Health Data Initiative
Strategy and Execution Plan

October 23, 2013
I. Strategic Overview

Data is a highly sought after resource that has demonstrated value to drive operational efficiency and provide cost reduction insights for multiple domestic and global industries - scientific research, finance, retail, and weather forecasting to name a few. This is no different for the health and human services ecosystem – in fact, given the massive scale of these industries, the importance of data as a catalyst of change is perhaps even more important. A recent report from the McKinsey Center for U.S. Health System Reform, "The Big Data Revolution in Health Care, Accelerating Value and Innovation"¹, points out that "new paradigms and pathways are emerging in a data enabled world" as companies "take advantage of data liquidity and embrace new value pathways enabled by big data". The report predicts that as traditional healthcare organizations evolve their thinking and maximize data liquidity, encouraging data to flow easily between systems and scaling the innovative tools and services that maximize the data flow can remove billions of dollars in costs from the nation’s healthcare system. In 2010 David Cutler indicated in his paper "Where Are The Health Care Entrepreneurs? The Failure Of Organizational Innovation In Health Care”² that the lack of access to data is one of the disincentives in the healthcare system that inhibits innovation. Secretary Kathleen Sebelius states that the vast stores of lazy data contained within the Department of Health and Human Services must be transformed into active data that informs the marketplace to contribute to improvements in health and health care.

These parallel positions about the importance of health data align closely with the Department of Health and Human Services (HHS) focus on realizing the full potential in HHS’ undervalued data and human capital assets. The now three-year-old Health Data Initiative (HDI), the collective effort to release vast stores of health data for innovation, is viewed across the federal government as a beacon in the advancement of data transparency and liquidity. The mission of the HDI is to help improve health, healthcare, and the delivery of human services by harnessing the power of data and fostering a culture of innovative uses of data in public and private sector institutions, communities, research groups and policy making arenas. The HDI’s goal is to make health data openly available, disseminate the data broadly across the health and human services ecosystem, and continuously educate internal and external participants in the ecosystem about the value of the data. The HDI’s efforts are demonstrating how health data may be used responsibly, privately and securely, while supporting healthcare innovations, and they provide a focal point for the movement that promotes a culture of data sharing and use. The value derived from sharing and exchange of data will propel problem solving toward creative, cost effective and efficient solutions in areas such as data driven decision-making in health care. The HDI’s vision encompasses a future where exchange and use of data will become commonplace, and involve all participants in the ecosystem, including patients, as vital contributors to individual, institutional, and population health problems.

The data products of the HDI have grown immensely from the initial efforts around health data liberation. This began in 2010 with the first 30 data sets listed centrally on Data.gov to the present day, with nearly 1000 data assets hosted on HealthData.gov, our website to help new data users discover resources they may not otherwise know exist. This site is a flexible platform that acts as a discovery resource for new and seasoned users across the healthcare ecosystem from researchers to tech/developers, and healthcare professionals to academia. Any organization or individual is free to employ the data to solve problems in the transformation of our nation’s healthcare system through data driven innovations.

To date, the primary focus of the HDI and the majority of the Department’s effort has gone into making a large volume of data available from various agencies in multiple formats for public consumption and innovation. As the concepts of data openness continue to become ingrained in the department, the HDI will continually adjust its execution plan to achieve the following goals five goals:

- Advancing the HealthData.gov site to a more efficient, user friendly, technically advanced platform for data discovery.
- Highlighting departmental assets that support achieving HHS strategic initiatives and an increased focus on strategic data liberation.
- Educating new and existing, internal and external participants in the health and human services ecosystem about data availability for innovative applications, and disseminating the data for problem solving.
- Enabling and incentivizing the health data ecosystem to utilize all data assets in innovative ways.
- Implementing administration and departmental policies that foster openness and transparency.

The tactical execution toward each of these goals and the resultant impact is integral to an overarching question:
What are the outcomes and impacts of the HDI and related activities on health care and the delivery of human services?

Identifying those outcomes and impacts, at least initially, requires that we consider the data from an organizational perspective, recognizing that its most important role is providing new understandings about complex issues in health and health care as well as new opportunities to create the means to solve them. By 2Q 2014, we will establish metrics to evaluate the impact of HDI by addressing questions such as:

- Is there a demonstrated increase in HHS programmatic focus on data to inform priority setting and goals, investment decisions, and the like?
- Have new opportunities been identified and/or innovative solutions to problems been discovered as a direct result of the uses of data?
- Are there established economic parameters of value creation from new applications, services or products established from using HHS data assets?

Ultimately it would be most valuable to have still deeper insights into the HDI’s impact on areas such as an individual’s access to health care, improvements in public health, greater efficiency and/or quality improvements in healthcare and human services or the data’s influence on public policy. The HDI will rely on the ecosystem’s participants to volunteer insights and provide feedback that will drive the direction of its work.

II. 2014-2016 HDI Strategy and Execution Plan

There are any number of tactics and initiatives which can be deployed to meet the goals described above. Given limited bandwidth and resources, we have prioritized our efforts in each area as detailed below. It is important to note that the HDI will attempt to measure the impact of any initiative relative to the goal it falls under; as a result, these initiatives are subject to change if the results are not measuring up to expectations or we can expect a greater return on investment in other areas.

A. Advancing the HealthData.gov site to a more efficient, user friendly, technically advanced platform for data discovery.

The HealthData.gov online platform is the central data access point and communications vehicle for the HDI
offering access to, dissemination of, and bi-directional communications about HHS and other sources health data. The goal for the platform is to be a highly useful, reliable, and well-supported platform for sharing datasets and fostering innovation. Success of the platform depends on broad-based enthusiasm for and commitment to the HDI, achieved through expanding partnerships, both internally and externally. Data is our primary asset and it is a strategic imperative to help our partners get all of the appropriate Strategically Relevant Data Assets (SRDA) published on the platform as quickly as possible. The platform is enabled by Application Programming Interfaces (APIs), programmatic functions that enable developers to easily embed and utilize data in their applications. Creating these automated connection points between the data and external computer systems will ease the data’s use and its transfer to outside systems creating true data liquidity to fuel algorithms, support data visualizations, or be used in other tools and services. HealthData.gov will continue to be at the forefront of HDI’s efforts to create market-based initiatives to use the data. The specific initiatives to enhance the usability of the platform include:

Focus Initiatives:

1. Determine and secure acceptable FY’14-16 funding commitment to support infrastructure upgrades and maintenance

   Metrics:
   - By 1Q 2014, secure commitment from departmental leadership for infrastructure funding.
   - By 2Q 2014, craft a two-year investment plan covering FY 2015 and FY 2016 for the healthdata.gov infrastructure.

2. Improve the metadata and the process to generate that metadata:

   a. Improve the metadata template for HHS colleagues to more efficiently and effectively catalog data assets.

      Metrics:
      - By 4Q 2013, solicit structured feedback from internal users about issues with the current metadata template.
      - By 2Q 2014, implement changes and re-survey users to validate improvements.

   b. Assist divisions in improving the metadata about data assets by providing guidance on principles of plain language writing for descriptions, updating the design of metadata templates for better usability, and updating older metadata entries for better clarity.

      Metrics:
      - By 4Q 2013, issue guidance to all Health Data Leads with timeframes and instructions for improving the metadata for their assets.
      - By 4Q 2013, recraft the descriptions and metadata for 10 existing assets to use as examples of plain language writing and better metadata.
      - Perform a random assessment of data assets to validate that by 2Q 2014 at least 25% adhere to these principles, 50% by 4Q 2014, and 100% by 4Q 2015.
      - By EOY 2014, increase the number of data assets updated to utilize common core metadata by 75%.

3. Accept feedback from users about how the platform should be altered to better suit their uses and/or address their needs.
4. Develop an area of the platform for non-HHS data (state/local data directory or catalog) for connectivity to other data libraries such as CMS navigator, Socrata, or others.

Metrics:
- Federate two (2) health-related data assets from at least two (2) federal agencies or open data initiatives (a total of four (4); e.g., Safety Data Initiative) outside of HHS 2Q 2014.
- Federate six (6) health data assets from the United Kingdom 4Q 2014.
- Federate health data assets from five (5) states by 2Q 2015.

Potential Future Initiatives

1. Developing, automating, and publicizing across HHS the platform’s ability to store/host data in CKAN from HHS agencies.

Metrics:
- Host 10 data assets in CKAN on behalf of HHS divisions.

2. Develop a nimble way for program managers to manage their own data - upload, management, and data quality improvement activities, lowering the barrier for making data available to the catalog.

Metrics:
- Solicit structured feedback from internal users about issues with the current data upload and management processes.
- Implement changes and re-survey users to validate improvements.

3. Improve data “sortability” and search results on the platform including results for costs of access, relevance, geographic specificity or other searchable factors.

Metrics:
- User feedback will reflect improved satisfaction with the user experience.
- Explore use of structured terminology standards for use as data keyword tags and in descriptive language about the data descriptions.

B. Highlighting departmental assets that support achieving HHS strategic initiatives and focusing on strategic data liberation

The catalog of data assets on HealthData.gov is continually growing, covering a wide array of health and human services topics. As the HDI attracts new and diverse audiences to the platform as a discovery zone of data resources it will be beneficial to focus on a subset of these data for directed communications efforts and promotion as they are strategically relevant data assets for Departmental goals. These SRDA will need to be characterized and defined, then publicized broadly through strategic communications to entice creative uses for these and other data on the platform. Conversely there are data assets that have not been emancipated
that are similarly strategically relevant. These data assets will be targeted for strategic data liberation efforts that focus resources on their responsible, expedient liberation after careful consideration of privacy and security.

**Focus Initiatives:**
Efforts to support strategic data liberation and highlighting departmental data assets will be:

1. Highlighting departmental assets that support achieving strategic goals:

   a. Establish a definition/criterion for SRDA, identify the datasets with input from HHS leadership, National Center for Vital and Health Statistics Work Group on Data Access and Use, the Health Data Consortium, and others and create an inventory of those defined datasets.

   Metrics:
   - By 1Q 2014, the HDI including the Health Data Leads will define the characteristics of strategically relevant data assets and identify a minimum of twenty data assets that meet the definition of SRDA.

   b. Allocate resources to create machine-readable formats for a subset of SRDA.

   Metrics:
   - During 2014, through the Health Data Leads, identify SRDA that are not available in machine-readable formats, and assist divisions transferring five (5) of them to machine-readable formats by EOY 2014.
   - Complete the conversion of twenty (20) SRDA data sets to machine-readable formats by Q3 2015.
   - By June 2015 make 20 new APIs available from non-SRDA data assets as part of a 15% overall growth in machine readable data sources available on HealthData.gov.

   c. Accumulate and highlight examples of where SRDA have been used for impactful or innovative solutions.

   Metrics:
   - By June 2014, identify and document at least three examples of innovative uses of SRDA for impactful or innovative solutions in healthcare or the delivery of human services.
   - By June 2014, identify and document at least five (5) examples of innovative uses of non-SRDA for impactful or innovative solutions in healthcare or the delivery of human services.

2. Create a communications and social media strategy to highlight the availability of the SRDA and promote their creative use.

   Metrics:
   - During 2014, as a measure of the efficacy of the SRDA communications and social media strategies, measure the increases in HealthData.gov website traffic to the associated SRDA.
   - During 2014, as a measure of the efficacy of the broader communications and social media strategy, measure common web statistics on the source websites from respective divisions where SRDA resides and changes in the number of downloads of those data.

3. Develop a pipeline of data assets, including SRDA, that are not publicly available but are strategic targets for focused data liberation efforts.
Metrics:
- Beginning in 1Q 2014, develop a pipeline of at least five (5) strategic data sets that will be the focus of targeted liberation efforts.
- By 4Q 2014, have a plan and tactical actions that will facilitate the liberation of at least three (3) of those assets by 2Q 2015.
- By 4Q 2014, have a plan and tactical actions that will facilitate the liberation of five (5) data assets that are not necessarily SRDA, and are not publicly available yet by 2Q 2015.

Potential Future Initiatives

1. Help divisions to recognize opportunities in data openness and transparency (e.g., transforming frequent Freedom of Information Act (FOIA) requests into open and transparent datasets that eliminate repetitive FOIA responses).

Metrics:
- Engage at least two of the department’s FOIA offices to develop a list of their top five requested data assets.
- Continue to work with those FOIA offices and their related divisions to draft an actionable plan for making at least two of those data assets publicly available in a usable form that meets the division’s and the department’s privacy and data quality standards.
- Create a ‘feedback’ loop that will inform divisions and their FOIA offices about the most requested data assets via HealthData.gov

C. Educating new and existing, internal and external participants in the health and human services ecosystem about data availability for innovative applications, and disseminating the data for problem solving.

Stakeholders from both inside and external to HHS are valuable contributors to the Health Data initiative. Internally, our HHS colleagues are the data liberators, however, it is important to engender understanding of the importance of openly available, easily accessible data for creative uses both for external innovations and internal operations. That understanding needs to penetrate beyond the front line data liaisons to program staff across the organization. Similarly external stakeholders need a broader and deeper understanding of why data assets were collected, the ways those assets are available to them, and how to interact with the data. Therefore a concerted effort to educate stakeholders about data availability, intended and possible uses, and examples of how data has been used by others are valuable educational components that will propel expanded appreciation for and uses of the data.

Focus Initiatives

Educating and informing internal and external stakeholders includes activities such as:

1. Spotlighting the value of openly available health data to health reform initiatives, health care quality improvement, cost containment, and delivery of human services, as well as its value back to HHS agencies or other public offices in order to paint the big picture of data’s importance in the development of tools and services that support these goals.

Metrics:
- During 2014 and beyond, collaborate with CMMI to highlight four (4) examples where qualified entities are utilizing the provision of data back to healthcare systems and infusing that data back into their operations for cost containment and quality improvement initiatives.
During 2014 and beyond highlight at least three (3) examples of how data has been utilized to stem the tide of waste, fraud, and abuse.

Collaborate with departmental human services agencies to highlight examples where data is utilized for program evaluation toward meeting milestones.

2. Increasing the percentage of data assets with machine-readable outputs for dramatically improved data dissemination (query tools will not count as machine readable).

Metrics:
- By 1Q 2015, increase the percentage of data assets on healthdata.gov available in machine-readable formats by 50%.
- Facilitate at least 25% of the department's recurring data outputs that are not available in machine-readable formats (.pdf files and graphs and reports) to machine-readable availability by 2Q 2015.

3. Continuing and expanding external outreach including conference attendance, speaking engagements, social media presence and online outlets for far reaching communications. External ecosystem influencers like the Health Data Consortium (see below) will be instrumental in this outreach.

Metrics:
- By 3Q 2014, health data leads will have posted at least three educational modules in the Codecademy catalog.
- By 4Q 2014, obtain structured feedback about the user ratings for each Codecademy module.
- By 4Q 2014, improve communications through the HealthData.gov blog with at least one (1) new blog post per week.
- By 4Q 2014, increase the @HealthDataGov twitter following to 15,000 and increase the number of retweets to sixty (60) per month.
- During 2014 representatives from the HDI will speak at or contribute materials to at least five (5) new conferences (audiences that we have never addressed before).
- By 2Q 2014, evaluate the utility of allowing HealthData.gov platform visitors to score datasets with usability ratings (required subject matter understanding, plain language description)
- By 2Q 2014, evaluate the utility of allowing HealthData.gov platform visitors to append linked media (articles, videos, etc.)

4. Develop use cases and an internal marketing approach to enhance workforce engagement and continue the culture shift toward an understanding of and appreciation for open data.

Metrics:
- By 4Q 2013, establish metrics for the health data leads program that will drive the productivity goals of the initiative.
- By 4Q 2013, solicit structured feedback from HHS employees about their awareness of the health data.gov platform, he and its availability as a resource to support their work.
- By 3Q 2014, implement changes and re-survey HHS employees to validate an internal communications campaign to improve awareness.

5. Insert the tenets of the HDI into IDEA Lab programs such as Ignites and Entrepreneurs to the extent that those programs are collecting or generating data make the data public.

Metrics:
- During the next round of each HHS innovation activities (HHS Entrepreneurs, HHS Ignite, HHS Innovates) ensure that at least one data derived output is the product of each activity
D. Enabling and incentivizing the health data ecosystem to utilize all data assets in innovative ways

The availability and analysis of data is one of the biggest drivers in the transformation of healthcare and there are tremendous opportunities for innovation in the health and human services ecosystems that will be fed by the vast stores of data made available via the HDI. An ongoing goal is to unleash the power of private-sector innovators and entrepreneurs to utilize HHS data in the creation of applications, products, and services that positively impact health and health care in the emergence of a decentralized, self-propelled "ecosystem" of innovators across America. That ecosystem includes organizations upon which the HDI will rely on for feedback, intelligence and insights that facilitate the democratization of health data and/or advocate for the innovative and responsible use of health data. Some specific groups are:

- The Health Data Consortium (HDC) is a non-profit collaboration among public and private entities working to foster the availability and use of health data to drive innovations. HDC will advocate across players in the healthcare continuum for data liberation and appropriate uses of data.
- Companies like Socrata, an example of a democratization facilitator, who is working with Federal, state, and local governments to provide access to open data.
- The HITECH Act empowered a wide array of health IT players (vendors, insurers, health care professionals, and delivery networks) to develop tools and services that support healthcare delivery including electronic health records (EHRs). On the patient side of healthcare, the Blue Button initiative has spawned a new generation of individuals who increasingly demand easy electronic access to their personal health information. Organizations serving these spaces can offer unique insights into the value and uses of health data.

Focus initiatives:

1. Continue to publicize the availability of the data and draw attention to uses of the data as one of the biggest drivers in the transformation of health, healthcare, and the delivery of human services. This includes highlighting the ample opportunities for innovation in the healthcare ecosystem that can be supported by the vast stores of data.

   Metrics:
   - By 2015, develop a series of at least 6 education and training events in partnership with non-government organizations focused on start-up and entrepreneurship education needs.

2. Seek new ways to engage entrepreneurs who may use the data as fuel for their business.
   - Develop and cultivate private sector and community relationships.
   - Provide insights about the data to the business community to highlight the myriad opportunities the data provide.
   - Generate market analysis for the department about how the data are being used, what additional data the market desires or requires, and accumulate recommendations that will help prioritize the HDI’s work.

   Metrics:
   - By 4Q 2014, the Department will identify resources to advance the goals of the HDI to the start up and development community and deliver measurable results in terms of community engagement and new applications.

3. Develop relationships and support the needs of federal and non-federal data projects, data enclaves,
repositories, and innovation test beds to meet the demand for more powerful analytic capabilities.

Metrics:
- By 3Q 2014, have an action plan with tactical objectives for how to utilize HHS data repositories currently under development.
- By 4Q 2014, have execution plans in place with at least two external data repositories to meet the needs of their analytics and knowledge development projects.

Potential Future Initiatives:
1. Establish new governance models for data sharing and promotion of innovative analytic methodology. This could include opportunities to explore data sharing practices, intellectual property arrangements, innovative privacy management practices, etc.
2. Forge partnership agreements with other application programming interface (API) development zones to capitalize on clinical data availability from EHRs and Health Information Exchanges (HIE).
3. Fund the development of applications and services through available vehicles.

E. Implementing administration and departmental policies that foster openness and transparency

There are several administration policies that support and impact the activities of the HDI and the internal stakeholders who produce and provide access to departmental data assets. Coordinating the current policy implementation agenda requires the department to understand the implications of each policy separately in order to harmonize the implementation of all policies for minimal disruption and maximum impact.

Focus Initiatives:

1. Draft an HDI Charter to institutionalize the Health Data Initiative as a vehicle for transparency, public access to health data, and feedback to the department’s senior leadership. The charter will clearly articulate a purpose, scope, and goals for the initiative.

Metrics:
- By 1Q 2014, draft, obtain feedback on, and complete departmental clearance for a charter to institutionalize the Health Data Initiative.

2. Implement the Open Data Policy—Managing Information as an Asset (May 9, 2013) which calls for the collection or creation of information in a way that supports downstream information processing and dissemination activities and must incorporate the following requirements: build information systems to support interoperability and information accessibility; strengthen data management and release practices; strengthen measures to ensure that privacy and confidentiality are fully protected and that data are properly secured; incorporate new interoperability and openness requirements into core agency processes.

Metrics:
- By Nov 1, 2013
  - Create and maintain an Enterprise Data Inventory and develop quarterly schedule to expand/update.
  - Create a Public Data Listing on HHS.gov/data according to the metadata schema.
  - Create a process to engage with customers to help facilitate and prioritize data release.
  - Document with the office of the General Counsel (OGC) if data cannot be released because it is deemed to be access level "private."
• By 4Q 2013 clarify roles and responsibilities for promoting efficient and effective data release.
• By 2Q 2014 disseminate Federal awards language that supports open data as a direct output of acquisitions, contracts, grants, etc.
• By 4Q 2014 incorporate new interoperability and openness requirements into core agency processes, such as machine-readable data formats for new datasets and uses of common core metadata.
• By 1Q 2015 Strengthen measures to ensure that privacy and confidentiality are fully protected and that data holders are properly implementing best practices from leading agencies with experience in privacy and confidentiality of public information including US Census Bureau.

3. Draft plans to address the impacts of the Holdren Memo on Increasing Access to the Results of Federally Funded Scientific Research (February 22, 2013) which requires Federal agencies with more than $100M in R and D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. Elements of the requirement include ensuring the public can read, download and analyze in digital form final peer-reviewed manuscripts or final published documents; facilitating easy public search, analysis and access to peer-reviewed pubs from fed-funded research; and ensuring full access to publications’ metadata. The HDI activities will support the strategy development and implementation plans drafted by AHRQ, NIH, CDC, ASPR.

Metrics:
• By 3Q 2013, submit a departmental strategy to OSTP describing the plan for compliance with the Holdren memo.

4. Continue to be a leading agency in the execution of the Executive Order on Making Open and Machine Readable the New Default for Government Information (May 9, 2013) which states that the default state of new and modernized Government information resources shall be open and machine readable; government information shall be managed as an asset throughout its lifecycle to promote interoperability and openness; and agencies shall ensure that they safeguard the individual privacy, confidentiality and national security.

a. Contribute to the formulation of the Cross Agency Performance (CAP) goal for open government data.

Metrics:
• By 2Q 2013, contribute feedback to OSTP and OMB representing HHS's strategy for meeting CAP goals.

b. Set incremental performance goals, ensuring they have metrics and milestones in place to monitor advancement toward the CAP Goal.

Metrics:
• By 3Q 2013, the HDI will have a strategic plan for meeting performance goals including metrics, milestones, and priorities for the initiative.

c. Distribute across the department procurement and acquisition language supporting the integration of open data requirements into Federal acquisition and grant-making processes (For civilian agencies, procedures for utilizing these clauses appear within FAR subpart 27.4 – Rights in Data and Copyrights...
(48 C.F.R. 27.400 et. seq.).

Metrics:
  ● By 1Q 2014, distribute to HHS agencies guidance and federal awards language for procurements and acquisitions supporting the integration of open data requirements in order to facilitate the management of data as an asset throughout its lifecycle.

d. Report on progress toward CAP goals including data inventory by Nov 9, 2013 (Quarterly updates thereafter).

  Metrics:
  ● By the end of 1Q 2014, the HDI will have a repeatable process in place for quarterly reporting on all CAP goals.

5. Clarifying relationship of the Open Data Policy, the Holdren Memo, and the Executive Order for the implications of each individually and as contrasted with each other:

  ● Re-purposing procurement language to be supportive of all policies
  ● Align metadata standards work across departmental systems and policy implementation efforts
  ● Infrastructure changes analysis
  ● Exploring the use of common repositories and registries to meet policy goals.
  ● Using of common core meta data across policy implementations and departmental systems
  ● Evaluating potential for Government Use and Minimal Public Use Licenses
  ● Obtaining stakeholder feedback on plans
  ● Developing strategies to ensure individual division’s compliance with agencies policies

Metrics:
  ● By 3Q 2014, the HDI will have a strategy in place for harmonizing requirements of each policy to maximize the efficiency of their individual and collective implementation.
  ● By 4Q 2014, the Department will have a long-term strategy for best practices derived from each strategy.

---

3 http://project-open-data.github.io/federal-awards-faq/