Open
Geospatial
Consortium

HEALTH SPATIAL DATA INFRASTRUCTURE CONCEPT DEVELOPMENT STUDY ENGINEERING REPORT

———

ENGINEERING REPORT

PUBLISHED

CONTENTS

LIST OF TABLES

LIST OF FIGURES

I ABSTRACT

Experts agree that access to, sharing, and application of location-enabled information is a key component in addressing health related emergencies. While the present COVID-19 pandemic has underscored a range of successes in dealing with the COVID virus, many gaps in supporting local to global preparedness, forecasting, monitoring, and response have been identified when dealing with a health crisis at such an unprecedented level. This study considers how a common, standardized health geospatial data model, schema, and corresponding spatial data infrastructure (SDI) could establish a blueprint to better align the community for early warning, response to, and recovery from future health emergencies. Such a data model would help to improve support for critical functions and use cases.

II EXECUTIVE SUMMARY

Experts agree that access to, sharing, and application of location-enabled information is a key component in addressing health related emergencies. While the present COVID-19 pandemic has underscored a range of successes in dealing with the COVID virus, many gaps in supporting local to global preparedness, forecasting, monitoring, and response have been identified when dealing with a health crisis at such an unprecedented level. A common, standardized health geospatial data model and schema would establish a blueprint to better align the community for early warning, response to, and recovery from future health emergencies. Such a data model would help to improve support for critical functions and use cases.

This Concept Development Study (CDS) aims to engage the health and geospatial communities across industry, government, academia, and research organizations in the evaluation of the current state and future design of a geospatially-enabled Health Data Model and corresponding Health Spatial Data Infrastructure (SDI). To achieve these purposes, this initiative emphasizes the examination of four health related data categories and three health emergency use cases. The results of this initiative include a notional health data model that can be the basis for piloting, prototyping, and use by the global community to improve detection, monitoring, and forecasting. It should also support improved planning, preparedness, response, and recovery for future health emergencies including epidemics / pandemics of infectious as well as environmentally related diseases and other impacts on population health.

A starting assumption of the study was that health information could be usefully organized into four categories including population and patient data; supply chain data; health facilities data; and foundation and contextual data. We also understood the importance of both bio-science data and clinical research data, which although generally not having spatial characteristics, never-the-less determined many aspects of the response to a health disaster, including the development of diagnostics; the design and production of vaccines; and the identification of effective treatments. These six data categories were not challenged and are used as the main building blocks of the health data model.

The CDS has also placed emphasis on the importance of examining the health spatial data value chain. The initiative participants agreed that it should be possible to demonstrate that every component of the data model served important purposes for applications and modeling, actionable intelligence, operations support, and the delivery of health benefits. The participants believe that the value of a data model hinges on the practical utility of the data elements identified.

Request For Information (RFI) responses and expert opinions agreed that many lower income countries lacked basic information about their health facilities and had very limited information about their populations. Often health records for large segments of the population did not exist at all. For these countries progress would entail the initial building of foundational GIS and health information. As a first step it would be important to establish a national basemap within a coordinate reference system, that could be adopted nationally and used by government, the private sector, and volunteer organizations that provide health support and assistance. The common basemap could then be used to identify the location and characteristics for health facilities of all kinds. A number of contributors pointed out that developing a complete inventory of health facilities, and keeping it up to date, would provide major benefits. Other key datasets that could be derived from and represented on a basemap include: the depiction of all transportation networks (roadway, air, rail, water), especially those connecting to health facilities; the identification of population centers ranging from settlements to cities including a focus on slum areas; the identification of clean water and wastewater infrastructure; and supply chains for diagnostics, PPE, medical equipment, vaccines, food, and other essential items. Finally, it was agreed that there should be a health spatial data infrastructure "maturity model" that while starting at the most basic level, provided low and middle income countries with a pathway towards further development and greater usefulness. Responders also agreed that smartphones with GPS could be used to obtain crowdsourced information about health status and local health needs, and could even start to be used as a diagnostic platform.

Responses largely referenced the pandemic use case and focused on the response to COVID-19. Noted was the high rate of viral spread, much of which was asymptomatic, making it very difficult to identify those infected so they could be isolated. Responders also identified the slow deployment of diagnostic tests, delays in getting test results, the slow ramping up of contact tracing operations, the failure to capture precision location information, the inadequacy of supply chains, and the reluctance to share information due to privacy concerns. Among the key recommendations for shaping a health data model included: the development of national address databases and geo-coding applications to ensure that precision addresses of those testing positive could be quickly captured and mapped either to the address point, or rolled up into any geography necessary to support applications and models; the automatic, digital capture of patient information at first interaction with the health system, with that information following the patient through all health system encounters; the compilation and integration of many kinds of demographic, housing, facility and infrastructure information for multiple uses by different segments of the response community; and the development of a geocoding module for contact tracing applications to enable rapid hotspot and micro-cluster identification.

A picture emerged that the cost for making basic health data improvements in countries at all income levels does not need to be great. For lower income countries accurate imagery and the development of a health facilities layer, does not come with a high price tag. For higher income countries, many of which already have spatial enterprise systems, the challenge is to bring together data, much of which already exists, for use in existing applications and models, and to support newer artificial intelligence capabilities. A Health Disaster Pilot Project to be conducted

with Peru will be used to test a number of these ideas within the context of a combined natural disaster event and disease outbreak.

III   SECURITY CONSIDERATIONS

No security considerations have been made for this document.

IV   SUBMITTERS

All questions regarding this document should be directed to the editor or the contributors:

Name	Organization	Role
Alan Leidner	NYC Geospatial Information Systems and Mapping Organization (GISMO)	Editor/Contributor
Mark Reichardt	Open Geospatial Consortium (OGC)	Editor/Contributor
Josh Lieberman	OGC	Editor/Contributor
Anna Gage	Harvard T.H. Chan School of Public Health (HSPH)	Contributor

TERMS, DEFINITIONS AND ABBREVIATED TERMS

1 TERMS, DEFINITIONS AND ABBREVIATED TERMS

This document uses the terms defined in OGC Policy Directive 49, which is based on the ISO/IEC Directives, Part 2, Rules for the structure and drafting of International Standards. In particular, the word "shall" (not "must") is the verb form used to indicate a requirement to be strictly followed to conform to this document and OGC documents do not use the equivalent phrases in the ISO/IEC Directives, Part 2.

This document also uses terms defined in the OGC Standard for Modular specifications (OGC 08-131r3), also known as the 'ModSpec'. The definitions of terms such as standard, specification, requirement, and conformance test are provided in the ModSpec.

For the purposes of this document, the following additional terms and definitions apply.

1.1. Terms and definitions

1.1.1. **spatial data infrastructure**

a comprehensive package of consensus and initiatives required to enable complete provision of data, access and privacy within the territory of the designated infrastructure (source: OGC http://www.opengis.net/def/glossary/term/SDIGDI)

1.2. Abbreviated terms

AQI Air Quality Index

CCVI COVID-19 Community Vulnerability Index

CDC Centers for Disease Control (U.S.)

CDS Concept Development Study

DOH Department of Health

ESF Emergency Support Functions

HIPAA	Health Insurance Portability and Accountability Act of 1996 (U.S.)
HMIS	Health Management Information Systems
ICS	Incident Command System
MSF	Médecins Sans Frontières (Doctors Without Borders)
NIMS	National Incident Management System
OSM	OpenStreetMap
PHI	Personal Health Information
RFI	Request For Information
SDI	Spatial Data Infrastructure
SVI	social vulnerability index

2

OVERVIEW

——

2 OVERVIEW

Clause 3 presents the background, justification, and goals for this Concept Development Study (CDS).

Clause 4 summarizes the questions posed in the Request for Information (RFI).

Clause 5 compiles the responses which were received from the RFI and subsequent response validation workshop.

Clause 6 presents the proposed health SDI data Model

Clause 7 covers the study conclusions and recommended next steps

3

BACKGROUND

3 BACKGROUND

3.1. The importance of data for disaster preparedness and response

Information is an essential part of being able to do almost anything. It is through information that we can pinpoint where things happen, understand what needs to be done, and then support the actual doing. That is why we are now in the "information age" where powerful tools of data creation, organization, analysis and application dominate almost everything that is done in society.

In the realm of disaster preparedness, mobilization, response, and recovery, data that provides situational awareness, a common operating picture, and the inputs into applications, operations, and decisions, has long been understood to be vital to success.

This Health Disaster Concept Development Study (CDS) is based on the certainty that the information associated with a disease outbreak needs to be systematically collected, organized, analyzed, shared, and used effectively in order to make its biggest possible contribution to the support of containment and suppression efforts.

Ever since Dr. Snow's mapping of Cholera cases in London, spatial (e.g., location) data and analysis has been an essential part of epidemiology.
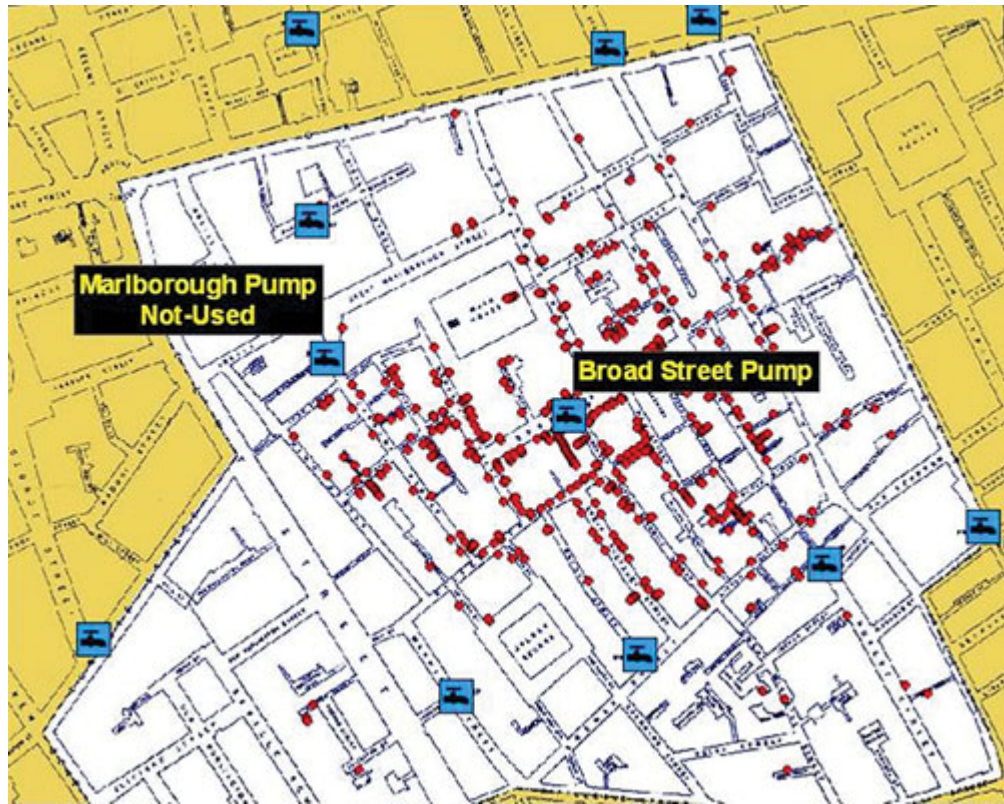
Figure 1 — Dr. John Snow's London Cholera Map, mapped to the address point, 1854

Recently (2000), the New York City (NYC) response to West Nile Fever demonstrated the effectiveness of advanced spatial analytics to successfully contain a deadly disease.
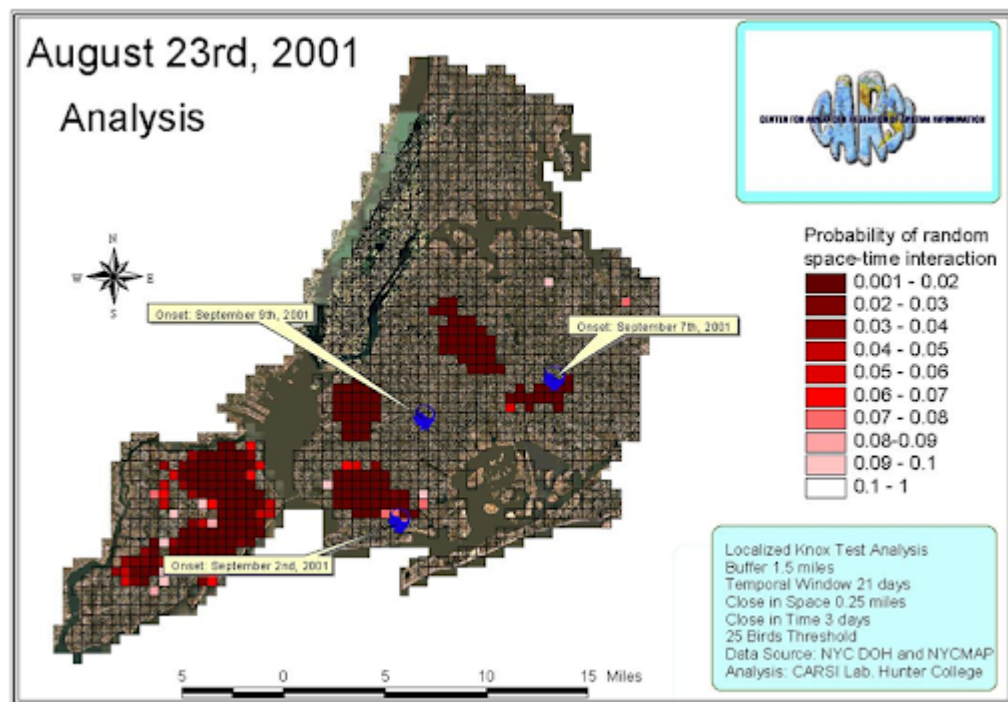
Figure 2 — NYC Department of Health and Dr. Sean Ahearn, Hunter College, West Nile Fever predictive model with address points anonymized and aggregated into a grid for analysis.

Disease is a spatial event, and location plays a role in almost every aspect of efforts to successfully deal with it. Infectious diseases spread from person to person and from place to place. Environmental health emergencies are also determined by the location and dispersion of dangerous substances across populated areas.

A health disaster such as a pandemic or epidemic, or a long-term chronic health condition, occurs against a backdrop of location and time. Where people get sick and how and where they spread disease to others is fundamental to understanding how to contain and suppress it. To answer the question "where" requires the support of a spatial data infrastructure (SDI). In high income countries this could be in the form of a robust national, state and local enterprise-wide SDI, including basemaps, foundation layers, and hundreds of functional layers from different agencies and organizations. Low- and middle-income countries (LMIC) also have national spatial data, but it may lack accuracy and comprehensiveness. It is important to understand the gap between health data needed and health data already on hand. It is also necessary to keep in mind that a large-scale health event generates huge amounts of new data that must be integrated with current data holdings. To manage and use this data properly requires a comprehensive data architecture.

The outbreak of a severe disease affects every sector of society, impacting the economy, jobs, education, government services, food supply, and transportation. Every person whether healthy, infected, or recovered will feel its effects. Therefore the spatial response to a major health event will need personal and population information as well as information about many different types of businesses, facilities, and infrastructures. Among the most important features of this information will be location, because only a common location framework will be able to relate all the information inputs to one another. It is combinations of data, drawn together by a common geography, that will make diverse data sets interoperable, allowing applications, models, and

other spatial analytic tools to create actionable intelligence that supports public health decision making and operations. In short, the integrative power of conferring spatial identity to almost every aspect of a major health event depends upon getting location right.
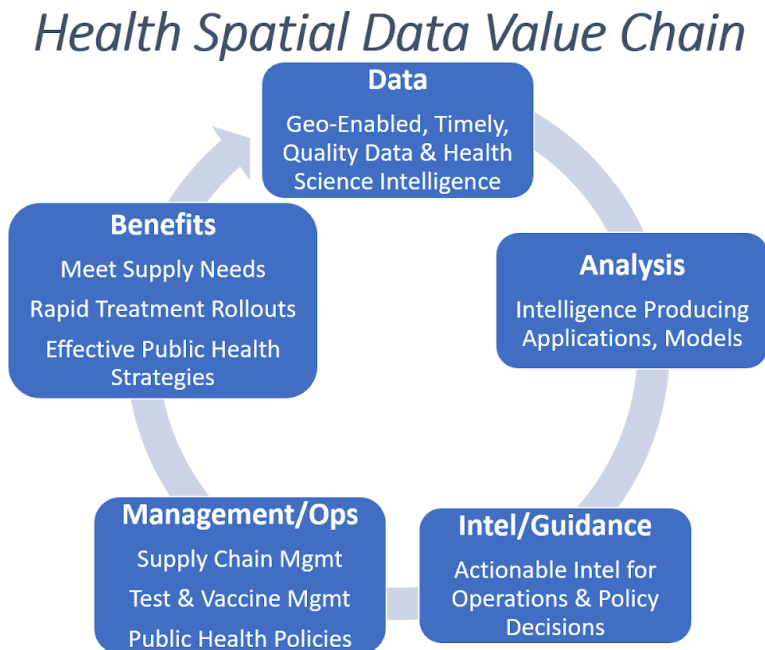
Figure 3 — Health Spatial Data Value Chain.

3.2. The role of a data model

An important step in the development of an information value/supply chain is to identify the kinds of data needed to achieve desired outcomes. A strong spatial data infrastructure for health should allow users to: 1) understand where health risks are located, 2) characterize populations that may be vulnerable to health risks, 3) predict and plan for health emergencies, 4) mount a health system response to address the health risk (i.e., contact tracing or vaccination campaigns), 5) continuously identify all individuals who are infected and recovered to determine disease prevalence and spread patterns, 6) identify all businesses and facilities likely to play a role in disease response, and 7) mobilize other non-health sectors to address the health risks and accommodate response measures (i.e., regulate air pollution emissions from power plants, enhance broadband access in rural areas for at-home work and education).

The listing above identifies a great many types of data and it is important to identify and organize as many of them as possible prior to a health emergency so that data repositories can be populated, data collection tools designed and deployed, and so applications and analytical tools can be designed in advance that make use of the data to produce results. We have found that it often requires data from multiple sources to yield actionable intelligence, and that any one data component is likely to be needed for multiple uses. This puts a premium on data interoperability and the data standards that make it possible.

It is the job of the data model to take that first critical step to specify, organize, standardize, and relate the different kinds of needed data; to identify the critical interactions between data; and to link the data to the tools that will create disease fighting value.

3.3. The COVID-19 Pandemic

The impetus for this initiative was the outbreak of COVID-19 in late 2019 in Wuhan, China and its rapid international spread. As of March, 2021, it is estimated that worldwide, the total number of COVID cases is more than 123 million with the total number of deaths above 2.7 million.https://coronavirus.jhu.edu/map.html[COVID-19 Map — Johns Hopkins Coronavirus Resource Center (jhu.edu)] The pandemic has disrupted regular health system functions, medical and non-medical supply chains, and the global economy.

- **Challenges presented by COVID-19:** There are a number of characteristics that make COVID-19 particularly difficult to control and why standard containment strategies worked poorly.

  - **Rate of spread:** As calculated by the Centers for Disease Control and Prevention (CDC), the R0 value for COVID-19 was estimated to be 5.7 with a doubling time of between 2.3 to 3.3 days. COVID is about three times more infectious than the seasonal flu. **Data Implications**: Because COVID spreads at a very high rate, it is essential that data be rapidly collected, assembled, analyzed, and turned into actionable intelligence, or run the risk of being irrelevant.

  - **Asymptomatic spread**: Many infected individuals never have symptoms and yet are able to spread the disease. Others who eventually show symptoms can be both asymptomatic and infectious in the days leading up to feeling ill. **Data Implications**: In the case of COVID-19, understanding disease prevalence requires information about both symptomatic and asymptomatic cases. This makes essential extensive sample collection, testing, and speedy data turnaround so that infected individuals, whether feeling sick or not, can be asked to isolate themselves and to identify their recent contacts.

  - **COVID-19 targeted the most vulnerable**: COVID proved to be particularly deadly to those sixty-five years and older and to those with pre-existing conditions such as obesity, diabetes, compromised immune systems, and heart problems. A large percentage of deaths occurred in group facilities such as nursing homes and senior care centers, likely carried into those facilities by asymptomatic visitors and staff. **Data Implications**: As a routine public health measure, it is essential to identify all group facilities and public places where a highly infectious disease would be expected to spread rapidly.

  - **Spread Pattern**: COVID-19 was found to spread in multi-generational homes and in dense apartment complexes. **Data Implications**: Census data and population health data needs to be used to anticipate the spread

of the disease to properly prepare and protect neighborhoods likely to be hit hard.

- **Health Supplies**: During the first months of the COVID-19 outbreak, and even extending to the present, the need for testing kits, PPE, ventilators, medical supplies, exceeded supply because of the rapid increase in the number of cases. Since COVID-19 quickly became a global pandemic, countries which normally produce these supplies kept them for their own internal use. An international scramble ensued and patients and medical personnel were put in jeopardy.

- **Data Implications**: It is essential that nations maintain a complete data inventory of essential health equipment and supplies, including where they are manufactured and where they are stored, and be ready to ramp up production quickly. This will require maintaining awareness of all critical supply chains starting at the raw materials stage through final delivery of finished products.

- **Health Facilities**: COVID-19 is characterized not only by rapid national and international spread, but by the intensity of spread within specific communities. This has resulted in some hospitals, medical personnel, and emergency response workers becoming overwhelmed even as hospitals in other parts of a jurisdiction had spare capacity. **Data Implications**: Vital information about all health facilities including staff and supplies, must be maintained so that patient loads can be properly balanced, ensuring the highest level of treatment possible.
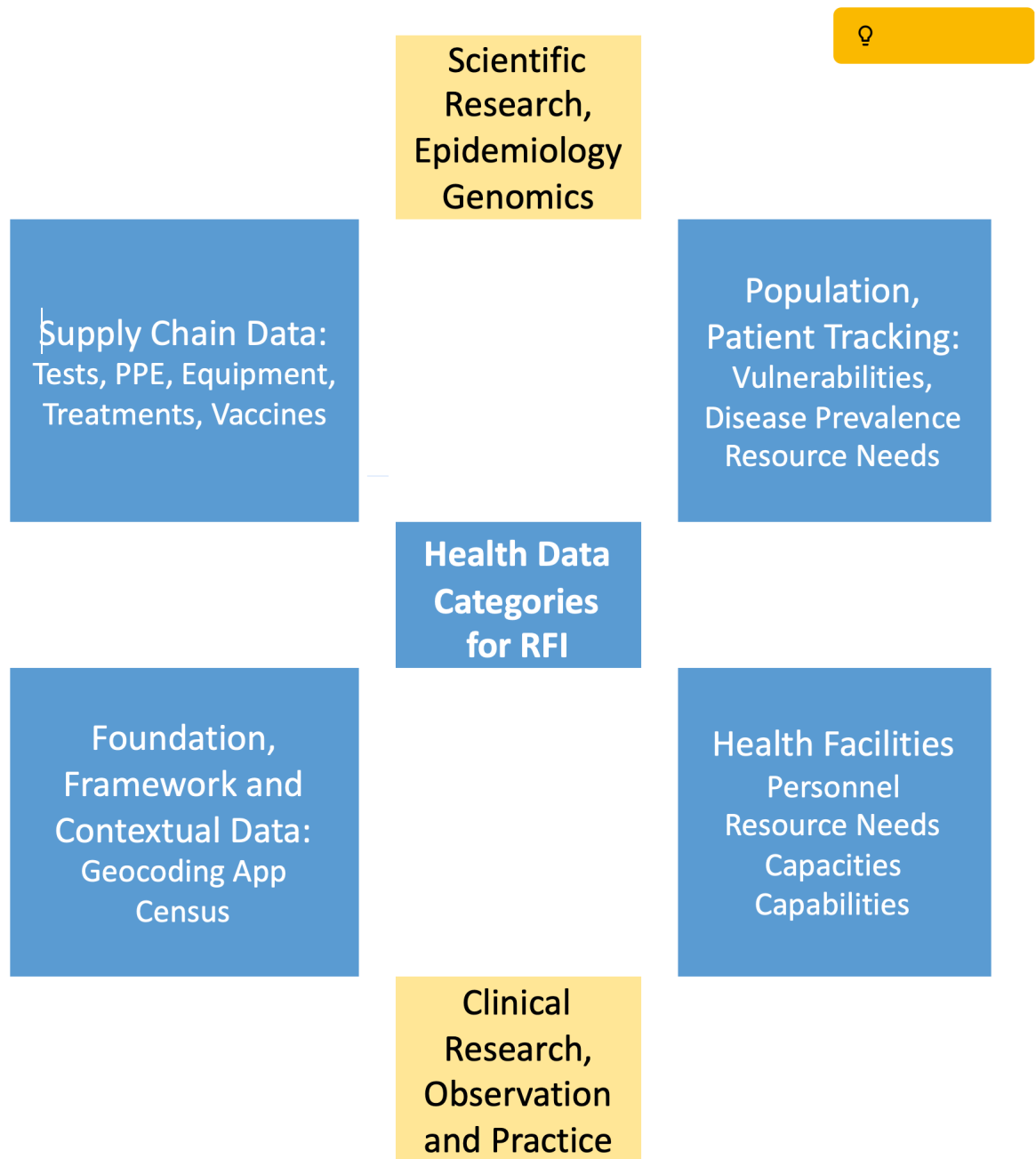
Figure 4 — Major health data categories

- **Categories of health-related data:** The Study Team identified broad information categories that would best characterize the types of health related data that would be needed for a Health SDI. Identifying these categories was a key step towards issuing a Request For Information (RFI), and was also the start of an effort to envision a health spatial data model.

- **Conclusion:** The COVID-19 pandemic exposed many of the existing challenges in gathering, sharing, and using spatial data to plan for monitoring and responding to health emergencies around the world. COVID-19 required data elements that were not previously considered as part of a health spatial data model (i.e.,

social gathering places), near real time data, and integration of data across many different platforms and sources which may have differing privacy requirements. Many of these challenges were present before the COVID-19 pandemic and will continue to be present for future health emergencies without action. The goal of this report is therefore to take stock of challenges and propose solutions and standards that will enable a stronger spatial data infrastructure for health.

3.4. The Challenge of a Health Spatial Data Infrastructure (SDI)

Experts agree that access to, sharing, and application of location-enabled information is a key component in addressing health related emergencies. While the present COVID-19 pandemic has underscored a range of successes in dealing with the COVID virus, many gaps in supporting local to global preparedness, forecasting, monitoring, and response have been identified when dealing with a health crisis at such an unprecedented level. A common, standardized health geospatial data model and schema would establish a blueprint to better align the community for early warning, response to, and recovery from future health emergencies. Such a data model would help to improve support for critical functions and use cases.

A major impetus for this Concept Development Study has been the continuing difficulties with acquiring and operationalizing data associated with the COVID-19 pandemic. Data collected at different levels of government are often not standardized, integrated or interoperable. Monitoring of critical supply chains has been an enormous challenge and there have been severe shortages of vital equipment and supplies for protracted periods of time. Patient data is often not digitized and geocoded at first contact with the health system – at test sites for example – making precision mapping and analysis of disease spread almost impossible. Health infrastructure data has not been comprehensively assembled, hindering the development of situational awareness and a common operating picture. A Health Spatial Data Infrastructure designed to address these issues, then extended to support other kinds of diseases and health problems, will raise the efficiency and effectiveness of health services, saving lives, protecting the public, and saving money.

3.5. Study Purpose

This Concept Development Study (CDS) aimed to engage the health and geospatial communities across industry, government, academia, and research organizations in the evaluation of the current state and the future design of geospatial data requirements for a Health Data Model and a Health Spatial Data Infrastructure (SDI). To achieve these purposes, this initiative emphasizes the examination of four health related data categories and three health emergency use cases. The results of this initiative include a notional health data model that can be the basis for piloting, prototyping, and use by the global community to improve detection, monitoring, and forecasting. It should also support improved planning, preparedness, response, and

recovery for future health emergencies including epidemics / pandemics of infectious as well as environmentally related diseases and other impacts on population health.

3.6. Methodology

The methodology applied for this CDS included community engagement that leverages both written and verbal (teleconference meetings and workshops) responses to a formal Request for Information (RFI). Specific steps in this CDS included the following.

a)   Pandemic like COVID-19;

b)   A natural disaster complicated by a contagious disease epidemic; and

c)   Health emergencies in general with an additional focus on the health effects of impaired air quality.

For each use case RFI respondents were asked to assess data requirements in the following categories:

- Useful **indicators**, metrics and measures that support detection, monitoring, forecasting and response priorities;

- **Occurrence and status data** required to identify and trace contacts of infected individuals; data to support critical analyses that identify and monitor disease clusters, identify vulnerable populations, forecast and predict future disease spread, and support effective public information;

- **Foundation and framework data** that provide critical context for health related data;

- **Resource data** related to determining the status and adequacy of health care facilities, services, resources, and staffing; and

- Data associated with **critical supply chains** for PPE, testing, equipment, and treatments.

Based on the responses to the Request For Information and review of related initiatives, analyze and document Health SDI core data model requirements and use cases in a draft CDS Report, including an initial Health SDI data model.

Convene a Health SDI summit session and CDS validation workshop as part of the March 2021 OGC Member Meeting to further engage experts from the health and geospatial communities, and to kick off the model refinement process.

Produce and post for OGC consideration and approval a completed CDS Report incorporating feedback from reviewers and the validation workshop. Include recommendations on ways in

which the results of the CDS, including a notional Health SDI Data Model can be prototyped in the upcoming Disasters 2021 Pilot Initiative.

4

REQUEST FOR INFORMATION AND VALIDATION WORKSHOP

———

4 REQUEST FOR INFORMATION AND VALIDATION WORKSHOP

4.1. Introduction

As part of the study, a RFI was developed and issued to gather all the best ideas for identifying and effectively using spatially-enabled health information through examination of three disease use cases: a pandemic like COVID-19; a contagious disease epidemic that accompanies a natural disaster, for example, the cholera epidemic that followed the 2010 Haiti earthquake; and environmental epidemics such as respiratory illnesses caused by air pollution. It was also determined that the RFI would address the development of a health spatial data model that could be relevant to high-income countries as well as to low- and middle-income countries. By looking at three different use cases across countries at different income levels it was hoped that the resultant data model could be used and adapted for all kinds of health emergencies anywhere in the world.

Responders to this RFI were asked to identify what data, products, and services are needed, the relationship between different types of data, and critical applications for interoperable data. Responders were also asked to provide information about data sharing between health organizations and other emergency response agencies and organizations, and about data safeguards needed to ensure compliance with privacy and other security requirements.

4.1.1. Health Science and Clinical Data and Research:

The role of the international bio-science and medical community is to meet the health needs of all countries and peoples. This community is formed into networks of government research centers, medical schools, universities, hospitals, private health care companies, and individual practitioners. Using the latest scientific and medical tools they are in position to identify newly emerging pathogens, map their genome, and rapidly design tests, treatments and vaccines to protect those at risk and to treat those who have been infected. Medical practitioners working at hospitals and other health facilities, provide care to infected individuals and find increasingly effective methods to reduce spread, severity, and save lives. They interact closely with the public health epidemiologists. The knowledge they develop is available to all countries regardless of income.

While these efforts provide essential understandings about the biology and treatment of a disease outbreak, the focus of the RFI was the spatially enabled data that guides a myriad of health activities and operations across nations ensuring that scientific and clinical findings are applied as effectively as possible. The RFI was not intended to address activities related to early scientific discovery in a health emergency, such as initial pathogen or toxin identification,

determination of exposure characteristics (i.e., origins or modes of transmission), pathogen gene sequencing, or development of diagnostics, treatments, or vaccines.

4.2. The Health Data Foundation

Health related data is the foundation for dealing with any major health emergency. Such health-related data can allow the health community to maintain oversight of critical supply chains (for testing, PPE, equipment and treatments, including antibiotics and vaccines); understand patterns of patient illness and spread; and provide us with intelligence to rapidly mobilize responses and suppress disease outbreaks. Four general data categories were identified as priorities in dealing with health emergencies. They are described here as a starting point for responders to identify what they themselves consider the most important data types and sources needed to support critical health work processes and analytics.

4.3. Patient Data Collection for Analytics, Situational Awareness and Public Information

Perhaps the most critical information for dealing with a health emergency comes from patient interactions with the health system. This information when analyzed, produces the intelligence that identifies where the disease is most prevalent, quantities of supplies needed, and medical resources required. This information is also used in public information dashboards, contact tracing operations and hotspot identification and micro-cluster analysis. Demographic and housing information is used as the basis for developing a social vulnerability index (SVI) that can help predict the geographical spread of a disease.

It is critical that individuals who have become infected and sickened are identified as soon as possible. From an infected individual's first contact with the health system, all the way through sample collection, testing, notification, contact tracing, and case resolution, location-enabled information should be rapidly captured in digital form, geocoded to ensure accuracy, and then aggregated with other standardized information to enable analytics that support disease suppression operations. Additional data are required to define and document a workflow and data flow methodology that fulfills these requirements. Another challenge involves providing security for personal health information, while also allowing this data to be shared with other government agency users who are assisting health agencies to manage the health emergency. This may require interpreting, and perhaps modifying, current privacy requirements for private health information established by different nations (e.g., U.S. HIPAA requirements).

4.3.1. Identifying individuals who test positive for infection

4.3.1.1. Administering Disease Diagnostic Tests:

At testing sites, collect specimens from individuals while also capturing names, home locations, and other relevant information to be identified by RFI responders. Specimens are then sent to laboratories for analysis. Self-administered, rapid tests may not need to be sent to laboratories.

4.3.1.2. Digitizing all Information and Geocode Location Data:

All testing stations should be equipped with the computer equipment and applications needed to ensure that all collected data is digitized and that location information is properly geocoded to an address point or building footprint.

4.3.1.3. Test Processing:

Identifying the laboratory locations where tests are processed to determine positivity. Each test processing location should have a rated daily processing capacity. Information about positive test results are sent to local and state departments of health, and may be shared with contact tracing teams.

4.3.1.4. Reporting Test Results in a Standardized Fashion:

Standardized test data enable results to be aggregated within and across jurisdictional boundaries and used for a wide variety of analytical and operational purposes.

4.3.2. Contact Tracing Information:

Contract tracers get in touch with those who test positive for infection and from those who have been in close contact with them. Various forms of location information are normally collected by contact tracing programs, including home locations, job locations, places visited, events attended, and transportation modes and routes taken. The location data collected by contact tracers should be derived from the use of an accurate base map and from accurate geocoding as needed. It must be possible to combine the data from all cases to identify geographical patterns of infection. This information is essential to identify disease hotspots, and to design containment strategies. One approach might be to retrofit current contact tracing applications with GIS tools that allow all data captured to be fully spatially enabled.

4.3.3. Crowd Sourced Information

Many types of health-related information can be collected through remote sensing, in-situ sensor feeds, IoT devices, smartphones, and other data collection devices and used to assess disease presence and spread characteristics. The use of anonymous Bluetooth-based proximity tracking options has shown promise, but has not yet been widely adopted nor proven its effectiveness in practice.

4.3.4. Data Required for Mapping, Analytics and Public Information

Vulnerable population mapping: Disease infection patterns, neighborhood characteristics, health information, census population and housing data, etc., are used to determine where rapid spread is likely to occur so that effective preventive measures can be deployed. The CDC has also defined methodologies for determining vulnerable populations (e.g., https://www.cdc.gov/nceh/hsb/disaster/atriskguidance.pdf).

4.3.5. Precision Case Mapping and Analysis

The best location data can map infections to an individual address point or building footprint using a geocoding application. In time, information may even be mapped to specific building floors and apartments. This level of precision data can support analytics that detect disease clusters and micro clusters. The precise location of infected individuals if properly collected can then be aggregated into larger geographical regions that are either pre-defined such as zip codes, census tracts, counties, states, and countries; or custom defined on the basis of the analyzed spatial trends. These regional trends might guide special restrictions imposed, for example, on communities with high disease counts as well as policies around the openings or closings of schools, commercial establishments, and other facilities. Such data could also help guide response to rising numbers of infections in specific areas with distinctive socio-economic characteristics, so that disease suppression strategies can be customized to be most effective for a particular neighborhood without being more severe or widespread than necessary. Advanced modeling tools including artificial intelligence can also be trained on this data to better predict disease spread and response effects.

4.3.6. Public Information

While much of this individualized and hyperlocalized microdata is and should be privacy constrained, appropriately aggregated and anonymized health data incorporated into user-friendly dashboards can be valuable for wider official and public awareness. Risk communications with the public are clearly an art and craft. A standard dashboard design could save hundreds of jurisdictions from expending scarce resources inventing their own dashboard versions; opinions vary on this feasibility. Examples of data aggregates currently being published include daily, weekly, monthly totals, seven day averages, and rate per 100,000 for individuals tested, testing positive, hospitalized, deaths, and immune/vaccinated.

4.4. Data Required for Managing Key Supply Chains

Data is required to comprehensively depict and track all the key supply chains involved in a health emergency. These clearly included the manufacture and supply of essential tests, PPE, equipment, supplies, treatments, and vaccines. Spatially enabled information is required for sources of raw material, facilities for manufacturing, assembly, shipping, storage, and delivery, and how they are all connected together. Information about national and international manufacturing capacity and the potential for increasing production in an emergency are also necessary.

The following are important supply chains for health emergency planning, preparedness, response, and recovery.

4.4.1. Diagnostic Test Supply Chain:

Adequate testing kits and support materials and chemicals are essential to identify those who have been infected, especially if they pose a danger of spreading a disease. Examples of test components include cotton swabs, specimen containers and chemicals used by laboratories to determine disease positivity.

4.4.2. Personal Protective Equipment:

Providing protective equipment and supplies such as disposable masks, gloves, gowns, and sanitizer to health care workers and first responders. PPE is essential to protect health care workers and to other essential workers likely to be exposed.

4.4.3. Essential Medical Equipment:

This includes equipment such as ventilators, oxygen delivery systems, hospital beds, refrigeration units, field hospital tents, and related infrastructure needed for proper treatment.

4.4.4. Medical Treatments:

Includes supplies of antibiotics, vaccines, and other essential medicines. Also includes the blood supply chain including blood donation centers, distribution networks, and storage facilities.

Maintaining supply chain situational awareness can make it possible to maintain inventory of needed supplies, equipment, and treatments at optimal levels and to quickly ramp production when necessary. The kinds of data that best identify issues in each supply chain such as inadequacies in raw materials, warehoused supplies, manufacturing and distribution capacity, over-dependence on non-domestic suppliers, may include the following.

4.4.5. Raw Materials and Component Parts:

For each supply chain, this refers to the sources of the materials and parts that go into manufacturing necessary testing, PPE, equipment, and treatments. Data may also include stockpile levels and the capacity to increase production and shipping during disease outbreaks.

4.4.6. Manufacturing:

This identifies places where materials and parts are fabricated to create finished products, the production capacity of these plants, and their ability to accelerate production during disease outbreaks.

4.4.7. Shipping and Storing:

This comprises data about major shipping methods and routes and any special shipping needs, such as a cold or cool chain refrigeration network. Also includes up-to-date information about inventory levels of key supplies kept at storage facilities in relation to recommended levels.

4.4.8. Additional Supply Chains:

Responders are encouraged to identify additional supply chains that are important to the development of a Health Spatial Data Infrastructure and the information, applications, and technologies that characterize them.

4.5. Data Pertaining to Health Care Facilities, Services, Resources and Staffing:

Data is needed about every hospital facility including overall patient capacity, current beds occupied, and the number of beds available. This information needs to be related to predictions of future load within hospital catchment areas. Information about medical staff, specialties, and supplies is also needed. Other types of health related facilities also need to be identified by name, address, functions, and capacities. Examples include clinics, pharmacies, testing stations, laboratories, vaccination sites, and doctors' offices. Because a major disease outbreak can bring other essential societal functions to a halt, it is also important to understand and monitor other critical supply chains such as food and home supply distribution.

4.5.1. Hospital and Health Care Facilities

a)   Patient Occupancy Capacity

b)   Resources: Critical supplies, medicines and equipment: Current levels vs. recommended levels

c)   Staffing Levels

d)   Health care facility catchment areas or hospital referral regions

4.6. Priority Framework, Foundation and Background Data

A wide variety of data, that is not directly health related, is needed to enhance health data and to determine the effects of a disease outbreak on the many different sectors of a country. In jurisdictions and countries with enterprise spatial data systems, many of these data layers already exist and simply need to be brought together. Examples include an address database linked to a geocoding application to provide validated location data for all individuals and facilities as well as facility layers identifying the businesses and institutions whose functions are vital to disease management or to the continuity of essential services. Infrastructure layers can also be useful. A sewer infrastructure layer can support efforts to detect disease spread by testing the wastewater from a building, neighborhood, or campus facility for genetic markers of disease presence. And just as Dr. Snow found, water infrastructure and hydrography is also necessary to understand the spread of waterborne diseases.

Many nations and jurisdictions have enterprise GIS systems, where dozens, if not hundreds, of data layers are registered to a common base map in a way that allows any combination of data layers to be used together to support an operation or solve a problem. The following is a listing of layers that are either essential or likely to be very useful for dealing with major health emergencies and are often found in existing enterprise GIS systems. Responders were asked to identify those of particular importance for a Health Spatial Data Infrastructure and to recommend additional layers that would add value.

4.6.1. Foundation/Framework Data That Provide Spatial Identity to Physical Objects or Boundaries:

a)   Imagery (x,y coordinates and national grid coordinates)

b)   Street centerlines, address ranges and street names

c)   Building footprints, Building IDs, building structure information

d)   Address points

e)   Roadways, transit routes, bus routes

f)   Census block and tract boundaries with population and housing information

g)   Health information by census tract and block

h)   Parcels and parcel IDs

4.6.2. State, County and Municipal Data Sets Useful in a Health Emergency:

a)   Commercial establishments (regulated) including gyms, sports venues, food services, bars, theaters, retail stores and malls, amusement parks, supermarkets, etc.

b)   Sewer systems with points where samples can be taken to detect pathogen presence

c)   Houses of worship and social and recreation centers

d)   Transportation stations, hubs and transfer points, airports

e)   Prisons and detention centers

f)   Community facilities: K-12 Schools, colleges and universities, libraries and other learning centers

g)   Healthcare Delivery centers: hospitals, clinics, community health centers, clinicians' offices including mental health clinicians

h)   Pharmacies, urgent care centers, testing laboratories

i)   Nursing homes and senior care facilities

j)   Health administrative centers and offices

k)   Telecommunications infrastructure including broadcast hubs, smart phone and broadband networks and other essential tele-services necessary to support pandemic communications, citizen remote access to work, e-learning and other key services

4.6.3. National Data, Including Infrastructure Data Encompassing the Entire Nation, Developed by the National Government, and National Datasets Developed by Private Sector and Non-Profit Organizations:

a)   Private sector basemaps with data layers depicting including retail establishments, cultural centers, landmarks, tourist destinations, etc.

b)   Private sector routing applications that include all roadways, traffic direction, and provide real time traffic congestion status

c)   Manufacturing, food processing and warehousing plants with high density workforces

d)   Hospitals and other health care facilities

e)   Drug manufacturing facilities

f)   Health equipment manufacturers

g)   Health research laboratories including those at universities and colleges

4.7. RFI Questions

Specific data questions were posed for each use case.

4.7.1. Pandemic / COVID-19

- What operations, applications, and technologies, and the data they rely upon, are needed to maintain awareness of PPE, test kits and materials, medical equipment, and medical treatments across their supply chains?

- What operations, applications, technologies, and the data they rely upon, are needed to identify and track those infected and to support contact tracing, analytics, identification of vulnerable populations, and public information? Also, how can personal health information be protected and shared securely within the health community and their supporting public safety agencies?

- What kinds of data, not directly related to health care, are needed to provide context and value to a health spatial data infrastructure?

- What kinds of data about the healthcare system should be collected, monitored, analyzed, and tracked?

4.7.2. Natural Disaster Complicated by Infectious Disease outbreaks

- What ways would you suggest to integrate government emergency plans with health plans including the integration of data to support a combined response to a natural disaster and an epidemic?

- How would you treat, hospitalize, and/or evacuate infected individuals while safeguarding first responders and other evacuees? How would you make active infection case data available to the first responder community in ways that also

preserve patient privacy and safety? What is the data required to support these operations?

- What information would you need to obtain or share to prepare disaster shelters so people with infectious diseases are isolated and properly attended to?

- What information would you need to work with in order to ensure the supply of tests, PPE, medical equipment, and treatments during a disaster event?

4.7.3. Health Effects of Air Pollution

- The eventual goal of this Health SDI effort is to build a comprehensive health spatial data infrastructure that encompasses many different kinds of health emergencies, including infectious diseases (TB, AIDS, Malaria, etc.) as well as non-infectious diseases such as heart disease or diabetes, and diseases related to environmental factors such as asthma. Please share your thoughts on what additionally would be needed to build such a comprehensive Health SDI. As a reference point, consider a use case in which air pollution, perhaps caused by wildfires or disaster-related industrial release of toxic materials, causes or exacerbates disease spread.

- What data sets not previously mentioned do you feel are important to include in a health spatial data infrastructure? Are there additional data needs that should be considered at local, state, regional, national, and international levels; and by various health related organizations including government, commercial, NGO, and academia/research? Data should relate to the needs of other health emergencies and can address additional supply chains, different characteristics of disease symptoms and spread, and other relevant framework and foundation data.

4.8. Validation workshop

The Health Data Modeling Team, after reviewing the RFI submissions and selected documents from key health-related organizations, developed a draft study report and summary presentation, then held a validation workshop on Wednesday, March 24th, that included experts from the Health and Geospatial fields.

- Overview of the Health SDI CDS aims, elements, and draft conclusions

  - Goals of the study

  - What is Health SDI and why does it need a common data model.

  - Public health / population health / individual health

- Use case scenarios:

  - Pandemics similar to COVID-19,

  - Natural disasters cascading to or from the spread of an infectious disease; and

  - Epidemics of non-infectious (e.g., air quality related) respiratory illnesses.

- Health data model:

  - Data that comprise useful spatial framework layers and themes for health specific data.

  - Data describing the healthcare delivery infrastructure including hospitals, hospital resources, and hospital staff;

  - Data relating to critical supply chains;

  - Data on populations tested, infected, and treated;

- Health SDI issues

  - Connecting events between people, places, resources

  - Issues of privacy and confidentiality — HIPAA, GDPR

  - Issues of information scale

  - Models of exchange and anonymization for sensitive information.

  - Analyzable (micro) data versus actionable (aggregate) indicators

- Key indicators to implement the scenarios in the study and the pilot

- First breakout sessions Suggested topics:

  - Scenarios validation for common data model

  - Spatial data lessons from the COVID-19 pandemic response

  - Regional and local issues in health spatial data (e.g., country specialization)

  - Health Indicators and interventions

  - Privacy, propriety, and utility

  - Microdata, geocoding, aggregation

  - Maturity models for health SDI (c.f. UNGGIM, SDG's)

  - Interplay between public health management and emergency management

- Breakout reports and next breakout topics

- Health SDI CDS second breakout sessions (REMO tables)

  - Topics to be selected by participants: e.g., data model components, next steps, avenues of engagement

- Second reports, next steps, and wrap up

5

RFI RESPONSES AND WORKSHOP OUTCOMES

———

5 RFI RESPONSES AND WORKSHOP OUTCOMES

5.1. RFI Respondents to the RFI

Respondents, either via written submissions or interviews, included the following individuals and organizations.

- Pandemic GIS Task Force: NSGIC, URISA, NAPSG

- NYS Regional Pandemic Committee: NYSGISA, NYC GISMO

- Io Blair-Freese: The Bill and Melinda Gates Foundation

- Raphael Brechard: Doctors Without Borders (MSF)

- Steeve Ebener: Health GeoLab Collaborative

- U.S. FEMA: Pandemic Response to Coronavirus Disease 2019, Initial Assessment Report

- U.N. Economic and Social Council: Geospatial data taxonomy for the sustainable development goals of Africa

- Prashant Hedao: World Health Organization

- Ajay K. Gupta: HSR.health

- Benjamin Malaya: International Community Foundation (ICF)

- Dana Thompson: IDEAMAPS Network

- U.S. Center for Disease Control: GIS Program Descriptions

- Professor Sean Ahearn: CARSI Lab and Hunter College Department of Geography

- Dr. Marcia Castro and Dr. S.V. Subramanian: Harvard T.H. Chan School of Public Health

- Eddie Oldfield: Quality Urban Energy Systems of Tomorrow

5.2. RFI Responses: Key findings, and Recommendations

In the following section we review the major findings and recommendations that we found in responder submissions and through interviews.

5.2.1. Bio-science and Clinical research

It should be noted at the beginning of this section that shortly after the initial outbreak of COVID-19, scientists were able to identify the virus, map its genome, and then within a matter of days, design COVID specific tests and begin the design and development of vaccines. On the treatment side, clinicians discovered that many COVID deaths were caused by cytokine storms produced by the infected person's own immune system, and began to use steroids as an effective treatment for some of the worst cases. These and many other findings continuously provide critical inputs into the public health response to COVID-19 and to every other disease. However, this essential information does not have a prominent spatial component. Therefore, these data sources will not be considered in detail for this report, but will be included, for reference purposes, in the data model.

5.2.2. Population and Patient Data

Many high income countries, along with their highly populated municipalities have advanced enterprise spatial data infrastructures, including hundreds of data layers identifying homes, properties, streets, buildings, businesses, utilities, and other facilities. This data is utilized as inputs for business applications that support government and private sector operations. These enterprise geospatial information systems have the power and flexibility to respond quickly to any emergency or disaster event. They can also manage the information generated by such events so it is rapidly and accurately collected, validated, compiled, integrated, and used to support response activities. When dealing with a health emergency, enterprise spatial systems can be used to pinpoint the location of people who test positive for a disease, and to relate individual patient data to service areas, critical facility catchment areas, and as the means to track disease prevalence and spread.

At this time some national health systems, such as the one in the United States, are splintered into many different components, operating within their own isolated silos. Consequently, it is exceedingly difficult to bring data together during a widespread disease event. Standardizing the location fields of health data bases would provide common data elements to support integration and interoperability. One excellent way to achieve this goal is through the development of a national address database and the implementation of interoperable geo-coding applications.

5.2.2.1. Comprehensive address database and geo-coding application:

A number of RFI responses, especially from GIS professionals at the state and local level cited the criticality of a comprehensive address database and a geo-coding application almost always associated with a photogrammetric basemap so that each address point has a precise geospatial

(x,y) coordinate associated with it. An address database is the compilation of all valid street names and addresses in a jurisdiction, each address associated with a building or property by an accurate coordinate and my unique identification numbers.

When an address is entered into an application that is connected to a geo-coding tool, the geo-coder corrects any errors in the spelling of the street name, and ensures that the address number is valid. The application prevents bad addresses from being entered into the database and ensures that all the addresses are reliable and can be utilized for analysis.

Geo-coding applications also contain digital boundaries for geographies of importance that can include block, census tract, neighborhoods, zip codes, health districts, and other administrative and election borders, and jurisdictional boundaries. This enables an address point to be automatically associated with all boundary areas, and allows information associated with address points to be rolled up into any of those areas. For example: by identifying and combining all the address points for positive disease tests captured within a census tract, census block group, or census block; the number of infections can be related to the socio-economic and housing data collected in the census. This serves a number of useful analytic purposes.

It should be noted that address databases and geo-coding applications are the information backbones of emergency response dispatch systems in many countries. During the COVID-19 pandemic, the U.S. CDC has approved providing dispatch centers with the addresses of infected individuals so that first responders can take precautions. Geo-coding applications are also a key component of almost all spatial applications to be found at state and local levels. Without geo-coded addresses the error rate of address data can be well in excess of 15 percent, and require enormous amounts of personnel time to correct. In the U.S., the inability to automate the collection and geo-coding of patient information is a major reason why zip codes were the go-to boundary areas for representing COVID-19 statistics and why zip code data could not be broken down into smaller areas for more detailed analysis. See: "Stop using zip codes for Geospatial Analysis" www.carto.com/blog/zip-codes-spatial-analysis/ Matt Forrest, August 28, 2019.

5.2.2.2. Create digital patient records at first patient contact with the health system:

Contacts with COVID-19 pandemic patients are marked by a series of interactions and data exchanges with the health system. These interactions occur at test stations, laboratories, medical care facilities, and vaccination centers. Often, patient information, including address, needs to be collected or re-entered manually at each of these points. A number of RFI submitters strongly recommended that personal information be collected and geo-coded at a patient's first contact with the health system and follow that patient – possibly through the use of a unique ID number – to avoid re-entering the same information over and over. By capturing patient information digitally at the first possible moment, the data can be made immediately available to public health officials who can rapidly analyze it, map it, model it, and use it to detect hotspots and to anticipate disease spread. With a fast moving disease, speed of data collection can be a life-saver.

Several RFI responses also suggested that a means be found to provide all citizens with a digital means to quickly and securely provide personal information, including their geo-coded address, so that data entry can be instantaneous, avoiding delays and minimizing the exposure of health

care workers to potentially infected individuals. This information could be in the form of a QR code issued from a secure government website and downloaded to smartphones or smart cards.

5.2.2.3. Access personal health records and population health studies:

In discussions about the information needed to better manage a large-scale health emergency, it was noted that there are instances when being able to access personal health records and population health studies could be valuable in anticipating who is likely to get severely ill and where large concentrations of vulnerable individuals are concentrated. This information could be very useful for public health planning for communities and the hospitals that serve them. It can also be used to effective vaccination strategies.

5.2.2.4. Use precision location information for analysis and modeling:

If personal information, including geo-coded addresses, is captured digitally at the first patient contact with the health system, it opens up an enormous number of analytic and modeling possibilities. At the hyper-local level, precision data mapped to an address point or building footprint can show which individual buildings are experiencing rapid spread of a disease and where high rates of disease incidence are occurring within blocks of single family homes. This information can be used for early detection of developing hotspots or micro-clusters, supporting containment actions. It can also be used to protect first responders dispatched to those locations. In the future, point address data may be enhanced by building floor and apartment information. In a twenty-story residential building, it can be extremely useful to identify a disease cluster on the top three floors.
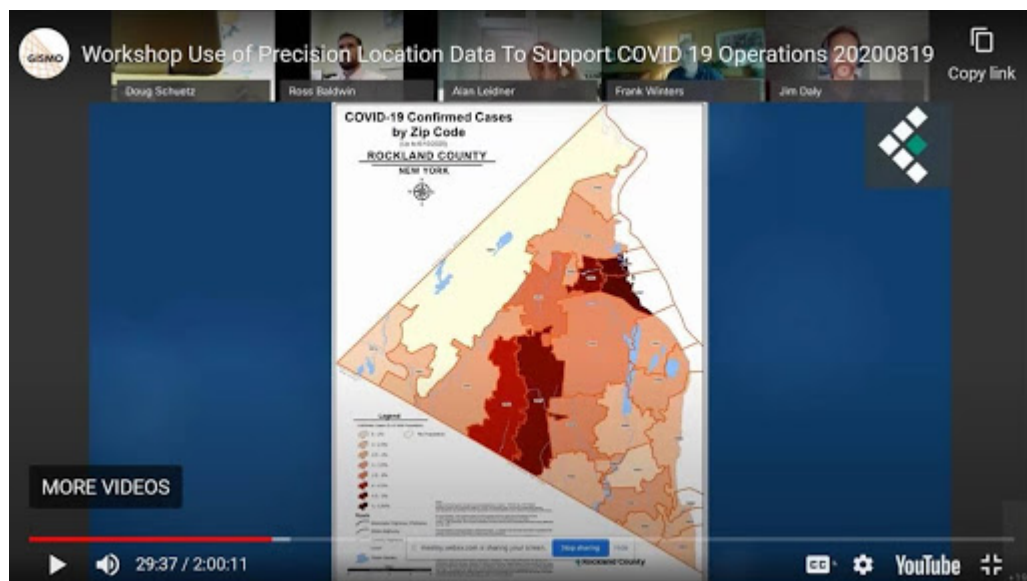
Figure 5 — Map showing COVID infections by Postal Code. From Rockland County, Doug Schuetz, Acting Commissioner for Transportation and Planning.
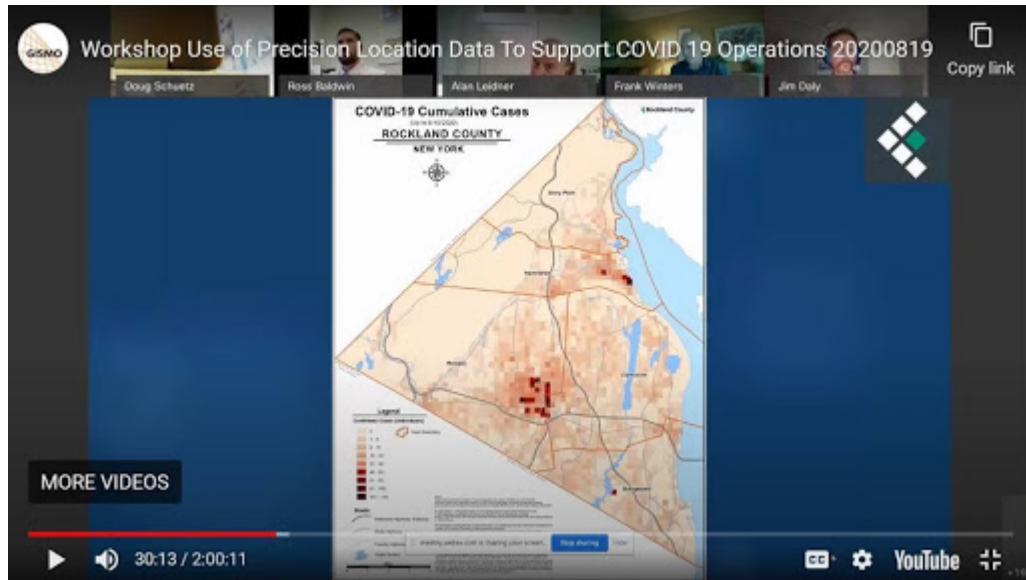
Figure 6 — Map showing COVID infections by micro-grid. For public health purposes, this map provides a far more actionable level of detail. From Rockland County, Doug Schuetz, Acting Commissioner for Transportation and Planning.

Also, with highly-accurate, precision location data, it is possible to quickly group infections into customized areas that can be delineated on the fly enabling disease counts to be calculated quickly. Geo-coded case data makes it possible to assemble and calculate data within any geography chosen, to support the requirements of dashboards, applications, and models. Geo-coded location data will also support data aggregation at the grid level. This is useful if you want to anonymize small data catchment areas.

The maps above from Rockland County demonstrate this principle: disease incidence data assembled by zip code shows a monolithic rate across many dozens of blocks. Disease occurrence assembled by micro-grids shows the same number of cases, but reveals small, intense clusters covering a small fraction of the zip code area. Clearly, the data aggregated to micro-areas presents a more precise target for actions aimed at suppressing spread.

5.2.2.5. Contact Tracing

An infectious disease control strategy that puts trained health personnel in touch with those who test positive for a disease. The contact tracers then collect information about other household occupants and people the infected individual has been in contact with. Contact tracers will then call those who were identified to suggest that they be tested and if they are symptomatic, to advise them to isolate themselves. Contact tracers also collect information about places the patient has visited and events attended. It is essential that all the location information about people and places identified through contact tracing be properly geo-coded. That makes it possible for the contact tracing information from many individuals to be analyzed to identify sources of infection spread such as a popular restaurant, gym, or sports venue. This greatly assists public health officials to develop effective policies related to seating restrictions, closings, masking, and social distancing requirements.

5.2.2.6. Digital Proximity Tracing

An important supplement for traditional contact tracing, based on personal interviews, has been automated: anonymized proximity contact tracing. Bluetooth and GPS equipped smartphones loaded with a contact tracing application identify interactions with others lasting more than a few minutes. If one of the persons involved tests positive, an automatic alert is sent to their contacts advising them to get tested. Digital proximity tracing can be totally anonymous.

5.2.2.7. Privacy Regulations

A number of RFI responses noted that in the United States, Departments of Health (DOHs) at State, District, and Local levels did not share case data with personnel from other agencies. Often Health Insurance Portability and Accountability Act of 1996 (HIPAA) requirements for protecting personal health information (PHI) were cited as the reason for not collaborating with outside organizations. Consequently, DOHs across the country, often overwhelmed by COVID-19 cases and short of staff, were unable to properly geo-code patient data nor perform the kinds of analysis and modeling that could have improved their situational awareness of disease spread and volumes. RFI responders suggest that there be a review of privacy rules to identify ways to have personnel from other agencies more involved with correcting and analyzing precision location data.

5.2.2.8. Social determinants of health and vulnerability

Social determinants of health vulnerability, including education, housing, crime, access to nutrition, and access to healthcare, have a strong spatial component. Mapping combinations of social determinants can help anticipate which neighborhoods will be hit hardest by a disease so that resources can be allocated on the basis of expected need so that residents in vulnerable neighborhoods can be encouraged to take protective measures. Examples of this kind of analysis include the CDC's social vulnerability index (SVI) and Surgo Venture's COVID-19 Community Vulnerability Index (CCVI). SVI is calculated by census tract and can easily be combined with geo-coded precision location data aggregated to census tract boundaries. This technique can be more widely adopted to address other non COVID-19 health risks. For example, the SVI was also used to predict heat-related health outcomes in Georgia.
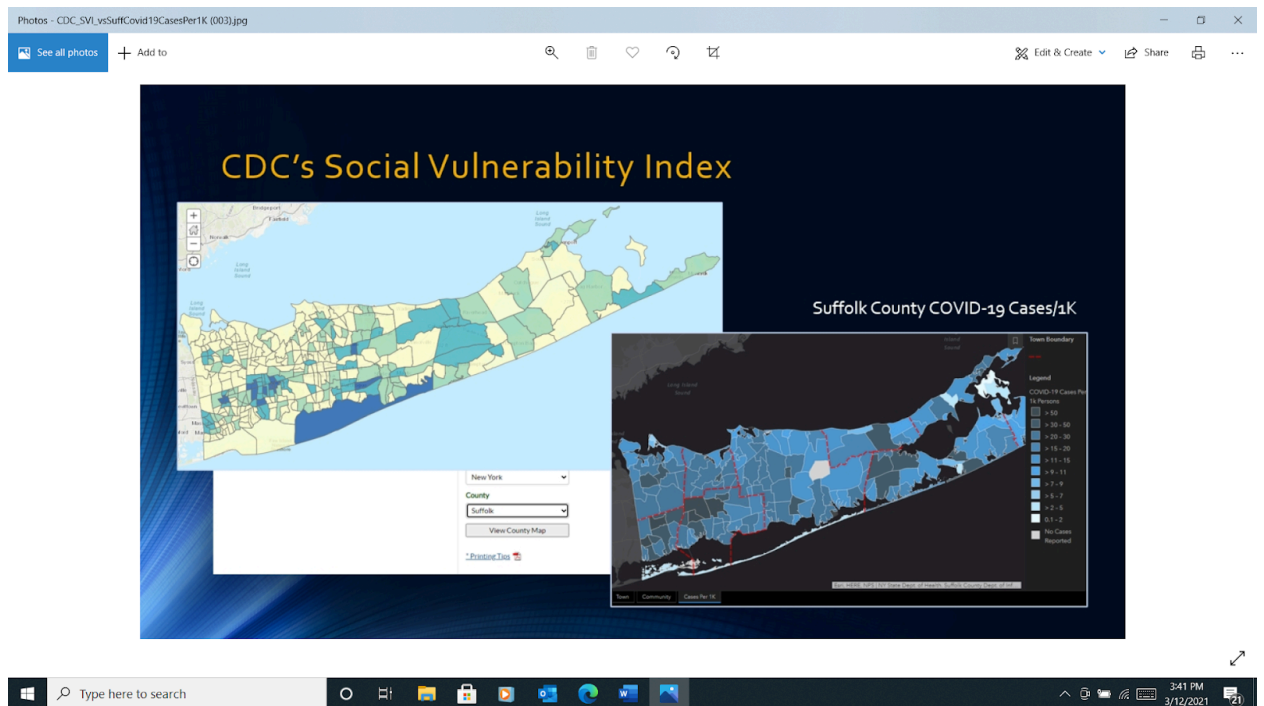
Figure 7 — Comparing Suffolk County COVID-19 cases and the CDC Social Vulnerability Index — From James Daly, Suffolk County, GIS Director.

The study team also encountered references to social resistance indexes, which measure the likelihood that specific groups of people will not comply with public health guidelines about behaviors to reduce disease spread. Public health officials have estimated that many COVID-19 deaths can be associated with individuals who spent time in crowded indoor spaces and who refused to wear masks and to maintain social distance. Social resistance indicators could give public health officials a heads up about groups unlikely to comply with disease avoidance measures. Social Resistance Framework for Understanding High-Risk Behavior Among Nondominant Minorities: Preliminary Evidence (nih.gov)

5.2.2.9. Crowd sourced, big data

Another likely source of valuable information about infected individuals is the data that can be harvested from smartphones, social media, and IoT. For example: the use of thermometers linked to the internet can provide indications in real time of people running fevers. Another example is tracking movements of individuals by private and public transportation to determine whether travel is occurring between areas of high infection and areas of low infection. Another information source can be the volume and location of internet searches about disease symptoms and remedies. Such data can give clues to public health officials about where a disease is spreading and strategies to contain it.

5.2.2.10. Population and patients: Special considerations for low and middle income countries

Low and middle income countries have a number of challenges in gathering spatially enabled health data about their population and patients. First, many low income countries lack individual health records, making individual level analysis of demographics, risk factors, and health conditions challenging. Second, defining health catchment areas and estimating populations within them is challenging in contexts with poor vital registration systems, particularly in urban slum areas and very remote regions. An accurate population estimate is critical for estimating disease incidence and prevalence, identifying hotspots, and mounting a health system response.

There are several groups making progress on these issues in low income countries. First, a health catchment working group led by Médecins Sans Frontières (MSF), Healthsites.io and grid3, is working on defining and updating catchment areas using open source platforms and utilizing crowd mapping techniques. Second, the IdeaMAPs network is developing methods and standards for mapping informal settlements.

Efforts should be made to bring together various information streams including health management information systems (HMIS), remote sensing imagery, cell phone data, social media, surveys, and crowd sourced information. Nationally representative data from the Demographic and Health Surveys and the Multiple Indicator Cluster surveys have been consistently geocoded over the past decade, and are increasingly used with other data sources for more precise mapping of health conditions and risk factors using modeling techniques, for example in mapping measles vaccination rates at 5 x 5 km precision. Satellite and remote sensing imagery has been used innovatively in order to respond to the Indian Ocean Tsunami in 2004, to detect the presence of very remote populations that may lack healthcare access and to estimate excess mortality during the Covid-19 pandemic by focusing on burial plots. Artificial intelligence and machine learning techniques can help fill in information gaps and raise the level of accuracy, reliability, and completeness where data gaps exist.

While many low income countries often have fewer codified regulations on health data privacy, spatial health data should still have protections in place in order to maintain privacy. This may include aggregating data to a higher administrative level using geo-displacement (similar to the Demographic and Health Survey approach), or using differential privacy methods. The Signal Code, developed by the Harvard Humanitarian Initiative, provides useful guidance for a human-rights based approach to data privacy during crises.

Over time, as demand for spatially-enabled information rises, it would be useful for international organizations to agree upon a maturity model that would guide low and middle income countries towards systematically strengthening their spatial data infrastructures in ways that support economic development, government operations, and efficient and effective emergency and disaster responses of all kinds.

5.2.3. Supply chains

Since the beginning of the COVID-19 pandemic in late 2020, it has been clear that nearly all countries have had difficulties obtaining and distributing critical health supplies. The

unprecedented number of people needing to be tested, getting sick, and requiring treatment quickly exhausted supplies on hand and revealed shortages of raw materials and manufacturing capacity. Some examples include the following.

- **Diagnostics:** Testing to identify those with disease is especially important since many people who had caught COVID-19 were infectious but either totally asymptomatic or asymptotic for days before feeling sick. Finding these individuals at an early stage is essential to minimize spread. It is also valuable to know how many individuals have recovered from a disease without being aware they had it. To accurately calculate disease prevalence, antigen tests must be used to identify asymptomatic cases after the fact.

- **Personal Protective Equipment (PPE):** Because the rate of spread of COVID-19 was so high and because a significant percent of those infected required hospitalization, hospitals in areas where the disease was most severe were sometimes stretched beyond their limits. Doctors, nurses, and other health care workers who treated severe cases of COVID had to have protective equipment to protect themselves and to prevent their spreading the disease to others. Supplies of N-95 masks, gowns, sanitizer, and other protective supplies quickly ran low. Foreign suppliers had difficulty meeting demands because of the needs of hospital workers in their own countries. In the U.S., manufacturers did not have the capacity to meet needs.

- **Medical Equipment:** In the early phases of the COVID pandemic, respirators were used to support the breathing of hospitalized cases with severe illness. Oxygen supplies and oxygen delivery systems were also in great demand. Given the need, there were not enough of this kind of equipment to go around and there ensued a desperate effort to find sufficient quantities. This led to competition between countries for these scarce resources and resulted in inflated pricing.

- **Vaccines:** Pharmaceutical companies were quickly able to design vaccines for COVID-19 based on rapid decoding of the virus' genome. However, vaccine production sometimes lagged due to shortages in raw materials and equipment as well as insufficient manufacturing capacity. Also, there were issues with providing enough cold chain shipping and storage facilities for vaccines requiring very low temperature refrigeration.

Respondents to the Health RFI noted the difficulty of getting these and other health supplies. They also mentioned that shortages developed in food supplies and household goods like toilet paper. For example: both agricultural workers and workers at food processing and manufacturing facilities, who often live and work at close quarters, got sick in large numbers hampering production.

It is clear that to more successfully deal with future large scale health events, there must be better supply chain management for all critical health related supplies and for important goods whose production and distribution might be affected by the disease outbreak. This awareness must begin by identifying the steps for the supply chains of each priority product, including the locations for sources of raw material, processing, assembly, manufacture, warehousing, and distribution. Also critical is maintaining awareness of the shipping requirements for products as

they proceed along the supply chain. It would be valuable to identify every facility associated with all critical supply chains in advance of a major health event.

A study made available by the U.S. Federal Emergency Management Agency (FEMA) to the study team described the steps FEMA staff took to implement supply chain management in support of the COVID response. These included coordinating shipments from overseas and working to increase manufacturing capacity domestically. FEMA has supply chain and logistics experts who should lead the effort of documenting important supply chains, specifying appropriate levels of inventory that must be maintained and having plans to ramp up or support production should shortages develop during a major disaster event. The emergency management agencies of all nations should work to do the same. Luckily, supply chain management and logistics are mature business practices that can be adapted to the needs of pandemics and other disaster and emergency events.

For countries with less evolved spatial data infrastructure good places to start the collection of supply chain related information include:

- The complete transportation network which should be accurately and comprehensively mapped including roads, water routes, train routes, and airports; capacities and reliability should be documented;

- Routes generally taken for the shipment of medical treatments, equipment, food, and water should be identified including key transportation hubs and transfer stations; and

- All farms and facilities associated with the growing, fabrication, assembly, and storage of critical supplies.

Additionally, inventory control and GPS tags can be put on health supplies and equipment. All vehicles involved in transporting these supplies should be equipped with GPS trackers so that their movements can be followed.

5.2.4. Health Facilities: High income countries

Medical personnel have been working heroically to care for victims of the COVID-19 pandemic at considerable risk to themselves and their families. Many have worked exhausting hours for extended periods of time. In the U.S., FEMA, DoD, and other federal agencies have been instrumental in putting up field hospitals and providing medical personnel to supplement local capabilities.

In some instances hospitals and treatment centers have operated without full awareness of the conditions in other hospitals within the same region. This has led to one hospital being overwhelmed with cases and undersupplied with PPE due to a local hot spot, while another facility several miles away is fully supplied and has available beds.

There needs to be better situational awareness and plans in place to coordinate a response. New York State was able to achieve a level of collaboration between its private and public hospital systems, shifting patients between hospitals in order to better meet demand.

Other components of the health infrastructure also play important roles in the response to a health disaster. This includes doctors' offices, pharmacies, laboratories, and medical supply stores. These facilities can be used to support a variety of strategies including testing and vaccinations. Each facility should be given a unique ID and be accurately geo-located. There should also be pre-arranged plans to use mobile health vehicles and pop-up test and vaccination sites to reach communities considered healthcare desserts with vital services.

5.2.5. Health Facilities: Special considerations for low and middle income countries

RFI respondents identified a lack of geo-coded master facility lists as one of the major problems confronting the health spatial data infrastructure in many low and middle income countries. With the resurgent focus on primary care in the last two decades, the health facility is the essential unit for launching many health system responses to challenges including vaccination campaigns, health education, and supporting community health workers. Without accurate, real-time access to where public and private facilities are located, health systems are not able to leverage all of their existing assets in responding to health disasters effectively.

There are varying levels of data challenges that health systems are trying to address; most frequently there is some master facility list but it is not geo-coded, it is incomplete, out of date, or lacks any data about the facilities such as ownership (public vs. private) or the services offered. Kenya's master facility list is often used as a strong example that is geo-coded with relevant attributes about the facility, yet it is challenging to keep it up to date with frequent facility closures, reopenings, and health worker strikes. In Chad, disaster responders found that many facilities were open seasonally, and would close during certain times of the year due to accessibility challenges. Finally, informal and unqualified providers are often not captured on master facility lists, which is of particular concern in South Asia where these providers are prevalent. Master facility lists are therefore a dynamic component of a spatial data model that need to be consistently updated.

There are several opportunities for strengthening master facility lists in low income contexts. First, using source platforms such as OpenStreetMap (OSM) can help to keep rapidly changing facility contexts up to date using crowd-sourced knowledge. While many attributes about the facilities may not be available via OSM, users could enter information about which facilities are operational, which then could be confirmed by the Ministry of Health. This would be particularly useful in the case of disaster response where real time information is critical. Second, DHIS2, the platform used for health management information systems in 73 low and middle income countries around the world, has geospatial functions that could be used more frequently. Using the DHIS2 platform to construct and update a geocoded master facility lists is a powerful strategy to incorporate spatial analytics not just about facilities, but also about the volumes of services and mapping outbreaks of disease.

Third, targeted surveys can be used to fill in health facility data gaps. For example, the Health Resources and Services Availability Monitoring System (HeRAMS) survey from the World Health Organization is a new rapid tool used to address the gaps. Its mission is to support countries with the standardization and continuous collection, analysis, and dissemination of information on the availability of essential health services and resources down to the point of service delivery and to strengthen health information systems, particularly through the compilation, maintenance, regular update, and continuous dissemination of an authoritative master list

of health facilities. HeRAMS includes a health facility rapid assessment with georeferenced coordinates. It started off in Yemen and Sudan and is now launching in Ethiopia.

HeRAMS is a particularly useful tool for addressing emergencies and resources in refugee communities because it is rapidly deployable, rather than much more intensive tools. However, it also contains less information about the facilities. More in depth tools such as the Service Availability and Readiness Assessment (former WHO tool) or the Service Provision Assessment (from the Demographic and Health Surveys), the Service Delivery Indicators (World Bank) contain more in depth information about staffing, equipment, services, and quality of care. These tools are currently deployed much less widely than population-based surveys, and could be a very useful tool for Ministries of Health looking to expand their understanding of the health system's capabilities and gaps.

5.2.6. Contextual Information

Those who work on the GIS and mapping side of an emergency response have been known to say that all data layers they created prior to a disaster event ended up being used, while all data that had not been developed was sorely missed. One of the best tools available to respond to a disaster of any kind is an enterprise geospatial system which has organized and integrated all spatially-enabled data utilized by a jurisdiction. In some cities, regions, and states, this could amount to hundreds and even thousands of geo-enabled, and therefore mappable, interoperable data sets, shared among the dozens of agencies. Many of these data layers and the applications that utilize them, while created without thinking of their potential use in an emergency, nevertheless can be repurposed in a disaster, yielding valuable benefits. These layers include the following.

- Foundation layers which provide location identity to all entities within an enterprise spatial system. These include imagery (with x,y coordinates for every pixel), street names and addresses, building footprints, and property parcels. The combination of these and other layers goes into the creation of an address database that can then be used by a geocoding application to support innumerable operations including emergency dispatch systems and customer relationship management (CRM) systems. Most spatial applications tie into a geocoding application to make sure that location information is collected correctly and accurately.

- Census information, including demographic and housing data, is vital to many health, human services, and planning operations and applications within a jurisdiction. Census information in the U.S. is organized by census tracts, blocks, and block groups: geographies that can be related to a jurisdiction's basemap allowing census demographic and housing information to be integrated with local data, including patient data.

- Enterprise spatial systems serve many agencies within a jurisdiction each with its own spatial enabled data needs. Often this means creating layers of facilities that a particular agency supports or regulates. For example: social service agencies regulate day care centers and other group facilities. Criminal justice agencies manage prison, court facilities, and police stations. Infrastructure agencies control or provide oversight to utilities, including water, sewer, and electric systems.

Transportation departments manage airports, subway and bus lines, roadways, bridges, and tunnels. Consumer agencies inspect and regulate restaurants, bars, and grocery stores. Education agencies maintain records of all private and public schools. This list can go on and on. Any of these layers may need to be accessed to support health disaster preparedness and response operations.

- Key environmental data are also required for health disaster response. Examples of such data may include elevation and water bodies, to respond to floods; air quality to address health risks from air pollution; temperature ranges and rainfall amounts, to map the reach of disease vectors such as mosquitos and ticks; and humidity readings combined with land cover to assess fire risks.

It is discouraging when, during a disaster response, significant resources are spent to build critical data layers that, it turns out, already existed but were not known by the response community. It is recommended that every jurisdiction create an inventory of existing data layers and that these listings be shared with higher levels of government. Those that are thought to be useful to any disaster event should be identified and positioned so that access to them can be rapid. Each must conform to standards that guarantees they can be integrated and used interoperably.

5.2.7. Contextual information: Special considerations for low and middle income countries (LMICs)

There are a few challenges particular to contextual spatial data in LMICs. First, gridded addresses are most often used where systems of street names and building addresses do not exist. As more countries shift to individual health record systems, building geo-enabled fields into the platforms will be critical. Several platforms such as Grid3 and WorldPop are developing finely gridded population and infrastructure maps that can be used as contextual layers in health disaster response.

Aggregation from the individual level to a higher spatial unit can also be challenging due to the differences in boundaries between villages, administrative areas (i.e., admin 3 level), or health catchment areas. Sharing data between partners and integration from different sources can be challenging when borders do not line up: it is recommended to use the lowest level of aggregation possible in order to facilitate operability.

Furthermore, in areas with non-Latin based alphabets, different village and district spellings can make unique identification challenging. It is recommended that countries shift to using unique identifiers to identify areas rather than anglicized spellings. Tamr is also working on developing natural language processing to assist in merging different spellings in non-Latin alphabets.

5.2.8. Conclusion

The responses to the Health Data Model RFI demonstrate that there are many measures to take to improve the data available for responding to a health disaster. A measure of the gap that currently exists in gaining access to this information is reflected in a survey taken by the Pandemic GIS Task Force response found in the figure below.

NSGIC, NAPSG, URISA Member Survey
COVID-19 Data Challenge Areas

Challenge Area	% of Entities Reporting This as a Challenge
Lack of Reliable Data	70%
Duplicative Manual Reporting Data Entry	60%
Restricted Data Access and Sharing	57%
Lack of Understanding Models	55%
Lack of Prior Knowledge About Available Technology Capabilities	54%
Lack of Trained Staff	46%
Need to Build Relationships with Technologists/GIS/Data Science	46%
Lack of Access to Technology or Licenses	28%

Figure 8 — NSGIC, NAPSG, URISA Member Survey — COVID-19 Data Challenge Areas

Much still needs to be done to ensure that the right data is at your fingertips when you need it. Ironically, much of this information exists along with the applications and models to derive essential intelligence, however an effort must be made to systematically bring all these data and tools together into an effective, functioning, integrated, accessible, interoperable health information system.

HIGH-LEVEL HEALTH CONCEPTUAL DATA MODEL

6   HIGH-LEVEL HEALTH CONCEPTUAL DATA MODEL

Using responses plus the personal and professional data modeling experience of the study team, a high-end conceptual model was developed to provide a framework for more detailed considerations. It is assumed that each health emergency would require its own, customized data model "profile" which addressed the unique characteristics of a particular disease outbreak.

Figure 9 — Health Conceptual Spatial Data Model

Towards the Development of Code Lists
Items in Bold denote location elements

Geo-Coding App	Personal Health	Health Supplies/Equipment	Health Facilities	Social Vulnerability Index (SVI)	Social Facilities	Supply Chains
Street Name	Sex	Test kits/Diagnostics	Hospitals	**Census Tract**	Schools	Raw Material Producers
Prefix	Age	Lab Chemicals	Clinics	Poverty	Religious Centers	Component Supply
Suffix	**Address**	Syringes	Drs. Offices	Vehicle Access	Social Clubs	Manufacturers
Street Number	Race	Sanitizer	Health Centers	Crowding	Child Care Centers	Warehouses
Floor	Ethnicity	N-95 Masks	Pharmacies	Employment	Hotels	Distribution Centers
Apartment	Household Members	Face Masks	Health Supply Outlets	H.S. Diploma	Airports	Transportation Network
Address Point (x,y)	Height	Gowns	Laboratories	Single Parent	Train Stations	Transportation Modes
Block Number	Weight	Treatments	Test Sample Collection	Native Speakers	Bus Stations	Production Capacity
Census Tract	Underlying	Respirators	Research Facilities	Mobile Home	Prisons	**Location**
Municipality	Conditions	Oxygen Supplies	Research Centers	Group Quarters	Group Homes	Foreign/Domestic
County		Vaccines	Administrative Offices	Income	Senior Living Centers	Delivery Points
District	Contextual Layers	Gloves	Medical Schools	Unemployment	Auditoriums	
State	Imagery	Cotton Swabs	Dental Offices	Disability	Theatres	
Health District	Buildings	Test Tubes	Nursing Homes	(Computer Access)	Bars and Restaurants	
Service District	Parcels		Rehabilitation Centers	(Mobile Phone Access)	Gyms	
Police Precinct	Roadways		Long Term Care Facilities		Sports Stadiums	
Fire District	Transportation Facilities		**Location**	Contact Tracing	Grocery Stores	
Local	Sewer System			Name	Museums	
Neighborhoods	Water System			**Location**	**Location**	
	Retail Businesses			Personal Health Details		
	Food Processing Plants			Household Members		
	Manufacturing Plantsg			Contacts, Names, **Locations**		
	Location			Places Visited		
				Events Attended		
				Routes Taken		
				Transportation Modes Used		

Figure 10 — Towards the development of Health Disaster Code Lists. Items in bold denote location elements that tie categories together making them interoperable

6.1. Interactions between Major Data Categories that Produce Effective Interventions/Uses

A major strength of spatially enabled data sets is their ability to be used interoperably. It is understood that combinations of data and analytics produced for one purpose, or for one area of operations, can be valuable to others. Below, we identify some of the useful interactions within and between different health data categories.

- **Scientific and clinical data**: The mapping of a pathogen's genome will drive the production of effective tests and vaccines. Understanding how a pathogen behaves and how it can be most effectively treated drives a number of different public health operations. It provides key inputs to predictive models that anticipate the volumes of patients that will need hospitalization. It also is instrumental in determining the kinds of treatments that are effective. This in turn determines the key pharmaceutical supply chains and helps quantify the amounts of drugs needed.

- **Use of population and patient data**

  - **Hotspot locations and spread patterns:** Having close to real-time knowledge of the spatial distribution of infected individuals, both symptomatic and asymptomatic, depends upon the geo-coding of their

address data collected at testing facilities. This information can be combined with social vulnerability information to identify current and future disease hot spots and micro-clusters, supporting strategies for disease suppression such as occupancy restrictions and lockdowns.

- **Contact Tracking Applications:** Rapid identification of positive cases enables contact tracers to pinpoint the location of individuals who might be infected, their home and work addresses, places visited, and travel routes. Contextual spatial data layers help contact tracers to more easily find the locations being mentioned and allows data from many contacts to be rolled up so that patterns of disease spread can be identified.

- **Predictive Models (e.g., SIR Model):** Information about patients and their locations, patterns of disease intensity, and pathways of spread allow public health officials to plan for expected patient volumes such as: securing additional hospital bed capacity, PPE supplies, equipment, treatments, and sufficient numbers of medical personnel.

- **Dashboards:** Providing public health information using an intuitive user-friendly design, informs the public about disease spread. Dashboards can also provide the public with information about the location of resources they may need, and facilities they can visit for testing, care, and vaccinations.

- **Use of supply chain data**

  - **Determine products and supplies needed:** The characteristics of a pathogen and the kinds and amounts of equipment, medicine, and supplies needed to successfully deal with the pathogen should be determined as early as possible. Quantifying disease prevalence by location also supports decisions about where supplies should be shipped, ensuring treatment facilities under the greatest patient pressure remain fully stocked.

  - **Supply chain logistics:** Once key products and supplies are identified, their supply chains need to be documented. Facilities at each stage of the production and delivery system will need to be identified. This includes sources of raw materials and facilities for manufacturing, assembly, warehousing, and shipping and the means of tracking supplies through different modes of transportation. This will require the use of a number of contextual spatial layers including transportation networks and a wide variety of facility layers. Supply chains may extend internationally requiring access to foreign data layers.

- **Health facilities management**

  - **Treatment Capacity Assessment and Monitoring:** Based upon the predicted numbers of infected individuals who will need the support of the health system, the capacity of hospitals and other treatment centers can be evaluated. This analysis will provide an early warning if capacities

are expected to be exceeded and allow time for extra facilities to be provided.

- **Sites for Tests and Vaccinations:** The selection of the most appropriate sites for testing and vaccinations will rely on data about numbers of infected individuals, vulnerabilities which point towards future increases in infections, and overall population based upon census demographic and housing data. This combined with mass transit routes and the identification of health deserts will enable good siting decisions based on need, that minimize individual travel time.

Table 1 — Data Types and Examples of Applications, Outputs and Benefits

DATA TYPE	APPS AND TOOLS	OUTPUTS	BENEFITS
Patient & Population	- Dashboards - Models - SVI - Analysis - Contact Tracing	- Awareness - Projected Patients *Vulnerable ID - Hotspot ID - Minimize Spread	- Informed Citizens - Volume and Speed of Spread - Direction of Spread - Containment Strategies - Reduce Illness and Death
Supply Chains	Supply Chain Tracking	Status Awareness	- Manage Supplies - Prevent Bottlenecks
Health Facilities	- Operating Status - Location Analysis	- Mutual Aid - Find Best Sites	- Reduce Overcrowding - Fair Distribution
Contextual Maps & Data	- Enterprise GIS - Geo-Coding Application	- Use Needed Data - Precision Mapping - Map Any Geography	- Rapid Response - ID Hotspots & Spread - Use with Census Tracts - Community & Other Boundaries

6.2. Conclusion

Practically speaking, every application, model, strategy, operation, or decision made related to a disease outbreak will require a combination of data from the patient/population, health facilities, supply chain, and contextual data categories. It should be possible to draw a direct line between any data type and the multiple benefits that can be derived through its use. Figure 6 is one way to illustrate this concept. It is therefore essential that all the data sets that are incorporated into this Health SDI data model be created to common standards so that they can be rapidly integrated.

Additionally, as noted by the RFI response submitted by the PanGIS Task Force: "When designing the connections between healthcare systems and GIS it is critical to not reinvent the wheel." GIS practitioners need to be aware of existing healthcare data standards including:

HAVE 1.0, HAVE 2.0, HL7 V2, TEP. GIS efforts should focus on the spatial location of facilities and include unique IDs to join in data from healthcare data systems to spatially enable the data. Automatic system to system integrations are critical. Too often data systems are connected by staff making extracts of databases and sending them as email attachments to be included in other applications and dashboards.

6.3. Considerations for advancing a Health SDI data model related to combined disaster and pandemic/ epidemic events

Whenever there is a natural disaster, there is always a risk of a disease outbreak. For example, natural disasters that damage water supply or sewer infrastructure may result in the spread of waterborne diseases. People fleeing the scene of a natural disaster may also be exposed to contaminated food and unsanitary conditions. Emergency responders will need to contend with not only people injured by the disaster itself, but those becoming sick from a variety of related causes.

Most responses to the Health SDI Data Modeling RFI did not directly address this type of health emergency, yet we feel that the data model we have outlined so far can be modified to be applicable to this kind of event. However, the management of health information in such a combined event, must be fully integrated with the information that is needed to manage the natural disaster itself. This will not be easy to do and will require additional study.

In the U.S., FEMA has developed a number of documents which detail methods for organizing a disaster response including the Incident Command System (ICS) and the National Incident Management System (NIMS). A component of NIMS describes the Emergency Support Functions (ESFs), detailing the differing missions of Federal and Local agencies responsible for major response components like food and water supply, transportation, communications, and utilities. ESF 8 pertains to the work of the Health and Human Resources Agency and contains health care guidelines for federal, state and local agencies. However, ESF 8 provides only limited guidance for addressing the spatial dimensions of illnesses caused by a disaster event. The U.S. Department of Homeland Security (DHS) has also developed the Geospatial Concept of Operations (GeoCONOPS), which details spatial data needs, collection methods, technologies, and applications. Many nations have similar operational guidelines. We are unfamiliar with instances where the natural disaster and the disease outbreak domains were fully integrated.

We are certain, however, that for a natural disaster complicated by a disease outbreak, it will be necessary to rapidly collect accurate, spatially enabled data and to turn it into operational intelligence that is shared between the emergency management community and the health community. As much as possible, critical data, strategic applications, and guidelines for collaboration need to be put in place well before a disaster strikes. Additionally, exercises involving natural disaster responders and health responders need to be held on a regular basis to identify and iron out coordination problems and data sharing issues.

6.4. Considerations for relating a Health SDI data model to environmental events such as the chronic/acute effects of dangerous levels and types of air pollution/contamination

While most RFI responses did not address this kind of health emergency directly, one data source stands out. The Air Quality Index (AQI) is an excellent data source on air pollution and represents a model for health spatial data to be like moving forward. The AQI data is available for the whole world, is in real time, and publicly available. There are apps and websites that use the AQI data and it is universally used, so that there is consistent measurement and definitions across the globe. Countries may measure air quality separately, but many rely on the AQI data.

Still, there is a great deal of research that needs to be done and there is much more data that needs to be collected including measurements of point air pollution (i.e., from power generation or steel plants) and ambient air pollution from vehicle exhaust or heating and cooking fuels. Additional research priorities include assessing the impact of weather conditions on air quality particularly as the climate crisis accelerates and understanding how people exposed to air pollution interact with the health system.

7

CONCLUSIONS AND NEXT STEPS

---

7 CONCLUSIONS AND NEXT STEPS

A starting assumption of the study was that health information could be usefully organized into four categories, including population and patient data; supply chain data; health facilities data; and foundation and contextual data. We also understood the importance of both bioscience data and clinical research data, which although generally not having spatial characteristics, nevertheless determined many aspects of the response to a health disaster, including the development of diagnostics; the design and production of vaccines; and the identification of effective treatments. These six data categories were not challenged and we felt comfortable using them as the main building blocks of the health data model.

The CDS has also placed emphasis on the importance of examining the health spatial data value chain. We agreed that we should be able to demonstrate that every component of the data model served important purposes for applications and modeling, actionable intelligence, operations support, and the delivery of health benefits. We believe that the value of a data model hinges on the practical utility of the data elements identified.

This study is a first step towards more fully delineating the spatially enabled data needed to more successfully deal with disease outbreak and other health emergencies. By identifying and describing the interplay between four data categories (population and patients; health related supply chains; health facilities; and contextual information such as imagery, addresses, demographics, housing, and transportation) we hope we have pointed the way for more in depth study to be supported by pilot projects and table top exercises. The international adoption of a common data model will not only provide for consistent, comprehensive, and interoperable health data within nations, but also between nations — essential when dealing with a widespread health event. It will also foster the sharing of data management methods, applications, models, and technologies; not leaving it for each nation to develop unique solutions for their own, individual data environment in the middle of a health crisis.

We envision a number of additional steps going forward, including the following.

- The addition of detailed data elements within each of the four data categories through the development of "code lists" customized, as necessary, for the characteristics of each type of health challenge and the uniqueness of each nation.

- The identification, documentation, and evaluation of dashboards, applications, models, and A.I. techniques, so that nations do not need to re-invent data management and analysis tools in the midst of an emergency.

- The development of a maturity model for Health Spatial Data Infrastructures that shows nations, whatever their level of income, the path to improving their current health data infrastructure or the way to build a better one. This should be done with the understanding that spatial data built for one specific use can likely be used for dozens of other uses as well. For example: a transportation spatial data layer developed for health supply chain tracking can be re-used for all other kinds of supply chains, economic development programs, urban and national planning, capital programs, tourism, and the delivery of government services.

- The evolution of a smartphone infrastructure to be used as a platform for location identification, public health messaging, individual health status reporting, and health diagnosis and treatment using sensors and video conferencing.

- The integration of Health spatial data based operations with those of the emergency response community as a whole. The effects of a major health disaster ripple across every sector of a society and requires a comprehensive, all-of-government approach, calling for the integration of data across many agencies and organizations.

As a final thought, we should keep in mind that the enormous power of spatially enabled data, when created to common standards across an entire nation, comes from its ability to integrate and make interoperable any combination of datasets needed to solve a problem or support an operation.

Vision For Integrated National, State and Local Data
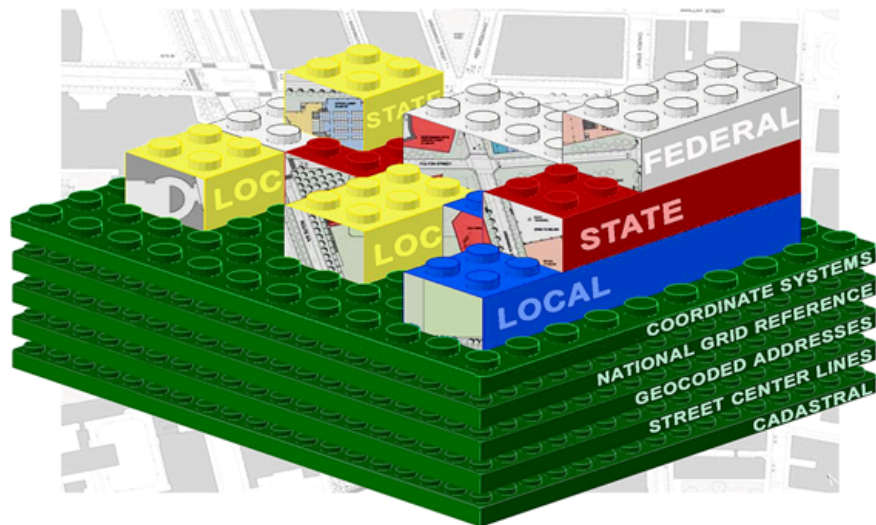
One Map To Bind Them All

Figure 11 — A Vision for National and International Spatial Data Integration

Spatial data breaks down information silos and increases the ability of applications and technologies to provide unprecedented levels of benefits for individuals and society.

A

ANNEX A (INFORMATIVE) REVISION HISTORY

A ANNEX A (INFORMATIVE) REVISION HISTORY

DATE	RELEASE	AUTHOR	PRIMARY CLAUSES MODIFIED	DESCRIPTION
March 18, 2021	.5	A. Leidner	all	validation draft version
April 1, 2021	.9	M. Reichardt	all	comments integrate
April 14, 2021	1.0	J. Lieberman	all	transfer to ER template asciidoc