EMBARGOED UNTIL NOON EDT ON TUESDAY, JULY 28, 2020

Strengthening the Public Health Infrastructure: The Role of Data in Controlling the Spread of COVID-19

By an Ad Hoc Pandemic-Response Subgroup of Former Members of
President Obama's Council of Advisors on Science and Technology

July 27, 2020

Introduction

The national response to the COVID-19 pandemic has exposed weaknesses in the United States public health system and has highlighted inadequacies in the nation's ability to anticipate, to prepare for, and to respond to serious infectious disease epidemics. Comprehensive preparedness includes prevention, and it also includes sound plans for responding to a pandemic. A large part of an effective response depends on infrastructure that is in place ahead of the onset of illness and spread of infection. An effective public health response depends on information based on reliable and timely data. The experience of the past few months has demonstrated the need to strengthen substantially the infrastructure for the collection, analysis, and sharing of those data.[1]

Inadequacy of data is not the only weakness exposed by the pandemic. The United States health care system has long been correctly characterized as costly, inefficient, inequitable, uncoordinated, and, thus, in need of improvement.[2] The underlying structural causes of these problems include the complexity of the combined public and private payment models, payment for volume rather than value, less than universal coverage, and separate silos of public health management and health care delivery. Our report focuses exclusively on the issues contributing to data availability, but many of these data problems impede needed improvements in the health system overall.

This report reviews the shortcomings of electronic data capture and use for public health purposes, outlines the recent history of changes in platforms and policies for digital health and health care data at CDC and in the broader health delivery system, and makes recommendations about immediate and mid-term policy changes that could accelerate the progress to a state-of-the-art, digital, data-science capability.

---

[1] Schneider, Eric C., *Failing the Test - The Tragic Data Gap Undermining the U.S. Pandemic Response*, New England Journal of Medicine, May 15, 2020, https://www.nejm.org/doi/pdf/10.1056/NEJMp2014836?listPDF=true; Laura Santhanam, *Data is key to fighting the coronavirus. Here's why it's so hard to find,* PBS website, June 26, 2020, https://www.pbs.org/newshour/health/data-is-key-to-fighting-the-coronavirus-heres-why-its-so-hard-to-find; Sittig DF, Singh H. COVID-19 and the Need for a National Health Information Technology Infrastructure. *JAMA.* 2020;323(23):2373–2374. doi:10.1001/jama.2020.7239

[2] Institute of Medicine (US) Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington (DC): National Academies Press (US); 2001, https://www.nap.edu/catalog/10027/crossing-the-quality-chasm-a-new-health-system-for-the; *Better Health Care and Lower Costs: Accelerating Improvement through Systems Engineering*, 66 pp, May 2014 https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_biodefense_letter_report_final.pdf

Establishing Electronic Health Records

Calls for better data collection for public health purposes are not new. About a decade ago, the United States made a push to digitize medical records from hospitals and doctors' offices. As part of the massive federal response to the 2008 financial crisis, Congress passed the American Recovery and Reinvestment Act of 2009 (ARRA),[3] part of which was the Health Information Technology for Economic and Clinical Health Act (HITECH Act).[4] The HITECH Act was intended to spur implementation of electronic health records (EHRs) by providing incentive payments to providers to transition to digital records systems and demonstrate their *meaningful use* in clinical practice.[5] The 2010 PCAST Report to the President on Health Information Technology observed, in this vein, that "If real time concurrent clinical data about every healthcare encounter were collected electronically, such data could be combined, without personal identifiers, from both regional and national perspectives to track public health developments and create timely prevention and amelioration strategies."[6]

The HITECH Act established a National Coordinator for Health Information Technology. One of the goals for the nationwide health-information-technology infrastructure for which the national coordinator is responsible is that it "improves public health activities and facilitates the early identification and rapid response to public health threats and emergencies, including bioterror events and infectious disease outbreaks".[7] The strategy that the Office of the National Coordinator (ONC) introduced was to implement Meaningful Use[8] in three stages. Stage 1 was to set the standards for the electronic capture of clinical data and the electronic access by patients to their personal health information. Later stages extended the scope to Health Information Exchanges (HIEs). HIEs are entities set up to share data among a local or regional group of health provider organizations. They have been described as a complex web of bilateral trade agreements.[9] Some HIEs were much more effective than others, but all required significant resources to sustain the ability to exchange patient data. As the initial HITECH funding ran low, and the program moved more slowly than anticipated, the goal of providing all currently available electronic information on a patient from any source suffered.[10]

Various attempts have been made over the years by the Office of the National Coordinator (ONC), the Centers for Disease Control (CDC)[11], and state and local public health

---

[3] Public Law No: 111-5 (02/17/2009), https://www.congress.gov/111/plaws/publ5/PLAW-111publ5.pdf

[4] Ibid., Division A, Title XIII—Health Information Technology.

[5] The program was renamed in 2019 as the Public Health and Promoting Interoperability Programs, https://www.cdc.gov/ehrmeaningfuluse/introduction.html

[6] *Realizing the Full Potential of Health Information Technology to improve Healthcare of Americans: The Path Forward*, Report to the President, President's Council of Advisors on Science and Technology, December 2010, p 21, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf

[7] P.L. 111-5, Division A, Title XIII, Subtitle A, Sec. 3001 (b)(7).

[8] Meaningful Use, as defined in the HITECH Act is now often styled as a proper noun.

[9] *Realizing the Full Potential*, pg. 40

[10] Gold, Marsha, and Catherine McLaughlin. "Assessing HITECH Implementation and Lessons: 5 Years Later." *The Milbank quarterly* vol. 94,3 (2016): 654-87. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5020152/pdf/MILQ-94-654.pdf

[11] Both are organizations within the Department of Health and Human Service (HHS).

organizations[12] to graft public-health data systems onto legislatively mandated EHR systems to achieve later stages of Meaningful Use. Some organizations have used other sources, such as hospital admissions records, to gather population-level data.. These efforts have been inadequately funded, and there has been insufficient coordination between federal efforts and those of other jurisdictions. As recent events have demonstrated, the nation cannot continue to do without a robust public-health data infrastructure. It is time for the United States government to act to create one, in coordination with state and local jurisdictions.

How Does the Public Health System Use Data?

Using data to discover and monitor the presence of disease

An essential first step in managing the response to any epidemic is the ability to identify early cases and clusters or outbreaks of disease. Ideally, centralized public health authorities would be able to spot outbreaks of characteristic symptom patterns even before diagnosis of identified cases occurs; then, identified cases would be reported immediately as the test results are returned. Information systems exist that would make this possible, but the United States is still not able to make optimal use of available digital science in the interest of public health in a pandemic situation. Traditional public health methodology is still too often manual – it depends on individual clinicians recognizing an unusual transmissible infection and then voluntarily reporting it by email, fax or phone to public health authorities. The transmission is electronic, but the reporting is manual. In a known epidemic situation, some hospitals and health systems will aggregate the data from a day, a week, or a month and then transmit the data to public health authorities. Electronic data retrieval and exchange has been enabled by a number of developments at local and state public health departments as well as at CDC, but these developments are widely variable in quality and far from universal.

As recently as May of 2020, the White House Pandemic Task Force was requesting Excel spread sheets of COVID-19 cases from hospitals by fax or email.[13] On July 10, 2020, the White House announced a controversial plan to have hospitals bypass CDC and send all COVID-19 patient information to a central database in Washington daily, still mostly by manual means.[14] The HHS guidance states that "The following data will greatly assist the White House Coronavirus Task Force in tracking the movement of the virus, identifying potential strains in the healthcare delivery system, and infoming [sic] distribution of supplies." Concern has been expressed that the change places an additional burden on hospitals, that it politicizes data collection, that it weakens CDC, and that it risks withholding information from the public.[15]

---

[12] In this report phrases referring to state and local public health entities encompass tribes, territories, and community organizations.

[13] Daris Tahir, *Virus hunters rely on faxes, paper records as more states reopen*, Politico, May 10, 2020, https://www.politico.com/news/2020/05/10/coronavirus-health-records-245483

[14] https://www.hhs.gov/sites/default/files/covid-19-faqs-hospitals-hospital-laboratory-acute-care-facility-data-reporting.pdf

[15] Sheryl Gay Stolberg, *Trump Administration Strips C.D.C. of Control of Coronavirus Data*, New York Times, July 14, 2020, https://nyti.ms/309Xtsi; Amy Goldstein and Lena H. Sun, *Hospital officials, experts say new federal rules for COVID-19 reporting will add burdens during pandemic*, Washington Post, July 15, 2020; https://www.washingtonpost.com/health/2020/07/15/coronavirus-trump-administration-data-change/; Thomas M. File, Jr., *Response from IDSA President to New COVID-19 Data Reporting Protocol*,

This reliance on antiquated data-entry and communication technology, along with burdensome clerical work from individual clinicians, is not how the United States should be retrieving such essential information in 2020. These antiquated approaches are fraught with shortcomings: incompleteness, errors, time lags, reporting bias, and added staff work and costs. Most importantly, they don't give health authorities the information they need to respond optimally to protect public health and reduce the economic impact of the kinds of major personal restrictions now in place.

Modern digital technology can provide a more complete, more timely, more efficient, and less onerous approach to creating a data environment adequate for local, state, and national leaders to make appropriate policy decisions and, thereby, reduce infections risk and deaths. Modern data systems will allow immediate tracing of symptoms as people seek help in doctor's offices, retail clinics, or emergency departments. This sort of *syndromic surveillance* becomes possible when the information from clinical encounters now recorded in EHRs is aggregated and automatically surveyed by regional, state, and national public-health offices. While a National Syndromic Surveillance Program (NSSP) already exists, it does not directly use EHR data and is limited in what symptoms are reported (as discussed below).

A significantly stronger syndromic surveillance would have allowed specific clusters of COVID-19 symptoms (such as dry cough, fever, lethargy, anosmia[16]) to be identified early, even if the clinicians did not make the right diagnosis or were not yet aware that COVID-19 was in their area. The traditional public health model is that clinicians report *cases* of an infectious disease, which means they must have correctly identified the cause of illness in their patient and taken the time to send a report to the local public health authorities. Syndromic surveillance, by reporting *encounters* rather than cases, allows more rapid identification of trends, and provides more rapid useful situational information to clinicians. Influenza, which has some symptoms similar to but also different from COVID-19 in its early phases, could be differentiated from COVID-19 in seasonal outbreaks, leading to different treatments and different implications for isolation policies.

This kind of surveillance is not just possible in theory. NSSP is a primitive implementation, but better systems are already used in other countries that have comprehensive health-data systems, including ten European nations.[17]

Using data to contain the spread of disease

In addition to data about symptoms, timely and complete information about testing prevalence (how widely diagnostic testing has been done) and test outcomes is essential to disease control. The United States has lagged behind other countries in the availability of diagnostic testing.[18] In addition, reliable information about testing prevalence and test outcomes has not been efficiently transmitted to public health authorities. The best databases on testing for COVID-19 were

https://www.idsociety.org/news--publications-new/articles/2020/response-from-idsa-president-to-new-covid-19-data-reporting-protocol/

[16] The loss of the sense of smell

[17] Guerrisi et al., "Participatory Syndromic Surveillance of Influenza in Europe," The Journal of Infectious Diseases, Volume 214, Issue suppl_4, December 2016, Pages S386–S392, https://doi.org/10.1093/infdis/jiw280

[18] Testing issues will be treated in more detail in a forthcoming report by our Ad Hoc Subgroup.

created by research universities and by news organizations, not by public health authorities, and they are limited to aggregated counts derived from a patchwork of inconsistent reporting by counties, states, and commercial labs, with unknown validity and reliability.[19] Other countries, such as Australia, South Korea, Germany, Denmark, Singapore, and Taiwan used a combination of widespread testing and contact tracing to make decisions to institute nonpharmacological interventions (NPI) such as social distancing, school and business closures, hand hygiene, and masking to limit the spread of the disease early and reduce case numbers and deaths.

If clusters of cases and regional outbreaks can be identified early, then NPI measures can be instituted in focused areas to limit spread and prepare for needed healthcare capacity, without requiring widespread shutdown of businesses and schools. Lacking this situational awareness, the United States has had to make public health decisions while "flying blind".

Data scientists use modeling to forecast trends and outcomes, in realms from weather to economics. The use of a diversity of models to forecast the spread and impact of COVID-19 has produced widely divergent predictions, ranging from tens of thousands to more than two million deaths during the first months of the United States epidemic. These wide variations have led some to be skeptical of any predictive analytics and have reinforced the apparent inevitability of flying blind. But statistical models and predictions are only as good as the data and assumptions on which they are based. Most modelers are quite expert, but their assumptions necessarily have been based on the data available, which were insufficient and often of uncertain validity.

Clear understanding is still lacking about the transmissibility, clinical course, and lethality of the SARS-CoV-2 virus that causes COVID-19, and about the impact of different approaches to NPI. Access to timely and reliable population-level data could enable much more reliable and accurate forecasts. Using whatever data and forecasts were available, states variably instituted use of NPI, some mandating closure of businesses and gatherings of people, others remaining more lenient—sometimes with dire subsequent consequences. Strong NPI measures slow the rates of spread of COVID-19 and lower fatality rates, but because of adverse economic and social consequences of strong NPI, all states struggle with the decisions of when and how best to loosen restrictions. With better data and awareness of specific outbreaks, more rational decisions about the pace and process of reopening would be possible.

What Are the Barriers to Using EHRs for Public Health?

EHRs have been designed for clinical use

EHRs in the United States were not designed to provide data for syndromic surveillance and other population-level public health purposes. They were developed primarily to capture the data necessary to document clinical services in order to issue bills to insurers and to issue prescriptions. In addition, the record systems are sold and maintained by several large commercial vendors, whose software is proprietary and who have been allowed to block information sharing with other entities. EHRs are able to aggregate data within a given system,

[19] Examples include the Johns Hopkins University Coronavirus Resource Center, https://coronavirus.jhu.edu, the New York Times Data Repository, https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html, and the Atlantic Magazine COVID Tracking Project, https://covidtracking.com/about-project

for analyses of preventive services for patient panels[20] or to identify patients with particular kinds of clinical needs. But with only a few exceptions, data from EHRs have not been able to be used directly for public health surveillance of an entire city, county, or state. Recently, work-around attempts to use other sources of data are being created, but none is likely to have the accuracy and completeness of EHRs.

The HITECH Act of 2009 focused on the export of data from EHRs. The enabling legislation and subsequent laws established a set of *data elements* that had to be accessible electronically. For example, Stage 1 of Meaningful Use measured data such as medications prescribed, vital signs, and smoking status. The legislation laid the groundwork for HIEs (see above) so that hospitals could access the data to send to CMS for Meaningful Use bonus payments, but the data flow is one-way – from the EHR to a client such as a billing system. Stage 2 required being able to share data with another local provider such as an emergency room and it too only required a limited scope of data transfer. Only in Stage 3 is there a provision for a trusted client to *add* information, such as a Summary of Care from a different health provider.

As regards public health, HITECH's vision for communication with public health organizations is modest, consisting mostly of the issuance of predefined reports. For example, Stage 3, which became mandatory in 2018 but is not universally implemented, requires that periodically the provider generate and share at least two out of these five reports: immunization registry, syndromic surveillance reporting, electronic case reporting, public health reporting (but only of so-called reportable conditions), and clinical-data registry reporting. In times of emerging health crises such as pandemics, much more than periodic predefined reports is needed. Access to the entirety of the EHR is needed, since, clues to the nature and spread of the pandemic may emerge from data such as co-morbidities, vaccination, medication, or family history, or other information contained in the EHR that cannot be anticipated in advance. Data from EHRs needs to be available without delay (i.e. in real time) so that public health organizations can respond to pandemics quickly.

Health information exchange, as called for in Stage 3, requires the transfer of a Summary of Care record and the reconciliation of clinical information from multiple records. To enable that, independent records systems must be able to exchange data (so-called *interoperability)*. Lack of interoperability has been a barrier to coordination of care, since patients may see providers who cannot access the records of other specialists, sometimes even in the same healthcare delivery system.

A more dynamic exchange of electronic information would be multi-directional, both for multiple health providers and for public health organizations. Conceptually, think of all the EHRs as a large distributed data base. Subject to privacy and security protections, authorized users should be able to make queries into that data base, rather than waiting to receive a predetermined set of reports.

[20] In some healthcare practices, each patient is assigned to a particular primary care physician who coordinates care of that patient. The group of patients assigned to a particular physician is known as a *patient panel.*

EHR access must conform to the Health Insurance Portability and Accountability Act of 1996 (HIPAA).[21] HIPAA required the Secretary of the U.S. Department of Health and Human Services (HHS) to develop regulations protecting the privacy and security of certain health information. To fulfill this requirement, HHS published what are commonly known as the HIPAA Privacy Rule and the HIPAA Security Rule. The Privacy Rule, or *Standards for Privacy of Individually Identifiable Health Information*, establishes national standards for the protection of certain health information irrespective of the form in which it is held. The *Security Standards for the Protection of Electronic Protected Health Information* (the Security Rule) establishes a national set of security standards for protecting certain health information that is held or transferred in electronic form.

The HIPAA Privacy Rule allows the disclosure of protected health information to public health authorities who are legally authorized to receive such reports for the purpose of preventing or controlling disease. Hospitals are required to limit the protected health information disclosed to the minimum amount necessary to accomplish the public health purpose. The interpretation of this rule leads to the possibility of a patchwork of accessible and inaccessible data, or even "malicious compliance" with the narrowest legal interpretation of a stated request hiding data that is needed and reasonably implied. The HIPAA Privacy Rule also specifies standards for the de-identification of data "*alone or in combination with other reasonably available information*", which many technologists believe to be unachievable given today's capabilities in data analysis, including artificial intelligence. Changes in the HIPAA Rules are needed to make protected disclosure more robust and more appropriate for electronic interchange and to make de-identification standards consistent with current technological reality. Public health, as a collective good, should require only best-practice measures that are not insurmountable roadblocks.

Advances in EHR technology are slow to be adopted

There has been continuing progress in technology for the accessibility, sharing, and use of EHRs for clinical care. Health Level 7 International (HL7) is a not-for-profit membership organization founded in 1987, whose mission is to develop a platform and standards for the exchange, integration, sharing, and retrieval of electronic health information[22]. Beginning in 2011, HL7 developed and has evolved the Fast Healthcare Interoperability Resources (FHIR) specification for the electronic exchange of healthcare information.

In October 2015, the HHS Center for Medicare and Medicaid Services (CMS) and the HHS Office of the National Coordinator for Health IT (ONC) published the final rule on Meaningful Use Stage 3, which became mandatory for all participants in 2018[23]. The rule requires that Summary of Care reports be transmitted using the Consolidated Clinical-Document Architecture (C-CDA) specified by HL7.

---

[21] Public Law 104-191, August 21, 1996, https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf

[22] https://www.hl7.org/about/

[23] Stage 3 Program Requirements for Eligible Hospitals, CAHs and Dual-Eligible Hospitals Attesting to CMS, https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Stage3_RequieEH

Meanwhile, the 21st Century Cures Act was signed into law in 2016[24]. Title IV of that Act requires that patients have access to their electronic health information. It mandates interoperability, specifically that Application Programmer Interfaces (APIs) be available "without special effort" (basically, compelling open APIs rather than proprietary ones)[25]. It also prohibits *information blocking*, the restricted availability of electronic health information. The National Coordinator is responsible for overseeing the implementation of those requirements.

Draft ONC rules to implement the Cures Act were finally issued in the fall of 2019, and the Cures Act Final Rule was issued in May 2020[26]. The API certification criterion requires the use of FHIR release 4. The rule has still not taken effect, however.

The long delay is an indication of the complex web of vested interests underlying the limitations of the EHR system. Originally scheduled to take effect in November, 2020, the law's implementation has regrettably been delayed even more in response to the COVID-19 pandemic, ostensibly because the changes necessary would cause additional burdens on the nation's hospitals and health-care delivery infrastructure. The national crisis we now face ought instead to be a reason to move more quickly, once and for all, to make the changes preparing our nation for a stronger, more data-capable future. The recent ONC rules are focused on opening the information for clinical and patient use but can be easily extended to public health, as we recommend below.

In the last two decades, CDC has experimented with several different models for the collection of public health data. None has achieved the scope that is now needed, but valuable lessons have been learned.

A surge of interest in bioterrorism defense followed the terrorist attacks on 9/11 and subsequent threats of anthrax being weaponized. Leaders at the time wisely combined bioterrorism planning and resources with efforts to prepare for a naturally occurring pandemic, which experts predicted was inevitable. These efforts led, among other steps, to the creation of the Strategic National Stockpile.[27] Leaders also understood the importance of data for biodefense, and CDC invested approximately $300M in BioSense, a "top-down", contractor-designed, HIE sentinel network.[28] The use of BioSense required a considerable amount of work and commitment by hospitals, requiring that data actually be transferred to a central entity. As reported in a paper by Gould *et. al*., by 2007 only 10% of civilian-hospital emergency departments (EDs) were participating.[29]

---

[24] Public Law 114-255, December 13, 2016, https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf

[25] APIs are the protocols for the interaction with software components such as data repositories or programs. They are often used to create user-level *applications* ("apps") that invoke system-level services and data access. For example, a capability that gives a patient access to personal health information uses an API to access the EHR.

[26] 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program, May 1, 2020, https://www.govinfo.gov/content/pkg/FR-2020-05-01/pdf/2020-07419.pdf

[27] http://adhocresponsegroup.org/OPCAST_Ad_Hoc_Subgroup_Stockpile_Recommendations_05-20-20.pdf,

[28] Public law 107-188, June 12, 2002, https://www.govinfo.gov/content/pkg/STATUTE-116/pdf/STATUTE-116-Pg594.pdf

[29] Gould DW, Walker D, Yoon PW. *The Evolution of BioSense: Lessons Learned and Future Directions.* Public Health Rep. 2017 Jul/Aug;132(1_suppl):7S-11S, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5676506/

Likely also relevant to its limited success was that BioSense bypassed local and state public health departments, who were thus disinclined to advocate for it.

In 2008, CDC embarked on a four-year plan to redesign the BioSense network. BioSense 2.0 allowed state and local health departments to access data that supported expansion of their syndromic surveillance systems in accordance with the Meaningful Use program. Rather than hospitals' sending all data directly to CDC (as was the case with the original BioSense), under BioSense 2.0 most data were sent to state and local health departments. Still, BioSense 2.0 lacked a sustainable governance model with buy-in from all parties. It appears that challenges in CDC's collaborations with state and local health departments, together with changes in personnel, caused some disruption to the Biosense 2.0 program. In 2014 the BioSense 2.0 program became the National Syndromic Surveillance Program (NSSP). NSSP replaced the Biosense 2.0 system architecture with a system called *Essence*, created by Johns Hopkins Engineering School, which had a more modern user interface but more limited capabilities, something of a return to a "top-down" architecture.[30]

What is the Path Forward?

A 21st century infrastructure for health information is a necessity for the United States. We believe that a national multi-app platform (defined below), federally funded but with collective governance involving state and local public health departments, should now be a top priority.

Platform and Applications

The distinction between platforms and apps is important. A *platform* is the software and communications infrastructure necessary for desired data flows to take place. The platform needed for public health data would take advantage of Application Program Interfaces (APIs) already being provided to their clients by EHR vendors to comply with Meaningful Use Stage 3 and Cures Act requirements. While these APIs have been designed for use in healthcare delivery, not for public health, a public health platform could utilize them as well. That platform would include *middleware* (software invisible to the end-user) that, depending on the platform's design, may run on local systems, federated systems, or in the cloud.[31] In addition to infrastructure for data flows, the platform must also provide infrastructure for security and the authentication of authorized users and programs.

The end-user exploits the public health platform through *apps*—software that presents a user interface designed for a specific purpose with corresponding convenience and functionality. In so-called *closed* platforms a central authority dictates, and generally itself develops, a single app, or small number of apps, for the user. *Open* platforms are designed to host apps written by more than one developer or entity (for example, state public health agencies or authorized private-

[30] Timothy C. Campbell *et.al.*, OpenESSENCE: An Open-Source, Self- Contained Disease Surveillance Software Application for Global Use, Johns Hopkins Applied Physics Laboratory Technical Digest 32 (4), 2014, https://www.jhuapl.edu/Content/techdigest/pdf/V32-N04/32-04-Campbell.pdf ; Johns Hopkins University Applied Physics Laboratory. OpenESSENCE user guide, 2013. http://www.jhuapl.edu/sages/guides/OE-User-Guide-Feb-2013.pdf.

[31] The *cloud* refers to large data centers that provide data storage and services accessible to multiple clients over the Internet.

sector companies), as long as their apps conform to the platform's security and governance policies.

The idea that U.S. public health needs would be best served by an open platform (subject to security and governance limitations) is not new. A 2013 report prepared for the Department of Health and Human Services (HHS) Agency for Healthcare Research and Quality (AHRQ) by the JASON advisory group proposed a platform to support the robust exchange of electronic health information[32]. The JASON proposal advocated interoperability and the implementation of apps that could enable the use of health data not only by hospitals and clinicians, but by public health organizations, researchers, and patients.

Some earlier efforts to provide systems for public health data embraced some aspects of the open platform approach. Beginning in autumn 2006, the Distributed Surveillance Taskforce for Real-time Influenza Burden Tracking and Evaluation (DiSTRIBuTE) project built a distributed system using a new model in which an individual's data were retained locally but aggregated data were reported centrally for activities such as syndromic surveillance. This was a move to an open platform model with *de-centralized* (i.e., distributed) data.[33] In spring 2009, as the H1N1 influenza pandemic emerged, the system was deployed nationwide under the auspices of CDC. By early 2011, the network had 43 reporting sites and captured over 40% of all emergency department visits. DiSTRIBuTE was a pilot project and was discontinued in 2012 after six flu seasons.

As part of the HITECH Act, ONC funded the Strategic Health IT Advanced Research Projects (SHARP) program in 2010. One of the projects funded under SHARP was an award to Harvard Medical School and Boston Children's Hospital to create the SMART (Substitutable Medical Application, Reusable Technologies) API to support a platform for apps that can be used on smart devices such as iPhones and Android phones. The SMART API uses the FHIR platform. The SMART project continued even after the end of the SHARP program. Technical support for the SMART API is provided by major EHR and Cloud vendors.[34]

CDC's 2002-vintage BioSense system, mentioned above, used an Internet-based, but not cloud-based, software architecture. This was a closed-platform model with *centralized data*. That is, CDC both held the data and provided the user-level interface. In the BioSense 2.0 program, however, CDC funded the Association of State and Territorial Health Officials (ASTHO) to host the BioSense 2.0 cloud-based infrastructure and application and to develop data use agreements with participating state and local health departments. Most data were sent from hospitals to a

[32] A robust health data infrastructure. (Prepared by JASON at the MITRE Corporation under Contract No. 13-717F-13). Rockville, MD: Agency for Healthcare Research and Quality. April 2014. AHRQ Publication No. 14-0041-EF. https://digital.ahrq.gov/sites/default/files/docs/publication/a-robust-health-data-infrastructure.pdf

[33] Olson DR, Paladini M, Lober WB, Buckeridge DL; ISDS Distribute Working Group. Applying a New Model for Sharing Population Health Data to National Syndromic Influenza Surveillance: DiSTRIBuTE Project Proof of Concept, 2006 to 2009. PLoS Curr. 2011 Aug 2;3:RRN1251, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3148528/

[34] The SMART API and services are described on the project website, SMART HealthIT, https://smarthealthit.org

secure, cloud-based, data-storage facility managed by ASTHO.[35] The BioSense 2.0 platform provided a web-based user interface and embodied some aspects of an open-platform design.

Essence, the successor to BioSense 2.0, is cloud-based and has strong support for data analysis and visualization, but it appears to be a return to a single-app, closed-platform model, rather than a platform capable of supporting multiple applications. While the decision to return to this earlier model may have been justified at the time it was made, we believe that this model is inadequate to the needs highlighted by the COVID-19 pandemic. This public health emergency as well as future ones, will require a national data network that is more flexible, dynamic and versatile.

A recent project based in Chicago exemplifies the feasibility of an expandable collection of apps built using modern technology.[36] The Chicago Department of Public Health (CDPH) and academic partners at the Rush University Medical Center have created their own platform, using a locally designed, cloud-based system that looks like Biosense 2.0 but is built on modern APIs, using components of the EPIC EHR platform.[37] Using the FHIR specification, the team was able to bring together analysis of clinical, lab, and capacity data in just a few days to support the COVID-19 response. The Chicago platform uses publicly available open-source software tools to convert proprietary data formats to or from FHIR formats.

Using a common FHIR representation, the platform's apps provide visualizations and downstream analytics that enable rapid connectivity of data and interoperability across multiple hospitals. With strong leadership from the CDPH and Rush Medical Center, all the major hospitals in the Chicago area agreed to share real-time capacity data to allow hospitals with overwhelmed ICUs to know immediately where they could transfer a patient who needed critical care or intubation and mechanical ventilation. They can use the same technology to share case numbers, prevalence of testing, results of testing, and other key information electronically. The Chicago experience demonstrates that a high technology-readiness level for the underlying necessary technologies already exists.

Governance

Previous attempts at achieving a national public health platform—whether closed or open—have often foundered on governance issues. First of all, it is the states, not the federal government, that are primarily responsible for public health. Yet, while we recognize that states are constitutionally responsible for public health law within their state, there is also an important role for national standards and guidance, especially in a major epidemic. For this reason, the nation's public health effort would be much strengthened by consistent and interoperable data systems that include federal participation.

States' data and practices are incomplete, uneven, and inconsistent. States too often collect public health data from manual reports from hospitals and clinics. The federal government

[35] Association of State and Territorial Health Officials website https://www.astho.org

[36] Chicago Data Portal, https://data.cityofchicago.org/browse?limitTo=datasets&sortBy=alpha&tags=covid-19

[37] According to an October 2019 report from EHR in Practice, a small number of for-profit vendors provide most of the EHR systems in common use. Epic and Cerner control over 50% of the U.S. acute care hospital market. Epic also controls over 25% of the ambulatory care market and 6 companies share most of the remainder. https://www.ehrinpractice.com/largest-ehr-vendors.html

collects public health data from the states and from regional entities through manual surveys and voluntary reporting of certain conditions by physicians and hospitals. The timeliness and completeness of these reports vary widely by municipality, by county and by state. Aggregated data from social networking and internet search add somewhat to this capability but are also subject to multiple sources of bias and data gaps. Real-time clinical data that would provide the best capability to respond optimally to national infectious public health threats are lacking.

While states have a statutory responsibility for public health in their jurisdiction, only the federal government has the (potential) capacity to coordinate data and accelerate collection for national planning and response to emergencies. CDC is the central federal agency responsible for this role. Even with the shortcomings described above CDC has until recently been the acknowledged national leader in the science of population health, respected globally, and a major source of both expertise and reliable data.[38] This administration has significantly weakened CDC, but considerable expertise remains. CDC's strengths can be restored, and its shortcomings can be addressed to create the unitary national resource that the nation and the global community need. The need for public health leadership will continue at both the national level and the state level, and for the organizations to work together closely and effectively. It is time for a national reinvestment in public health skills at both levels and in building bridges between them.

The development of a national public health infrastructure requires careful consideration of what its governance model should be. It is clear that top-down models, where CDC interacts directly with hospitals, bypassing state and local authorities, do not work. CDC itself has recognized this reality, for example by working more collaboratively with ASTHO and other stakeholders in the development of the now-defunct Biosense 2.0 program.[39]

A more recent example of a workable governance model is the Digital Bridge Initiative, a collaboration among public health organizations, healthcare delivery organizations, and industrial healthcare information technology providers founded in 2016. Its initial focus has been on electronic case reporting. Using existing EHR services, potentially reportable disease cases are sent electronically to a central decision support service managed by the Association of Public Health Laboratories (APHL) and the Council for State and Territorial Epidemiologists (CSTE). Reportable cases are forwarded to public health organizations from one of the seven pilot sites.[40] The Initiative has paid careful attention to its governance structure and the process by which decisions are made. The CDC Office of Public Health Science and Surveillance is an active participant in Digital Bridge. Alas, while possibly offering lessons on governance, the pilot programs appear hobbled in technology, for example, not implementing any of the technology used in the Chicago project, the SMART platform, or other such efforts, and choosing instead to use software developed by APHL and the EHR vendors in seemingly a closed fashion.

---

[38] Charles Duhigg, *Seattle's Leaders Let Scientists Take the Lead. New York's Did Not* , The New Yorker, April 26, 2020, https://www.newyorker.com/magazine/2020/05/04/seattles-leaders-let-scientists-take-the-lead-new-yorks-did-not

[39] Gould DW, Walker D, Yoon PW. *The Evolution of BioSense: Lessons Learned and Future Directions.* Public Health Rep. 2017 Jul/Aug;132(1_suppl):7S-11S, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5676506/

[40] Digital Bridge website, https://digitalbridge.us/about/

A satisfactory governance model, coupled with a suitable platform-architecture design and adequate incentives for the development of apps, is a necessary prerequisite for the success of a national public-health infrastructure. Given the inherent turf issues and conflicts of interest among many of the stakeholders (e.g., CDC, states, EHR vendors), a dispassionate study of governance by a neutral party such as the National Academy of Medicine seems desirable.

Information Technology Expertise

Many public health departments do not have the level of technology expertise that the Chicago project enjoys, nor does it make sense for every organization to implement its own solutions. Federal, as well as state, leadership is needed. Even if a shared platform supports multiple apps, there is an immediate need for a co-developed single app with a high-quality user interface that all public health organizations can readily use. This is both a public health issue and a national security issue.

As part of the restoration of a strong collaborative Federal, state and local public health infrastructure. HHS and the Congress need to ensure that technical personnel are embedded in all levels of HHS, and included in the highest-level policy discussions, and that expertise is provided to state and local public health organizations that may not have that expertise within their own ranks.

Short-term and Long-term Funding

The CDC and state and local jurisdictions have all been hampered for decades by underfunding.[41] In the shadow of the financial dominance of the medical-care system, including CMS, and the congressional excitement generated by the biomedical research mission of the National Institutes of Health (NIH), public health has long been undervalued at both the national and the state level.[42] Relatively little of the funding that is allocated for public health is used for infectious diseases. Recent years have seen public health budgets go from small to smaller, as parts of the public health mission became political targets in the context of a more general devaluing of scientific expertise.[43]

Federal funds, mostly from CDC and the Department of Agriculture (USDA), are the largest source of funding for state public health departments. The funding is almost always targeted to specific programs – particular diseases, maternal and child health, and food insecurity are examples. There is little core funding that might be used for infrastructure development or for emergencies such as the COVID-19 pandemic.

Funding for CDC's Public Health Emergency Preparedness (PHEP) cooperative agreements, which support core public health capabilities in states, territories, and local areas, have decreased

---

[41] Trust for America's Health, *The Impact of Chronic Underfunding on America's Public Health System: Trends, Risks, and Recommendations, 2020*, April 2020, https://www.tfah.org/report-details/publichealthfunding2020/

[42] Institute of Medicine. 2003. *The Future of the Public's Health in the 21st Century*. Washington, DC: The National Academies Press. https://doi.org/10.17226/10548

[43] Michael D. Shear, *'They Let Us Down': 5 Takeaways on the C.D.C.'s Coronavirus Response*, New York Times, June 3, 2020, https://www.nytimes.com/2020/06/03/us/cdc-virus-takeaways.html

from $940 million in FY 2002 to $675 million in FY 2020. The 2010 Affordable Care Act[44] established the Prevention and Public Health Fund (PPHF) and funded it with a permanent appropriation that was to rise to $2B per year in 2015.The Middle Class Tax Relief and Job Creation Act of 2012 (P.L. 112-96) reduced PPHF appropriations for FY2013 through FY2021 and the 21st Century Cures Act reduced PPHF appropriations for FY2018 through FY2024, each time diverting the appropriated funds for other purposes. The Fund reached $1B in FY 2012 and again in FY 2020; it is not expected to reach $2B until FY 2025.[45]

It appears that funding by state governments, which typically comes from allocations from General Funds, has decreased at an even faster rate than Federal funding over the last decade. Exact figures are not available, however; there is little quantitative data that documents state public health expenditures. There is little consistency among states, moreover, in part because states use their public health funds for differing purposes. Some state funding for specific purposes such as immunizations is acquired through state taxes on public and private health insurance.[40]

Given the inadequate and fluctuating funding, it is no surprise that both the Federal government and the states fail to invest in preparedness and infrastructure, choosing instead to use their limited funds for more immediate needs. Congress must take the lead in ensuring that the nation creates and maintains modern infrastructure and is well-prepared to respond to the present pandemic and the inevitable future public health emergencies.

Recommendations

Cures Final Rule

Recommendation 1. The Office of the National Coordinator (ONC) should rescind its announced delay in implementation of the Cures Act Final Rule on interoperability, and, during the COVID-19 pandemic, seek to accelerate its implementation.

Recommendation 2. In light of the COVID-19 pandemic, EHR vendors should accelerate their programs to support the ability to connect any authorized app to their systems, using the FHIR specification, as required by the Cures Act Final Rule.

Technology Improvements

Recommendation 3. The recent CARES act appropriates $500M to CDC to upgrade its IT systems. A major focus of that spending should be the planning, at a national level with collaboration by the states and localities, of a platform for real-time access by public health organizations to data from EHR records, mortality records, demographic data, and other electronically available information that can be used for pandemic preparedness. Such a

[44] Public Law No: 111-148. https://www.govinfo.gov/content/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf

[45] NORC at the University of Chicago, *An Examination of Public Health Financing in the United States*, March 2013, prepared for the HHS Office of the Assistant Secretary for Planning and Evaluation (ASPE), https://www.norc.org/PDFs/PH%20Financing%20Report%20-%20Final.pdf

platform should support multiple apps and decentralized data. The plan should be presented to Congress by June 2021.

Recommendation 4. CDC should lead the creation of the platform, drawing on the expertise provided by Recommendations 8 and 9, and funding appropriated by Congress.

Data Availability

Recommendation 5. To exploit the platform at the earliest possible time, including appropriate participation in its development and the development of apps for its use by public health officials, states and localities should develop individual policies for which public health data should be accessed by the platform from hospital/clinician EHR systems and which should be held in data repositories under state or local control. Regardless of location, however, the data must be available for authorized uses by national-scale apps on the platform.

Recommendation 6. The Centers for Medicare and Medicaid Services (CMS) and the Veteran's Administration (VA) should seek to accelerate the collection of public health data, from hospitals and providers in their respective patient bases, using the mechanisms of the Cures Act Final Rule, even before that rule takes final effect. Congress should appropriate funds that can be allocated to hospitals and providers under supervision by CMS and VA for this purpose.

Recommendation 7. HHS should issue clarifying guidance (or if necessary amend) 45 CFR Section 164.515 to clarify that, for purposes of the HIPAA Privacy Rule, (i) determinations by CDC on "minimality" for the release of EHR records to state and local public health authorities should be considered sufficient for their release, and (ii) for public health purposes, a combination of reasonable de-identification and good cybersecurity practices in data storage will be deemed sufficient to satisfy the rule.

Information Technology Expertise

Recommendation 8. CDC, ONC, and HHS should make use of existing hiring authorities, significantly increased by the Cures Act, to strengthen agency information technology leadership and expertise.

Recommendation 9. The Office of the National Coordinator (ONC) should establish immediate workforce programs designed to bring needed IT technical expertise to the states and localities. This might include hiring a pool of experts at the federal level and deploying them under the Intergovernmental Personnel Act (IPA) to states and localities on short-term assignment as "IT Tiger teams" under temporary state control. The pool should include high-level executives and managers with systems expertise as well as an IT service corps. Congress should appropriate funds for this purpose.

Infrastructure Governance

Recommendation 10. HHS should ask the National Academy of Medicine to convene a consensus study on governance issues associated with a national public-health data infrastructure under several scenarios, considering the views of all stakeholders. The study should be specifically tasked to recommend an actionable governance model.

Funding

Recommendation 11. Congress should pass legislation that restores funding for the Prevention and Public Health Fund to $2B per year from FY 2021 onward and provides for inflationary increases.

Recommendation 12. States should explore sources of funding for public health other than allocations from the General Fund. Taxes on health-insurance providers to support public health, for example, might actually decrease the provider' net expenditures as a result of better disease prevention .

Recommendation 13. Congress should ask the National Academy of Medicine to convene a consensus study on funding issues associated with a national public-health data infrastructure under several scenarios, considering the views of all stakeholders. The study should be specifically tasked to recommend an actionable and stable funding model.

Conclusion

A strong national public health resource is necessary to help the states when needed, provide national guidance where it is essential, and coordinate technical capabilities for the nation to respond to epidemics and pandemics as well as bioterrorism threats. Abundant, accurate, and real-time data are essential to evidence-based decision making in times of pandemics, as well as in the ongoing responsibility of government for the health of its residents. The United States has fallen behind in its access to data and the technology to provide decision-support. Catching up must be a very high and immediate priority.

The Ad Hoc Group

The authors are a subset of the members of President Obama's Council of Advisors on Science and Technology (OPCAST) who were involved in producing the six reports dealing with issues related to viral pandemics that his PCAST delivered between 2009 and 2016. In alphabetical order, they are:

Christine Cassel, University of California, San Francisco
Christopher Chyba, Princeton University
Susan L. Graham, University of California, Berkeley
John P. Holdren, Harvard University (OPCAST Co-Chair)
Eric S. Lander, Broad Institute of MIT and Harvard (OPCAST Co-Chair)
Rick Levin, Yale University
Ed Penhoet, University of California, Berkeley
William Press, University of Texas, Austin (OPCAST Vice Chair)
Maxine Savitz, National Academy of Engineering (OPCAST Vice Chair)
Harold Varmus, Weill Cornell Medicine (OPCAST Co-Chair)

The six indicated reports by the Obama PCAST are:

U.S. Preparations for 2009-H1N1 Influenza, 88 pp, August 2009
https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-h1n1-report-final2.pdf

Reengineering the Influenza Vaccine Production Enterprise to Meet the Challenges of Pandemic Influenza, 87 pp, August 2010
https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST-Influenza-Vaccinology-Report.pdf

Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward, 108 pp, December 2010
https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf

Propelling Innovation in Drug Discovery, Development, and Evaluation, 110 pp, September 2012
https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-fda-final.pdf

Better Health Care and Lower Costs: Accelerating Improvement through Systems Engineering, 66 pp, May 2014
https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_biodefense_letter_report_final.pdf

Preparing for Biological Threats, 18 pp, November 2016
https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_biodefense_letter_report_final.pdf

The reports issued by the Ad Hoc Group can be found at http://adhocresponsegroup.org/. In the coming weeks and months, the Ad Hoc Group will be issuing additional reports on other aspects of responding to COVID19 and future pandemics, drawing on these earlier studies and subsequent research and experience.

Appendix: Glossary of Acronyms

ACA	Patient Protection and Affordable Care Act of 2010, known colloquially as the Affordable Care Act, or Obamacare.
AHRQ	Agency for Healthcare Research and Quality. Federal Agency that supports research about how health care systems work and its impacts on patient care.
APHL	Association of Public Health Laboratories. Private association of Public Health Laboratories.
API	application programming interface. A specification of possible interactions with a computer program, allowing other programmers to use it without needing to know its internal details.
ARRA	American Recovery and Reinvestment Act of 2009. Law authorizing economic relief after the 2008 market collapse.
ASTHO	Association of State and Territorial Health Officials. Private professional association of state and territorial health officials.
CARES	Coronavirus Aid, Relief, and Economic Security Act of 2020. Act authorizing major expenditures to support recovery from the Covid19 pandemic economic effects.

C-CDA	Consolidated Clinical-Document Architecture. Software providing the ability to generate industry standard clinical summary, transitions of care, and other documents that meet HL7 standards.
CDC	Centers for Disease Control and Prevention. Federal Agency that supports public health, including data collection from states and municipalities.
CDPH	Chicago Department of Public Health. Municipal agency responsible for public health of the city of Chicago.
CMS	Centers for Medicare & Medicaid Services. Federal agency within HHS that administers the Medicare and (partnering with the states) Medicaid programs.
COVID-19	Coronavirus Disease 2019. Disease caused by the novel coronovirus SARS-CoV2.
CSTE	Council for State and Territorial Epidemiologists. Voluntary association of government epidemiology agencies.
DiSTRIBuTE	Distributed Surveillance Taskforce for Real-time Influenza Burden Tracking and Evaluation. A syndromic surveillance system implemented for community-based monitoring of influenza-like illness.
ED	emergency department of a hospital
HER	Electronic Health Record. Software platform for recording data about patient healthcare, including notes of doctor visits, lab tests, and hospitalizations.
EPIC	One of the major companies selling and supporting electronic health records.
FHIR	Fast Healthcare Interoperability Resources. A standard developed by HL7 for exchanging healthcare information electronically.
HHS	U.S. Department of Health and Human Services. Cabinet level federal agency focused on health, it includes CMS, CDC , NIH, AHRQ, ONC and other agencies.
HIE	Health Information Exchange. Regional organization set up to allow bilateral exchanges of health information by local providers.
HIPAA	Health Insurance Portability and Accountability Act of 1996. Federal law establishing standards for privacy protection of health care information.
HITECH	Health Information Technology for Economic and Clinical Health Act. Part of ARRA focused on accelerating adoption of electronic health records by bonus payments to providers.
HL7	Health Level 7 International. Voluntary organization that sets standards for electronic health care data formats and exchange protocols.
ICU	intensive care unit. Special unit in a hospital caring for extremely sick patients using breathing machines and other high technology.
IPA	Intergovernmental Personnel Act Mobility Program. Program providing for the temporary assignment of personnel bidirectionally between the federal government and state and local governments, universities, and certain nonprofit research organizations.
IT	Information Technology

JASON	Longstanding independent group of scientists that advises the federal government on matters of science and technology, especially national-security related.
NIH	National Institutes of Health. Federal agency that supports basic and clinical research in human biology and health.
NPI	nonpharmacological interventions. Behavioral ways to prevent the spread of infectious illnesses, including social distancing and wearing face masks.
NSSP	National Syndromic Surveillance Program. Use of electronic data for early identification of people with symptoms of an illness that might be infectious.
ONC	Office of the National Coordinator. Office within HHS responsible for overseeing electronic health records.
OPCAST	Obama Administration PCAST. Scientists who advised the president during the Obama Administration, now acting in their individual capacities.
PCAST	President's Council of Advisors on Science and Technology. White House office comprising a group of scientists who formally advise the president.
PHEP	Public Health Emergency Preparedness cooperative agreements. CDC program of financial assistance to state and local public health departments for emergency preparedness.
PPHF	Prevention and Public Health Fund. Appropriations funding stream to CDC for a range of purposes, established as part of ACA.
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2. The virus that causes Covid-19. Related to SARS and MERS, previous epidemics causing respiratory illness.
SHARP	Strategic Health IT Advanced Research Projects. Former university grants program for the development of advanced health information technology administered by ONC.
SMART	Substitutable Medical Application, Reusable Technologies. Open source software that allows developers to create secure apps to access authorized healthcare data directly within an EHR using FHIR protocols.
USDA	U.S. Department of Agriculture. Cabinet level agency overseeing agriculture.
VA	U.S. Department of Veterans Affairs. Cabinet level agency overseeing Veterans Affairs, including a major health care program for veterans.