THE NCVHS FRAMEWORK PROJECT

Overview, Workshop Summary, and Framework v1

October 2014

INTRODUCTION

This status report describes the Framework Project, a new initiative of the National Committee on Vital and Health Statistics (NCVHS) that joins other targeted NCVHS efforts to enhance communities' data use capacities. The project focuses on data classification, use, and analysis. It is designed to generate recommendations for the Federal government and the entire ecosystem, along with an evolving set of classification resources to help communities and other data users to systematically, rigorously, and appropriately use data from all relevant sources to solve local problems. The Committee held a workshop in June 2014 at which participants toggled between a use case exercise and work on fleshing out a Data Continuum and Methods Categorization that ultimately will comprise the interlocking parts of a data Framework. Through the activities and plans described below, NCVHS seeks to lay the foundation for an iterative and expanding Framework development process in which it hopes many partners will participate.

Background

America's communities face a growing set of opportunities and pressures to use data effectively in their local health improvement efforts. In addition to the vigorous Federal data liberation initiative, for example, there are new forms of accountability for non-profit hospitals and public health departments, and incentives to share data for collective impact. A network of supportive organizations and websites offers a rich array of data and support. This confluence of forces gives communities ever-increasing prospects for leveraging diverse sources and types of data to better understand community health and its determinants.

Despite these influences, however, many communities lack the capacity to take advantage of the expanding resources. Most data users approach analysis from a single perspective and may not know how to draw on other data sources. They may be unaware of sources outside their own arena (health care, public health, education, the private sector, and so on); or they may be aware that other data exist but not know how to combine data from multiple sources. Perhaps they work with data at a single level of aggregation (individual, healthcare catchment area, county population) and don't know how to move among several levels, or how to look at data on upstream determinants such as economic resources or the built environment in conjunction with data on health outcomes, or how to choose the best data for evaluating the impact of interventions. The realities of non-interoperable data, data gaps, lag times, and uneven data quality, plus the shortage of local analysts, can add challenges to these already complex tasks. When data are combined across perspectives, domains, and sources, the complexities multiply. And all of these challenges are compounded by the absence of a common language and concepts that would enable effective communication about health data.

Helping Communities Become Learning Systems for Health

The optimal use of data for community health purposes requires a broad range of knowledge and technical skill. For example, it involves systematically locating relevant available data; applying standardization and data stewardship techniques for using multiple types, levels, and sources of data; identifying data gaps and designing strategies for filling them; and understanding the appropriate uses and limitations of the data. Broadly speaking, local actors need access to a range of data and the analytical tools to convert diverse data into useful information, as well as multi-dimensional partnerships and a supportive national infrastructure to turn to for technical and analytical support. The benefits of the current data explosion can only be realized if the systems for making sense of data keep pace with their burgeoning volume and complexity.

The National Committee on Vital and Health Statistics, which advises the Department of Health and Human Services on health information policy, is engaged in a multi-year effort to understand how communities can become learning systems for health, with several projects to help them address the needs outlined above.[1] One of these is the Framework Project.

The Framework Project

The NCVHS Framework Project focuses on data classification, use, and analysis. It grows out of the Committee's urgent sense that without appropriate systems and resources, even sophisticated communities could be overwhelmed by the pace and volume of data release and the complexities of using the data. The Framework Project is designed to generate recommendations and an evolving set of classification resources. Its goal is to help communities and other data users systematically, rigorously, and appropriately use data from all relevant sources to solve local problems.

The project is developing two complementary resources that, together, will compose the Framework:

1) The **Data Continuum**, a multi-dimensional structure for organizing information about populations at different levels or scales; and
2) The **Methods Categorization**, a family of methods to characterize data sets, secondary uses, techniques, and responsibilities.

Once it is developed, the Framework will relate the data elements or sets classified in the Data Continuum to the relevant content in the Methods Categorization, providing a way of

---

[1] NCVHS serves as the statutory (42U.S.C.242k[k]) public advisory body to the Secretary of Health and Human Services on health data and statistics. In that capacity, it provides advice and assistance to the Department and serves as a forum for interaction with interested private sector groups on key issues related to population health, standards, privacy and confidentiality, quality, and data access and use. Its 18 members have distinction in such fields as health statistics, electronic interchange of health care information, privacy 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. http://ncvhs.hhs.gov/

annotating data or datasets to clarify appropriate uses, limits, techniques, and responsibilities. The information can be used to document the biases of data, show how to use and repurpose the data, and provide a context in which to identify gaps. Together, these resources are intended to help community leaders determine what types of data are available to answer their questions and where to find them, as well as how to link and analyze them. An overview of the evolving Framework at its current stage of development is presented in Appendix 1.

The Framework Project is being carried out by the NCVHS Framework Work Group, coordinated by the Population Health Subcommittee. In addition to crafting classification resources, the project will develop recommendations to the Department and the broader ecosystem on how it can help achieve these goals. The project's deliverables include a white paper describing the Data Continuum and Methods Categorization, recommendations to the Secretary on addressing high-impact gaps in data sources and methods, and advice to the Department on online tools it could develop to support communities in these areas.

Other NCVHS subgroups are working on complementary products with which the Framework Project is closely aligned: a toolkit with guidelines and best practices for stewardship of community-level data (Subcommittee on Privacy, Confidentiality and Security); exploration of new data sources and dissemination modalities and ways to combine them with traditional Federal and state ones (Working Group on Health Data Access and Use); recommendations to the Secretary for enhancing the usability of HHS data (Working Group on Health Data Access and Use); and proposed content and transactional standards for population health data (Subcommittee on Standards).

THE FRAMEWORK WORKSHOP

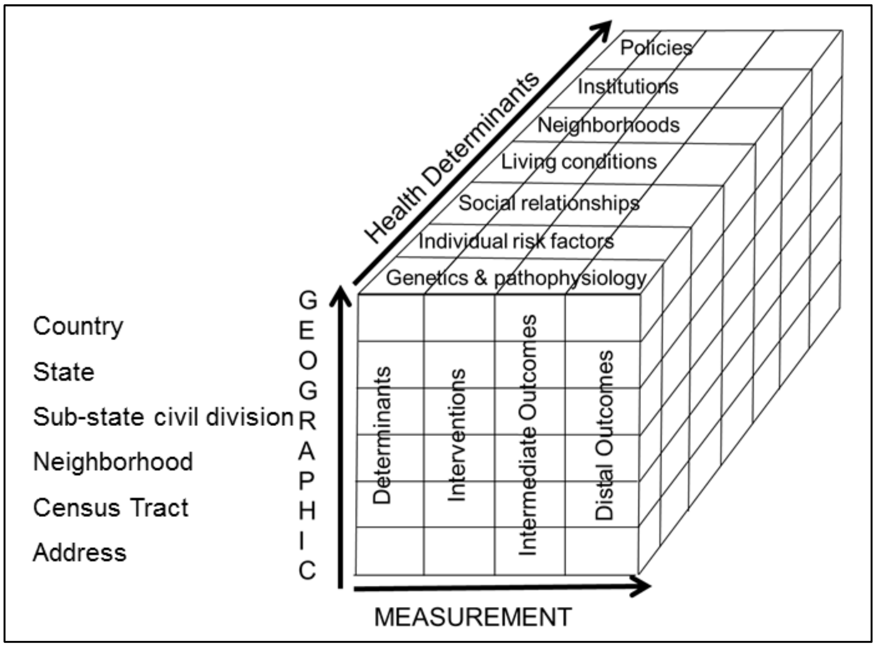
NCVHS took a large step forward in this project at a workshop held on June 13, 2014 in Washington, DC. The purpose of the workshop was to build a concrete example of the Data Continuum and Methods Categorization, which had only been described theoretically and in the abstract to that point. The meeting revolved around an extended exercise in which the diverse participants used Framework concepts and organizing principles to think through a real-world scenario from several perspectives. They used a familiar community health use case concerning childhood obesity reduction to explore data needs and craft an early application of the Framework. A special session was devoted to exploring the complex issues associated with data granularity and timeliness. This initial work will inform the further development of the Data Continuum and Methods Categorization.

Overview of the Data Continuum and Methods Categorization

The meeting began with an introduction to the conceptual structure of the Framework, with its layered and linked dimensions of data and methods. It described the Data Continuum using three dimensions as an example (Exhibit 1), with levels of geographic aggregation on one axis, measures on another, and health determinant variables on a third. This picture of the way data fit together is meant to help systematically identify gaps in data sources and point to the methods for developing strategies to fill the gaps, while applying relevant standards and stewardship principles.

The Data Continuum is a multidimensional picture of the data space. Exhibit 1 shows three of many dimensions, each of which builds from the most granular (individual) or proximal (determinant) to something that is more summative or distal. The dimensions represent different social-structural-biological variables, with upper levels of a dimension being more general than lower levels. Each dimension is a spectrum—that is, the boundaries of the cells are not fixed. Subdivisions may be added to clarify distinctions, or removed if a division is misleading.

Exhibit 1. Data Continuum



The Data Continuum describes the *data space* rather than the data that we have. This orderly depiction shows how particular data relate to other categories of data in the same dimension, while retaining their distinctive characteristics and/or zones.[2] Placing or linking the data we have into the relevant data space(s) would generate a multi-dimensional "map" showing both the data we have and the data we don't (yet) have, thereby highlighting data gaps.

The Methods Categorization shown in Exhibit 2 is an outline or taxonomy of dataset characteristics, uses, users, techniques, responsibilities, and standards.

Discussion Vignette: Outcome data

Participants expressed differing views about the best way to represent outcome data in the Data Continuum. While the figure in Exhibit 1 shows intermediate and distal outcomes, some argued that to be consistent with Donabedian's framework, outcomes should be represented in a single column. Others, however, countered that intermediate outcomes warrant their own column because that is the space in which government does much of its work. Further, communities need to look at intermediate outcomes to know if their interventions are having any effect on targeted aspects of community health. There was agreement that, however they are sliced in the model, outcomes are actually on a continuum.

---

[2] For example, a national program of screening newborns for genetic abnormalities would be located in the Country-Intervention-Genetic/Pathophysiology cell.

Exhibit 2. Methods Categorization

Methods Categories Include:

- Data source characteristics
- Secondary data use characteristics
- Framing questions
- Assembling and analyzing data from multiple sources and levels
- Data stewardship responsibilities
- Enhancing privacy
- Applying standards
- Dissemination

The Methods Categorization is extensible (that is, it takes future growth into consideration). Its categories could be used as metadata to tag datasets, analytic techniques, standards, and privacy-enhancing techniques, clarifying which ones work together and where they apply in the Data Continuum. (See the example below.)

Example: Tagging Protected Health Information

Under federal law and regulations, a health care provider may collect an individual's social and behavioral determinants, provided the data are to be used for a purpose related to the patient's health. In that case, identified data are *protected health information* (PHI). Federal law (HIPAA) allows the provider to disclose PHI to other health care providers and to a legally defined public health authority or for law enforcement, among other defined recipients. By tagging a data set with the category of original collector (health care provider), the purpose (individual's health), the identification status (identified), and also tagging the disclosure with the category of secondary user (public health authority) and use (public health), the combination of tags provides the metadata needed to systematically comply with the law and regulations.

The Use Case Exercise: A Community Childhood Obesity Reduction Project

The workshop participants worked on a use case based on the community health needs assessment and improvement programs of the Centers for Disease Control and Prevention. The scenario cast the workshop participants as coalition members in a community that has already gone through an assessment process; talked with community members; analyzed the data on health, disparities, and assets; and, on that basis, targeted childhood obesity reduction as its top priority. (For the purposes of the exercise, the group agreed to set aside the many issues associated with a focus on childhood obesity, including definitions, cultural factors, and stigma.)

The participants were asked to role-play community actors from four (of many possible) perspectives, each encompassing a constituency and set of actors, an institution or sector, an area of expertise and responsibility, and/or a set of information assets. The four perspectives used were: community members, schools, health care teams, and public health. The complexities associated with capturing a range of perspectives on community health was a major theme of the workshop, and the chief reason for engaging in this exercise. The participants adopted each of the perspectives in turn as they thought through each of the subjects outlined in the worksheet shown in Exhibit 3.

Exhibit 3. Use Case Worksheet

Goals				
Program Design Questions • Target population(s) • Intervention(s) • Process and outcome measures				
Available & Needed Data • Relevant data they have (and may supply to partners) • Additional obtainable primary data • Needed data outside usual sources				

After identifying goals, targets, interventions, and process and outcome measures relevant to reducing childhood obesity in their imaginary community, the group moved on to the key step of identifying where the community can find the information needed to answer their questions and guide decision-making. The worksheet divided this topic into three categories: (1) data that community agents *already possess* from existing internal and external sources; (2) other data that they are *able to obtain*; and (3) *needed* data that must be found from new and potentially unusual sources. Again, the group worked through this thought process from each of the four perspectives. It quickly became apparent that the variations in sources and data availability added greatly to the complexity of the task.

The results of the exercise were captured and documented in the Use Case worksheet presented in Appendix 2. As this was a rapid brainstorming exercise, these lists are illustrative rather than exhaustive.

An Expanding Frame of Reference

As noted, the purpose of fleshing out the use case table was to explore how the draft Framework might be revised to make it as valuable a resource as possible for community members in their uses of data to understand and improve community health. Thus, the brainstorming on the use case alternated with opportunities for participants to process emerging theoretical issues and choices and discuss their implications for the Framework.

Members introduced themes in the initial discussion that resonated throughout the day. A major one was that tools and analysis always must be contextualized by the needs, purposes, preferences, and capabilities of specific users. The group kept this in mind as it worked, and agreed to return to this fundamental point in future work.

Another critical aspect of the framing concerned the necessity of including in the analysis socioeconomic, environmental, and other contextual factors and secular trends outside the policy realm. This theme introduced additional data-relevant factors such as societal funding, societal wealth, cultural norms, insurance coverage, social/economic context, institutional racism, and the economic system.

While the participants accepted the discipline of adopting only some of many possible perspectives for the use case assignment, they also felt it important to acknowledge those that were missing. Given the importance of including information on all the determinants of health, there were suggestions about including the perspectives of law enforcement, local employers, restaurants and merchants, among others, as well as references to the natural and built environments and to social, political, and economic influences. The group agreed that the conceptual and data structures being developed should facilitate looking for data beyond governmental, public health, and health care sources and should include private, corporate, and non-profit sources and stewards.

Implications for the Framework

The participants applied the insights and learning they gained through the use case exercise to developing the Data Continuum and Methods Categorization. Some early observations led to these modifications:

- The group broadened the framing of the health care team to include non-traditional professionals such as roving community nurses, school nurses, and other community health workers; and
- Using schools and/or school districts as the point of reference, they agreed on the need to add an “organization” dimension to the Framework, inserting it between the geographic and population health measurement dimensions.

Discussion Vignette: Organization and Geography

The workshop participants brought their different perspectives to bear on the question of how to represent organizations in the data hierarchy—on the same level as geography and population health measures, or as a sub-level within geography. Schools and health care organizations served as examples. Those using a large integrated health plan as the paradigm favored embedding organizations within geography because that type of organization has a geographic dimension. However, others argued for the importance of being able to independently vary organization and geography. In further discussion, the Committee agreed to expand the concept of organization to include three types: single, aggregate, and virtual.

A participant clarified a key distinction between the level at which data are collected and the level at which they are made available by sketching a matrix with the data collector on one axis and the aggregation level on the other (Exhibit 4). It was noted that the collectors in these categories could be private or public (governmental), and that community groups and institutions themselves might contribute their information to a virtual database made accessible to others.

Exhibit 4. Levels of Data Collection and Aggregation, with examples in cells

	COLLECTOR AND LEVEL		
AGGREGATION LEVEL	**Sub-Community**	**Community**	**External**
Community	Data collected by individual schools & reported to the district	Police department's crime data	National survey
Sub-Community	A school's internal data	Crime data aggregated by neighborhood	Sociodemographic characteristics by census tract or block group

Zooming in on Timeliness and Granularity

The workshop participants took a deeper dive into two issues—timeliness and granularity—that surface frequently in NCVHS discussions of data, particularly as they affect communities. They divided into two groups and worked through some of the challenges in these areas and what they indicate about the background information needed on data and/or for data uses. After the breakouts, they reconvened to report on their findings. There was agreement that timeliness and granularity are only two of several characteristics pertaining to data usefulness, along with accuracy, sensitivity, and completeness, and that all of these filters interact with each other.

Timeliness

NCVHS addressed the issue of data timeliness most recently in a March 2014 letter to the Secretary that presented observations and recommendations developed by its Working Group on HHS Data Access and Use.[3] The timely availability of data is one of three aspects of data usability addressed in the letter. Working Group members have suggested that datasets should be tagged with metadata describing timeliness, and also that data may be "fit to use" for some purposes before they are adequate for others.

Following that logic, every dataset could be tagged with metadata describing its timeliness. Similarly, uses could be tagged with metadata describing the timeliness required for each use. The tags on available datasets could be matched to the tags on the proposed use to determine when the data set was ready for the use. The timeliness breakout group was asked to pick a dataset from the ones discussed earlier and brainstorm a taxonomy for metadata that describes its timeliness, with an eye to what would be most useful for data users to know about the data

---

[3] http://www.ncvhs.hhs.gov/140320lt.pdf

to judge its "fitness" for the intended purpose. It also was asked to think about a second set of metadata to describe how timely a data set needs to be for a given purpose such as clinical treatment or population-level measurement.

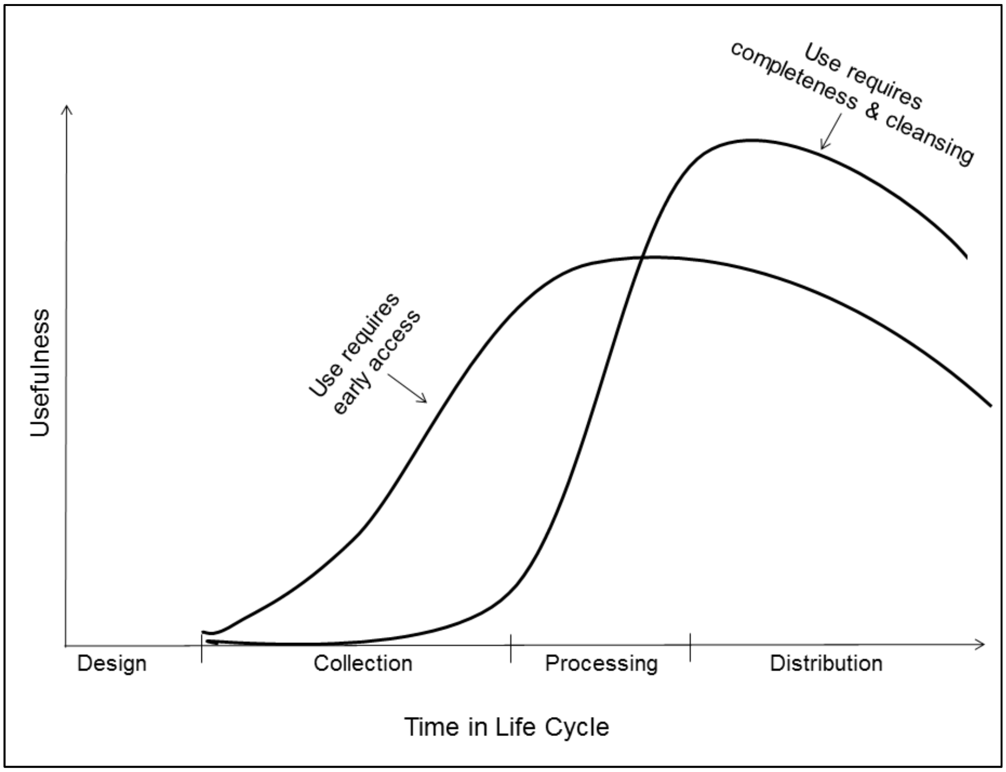
After the breakout sessions, the timeliness team reported that they had identified the following variables for characterizing data timeliness:

- Rate of change (how often the subject of the data needs to be measured)
- Shelf-life of the data (how long the data are good)
- Lag-time for validity (how long it takes for the data to become good)
- Acuity of need for the data (a major event starting a new cycle of data collection)
- Background rate of change (knowledge of secular trends that contextualize the significance of the data)

The group also pointed to the challenge of the Heisenberg effect, which says that when you measure something, it changes; so the timing of the release of data can have an effect on the rest of the cycle (and the need to collect data) by spurring action.

Team members also developed a "timeliness lifecycle" diagram (Exhibit 5) to differentiate the concepts of shelf-life and lag-time, showing how the usefulness of data increases and diminishes. The notion of "fit for use" points to the diminishing value (for some cases or purposes) of waiting for data completion.

Exhibit 5. Data Timeliness Lifecycle



The group briefly discussed differences such as longitudinal and time-slice data, noting that while this and many other data attributes (e.g., provisional, preliminary, open, closed) need to be understood, they were not considered during this focused exercise. However, they were flagged as important for the future work on the Framework.

Granularity

Community leaders raised the issue of data granularity at the seminal February 2011 NCVHS workshop on communities as learning systems for health.[4] While communities have become adept at working around gaps in available data, demand is growing for more detailed data on smaller populations and geographic areas. The centrality of neighborhood-level information for meaningfully addressing community health was a major theme of the 2013 NCVHS Roundtable on community health data needs.[5] And increasingly, community groups and agencies are collecting their own primary data or finding and creatively repurposing existing local data to augment secondary sources.

NCVHS has observed that "growing linkages and granularity can—and should—heighten privacy concerns" when there is a risk of identifying individuals or otherwise compromising privacy.[6] As noted above (page 3), the Committee is developing a toolkit to help communities practice data stewardship to protect the privacy of individuals and sub-groups. The Health Data Access and Use Working Group, too, has explored the need for more granular data, but deferred sending recommendations to the Secretary pending further exploration.

At the June 2014 workshop, the group working on this topic kept these perspectives in mind as it explored the data needed to tackle community health concerns such as childhood obesity. The participants were asked to choose two sample communities of differing sizes and densities (e.g., a rural county of 500 and an urban community of 600,000) and then think through a series of questions relevant to granularity, including estimation and stewardship. They also were asked to consider how techniques might change when applied to a small set of 100 individuals with a rare disease drawn from a large geographically dispersed population of 10 million. The goal of identifying relevant methods for moving among different levels of granularity is to make it possible to work with data at multiple levels of aggregation while taking into account relevant social constructs and constraints.

After the breakouts, the granularity team observed that the need for granular data is a function of the specific uses of the data, such as evaluation, research, or action. Also, the nature of the data source, methods, requirements, whether the focus is an individual or an institution, and social structures all have an influence on how granular the data are or need to be. Stewardship responsibilities also vary across these dimensions and others, with differences in sensitivity and tolerance.

This group looked at the approach to analyzing small area variations in the context of "who you are dealing with." The differences between dense, sparse, large, and small populations have an

---

[4] http://www.ncvhs.hhs.gov/111213chip.pdf (p. 23)

[5] http://www.ncvhs.hhs.gov/130430sm.pdf

[6] http://www.ncvhs.hhs.gov/111213chip.pdf (p. 28)

impact on the appropriate data collection infrastructure and consent mechanism as well as on the risk of harm. The team's spokesman used the following example to talk about some of these differences:

> The Amish and Mennonite communities in Lancaster County, PA, have a high incidence of a rare metabolic disease. People from all over the world come to see a physician who practices there.[7] The extensive data on these individuals (as on others with rare events and conditions) not only were collected by different means but require special handling and stewardship that are different from what is required for data on large populations. The appropriate methods are a function of the purpose of data use and other variables.

Some of the relevant variables are outlined in the table below (Exhibit 6).

Exhibit 6. Granularity Variables

Variables	Analysis of geographic communities of any size or density	Analysis of a sub-population with a rare condition
Data type	Passive collection Environmental Socioeconomic & cultural	High individual density of data Environmental
Geography	More relevant	Less relevant
Infrastructure	Common	Specialized
Analysis	Large: New methods, new data types Population intervention	Statistical methods for small groups "Classic" analysis Individual intervention
Stewardship	Population/political accountability	Higher risk of exposure, but maybe also need-based tolerance

Implications for the Methods Categorization

Once it is developed, the Framework will provide a systematic way of determining how to collect and protect individual data under different circumstances, depending on the purpose. It could serve as a filter to enable work at the appropriate level, given the balance between the analysis required and the sensitivity and risk associated with the data. Thus, it could provide a way to control and structure the process.

The group considered the implications for the Methods Categorization of their work on timeliness and granularity. For timeliness, further detail and categories were added to the Methods Categorization to address issues related to timing (1p), timeliness (1s), the aggregator's judgment of fitness (1t), and the timing required by type of analysis (4a). The resulting iteration of the Methods Categorization can be seen in Appendix 1.

Framework Project leaders observed that technology and the Framework will make it possible to represent the view of the data supplier and put filters on the data that are appropriate to the purpose. They declared the results of this focused effort on timeliness "a real victory."

---

[7] Holmes Morton, MD, at the Clinic for Special Children.

The participants then turned to the more complex subject of granularity as it applies to the Framework. This topic manifests both in the Data Continuum—which will be built out further as granularity expands—and in the blocks of methods related to data source characteristics (1f) and assembling and analyzing data from multiple sources and levels (4b). (See Methods Categorization in Appendix 1.)

Because of the privacy issues outlined above, the Methods Categorization also was expanded in relation to identification status (1j) and consent (1k); and the section on applicable regulations (1i) is also relevant. The idea here is to identify the characteristics needed to be able to analyze data at a given level or grouping (block, census tract, community, county, etc.). The distinction between analysis and reporting led to another change, reflected in section 8. In addition, members noted the implications for the level of confidence as it affects analysis. This was captured in methods item 1q, which relates to accuracy of the data source.

Discussion Vignette: Engaging the Community

A participant commented on the need to distinguish data collection from data analysis when considering granularity. Collection and analysis are linked by *purpose*, which determines the comfort level with various forms of analysis. Similarly, it is important to consider not only statistical significance but also meaningfulness. Questions such as these must be worked out not by analysts but in substantive conversations with community members. In a community decision-making process, data are filtered through values, judgments, and other priorities along with a sense of what people feel can be accomplished.

Due to a lack of time, discussion of methods associated with the purpose of data use (which appears in several contexts in the Methods Categorization) was deferred for another time. As this portion of the meeting drew to a close, members also flagged other topics for future consideration, including:

- Emerging methods of analysis as data source characteristics change (e.g., critical analysis of the use of spatial data analysis in GIS, and dealing with duplicate entries in big data); and
- Existing analytic methods that are not appropriate to specific levels of granularity.

One participant offered to look at possible applications to this project of findings from AHRQ-funded research on electronic data methods, in connection with the Patient-Centered Outcomes Research Institute (PCORI).

To conclude this session, after agreeing that much work lies ahead to develop the Data Continuum and Methods Categorization, the group paused to celebrate the success of this initial effort.

Key Learnings

A Fruitful Process

The workshop exercises demonstrated the breadth and depth of information needed to address community health issues and the complexity of moving among different perspectives and levels with respect to that information. By highlighting the need for rigorous classification and methods, these findings showed the potential usefulness of a tool that would help community data users systematically obtain, organize, protect, use, and reuse data.

The participants appreciated the benefits of exploring an abstract theoretical structure and set of concepts at such a concrete level. Working through a use case brought to light nuances that will influence the further development of the Framework. Based on this experience, the group encouraged those planning the Fall Roundtable (described below) to continue the concrete and detailed work. The experience also highlighted the benefits of bringing differing perspectives to bear on the many questions considered in the course of the day. It suggests that, when they are fully developed, the Framework resources now in an embryonic stage will help facilitate convergence among multiple sectors and endeavors and their information siloes—another major priority for the National Committee.

The Tower of Babel Problem

The day's activities and discussions shed light on what one participant called the Tower of Babel problem: that people in the health data world (as in others) sometimes mean quite different things by the same terms, without being aware of these differences.[8] Clearly, communication is not possible when people are unaware that they are speaking different languages. The workshop revealed that perspectives that may seem the same on the surface can turn out, on further examination, to be quite different. A major purpose of the evolving Framework is to create a "cross-walk" among vocabularies that will make it possible for everyone to understand one another.

A Path to Interoperability

It is widely recognized that the lack of interoperability is a major obstacle to the efficiencies and convergence that are critical for achieving the Triple Aims of health care quality, affordability, and population health.[9] Some of the critical types of data interoperability that can be enhanced by the Framework are shown in Exhibit 7 below. In the potential interplay between the Framework's Data Continuum and Methods Categorization, workshop participants could see the path toward enhancing interoperability.

---

[8] Petrie H, Do You See What I See? The Epistemology of Interdisciplinary Inquiry. Journal of Aesthetic Education, Vol. 10, No. 1 (Jan., 1976), pp. 29-43. University of Illinois Press. http://www.jstor.org/stable/3332007

[9] Interoperability is defined here as the ability of all of the actors who work to improve the health of individuals and populations (from the community to the international level), including patients and other lay people, and of different information systems and applications, to communicate, exchange data, and use the information that has been exchanged.

Exhibit 7. Types of Data Interoperability

- **Syntactic interoperability**: Linking industry-adopted standards formally recognized by a standard-making body to the data set being collected/exchanged (i.e., the message format and content).
- **Semantic interoperability**: Synchronizing definitions of concepts, terms, and variables (e.g., defining smoking or functional status).
- **Privacy interoperability**: Aligning health information privacy policies across health and information systems to allow the collection, use, and disclosure of information (e.g., matching primary data source restrictions to threshold for secondary use)(see the tagging example on page 6).
- **Security interoperability**: The use of comparable health information security policies and practices across systems to ensure consistent availability, confidentiality, and integrity of health information.
- **Granular interoperability**: Coordinating units of geography for which data are available (e.g., individual through national).
- **Time interoperability**: Currency of data and periodicity of data collection (e.g., real time data and how often collected).
- **Content domain interoperability**: Areas of focus (e.g., clinical indicators, risk behaviors, social/economic context, environmental factors, community assets).
- **Analytic interoperability**: Availability of tools for data manipulation (e.g., GIS, simple statistical software, WDQS).

NEXT STEPS FOR THE FRAMEWORK PROJECT: ROUNDTABLE ON COMMUNITY DATA ENGAGEMENT

The Framework development process will continue as a byproduct of the NCVHS Roundtable on Community Data Engagement in Washington, DC, on October 27-28, 2014. The Roundtable will bring together the perspectives of communities, data connector organizations, and data suppliers, and showcase solutions being developed in communities around the country. It will highlight where resources and support exist and where they are needed, with a major goal of shedding light on priorities for federal action.

The Roundtable will provide an opportunity to place the largely technical and theoretical questions explored during the June workshop into the broader context of community values and priorities, which determine the meaning and significance of data. This context makes it possible to convert data into information that communities can use to achieve their goals. A participant predicted that the workshop experience would help prepare the Committee to “listen from a more informed position to the people on the ground who are trying to get this done at the community level.”

The agenda, presenter list, and background materials for the upcoming Roundtable (including this report) will be posted on the NCVHS website, along with information on how to participate or listen to the proceedings.[10]

---

[10] http://ncvhs.hhs.gov/

Appendix 1. NCVHS Framework Overview – DRAFT v1, 9/30/14

The Framework organizes information about what types of data are available, where to find them, and what methods for accessing, analyzing, linking, etc. are appropriate for each source. It also provides a way to link relevant methods at each scale for different users and uses, document the biases of data and identify appropriate methods for repurposing data, and provides a context in which to identify gaps in both data and methods. The Framework includes two data classification resources, the *Data Continuum* and the *Methods Categorization*.

Data Continuum

1. **Geographic dimension** (geocoded from address)
    a. Individual
    b. Census tract
    c. Neighborhood
    d. Sub-state civil division
    e. State
    f. Country
2. **Organization dimension**
    a. Single
    b. Aggregate (roll-up)
    c. Virtual
3. **Population health measurement dimension** (IOM – for the public's health – measurement)
    a. Determinants
    b. Processes & interventions
    c. Intermediate outcomes
    d. Distal outcomes
        i. Disease specific scales
        ii. Health related quality of life
        iii. Summative (Health adjusted life years)
4. **Determinants of health**
    a. Pathophysiologic pathways
    b. Genetic & constitutional pathways
    c. Individual risk factors
    d. Social relationships
    e. Living conditions
    f. Neighborhood and community
    g. Institutions
    h. Social & economic policies

5. **Expression of pathology dimension**
    a. Acute illness
    b. Chronic disease
    c. Pre-disease
    d. Risk groups
6. **…**

Methods Categorization

1. **Data source characteristics** *[could be used as meta-tags]*
    a. Type of data source
        i. Electronic health records
            1. Care provider
            2. Health information exchange
                a. Regional
                b. Laboratory reporting
                c. E-prescribing
                d. ..
            3. …
        ii. Personal journal or health record
            1. …
        iii. Domain-specific measurement instrument
            1. …
        iv. Community data sets
            1. ….
        v. National surveys
            1. …
        vi. Payor datasets
            1. Medicare
            2. Medicaid
            3. …
        vii. Social network data sets
            1. …
        viii. Economic actor dataset
        ix. …
    b. Original collector and aggregator [*synonym – supplier, categories also applicable to secondary users*]
        i. Government
            1. Jurisdiction

            a. Federal
            b. State
            c. Sub state
        2. Type of authority
            a. Public health authority
            b. Non-public health agencies (e.g. social services)
            c. Law enforcement
            d. Environmental authority
    ii. Health plan
    iii. Healthcare provider
    iv. Individual member of the public
    v. Economic actors—corporate and private (e.g., businesses, credit scoring agencies, lenders, communications, grocery stores, devices)
    vi. …
c. Purpose of collection
    i. …
d. Method of collection
    i. …
e. Voice
    i. Self-report
    ii. Administrative staff
    iii. Trained observer
    iv. Passive collection (devices)
    v. …
f. Granularity
    i. Collection level [*link to DC geographic and organization dimensions*]
    ii. Aggregation level [*link to DC geographic and organization dimensions*]
    iii. Minimum # of individuals represented in a sample
g. Primary users
    i. …
h. Primary uses
    i. …
i. Applicable regulations
    i. Protected health information (HIPAA privacy rule)
    ii. Electronic identifiable health information (HIPAA Security rule)
    iii. Family educational rights and privacy act (FERPA)
    iv. State regulations
    v. Institutional review board (IRB)
    vi. …

- j. Identification status
    - i. Individually- identifiable data
    - ii. De-identified data (HIPAA definition)
    - iii. Anonymized data
        - 1. No linkage possible (alteration precluding linkage)
        - 2. Re-linkable data
        - 3. Linked with protected key (trusted third party)
- k. Consent provided at the time of data collection
    - i. No consent by the individual
    - ii. Consent by the individual
        - 1. Broad and unspecified
        - 2. Time-limited consent
        - 3. Consented for partial, source specific use (e.g., no psychiatric data)
        - 4. Consented for the particular type of use
- l. Applicable standards
    - i. Content
        - 1. …
    - ii. Messaging
        - 1. …
    - iii. …
- m. Demographic representation
    - i. Age
    - ii. Race
    - iii. Gender
    - iv. SES
    - v. Insurance status
    - vi. …
- n. Vulnerable populations included
    - i. Prisoners
    - ii. Pregnant women
    - iii. Undocumented immigrants
    - iv. …
- o. Population health measures included
    - i. [*Link to appropriate levels in DC population health dimension*]
    - ii. [*Link to appropriate levels in DC health determinants dimension*]
    - iii. [*Link to appropriate levels in DC expression of pathology dimension*]
- p. Timing

        i. Cross-sectional
        ii. Longitudinal
    q. Accuracy
        i. Level of confidence
        ii. …
    r. Completeness
        i. …
    s. Timeliness
        i. Rate of change
        ii. Shelf life
        iii. Acuity of need
        iv. Lag time
        v. Background rate of change
    t. Aggregator’s judgment of fitness
        i. Provisional vs preliminary
        ii. Open vs. closed
    u. Biases
        i. …

2. **Secondary data use**
    a. Users
        i. *[re-use categories under original collector or aggregator 1.b]*
    b. Uses
        i. Healthcare
        ii. Public health
        iii. Social services
            1. Abuse neglect or domestic violence
                a. Child abuse or neglect
            2. Workplace safety…
        iv. Law enforcement
        v. …
    c.
3. **Framing questions**
    a. Worksheet for framing the problem (Davenport)
    b. …
4. **Assembling and analyzing data from multiple sources and levels**
    a. Timing required by type of analysis
    b. Level of granularity (affecting ability to analyze at different levels)
        i. Infrastructure
        ii. Consent

            iii. Risk of harm
            iv. …
        c. Purpose-related factors [of data use]
        d. Proprietary ownership issues
        e. Relevance to analytic method[s]

5. **Data stewardship responsibilities**
    a.
6. **Enhancing privacy**
    a. …
7. **Applying standards**
    a. …
8. **Reporting and dissemination**
    a. …
9. **…**

Appendix 2. Use Case Worksheet: Community Plan—Childhood Obesity

Perspectives ➔	**Community**	**Schools**	**Health Care Teams**	**Public Health**
Goals	▪ Community culture of wellness ▪ Shift bmi distribution in targeted age groups in 5 years, e.g. Fewer overweight kids entering kg ▪ Increase activity level in families with young children in 5 years ▪ Change in attitudes toward diet & exercise in kids entering hs ▪ Healthier kids	▪ Kids that are “fit to learn” ▪ Shift BMI distribution of kids in the district ▪ Increase activity of kids in district ▪ Increase in healthy lunches	▪ Reduce incidence of obesity related co-morbidities for patients in their care ▪ Stable bmi appropriate to body frame ▪ Trusted point of access to health care for all members of families in their care ▪ Increase awareness of clinical team about community resources	▪ Decrease disparities in nutrition, activity & obesity ▪ Decrease morbidity & mortality ▪ Increase community awareness of obesity risk & trends ▪ Increase community awareness of barriers to proper nutrition & activity
Program Design Questions				
Target population(s)?	▪ Pre-KG ▪ Schools ▪ Day care ▪ Churches ▪ Malls	▪ KG ▪ Lower elementary ▪ Upper elementary ▪ HS	▪ Perinatal families ▪ Pediatric age groups ▪ Adolescent patients	▪ State ▪ County ▪ Neighborhoods
Intervention(s)?	▪ ID most effective “upstream” interventions ▪ Public awareness campaigns - wellness is “hip”, targeted at young parents and kids; obesity risk ▪ Ban ads for unhealthy foods targeting youth ▪ Programs for key life transitions, birth, entry into kg, elementary, health care ▪ Day care nutrition guidelines ▪ Fresh food markets near school ▪ Healthy supper clubs in churches, community centers, grocery stores ▪ Clean up, light & monitor parks & playgrounds ▪ Access to safe recreation areas for kids & families ▪ 1k steps/day campaign with pedometers ▪ Provide web-based referral to community wellness resources for public, school nurses, HCPs	▪ ID most effective school based interventions ▪ Affordability in healthy lunches ▪ Healthy snack machines ▪ Student/parent healthy supper classes ▪ Physical education interventions and activities ▪ Increase activity in extracurricular activities ▪ School based wellness coordinators & nurses	▪ ID most effective practice-based interventions ▪ Training & information for clinicians, technical resources & incentives ▪ Screen for diet and activity ▪ Include nutritional & activity coaching in assessment of developmental milestones ▪ Referrals to community resources for wellness & life change ▪ Know who adolescent patients are; demonstrate sustained long relationship with them; review their “journals” of relevant data	▪ ID most effective public health interventions ▪ Disseminate data on prevalence & risks of C.O. ▪ Review literature on what works & provide good information for all coalition members ▪ Promote awareness & convene stakeholders to share perspectives ▪ Improve parent awareness of obesity risks ▪ Establish trust as data steward for community, collect missing data & convene discussion of meaning of data

Appendix 2. Use Case Worksheet: Community Plan—Childhood Obesity

Perspectives →	**Community**	**Schools**	**Health Care Teams**	**Public Health**
Process and outcome measures?	▪ O: Weight, BMI by age cohort & neighborhood ▪ P: media appearances; O: awareness of risks, attitudes ▪ P: minsters pitched to; O: churches adopting interventions ▪ P: # gardens planted ▪ O: time parks and playgrounds available, # kids participating ▪ O: pre-post survey family perceptions of change	▪ P: % of schools in district participating ▪ P: % school measurement of BMI at start of yr, O:% elevated BMI entering next grade ▪ P: # students participating by type of intervention ▪ O:% healthy lunches ▪ O:Minutes of in-school activity trends by school & age cohort ▪ O:BMI trends for studies by school & age cohort	▪ P: % screened, % coached, % referred to community resources ▪ O: Distribution of # days w active exercise & minutes/day in patients under care ▪ O:Distribution of fruit and vegetable consumption in patients under care ▪ O:Distribution of BMI trends in families under care ▪ O: Diabetes-2 prevalence in families under care for 2 yrs, 5 yrs, 10 yrs	▪ P: % of effective interventions implemented ▪ O: Weight, BMI by age cohort & neighborhood
Available and Needed Data				
Relevant data they have (& may supply to partners)	▪ Bike trails, walking, other rec opportunities & spaces, safe & clean	▪ Location, resources, staff, programs ▪ Catchment area served	▪ BMI, ht, wt, BP for patients under care ▪ Payer	▪ Incidence and prevalence of obesity by census tract ▪ # & types of providers avail ▪ # households ▪ WIC kids' BMI ▪ SES data on community, demographics, housing stock, pop density, public safety ▪ Un/employment, Medicaid, free lunch, etc.
Additional primary data they can obtain	▪ Snack food revenues ▪ Neighborhood assessment of kid activity, qualitative data, # gunshots, space avail.	▪ BMI ▪ Belly circumference ▪ Hours of exercise/day ▪ Calories of school meals ▪ # Vending machines SES factors	▪ Complications of C.O. ▪ Activity & nutrition screening ▪ More thorough information on family, incl history of diab, culture/attitudes	▪ Survey available space for recreation & current activity there ▪ % w/in 1 mi of walking ▪ School, healthy food
Needed data outside usual sources	▪ Retail ▪ Church health fairs ▪ Salons ▪ Malls ▪ Barber shops	▪ School-based BMI measures ▪ Children at risk for obesity	▪ Comparison data on children in their practice vs others; ▪ Trend data on CO	▪ % households with trusted places of entry ▪ SES predictors; ▪ Accurate information on comm resources to deal with CO; ▪ Social media sources, every kid ▪ Need for detailed, comparative data across neighborhoods to show disparities

Appendix 3. Framework Workshop Attendees

NCVHS Members and Data Working Group Members

Larry A. Green, M.D., Chair

Bruce B. Cohen, Ph.D. (Co-Chair, Subcommittee on Population Health)

William W. Stead, M.D. (Co-Chair, Subcommittee on Population Health)

John J. Burke, M.B.A, MSPharm.

Leslie Pickering Francis, J.D., Ph.D.

Alix Goss

Linda L. Kloss, M.A.

Walter G. Suarez, M.D., M.P.H.

Leah Vaughan, M.D.

James M. Walker, M.D., FACP

NCVHS Staff

Debbie Jackson, M.A.

Tammara Jean-Paul, Ph.D.

Others

Oscar Morgan, Facilitator

P. Jonathan White, M.D., AHRQ

Michael Fitzmaurice, Ph.D., AHRQ

Vickie Boothe, CDC

Susan Baird Kanaan, Consultant Writer

Appendix 4. Population Health Subcommittee

MEMBERS

Bruce B. Cohen, Ph.D., Co-Chair
William W. Stead, M.D., Co-Chair
John J. Burke, M.B.A., M.S.Pharm
Llewellyn Cornelius, Ph.D.
Leslie P. Francis, J.D., Ph.D.
Vickie M. Mays, Ph.D., M.S.P.H
Sallie Milam, J.D., CIPP, CIPP/G
Len M. Nichols, Ph.D.
Walter G. Suarez, M.D., M.P.H.

STAFF

Vickie Boothe, CDC
Virginia Cain, Ph.D.
Tammara Jean Paul, Ph.D., NCHS
Jacqueleine Lucas, NCHS
Susan Queen, Ph.D., ASPE

Susan Baird Kanaan, Consultant Writer