Executive Perspective

Modernizing Centers for Disease Control and Prevention Informatics Using Surveillance Data Platform Shared Services

Brian Lee, MPH[1], Tonya Martin, MPH[2], Agha Khan, MD, MPH, MBA[3],
Kathleen Fullerton, MPH[4], Wil Duck, MS, MPH[5], Teresa Kinley, MS[5],
Suzette Stoutenburg, PhD[6], Jason Hall, BS[4], Melvin Crum, MS[7],
Macarena C. Garcia, DrPH, MIS[5], Michael F. Iademarco, MD, MPH[5],
and Chesley L. Richards, MD, MPH[1]

Keywords
public health surveillance, public health informatics, interoperability, change management

Public health surveillance is the cornerstone of public health practice.[1] In the United States, the Centers for Disease Control and Prevention (CDC) and states share responsibility for the surveillance enterprise. States have primary responsibility for traditional name-based disease reporting, and they subsequently share anonymized data with CDC. At the same time, CDC maintains many surveillance systems at the federal level. For decades, the number of these single-disease or condition, single-purpose surveillance systems has grown as CDC has needed to expand surveillance data collection to address new public health problems. Currently, CDC has more than 110 surveillance systems. Although these systems provide CDC with the surveillance data needed by the agency, many are experienced as repetitive and burdensome by state and local public health departments, with little or no coordination. This profusion of CDC systems results in duplication of effort, discrepancies among the data elements collected by various programs, and the need to use multiple information technology (IT) systems, which may not be interoperable.

Recognizing these problems, multiple CDC surveillance stakeholders (eg, the US Congress, the Council of State and Territorial Epidemiologists, and the Federal Advisory Committee to the CDC Director) have urged CDC to create a more integrated approach to national public health surveillance and an approach that makes greater use of established data standards.[2] In 2014, CDC launched a new surveillance strategy to streamline and improve the surveillance capabilities of the agency and the public health system.[3] This new strategy is guiding the agency's efforts to eliminate unnecessary redundancies in surveillance data collection and reporting and to increase surveillance IT system interoperability.

The goal is to improve the quality, timeliness, and availability of surveillance data in the public health system (Figure 1). Importantly, the new CDC surveillance strategy also aims to reduce the burden of data exchange on state and local health departments.

As part of this strategy, in May 2016, CDC launched a new public health informatics initiative that is intended to reduce the use of single-purpose (siloed) systems for surveillance and to develop reusable, shared services that can be plugged into multiple surveillance activities and programs in the agency. Shared services are business, science, and IT functions that are standardized and built once for reuse by multiple parts of an organization. At CDC, shared services will allow the agency to use fewer

[1] Office of Public Health Surveillance Services, Centers for Disease Control and Prevention, Atlanta, GA, USA
[2] Office of Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA
[3] National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA
[4] National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA
[5] Center for Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention, Atlanta, GA, USA
[6] The MITRE Corporation, McLean, VA, USA
[7] National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA, USA

Corresponding Author:
Brian Lee, MPH, Centers for Disease Control and Prevention, Office of Public Health Surveillance Services, 1600 Clifton Rd NE, Atlanta, GA 30329, USA.
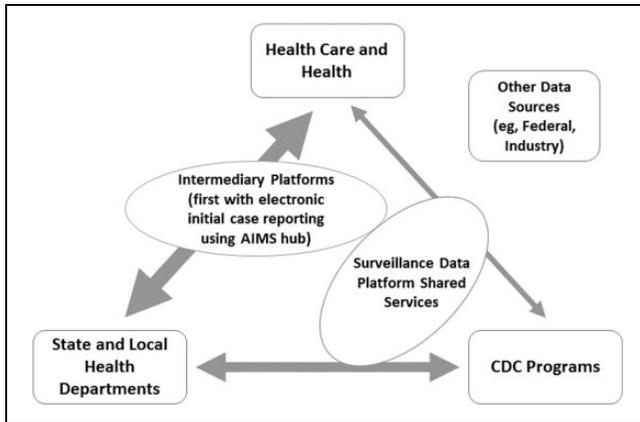Email: brian.lee@cdc.gov

Figure 1. Role of Centers for Disease Control and Prevention (CDC) Shared Data Platform and Shared Services in high-level public health surveillance. Shared Data Platform supports data flow between health care providers and organizations, state and local health departments, and CDC programs. Intermediary platforms (eg, Association of Public Health Laboratories Informatics Messaging Services [AIMS] hub) support data flow between health care providers and organizations and state and local health departments. Other data sources also support surveillance. Arrow width represents the relative volume of bidirectional data flow (ie, thicker width indicates higher volume).

software programs and to reduce the need to create new software programs. Multiple IT and related services can be shared and reused for many different surveillance purposes.

In this *Executive Perspective*, we, members of the CDC team that is creating the new approach, describe the new CDC public health informatics initiative and outline how it aims to modernize the way CDC builds and uses surveillance systems, thus reducing the burden of data exchange on the states. The goal of the initiative is to improve the efficiency of data collection and the availability and interoperability of national surveillance data. By creating new shared data services, CDC aims to improve integration and interoperability of independent, siloed systems; reduce the reporting burden on states; reduce the time needed to define data of interest for national surveillance or to support emergency response (eg, outbreak investigations); provide valuable surveillance data to CDC programs; and enable quicker public health action. CDC is designing these shared services to better connect CDC to the growing number of services and systems that can provide data on public health problems from hospitals, physicians' offices, pharmacies, and other health care sources (Figure 2).[4,5]
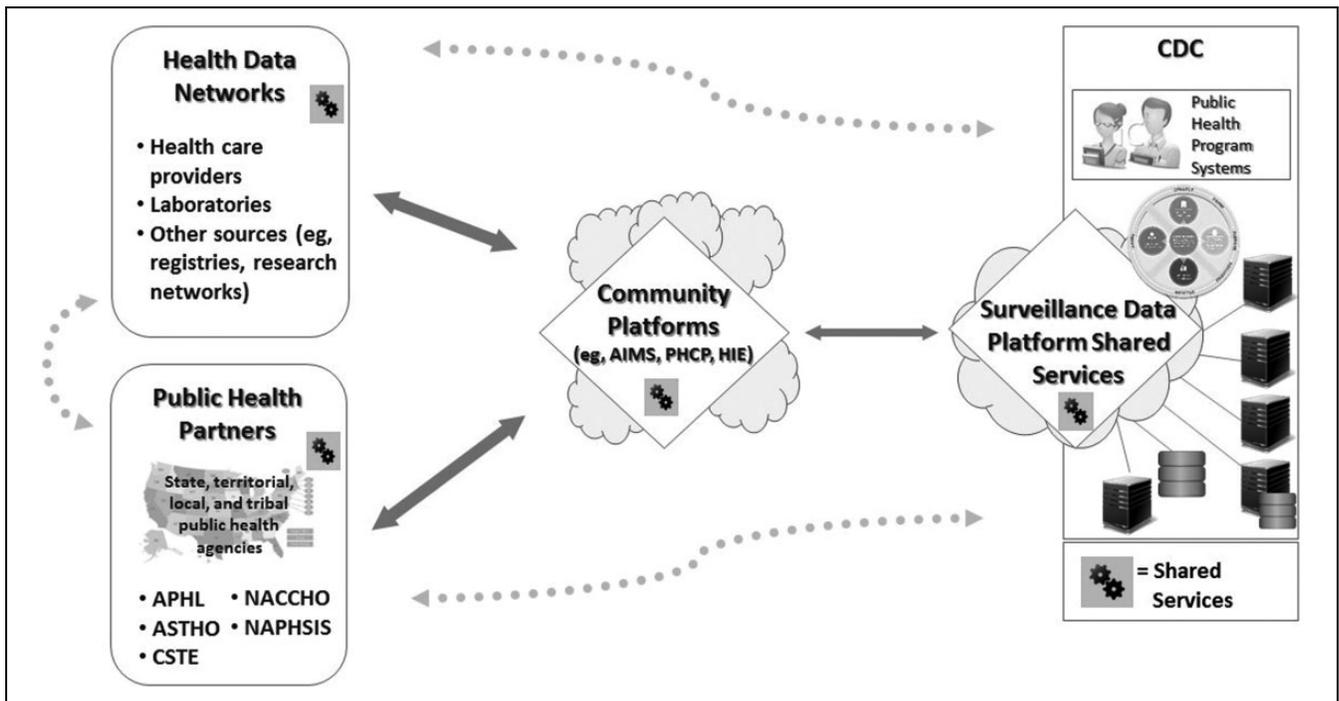

Figure 2. Centers for Disease Control and Prevention (CDC) Surveillance Leadership Board Workgroup vision for public health surveillance data flow. The figure shows data arising from health care and public health partners, transmitted through public health community platforms, entering CDC Surveillance Data Platform and Shared Services, and ultimately routed to appropriate CDC systems, databases, and programs. Shared services (business, science, and information technology functions that are standardized and built once for reuse by multiple parts of an organization) are configured based on appropriate data use for public health program needs. Abbreviations: AIMS, Association of Public Health Laboratories Informatics Messaging Services; APHL, Association of Public Health Laboratories; ASTHO, Association of State and Territorial Health Organizations; CSTE, Council of State and Territorial Epidemiologists; HIE, health information exchange; NACCHO, National Association of City and County Health Officials; NAPHSIS, National Association for Public Health Statistics and Information Systems; PHCP, public health community platform.

Box. Initial list of cross-program shared services[a] identified by the Centers for Disease Control and Prevention Surveillance Leadership Board Workgroup, by tiers and functional areas, January–May 2016

Tier 1

Routing[b]
Secure data exchange[c]
Security[b]

Tier 2

Enterprise licensing[c]
Message testing and validation[b]
Partner and public access to data[d]
Personally identifiable information and protected health information scrubbing[b]
Reference data[e]
Repository, registry, and catalog[f]
Vocabulary[f]

Tier 3

Analysis and visualization tools[e]
Consultation[f]
Data collection tools[c]
Data management and storage[f]
Extract, transform, load[b]
Geocoding[b]
Message mapping guide development[c]
Vocabulary translation[b]

Tier 4

Collaborative tools[f]
Decision support[d]
Evaluation[f]
Geospatial analysis[e]
Linkage[b]
Metadata portal[d]
Metrics and indicators[e]
Natural language processing[b]
Public health decision support[b]
Technical assistance and consultative services[d]

[a]Shared services consist of people, processes, and technology that could meet common needs and that could be sharable, scalable, and easily standardized, while being used simultaneously by more than 1 CDC program, system, or organization. The services are grouped into 4 tiers according to priority (1 = highest priority, 4 = lowest priority).
[b]Core functional area.
[c]Data collection and submission functional area.
[d]Data sharing and dissemination functional area.
[e]Data analysis and visualization functional area.
[f]Data parsing, provisioning, fusing, linking, and preparing functional area.

Choosing the First Shared Services

As part of its new surveillance strategy, CDC established the Surveillance Leadership Board, which comprised leaders from across the agency, and chartered a workgroup to provide recommendations for a new integrated system called the Surveillance Data Platform and Shared Services (Recommendations to the CDC Surveillance Leadership Board from the Platform Workgroup, unpublished memorandum, May 19, 2016). The workgroup was charged with prioritizing which shared services would have the greatest impact on the greatest number of CDC programs. The workgroup included CDC experts with skills in surveillance, epidemiology, data management, IT, information security, and public health informatics. The workgroup initially met in January 2016 and held 6 additional sessions through May 2016 to develop recommendations on the scope, priorities, and design principles for shared services. This process brought stakeholders together and involved the use of a new development technique created by Google Ventures called Design Sprint,[6] a 5-phase framework that helps teams answer critical business questions through rapid prototyping and user testing.

As a result of these initial meetings, the workgroup determined that the new system's platform would exist in an online- and cloud-based informatics environment. We, the team members, also agreed on a set of common design principles that reflected a commitment to transparency, collaboration, quality, and communication.[7] Furthermore, we established that shared services would consist of people, processes, and technology that could meet common needs and that could be sharable, scalable, easily standardized, and usable by more than 1 CDC program, system, or organization.[8] Examples of shared services include vocabulary and data standardization, secure messaging, partner data use, access control, data management, analytics, and visualization. We identified 28 of these cross-program shared services, organized them by priority into 4 tiers, grouped them into 5 functional areas (core; data collection and submission; data analysis and visualization; data parsing, provisioning, fusing, linking, and preparing; and data sharing and dissemination), and validated them as likely to be useful in monitoring multiple conditions and diseases (Box).[9]

Building the First Shared Services

We designed the new informatics initiative as a marketplace approach that would allow CDC programs to select and use the right shared services for their own activities and then to build upon and use the services for specialized surveillance tasks. Planning work included determining platform requirements and service definitions; service infrastructure, cybersecurity, and technical support needs; performance management, evaluation, and monitoring techniques; standards development and harmonization approaches; governance and integration policies; and communication guidelines. We used a phased approach to support quick wins and build on small successes, so our initial efforts focused primarily on creating the
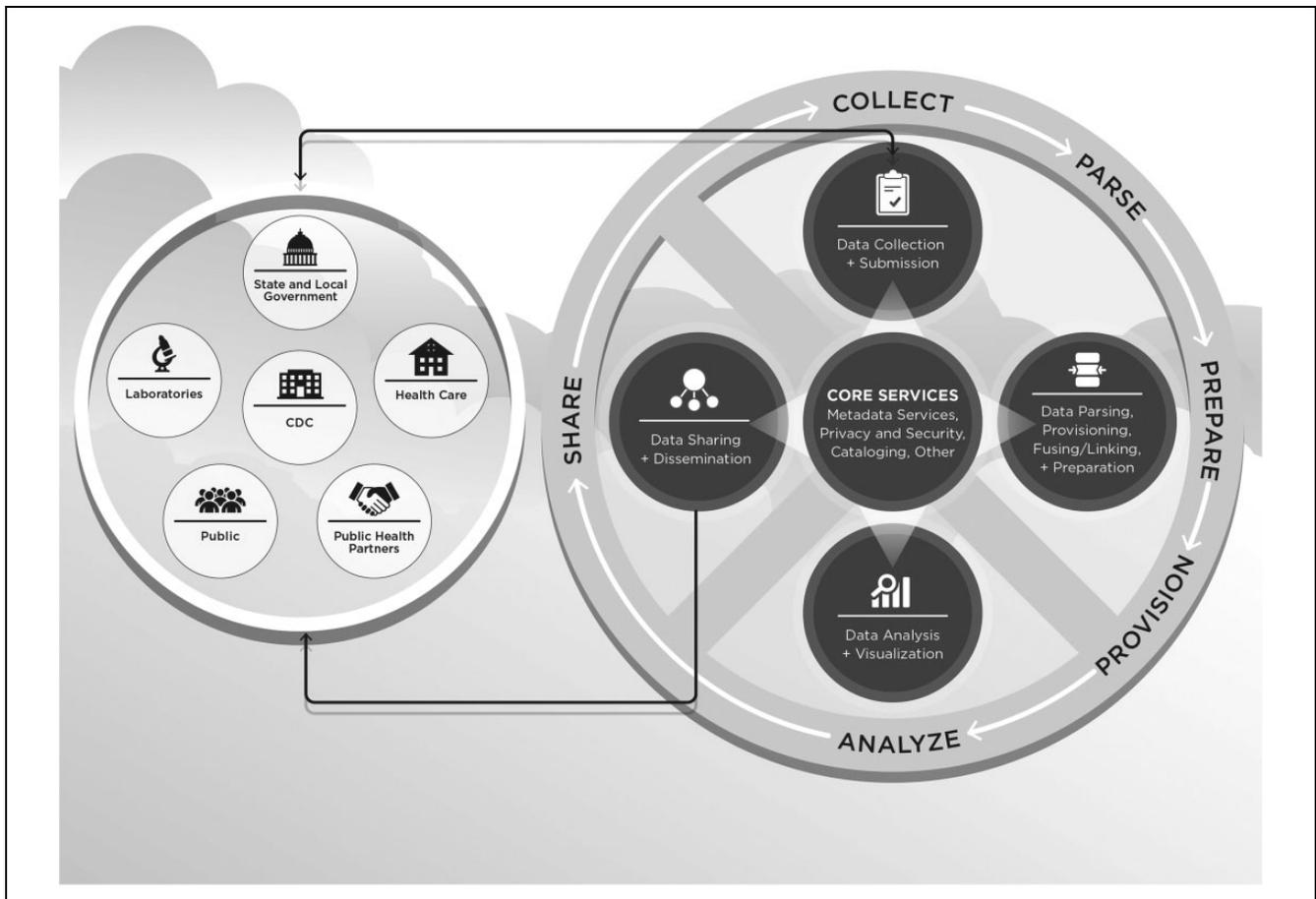
Figure 3. Public health surveillance data lifecycle: Centers for Disease Control and Prevention (CDC) Surveillance Leadership Board Workgroup concept model showing interactions between diverse public health stakeholders (left) and shared services functional areas (right). Shared services consist of people, processes, and technology that could meet common needs and that could be sharable, scalable, and easily standardized, while being used simultaneously by more than 1 CDC program, system, or organization. Shared services may be used within different phases of a lifecycle, are modular and so may be used independently of lifecycle phase, and are run in an online cloud for efficiency and security.

surveillance data platform and an initial set of shared services.

We designed the shared services to (1) use existing health and technology standards for interoperability, including Health Level 7 Fast Healthcare Interoperability Resources,[10] Representational State Transfer,[11] and Open Authentication 2.0[12]; and (2) individually support various activities in the public health surveillance lifecycle (Figure 3).

User-Centered Design of Shared Services

Creating shared services requires close collaboration with the public health practitioners who will use the services. To determine user needs, we conducted an agency-wide landscape analysis to identify common challenges in surveillance activities, discover potential existing services or technology that could be reused, and create partnerships with stakeholders to develop the initial set of services. We interviewed surveillance program staff members, conducted in-house surveys, and classified services by their potential impact and feasibility.

We then began to identify a short list of shared services priorities that would generate support among end users for our shared services approach. The short list focused on common "pain points" of surveillance systems (eg, reporting burden, labor-intensive validation, data harmonization, siloed data sources, and time-consuming visualization). To identify the services on the short list, we considered the potential impact, cost, development time, scalability, and visibility to end users (eg, epidemiologists). By July 2016, we had identified 3 services as the initial targets: (1) data routing (receiving data and determining the programs that needed to receive copies of the data), (2) transformation (automatically modifying the format and structure of incoming data so that they conform to the needs of each surveillance program), and (3) vocabulary (harmonizing and organizing data inquiries and determining acceptable responses).

In September 2016, we held a collaborative design session as part of the 2016 Public Health Informatics Conference of the National Association of County and City Health Officials. We invited state and local public health department professionals, health IT developers, and other public health practitioners to this session.[13] During the session, we received input about our approach and potential use examples about the short list of services.

We completed the initial phase of this public health informatics initiative in August 2017 with the release of the short list of shared services, and we have a plan to add new shared services. As the software moves into production and end users begin training on and working with these new services, we plan to measure the impact of the new system on users and surveillance programs.

Lessons Learned

We found that evaluating and coordinating the IT and surveillance needs of a large agency such as CDC were more challenging than we expected. The work required intense collaboration and relationship building between technical IT specialists and public health practitioners—not always an easy process. The transition from the use of siloed systems to shared services was an exercise in organizational change management. Coordinating with multiple programs, all with different needs and priorities, was challenging. To make it work, we focused heavily on working in a cross-agency manner that included repeated interactions with multiple user groups.

Building the software of the new system required many iterations and refinements to the platform architecture and software source code. Each iteration tested new approaches and new software products to find the right fit for the services that finally made it into the August 2017 release. To build the software quickly and respond rapidly to end-user feedback, we relied on Scrum, an agile software development methodology.[14] It allowed us to iteratively develop the services, with users participating in each small release. Each iteration focused only on a small subset of functionality that end users had identified as a priority. To engage users, we attended monthly workgroup meetings; held "service product owner hours," where designers worked with users for feedback; and held "requirements sessions" with program partners to identify and prioritize service features and functions. We also used collaborative tools such as JIRA (for issue tracking, feature requests, and project management),[15] Confluence (for content collaboration software and wikis),[16] and GitHub (for software source code collaboration).[17]

Next Steps and Conclusion

We think that this short list of shared services has the potential to reduce state and local data reporting burdens and provide infrastructure building blocks for additional services to be developed later. For future phases and additional shared services, we think that an ongoing feedback loop with public health partners outside the agency should be created. CDC will develop new services collaboratively, by encouraging feedback from multiple stakeholders and pursuing dialogue with public health partners, so that ultimately these open-source shared services can be used by many different public health professionals and organizations. We plan to stick with existing health and IT standards so that future shared services will be interoperable and reusable across CDC programs, health care providers, and state, tribal, local, and territorial agencies. Future shared services will be developed in collaboration with other emerging platforms, information exchanges, and public health intermediaries—organizations that facilitate the exchange and routing of health information in the health system—which are all working to connect the health care and public health systems.[18]

References

1. Thacker SB, Qualters JR, Lee LM; Centers for Disease Control and Prevention. CDC's vision for public health surveillance in the 21st century. *MMWR Suppl*. 2012;61(3):3-9.
2. Richards CL, Iademarco MF, Anderson TC. A new strategy for public health surveillance at CDC: improving national surveillance activities and outcomes. *Public Health Rep*. 2014;129(6):472-476.
3. Centers for Disease Control and Prevention. Surveillance strategy: a strategy for improving the Centers for Disease Control

and Prevention's activities in public health surveillance. http://www.cdc.gov/ophss/docs/CDC-Surveillance-Strategy-Final.pdf. Published 2014. Accessed August 9, 2016.

4. Office of the National Coordinator for Health Information Technology. Connecting health and care for the nation: a shared nationwide interoperability roadmap. https://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf. Published 2015. Accessed August 9, 2016.

5. Office of the National Coordinator for Health Information Technology. Connecting health and care for the nation: a 10-year vision to achieve an interoperable health IT infrastructure. https://www.healthit.gov/sites/default/files/ONC10yearInteroperabilityConceptPaper.pdf. Published 2014. Accessed August 9, 2016.

6. Knapp J, Zeratsky J, Kowitz B. *Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days*. New York, NY: Simon & Schuster; 2016.

7. Surveillance Leadership Board Platform Workgroup. Design principles. https://publichealthsurveillance.atlassian.net/wiki/display/PLAT/Design+Principles. Accessed August 29, 2016.

8. Executive Office of the President of the United States. Federal information technology shared services strategy. https://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/shared_services_strategy.pdf. Published 2012. Accessed January 12, 2017.

9. Surveillance Leadership Board Platform Workgroup. Service definitions. https://publichealthsurveillance.atlassian.net/wiki/display/PLAT/Service+Definitions. Accessed August 29, 2016.

10. Health Level Seven International. Fast healthcare interoperability resources. https://www.hl7.org/fhir. Accessed August 9, 2016.

11. Fielding R, Reschke J. Hypertext transfer protocol (HTTP/1.1): semantics and content. https://tools.ietf.org/html/rfc7231. Published 2014. Accessed January 12, 2017.

12. Hardt D. The OAuth 2.0 authorization framework. https://tools.ietf.org/html/rfc6749. Published 2012. Accessed August 9, 2016.

13. Jellison J, Lee B. *Collaborative Design Session for CDC Surveillance Data Platform*. Atlanta, GA: Public Health Informatics Conference; 2016. https://www.eventscribe.com/2016/Informatics/QRcode.asp?Pres=165011. Accessed May 4, 2017.

14. Scrum Alliance. Scrum guide. https://www.scrumalliance.org/why-scrum/scrum-guide. Accessed August 26, 2016.

15. Public Health Surveillance Story Tracker. Jira system dashboard. https://publichealthsurveillance.atlassian.net/secure/Dashboard.jspa. Accessed January 12, 2017.

16. Public Health Surveillance Wiki Confluence. Platform. https://publichealthsurveillance.atlassian.net/wiki. Accessed January 12, 2017.

17. Surveillance Data Platform. GitHub repository for CDC's surveillance data platform. https://github.com/CDCgov/SDP. Accessed January 12, 2017.

18. MacKenzie WR, Davidson AJ, Wiesenthal A, et al. The promise of electronic case reporting. *Public Health Rep*. 2016;131(6):742-746.