

HealthData.gov, Its Early Successes, and Recommendations for Maximizing Its Value

National Committee on Health and Vital Statistics

Work Group on HHS Data Use and Access

DRAFT

February 16, 2017

Separate 2-page letter for the Secretary will be written based on the report

HealthData.gov is the US Department of Health and Human Services (HHS) online health information portal that drives economic empowerment, innovation, and transformation in health. This national resource powers visibility, competition, community learning, research, and evidence-based decision making by allowing data users to search for and access publicly available federal, state, and city data resources in one centralized location. Increased accessibility and efficient use of high-value public data by diverse users is critical to creating the information needed to improve population health and the delivery of value-based health care through activities such as: business innovations; identifying inefficiencies in health services; monitoring fraud and abuse; performing data-driven community health needs assessments, planning, and monitoring; empowering health consumers; and enabling improved decision-making by patients, providers, and policymakers. Data are a critical commodity in today's economy, and problems and solutions cannot be fully discovered and evaluated when data are absent.

HealthData.gov powers visibility, competition, community learning, research, and evidence-based decision making by allowing data users to search for and access publicly available federal, state, and city data resources in one centralized location.

The purpose of this report is to identify opportunities to further enhance HealthData.gov as a critical national resource that will enable data users to transform public data into health innovation that will benefit all Americans. This will require the continued availability of information on HealthData.gov; stringent privacy, security, and confidentiality measures so that data may be 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. The recommendations provided are supported by peer-reviewed research, experiences from open data innovators in the public and private sectors, and the consensus of the NCVHS expert work group.

In less than a decade, HealthData.gov has already made considerable strides towards enabling diverse stakeholders including entrepreneurs, health professionals, researchers, policy makers, community groups, and others to use public data to drive economic empowerment, innovation, and transformation in health. Providing data customers with a “one-stop shopping” experience to efficiently discover and access public health data in a centralized location has already led to important innovations that can improve the delivery of high value health care, thereby stimulating competition within health care markets for quality care while simultaneously improving population health. HealthData.gov can also be an engine for national and community learning health systems. These five recommendations and associated strategies can help HealthData.gov meet its full potential, further increasing the value of public data. This vision and goals also align with charge to the bipartisan Commission on Evidence-Based Policymaking.

HealthData.gov: One-Stop for Health Data

HealthData.gov facilitates the discovery, access, and use of health data that is publicly available by providing a searchable online directory of data resources from all HHS operating divisions and several state and city open data portals. It also allows data stewardship practices to be coordinated to permit data release while minimizing risks. This innovation provides data customers with a “one-stop shopping” experience for discovering and accessing health data, and determining their fitness for use. HealthData.gov does not replace operating divisions making their data available through their own websites and it does not require the collection of new data. However, its searchable data directory and provision of standardized metadata, or “data about the data” (such as a dataset description, publisher, and release date) is a new centralized mechanism allowing users to quickly locate data from each operating division and learn how to procure data. The one-stop shopping experience makes it more efficient for data customers to search for and discover data related to their interests across all HHS operating divisions and also in selected state and cities. This enables data customers without detailed content expertise on HHS and health data to identify relevant health data, thereby bringing in new data users. It also enables data customers to discover additional data that may be related to their interest, thereby breaking down data silos. HealthData.gov enhances the value of existing HHS data by improving process efficiency for locating data and promoting increased data usage and the combination of data in new ways among diverse data consumers.

At \$100,000 per year, HealthData.gov is inexpensive to maintain.

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.”¹ Many of these data are released to the public consistent with government transparency efforts employing stringent privacy, security, and confidentiality protections in order to fulfill multiple objectives: identify quality gaps and health care inefficiencies, empower data consumers to address health issues, enable researchers to make scientific discoveries, provide communities information about their health problems to inform their community planning efforts, and foster innovation by entrepreneurs such as applications to search for physicians.

Prior to HealthData.gov, HHS operating divisions had their own policies and procedures for accessing their public health data, typically either through a website for free download or through a request process. In addition to downloadable data, users seeking datasets with personally identifiable health information would contact specific operating divisions to submit a formal request. As good data stewards, operating divisions reviewing these requests assessed risk of identification, adherence to data protection requirement, and approval by an Institutional Review Board if needed. To comprehensively identify data related to a specific topic such as prescription drug abuse or smoking behavior, a data user may need to visit multiple HHS websites to determine what data are available, whether they are fit for use, and how to access them. This process to search multiple sites for potentially relevant data and assess their quality, usability, and fitness for use had a considerable search cost, limiting the pace of data-driven health innovation. This process is similar to patients who, in the absence of

integrated electronic health records available in patient portals, need to piece together their health status by accessing various pieces of electronic or paper-based health information from multiple clinics and locations.

HealthData.gov has adopted a “do no harm” policy, in that it does not provide users with unrestricted access to all datasets listed in its directory. Access remains under the control of the HHS operating divisions. When it is necessary to restrict data access to protect privacy and other public interests, the user is referred to the appropriate mechanism to request access. This allows the operating divisions to retain control over data to release to assure that the public trust is not violated.

Figure 1. Processes to Search for, Browse, and Examine Data for Relevance and Fitness for Use Before and After HealthData.gov

[INSERT INFOGRAPHIC HERE: For a given health topic, show the process that a user would need to undergo to search for, browse, and examine data to determine what might be relevant and then how to access the data under the “old” pre-HealthData.gov method and the “new” HealthData.gov method.]

Current State of HealthData.gov

HealthData.gov was developed in 2010 as an open-source platform that “federates” information about data offerings from HHS operating divisions and some state and local governments. It does not host data, but its software continuously reads output files from HHS operating divisions and other government agencies’ existing platforms and automatically updates the HealthData.gov director as data are added or refreshed. A few smaller HHS operating divisions contribute their information manually. HealthData.gov contracts with a vendor to support its directory and federation activities. The current five-year contract is approximately \$100,000 annually; to put this in context, the 2015 allocations to the 50 states and District of Columbia to collect data for the most recent waves of the National HIV Behavioral Surveillance survey, National HIV Surveillance System, and Medical Monitoring Project are \$7.9 million, \$56.9 million, and \$13.6 million, respectively.ⁱⁱ

There are currently over 3,000 datasets in HealthData.gov that provide access to diverse information on health issues, in the areas of clinical outcomes, administration, and monitoring and evaluation. Examples include disease surveillance and mortality data, health cost and quality, and specific health topics such as breast cancer or prescription drug overdoses.ⁱⁱⁱ As shown in Table 1, searching for three salient health topics (overdose, hepatitis, and Medicare) yields varied data offerings from multiple HHS operating divisions and states. They include multi-year datasets (e.g. the Drug Abuse Warning Network), links to existing dashboards (e.g. the Maryland Department of Health and Mental Hygiene Dashboard Measures and CDC WONDER), surveillance data (e.g. overdose deaths), health care quality measures (e.g. Medicare preventable hospitalizations), and administrative data (e.g. enrollment into Pre-Existing Insurance Condition plans).

Table 1. Sample HealthData.gov Queries and Resulting Data Sources

Query	Dataset Names and Sources
"Overdose"	<ul style="list-style-type: none"> • Maryland Department of Health and Mental Hygiene Dashboard Measures (source: State of Maryland) • Drug Abuse Warning Network (DAWN), multiple years (source: Substance Abuse & Mental Health Services Administration) • Overdose Deaths, Impaired Motor Vehicle Crashes, and Individuals in State-Supported Substance Abuse Programs by Month (source: State of Maryland) • Total Number of Drug- and Alcohol-Related Intoxication Deaths Occurring in Maryland, 2007-2015 (source: State of Maryland) • Total Number of Drug Intoxication Deaths by Selected Substances: 2007-2015 (source: State of Maryland) • VSRR – Quarterly Provisional Estimates for Selected Indicators of Mortality (source: Centers for Disease Control and Prevention)
"Hepatitis"	<ul style="list-style-type: none"> • National Notifiable Diseases Surveillance System disease tables for notifiable diseases, multiple (source: Centers for Disease Control and Prevention) • Infectious Disease Cases by County, Year, and Sex, 2001-2014 (source: State of California) • Hepatitis B Birth Dose Vaccination Rates: Beginning 2012 (source: State of New York) • Cancer Incidence – Surveillance, Epidemiology, and End Results (SEER) Registries Limited-Use (source: National Cancer Institute) • CDC WONDER Online Tuberculosis Information System (OTIS) (source: Centers for Disease Control and Prevention) • CDC WONDER Sexually Transmitted Disease (STD) Morbidity (source: Centers for Disease Control and Prevention)
"Medicare"	<p>Of the 920 search results, the first five results are:</p> <ul style="list-style-type: none"> • Medicare Preventable Hospitalizations (source: State of Oklahoma) • Monthly Pre-Existing Condition Insurance Plan Enrollment (source: Centers for Medicare & Medicaid Services) • Comprehensive Primary Care Initiative – Participating Primary Care Practices (source: Centers for Medicare & Medicaid Services) • Number of Estimated Eligible Uninsured People for Outreach Targeting (source: Centers for Medicare & Medicaid Services) • Medicare Provider Data – Physician and Other Supplier Public Use File (source: Centers for Medicare & Medicaid Services)

Source: NCVHS Data Access and Use Work Group Analysis of the HealthData.gov Search Engine, February 2017

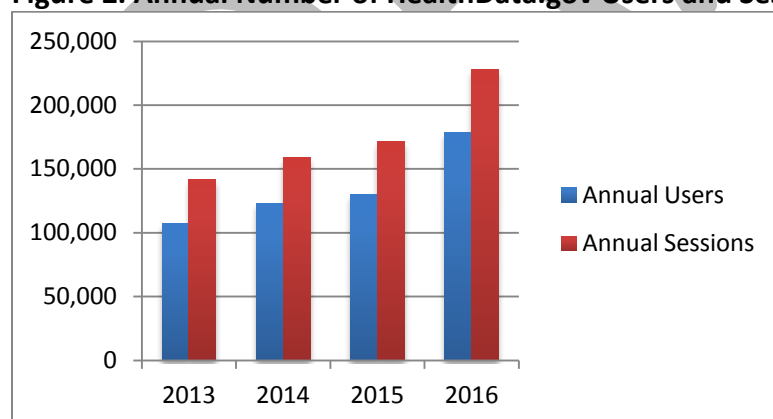
In addition to covering all HHS operating divisions, one-third of HealthData.gov resources are federated from state and local portals. The inclusion of non-federal health data increases the value of HealthData.gov by helping consumers discover related datasets from geographic areas that could be linked to develop new insights and innovations. Covering data outside HHS can also enhance the quality, amount, and timeliness of HHS data as more data customers become acquainted with varied data sources and express desires for additional data or alternative data formats.

HealthData.gov provides multiple features that allow users to browse data. The main web page contains icons that link users to data offerings grouped by topic, such as quality, Medicare, hospital, and inpatient. Users can search for data offerings by topic either through a search engine or clicking on tags (key terms associated with datasets, such as “Medicaid,” “substance abuse,” “tobacco,” and “inpatient”). After reaching a data page, the users can review metadata, or “data about the data” (such as publisher, dates of modification and release, and a contact email) and click links to be redirected to the source data. Many data are available in multiple formats. Overall these search capabilities, improved process efficiency, and provision of standard metadata allow for faster and broader data discovery by diverse users.

HealthData.gov had three-quarter million sessions over 5 years, and the number of visitors increased ten-fold from 2013 to 2016

From March 30, 2012 through January 3, 2017, there were 748,743 sessions encompassing 3.31 million page views and 2.35 million unique page views. Web traffic has increased steadily from 17,561 visitors in 2013 to 178,334 visitors in 2016, and sessions have doubled from 508,705 sessions in 2013 to 1,168,138 sessions in 2016 (see Figure 2).

Figure 2. Annual Number of HealthData.gov Users and Sessions, 2013 to 2016

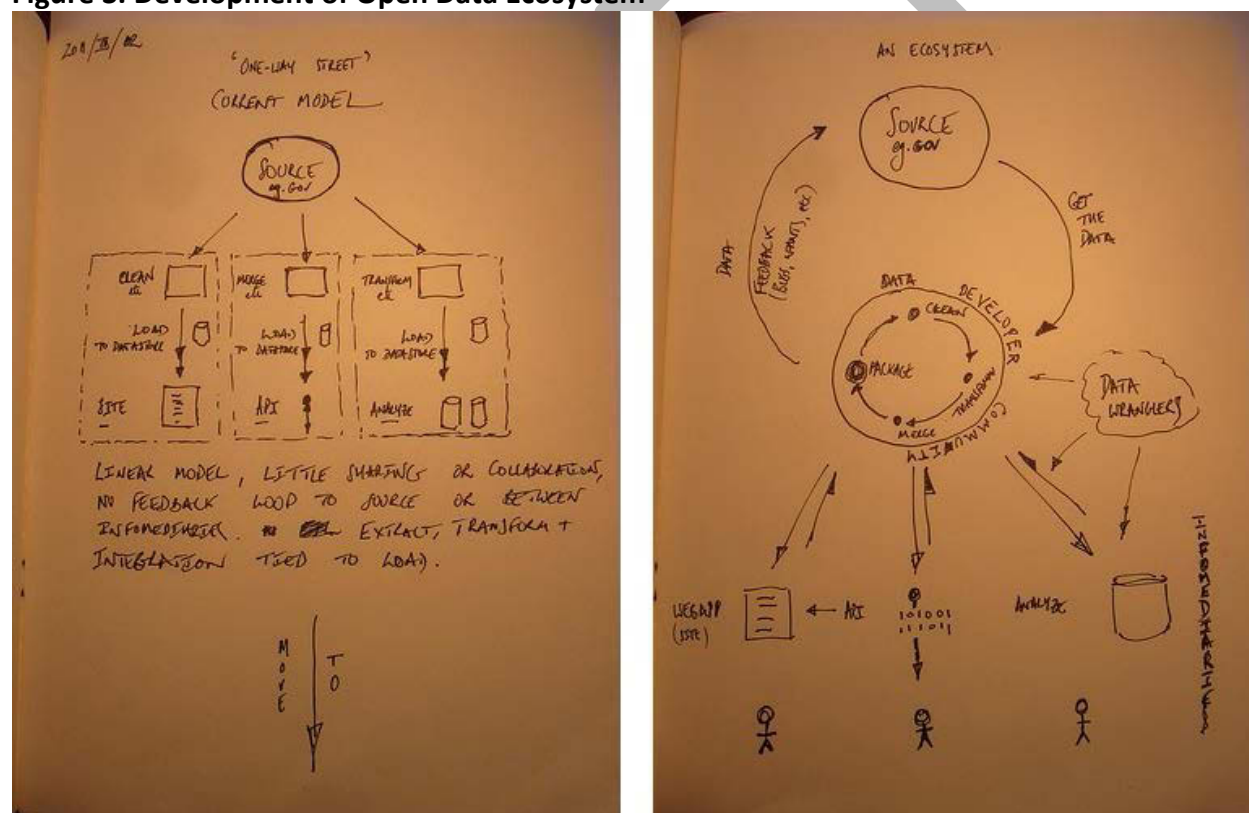


Source: NCVHS Data Access and Use Work Group Analysis of HealthData.gov Google Analytics data, February 2017

Beyond increased web traffic, HealthData.gov has fostered the development of an open health data ecosystem whereby data customers repurpose data to develop applications and other innovations and share back their data in a reusable format. This cycle improves the value of

data as they are repurposed and repackaged in better ways (see Figure 3). One metric is the size of 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 developments supported by data accessed through HealthData.gov. By drawing together a diverse group of data consumers interested in open health data, the conference aims to “build momentum toward a vibrant health information economy.”^{iv} Its inaugural meeting included approximately 200 participants and featured the work of 21 developers who were granted access to 10 federal datasets with which they created applications from the data. It has since grown exponentially, 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 businesses and communities using health data to improve health outcomes. The conference also includes a challenge competition for developers.

Figure 3. Development of Open Data Ecosystem



Source: Adapted from Open Knowledge International¹

Early Successes of HealthData.gov

HealthData.gov drives economic empowerment, innovation, and transformation in health care by powering visibility, competition, community learning, research, and evidence-based decision

¹ <https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/> ** This needs to be redone and permission to adapt is needed, it is copied from website. **

making. Since its creation in 2010, HealthData.gov has grown organically to expand the coverage of its data repository and cultivate a health data ecosystem. This has already facilitated the production of health innovations that would not have been possible without its existence. Below are specific use cases of open data for health innovation, using data that are all currently available in the HealthData.gov repository.

Visibility

Highlighting trends in health care utilization and costs: Journalists are increasingly using public data from HealthData.gov to highlight and communicate trends in health care costs to the lay public. Examples include documenting parallel trends with declining Medicare fee-for-service spending and increased popularity of Medicare Advantage,^v and showing that incorrectly estimating hospitals' costs caused Medicare to overpay hospitals by \$2.6 billion.^{vi}

Medical hot-spotting: Crain's New York Business used state hospital discharge data, accessible through HealthData.gov, to document substantial disparities in the median cost of procedures such as hip replacements being 570% higher at New York University Hospitals Centers compared to Bellevue Hospital Center despite being affiliated institutions. Similarly, markup rates for appendectomies ranged from 1% (Lincoln Medical and Mental Health Center) to 324% (Beth Israel Medical Center).^{vii} The easy accessibility of school district-level childhood obesity data in New York, also on HealthData.gov, highlighted geographic disparities and the wide media coverage influenced school wellness policies.^{viii}

Competition

Improved food safety and reduced foodborne illness: San Francisco's Department of Health rates the restaurants' food safety on a 0 to 100 scale. To make this more visible to consumers, thereby spurring restaurants to improve their safety, it partnered with data entrepreneurs to include ratings in Yelp, a national crowd-sourced restaurant rating website.^{ix} Past research has shown that making restaurant health inspection grades publicly available led to a 13% decrease in hospitalizations related to foodborne diseases in Los Angeles County^x and fewer serious restaurant violations in Salt Lake City.^{xi} Chicago has also used its food safety data to do predictive modeling on restaurants at highest risk for violations, thereby using its public health resources more efficiently.^{xii} San Francisco's new partnership, centered around publicly accessible data that are also available through HealthData.gov, represents a novel innovation in allowing the public to make more effective use of health data and spurn competition to improve public health outcomes such as food safety.

Improved cardiac surgery outcomes in New York: Since the late 1980s, New York has curated a registry of cardiac surgery and released data on differences in risk-adjusted mortality rates for percutaneous coronary interventions, cardiac valve surgery, and pediatric cardiac surgery across hospitals and surgeons. The public release of this information, which is currently accessible through HealthData.gov, has led to improved patient outcomes including lower mortality rates, low-volume and high-mortality surgeons no longer practicing cardiac surgery, insurance companies changing their contracted surgeons, and altered cardiac hospital markets.^{xiii}

Community learning

Community planning: The New York State Department of Health maintains an interactive “Prevention Agenda” dashboard for communities to view numerous health metrics by county and compared to the rest of the state, which is critical to local planning and needs assessments. Although its main data audience will likely continue to use the dashboard site, including these data in the HealthData.gov repository allows broader audiences to discover the Prevention Agenda site to collect relevant data and generate ideas for presenting data to their own community groups outside New York. 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 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.^{xiv}

Workforce training: Public data from open data platforms—which can be discovered through HealthData.gov—have already been used to improve classroom experiences to train the future workforce. New York University’s medical school uses New York State hospital discharge data to document the relative burden of disease at different facilities and regions. These are uploaded into students’ personalized medical training dashboards, and their medical training mentors compare students’ cases to the data-based benchmark to identify training gaps and adjust their future assignments to make them better prepared to serve their patient populations.^{xv} University at Albany recently modernized its Masters of Public Administration curriculum to require data management skills. Over six weeks, first-year masters students worked with open data from New York’s open data platform and HealthData.gov to learn about different types of data and data production cycles, how to evaluate data quality, how to manipulate data to summarize them meaningfully for managerial decision-making, and core principles of data management. As governments continue to collect data and information and communication technologies make it possible to integrate data analytics into real-time decision making, these data skills are needed to ensure that government agencies can deliver services more effectively and communicate with its citizenry.

Sharing ideas across sectors: A third example of community learning is through the exchange of ideas at Health Datapalooza and regional events inspired by this national conference such as the New York State Department of Health Innovation Challenge and California HHS Open DataFest.

Research

Pilot studies and enhancing the value of existing data: Although individual-level data will always be important for researchers, easily accessible open versions of these datasets can reduce the cost of research by allowing investigators to conduct inexpensive pilot studies. After attending a

workshop about New York State's new open health data platform, State University of New York researchers reported plans to use New York's public health data to conduct exploratory analyses and locate data that can be merged with other sources.^{xvi} For example, a researcher could easily explore data on childhood blood lead levels by zip code to assess whether they could be overlaid onto Zip code-level measures of the built environment, vital statistics information, educational outcomes, and neighborhood socioeconomic indicators. This preliminary work, which is more feasible through the use of data portals, can help researchers assess whether it is worthwhile to pursue a request for more granular data with individual-level information. These data are federated in the HealthData.gov repository, making them readily available to other researchers nationally.

New discoveries: New York's publicly available student obesity surveillance system, which assess the percentage of elementary and middle/high school students are overweight or obese, was merged with other publicly-available data from state and federal open data platforms to identify regional patterns in the relationship between the built environment and childhood obesity in New York. This led to a new discovery that these patterns between childhood obesity and the built environment differ by grade level and across counties.^{xvii} Although these data were ecological and could not be linked at an individual level, they were still able to identify new patterns that can be explored in follow-up research with more granular data.

Evidence-based decision making

Real-time decision making during disasters: In 2011, nursing home administrators in New York City to identify empty beds in nearby nursing homes where they could move their patients during the evacuation. These data made it possible because they were readily accessible and refreshed weekly.^{xviii} As an another example, researchers have demonstrated how Behavioral Risk Factor Surveillance System data can be combined with geographic information systems technology to improve planning public health responses for natural disasters such as coastal hurricanes in the southeast.^{xix}

Better information for health consumers: HealthData.gov includes data that can be used to rate doctors and hospitals, allowing patients to make more informed decisions. Although US News & World Report has long published annual reports on the best doctors and hospitals, a large portion of the evaluations are based on perceived reputation among experts. The publisher is now partnering with data entrepreneurs to incorporate quantitative health data from HealthData.gov to provide patients with more accurate information on local physicians providing "no- or low-value" health treatments that are unnecessary or inappropriate.^{xx} In addition to this public information, innovative startup companies are tailoring public health data, such as those available through HealthData.gov's repository, to specific health plan contexts and internal business rules to help plan members make better health choices.

Enhanced government decision-making: One benefit of open data portals is that they break down data silos within government operating divisions. These data silos limit staff access to data from other program areas.^{xxi} By including these diverse datasets in its repository, HealthData.gov can make it easier for HHS staff to see what other data are available and how to

procure data, thereby allowing them to make better use of data for internal decision-making. No similar inventory of HHS data is available within HHS for analysis by internal staff.

Recommendations for Maximizing the Value of HealthData.gov

HealthData.gov has already transitioned from its initial developmental phase to covering all HHS operating divisions and multiple state and local agencies. Actively outreaching to non-government data consumers through key stakeholder forums, such as the annual Health Datapalooza, has contributed to the development of a health data ecosystem. As HealthData.gov continues to evolve, there are several recommendations for how HealthData.gov can improve the accessibility and usability of high-value health data accessible to a wide range of customers to drive economic empowerment, innovation, and transformation in health. Underlying all recommendations is a continued commitment to good data stewardship, including adhering to best practices for data protection, security, and confidentiality.

Recommendation 1: Develop an integrated and coordinated strategy with HHS operating divisions, states, and local municipalities to advance the HealthData.gov vision and mission.

Open data platforms are a new technology to facilitate access to and efficient use of public data, and HealthData.gov was a groundbreaking undertaking in 2010. Despite developing and adapting organically to a rapidly evolving health data environment, it has already succeeded in achieving valuable outcomes related to its mission. The HealthData.gov aspiration to drive economic empowerment, innovation, and transformation in health can be attained at a higher level by focusing future activities around a focused mission. Specific strategies include:

- Promote HealthData.gov as a brand, reinforcing a consistent message about its vision and mission, and add a tag line such as “Public Data Unleashed” to materials to make the website recognizable and memorable.
- Facilitate a learning collaborative with state and local municipalities to share lessons, such as sustainable business models for public data platforms.
- Inspire data customers 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 facilitating user group meetings.
- Launch a public data accelerator to increase the number of innovators and data entrepreneurs and successful American businesses using public data.
- Support developer challenge competitions; sponsor events such as the Health Datapalooza; and provide seed funding opportunities to support solutions that make government more efficient and effective, cultivate the health data ecosystem, and foster a culture of innovation.
- Establish mechanisms to better integrate HealthData.gov into broader HHS data efforts.

Recommendation 2: Enhance the platform’s capabilities to make the data more meaningful to a range of data customers and extend its reach.

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 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. Specific strategies include:

- Keep the website in a modern design so that users view it as fresh and well-maintained platform.
- Implement additional web features previously recommended by the National Committee on Health and Vital Statistics, 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.^{xxii}
- Improve the automated processes currently in use to extract information from the open data portals that house the data covered in HealthData.gov's directory to ensure that HealthData.gov entries are timely and accurate.
- Show relations between datasets to better represent datasets in a family; for example, datasets with multiple years should be included in one place.
- Publish corresponding data models, to help users maximize the information.
- Make data documentation consistent across HHS datasets, consistent with international standards and following the National Committee on Health and Vital Statistics' past recommendations.^{xxiii}
- Develop a toolkit to enable communities to synthesize and integrate multiple datasets to address local health challenges consistent with the Office of the National Coordinator for Health IT's national interoperability roadmap,^{xxiv} thereby being a national resource and engine for community and national learning health systems.

Recommendation 3: Develop a long-term HealthData.gov infrastructure that allows it to tap into state-of-the-art knowledge about big data, machine learning, and algorithms for increasing customer use.

A strong infrastructure is required to ensure that HealthData.gov can be sustainable in the long-term and stay fresh so data customers perceive the site to be a valuable source of information. One component of infrastructure is ensuring that HHS has the technical capacity to use rapidly evolving state-of-the-art technology and business practices to make the data meaningful, timely, and usable. Another equally important component of infrastructure is ensuring organizational sustainability, which is required to sustain momentum, control the quality and accuracy of content, and continuously engage HHS staff and external users to promote customer use. Specific strategies include:

- Provide a foundational infrastructure to enable long-term planning and use of modern technologies to support the mission of HealthData.gov.

- Ensure that HHS has the technical capacity to use machine learning, data analytics, information and communication technologies, and other state-of-the-art scientific techniques to support users' needs.
- Establish a governance process to establish standards for the datasets that are referenced and other aspects of HealthData.gov.
- Develop business practices to allow operating divisions to keep the data comprehensive and timely, better understand and support user needs, and meet cyber security and privacy regulations.
- Create a Chief Data Officer who can oversee HealthData.gov activities and routinely engage staff within HHS operating divisions to solicit their input, provide feedback on how data have been used, and have a sustained commitment to releasing public data.
- Redefine HHS workforce roles to enable more active data customer support and foster health data community growth and a learning health system.
- Engage HHS in long-term data standardization activities to increase the ability to link datasets across operating divisions and decrease duplication.

Recommendation 4: Implement mechanisms to solicit data customers' input regularly to optimize their use of HealthData.gov and development of data-driven health innovations.

As health data reach new audiences with diverse skills and ideas, a better understanding of who the data customers are and how to release data to have a maximum impact is needed.

Traditional audiences for public health data include researchers and community health groups. Fostering health innovation from public data requires the continued engagement of new data consumers, who have different data requirements, intended uses, technical skills, and content expertise. For example, researchers frequently desire historical data over many years to document trends whereas an entrepreneur developing an application to help patients locate a medical provider may only require the most recent data. Specific strategies to better meet HHS data customer needs include:

- Create a "data product team" to engage regularly with diverse data customers from the private, public, academic, and other sectors to solicit their input on how to improve usability, access, and fitness for use.
- Solicit stakeholders' input systematically through surveys and focus groups.
- 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.

Recommendation 5: Implement data stewardship practices to keep the commitment to the public to protect health data security and confidentiality.

As data become more readily accessible and users combine data in novel ways, the risk of re-identification may increase due to the mosaic effect. As data are released, it may also become more challenging to identify recombinations or uses of the data in ways that increase the risks

of re-identification or stigmatization of small groups. Proper data stewardship is critical to avoid both over- and under-protection. Specific strategies include:

- Assign responsibility for data stewardship.
- Design processes for monitoring data recombination and the risk of mosaic and other similar effects that would permit identification or stigmatization of individuals or small groups.
- Design processes for monitoring downstream data transfers and use to avoid re-identification risks or violation of other principles of data stewardship.

ⁱ 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.

ⁱⁱ Martin E, Khalid H, Dadayan D. *Distribution of HIV prevention and surveillance funding to states, District of Columbia, and metropolitan areas, 2015*. Albany, NY: Nelson A. Rockefeller Institute of Government; 2016.

ⁱⁱⁱ Davis D. HealthData.gov reaches 3,000 datasets: a look back and what's ahead. <https://www.healthdata.gov/blog/healthdatagov-reaches-3000-datasets-look-back-and-what%E2%80%99s-ahead>. Published June 6, 2017. Accessed February 3, 2017.

^{iv} AcademyHealth. [Health Datapalooza conference website.] <http://www.academyhealth.org/healthdatapalooza>. Accessed February 3, 2017.

^v Frakt A. The two mysteries of Medicare. *New York Times*. October 3, 2016. <https://www.nytimes.com/2016/10/04/upshot/the-two-mysteries-of-medicare.html>. Accessed February 3, 2017.

^{vi} 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.

^{vii} *Crain's New York Business*: Health Pulse Extra. January 8, 2014.

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

^{ix} 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.

^x Simon PA, Leslie P, Run G, et al. Impact of restaurant Hygiene grade cards on foodborne-disease hospitalizations in Los Angeles County. *J Environ Health* 2005; 67(7): 32-6.

^{xi} Waters AB, VanDerslice J, Porucznik CA, et al. Impact of internet posting of restaurant inspection scores on critical violations. *J Environ Health* 2013; 75(10): 8-12.

^{xii} 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.

^{xiii} Hannon EL, Cozzens K, King SB 3rd, et al. The New York State cardiac registries: history, contributions, limitations, and lessons for future efforts to assess and publicly report healthcare outcomes. *J Am Coll Cardiol* 2012; 59(25): 2309-2316.

^{xiv} AcademyHealth. CHP program participant communities. <http://www.academyhealth.org/node/4901>. Accessed February 17, 2017.

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

^{xvi} Martin EG, Helbig N, Birkhead GS. Opening health data: what do researchers want? Early experiences with New York's open health data platform. *J Public Health Manag Pract*, 2015; 21(5): e1-7.

^{xvii} Dwicaksono A, Brissette I, Birkhead GS, et al. Evaluating the contribution of the built environment among New York State students. Poster presented at the Academy Health Annual Research Meeting, Boston, MA; June 26, 2017.

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

^{xix} Holt JB, Mokdad AH, Ford ES, et al. Use of BRFSS data and GIS technology for rapid public health response during natural disasters. *Prev Chronic Dis* 2008; 5(3). http://www.cdc.gov/pcd/issues/2008/jul_07_0159. Accessed February 16, 2017.

^{xx} 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.

^{xxi} Martin EG, Begany GM. Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. [Published online ahead of print August 7, 2016.] *J Am Med Inform Assoc* 2016. DOI: 10.1093/jamia/ocw076.

^{xxii} 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.

^{xxiii} 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.

^{xxiv} Office of the National Coordinator for Health Information Technology. Connecting health and health care for the nation. A shared nationwide interoperability roadmap. FINAL Version 1.0. <https://www.healthit.gov/policy-researchers-implementers/interoperability>. December 22, 2015. Accessed February 17, 2017.