gather information over the following Supply Chains: COVID-19 Vaccination, PPE, Healthcare Workforce, Auto Manufacturing Workforce, and the Agriculture Sector Workforce. In addition, the team intends to automate the process in order to provide the most recent and relevant statistics.

Best practices for conducting Bibliometric Analysis were identified from the literature. This analysis is done through topic modeling, a method that discovers similar ideas in a group of documents. The team established that Latent Semantic Analysis (LSA) would be the best strategy for generating a topic model [10]. In addition, the team has been able to compare different software tools for producing Bibliometric Analysis [11]. Members of the project conducted experiments using several tools in order to observe the benefits and limitations of each one. Through literary review, the team noted which software tools were preferable based on the types of operating systems, programming languages, and visualizations they could offer.

The methodology that the team has set is the following: Data Extraction, Data Preprocessing, Data Analysis, Data Visualization, and Automating the Process. Ultimately, the goal is to create a system that is self-updating and self-serving where our end user is able to visualize and predict changes and activity in the Supply Chain.

For the data extraction phase, members will gather data from various sources such as Panjiva, Eikon, TAMU ProQuest etc. Once the data has been gathered, the material will be organized in blocks of information which will make this more manageable. Using a variety of techniques, the team will categorize the extraction and normalize the gathered data, setting it up for statistical analysis and manipulation. The latent semantic analysis technique will be used in order to consume the stored data. The members plan to create visual graphs in the fourth step that can portray trending topics found in documents, connections in networks, occurrences of authors, occurrences of keywords, and more. For now, several steps in the process involve manual entry from team members, but as mentioned before is moving to full automation and self-service.

Currently, R13 is continuing to investigate and experiment with other bibliometric analysis tools. This work in progress is helping define and strengthen the means of completing this effort. R13 has requested access from Panjiva and Eikon for use of their content, and the team is waiting for their permission.

III. Performance Discussion:

Weekly research updates corresponding to the project's outputs have been delivered in the form of PowerPoint presentations summarizing the project progress by members of the Stochastic Geomechanics Laboratory (SGL). Minutes of every work meeting have been captured on a word document, capturing the key contributions to each Milestone of the project. Monthly summaries have consisted of PowerPoint presentations and word documents of the advances of the project. In addition, CBTS PIs have attended monthly status updates meetings. Moreover, both types of document updates and summaries have been stored in a dedicated Microsoft Teams channel to the project, available to all PIs, graduate students and supporting staff. Content of these documents were used as basis to populate each monthly report. Furthermore, the dedicated Microsoft Teams channels have allowed SGL's team to collaborate, expedite communications, manage information exchange, and improve resources' archiving, among other benefits. The key contributions and tasks by month and milestone have been delivered to CBTS on a monthly basis with expanded details of our progress.

Milestone 1: ***Project kickoff***
Kickoff meeting held on August 26th, 2020

Participants: Tom McGinn, Burke Michael, Alexander Eastman, Heather Manley, Theophilos Gemelas, Wittrock Mark, Chris D. Scarmardo, Caneva Duane, Christine Kim, Beason Valerie, Gregory Pompelli, Matt H. Cochran, Zenon Medina-Cetina, Miriam Olivares, Maria Perez-Patron, Victor Guitierrez, Oscar Sanchez, Guillermo Duran, Enrique Z. Losoya, Juan Pablo Alvarado, Araceli Lopez, Yuliana Razo, Audrey Guzman.

Milestone 2: ***Data Acquisition and Management Plan*** (within 30 days from initiation of work)
Data Acquisition and Management Plan was completed in September 2020.

Data Acquisition from public databases:

- John Hopkins University COVID-19 data.
- Secretaria de Salud México – COVID-19 data.
- CDC Social Vulnerability Index (SVI) data.
- Mexico CDC-based SVI data.
- CENSUS' Community Resilience Estimates (CRE) data.
- Mexico's COVID-19 Social Vulnerability Index (MCVI) data.
- Google Mobility data.
- US Excess Deaths data.
- Mexico Excess Deaths data.

Data Acquisition from private databases (see Table 5 for extended description):

- Panjiva.
- Thomson Reuters's Eikon.
- Texas A&M Library system

Type	Database/Resource	Description
General DB	Web of Science	Comprehensive Citation Index for multiple disciplines
	ProQuest	Databases with access to dissertation and theses, eBooks, newspapers, periodicals, historical collections and other aggregated databases
	ProQuest TDM Studio	Pending application for academic access. This service provides a direct access to the whole collection of ProQuest's databases using R & Python scripting programming languages to accelerate the automation and ingestion of copyrighted cleared and full text of News from major national newspapers and publications.
News aggregator	Factiva	Current international news database produced by Dow Jones, one of the leading global providers of economic and financial information. combines over 32,000 sources to give students, faculty, and librarians access to premium content from 200 countries, in 28 languages.
	Nexis Uni	Nexis Uni features more than 17,000 news, business, and legal sources—including U.S. Supreme Court decisions dating back to 1790.
	News Bank Inc	News Bank consolidates current and archived information from thousands of newspaper titles, as well as newswires, web editions, blogs, videos, broadcast transcripts, business journals, periodicals, government documents and other publications
	Newspaper Source Plus EBSCO	Access to full-text newspapers
Industry reports and statistics	IBIS World	IBISWorld provides reports on industries, including statistics, analysis, and forecasts.
	U.S Bureau of Labor Statistics	Principal fact-finding agency for the U.S. government in the broad field of labor economics and statistics and serves as a principal agency of the U.S. Federal Statistical System.
	Statista	Database that contains international statistics, facts, and market data taken from free and proprietary sources.
	BCC Research	Market research reports that includes major economic, scientific, and technological developments in industrial, pharmaceutical, and high technology organizations.
	RKMA	Market research handbooks focused on various consumer-related markets. Each handbook includes market forecasts, sector trends, and statistics.
	Fitch Solutions	Country and industry business forecasts, company profiles, personnel listings, risk-assessments, growth projections, analysis of the business operating environment, and more.
Financial data vendors	Thomson Reuters Eikon	Software package that provides access to a range of data, and market-leading Reuters news
	Panjiva Platform	Panjiva Inc. is a global trade data company based in New York City. It is a subscription-based website with import and export details on commercial shipments worldwide.

Table 5.- Sources of evidence identified to support the development of risk-based analytics.

Each dataset was presented to the Data-Lake team in order to approve or reject its usability.

Milestone 3: ***Taskforce*** (award + 2 months):
Completed the definition of the Taskforce members, and the first Taskforce meeting was held on October 23rd, 2020.
As of June 30th, 2021, a total of thirteen Taskforce meetings took place.

Milestone 4: ***Data-Lake*** (award + 6 months):
Data-Lake design, development, implementation was completed on February 2020 for both on-premise and cloud-based systems.

Milestone 5: ***Bulletin*** (award + every month):
As of June 30th, 2021, a total of seven Risk Bulletins have been published at R13 project website. April, May, and June 2021 bulletins are still in review and waiting to be released. The R13 team plans to release these bulletins at the beginning of August 2021.

Report card grading scheme methodology was completed on July 2021. The team is working on a few refinements and it's expected to be completed at the beginning of August 2021.
Risk communication platform is currently in the development phase, with the database display feature finished but the dynamic search feature still being implemented. Students are focusing their efforts on finishing this project before the closeout.

Milestone 6: ***Project closeout (award + 12 months)***
The closeout date of the project is August 26th, 2021.

IV. Stakeholder Engagement:

Project champions – Ivan Zapata, Global Health Security Advisor, DHS Chief Medical Officer Team, U.S. Department of Homeland Security, Washington, DC. Tom McGinn, DVM, Senior Health Advisor, Countering Weapons of Mass Destruction, Department of Homeland Security[1].

Mr. Zapata provided guidance and feedback in the last two Taskforce meetings to help R13 team to prioritize research goals for the project.

CBTS has engaged with the project team, and DHS champions, customers, and stakeholders continually; The PIs held weekly interactions through research status updates presentations and meetings with the subcontractors. In addition, students working on the project met as needed to provide research outcomes. Lastly, the binational Taskforce met regularly every other week (twice a month) where project's progress was presented and which created an opportunity to provide advice on the R13's direction.

[1] Dr. McGinn main engagement took place at the kickoff meeting and at the presentation of the R-7 and R13 Data-Lake System shared milestone during the first R13 Taskforce meeting.

V. Transition Progress:

The project poses a significant potential for transition of technology for organizations similar to DHS where decision-making based on mapping of Risk is critical. This project aims at setting the stage to produce an umbrella Data-Lake System for North-America and to concentrate all other research efforts within CBTS. It is anticipated that user-specific platforms will be developed on subsequent phases of this project where full quantitative Risk assessment and management can be conducted, as coordinated with R7 (provided both R7 and R13 are renewed). This opens significant opportunities for transitioning to quantitative simulations of risk scenarios of multiple supply chains.

The inclusion of sector-oriented supply chains is also anticipated as part of R7 and R13 future collaboration, which opens a wide spectrum of opportunities for multiple sectors to trace supplies and services across U.S.-Mexico-Canada ports of entry/exit. The relevance of expanding the composition of the Taskforce to account trade with Canada, and to account for specific sectors is imminent. The collection of Risk-analytics for multiple sectors when presented as inter-dependent supply chains, can offer a very powerful tool which value is anticipated to grow as other supply chains are included into the Risk analytics.

VI. Project Risks:

The major potential limitation for successful completion of the project was budgets' management and the time constraint, which is why this effort was thought of as a 'proof of concept', and has proven successfully to be a high impact high reward project.

The team is confident on finishing pending report card grading scheme, Risk communication platform, and Risk bulletins before the closeout of the project. And to follow a consistent 'on-schedule' approach for all milestones and activities described in the original scope of work.

06/30/2021

VII. Project Timeline:

Texas A&M University						Year 1						
Tasks /Month	1	2	3	4	5	6	7	8	9	10	11	12
Milestone 1. *Project kickoff*	✓											
Milestone 2. *Data Acquisition and Management Plan in place*	✓											
Output 1. Weekly updates, monthly summaries, and quarterly IPRs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Milestone 3. *Taskforce*		✓										
Output 2. Taskforce		✓										
Milestone 4. *Data-Lake*						✓						
Output 3. Data-Lake						✓						
Milestone 5. *Bulletin*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Output 4. Bulletin	✓	✓	✓	✓	✓	✓	✓					
Output 6 Semi-annual report						✓						
Output 6. Final report												
Milestone 6. *Project closeout*												

✓ Actual month of completion.

VIII. Intellectual Property:

Not applicable.

IX. References:

[1] Centers for Disease Control and Prevention (CDC). (2020). "CDC SVI 2018 Documentation." *CDC*, https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/pdf/SVI2018Documentation-H.pdf

[2] U.S. Census Bureau; American Community Survey. "2014-2018 American Community Survey 5-Year Estimates." using *data.census.gov*, https://data.census.gov/cedsci/

[3] White House COVID-19 Team, Join Coordination Cell & Data Strategy and Execution Workgroup. "COVID-19 Community Profile Report." *U.S. Department of Health & Human Services*, https://healthdata.gov/Health/COVID-19-Community-Profile-Report/gqxm-d9w9

[4] Surgo Ventures. (2020). "COVID-19 Community Vulnerability Index (CCVI) Methodology." A*mazon AWS*, https://covid-static-assets.s3.amazonaws.com/US-CCVI/COVID-19+Community+Vulnerability+Index+(CCVI)+Methodology.pdf

[5] Centers for Disease Control and Prevention BRFSS. (2018). "Survey Data & Documentation." PolicyMap, https://www.policymap.com/

[6] Centers for Medicare & Medicaid Services. (2018). "Percent of Medicare fee-for-service beneficiaries who are diagnosed with chronic kidney disease." *CMS*, PolicyMap, https://www.policymap.com/

[7] Health Resources and Services Administration. (2016). "Healthcare data." *HRSA*, PolicyMap, https://www.policymap.com/

[8] Medina-Cetina, Zenon, and Farrokh Nadim. (2008). "Stochastic design of an early warning system." *Georisk*, 2.4: 223-236. https://www.tandfonline.com/doi/full/10.1080/17499510802086777

[9] Cross-Border Threat Screening and Supply Chain Defense; Stochastic Geomechanics Laboratory. “Binational Dashboard”. https://cbts-sgl.engr.tamu.edu/

[10] Prateek, Joshi. “Topic Modelling in Python Using Latent Semantic Analysis.” Analytics Vidhya, 7 May 2020, www.analyticsvidhya.com/blog/2018/10/stepwise-guide-topic-modeling-latent-semantic-analysis/.

[11] Moral-Muñoz, José A.; Herrera-Viedma, Enrique; Santisteban-Espejo, Antonio; Cobo, Manuel J. (2020). “Software tools for conducting bibliometric analysis in science: An up-to-date review”. El profesional de la información, v. 29, n. 1, e290103. https://doi.org/10.3145/epi.2020.ene.03