June 12, 2017

Honorable Thomas E. Price, M.D.
Secretary
Department of Health and Human Services
200 Independence Avenue, S.W.
Washington, D.C. 20201

Re: Recommendations to Maximize the Value of HealthData.gov

Dear Secretary Price:

This letter summarizes the findings of an accompanying National Committee on Vital and Health Statistics (NCVHS) report entitled "Recommendations to Maximize the Value of HealthData.gov," regarding the initial success of HealthData.gov and opportunities for its enhancement. NCVHS is your advisory committee on health data, statistics, privacy, and national health information policy.

HealthData.gov is the U.S. Department of Health and Human Services (HHS) online health information portal. HealthData.gov facilitates the discovery, access, and use of publicly available health data by providing a searchable online directory of data resources across all HHS operating divisions and several state and municipal open data portals. These data are released to the public under stringent privacy, security, and confidentiality protections. They are used by a diverse set of stakeholders to contribute solutions toward HHS and state government objectives such as identifying quality gaps and health care inefficiencies, enabling researchers to make scientific discoveries, providing communities information about their health problems to inform their community planning efforts, and fostering innovation by entrepreneurs such as for development of applications to assess physicians' services.

NCVHS finds that HealthData.gov has developed into a national resource since its launch in 2010. In a report released in March 2017, HealthData.gov ranked first overall among 297 Federal Government websites based on an analysis of a defined set of factors.[1] Since its development, HealthData.gov has expanded to reflect over 3,000 datasets in diverse areas of clinical outcomes, administration, and health monitoring and evaluation.[2] From 2013 to 2016, web traffic steadily increased from 107,561 to 178,334 visitors annually, and the number of sessions doubled from 508,705 to 1,168,138.[3] Data that are currently cataloged in HealthData.gov and discovered through HealthData.gov or other open data portals, are being used to provide medical consumers with more accurate information on local physicians providing the best valued healthcare treatments,[4] supporting community learning initiatives whereby local communities synthesize data across sectors to identify and address locally-

defined issues,[5] illustrate that incorrectly estimating hospitals' costs caused Medicare to overpay hospitals by $2.6 billion,[6] and numerous other examples. HealthData.gov has cultivated a health data ecosystem comprised of application developers, community groups, private industry, and other entrepreneurs that improve the data's value by repurposing them in creative ways and improving their quality.

The Committee finds HealthData.gov to be a critical infrastructure with a high return on investment. It is meeting its goals of increasing visibility and accessibility of HHS data while simultaneously facilitating the production of health innovations that would not have been possible without its existence. After a thorough review of the current status of HealthData.gov, NCVHS offers the following recommendations to further its potential and enhance its utility:

- **Recommendation 1: HHS should continue to support HealthData.gov and formalize the governance, stewardship, and business operations of HealthData.gov.**

- **Recommendation 2: HHS should maximize the value of HealthData.gov by promoting it, understanding its user-base, and enhancing the platform.**

    - **Recommendation 2a: HHS should develop an integrated and coordinated strategy within its operating divisions to advance the HealthData.gov vision and mission.**

    - **Recommendation 2b: HHS should implement evaluation and performance metrics and solicit data customers' input regularly to increase the use and usefulness of HealthData.gov for diverse consumers, support the work of HHS staff, and facilitate the development of data-driven health innovations.**

    - **Recommendation 2c: HHS should enhance the HealthData.gov platform's capabilities to make the data more meaningful to a range of data customers to extend its reach.**

The implementation of these recommendations will: 1) strengthen HealthData.gov in the near-term to enhance and increase its use by a diverse user base; 2) assist HHS staff in leveraging data across operating divisions for internal use, and; 3) of greatest importance, ensure its long-term success and sustainability as a next-generation platform.

In the attached full report, we include a rationale and specific opportunities for each recommendation. Through its original launch, HealthData.gov spurred healthcare innovation and significantly contributed to the development of an open health data ecosystem. Through these recommendations, NCVHS has identified specific strategies to increase HealthData.gov's ability to facilitate greater improvement in the value and quality of U.S. healthcare and provide significant opportunities to transform publicly available data to innovate, educate, and provide evidence for decision-making to improve the health and well-being of all.

In addition to these actionable short-term recommendations to maximize HealthData.gov's potential, NCVHS identified a number of longer-term transformative strategies to enhance

HealthData.gov and the data sources that its users connect and access. These opportunities fall into areas such as data standardization, data stewardship, developing the capacity to use state-of-the-art technology to accelerate users' ability to understand and develop insights from HHS data, and further developing resources for communities to become learning health systems.

NCVHS remains available to answer questions and is ready to provide you or your staff with additional insights and results that have emerged through this initial analysis and to pursue additional ideas of a long-term vision for HealthData.gov.

Sincerely,
/s/
William W. Stead, MD, Chair
National Committee on Vital and Health Statistics

Attachment

CC: HHS Data Council Co-Chairs

---

[1] McQuinn A, Castro D. Benchmarking U.S. Government Websites. Information Technology and Innovation Foundation. March 8, 2017: https://itif.org/publications/2017/03/08/benchmarking-us-government-websites.

[2] Examples include: disease surveillance and mortality data (e.g. Drug Abuse Warning Network (DAWN); Surveillance, Epidemiology, and End Results (SEER) cancer registry; and Nationally Notifiable Diseases Surveillance System); health care access, cost, and quality data (e.g. Medicare Preventable Hospitalizations, HCUPNet query system of publicly available all-payer health care databases; Minimum Data Set Quality Indicator Report; and address tool to locate Health Professional Shortage Areas); and consumer-oriented topics (e.g. Household Products Database on the health safety of consumer brands; Genetics Home Reference on genetic conditions; and DailyMed compilation of pharmaceutical package inserts).

[3] From NCVHS Data Access and Use Work Group analysis of HealthData.gov Google Analytics data, February 2017.

[4] Lyons L. U.S. News and RowdMap, Inc. team up to help patients make more informed health care decisions. http://www.usnews.com/info/blogs/press-room/2016/01/19/us-news-and-rowdmap-inc-team-up-to-help-patients-make-more-informed-health-care-decisions. Published January 19, 2016. Accessed January 12, 2017.

[5] The Office of the National Coordinator for Health Information Technology is currently supporting 10 local communities to synthesize data across sectors (such as clinical health, public health, criminal justice, housing, schools, and the environment) to help them identify and address locally-defined community health challenges. For example, Cleveland, Ohio is combining weather, housing, school, and clinical data to target and reduce childhood asthma attacks. Other community goals include better coordination of services for Chicago residents who are unstably housed, developing a more streamlined system of care for children with complex behavioral and other medical conditions in Austin, Texas, and reducing inpatient hospitalizations among pediatric patients in Cincinnati, Ohio. AcademyHealth. CHP program participant communities. http://www.academyhealth.org/node/4901. Accessed February 17, 2017.

[6] Weaver C, Mathews AW, McGinty T. Medicare overpays as hospital prices rise. *Wall Street Journal*. April 15, 2015. https://www.wsj.com/articles/medicare-overpays-as-hospital-prices-rise-1429151451. Accessed February 3, 2017.

National Committee on Vital and Health Statistics

Recommendations to Maximize the Value of HealthData.gov

U.S. Department of Health and Human Services
June 2017

This report was written by Erika Martin, PhD, MPH, NCVHS Workgroup on HHS Data Access and Use member, Director of Health Policy Studies Rockefeller Institute of Government-State University of New York, and Associate Professor Department of Public Administration & Policy University at Albany-State University of New York – in collaboration with the NCVHS Data Access and Use Workgroup chair, members and staff.

The National Committee on Vital and Health Statistics (NCVHS), established in 1949, serves as the advisory committee to the Secretary of Health and Human Services on health data, statistics, privacy, and national health information policy. It provides advice and assistance to the Department and serves as a forum for interaction with interested groups on key issues related to population health, standards, privacy and confidentiality, and data access and use. Its members have distinction in such fields as health statistics, electronic interchange of health care information, privacy, confidentiality, and security of electronic information, population-based public health, purchasing or financing health care services, integrated computerized health information systems, health services research, consumer interests in health information, health data standards, epidemiology, and the provision of health services. NCVHS reports and recommendations, and additional information on members, are available online at www.ncvhs.hhs.gov.

Overview

This report summarizes the findings of the National Committee on Vital and Health Statistics (NCVHS) regarding the initial achievements and successes of HealthData.gov and provides recommendations to maximize its value and potential moving forward.

NCVHS is the advisory Committee to the Secretary of HHS on health data, statistics, privacy and national health information policy. The Committee drew on input from the NCVHS Work Group on Data Access and Use,[1] interviews from open data innovators in the public and private sectors, and peer-reviewed research to develop the findings and recommendations in this report.

HealthData.gov is the U.S. Department of Health and Human Services (HHS) online health information portal designed to make "….HHS data easily available and accessible to the public…This information includes clinical care provider quality information, nationwide health service provider directories, databases of the latest medical and scientific knowledge, consumer product data, community health performance information, government spending data and much more..." The website "promotes the availability of HHS data to innovators across the country…to utilize HHS data to create applications, products, services and features that help improve health and health care…By opening up our data, HealthData.gov helps catalyze the emergence of a decentralized, self-propelled 'ecosystem' of innovators across America who leverage HHS data to help consumers, care providers, employers, journalists, local policymakers and others in ways that no one organization could possibly even imagine – let alone build, deploy and scale."[2]

NCVHS finds that HealthData.gov has developed into a national resource since its launch in 2010. In a report released in March 2017, **HealthData.gov ranked first overall** among 297 federal government websites based on an analysis of a defined set of factors.[3] This resource powers visibility, competition, community learning, research, and evidence-based decision-making by allowing data users to search for and access publicly available data resources through one portal. Data are a critical commodity in today's economy, and problems and solutions are best discovered and evaluated when quality data are publicly available and easily accessible.

Further enhancing HealthData.gov as an essential national resource will enable data users to transform public data into health innovations that will increase its benefit for the American public. This requires the continued accessibility of diverse data sources through HealthData.gov;

---

[1] See the workgroup charge and activities at: https://www.ncvhs.hhs.gov/subcommittees-work-groups/working-group-on-hhs-data-access-and-use/charge-of-the-working-group-on-hhs-data-access-and-use/

[2] From HealthData.gov: https://www.healthdata.gov/content/about.

[3] McQuinn A, Castro D. Benchmarking U.S. Government Websites. Information Technology and Innovation Foundation. March 8, 2017: https://itif.org/publications/2017/03/08/benchmarking-us-government-websites.

stringent privacy, security, and confidentiality measures to ensure that data are used with proper stewardship practices; expanding the efficient access and use of publicly available health data; and sustaining the growing community of data entrepreneurs, businesses that innovate using these data, researchers, public health practitioners, medical consumers, communities, health care organizations, and other data consumers interested in using government-produced data for health innovation.

NCVHS HealthData.gov Review Process

The Committee solicited input from the NCVHS Work Group on Data Access and Use, whose members are nationally-recognized in diverse fields including bioinformatics, application development, population health, web analytics, and data stewardship (see Appendix A); conducted a review of relevant peer-reviewed research, government reports, and other gray literature; solicited input from HHS staff who manage HealthData.gov and two state open health data directors whose data are included in HealthData.gov; and analyzed HealthData.gov's web analytics data.

Findings on the Current State of HealthData.gov

HealthData.gov facilitates the discovery, access, and use of publicly available health data by providing a searchable online directory of data resources from all HHS operating divisions and several state and municipal open data portals. All HHS operating divisions collect data that are congruent with the HHS mission to "enhance the health and well-being of Americans by providing for effective health and human services and by fostering sound, sustained advances in the sciences underlying medicine, public health, and social services."[4] Many of these data are released to the public under stringent privacy, security, and confidentiality protections. Their availability may fulfill multiple objectives including identifying quality gaps and health care inefficiencies, empowering data consumers to address health issues, enabling researchers to make scientific discoveries, providing communities with information about their health problems to inform their planning efforts, and fostering innovation by entrepreneurs such as applications to search for physicians meeting specific criteria. Prior to HealthData.gov, a user would search for HHS data related to a specific topic such as prescription drug abuse or obesity by accessing multiple HHS websites and searching each one individually to determine what data were available, whether they were fit for use, and how to access them. This process incurred considerable costs for users, limiting the pace of data-driven health innovation.

[4] US Department of Health and Human Services. [HHS website] https://www.hhs.gov/about/strategic-plan/introduction/index.html#mission. Published March 10, 2014. Accessed February 16, 2017.

HealthData.gov enhances the value of existing HHS data by reducing the effort to locate data and promoting increased data access and use by data consumers who reuse and combine data in new ways to improve health and well-being.[5] Data users can quickly browse resources using the built-in search engine and key terms or navigating the site's icons corresponding to topics such as "quality" and "hospital." This allows them to more easily locate relevant datasets, obtain metadata ("data about the data," such as publisher, years covered, and a description), and access the data by clicking hyperlinks to external HHS or state websites where the data can be downloaded. HealthData.gov does not replace operating divisions making their data available through their own websites and does not require the collection of new data. Rather, this innovation allows diverse internal and external data users to more efficiently discover, browse, access, and understand publicly available health data. As the only public inventory of all HHS data resources, it reduces data silos and allows HHS staff to make better use of data for internal programmatic, planning, and evaluation purposes. It enables data consumers without detailed content expertise on HHS and health data to identify relevant health data, thereby bringing in new data users including application developers, businesses, and community planning groups. It enables data customers to discover and understand additional data that may be related to their interest, thereby expanding the appropriate use of data.

Since its development in 2010, the HealthData.gov directory has expanded to cover over 3,000 datasets in diverse areas of disease surveillance; healthcare access, cost and quality; and consumer oriented datasets.[6] From 2013 to 2016, web traffic steadily increased from 107,561 to 178,334 visitors annually, and the number of times a web page was viewed doubled from 508,705 to 1,168,138.[7] In addition to including data from all HHS operating divisions, one-third of HealthData.gov resources are federated from state and municipality portals; this increases the value of HealthData.gov by helping data users discover related datasets from defined geographic areas that could be linked to develop new insights and innovations. The inclusion of non-HHS sources in HealthData.gov also enhances the quality, amount, and timeliness of HHS

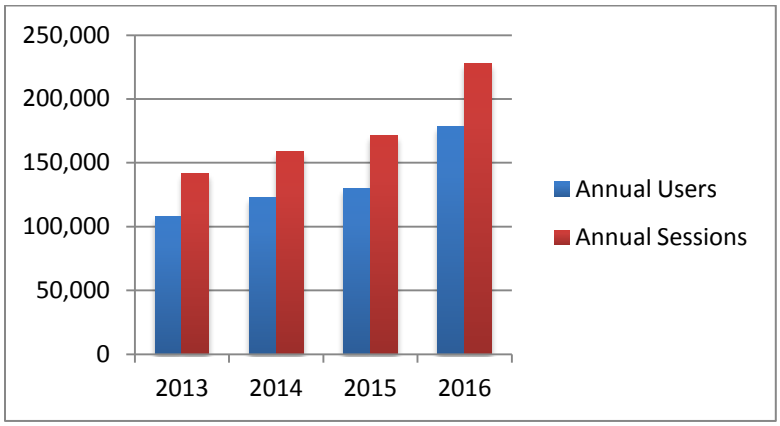
[5] The HealthData.gov catalog does not host data. The data are located on external websites including open data portals (e.g. https://health.data.ny.gov/) and webpages tailored to specific topics (e.g. https://dailymed.nlm.nih.gov/dailymed/index.cfm). The HealthData.gov software continuously reads output files from HHS operating divisions and several state and municipality open data platforms, and updates the HealthData.gov index as data are added or refreshed. The HealthData.gov catalog provides standardized metadata for each data offering and data users can click hyperlinks to be redirected to external websites where they can access data and additional information about the resource.

[6] Examples include: disease surveillance and mortality data (e.g. Drug Abuse Warning Network (DAWN); Surveillance, Epidemiology, and End Results (SEER) cancer registry; and Nationally Notifiable Diseases Surveillance System); health care access, cost, and quality data (e.g. Medicare Preventable Hospitalizations, HCUPNet query system of publicly available all-payer health care databases; Minimum Data Set Quality Indicator Report; and address tool to locate Health Professional Shortage Areas); and consumer-oriented topics (e.g. Household Products Database on the health safety of consumer brands; Genetics Home Reference on genetic conditions; and DailyMed compilation of pharmaceutical package inserts).

[7] From NCVHS Data Access and Use Work Group analysis of HealthData.gov Google Analytics data, February 2017.

data as more data users become acquainted with varied data sources and express desires for additional data or alternative data formats.[8]

Annual Number of HealthData.gov Users and Sessions, 2013 to 2016



Source: NCVHS Data Access and Use Work Group Analysis of HealthData.gov Google Analytics data, February 2017

Data indexed in HealthData.gov have driven economic empowerment, innovation, and transformation in health. Data that are currently cataloged in HealthData.gov and discovered through HealthData.gov or other open data portals have been used to provide medical consumers with more factual and critical information on local physicians providing high-value healthcare treatments,[9] document disparities in the median cost of orthopedic surgeries and appendectomies between hospitals,[10] illustrate that incorrectly estimating hospitals' costs caused Medicare to overpay hospitals by $2.6 billion,[11] incorporate food safety inspection data into Yelp's crowd-sourced restaurant reviews to make timely information readily available to consumers and encourage safer food handling,[12] make food restaurant inspections more efficient through data analytics,[13] support community learning initiatives whereby local communities synthesize data across sectors to identify and address locally-defined issues,[14]

---

[8] Martin EG, Begany GM. Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. J Am Med Inform Assoc 2017; 24(2): 345-351.

[9] Lyons L. U.S. News and RowdMap, Inc. team up to help patients make more informed health care decisions. http://www.usnews.com/info/blogs/press-room/2016/01/19/us-news-and-rowdmap-inc-team-up-to-help-patients-make-more-informed-health-care-decisions. Published January 19, 2016. Accessed January 12, 2017.

[10] *Crain's New York Business*: Health Pulse Extra. January 8, 2014.

[11] Weaver C, Mathews AW, McGinty T. Medicare overpays as hospital prices rise. *Wall Street Journal*. April 15, 2015. https://www.wsj.com/articles/medicare-overpays-as-hospital-prices-rise-1429151451. Accessed February 3, 2017.

[12] Boiles J. San Francisco promotes its restaurant inspection data on Yelp to improve food safety. https://www.codeforamerica.org/featured-stories/san-francisco-puts-restaurant-safety-data-on-yelp. Accessed February 11, 2017.

[13] Wold C. In plain sight: is open data improving our health? http://www.chcf.org/publications/2015/01/in-plain-sight-open-data. Published January 2015. Accessed February 17, 2017.

[14] The Office of the National Coordinator for Health Information Technology is currently supporting 10 local communities to synthesize data across sectors (such as clinical health, public health, criminal justice, housing, schools, and the environment) to help them identify and address locally-defined community health challenges. For example, Cleveland, Ohio is combining weather,

improve physician training by benchmarking medical students' case assignments against the local disease burden,[15] and improve real-time decision making during a natural disaster to save patients' lives.[16] In addition to these uses, HealthData.gov has cultivated a health data ecosystem comprised of application developers, community groups, private industry, and other entrepreneurs who did not previously engage with HHS data.[17] After re-purposing the data to develop applications and other innovations, ecosystem members share back their data in reusable formats. This cycle improves the value of data as they are re-purposed and re-packaged in ways that increase its use by larger audiences while furthering the HHS mission.[18] Beyond the federal open health data ecosystem, HealthData.gov and its growing user community contributed to an "open data revolution," inspiring the development of open health data portals within states and municipalities and further data-driven health innovation.[19]

Recommendations

HealthData.gov is meeting its goals of increasing visibility and accessibility of HHS data while simultaneously facilitating the production of health innovations that would not have been possible without its existence. The Committee finds HealthData.gov to be a critical infrastructure with a high return on investment. Moreover, there are a number of resource-prudent opportunities to strengthen HealthData.gov over the short-term to enhance and increase its use by diverse users, assist HHS staff in leveraging data across operating divisions for internal use, and of greatest importance, ensure its long-term success and sustainability as a next generation platform.

---

housing, school, and clinical data to target and reduce childhood asthmatic attacks. Other community goals include better coordination of services for Chicago residents who are unstably housed, developing a more streamlined system of care for children with complex behavioral and other medical conditions in Austin, Texas, and reducing inpatient hospitalizations among pediatric patients in Cincinnati, Ohio. AcademyHealth. CHP program participant communities. http://www.academyhealth.org/node/4901. Accessed February 17, 2017.

[15] Martin EG, Helbig N, Shah NR. Liberating data to transform healthcare: New York's open data experience. *JAMA* 2014; 311(24): 2481-2482.

[16] Ibid.

[17] One metric of the expanding open health data ecosystem is attendance at the annual Health Datapalooza conference for entrepreneurs, researchers, community groups, the private sector, and other users. This annual conference allows for the exchange of ideas and updates on innovations and technology develops supported by data accessed through HealthData.gov. Its inaugural meeting included 400 participants and featured the work of 21 developers who were provided with web links to a dozen federal datasets and tasked with developing applications to present publicly. It has since grown significantly, with a current attendance of nearly 2,000 individuals from multiple sectors. Presentations communicate the state-of-the-art scientific developments in health data use, government priorities, new health data applications, and examples of business and communities using data to improve health outcomes. (G Downing, personal communication, February 22, 2017.)

[18] National Committee on Vital and Health Statistics, Toolkit for communities using health data: how to collect, use, protect, and share data responsibly. http://www.ncvhs.hhs.gov/wp-content/uploads/2013/12/Toolkit-for-Communities.pdf. Published May 2015. Accessed March 24, 2017. Martin EG, Begany GM. Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. J Am Med Inform Assoc 2016; Epub ahead of print. Harrison TM, Pardo TA, Cook M. Creating open government ecosystems: a research and development agenda. Future Internet 2012; 4(4): 900-28.

[19] Martin EG, Helbig N, Shah NR. Liberating data to transform healthcare: New York's open data experience. JAMA 2014; 311(24): 2481-12; Martin EG, Shah NR, Birkhead GS. Unlocking the power of open health data: a checklist to improve value and promote use. 2017 Mar 1. doi: 10.1097/PHH.0000000000000561. [Epub ahead of print].

Recommendation 1

HHS should continue to support HealthData.gov and formalize the governance, stewardship, and business operations of HealthData.gov.

HealthData.gov has transitioned from its initial developmental phase to being inclusive of all HHS operating divisions, as well as multiple state and local agencies. By actively outreaching to non-government data consumers through key stakeholder forums, such as the annual Health Datapalooza, HHS has contributed to the development of a health data ecosystem.

The Committee noted other U.S. state and local jurisdictions, which regularly attract diverse users to their open data sites, have established strong governance to make explicit standards for how the data they provide are produced and presented; recommended business rules for data access, use, and publication; and made the release of easily discoverable, accessible, and usable public data a routine public health activity.[20] Clear governance sets ground rules helpful to streamlining operations, facilitating the implementation of strategies to more effectively target the site to diverse users' needs, and ensures long-term sustainability. They are also critical to consumer trust in data use.

Building from prior work in this area,[21] the specific strategies HHS could implement to formalize governance and stewardship of HealthData.gov include:

- Establishing standards both for datasets and other aspects of HealthData.gov – and piloting these standards for datasets that are currently indexed.
- Creating a feedback loop to inform HHS staff and data stewards about how their data are being used and protected in ways that are responsive to consumer demand.
- Developing a centralized HHS repository of policies and recommended guidelines for preparing data that are more discoverable and usable (such as developing metadata and tags, and providing alternative formats), privacy and security management, and other constructive data policies.
- Establishing mechanisms to better integrate HealthData.gov into broader HHS data efforts. This could be achieved through establishment of a governance process within HHS.

[20] Martin EG, Shah NR, Birkhead GS. Unlocking the power of open health data: a checklist to improve value and promote use. 2017 Mar 1. doi: 10.1097/PHH.0000000000000561. [Epub ahead of print]; Martin EG, Begany GM. Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. J Am Med Inform Assoc 2017; 24(2): 345-351.
[21] National Committee on Vital and Health Statistics, Toolkit for communities using health data: how to collect, use, protect, and share data responsibly. http://www.ncvhs.hhs.gov/wp-content/uploads/2013/12/Toolkit-for-Communities.pdf. Published May 2015.

Recommendation 2

HHS should maximize the value of HealthData.gov by promoting it, understanding its user-base, and enhancing the platform.

NCVHS identified opportunities to make HealthData.gov more efficient for diverse users to discover, browse, and access data from HealthData.gov; and encourage visitors to return to the site thereby bringing in additional web traffic from new users. Underlying all these strategies is a continued commitment to good data stewardship, including adhering to best practices for data protection, security, and confidentiality. The three parts of this recommendation are intended as short term and practical actions that HHS could undertake within an 18-month period.

Recommendation 2a

HHS should develop an integrated and coordinated strategy within its operating divisions to advance the HealthData.gov vision and mission.

To maximize the use and benefits of HealthData.gov, a more systematic approach is needed to promote and increase awareness. Open data platforms are a relatively new technology to facilitate access to and efficient use of public data, and HealthData.gov was a groundbreaking undertaking in 2010. While developing and adapting organically to a rapidly evolving health data environment, it has succeeded in achieving valuable outcomes related to its mission. The HealthData.gov aspiration to drive economic empowerment, innovation, and transformation in health and health care can be attained at a higher level by focusing future activities around a focused mission. Specific strategies could include to:

- Promote HealthData.gov as a brand, reinforcing a consistent message about its vision and mission, adding a tag line such as "Public Data Released" to materials to make the website recognizable and memorable.
- Inspire data users to do more with the data by highlighting success stories, encouraging HHS data custodians to post blogs and other reports about their data on HealthData.gov, and organizing regular user sessions where HHS staff and external users can share ideas.
- Continue to support developer challenge competitions; and provide seed funding opportunities to support solutions that make HHS activities more efficient and effective, cultivate the open health data ecosystem, and foster a culture of consumer engagement in driving health innovation.

- Develop data sharing partnerships with as many states as possible to support the diffusion of best practices related to the use of these data to improve state-based programs such as public health and Medicaid.

Recommendation 2b

HHS should implement evaluation and performance metrics and solicit data customers' input regularly to increase the use and usefulness of HealthData.gov for diverse consumers, support the work of HHS staff, and facilitate the development of data-driven health innovations.

There is a limited understanding of current HealthData.gov users and how to release data to achieve maximum impact. This is relevant as HealthData.gov draws in new data users who have not traditionally interacted with HHS, including application developers, businesses, and community planning groups. As health data reach new audiences with diverse ideas and skill-sets, a greater understanding of who the data customers are and how to release data to maximize their experience and encourage their regular participation is needed. Traditional audiences for public health data include researchers and community health groups. Fostering new health innovation from public data requires the continued engagement of new and diverse data consumers who have different data requirements, intended uses, technical capabilities, and content expertise. For example, researchers frequently desire historical data over many years to document trends whereas an entrepreneur seeking to connect air quality control, education, transportation networks, and environmental health may only require the most recent data and needs to efficiently understand the structure and format of the data to determine whether these data can be combined meaningfully. Specific strategies to enhance HHS data customer needs could include to:

- Implement a mechanism to solicit input systematically from diverse internal and external data users.
- Participate in the Digital Analytics Program[22] and use web analytics to improve understanding of HealthData.gov customers and their use patterns, thereby facilitating better tailoring of content to users, evaluation of dissemination strategies, marketing of data to specific groups, increasing the "findability" of data, and other benefits that result from enhanced data for decision-making and customer engagement.
- Develop a mechanism for engaged data users to share their solutions and coding resources to educate and accelerate other users' abilities to access and use HealthData.gov resources.

[22] See https://www.digitalgov.gov/

Recommendation 2c

HHS should enhance the HealthData.gov platform's capabilities to make the data more meaningful to a range of data customers to extend its reach.

HHS can adapt the HealthData.gov site to improve the user experience for diverse audiences. Efficient processes for users to discover, browse, and access data encourage the innovative use of data. The web industry is continuously updating how users interface with information to ensure that consumers find sites engaging, usable, and relevant for their purposes—and are likely to return. Similarly, HealthData.gov needs to use state-of-the-art technology in line with web industry standards so that it continues to be acceptable and usable to a range of data customers, to help users more efficiently and effectively discover and understand data, and to ensure data customers view the HealthData.gov brand as a leading source of publicly available health data. One aspect of enhanced capabilities is improved indexing, which can improve the discoverability of data within HealthData.gov for internal and external users and potentially benefit the HHS data enterprise more broadly by limiting redundancies. Specific strategies include to:

- Maintain the website in a modern design so that users view it as a current and well-maintained platform.
- Implement additional web features previously recommended by NCVHS, including: providing recommendations for other data offerings that may be relevant based on other users' experiences and interests ("other users who used this also look at…"), enhancing the applicability of the content by allowing users to identify their roles (e.g., provider, employer, consumer, etc.), and providing users with drill-down menus that display characteristics such as the format and metadata, allowing for customer ratings of datasets.[23]
- Improve the automated processes currently in use to extract information from the websites and open data portals that host data to ensure that HealthData.gov entries are timely, complete, and reflect all updates.
- Improve the quality of HealthData.gov's metadata tagging so that information can be more reliably indexed and discovered; in particular, consider feasibility studies of using existing curated vocabularies such as the National Library of Medicine's Medical Subject Heading (MESH) to tag HHS datasets.[24]

[23] Green LA. Steps to improve the usability, use and usefulness of selected online HHS data resources. www.ncvhs.hhs.gov/wp-content/uploads/2014/05/140320lt.pdf. Published March 20, 2014. Accessed January 12, 2017.
[24] Marc DT, Beatti J, Herasevich V, et al. Assessing metadata quality of a federally sponsored health data repository. AMIA Annu Symp Proc 2016: 864-873. Marc DT, Khairat SS. Medical subject headings (MeSH) for indexing and retrieving open-source healthcare data. Stud Health Technol Inform 2014; 202: 157-60.

- Add capabilities to allow users to post application programming interfaces (APIs) and other products that can be searched by other users.

A Strengthened HealthData.gov Benefits the HHS Mission

In less than a decade, HealthData.gov has made considerable strides towards enabling diverse stakeholders including entrepreneurs, health professionals, researchers, policymakers, community groups, and others to use public data to drive economic empowerment, innovation, and transformation in health and healthcare. By allowing users to more efficiently discover, browse, and access publicly available health data in a centralized portal, HHS increases the value of its existing resources by empowering the development of new health and healthcare innovations and catalyzing an open health data ecosystem. HealthData.gov can also be an engine for learning health systems at both the national and community levels, which the Committee has identified as consisting of a wide range of organizations with significant unmet data needs.[25]

Implementing these recommendations would have broad and beneficial effects for both HHS as well as external HealthData.gov stakeholders. Anticipated future benefits of an enhanced HealthData.gov include:

- Policymakers and other data consumers within HHS will have significantly improved access to timely and interpretable data across the entirety of HHS databases.
- Database developers within HHS can begin to share data models, data dictionaries, and data standards across HHS operating divisions, thus reducing redundancies and costs associated with building and maintaining databases. Sharing database designs will foster interoperability and information exchange across HHS agencies and types of data (such as administrative, survey, or clinical).
- Groups with deep expertise in metadata such as the National Library of Medicine could facilitate the implementation of metadata standards, and best practices for data analysis across HHS databases that can also be referenced in the HealthData.gov index.[26]

In addition to these actionable short-term recommendations to help HealthData.gov reach its full potential, NCVHS identified a number of longer-term transformative strategies to enhance

[25] NCVHS Letter to the HHS Secretary, "Recommendations on Measuring Health at the Community Level – Opportunities for HHS Leadership," May 2, 2017: https://www.ncvhs.hhs.gov/wp-content/uploads/2013/12/Framework-from-Measurement-Report-FINAL-with-cover-letter-optimized.pdf

[26] The HealthData.gov software refreshes its index by continuously reads output files from HHS operating divisions. Increasing the breadth, quality, and usefulness of metadata in HealthData.gov requires developing metadata standards for all data collected within HHS operating divisions.

HealthData.gov and the data sources that its users connect and access. Moving toward the future, the Committee identified additional improvements for consideration such as:

- Including links in HealthData.gov to datasets published with articles in PubMed Central and leveraging the National Library of Medicine's expertise to enhance best practices for HealthData.gov data curation and ontologies (e.g., types, properties, and interrelationships among data).
- Developing a mechanism for user-driven forms of indexing and labeling the resources through social tagging to increase efficiency in finding desired materials, reduce redundancy in searches and increase transparency of data resources.
- Developing capacity through toolkits, online applications, and other modalities to enable communities to synthesize and integrate multiple datasets to address local health challenges consistent with the Office of the National Coordinator for Health Information Technology's national interoperability roadmap, thereby being a national resource and engine for community and national learning health systems.

In summary, it is the Committee's finding that HealthData.gov has proven value and that the system offers great potential to further benefit the health and well-being of all Americans.

Appendix A

NCVHS Work Group on HHS Data Access and Use

Vickie M. Mays, PhD, MSPH, Chair
Professor and Director
UCLA Department of Psychology & Health Services
Los Angeles, CA

James Sorace, MD MS, Lead Staff
Senior Medical Officer
Office of Science and Data Policy
Office of the Assistant Secretary for Planning and Evaluation, DHHS

Helga Rippen, MD, PhD, MPH, FACPM
Chief Medical Officer
Alertgy
Chapin, SC

Gillian R. Barclay, DrPH, MPH, DDS
Troy, MI 48098

Chris Boone, Ph.D., MHA, FACHE
Vice President, Real World Informatics
Avalere Health, An Inovalon Company
Washington, DC

P. Kenyon Crowley, MBA, MS, CPHIMS
Deputy Director
Center for Health Information and Decision Systems
Program Director, Center of Excellence in Health IT Research
Robert H. Smith School of Business
University of Maryland
College Park, MD

Leslie Pickering Francis, JD, PhD
Distinguished Professor of Law and Philosophy
Alfred C. Emery Professor of Law
SJ Quinney College of Law
University of Utah
Salt Lake City, UT

Mohit Kaushal, MD, MBA
Partner

Aberdare Ventures
San Francisco, CA

Richard Leadbeater
Esri Inc.
Global Manager: State/Provincial Government
Industry Solutions and Government Trade Associations
Laurel, MD

Erika Martin, PhD, MPH
Director of Health Policy Studies
Rockefeller Institute of Government-State University of New York
Associate Professor, Department of Public Administration & Policy
University at Albany-State University of New York
Albany, NY

Joshua Rosenthal, PhD
Co-Founder & Chief Science Officer
RowdMap
Louisville, KY

Mark Savage, JD
Director, Health Information Technology Policy and Program
National Partnership for Women & Families
Washington DC

Paul C. Tang, MD
Vice President
Chief Medical Information Officer
Palo Alto Medical Foundation
Mountain View, CA