clinovations
GOVERNMENT + HEALTH

NATIONAL VITAL STATISTICS SYSTEM MODERNIZATION – NEW OPPORTUNITIES FOR INTEROPERABLE DATA

SEPTEMBER 27, 2024 – FINAL REPORT

Prepared By: Clinovations Government + Health

Contributing Team: Cajay Jacobs, Nicole Kemper, Tatyana Roberts, Anita Samarth, Katie Schultz, Justin Wang

Prepared For: National Science Foundation (NSF) National Center for Science and Engineering Statistics (NCSES) in collaboration with the Centers for Disease Control (CDC), National Center for Health Statistics (NCHS)

Contract Number: ADC-NVSS-23-N02

clinovations
GOVERNMENT + HEALTH

CONTENTS

The America's DataHub Consortium (ADC), a public-private partnership is being utilized to implement research opportunities that support the strategic objectives of the National Center for Science and Engineering Statistics (NCSES) within the U.S. National Science Foundation (NSF). This report documents research funded through the ADC and is being shared to inform interested parties of ongoing activities and to encourage further discussion. Any opinions, findings, conclusions, or recommendations expressed in this report do not necessarily reflect the views of NCSES or NSF. Please send questions to ncsesweb@nsf.gov. The OMB control number for this collection is 3145-0215. This product has been reviewed for unauthorized disclosure of confidential information under NCSES-DRN24-065.

clinovations
GOVERNMENT + HEALTH

EXECUTIVE SUMMARY

Background and Purpose

This project is part of the National Secure Data Service (NSDS) Demonstration (NSDS-D) project) sponsored by the National Center for Science and Engineering Statistics (NCSES) and the National Center for Health Statistics (NCHS) within the Centers for Disease Control and Prevention (CDC). As NCSES considers a potential, future NSDS, demonstration projects that study successful models that transfer state and local data to the federal government can offer insights and lessons learned in determining shared services to support data for evidence-building.

Clinovations Government + Health (Clinovations) was awarded this project to perform an environmental scan of the National Vital Statistics System (NVSS) to identify shared services to consider within a potential, future NSDS. NVSS is often cited as an exemplary model of intergovernmental data sharing because it effectively involves collaboration between federal, state, and local government entities.

NVSS has undergone significant modernization efforts to enhance the timeliness, accuracy, and accessibility of vital statistics data. Over the past few decades, NVSS modernization initiatives have focused on implementing electronic vital records systems, standardizing data exchange formats, and leveraging technology advancements to enable real-time data reporting and interoperability between state and federal systems. NVSS has achieved significant progress in use of high quality and timely interoperable data for public health, policy-making, and research through continuous evolution of infrastructure and adoption new technologies. The NVSS ecosystem could serve as a potential interoperable data and evidence model for a NSDS because of its experience implementing intergovernmental data sharing amongst a multitude of stakeholders, implementation of governance considerations and authorized roles and responsibilities, and tiered data access structure.

Approach and Methods

To collect input for the environmental scan, Clinovations conducted a literature scan, a NVSS modernization scan, interviews with federal stakeholders, and a series of virtual non-federal stakeholder workshops. The literature scan was used to understand common uses, data quality considerations, standards and interoperability needs, and identify key stakeholders for stakeholder workshops. For the NVSS modernization scan, Clinovations reviewed publicly available information describing NVSS modernization projects, interoperability approaches, and infrastructure. Interviews with NVSS leadership and staff with responsibilities for NVSS data collection from jurisdictions and processing for statistical dataset generation and federal users of NVSS data offered a foundation of information about federal roles, responsibilities, and needs. Four stakeholder workshops with data entry professionals, data custodians, data users, and analytics/innovator vendors provided diverse perspectives for a landscape analysis of interoperability considerations for NSDS consideration.

To develop considerations for a potential NSDS, Clinovations aggregated findings across environmental scan activities, compiled themes, identified and classified themes that could be applied beyond NVSS and health data, developed considerations informed by NVSS, and identified NVSS examples that can be applied to a potential future NSDS for each consideration.

Considerations for a Potential, Future NSDS

By drawing insights from the NVSS environmental scan, a potential NSDS can leverage process and infrastructure key success factors and interoperability enablers to meet its purpose and core functions of providing federal, state, and local partners the services they will need to securely use data to support evidence-building in policymaking and research. The NSDS will act as a shared service to provide support through resources for communication, data standardization, coordination, and research and development. Based upon the environmental scan of NVSS, a potential, future NSDS could support data interoperability and modernization of intergovernmental data sharing for evidence-building by developing and offering shared services that support and enable:

- **Navigating and Comparing Statistical Data Sources.** Online resources (e.g., GitHub repository, local/state laws and regulations comparisons and analyses, tools and links to downable datasets) and data navigators or data concierges to guide data requesters to potential data sources.
- **Cataloging and Harmonizing Value Sets for Common Data.** Coordination and navigation across available value sets, such as: publishing or collating value sets, data concierge to evaluate values across different value sets, or resources to harmonize and guide selection of standards-based value sets.
- **Developing Data Entry Stakeholder Training and Education Resources.** Guidance and resources for cohorts of communities that enter data into information systems that are transformed or mapped to coded values for development of statistical datasets or used in evidence-building.
- **Advancing Skills Through Modern Interoperability Infrastructure.** Implementation guidance, test environments, sandboxes, pilots, and real-world and synthetic test data for upskilling of state, local, and federal staff by providing tools that build interoperability knowledge and capabilities.
- **Streamlining Common Data Access Tiers.** Best practices for streamlined and simplified data access tiers that apply to statistical datasets that use data generated at state and local levels. Data concierge and informational services to support navigating state/local regulations in developing access tiers.
- **Expanding Standard Application Process (SAP) Awareness.** Increase SAP awareness and coordinate across states and jurisdictions to foster a more standardized application process modeled on the and consider a metadata catalog of data that are available for discovery within or across states.
- **Modernizing Platforms for Data Access and Analyses.** Navigate limitations of data assets related to conditions of use and feasibility of using modern computing technologies such as cloud-based tools, secure computing, APIs, ethical hacking sandboxes, and privacy preserving technologies.
- **Participating in Standards Development and Advancement.** Actively engage with standards development organization and industry efforts to align and foster the development of standards that supports needs at the point of capture (data entry) and statistical dataset standardization.
- **Providing Templates for Common Queries and Computational Analyses.** Templates for queries and computational analyses tools for data or metrics that are commonly requested. Develop a "fast statistics" dashboard with downloadable set of data elements and metrics and visualization tools.
- **Communicating Best Practices for Rapid Release of Data.** Develop best practices, guidance, and strategies for rapid access to provisional data and mobilizing and connecting with research and industry data users who need access to rapid response data.

NVSS Key Takeaways for a Potential NSDS

An environmental scan of the NVSS revealed national statistical dataset programs that have achieved significant modernization progress, such as NVSS, can offer insights and models for services, processes, and programs for future potential future NSDS consideration.

Establish and lead communities of practice. As a governing body, NCHS used its influence and reputation to generate collaboration and coordination with multiple stakeholders. A potential future NSDS could consider how to use learning communities to strategically support other core functions.

Provide stakeholder specific guidance and communication. Through continuous interactions with stakeholders, NCHS has developed tailored communication materials with partners. potential future NSDS could consider proactively supporting stakeholders to identify pain points and define next steps.

Update technology and operations to support urgent needs and demands. To meet evolving public health data demands, NCHS committed to continuous improvements to statistical datasets within NVSS. A potential future NSDS could consider support for more agile systems and processes for timely response.

Use modern data capabilities and advanced data analytics. NVSS incorporated natural language processing, artificial intelligence, application programming interfaces, and open-source data standards. A potential future NSDS could investigate and test new tools and technologies that improve interoperable data use.

1 INTRODUCTION

1.1 PROJECT GOALS AND OBJECTIVES

This project is part of the National Secure Data Service (NSDS) Demonstration (NSDS-D) project sponsored by the National Center for Science and Engineering Statistics (NCSES) within the United States (U.S.) National Science Foundation (NSF) and the National Center for Health Statistics (NCHS) within the Centers for Disease Control and Prevention (CDC). America's DataHub Consortium, a public-private partnership, is being utilized to implement this project. Clinovations Government + Health (Clinovations) was awarded the National Vital Statistics System Modernization - New Opportunities for Interoperable Data (NVSS-Interoperability Opportunities) project, which aims to identify data interoperability accomplishments and opportunities for the National Vital Statistics System (NVSS) ecosystem that could also be applied more broadly to support data ecosystems beyond NVSS to use for evidence-building and decision-making. The NVSS-Interoperability Opportunities project is just one of many NSDS-D projects, mandated by the 2022 Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science Act, to inform a federal effort on strengthening data linkage and access infrastructure.

Project Objective: Inform planning for a potential, future NSDS via a scan of NVSS

This project seeks to identify untapped opportunities for leveraging NVSS approaches for data interoperability, which can be used to inform a potential, future NSDS. The results of this project should enhance collaboration by informing data interoperability standards and modernization opportunities for data and evidence ecosystems. The NVSS ecosystem may offer a strong model to inform a potential NSDS because of its long-standing value and use in public health and research, well-standardized datasets and data exchange partners, and progress in adopting modern standards for interoperable health data.[i,ii]

An environmental scan of NVSS that reviews the significant progress to date and explores current gaps, data access needs, privacy risks, and infrastructure considerations can inform future NVSS planning and offer considerations for a future potential NSDS.

[i] About the National Vital Statistics System: https://www.cdc.gov/nchs/nvss/about_nvss.htm

[ii] Hetzel AM. History and organization of the vital statistics system. Hyattsville, MD: National Center for Health Statistics. 1997;14.

1.2 NATIONAL VITAL STATISTICS SYSTEM BACKGROUND

The NVSS is the oldest, and represents a successful example of, intergovernmental data sharing in public health.[i] It is the data and evidence ecosystem maintained by NCHS that collects and disseminates birth and death data, using the U.S. Standard Certificate of Birth (2003 Revision) and the U.S. Standard Certificate of Death (2003 Revision) as a model for individual state birth and death certificates. The NVSS comprises many systems that support the data exchange of roughly 6.5 million records a year between the local, state, and federal levels. Over the last decade, NCHS has been collaborating with its jurisdictional partners (i.e., 50 states, Washington, District of Columbia, New York City, and five territories (Puerto Rico, the Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands) to modernize by improving the collection, processing, and use of vital statistics data.[iii]

The NVSS highlights four areas that serve as potential examples of modern approaches for data exchange: 1) interoperable data (i.e., data that are exchangeable via electronic services between two IT systems through the use of standards), 2) required access levels for the use of interoperable and statistical data, 3) associated considerations related to privacy and confidentiality, and 4) infrastructure and metadata areas for future exploration. Exploring these areas, outlined in Figure 1, can help inform future exploration for uses of interoperable data and shared services. The results of the project will inform planning for both the NVSS and a potential future NSDS and offer recommendations for advancing interoperability and data exchange to expand data available for evidence-building. Furthermore, insights from recent NVSS efforts to modernize data interoperability may inform best practices for other statistical data systems and state/local data submissions to federal statistical agencies.

Figure 1. NVSS Areas to Explore in Environmental Scan

Uses & Gaps	• Methods for timely research and surveillance • Criteria for use and expansion of interoperable data • Impact of variable needs across Federal, state, local levels
Data Access	• Levels of data access needed by data users • Mechanisms for data access
Privacy Risks	• Approaches to address to privacy and confidentiality • Limitations and considerations to mitigate risk
Infrastructure & Metadata	• Modernization of technology and process infrastructure • Use of metadata to support interoperability • Best practices to establish consistent standards

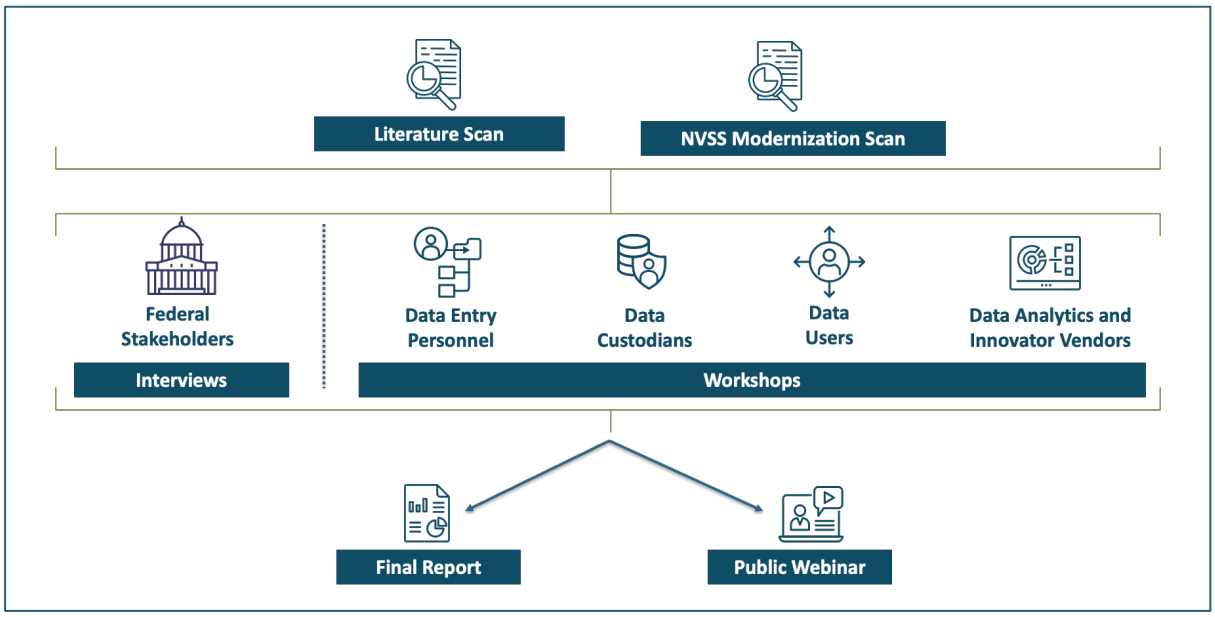
The NVSS ecosystem serves as a potential data and evidence model because of four characteristics: a multitude of stakeholders, governance considerations, authorized roles and responsibilities, and a tiered data governance and access structure for evidence-building. Each of these characteristics will be explored more fully throughout this report.

[iii] NVSS Modernization Projects and Initiatives: https://www.cdc.gov/nchs/nvss/modernization/projects.htm

1.3 PROJECT APPROACH

To establish a foundation and understanding of NVSS modernization efforts and opportunities, Clinovations used a range of methods for gathering input for the environmental scan, which included a literature scan, a NVSS modernization scan, interviews conducted with federal stakeholders, and a series of virtual non-federal stakeholder workshops. Figure 2 depicts the environmental scan project approach.

Figure 2. Environmental Scan Approach



Literature Scan

The primary goal of the literature scan activities was to understand common uses, data quality considerations, standards and interoperability needs, and other insights already reported in the literature about NVSS data. The literature scan was intended to support and inform the development of discussion topics that expand beyond known findings in the literature and to identify participants for project outreach to engage in stakeholder workshops to discuss strengths, uses, gaps, challenges, and modernization opportunities using interoperable data. More information, including a summary of the literature scan findings, is included in Appendix A.

NVSS Modernization Scan

Clinovations reviewed the CDC website, public presentations, and peer-reviewed publications to assess publicly available information describing NVSS modernization projects. These sources served as background for interoperability approaches and infrastructure to support statistical datasets and potential NSDS considerations for shared services. Clinovations also reviewed draft Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) Implementation Guides (IGs) for Vital Records Death Reporting (VRDR), Birth and Fetal Death Reporting (BFDR), and Medicolegal Death Investigation (MDI). HL7 FHIR is a healthcare data standard that is becoming a commonly used standard for exchanging health information and is increasing in adoption amongst NVSS data source and exchange systems.[iv,v] Review of the IGs contributed to an understanding of how legacy and long-standing data requirements via paper-forms are modernized for use by interoperable information technology (IT) systems to exchange data using HL7 FHIR standards.

[iv] HL7 FHIR Overview: https://www.healthit.gov/topic/standards-technology/standards/fhir

[v] NVSS Modernization Tools and Technologies: https://www.cdc.gov/nchs/nvss/modernization/tools.htm

NVSS modernization projects also included the development of trainings, communities of practice, and other educational forums to develop interoperability and standards expertise amongst state and local jurisdiction staff. A summary of the NVSS modernization scan and reviewed projects is included in Appendix C.

Federal Stakeholder Interviews

Clinovations worked in collaboration with project sponsors from NCSES and NCHS Division of Vital Statistics (DVS) to identify federal stakeholders to invite to one-hour-long interviews. Two types of federal stakeholder interviews were conducted: 1) select NCHS leadership and staff with responsibilities for NVSS data collection from jurisdictions and processing for statistical dataset generation, and 2) a small sampling of NVSS public or restricted dataset users within federal agencies. Interview findings are summarized within Appendix B.

Stakeholder Workshops

Clinovations designed and facilitated four stakeholder workshops to gather perspectives across various non-federal government stakeholders to understand opportunities for efficiencies and potential uses of interoperable vital statistics data. These stakeholder workshops provided diverse perspectives for a foundational landscape analysis of interoperability considerations for further investigation and analysis. The stakeholder workshops were not used to develop recommendations or provide consensus advice but, rather, to capture a variety of perspectives to inform future NVSS and potential NSDS planning. More information, including summaries, stakeholder perspectives, and findings from the stakeholder workshops, can be found in Appendix B.

Public Project Outputs

The results of this project are detailed within this final project report. This project final report summarizes findings from the environmental scan and provides interoperable data modernization considerations for both NVSS and a potential future NSDS. Clinovations also conducted a public webinar containing a summary of project activities and findings on July 23, 2024.

clinovations
GOVERNMENT + HEALTH

2 CONSIDERATIONS FOR A POTENTIAL FUTURE NATIONAL SECURE DATA SERVICE

2.1 USE OF NATIONAL VITAL STATISTICS SYSTEM AS A MODEL

Project environmental scan findings supported the Advisory Committee on Data for Evidence Building (ACDEB) Year 2 Report Recommendation 3.13, the NSDS should coordinate with stakeholders to develop and promote standards for government data at all levels. Moreover, in the Year 1 Report, ACDEB acknowledges the importance of data standards, consistency, and interoperability for a potential future NSDS, based on the committee's fact-finding process, which included the NVSS. Specifically, this project supports and investigates the recommendation that:

NVSS *processes, systems, and technologies* serve as a prime model of a statistical system that can be used to derive findings and insights by a potential future NSDS.

Figure 3 describes areas where NVSS could inform functions and services of a potential future NSDS in the areas of processes, systems, and technologies.

Figure 3. Areas Where NVSS Can Inform a Potential Future NSDS

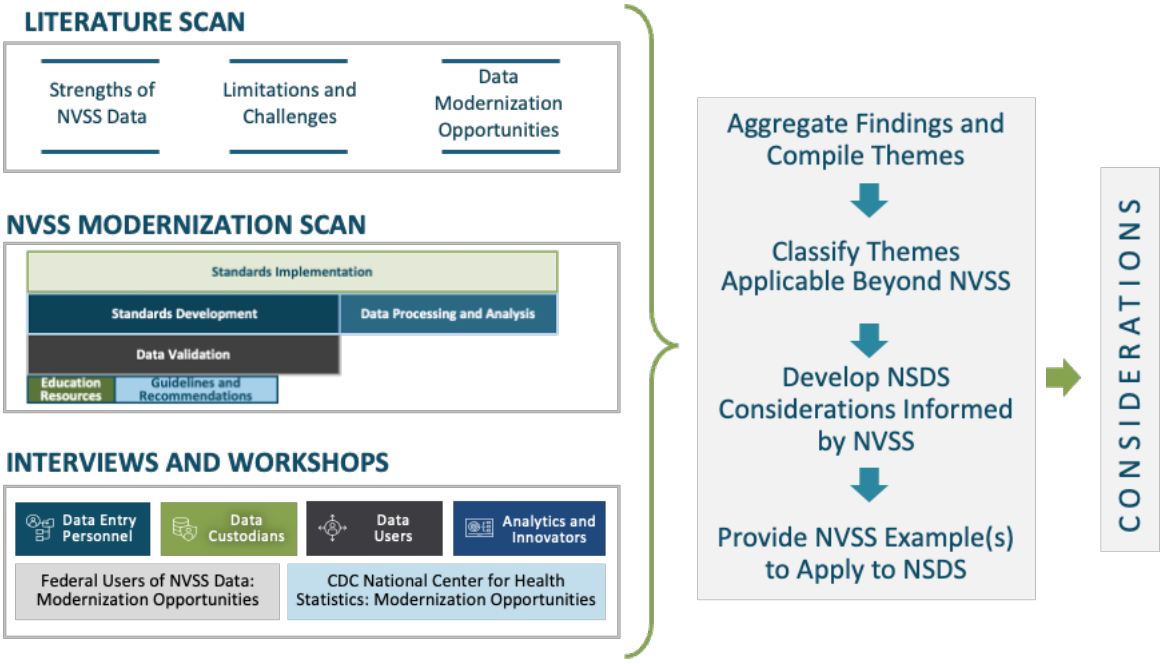
PROCESSES	• NVSS processes involve several traits and characteristics that align with other federal statistical data ecosystems ⇒ ***Established governance structure:*** NVSS does not own the data – the state owns the data ⇒ ***Clear reporting requirements:*** NVSS requires the reporting of data elements at regular intervals with a defined reporting structure ⇒ ***Standardized data elements:*** NVSS requires the reporting of a standardized set of data elements ⇒ ***Multi-stakeholder engagement:*** NVSS collects data from multiple systems and data sources, and is used by both healthcare and non-healthcare stakeholders • NVSS processes can serve as best practices for open standards, support, testing and conformance, data element standardization, and education
SYSTEMS	• NVSS modernization and interoperability implementation efforts offer systems with ready-insights into metadata for a potential NSDS: descriptive, structural, administrative, reference, and statistical • NVSS experience in adapting to new data needs (e.g., Covid-19 pandemic reporting) demonstrates how a very standardized data ecosystem can adapt to changing and evolving needs
TECHNOLOGIES	• NVSS is advancing use of standards-based Application Programming Interfaces (APIs) for interoperable data exchange • NVSS brings experience in using modern computing technologies to code text-based data

Developing Considerations for a Potential Future NSDS

By drawing insights from the NVSS environmental scan (i.e., literature scan, NVSS modernization scan, federal stakeholder interviews, and stakeholder workshops), a potential future NSDS can leverage process and infrastructure key success factors and interoperability enablers to meet its purpose and core functions of providing federal, state, and local partners the services they will need to securely use data to support evidence-building in policymaking and research.

To develop considerations for a potential future NSDS, Clinovations aggregated findings across environmental scan activities, compiled themes, identified and classified themes that could be applied beyond NVSS and health data, developed considerations informed by NVSS, and identified NVSS examples that can be applied to a potential future NSDS for each consideration. Figure 4 depicts this process for developing considerations.

Figure 4. Process for Analyzing Scan Findings to Develop Considerations



2.2 10 CONSIDERATIONS: CHALLENGES, OPPORTUNITIES, AND EXAMPLES

Table 1 summarizes project considerations for a potential future NSDS that are described in detail within this section. Each of the ten considerations includes a description of the associated challenge, opportunity, and a relevant NVSS example. Each consideration is mapped to a potential NSDS function (i.e., communication, coordination, data standardization, and research & development), described within the ACDEB Year 2 Report.

Table 1. Summary of Considerations for a Potential Future NSDS

1.	Navigating and Comparing Statistical Data Sources
2.	Cataloging and Harmonizing Value Sets for Common Data
3.	Developing Data Entry Stakeholder Training and Education
4.	Advancing Skills Through Modern Interoperability Infrastructure
5.	Streamlining Common Data Access Tiers
6.	Expanding Standard Application Process (SAP) Awareness
7.	Modernizing Platforms for Data Access and Analyses
8.	Participating in Standards Development and Advancement
9.	Providing Templates for Common Queries and Computational Analyses
10.	Communicating Best Practices for Rapid Release of Data

1. Navigating and Comparing Statistical Data Sources

Challenge: To maximize use of available (e.g., public-use and restricted-use) datasets, data requestors should research and review a multitude of federal websites, private sector offerings, and research papers to learn about a potentially common data type or dataset. There is considerable burden in reviewing this information and evaluating whether a particular dataset is fit-for-purpose, cost effective, and accessible for use. In addition, desired data or uses may be subject to state laws that restrict data availability or use. Federal agencies and researchers do not have a single source to review state laws related to data that is used in statistical datasets.

Opportunity: The NSDS could serve as a single point of knowledge for commonly requested data types and datasets by providing online resources (e.g., GitHub repository, local/state laws and regulations comparisons and analyses, navigation tools and links to downable datasets, online portals) as well as data navigators or data concierges to guide data requesters to potential data sources. A potential NSDS could serve as a resource to discover state laws that pertain to priority data that may impact data availability for exchange, linking, research, or surveillance. NSDS could maintain metadata on datasets that could help data requesters determine the best data asset for their intended purpose. Potential metadata could include: Data Sources; Authorized Uses and Data Access Tiers; Costs; Population Coverage; Data Storage; Retention; Protection Requirements; Other Conditions of Use, and Applicable State Laws or Restrictions.

☑ **Communication** ☐ **Data Standardization** ☑ **Coordination** ☐ **Research & Development**

Example: Death Information. There are multiple options for death information, from both federal and non-federal (commercial) sources. The data elements available, standards used, and authorized uses vary across these data sources. Examples of death data sources (not exhaustive) that have different data use, costs, coverage, or storage requirements include:

- *CDC NVSS Restricted Use Datasets* can be used for authorized and approved purposes only, set their own requirements for secure storage and use of the data, and are available at no cost. If NVSS restricted use datasets are used in a Research Data Center (RDC) or a Federal Statistical Research Data Center (FSRDC), costs are incurred.
- *CDC National Death Index (NDI)* offers linkage services based upon a data file with identifiable data. NDI linkage is solely for statistical use by public health and medical investigators and incurs costs.
- *Department of Commerce's National Technical Information Service (DOC NTIS)* sells a public file of death information obtained from SSA, by charging fees, for use by agencies and private organizations (such as banks and credit card companies).
- *National Bureau of Economic Research (NBER) Public Use Data Archive* is a resource offering a mix of public-use economic, demographic, and enterprise data obtained over the years to satisfy the specific requests of NBER-affiliated researchers. Files here are often in more convenient formats than the original data source. For NVSS data, NBER uses original NVSS files to generate Stata, SAS, and .CSV files to facilitate use by analytic tools and statistical software.
- *Social Security Administration (SSA)* offers access to death information to federal benefit-paying agencies and states, each with its own data exchange request process for a variety of use cases including surveillance, statistical purposes, benefit information tracking, and other purposes. SSA provides vital status data for health researchers contributing to national interest (incurs costs).
- *Veritas Fact of Death Dataset* is a commercial ready to use dataset that contains over 38 million records and is linkable to deidentified real-world data partners (i.e., Healthverity and Datavant) and within secure cloud environments.

2. Cataloging and Harmonizing Value Sets for Common Data

Challenge: Value sets for common data elements can vary greatly across data source organizations, state systems, and federal statistical data processing systems. A value set is a set of codes, values, or data field input options that are used to capture or store data elements within a common use case. Value sets may be standards-based, such as for medications, procedures, states and jurisdictions, race, ethnicity, or may be a collection of commonly used values. There are cases where multiple standards exist, and industry specific efforts may or may not converge on a single standard or select the most appropriate standard based upon the use case. Often source data systems, state/local data custodians, or federal data aggregators will need to map or configure local codes to the desired value set used by receiving systems or databases. Certain datasets may be collected or aggregated for statistical use for different purposes or domains (such as health, education, income, crime data). Currently, for NVSS, source systems or data intermediaries (e.g., states) could be required to report a smaller value set than is available in source systems. Therefore, granularity and specificity available at the source is not available to future data users once the data have been mapped and transformed to the federal standard value set. There is no single, authoritative source of value sets used by state/local systems, federal systems, and industry standards and researching available options can be time-consuming and burdensome.

Opportunity: An NSDS could serve as a single or central resource for coordination and navigation across available value sets. NSDS functions could involve publishing or collating value sets, serving as a data concierge to evaluate values across different value sets, and/or initiate efforts to harmonize and guide selection of standards-based value sets used in federal statistical data systems and other federal and state reporting programs. In addition, an NSDS could identify value sets for which established standards are not available and could champion efforts to work with standards development organizations (SDOs) to develop and advance standards.

Promoting standards-based adoption or mapping at the source, or data intermediaries, can support alignment across data for evidence building and allow for local/regional flexibility by exchange of provenance information from the source. In this model, federal datasets could develop datasets that align with a selected standard, but also have source data available for researchers (public or restricted-use, depending upon datasets).

☑ **Communication** ☑ **Data Standardization** ☑ **Coordination** ☐ **Research & Development**

Example: Race and Ethnicity Value Sets. Race and ethnicity data are used in multiple statistical datasets. However, there are many standards or commonly used value sets utilized by source and data systems. Some source systems support selection of more than one.

1. *Office of Management and Budget (OMB)* reviews and maintains national standards for race and ethnicity data selected and used by many data source systems. For statistical purposes, data may not be reported at this level of specificity for all localities. Some federal statistical programs select a smaller subset of values for reporting to federal agencies, including the American Community Survey (ACS) and 2000 and 2010 Decennial Census. U.S. Department of Health and Human Services (HHS) standards for race and ethnicity include additional granularity that roll-up to the OMB standard.
2. *CDC Race and Ethnicity Code Set* is based upon the minimum race and ethnicity categories defined by OMB and includes a more detailed set of race and ethnicity categories maintained by the U.S. Census Bureau. The Public Health Information Network Vocabulary Access and Distribution System (PHIN VADS) uses the HL7 value set based upon the CDC Race and Ethnicity code system.
3. *Indian Health Service (IHS)* collects and maintains more granular data on its population, including tribal affiliations.
4. *HL7* maintains its own code and value sets used by health IT products and implementation guides. *HL7* maintains its own code and value sets used by health IT products and implementation guides. The HL7 Vocabulary Work Group evaluates and harmonizes proposed changes in other national standards.

3. Developing Data Entry Stakeholder Training and Education

Challenge: Individuals who enter data into source systems or originating systems for statistical datasets may be capturing data for other purposes and workflows that align with their primary roles and responsibilities. These individuals are often not aware of the longer-term implications of the data they enter and how it will be used in an aggregated dataset for statistical analyses by researchers, national reporting, or other purposes. For example, without context, a clinical team member may provide different level of detail for a narrative text-based data element than a birth clerk at a hospital. By providing additional training and education for individuals who enter data into source systems, primary data enterers and organizations could improve their comprehensive input, data collection workflows, and understanding of the importance for accurate information.

Opportunity: The NSDS could develop guidance for cohorts of communities that enter data into information systems that are transformed or mapped to coded values for development of statistical datasets or used in evidence-building. Collect insights on data entry-driven variability to support consistency across stakeholders. Develop short, easy-to-use resources (e.g., tip sheets, short videos, case studies) that highlight and explain how the data captured by these cohorts will be used. Resources could be shared with professional organizations that serve these cohorts and used by individual states, communities, or private sector organizations to inform and educate individuals so that the data quality, completeness, and comprehensiveness is improved once aggregated for statistical use.

☑ **Communication** ☐ **Data Standardization** ☑ **Coordination** ☐ **Research & Development**

Example: Underlying Cause of Death Information. Medical examiners and coroners participating in this project's stakeholder workshop indicated that variability exists across their colleagues and jurisdictions in completing underlying cause of death information. The example provided was in a case of a motor vehicle accident that resulted in a death, where participants indicated variability where some individuals may add toxicology result related information to the underlying cause of death (such as drugs present in the decedent's body), while others may not include this information and indicate that the underlying cause of death was a motor vehicle accident. By understanding what data and information is needed, both for clinical and non-clinical use cases, individuals who input data into source systems or originating systems for statistical datasets will be more likely to identify and accurately input relevant necessary data and information.

NCHS developed tools and training to support physicians, medical examiners, and coroners in understanding how information written in a cause of death certificate can be as important as information written in a medical record, as the information is used to detect trends that determine public health programs and funding allocations as well as the clinical record. The patient's medical doctor or primary provider of care may not be the individual who is completing the cause of death section. NCHS has developed training and instructional materials for Death Certification: Writing Cause-of-Death Statements that include tailored handbooks for physicians vs. medical examiners and coroners, online self-guided training course, and a Cause of Death Mobile Application that provides a quick reference guide for physicians, medical students, and other individuals.

4. Advancing Skills Through Modern Interoperability Infrastructure

Challenge: New analysis and visualization tools and modern computing methods are rapidly emerging to support user interests to consume and navigate data that are being exchanged and available in public and private datasets. New toolsets are becoming de facto analytic platforms within certain public and private organizations and individuals with proficiency in these toolsets can demand high compensation, significant training, and certifications. Meanwhile, public sector analytics staff, especially at state and local jurisdictions, are highly skilled at using traditional statistical packages or legacy technology platforms and are compensated at more modest levels than their counterparts. While many states and local jurisdictions have interest in modernizing their platforms and capabilities, workshop participants indicated that financial resources and skills limitations have hindered their ability to adopt modern technologies. While online or course-based training may be available, it may require investment in vendor technologies or high costs and only offer didactic training rather than the experiential training.

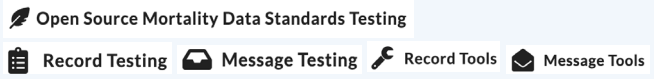
Opportunity: An NSDS could support development of implementation guidance, test environments, sandboxes, pilots, and test data (e.g., real world data or synthetic data) for upskilling of state, local, and federal staff by providing resources available prior to making state/local level investments. Provide federal support to states and local jurisdictions to implement and converge upon modern standards for data exchange (e.g., HL7 FHIR for RESTful API-based data exchange). Host connectathons, ethical data hacking, "war games" scenarios, data analytic workshops, and collaboration venues with established test data and authorization protocols to advance modern methods for data access, exchange, and security. By providing infrastructure and forums for "hands on" learning, government staff can advance their capabilities to plan, procure, implement, and use modern platforms and new standards for data exchange.

☐ Communication ☐ Data Standardization ☑ Coordination ☑ Research & Development

Example: Funding and Implementation Support to NVSS States and Jurisdictions to Implement FHIR. NCHS funded and directly provided support to state and local jurisdictions to implement the new FHIR standard for exchange of vital records information. In addition to providing funding, NCHS conducted pilots with a small group of states with high readiness to participate and then supported the onboarding of additional states.

NCHS provides direct technical assistance and support to states and jurisdictions and conducts testing individually with each jurisdiction. States have used the federal funding and skills building support to build in-house FHIR expertise and experience, which is costly and difficult to procure from external contractors and consultants that are in high demand across the healthcare IT market. By providing federal support, states and jurisdictions were able to prioritize and support this NVSS modernization initiative.

Example: Canary EDRS Testing Framework. Canary is an open-source testing framework that supports development of systems that perform standards-based exchange of mortality data. Canary provides tests and tools to aid developers in implementing the Vital Records Death Reporting (VRDR) with tools that support synthetic record generation, ability to import/export test records, and validation of test and synthetic records that state and local jurisdictions can use to test their systems' FHIR conformance.



5. Streamlining Common Data Access Tiers

Challenge: Restricted data use applications and conditions of use can be cumbersome to researchers and pose limitations for data analysis and evidence generation. Researchers reported avoiding projects that require restricted data use applications due to effort and timeline impacts. Data entry professionals indicated that as multiple data sources may exist to submit data used for statistical datasets and evidence-building, they do not have access to the source data due to access limitations. Therefore, certain data are still obtained via phone calls, faxes, and other traditional and time-consuming methods, rather than modern, access-tier driven electronic data access and exchange.

Opportunity: A potential future NSDS could develop best practices and recommendations for streamlined and simplified data access tiers that apply to statistical datasets that use data generated at state and local levels. NSDS could offer data concierge and informational services to support navigating state/local regulations in developing their access tiers. Potential opportunities include:

- A potential future NSDS could convene federal agencies that generate statistical datasets to determine if a "superuser" tier could streamline repeated access to the same restricted data at defined intervals for researchers from known entities. A potential future NSDS could also consider whether this or an additional "superuser" tier could apply to federal users.
- An analysis of data access and data use policies across statistical datasets could identify whether some data access requests (or steps within processing requests) could be centrally managed or if resource burden could be leveled across agencies that review data access requests (e.g., SAP).
- A future NSDS could offer support to data custodians at state/local levels to facilitate inter-state and cross-jurisdictional data sharing for areas where researcher needs are not met using federal restricted datasets. This support could involve communication of examples and best practices across data areas (e.g., vitals data) that use systems that can provision levels of authorization and access based upon user type.

Refer to the "Data Access Tiers" subsection within Section 3 "Analysis and Discussion" for further discussion regarding potential NSDS services to support expanded data access tiers.

☑ **Communication** ☐ **Data Standardization** ☑ **Coordination** ☐ **Research & Development**

Example of Data Access Tier Implementations – Vital Records Systems. Many vital records systems can provision tiers of users by system user type (e.g., data entry personnel and data user). Key factors for managing data governance and the access tiers at state/jurisdiction state vital record offices (VROs), vital records health IT system vendors, and at federal agencies include:

- Identifying which users/roles receive access to data files/sets,
- Identifying which users/roles obtain access to appropriate modules and screens within the health IT solution,
- Identifying which users/roles obtain access to identified vs. de-identified data, and
- Management of data governance and access tiers across different vital records and vital records statistics stakeholders (e.g., health IT systems, state/federal agencies, county coroner's offices).

6. Expanding Standard Application Process (SAP) Awareness

Challenge: Researchers have individually applied for restricted-use data assets from the federal statistical agencies , as well as state-level restricted datasets, and have indicated that the process is time-consuming and cumbersome. Researchers were experienced with established application processes and were not always aware of other application options.

Opportunity: SAP offers a streamlined "front door" to restricted-use datasets from the federal statistical agencies. A potential, future NSDS could expand communication and awareness amongst known users of restricted data files across recent applicants.

Increased awareness amongst states and local jurisdictions could serve as a model for other areas where an SAP model could benefit data users, such as requests for state/jurisdiction-level restricted-use datasets. A potential future NSDS could coordinate across states and jurisdictions to foster a more standardized application process modeled on the and consider a metadata catalog of data that are available for discovery within or across states.

☑ **Communication** ☐ **Data Standardization** ☐ **Coordination** ☐ **Research & Development**

Example: NVSS Data Users Participating in Project Workshops Did Not Use the SAP. In reviewing the NVSS NCHS Restricted Vital Statistics Data Request Application Form and the SAP, there is some streamlining of information requests available within the SAP. However, researchers and state public health agencies participating in project workshops did not indicate use of the SAP and discussed the burden of applying for multiple national and state-level datasets. Frequent users of NVSS data may not be aware of new alternatives such as the SAP.

Interagency Council on Statistical Policy
Leaders of the United States Federal Statistical System
Standard Application Process

7. Modernizing Platforms for Data Access and Analyses

Challenge: Some federal statistical datasets include conditions of use for restricted-use data that may need to be evaluated in the context of modern data analytic tools and computing platforms. Some federal statistical datasets (e.g., NVSS) have limitations on use of cloud-based tools and do not support de-identified individual level data-linking across data assets. Researchers in stakeholder workshops identified needs for individual linked data where research teams do not require access to identifiable information, but the research capability is limited without linked data. Researchers also indicated that they are not able to use and leverage commercial analytic tools and platform investments due to limitations that prevent use of cloud-based tools.

Opportunity: A future NSDS could perform an assessment of limitations of data assets from federal statistical agencies related to conditions of use and feasibility of using modern computing technologies and platforms. Where restrictions prevent use of emerging tools (e.g., cloud-based tools), a future NSDS could lead evaluations and studies of these tools and provide guidance to federal statistical agencies that may support more flexible conditions of use, while preserving privacy and confidentiality. Modern tools and platforms to support and evaluate could include:

- Cloud-based analytic tools and storage environments,
- Data lakes for secure computing – centralized and decentralized models,
- API-based data access – authentication, authorization, and access protocols,
- Ethical hacking sandboxes and scenario-based testing environments, and
- Privacy-preserving technologies to support data linkages.

☐ Communication ☐ Data Standardization ☐ Coordination ☑ Research & Development

Example: NVSS Restricted Data Conditions of Use Prevents Use of Cloud-Based Systems. Storage and access of data files from the NVSS using a cloud-based system is not permitted by NCHS. NCHS could reexamine the common agreement between NCHS and the 57 states and jurisdictions and its internal policies to identify if this restriction can be revised.

The industry is gradually adopting cloud-based architectures. To modernize the NVSS data ecosystem, NCHS could consider reviewing its policy on the storage of restricted-use data files and determine whether the tools that are supported by cloud-based architecture is necessary for NVSS data users. Cloud-based architectures support capabilities such as:

- Secure computing environments,
- Modern computing analytic tools that employ artificial intelligence (AI), machine learning (ML), and/or natural language processing (NLP),
- Interoperable data sharing and exchange, and
- Authorization and provision of data access tiers.

8. Participating in Standards Development and Advancement

Challenge: Federal agencies that generate statistical datasets often standardize their forms and data collection instruments with defined data element structures that align with statistical groupings. Examples of these forms include the U.S. Standard Certificate of Birth, the U.S. Standard Certificate of Death, and the Decennial Census of Population. The data collected in these forms may have been standardized decades ago, as implementation and adoption of updates take additional years. In the past, much of this information may have been collected manually using paper-based forms and records, evolving to electronic text-based forms. However, much of the data collected originates in source data systems, where industry standards – or commercial product conventions – may offer a standard or default value set that differs from federal forms and certificates.

Opportunity: With a modern, data-driven, electronic ecosystem, the use of standards can support faster updates and changes to data elements or value sets collected. These techniques could support consistent standards adoption from local to national levels. By actively engaging with SDOs and industry standards efforts, the federal government could align and foster the development of standards that supports needs at the point of capture (data entry) and statistical dataset standardization. OMB is one example of a federal agency informing policymaking through its Federal Data Strategy, and as a value set custodian for areas such as race/ethnicity and standard occupational classification. A potential future NSDS could:

- Identify and participate in national standards efforts,
- Develop a playbook for standards advancement and participation within federal agencies,
- Provide technical assistance to agencies to guide standards development, and
- Compile lessons learned from agency-supported pilots and standards efforts.

☐ Communication ☑ Data Standardization ☑ Coordination ☐ Research & Development

Example: NVSS FHIR Implementation Guidance and Specifications Development. NCHS has been actively involved in the development of multiple FHIR implementation guides (IGs) to facilitate vital records data exchange using standards. This includes IGs and associated testing frameworks, such as the VRDR (VRDR FHIR IG) and Canary; MDI (MDI FHIR IG) and Raven; and BFDR (BFDR FHIR IG); and other proof-of-concept demonstrations and reference implementations. NCHS offers a number of technical resources and a NVSS Modernization Toolkit to support use of FHIR and other modern tools and standards. NCHS coordination with the Assistant Secretary for Technology Policy/Office of the National Coordinator for Health IT (ASTP/ONC) has resulted in a proposed rule (July 2024) for certified health IT to use the BFDR FHIR IG to transmit, receive, validate, parse, and filter birth records to/from public health agencies.

CDC has developed an HL7 FHIR IG Checklist, for use by CDC programs to help others navigate available resources around FHIR adoption and API-based approaches to interoperability. CDC's YouTube Channel offers many videos to support its modernization community of practice and other stakeholders using FHIR.

By participating in the standards developments process, NCHS and CDC would be positioned as the definitive authority on the developed data standards, enabling them to provide in depth implementation guidance and ensure usage by federal and industry data users via policymaking.

9. Providing Templates for Common Queries and Computational Analyses

Challenge: State and local government departments and other researchers use statistical datasets and other data assets to perform regular analyses at scheduled intervals (e.g., annually, quarterly, monthly) to generate reports and public dashboards that are similar from state to state or across jurisdictions. These data analysts and researchers must access, download, and analyze the same data locally for each interval for which public dashboards or reports are published. A common set of options to rapidly communicate state, county, and jurisdiction-level data could reduce repetitive efforts.

Opportunity: Established views of community and regional needs or metrics could be incorporated into a potential future NSDS using templated queries and computational analyses tools. "Fast statistics" can be defined as a set of commonly requested data elements or metrics. This modality could be employed for data elements or metrics that are commonly requested. A "fast statistics" dashboard could fill two niches within a data ecosystem: 1) a downloadable set of commonly queried data elements and metrics, and 2) a visualization of these commonly requested data elements or metrics. A "fast statistics" dashboard hosted by a potential future NSDS, which is already being explored by the Federated Data Usage Platform (DUP) NSDS-D project, could:

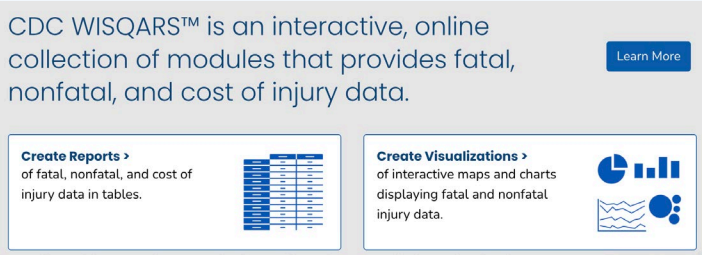
- Reduce data user burden when processing data,
- Remove needs for users to visualize commonly requested data,
- Create a one-stop-shop for various federal data sets, and
- Improve data transparency and public trust of federal data sets.

"Templated code library" can be defined as a set of template code scripts that data users can modify for commonly completed computational analyses techniques. This modality could be used for more obscure data elements or metrics that are less likely to be of interest to the average data user.

To further reduce the technical knowledge threshold, a potential future NSDS could consider the development of "plug-and-play" modules for common computational analysis techniques (e.g., regressions, cluster analyses, and Monte Carlo statistical models). These "plug-and-play" modules require even less technical knowledge to use, because inputting computational analysis parameters that are based upon the data user's sample population would not require the data user to review code. The Democratizing Data Initiative is working with government agencies to support effective use of public data and may provide a venue for exploring these capabilities. In addition, the recently awarded Secure Compute Environment Testbed for a National Secure Data Service (SCET) Demonstration project may offer insights for templating common queries and computational analyses techniques.

☐ Communication ☐ Data Standardization ☑ Coordination ☑ Research & Development

Example: CDC Web-based Injury Statistics Query and Reporting System (WISQARS) - NVSS Data Visualization. CDC WISQARS is an online, interactive collection of modules that provides web-based injury statistics query and reporting data. This site offers interactive maps and charts, state-to-state comparisons, infographics, and visualization and reporting tools.



Example: Community Health Needs Assessments (CHNAs). Many states conduct and publish CHNAs at regular intervals. Tax-exempt hospitals must conduct CHNAs every three years, and the U.S. Internal Revenue Service (IRS) also requires tax-exempt hospitals to obtain input from at least one state, local, or tribal public health department, as well as from medically underserved, low-income, and minority populations. Population profiles could be harmonized across states and IRS requirements to readily visualize or download common templates.

Example: CARES Map Room & SparkMap– NVSS Data Visualization. The Center for Applied Research and Engagement Systems (CARES) has ingested NVSS public use file data using APIs and serves as a data aggregator across different data types (e.g., NVSS, census, Environmental Protection Agency (EPA) data) to package for public health and population health users (free). SparkMap is a subscription (fee) based solution for CHNAs.

10. Communicating Best Practices for Rapid Release of Data

Challenge: Data for evidence building, such as data published by federal statistical agencies, are often published annually, with an inherent delay due to finalization of source data, analysis, and processing. COVID-19 demonstrated the need for more rapid access to data in support of public health needs, natural disasters, and other potential force majeure and crises circumstances.

Opportunity: A potential future NSDS could evaluate rapid response programs and efforts across federal agencies that provide rapid access to provisional data. An NSDS could develop best practices across agencies, guide agencies in developing their rapid response programs, and develop strategies for mobilizing and connecting with research and industry data users who need access to rapid response data.

☑ Communication ☐ Data Standardization ☐ Coordination ☑ Research & Development

Example: Vital Statistics Rapid Release program. The Vital Statistics Rapid Release program provides timely access to vital statistics provisional data for public health surveillance, through releases of: 1) focused surveillance activities, 2) Quarterly Provisional Estimates, 3) Vital Statistics Rapid Release Reports, and 4) state and national provisional control counts of births, deaths, and infant deaths based on a current flow of vital statistics data from state vital records offices. This program provides dynamic visualizations of provisional data that estimates health indicators for public health practitioners, researchers, and health policymakers, offering more timely insights than possible using final data which are published annually.

Example: NCHS COVID-19 Mortality Data Response. Within weeks of the first U.S. reported cases of COVID-19, NCHS mobilized to support rapid-response to support the growing need for real-time data on COVID-19 related deaths. NCHS applied a standard ICD-10 (International Classification of Diseases, Tenth Revision) disease classification code for underlying cause of death, by which 92% of COVID-19 related deaths were identifiable using the U07.1 code. NCHS manually reviewed and coded 100% of initial COVID-19 records because it was a novel cause of death to enhance its data processing and education and training for individuals that document and enter death certificate data. NCHS was able to release NVSS COVID-19 real-time provisional death data through daily, weekly, and ad hoc analyses through:

- Decades of standardized data elements for all US deaths,
- Use of interoperable data standards to guide automated and manual coding processes, and
- New statistical methods for more context on provisional data beyond COVID-19.

3 ANALYSIS AND DISCUSSION

This section provides perspectives for further consideration based upon reflections and insights obtained during the project and development of this final report. This section is organized to directly respond to questions posed during the inception of this project and is intended to seek continued discussion and analysis as the NVSS team continues its modernization journey and NCSES seeks to draw insights and lessons learned from NCHS accomplishments in, and opportunities for, NVSS modernization through use of interoperable data.

3.1 PROMISING USES AND UNMET NEEDS

What are the most promising uses for interoperable vital statistics to support timely research and public health surveillance?

The most promising uses for interoperable vital statistics can be categorized into two buckets: population-level insights and individual-level insights. Interoperability is especially key when a data user is utilizing dataset(s) from both healthcare and non-healthcare data sources especially in areas where data standards efforts may vary across sectors.

Population-level insights may be used to test and/or validate hypotheses based on a sample population to then describe a larger population. Interoperable aggregated, statistical datasets are commonly used to obtain population-level insights, because the dataset predefines the sample population and provides basic demographic information about the sample population.

Promising uses for interoperable vital statistics data include:

- Clinical Care,
- Clinical Trials,
- Economic Analyses,
- Policymaking,
- Public Health,
- Research,
- Syndromic Surveillance (e.g., Pandemics, Epidemics, and Disease Outbreak), and
- Training of Machine Learning (ML) and/or Artificial Intelligence (AI) models and algorithms.

Interoperable vital statistics data could also provide individual-level insights. In many of these scenarios, the interoperable vital statistics dataset serves as the population baseline for individual-level metrics. Value(s) from or calculated from the interoperable vital statistics dataset are inputted into individual-level insights metrics as the "denominator". A primary clinical example of individual-level insights are risk calculators, which use interoperable vital statistics to generate the baseline risk (i.e., the denominator).

In addition to the use of interoperable vital records datasets, some data users require the use of interoperable record-level vital records datasets for both healthcare and non-healthcare use cases. In scenarios where a data user is researching a hypothesis or problem and may be required to link multiple datasets, data users may require interoperable record-level data to link the healthcare dataset with the non-healthcare dataset.

What are the key criteria for this determination?

There are several key criteria for determining whether a use case requires interoperable data. Generally, use cases require interoperable data when the data need to be:

- Accessible by different data users,
- Captured in different data standards (e.g., HL7 Version 2 (HL7v2) vs. HL7 FHIR),
- Duplicated to preserve the legal version of the document and/or dataset,
- Exchange between two disparate data platforms,
- Housed in different architectures (e.g., local storage vs. cloud-based storage),
- Obtained in real-time or in a timely manner, and
- Updated periodically (e.g., hourly, daily, monthly, yearly).

Certain use cases require either interoperable statistical-level data or record-level data. Data users require interoperable record-level data when datasets need to be linked or aggregated within a larger data lake or repository. When large amounts of data need to be collected or accumulated to generate a dataset that satisfies the use case's requirements, data users require interoperable statistical-level data.

How do unmet needs vary across the federal, state, and local levels?

Unmet needs can be categorized into six general themes – data access and exchange, data granularity and quality, data standardization, data sources, data timeliness, and interoperability between IT systems. These needs are met at various levels across the federal and state and jurisdictional levels.

Data access and exchange for restricted-use datasets are burdensome due to non-standardized data request applications across federal statistical agencies and between states, resulting in data request applications that contain different fields. At the federal level, progress is being made with the development and adoption of the SAP, which standardizes the data request applications across all federal statistical agencies. However, no observable progress is occurring at the state and jurisdictional levels. One potential solution to ameliorate this burden is to use the SAP as a potential model to develop resources to navigate and compare access to state and jurisdiction-level data or develop and publish best practices, guidance, and state-by-state comparative analyses so that states and technology vendors that support states and jurisdictions could implement more standardized application processes.

User needs are generally met for public-use datasets with communities of researchers and data users that are well-familiar with longstanding datasets. However, data users noted that some file formats are more difficult to import and use compared to others. One example of a challenging file format is .DUSMCPUB, a legacy custom extension for Detailed U.S. Multiple Cause Public Use files (the underlying file is a fixed format text file), which is difficult for data users to import into R. Two examples of easy-to-use file formats are .CSV and .TXT files. All data custodians should release datasets in easily importable and commonly used file formats.

Data granularity and quality are commonly cited unmet needs at both the federal and state levels. For many data users, statistical datasets are not granular enough because they are not record-level datasets. Additionally, due to code and value set differences, data granularity is lost for certain data elements (e.g., race and ethnicity data) once they are aggregated at the national level. At the state level, some states provide more granular data elements due to having populations that require additional stratifications. When the data are received at the federal level, the data must be mapped to common value sets (e.g., the lowest common denominator of codes and/or values that exists within the aggregated dataset), so that the data can be aggregated into datasets that preserve confidentiality.

Occasionally, data users question the quality of their data. Datasets pass through several different "checkpoints" (e.g., data are captured, data are reviewed and confirmed, data are reported to the state, state reviews and confirms, state enters the data into a state-level dataset, and state reports entire dataset to the federal level) before reaching the federal data ecosystem. The chance of errors being introduced to the dataset increases each time a mutable dataset is exchanged or is modified. It is common for aggregated datasets to lose some granularity and quality once they reach the federal level. Progress is being made at both the state and federal levels by introducing data validation modules into IT systems to ensure that data elements are entered correctly.

Data standardization is challenging both at the state and federal levels. While vital records data elements are standardized at the federal level (i.e., U.S. Standard Certificate of Birth (2003 Revision) and U.S. Standard Certificate of Death (2003 Revision)), code and value sets are not standardized across federal agencies and divisions and departments within the same federal agency. Code and value sets are often determined by use cases. For example, historically, ICD codes have been used for billing purposes, while SNOMED codes have been used for clinical purposes, even though both codes may map to the same literal text. However, progress is being made at the federal level with the introduction of automated data transformation tools. One example is the NVSS MedCoder tool, which transforms literal cause of death text into the ICD-10 standard. This concept could potentially be expanded to transform data elements that are in one data standard or value set into another, provided that the mappings between the two data standards or value sets have been completed.

Data elements collected are not standardized across states, as states may collect additional vital records data elements for various purposes (e.g., required by law, assistance with a state program or project that requires vital records data). Data standards amongst state data repositories and registries (e.g., vital records office, department of motor vehicles, immunization registries) may not align. One example is the lack of interoperability between state vital records offices and state immunization registries. Currently, state vital records offices exchange vital records data primarily in the inter-jurisdictional exchange format (IJE), with some exchanging data in the HL7 FHIR, while state immunization registries are exchanging data in HL7v2, specifically HL7v2.5.1. Additionally, code and value sets are not standardized across states for standardized data elements (e.g., race/ethnicity and marital status).

Advancements are occurring as some state data custodians have supported expanding the U.S. Standard Certificate of Birth (2003 Revision) and U.S. Standard Certificate of Death (2003 Revision) for research purposes. States have also been open to the adoption of HL7 FHIR, which is a federally recognized data format standard that is gaining traction across all healthcare stakeholders (e.g., providers, patients, payers, states, federal agencies, health information exchanges, qualified health information networks, and third-party app developers). Federally, there is support for HL7 FHIR to be the healthcare data standard lingua franca and Rosetta Stone.

Federal users and external, non-federal data users both reported data timeliness as an unmet need, which is an area of ongoing modernization and focus within NCHS NVSS modernization efforts. The majority of states report their annual vital records datasets in a timely manner. However, year-to-year, there are some states that do not report their data in a timely manner. This delays federal agencies, because federal agencies do not have a national dataset to release.

Some federal agencies have aimed to mitigate frustration with delays by releasing provisional datasets, which are released monthly. Furthermore, some federal agencies release provisional annual datasets, as finalization of annual datasets requires additional data processing and time. However, these delays may be due to the states' inability to obtain data from the point of data collection. As a result, additional education from federal agencies (e.g., toolkits and webinars to assist states in setting up a data collection infrastructure and workflow to support the national reporting program) and from states (e.g., educating the data collection workforce on collecting the data and the importance of data collection) may improve data timeliness.

The lack of interoperability between IT systems is cited as an unmet need at both the federal and state levels; many IT systems are not interoperable with one another, as they were built independently. Thus, these systems tend to be disparate and siloed, resulting in repeat data collection efforts, higher IT maintenance costs, slower data report out times, and less overall IT infrastructure flexibility. Progress is being made through the adoption of HL7 FHIR, which has regulatory backing to be the interoperability data standard of choice for health data. Additionally, some states and federal agencies have explored the adoption and implementation of cloud-based architectures for IT systems, which are by nature more interoperable compared to IT system architectures that save and stores data locally.

Data Access Tiers

What would be the minimal level of data access (e.g., aggregated data or confidential microdata) to meet the needs identified by these opportunities?

Requirements for data governance and access can be grouped into two buckets: data for input (Data Entry) and data use (Data User).

DATA ENTRY

To manage data access for individuals who capture and input data into IT systems and data platforms, many systems can assign, permit, and monitor by system user type (e.g., data entry personnel and data user). Everyone who interacts with an IT system and/or a data platform should be assigned a role. Each role should have a predetermined level of access to a set of data and areas within the IT system and/or data platform. Thus, each system user will be granted the appropriate and minimum level access to the data, based on their assigned system role.

Stakeholders (e.g., providers, medical examiners, coroners, funeral home directors, state agencies, federal agencies) who access data to input or modify data elements in the source system primarily require record level data elements. State and federal agencies who assist organizations (e.g., health systems) with research activities (e.g., data mapping and standardization) require record-level data to transform the pre-existing data element into a conformant data standard.

These settings are configurable within the IT system and/or data platform during the implementation and configuration stage of an IT system implementation. However, support is needed around best practices for managing data access for individuals who capture and input data into IT systems and data platforms that are not owned by the individual's organization. While external users should have more restricted access to IT systems than internal users, there is an opportunity to update current access and governance models for statistical data, some of which are overly restrictive and may prevent individuals from completing their duties and obligations. One example is when a coroner or medical examiner requires outside data from a hospital medical record to complete a death certificate. Potential opportunities include development of best practices for hospitals, health systems, and state/local health information exchanges to provide access to individuals responsible for completing death certificates. In addition, case management systems that support these users can adopt interoperable data standards (e.g., HL7 FHIR, API-based query or subscription capabilities) to offer technology-driven integration between case management systems and electronic health records (EHRs).

These are two examples of potential solutions to decrease data entry burden for external IT systems or data platform users:

1) One potential solution is to grant authorized and authenticated external users access to the data within the IT system or data platform by assigning roles in the manner of other current cloud-based data/document repositories (e.g., shared Google Drive, shared MS Teams/SharePoint). In addition to the roles that exist within the IT system and/or data platform, a "guest" tag should be incorporated, so that data access audits and provenance can be monitored, managed, and maintained. If implemented, IT system managers should thoroughly review its authorization and authentication workflows and role-based access configurations (to ensure "guest" or external users only have access to the minimum data necessary and are not permitted to update or modify data) to ensure privacy, confidentiality, and security is maintained.
2) A potential longer-term solution is to expand data access for external data entry users by requiring the user to agree to a business associate agreement (BAA) or equivalent agreement with data source providers (e.g., hospitals and health systems, health information exchanges (HIEs)). For example, while medical examiners and coroners have access to decedent records when invited to a case, they do not have user rights to login remotely or query EHR systems for records. This would provide authorized individuals with the right to query or access EHR or HIE systems as and could reduce data gathering and review burden.

DATA USER

Within current federal statistical systems, federal interview and stakeholder workshop participants noted three common access options for available datasets: 1) public-use, 2) restricted-use, 3) federal research data centers. The determination of a dataset's access level is based primarily on the sensitivity of the data elements contained within the dataset. The access level assigned for datasets increases depending on the sensitivity of the dataset. For datasets that include data that are more sensitive or protected by nature (e.g., personal identifiable information (PII), geographic location, Heath Insurance Portability and Accountability Act (HIPAA) Safe Harbor data elements), a more restrictive data access level (e.g., restricted-use or federal research data center) is assigned.

Aggregated, statistical datasets should suffice for many data users who only require a dataset that describes the demographic of a sample population. Use cases for these types of datasets include research, public health, syndromic surveillance, evidence for policymaking, and evidence for financial investments. Data users who wish to link or combine datasets require more granular, record level data that includes identifying information. In these use cases, the identifying information can be hidden from the data user by employing privacy preserving technology (PPT) (e.g., encryption). An example use case is when a data user is researching the relationship between a healthcare and a non-healthcare topic (e.g., the effects of disease on an individual's finances, the effects of chronic illnesses on an individual's education performance, and the effects of therapy as a correctional method for perpetual inmates). For use cases at an individual level (e.g., risk calculators, clinical decision support, clinical trials, modification of a particular record or data element), data users require record level data element data feeds.

Provided below are two examples of potential solutions to decrease data access and use burden for external IT systems and/or data platform users.

1) A potential solution is to develop a combined "top-down" and "bottom-up" federal-level data access management platform. In this model, states and federal agencies would still be the key decisionmakers in creating their data access and use policies, data request applications, data elements, and datasets offered, and they would make the final decision on whether a data user is granted access to the requested dataset. Here, the platform would serve as a mediator between the data custodian (e.g., states or a federal agency) and potential data users, enabling data users to submit, manage, and track their many data access requests in one location. Within the current data access request workflow for NVSS, many data users must monitor their email inboxes to track the acceptance of their data request applications. (NCHS is currently developing an application portal where application status can be tracked online.) While the federal statistical system recently adopted the SAP, which standardizes data request application forms across 13 principal statistical agencies, SAP does not apply to states and their datasets.[vi] There is an opportunity to use SAP as a model to inform or support data governance and access standardization across restricted-use state and local data.
2) Another potential solution is to enable authorized external users to access the IT system and/or data platform but hide PII and protected health information (PHI) by tokenizing those data elements. In this solution, a middle-layer would need to be introduced to the IT system and/or data platform to serve as an encryption layer. By using homomorphic encryption, PII and PHI data elements, and potentially the entire record, will be redacted with tokens, but external data users will be able to conduct computational analysis on the encrypted data without needing to first decrypt the data – this is a trait for homomorphic encryption.[vii] This analysis will result in a format that is not human-readable. To turn the computational analysis results into a human readable format, the external users must obtain the "key" to the homomorphic encryption. In this model, the key to the homomorphic encryption should be provided once the data user completes all necessary agreements (e.g., BAA, data use agreement (DUA), or memoranda of understanding (MOU)). Access to tokenized data would require development of a program with governance and controls to ensure that tokenization of PII and PHI do not risk disclosure (i.e., HIPAA expert determination method).

[vi] NCSES Standard Application Process: https://ncses.nsf.gov/initiatives/standard-application-process

[vii] Alloghani M, Alani MM, Al-Jumeily D, Baker T, Mustafina J, Hussain A, Aljaaf AJ. A systematic review on the status and progress of homomorphic encryption technologies. Journal of Information Security and Applications. 2019 Oct 1;48:102362.

3.2 PRIVACY AND CONFIDENTIALITY RISKS

What are the current and foreseeable risks to privacy and confidentiality in meeting these needs? How could those risks be mitigated while maximizing the utility of interoperable health data for better decision-making?

A current risk to privacy and confidentiality are unintentional bad actors (e.g., employees who use the IT system) who are unaware of cybersecurity data confidentiality and security best practices and malicious bad actors (e.g., hackers) who purposely try to bypass the IT systems security features and firewalls. Organizations can assist unintentional bad actors by providing additional education (e.g., manuals, webinars, toolkits, example scenarios) to ensure that these individuals are aware of the consequences. Additionally, organizations can reduce privacy and confidentiality risks by managing employees access roles through a central data access and governance management platform, so that employees are limited to only the minimum set of data that is required to complete their duties. This reduces an organization's overall exposure to privacy and confidentiality risks.

In cases involving malicious bad actors, organizations should 1) employ best-in-class IT system security features, 2) encrypt their data when the data are in storage using a NIST-approved encryption method (e.g., Triple DES, AES, RSA, CRSYTALS-Kyber, CRSYTALS-Dilithium, SPHINCS+, FALCON)[viii, ix], and 3) employ IT privacy and security best practices. However, given enough computing power, processing power, and the necessary technical skills, malicious bad actors have occasionally been able to bypass stringent IT system defenses. While the majority of current encryption method can be cracked – the one-time pad encryption method is the only "uncrackable" encryption method, though it is not easily implemented[x] – current encryption methods are relatively secure but are still susceptible to the brute force hacking methods employed by malicious bad actors.[xi, xii]

In addition to the mentioned mitigation strategies to limit current and foreseeable risks, participants in the Analytics and Innovators stakeholder workshop posed that data custodians at all levels (e.g., point of data capture, state, and federal) could conduct ethical hacking and "war game"-like activities. These activities could be used to develop new strategic, contingency plans and test current plans against potential current and future scenarios (e.g., foreign, state-backed hackers, weaponization of electromagnetic pulse (EMP) to shut down electrical grids, and attack on upstream service providers such as internet service providers (ISPs)).

[viii] NIST Algorithm Validation Program: https://csrc.nist.gov/projects/cryptographic-algorithm-validation-program

[ix] NIST Cryptographic Standards and Guidelines Development Process: https://csrc.nist.gov/projects/crypto-standards-development-process

[x] Agunbiade A. Big Data Security on Hadoop Open Source Frame for Healthcare Data Management using One-Time-Pad Encryption Algorithm. European Journal of Computer Science and Information Technology. 2024 Mar 6;12(1):68-82.

[xi] Tibbetts J. Quantum computing and cryptography: Analysis, risks, and recommendations for decisionmakers. Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States); 2019 Sep 10.

[xii] NIST Security and Privacy Controls for Information Systems and Organizations https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf

3.3 APPLYING PROJECT FINDINGS

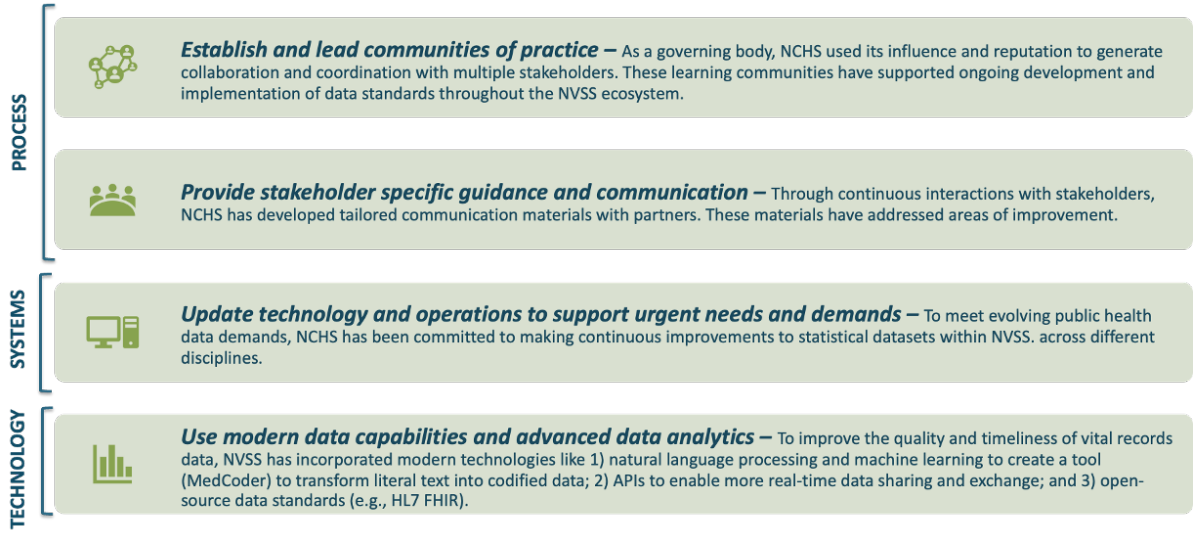
How could this project inform the functions and services of an NSDS?

The NVSS system is a prime model of a statistical system that can be used to derive findings and insights for a potential future NSDS. Several traits and characteristics of the NVSS data ecosystem align with other federal statistical data ecosystems. Examples include:

- NVSS does not own the data – the state owns the data.
- NVSS requires the reporting of data elements for reporting purposes at regular intervals.
- NVSS requires the reporting of a standardized set of data elements (i.e., U.S. Standard Certificate of Birth (2003 Revision) and U.S. Standard Certificate of Death (2003 Revision).
- NVSS is used by both healthcare and non-healthcare stakeholders.
- NVSS data has a wide range of use cases and is used for both healthcare and non-healthcare use cases.

These traits and characteristics enable the NCSES to translate and apply findings and insights from this project to other federal-level statistical data ecosystems. The NSDS demonstration projects shall inform use of "***processes, systems, and technologies*** to protect restricted data, statistical products, privacy, and confidentiality under the Confidential Information Protection and Statistical Efficiency Act of 2018".[xiii] Specific considerations for a potential future NSDS are further described in Section 2, based on NVSS environmental scan findings and insights. Takeaways for a potential future NSDS based upon NVSS processes, systems, and technologies include examples provided in Figure 5.

Figure 5. NVSS Examples to Apply to a Potential Future NSDS



How could this project highlight metadata infrastructure requirements? What are the minimal and ideal metadata required for interoperability?

Metadata are data that provide information on data. This definition can be interpreted in multiple ways, and metadata can be categorized into buckets, such as descriptive metadata (information about the data itself), structural metadata (information about the data's container), administrative metadata (information about the management of the data), reference metadata (information about the contents and quality of statistical data), and statistical metadata (information about the process to create the dataset). Specific examples of metadata that an NSDS could develop best practices guidance for access, exchange, and use of data for evidence building are included below, in Table 2:

[xiii] NCSES Authorizing Legislation: https://ncses.nsf.gov/initiatives/national-secure-data-service-demo/authorizing-legislation

Table 2. Sample Metadata to Consider for Best Practices Guidance

Descriptive Metadata • Dataset submitter • Time of publication • Unique identifier for the dataset • Description of the dataset Structural Metadata • Dataset file type • Dataset location • Key words for finding the dataset Administrative Metadata • Access role tier requirement for accessing • Audit log of number of times the dataset was accessed or downloaded • Audit log of individual who accessed or downloaded the dataset • Creation of net new/duplicate datasets • Audit log of individuals who modified the dataset • Audit log of the times when the dataset was modified	Reference Metadata • Number of rows and columns in the dataset • List of column headers • List of inputted/missing data elements • Number of inputted/missing data elements • Data field type (e.g., integer, string, Boolean) • Code and/or value set for each data field Statistical metadata • Origin of data element • Individual who inputted the data at the point of capture • Time of submission from data source • Role/organization that coded the data • Time at which the data was codified • Tracking of changes in coding practices or versions • Reasons and timings of error corrections or other changes • Dataset sampling plan • Mappings between data standards and value sets • Audit log of techniques used to process data elements (e.g., ETL)

All metadata types described above would apply to NVSS and to both statistical data ecosystems and record-level data ecosystems. In addition to creating IT infrastructure to capture and support the computational analysis of the data elements and datasets themselves, one modernization opportunity is capturing and reporting the data exchange pipeline (e.g., an API). This would provide public transparency and build trust amongst data users. Using APIs as an example, metadata data elements could include the identity of API query requester, frequency of API queries, number of successful/unsuccessful data transactions.

For NVSS specifically, a metadata infrastructure could form a foundation for a more modernized NVSS ecosystem that allows a more flexible approach to standardization. While certain data may be required, NVSS could consider incorporating state-specific data elements and test future or optional data fields. While FHIR and associated implementation guides have provided a strong mechanism for standards-based interoperability, NVSS could collect EHR native codes or obtain the FHIR-mapped or FHIR-native values from source systems like EHRs and case management systems. With FHIR-based exchange, NVSS stakeholders such as state vital records offices and NCHS could understand the source and documenter of a data element and apply analyses and decision support to prioritize data sources and roles. With a strong metadata architecture, NVSS could both collect more standardized as well as more granular or flexible data from multiple data sources and apply NVSS-specific mapping and coding components in its dataset processing. In addition, NCHS could work with the standards development community to promote new data elements and standards and implement a more rapid rollout of enhancements and modifications.

What are best practices for establishing consistent standards across metadata types, datasets, and documentation?

Best practices for establishing consistent standards across metadata types, datasets, and documentation can be categorized into five categories: open standards, incentives, education, testing and conformance, and standardization data standards of vital records data across states and jurisdictions.

OPEN STANDARDS

When establishing a metadata standard, it is important to consider open data standard formats (i.e., data standard formats that have public documentation). A metadata standard is a data standard that establishes guidance that creates an expectation of available data elements and their format (e.g., code/value set, minimum set of data element, and data field type). There are multiple different metadata standards including standards created by the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), and the Internet Engineering Task Force (IETF).

In addition, it is important to consider adopting extensible markup language (XML) as the metadata standard schema of choice, because XML is a commonly used data schema for other prominent healthcare data standards, such as HL7 FHIR. Additionally, XML is both human and machine readable, making it easy for humans to understand. Several data standards exist to support the design and deployment of XML, such as IETF RFC 7303, IETF RFC 3023, and IETF RFC 3470. These standards should be adopted to ensure that the systems built to support the capture of metadata are interoperable with the IT systems in which the data are housed in.

FEDERAL SUPPORT AND SUSTAINABILITY

Federal support could assist the establishment of consistent standards usage for metadata types, datasets, and documentation across the 13 federal principal statistical agencies and between states and jurisdictions. Federal modernization efforts and associated funding could support development of additional data standards to address current gaps in healthcare data exchange and interoperability. Moreover, technical assistance, pilot, and demonstration project funding could encourage implementation of specific technologies or functionalities. Pilot and demonstration projects should consider conducting assessing sustainability to understand the level of effort needed to support newly implemented technologies, functionalities, and infrastructure long-term.

STANDARDIZATION OF DATA ELEMENTS AND EXCHANGE

Standardization of data elements across the federal, state, and jurisdictional levels improve the collection of metadata data elements. Depending on the data element type, certain data elements may inherently require additional metadata data elements to describe the data element stored in the IT system or data platform. By standardizing the data collected from various data sources, custodians should fully understand the metadata they expect to collect. Metadata collection may vary if the dataset within the IT system or data platform is not standardized.

Federal agencies could participate in, and provide technical assistance for, standards development efforts with SDOs (e.g., HL7), support HL7 connectathons and pilot events, and develop educational materials to support standardization and standards adoption.

TECHNICAL ASSISTANCE, TESTING, AND CONFORMANCE

Testing and conformance will ensure that implemented IT systems are compliant with data standards identified by the federal and/or state government. Similar to the testing and development of standards for regular data, there are two methods by which testing and conformance could be accomplished: test-driven and data flow-driven. Traditional standards and systems development focuses on testing whether a standard or system is implemented correctly by determining if a computational action has been completed (e.g., a query, an accepted input, or a FHIR Bulk Data complete message). In contrast, a data flow-driven testing and conformance process would instead monitor the data elements, including the metadata, that are being exchanged in the data exchange mechanism. By monitoring the data flow, data custodians and regulators can proactively identify errors. This approach could test standards and systems as they are in development, rather than at the end. Metadata standards that align with currently implemented IT systems and data platforms are critical when assessing testing and conformance.

EDUCATION

When a metadata standard is selected, it is important to provide education and training for data stakeholders on the selected metadata standard. This is critical as it will ensure that key metadata fields are either input by an individual or automatically input by the IT system or data platform. To ensure that an IT system or data platform is capable of automatically inputting the required data into the metadata fields, IT systems and data platform administrators should review their systems' configurations to ensure that these critical metadata data elements are being filled.

4 CONCLUSION

An environmental scan of the NVSS revealed significant modernization progress, offering insights and models for services, processes, and programs for potential future NSDS consideration. The environmental scan of NVSS yielded key takeaways for consideration in developing a potential future NSDS:

- *Establish and lead communities of practice.* As a governing body, NCHS used its influence and reputation to generate collaboration and coordination with multiple stakeholders. These learning communities have supported ongoing development and implementation of data standards throughout the NVSS ecosystem. A potential future NSDS could consider how to use learning communities to strategically support other core functions.
- *Provide stakeholder specific guidance and communication.* Through continuous interactions with stakeholders, NCHS has developed tailored communication materials with partners. These materials have addressed areas of improvement. A potential future NSDS could consider proactively meeting with stakeholders (or providing support) to identify pain points and collaboratively define next steps.
- *Update technology and operations to support urgent needs and demands.* To meet evolving public health data demands, NCHS has been committed to making continuous improvements to statistical datasets within NVSS. A potential future NSDS could consider structuring data systems to be more agile and timely to respond to urgent data needs across different disciplines.
- *Use modern data capabilities and advanced data analytics.* To improve the quality and timeliness of vital records data, NVSS has incorporated modern technologies like 1) natural language processing (NLP) and artificial intelligence to create a tool (MedCoder) to transform literal text into codified data, 2) application programming interfaces (APIs) to enable more real-time data sharing and exchange, and 3) open-source data standards (e.g., HL7 FHIR). A potential future NSDS could consider investigating and testing new tools and technologies to improve data capabilities and promoting wider adoption across its ecosystem. Leveraging interoperable data can also support the ability to implement more rapid changes to long-standing data collection standards.

By drawing insights from NVSS modernization efforts to date, a potential future NSDS can leverage key success factors and enablers to meet its purpose and core functions to provide federal, state, and local partners the services they will need to securely use data to support evidence-based policymaking and research. The implementation of a potential future NSDS informed by NVSS modernization efforts could consider the following activities and services to support advancement and modernization of datasets for evidence generation and decision-making:

- Navigating and Comparing Statistical Data Sources
- Cataloging and Harmonizing Value Sets for Common Data
- Developing Data Entry Stakeholder Training and Education Resources
- Advancing Skills Through Modern Interoperability Infrastructure
- Streamlining Common Data Access Tiers
- Expanding Standard Application Process (SAP) Awareness
- Modernizing Platforms for Data Access and Analyses
- Participating in Standards Development and Advancement
- Providing Templates for Common Queries and Computational Analyses
- Communicating Best Practices for Rapid Release of Data

clinovations
GOVERNMENT + HEALTH

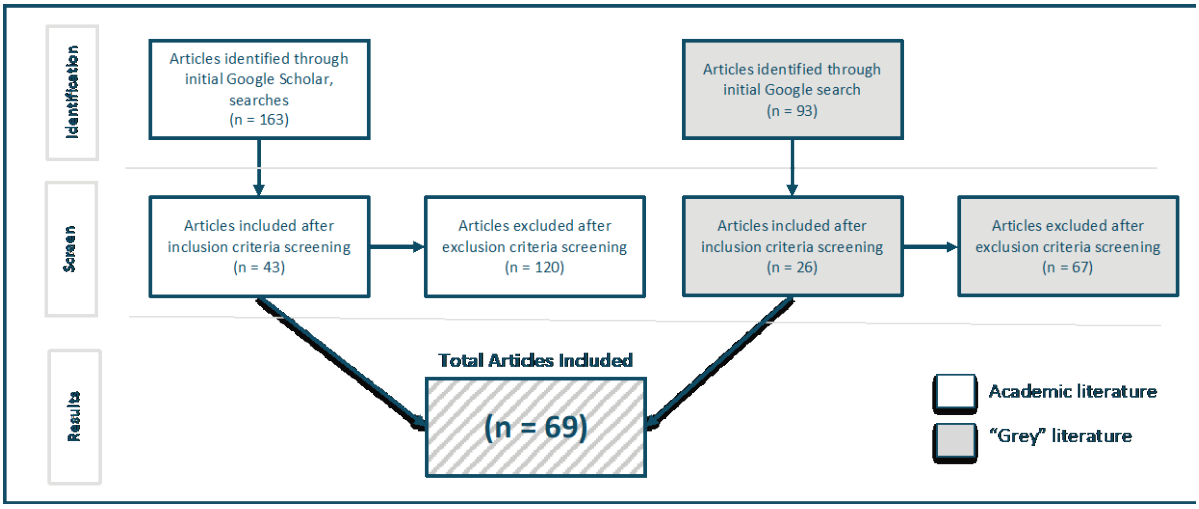
APPENDICES

APPENDIX A - LITERATURE SCAN SUMMARY

Methodology

The methodology of the literature scan involved a review of publications released between January 2019 to December 2023. To survey the landscape of uses of NVSS data, Clinovations gathered peer-reviewed sources from Google Scholar using search queries created by combining search terms. Then, the project team inserted the same search queries into Google to conduct the "grey" literature landscape scan, and this methodology resulted in diverse insights that may not have been discovered in peer-reviewed literature, as is common in use of emerging modernization technologies. Figure 6 depicts the literature scan inclusion process that resulted in review of 69 articles.

Figure 6. Literature Scan Inclusion Process



The primary goal of the literature scan was to identify NVSS data users to participate in stakeholder workshops. This was expanded into a brief literature scan of the NVSS data ecosystem to identify known strengths and challenges. Interoperability and modernization of vital statistics data in peer-reviewed and "grey" literature publications were present in publications with a policymaking goal related to the Evidence Act, emergency response to the COVID-19 pandemic, and the CHIPS and Science Act[xiv]. The literature scan focused on publications from researchers and data users.

Strengths

Scans of both peer-reviewed and "grey" literature revealed that, due to decades of collecting standardized information from birth and death certificates, the NVSS datasets currently provide more complete birth and death data in comparison to other mortality databases. By 2018, all states and the District of Columbia adopted the 2003 edition of the U.S. Standard Certificate of Birth and U.S. Standard Certificate of Death, which has increased data uniformity across different states and counties.

[xiv] Crowley R, Mathew S, Hilden D, Health and Public Policy Committee of the American College of Physicians. Modernizing the United States' public health infrastructure: A position paper from the American College of Physicians. Annals of Internal Medicine. 2023 Aug;176(8):1089-91.

In the literature reviewed, the NVSS dataset was the de facto and definitive dataset for analysis of vital statistics data as it remains the most comprehensive, complete, and publicly available birth and death statistical dataset source. As NVSS data are published for public health surveillance, publicly available, and de-identified, it is exempt from Institutional Review Board (IRB) approval,[xv] making it less burdensome while being the most comprehensive.[xvi,xvii] A majority of the peer-reviewed publications cited this rationale in the methods or ethics section of their publication. However, some researchers obtained IRB approval for their study when using NVSS as the only study sample.[iii] Therefore the literature reviewed did not offer much discussion on rationale for use of NVSS, or insights into the strengths of NVSS data.

NVSS dataset identified within the literature scan are summarized in Table 3.

Table 3. NVSS Dataset Strengths from the Literature

Strengths	Description
Death Data Elements Not Found in Other Public Datasets	Contains death data elements, such as cause of death, that are not found in other datasets (e.g., National Death Index)
High Data Quality Across Data Sources and Jurisdictions	Data are processed in a standardized manner, even though the data source for received datasets are different
More Complete Birth and Death Data	Includes data from all 50 states, DC, New York City, and five United States territories which is more comprehensive than other available datasets
More Timely Birth and Death Data	Offers provisional birth and death datasets for the previous month via CDC Wide-ranging ONline Data for Epidemiologic Research (WONDER), reducing need to wait for final, annual datasets
Resources for Monitoring and Understanding Population Health	NVSS data commonly cited by peer-reviewed articles and data visualizations in public health and population health publications
Advancement and Adoption of Interoperable Data Standards	Data standardized for decades with leadership and support to develop, advance, and use modern health data standards reduces burden
High Engagement from State and Jurisdictions due to Federal Support	Longest established data sharing between states/jurisdictions and the federal government, with multiple types of federal support

Limitations and Challenges

The literature scan revealed common and unique limitations and challenges around the uses of NVSS datasets and within the current NVSS ecosystem, which are summarized in Table 4. Many of the limitations and challenges identified in the literature scan have been addressed, or involve active efforts, by NVSS modernization and improvement projects reviewed in the NVSS modernization efforts scan.

[xv] 45 CFR Part 46

[xvi] Ncube CN, McCormick SM, Badon SE, Riley T, Souter VL. Antepartum and intrapartum stillbirth rates across gestation: a cross-sectional study using the revised foetal death reporting system in the US. BMC Pregnancy and Childbirth. 2022 Nov 29;22(1):885.

[xvii] HHS Regulations for Protection of Human Subjects Research (45 CFR 46) 2018 Common Rule: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html

Table 4. NVSS Dataset Limitations and Challenges

Limitations and Challenges	Description
Bridged-Race Reporting Implications	All states have used the standard birth certificate since 2016 and standard death certificate since 2018 - Data for race and ethnicity prior to these dates required bridging back to the older standard until all states had adopted the new certificates (NCHS stopped releasing bridged race data in 2020)
Data Granularity	Uses data elements specified by the U.S. Standard Certificate of Birth and U.S. Standard Certificate of Death; however, some users seek more granular data (e.g., gestational age and singleton/twin pregnancy) and/or data standards (e.g., race/ethnicity) in public use files, rather than having to use restricted use files
Data Misclassification	May include misclassified (e.g., misreported or miscodified) data, especially when translating text data elements into a coded data element (e.g., text-reporting cause of death into an ICD-10 code). Also, cause of death text reported by the certifier may be deficient resulting in data misclassification
Data Sources and Inconsistencies	Data provided by states are subject to variation by reporting facility and variations in adherence to reporting requirements, resulting in inconsistencies in data collection, report-out, and format
Geographic Disparities	Will not report out data from counties with either a population smaller than 100,000 or fewer than 10 cases in public use files and CDC WONDER. Publications must suppress data with counts less than 10 in a specific category to preserve privacy and confidentiality
Interoperability at the State Level	NVSS data reporting interoperability at the state-level (e.g., state VRO systems) is still evolving, which can delay dataset releases
Missing Data	Data may be missing due to unknown, incomplete, or misreported data from source systems (e.g., point of capture systems like EHRs, state VRO systems)

Future Enhancement Opportunities

The limitations and challenges noted in both the peer-reviewed and "grey" literature highlight specific areas where NVSS and NVSS datasets could be updated, improved, and modernized through use of interoperable and standards-based data capture and exchange.

Table 5. NVSS Dataset Modernization Opportunities

Opportunity	Description
Improved documentation of Certificate of Birth and Certificate of Death data	NVSS data could be improved by 1) reviewing the process by which data are collected at the point of contact, 2) providing education on the importance of accurate data collection to data entry personnel, and 3) reviewing the process for the translation of literal, text-based data into a coded dataset.
Improved classification of race and ethnicity data	Race/ethnicity data can be influenced by other information that is collected within the U.S. Standard Certificate of Birth and U.S. Standard Certificate of Death (e.g., cause of death information). While NCHS stopped releasing bridged race data in 2020, some researchers are looking to analyze data from years that pre-date the latest standard and may still seek bridged race data.
Centralized location for NVSS data and supporting documentation	The "grey" scan revealed multiple data access portals for NVSS data. Examples included: CDC WONDER, CDC WISQARS, NCHS data briefs, and third-party websites. NCHS and NVSS have an opportunity to promote a centralized location from which NVSS data users may access NVSS data and dataset guides to increase use of NVSS data.
Forum to foster a community for NVSS data users	The "grey" literature scan revealed that users of NVSS data utilized online forums (e.g., StackOverflow.com) to ask questions and find answers when using statistical analysis software. NCHS and NVSS have an opportunity to foster an online community for NVSS users who conduct more technical statistical methods using statistical analysis software or other data analytic tools.
Improved interoperability of Vital Records Office commercial systems	NVSS has an opportunity to lead and guide states and jurisdictions in procuring and implementing a full spectrum vital records data system that serves data collection, data exchange, data processing, and data storage purposes.

References – Reviewed Articles

Table 6. Articles Reviewed for Literature Scan

Literature Reference Article	Article Type
1. Hata, J., & Burke, A. (2020). A Systematic Review of Racial and Ethnic Disparities in Maternal Health Outcomes among Asians/Pacific Islanders. Asian/Pacific Island Nursing Journal, 5(3), 139–152. https://doi.org/10.31372%2F20200503.1101	"Grey" Literature
2. Ahmad, F., Anderson, R. N., Knight, K., Rossen, L. M., & Sutton, P. D. (2021). Advancements in the National Vital Statistics System to meet the Real-Time data needs of a pandemic. American Journal of Public Health (1971), 111(12), 2133–2140. https://doi.org/10.2105/ajph.2021.306519	Peer-Reviewed
3. Nair, D. R., Chauhan, A., & Vaidya, D. (2022). Are US Asian Indians Dying with Atherosclerosis More Likely to have Concurrent Diabetes Mellitus: Analysis of National Multiple Cause of Mortality Data (2012 - 2019). Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-1894160/v1	Peer-Reviewed
4. Kleinman, R. A., & Weiss, R. D. (2022). Benzodiazepine-Involved overdose deaths in the USA: 2000–2019. Journal of General Internal Medicine, 37(8), 2103–2109. https://doi.org/10.1007/s11606-021-07035-6	Peer-Reviewed
5. Dunietz, G. L., Holzman, C., Zhang, Y., Li, C., Todem, D., Boulet, S. L., McKane, P., Kissin, D. M., Copeland, G., Bernson, D., & Diamond, M. P. (2017). Assisted reproduction and risk of preterm birth in singletons by infertility diagnoses and treatment modalities: a population-based study. Journal of Assisted Reproduction and Genetics, 34(11), 1529–1535. https://doi.org/10.1007/s10815-017-1003-6	"Grey" Literature
6. Golabi, P., Paik, J., Eberly, K., De Avila, L., Alqahtani, S. A., & Younossi, Z. M. (2022). Causes of death in patients with Non-alcoholic Fatty Liver Disease (NAFLD), alcoholic liver disease and chronic viral Hepatitis B and C. Annals of Hepatology, 27(1), 100556. https://doi.org/10.1016/j.aohep.2021.100556	Peer-Reviewed
7. Hedegaard, H., Miniño, A. M., Spencer, M. R., & Warner, M. (2021). Drug overdose deaths in the United States, 1999–2020. https://doi.org/10.15620/cdc:112340	Peer-Reviewed
8. Report on the environment | US EPA. (n.d.). US EPA. https://cfpub.epa.gov/roe/indicator.cfm?i=72	"Grey" Literature
9. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK. Births: Final data for 2018. National Vital Statistics Reports; vol 68, no 13. Hyattsville, MD: National Center for Health Statistics. 2019.	"Grey" Literature
10. Osterman, M. J. K., Hamilton, B. E., Martin, J. A., Driscoll, A. K., & Valenzuela, C. (2021). Births: Final data for 2020. https://doi.org/10.15620/cdc:112078	"Grey" Literature
11. JK, M., E, B. H., Joyce, A., Driscoll, A., & Valenzuela, C. (2023). Births: Final data for 2021. https://doi.org/10.15620/cdc:122047	"Grey" Literature
12. Hedegaard, H., Miniño, A. M., & Warner, M. (2022). Drug overdose deaths in the United States, 2001–2021. https://doi.org/10.15620/cdc:122556	Peer-Reviewed
13. Drake, K. M., & Ruhm, C. J. (2023). Estimating drug involvement in fatal overdoses with incomplete information. American Journal of Preventive Medicine, 65(6), 1172–1180. https://doi.org/10.1016/j.amepre.2023.06.019	Peer-Reviewed
14. Kochanek, K. D., Murphy, S. L., Xu, J., & Arias, E. (2023). Deaths: Final data for 2020. https://doi.org/10.15620/cdc:131355	"Grey" Literature
15. Stevens, B. R., & Ashley, W. S. (2022). Fatal Weather-Related carbon monoxide poisonings in the United States. Weather, Climate, and Society (Print), 14(2), 373–386. https://doi.org/10.1175/wcas-d-21-0130.1	Peer-Reviewed
16. Milam, A. J., Furr-Holden, D., Wang, L., & Simon, K. M. (2021). Health data disparities in Opioid-Involved overdose deaths from 1999 to 2018 in the United States. American Journal of Public Health (1971), 111(9), 1627–1635. https://doi.org/10.2105/ajph.2021.306322	Peer-Reviewed
17. Hedegaard, H., Curtin, S. C., & Warner, M. (2020, April 1). Increase in suicide mortality in the United States, 1999–2018. https://stacks.cdc.gov/view/cdc/86670	Peer-Reviewed

Literature Reference Article	Article Type
18. Lynn, M. M., Salemi, J. L., Kostelyna, S. P., Morris, S. A., Tejtel, S. K. S., & Lopez, K. N. (2022). Lesion-Specific congenital heart disease mortality trends in children: 1999 to 2017. Pediatrics, 150(4). https://doi.org/10.1542/peds.2022-056294	Peer-Reviewed
19. Economic Analysis for the Proposed Per- and Polyfluoroalkyl Substances National Primary Drinking Water Regulation | US EPA. (2023, March). US EPA https://www.epa.gov/system/files/documents/2023-03/Proposed%20PFAS%20NPDWR%20EA_final_03_09_2023_0.pdf	"Grey" Literature
20. Powell, D. (2023). Educational attainment and US drug overdose deaths. JAMA Health Forum, 4(10), e233274. https://doi.org/10.1001/jamahealthforum.2023.3274	"Grey" Literature
21. Liang, D., & Shi, Y. (2019). Prescription drug monitoring programs and drug overdose deaths involving benzodiazepines and prescription opioids. Drug and Alcohol Review, 38(5), 494–502. https://doi.org/10.1111/dar.12959	Peer-Reviewed
22. Hedegaard, H., & Warner, M. (2021). NVSR 70-13. Evaluating the cause-of-death information needed for estimating the burden of injury mortality: United States, 2019. https://doi.org/10.15620/cdc:110638	"Grey" Literature
23. Garnett, M. F., Curtin, S. C., & Stone, D. M. (2022). Suicide mortality in the United States, 2000–2020. https://doi.org/10.15620/cdc:114217	Peer-Reviewed
24. Research Guides: A social Work Research Guide: Free Datasets subpage. (n.d.). https://libguides.wustl.edu/social-work/datasetssubpage	"Grey" Literature
25. Hedegaard, H., & Spencer, M. R. (2021). Urban–Rural Differences in drug Overdose death Rates, 1999–2019. https://doi.org/10.15620/cdc:102891	Peer-Reviewed
26. Garnett, M. F., Spencer, M. R., & Hedegaard, H. (2021). Disparities in stressful life events among children aged 5–17 years: United States, 2019. https://doi.org/10.15620/cdc:110040	Peer-Reviewed
27. Lopez, K. N., Morris, S. A., Tejtel, S. K. S., Espaillat, A., & Salemi, J. L. (2020). US mortality attributable to congenital heart disease across the lifespan from 1999 through 2017 exposes persistent Racial/Ethnic disparities. Circulation (New York, N.Y.), 142(12), 1132–1147. https://doi.org/10.1161/circulationaha.120.046822	Peer-Reviewed
28. How do I import a file into r with extension .DUSMCPUB? (n.d.). Stack Overflow. https://stackoverflow.com/questions/70657014/how-do-i-import-a-file-into-r-with-extension-dusmcpub	"Grey" Literature
29. Ncube, C. N., McCormick, S., Badon, S. E., Riley, T., & Souter, V. (2022). Antepartum and intrapartum stillbirth rates across gestation: a cross-sectional study using the revised foetal death reporting system in the U.S. BMC Pregnancy and Childbirth (Online), 22(1). https://doi.org/10.1186/s12884-022-05185-x	Peer-Reviewed
30. Burchard, J., Saade, G. R., Boggess, K., Markenson, G., Iams, J. D., Coonrod, D. V., Pereira, L., Hoffman, M., Polpitiya, A. D., Treacy, R., Fox, A., Randolph, T. L., Fleischer, T. C., Dufford, M. T., Garite, T. J., Critchfield, G. C., Boniface, J. J., & Kearney, P. (2022). Better Estimation of Spontaneous Preterm Birth Prediction Performance through Improved Gestational Age Dating. Journal of Clinical Medicine, 11(10), 2885. https://doi.org/10.3390/jcm11102885	Peer-Reviewed
31. K, M. O. J. (2022). Changes in primary and repeat cesarean delivery: United States 2016-2021. https://doi.org/10.15620/cdc:117432	Peer-Reviewed
32. Greenberg, D. S., Khandwala, Y. S., Lü, Y., Stevenson, D. K., Shaw, G. M., & Eisenberg, M. L. (2019). Disease burden in offspring is associated with changing paternal demographics in the United States. Andrology (Print), 8(2), 342–347. https://doi.org/10.1111/andr.12700	Peer-Reviewed
33. Tahmasebifard, N., Briley, P. M., Ellis, C., & Perry, J. L. (2021). Early Nutrition among Infants Admitted to the NICU with Cleft Lip and Palate. the Cleft Palate-craniofacial Journal (Print), 60(3), 299–305. https://doi.org/10.1177/10556656211059371	Peer-Reviewed
34. Ely, D. M., & Driscoll, A. K. (2022b). Infant mortality in the United States, 2020: Data from the Period Linked Birth/Infant Death File. https://doi.org/10.15620/cdc:120700	"Grey" Literature
35. Ely, D. M., & Driscoll, A. K. (2023). Infant mortality in the United States, 2021: Data from the Period Linked Birth/Infant Death File. https://doi.org/10.15620/cdc:131356	"Grey" Literature

Literature Reference Article	Article Type
36. Brown, C. M., Moore, J. E., Felix, H. C., Stewart, M. K., & Tilford, J. M. (2020). Geographic Hotspots for Low Birthweight: An analysis of counties with persistently high rates. Inquiry (Chicago), 57, 004695802095099. https://doi.org/10.1177/0046958020950999	Peer-Reviewed
37. Wang, M. C., Shah, N. S., Petito, L. C., Gunderson, E. P., Grobman, W. A., O'Brien, M. J., & Khan, S. S. (2021). Gestational Diabetes and Overweight/Obesity: Analysis of nulliparous women in the U.S., 2011–2019. American Journal of Preventive Medicine, 61(6), 863–871. https://doi.org/10.1016/j.amepre.2021.05.036	Peer-Reviewed
38. Yusuf, K. K., Dongarwar, D., Maiyegun, S. O., Ikedionwu, C., Ibrahimi, S., & Salihu, H. M. (2021). Impact of maternal age on the Foreign-Born Paradox. Journal of Immigrant and Minority Health, 23(6), 1198–1205. https://doi.org/10.1007/s10903-021-01157-z	Peer-Reviewed
39. Dongarwar, D., Maiyegun, S. O., Yusuf, K. K., Ibrahimi, S., Ikedionwu, C., & Salihu, H. M. (2021). Incidence and risk of stillbirth among various Asian-American subgroups. Journal of Maternal-fetal & Neonatal Medicine, 35(25), 6638–6643. https://doi.org/10.1080/14767058.2021.1918669	Peer-Reviewed
40. Rauscher, E., & Song, H. (2022). Learning to Value Girls: Balanced infant sex Ratios at Higher Parental Education in the United States, 1969–2018. Demography, 59(3), 1143–1171. https://doi.org/10.1215/00703370-9968420	Peer-Reviewed
41. Obsekov, V., Kahn, L. G., & Trasande, L. (2022). Leveraging systematic reviews to explore disease burden and costs of per- and polyfluoroalkyl substance exposures in the United States. Exposure and Health (Print), 15(2), 373–394. https://doi.org/10.1007/s12403-022-00496-y	Peer-Reviewed
42. Dongarwar, D., Mercado-Evans, V., Adu-Gyamfi, S., Laracuente, M., & Salihu, H. M. (2022). Racial/ethnic disparities in infertility treatment utilization in the US, 2011–2019. Systems Biology in Reproductive Medicine, 68(3), 180–189. https://doi.org/10.1080/19396368.2022.2038718	Peer-Reviewed
43. Jackson, K., Hamad, R., Karasek, D., & White, J. S. (2023). Sugar-Sweetened Beverage Taxes and Perinatal Health: A Quasi-Experimental Study. American Journal of Preventive Medicine (Print), 65(3), 366–376. https://doi.org/10.1016/j.amepre.2023.03.016	Peer-Reviewed
44. Marazzi, M., Miloucheva, B., & Bobonis, G. J. (2022). Mortality of Puerto Ricans in the USA post Hurricane Maria: an interrupted time series analysis. BMJ Open, 12(8), e058315. https://doi.org/10.1136/bmjopen-2021-058315	"Grey" Literature
45. NCHS Stillbirth Activities | CDC. (2022, October 4). Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/stillbirth/activities-nchs.html	"Grey" Literature
46. NCHS - Age-adjusted death Rates for selected major causes of death | HealthData.gov. (2021, February 25). https://healthdata.gov/dataset/NCHS-Age-adjusted-Death-Rates-for-Selected-Major-C/67cg-fthd/about_data	"Grey" Literature
47. NCHS - Death rates and life expectancy at birth - Catalog. (2022, April 21). https://catalog.data.gov/dataset/nchs-death-rates-and-life-expectancy-at-birth	"Grey" Literature
48. NCHS - Teen birth rates for age group 15-19 in the United States by county | HealthData.gov. (2021, February 25). https://healthdata.gov/dataset/NCHS-Teen-Birth-Rates-for-Age-Group-15-19-in-the-U/v3q8-qefr/about_data	"Grey" Literature
49. NCHS - Top five leading causes of death: United States, 1990, 1950, 2000 - catalog. (2022, April 21). https://catalog.data.gov/dataset/nchs-top-five-leading-causes-of-death-united-states-1990-1950-2000	"Grey" Literature
50. Paik, J., Golabi, P., Biswas, R., Alqahtani, S. A., Venkatesan, C., & Younossi, Z. M. (2020). Nonalcoholic Fatty Liver Disease and Alcoholic Liver Disease are Major Drivers of Liver Mortality in the United States. Hepatology Communications, 4(6), 890–903. https://doi.org/10.1002/hep4.1510	"Grey" Literature
51. PFAS National Primary Drinking Water Regulation Rulemaking. (2023, March 29). US EPA. https://www.govinfo.gov/content/pkg/FR-2023-03-29/pdf/2023-05471.pdf	"Grey" Literature

Literature Reference Article	Article Type
52. Bruno, A. M., Allshouse, A. A., Metz, T. D., & Theilen, L. (2022). Trends in hypertensive disorders of pregnancy in the U.S. from 1989 to 2018. American Journal of Obstetrics and Gynecology (Print), 226(1), S495–S496. https://doi.org/10.1016/j.ajog.2021.11.819	Peer-Reviewed
53. Ely, D. M., & Driscoll, A. K. (2022). Infant mortality in the United States, 2020: Data from the Period Linked Birth/Infant Death File. https://doi.org/10.15620/cdc:120700	Peer-Reviewed
54. Xie, L., Boudreaux, M., & Franzini, L. (2021). Maryland's global Budget Revenue program. Medical Care, 59(8), 663–670. https://doi.org/10.1097/mlr.0000000000001534	Peer-Reviewed
55. RDC - Restricted Data. (n.d.). https://www.cdc.gov/rdc/b1datatype/Dt122.htm	"Grey" Literature
56. Cook, A., & Stype, A. (2020). Medicaid expansion and infant mortality: the (questionable) impact of the Affordable Care Act. Journal of Epidemiology and Community Health, jech-213666. https://doi.org/10.1136/jech-2019-213666	Peer-Reviewed
57. Shapiro-Mendoza, C. K., Woodworth, K. R., Cottengim, C., Lambert, A. B. E., Harvey, E., Monsour, M., Parks, S. E., & Barfield, W. D. (2023). Sudden unexpected infant deaths: 2015–2020. Pediatrics (Evanston), 151(4). https://doi.org/10.1542/peds.2022-058820	Peer-Reviewed
58. State of California Budget Change Proposal - Cover Sheet DF-46 (REV 10/20). (2022, April 8). State of California. https://esd.dof.ca.gov/Documents/bcp/2223/FY2223_ORG4265_BCP5863.pdf	"Grey" Literature
59. Moore, M. D., Brisendine, A. E., & Wingate, M. S. (2020). Infant Mortality among Adolescent Mothers in the United States: A 5-Year Analysis of Racial and Ethnic Disparities. American Journal of Perinatology, 39(02), 180–188. https://doi.org/10.1055/s-0040-1714678	Peer-Reviewed
60. Ikedionwu, C., Dongarwar, D., Yusuf, K. K., Ibrahimi, S., Salinas-Miranda, A., & Salihu, H. M. (2020). Pre-pregnancy maternal obesity, macrosomia, and risk of stillbirth: A population-based study. European Journal of Obstetrics, Gynecology, and Reproductive Biology, 252, 1–6. https://doi.org/10.1016/j.ejogrb.2020.06.004	Peer-Reviewed
61. Dongarwar, D., & Salihu, H. M. (2020). Risk of Stillbirth after Infertility Treatment in the United States: 2014-2017. International Journal of MCH and AIDS, 9(1), 149–152. https://doi.org/10.21106/ijma.345	Peer-Reviewed
62. Maiyegun, S. O., Yusuf, K. K., Dongarwar, D., Ibrahimi, S., Ikedionwu, C., & Salihu, H. M. (2021). Risk of stillbirth among Foreign-Born Mothers in the United States. Journal of Immigrant and Minority Health, 24(2), 318–326. https://doi.org/10.1007/s10903-021-01164-0	Peer-Reviewed
63. Carney, M. H. (2021). The impact of mental health parity laws on birth outcomes. Health Economics (Print), 30(4), 748–765. https://doi.org/10.1002/hec.4217	Peer-Reviewed
64. Bruno, A. M., Allshouse, A. A., Metz, T. D., & Theilen, L. (2022). Trends in hypertensive disorders of pregnancy in the U.S. from 1989 to 2018. American Journal of Obstetrics and Gynecology (Print), 226(1), S495–S496. https://doi.org/10.1016/j.ajog.2021.11.819	Peer-Reviewed
65. Adair, T., Temple, J., Anstey, K. J., & López, A. D. (2022). Is the Rise in Reported Dementia Mortality Real? Analysis of Multiple-Cause-of-Death Data for Australia and the United States. American Journal of Epidemiology, 191(7), 1270–1279. https://doi.org/10.1093/aje/kwac047	Peer-Reviewed
66. Winant, C., Alexander, M., Dharamshi, D., Barbieri, M. (2021). METHODS PROTOCOL FOR THE UNITED STATES MORTALITY COUNTY DATABASE (Online). https://usa.mortality.org/uploads/counties/USCountyBayesianEstimationMethodsProtocol20210927.pdf	Peer-Reviewed
67. Yu, V. Y. H., Ramsay, J. M., Horns, J. J., Mumford, S. L., Bruno, A. M., & Hotaling, J. M. (2023). The association between parental age differences and perinatal outcomes. Human Reproduction (Oxford. Print). https://doi.org/10.1093/humrep/dead236	Peer-Reviewed
68. Lopez, K. N., Morris, S. A., Tejtel, K. S., Espaillat, A., & Salemi, J. L. (2020). US Mortality Due to Congenital Heart Disease Across the Lifespan from 1999-2017 Exposes Persistent Racial/Ethnic Disparities. medRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2020.03.15.20036525	"Grey" Literature
69. What's new – CARES HQ. (n.d.). https://careshq.org/whats-new/	"Grey" Literature

APPENDIX B - FEDERAL STAKEHOLDER INTERVIEWS SUMMARY

NCHS Interview Summary

NCHS INTERVIEW DISCUSSION TOPICS

Clinovations conducted interviews with six NCHS team members. The purpose of these interviews was to review the project goals and objectives and obtain insight into vitals data and NVSS modernization efforts, ongoing challenges, and planned future activities. DVS, a division within NCHS, is responsible for the oversight and management of the NVSS and the National Death Index (NDI). It comprises of four branches.

- Data Acquisition, Classification, and Evaluation Branch (DACEB): DACEB is responsible for acquiring the datasets from the data source (i.e., state VROs) and analyze collected datasets for data quality and data missingness.
- Information Technology Branch (ITB) Data Governance: ITB's charge is to conduct research into the design, development, and administration of vital statistics information technology systems, including the processing of vital records data.
- Partner Engagement and Data Dissemination Branch (PEDDB): PEDDB is responsible for the management of the NDI and supports the dissemination of vital statistics.
- Statistical Analysis and Surveillance Branch (SASB): SASB's charge is to establish the research agenda for natality, mortality, and fetal death statistics.

For NCHS interviews, discussion topics included:

- *Sources:* Discuss characteristics of data sources and received data (e.g., complete, timely, and standards).
- *Acquisition and Transformation:* Discuss extraction, translation, and cleaning procedures.
- *Formatting and Processing:* Discuss opportunities to enhance value sets, preprocessing, and data linkages.
- *Systems and Infrastructure:* Discuss the NVSS infrastructure, maintenance, and challenges.
- *Governance and Management:* How do you manage access of outputs?
- *Analytics and Outputs:* Describe the tools/statistical packages used.
- *Technical Assistance:* What technical capabilities are needed to use modern tools and methods?
- *Current and Future Opportunities:* Discuss future goals and requests from NVSS stakeholders and users.

This section summarizes discussion findings and insights related to the NVSS data source, data processing, data governance, current uses, and modernization opportunities.

DATA SOURCES

When discussing NVSS data sources, participants emphasized that the vital records data are not owned by DVS but rather each state and jurisdiction owns their own data. DVS has an agreement with all 57 states and jurisdictions, enabling them to use vital records data from each state and jurisdiction to create aggregated, statistical datasets. This agreement is known as the collective agreement. DVS must renew the collective agreement with all 57 states and jurisdictions every five years. DVS has the capability to request amendments to the collective agreement, such as adding or removing vital records data elements. However, all 57 states and jurisdictions must agree to the proposed amendments. Per the collective agreement, the DVS and the 57 state and jurisdictional VROs communicate to ensure that the other party is aware if any changes are made.

While NCHS collects data at the state and jurisdictional level, one participant acknowledged the interest in data and insights from the point of data capture. One participant stated that, currently, over 99 percent of all natality data comes from the birthing hospital. Most death data come from the medical examiner, coroner, and funeral home case management systems or hospital systems. Given that these systems have some federally or state coordinated requirements, with certified EHRs serving hospital systems and ambulatory clinics, and case management systems serving medical examiners and coroners that are procured by state and local jurisdictions, there may be an opportunity to advance interoperability to collect data from these point of capture data systems.

CDC and NAPHSIS developed the State and Territorial Exchange of Vital Elements (STEVE) to revolutionize the timeliness of data received from states and jurisdictions. STEVE serves as a routing network for inter-state and inter-jurisdictional data exchange and to enable states and jurisdictions to exchange vital records data with NVSS in a quick and secure manner. Previously, the delay for birth data was up to 30 days and up to 60 days for death data. With STEVE, the delay for both birth and death data are one day. With the advancement of modern API-based capabilities, such as FHIR, there may be more opportunities to directly exchange data with data sources, states/jurisdictions, and other intermediaries and partners.

DATA PROCESSING

In addition to the overseeing and managing of data acquisition and collection effort, NCHS is responsible for analyzing the collected dataset for missing data elements and to follow up with each state or jurisdiction to obtain data collection information, discuss current trends for rate of missing data, and identify data collection barriers and challenges.

If there is significant missing data or missing data elements in the dataset, NVSS will delay the release of the published dataset and statistical data file. To resolve the issue, NCHS communicates with the state or jurisdiction to understand the cause for the missing data elements and to identify a solution to issues that prevented a complete dataset.

Two common reasons for missing data elements are if the dataset is still being processed, or if the dataset was never sent. Occasionally, missing data elements are due to state or jurisdictional policies and/or regulation changes. However, another participant noted that state and jurisdictional laws and regulations may improve the timeliness of obtaining data from state VROs due to time limits of reporting the data to either the state or jurisdictional-level or the federal-level.

Vital records datasets from state and jurisdiction VROs are typically received in a standardized manner, although data entry, coding, and logic errors may occur. NCHS notifies the state or jurisdiction who owns the data when errors are found.

In conjunction with their data quality check, NCHS also provides an automated ICD-10 coding service for the cause of death data element. NVSS receives both literal (free text) and structured data in the datasets that they receive. One example of a free text field is the cause of death data element. While some states have coding capabilities and nosologists on staff to code the data from literal text into ICD-10 codes, many states send their cause of death information in a free-text field. Seeking faster codification turnaround times, ITB developed the tool MedCoder, which enables DACEB to transform data elements from 140 to 240 records in 15 minutes. One participant noted that, currently 85 percent of cause of death transformation is performed using MedCoder. MedCoder is a prime example of machine learning and natural language processing in real world use, to reduce burden on data stakeholders while improving data timeliness.

STANDARDS ADOPTION AND USE

NCHS has adopted HL7 FHIR Release 4 (R4) and developed a FHIR R4 NVSS API (Nightingale) to support state and jurisdiction NVSS data exchange. State and jurisdiction VROs can now submit mortality data directly to the NCHS and receive acknowledgments, errors, and coded data in response. The states noted that there are now some states and jurisdictions who are submitting their mortality data via the NVSS API. Although data exchange via APIs can make data timelier, some states will continue to use STEVE to submit mortality data to NCHS. NCHS will need to map data elements to a common data standard once the data is received from states and jurisdictions.

While NCHS has successfully adopted APIs and HL7 FHIR, many states are budget constrained, thus limiting ability to adopt technologies that can support these modern capabilities. Additionally, many states and jurisdictions require assistance in staffing, workforce development, and organizational support to implement APIs and HL7 FHIR. NCHS has worked in partnership with states and jurisdictions to provide resources and support and could expand training and support to states as they develop API and HL7 FHIR capabilities.

INFORMATION TECHNOLOGY ARCHITECTURE

The ITB is responsible for management of NVSS IT infrastructure and data processing capabilities. This branch also develops, implements, monitors, and enforces policies and models for data governance and access tiers. The ITB has responsibility to:

- Develop technologies, data architectures, security infrastructure, and database management for vital records data,
- Provide consultation and expert technical assistance on database technologies (e.g., SQL server, web services, networking applications),
- Provide consultation, policy guidance, and expert technical assistance to federal stakeholders for vital statistics systems design, administration, and usage,
- Prepare and maintain population databases, and
- Manage national vital statistics data files and databases.

STATISTICAL DATASET GENERATION

NCHS assesses disclosure risk and develops optimal data release strategies to maximize the utility of vital statistics data for policy analysis and decision-making purposes. While NCHS receives and processes PII, the NVSS statistical systems and datasets do not include or release any PII.

NCHS statistical analysis and surveillance teams bridge the gap between record-level data and syndromic surveillance data. User and public health data needs are translated into a potential mode for NVSS data ecosystem stakeholders to obtain, evaluate, analyze, and disseminate natality, mortality, and fetal death statistics data. This is accomplished by its charge to:

- Provide nosological assistance and training for the ICD code set,
- Develop and conduct training activities related to the collection, production, use and interpretation of natality, mortality, and fetal death statistics,
- Develop and conduct training activities for cause of death coding, and
- Provide technical assistance.

The research agenda for natality, mortality and fetal death statistics is established in response to national public health priorities. An example is NCHS collaboration with the CDC Flu Division. Previously, the CDC Flu Division only had data access to the flu and pneumonia death database, which was data gathered from a separate project from 122 cities. However, this dataset was not integrated with NVSS mortality data. In 2017, SASB assisted the CDC Flu Division with integrating the flu and pneumonia death database with the NVSS data ecosystem. NCHS developed and provides subject matter expertise for four surveillance dashboards: Flu View, Respiratory Virus Surveillance, Drug Overdose Death Surveillance, and COVID-19.

While many data elements provided by states and jurisdictions are discrete data elements, a subset of the data fields are open-text, string fields. These data must be parsed so that nosologists can identify the correct cause of death code. For example, for a case related to drug mortality, the spelling of drugs (i.e., official or street term for a drug) can make it more challenging to analyze the cause of death data element.

DATASET LINKING CONSIDERATIONS

The NDI offers a dataset linkage service via a database with identifiable record-level mortality data to data users at a cost. NDI contains death data elements such as the names of the state in which those deaths occurred, dates of death, corresponding death certificate numbers, and cause(s) of death. When data linkage activities are conducted, privacy and confidentiality remain at the forefront. While there is value added by linking disparate datasets, data linkage activities require a workforce that is knowledgeable about both the NDI dataset and the record level NVSS mortality dataset that is received from state and jurisdiction VROs. Expansion of dataset linking would require the development of training programs, education for data users on data linkage activities, and an increase in technical capacity and capabilities.

MODERNIZATION OPPORTUNITIES

NVSS modernization opportunities for consideration noted by NCHS interview participants are summarized below:

Modernization Opportunities Identified

1) Explore solutions and technologies that support linking, data use rights, and privacy/security considerations across states
2) Expand data collection training program and modules to ameliorate impact of high turnover rates of data entry and VRO personnel
3) Improve interoperability between NVSS data ecosystem stakeholders (e.g., states, jurisdictions, and NCHS)
4) Provide technical assistance to states and jurisdictions in purchasing, implementing, and adopting IT systems
5) Apply experience in modern computing (e.g., machine learning) to automate transformation of text to codified data for additional data elements
6) Expand use of APIs for data sharing from additional data sources

Federal Users of NVSS Data Interview Topics

Clinovations conducted group interviews with a total of 23 individuals (external to NCHS) that use NVSS datasets. The project team invited participants to share their experiences in accessing, processing, and using vital statistics data from the NVSS data ecosystem and provide perspectives on future needs and modernization opportunities. Individuals participating in interviews represented the following agencies, offices, and divisions:

- Centers for Disease Control and Prevention (CDC)
  - Agency for Toxic Substances and Disease Registry (ASTDR)
  - National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP)
  - National Center for Infectious Diseases (NCID)
  - National Center for Injury Prevention and Control (NCIPC)
- Food and Drug Administration (FDA)
  - Center for Drug Evaluation and Research (CDER)
- Indian Health Service (IHS)
- National Institutes of Health (NIH)
  - National Institute on Aging (NIA)
  - National Institute of Mental Health (NIMH)
  - Office of Behavioral and Social Sciences Research (OBSSR)

Interview topics included:

- ***Current Use Cases:*** *How do you use vital statistics information?*

- *Data Access, Format, Interoperability, and Processing:* *What works well? What can be improved?*
- *Data Governance and Access:* *How is data access and use managed?*
- *Current Modernization Opportunities and Future Use Cases:* *Discuss future goals, needed skills and tools.*

Participants from federal agencies offered perspectives on obtaining NVSS datasets and their current use within their own federal agency. In some use cases, participants from federal agencies may remove pre-existing data elements from a data field and repopulate it with data elements from their own database, given that the agency has permission from NCHS.

Some participants discussed obtaining record-level datasets from states and jurisdictions. Many participants noted the need for record-level data. Because NVSS does not release record-level datasets – NVSS only releases aggregated, statistical datasets – vital records data users must either use NDI datasets or request access to data from each state.

One participant noted that each of the 57 states and jurisdictions has their own process and data release policy. Standardizing the data request application form and process is a modernization opportunity, and it would reduce data user burden and time. However, this would require coordination and alignment across states and jurisdictions. A potential future NSDS could offer best practices and recommendations for a common state-level agreement that could be used by states and jurisdictions.

Two participants noted that they occasionally cannot obtain datasets they require due to the data governance policies underlying those datasets. One of the participants noted that their organization cannot use SSA's Death Master File dataset due to the underlying data access and use policies. Both participants noted their challenges and burden to obtain data from the 57 states and jurisdictions due to a lack of a common agreement. Participants noted that NVSS dataset data governance and data use policies can be restrictive and limit the adoption or use of modern technologies (e.g., storage of restricted use datasets in cloud-based architectures and data linking technologies).

Federal agencies use various data portals to access NVSS datasets: CDC WONDER and CDC WISQARS were the two most frequently mentioned online data portals. Federal agencies also reported using the CDC State Unintentional Drug Overdose Reporting System (SUDORS) data portal, the Substance Abuse and Mental Health Services Administration (SAMHSA) Public-use Data Analysis System (PDAS), and the SAMHSA Restricted-use Data Analysis System (RDAS).

MODERNIZATION OPPORTUNITIES

NVSS modernization opportunities for consideration noted by federal stakeholder (non-NCHS) interview participants are summarized below.

Modernization Opportunities Identified

1) Develop a specialized access tier for federal "superusers" that support extended access timelines for ongoing and recurring dataset requests
2) Investigate ability to securely link identifiable data to other federal datasets and state-level datasets
3) Support streamlining of data request and application process across states and jurisdictions
4) Expand standard birth and death data elements to include data needed for research and surveillance
5) Develop data standard mappings to support data partner needs – some data users need different value sets or seek levels of granularity not available within ICD-10
6) Update data collection workflows to delay processing of records until all documents and tests are received from ambulatory departments – these documents and tests may influence the codification or entry of specific data elements, such as cause of death

clinovations
GOVERNMENT + HEALTH

APPENDIX C - NATIONAL VITAL STATISTICS SYSTEM MODERNIZATION SCAN SUMMARY

NVSS Modernization Scan Objectives

Clinovations conducted a brief scan of NVSS modernization and interoperability activities to date. Publicly available NVSS related modernization information available on CDC and partner websites was reviewed. Previously, NCHS has categorized modernization activities by data type (e.g., modernizing fetal data and modernizing death data). To obtain insight on possible considerations for a potential future NSDS, activities were classified based on the type of modernization activity, interoperability stage, and impact.

Many of these activities have evolved over time and may no longer operate as distinct modernization projects, as some projects have changed in scope and others have merged or folded into other initiatives. The main purpose of the scan of documented activities is not to produce a definitive catalog, but rather to assess the types of modernization activities and approaches that could inform modernization of other federal statistical research data and other data assets used for evidence-building.

Review of NVSS Modernization Efforts

The NVSS ecosystem is composed of many stakeholders, including data enterers, state VROs, NAPHSIS, CDC/NCHS Division of Vital Statistics (CDC/NCHS DVS), and data users.

Data enterers, such as birth clerks, labor and delivery nurses, physicians, medical examiners, coroners, and funeral home directors, are responsible for capturing vital records data. Once the data have been collected, they are transferred to state VROs, which are government institutions responsible for filing, issuing, preserving, protecting, and amending copies of viral records. NAPHSIS is a nonprofit organization that represents the U.S.'s vital records and public health statistics offices and operates technology infrastructure for routing vital records from state VROs to the CDC. NAPHSIS also serves as a membership organization for jurisdictions and provides various support services and resources to its membership. Collectively, state VROs and NAPSHIS are referred to as "data custodians" in this report to reflect their roles in collecting, storing, and routing data to the CDC.

CDC/NCHS DVS collects, processes, and analyzes data received from jurisdictions and is responsible for disseminating timely, relevant, and accurate health data and statistics. DVS performs data processing functions that can involve actions to clean the data for analysis and coding and then generating public and restricted-use datasets for data users. Data users include individuals that use DVS products, such as policymakers, academic researchers, and public health agency statisticians.

As interoperability refers to the standards-based exchange of data between two data exchange partners, the NVSS ecosystem can be categorized into three types of interoperable information exchange:

1. *Data Capture:* The exchange of information captured either by data entry professionals (e.g., birth clerks, clinicians, and coroners) who document birth or death vital information at the point of information capture within setting specific systems (e.g., EHRs, case management systems, VRO birth/death certificate systems) with the custodian or data owner, the state VRO.
2. *Data Exchange:* The exchange of information from state VROs with the CDC, supported by NAPHSIS for data routing. While the U.S.'s Standard Certificate of Birth, Standard Certificate of Death, and Standard Report of Fetal Death have long been standardized, individual jurisdictions maintain autonomy over their jurisdiction-specific data collection and systems.
3. *Data Processing:* The exchange of statistical datasets developed from individual records received from jurisdictions with end-users of NVSS datasets. These data are first processed to analyze, code, and standardize data received across 57 states and jurisdictions and then aggregated to develop statistical datasets available for download by data users.

Using the CDC website, NCHS presentations, and published literature as information sources for the NVSS modernization scan, Clinovations catalogued 25 NCHS modernization activities: nine data capture, eight data exchange, and eight data processing activities. The data processing activities were further classified into two groups: analysis and coding and dataset generation. Figure 7 below depicts a summary of NVSS modernization efforts to visualize associated stakeholders and interoperability areas.

At the end of this section, Table 7 provides a listing of the modernization projects reviewed that indicate the impact of each modernization activity improving the timeliness and/or data quality of national vital statistics data. Timeliness efforts focused on increasing the rate of release of vital records data. Data quality projects aimed at improving accuracy, completeness, and validity of viral records data. Some projects aimed to improve both timeliness and data quality. Figure 7, below, provides a listing of NVSS projects reviewed by project type followed by Figure 8 that provides a list of the projects reviewed within each project type.

Figure 7. NVSS Modernization Efforts by Project Type

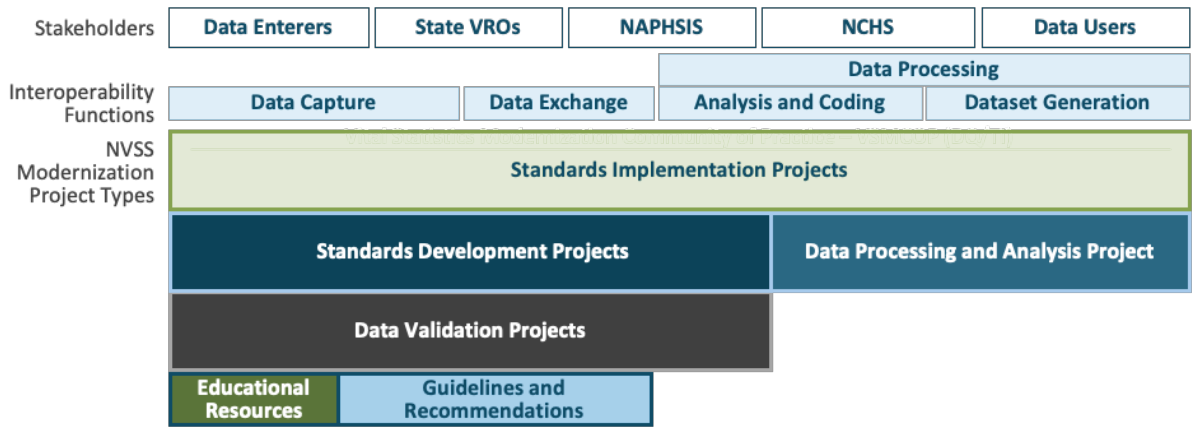
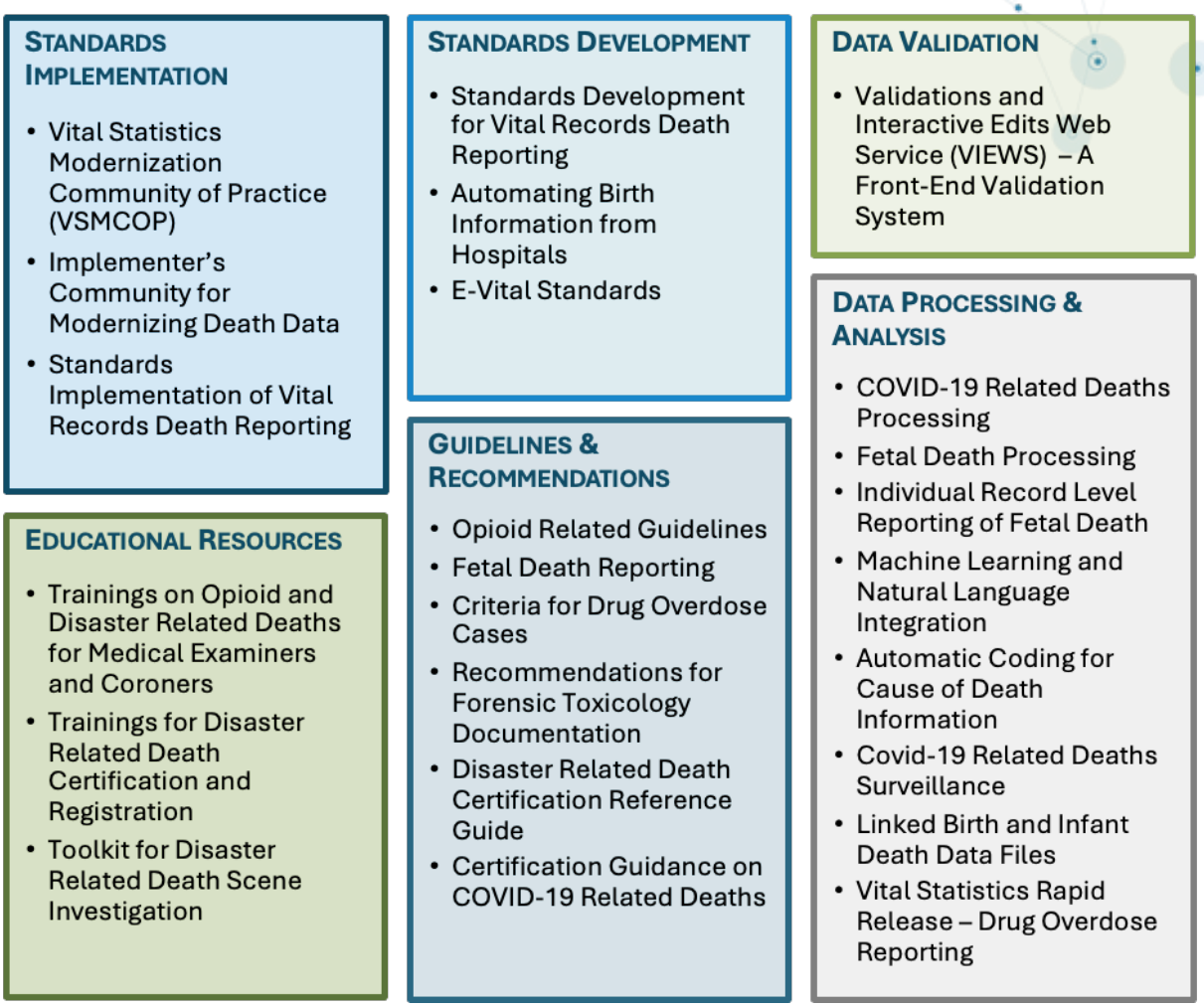
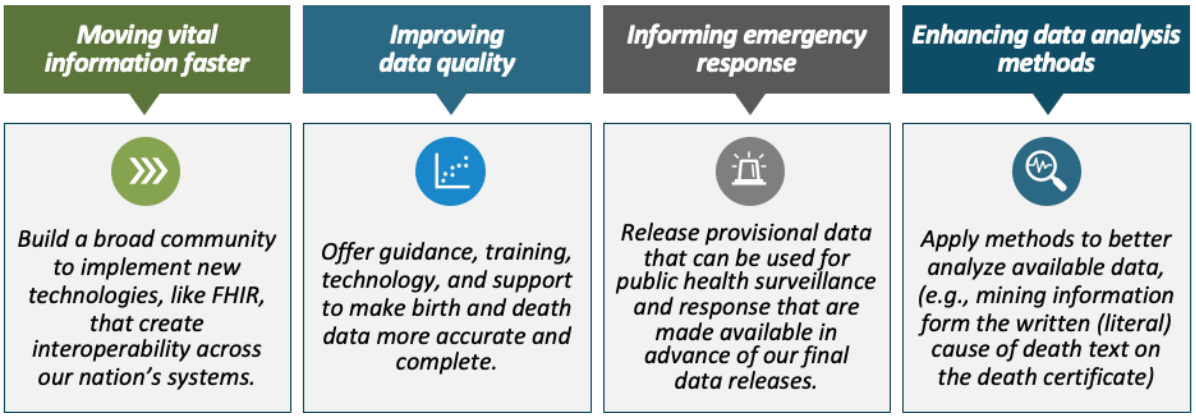


Figure 8. NVSS Modernization Projects by Type



The NCHS Data Modernization Initiative continues to evolve and expand to address new and emerging ways to make vital information available for research, evidence-building, and public health. While Figure 7 and Figure 8 above represents Clinovations review and classification of NVSS modernization projects, Figure 9 below published by NCHS DVS notes four areas for NVSS modernization activities.

Figure 9. NVSS Modernization Areas



clinovations
GOVERNMENT + HEALTH

Vital Statistics Standards Review

Clinovations reviewed the U.S. Standard Certificate of Birth (2003 Revision) and the U.S. Standard Certificate of Death (2003 Revision) to identify the data elements that are collected from those two documents. The 2022 Mortality Multiple Cause-of-Death and Natality public use files were reviewed to understand NCHS's data mapping process, which transforms raw data into codified, de-identified record-level files and aggregate datasets for public use.

Two activities were conducted: 1) reviewed the code and/or the value set(s) to understand the differences between standards, and 2) created a data mapping across identified data standards for one vital records data element: "marital status". Marital status was chosen as a representative data element for review because it did not include a large number of values within standards-based value sets; also, it could serve as an example of the type of standards harmonization activities that could advance NVSS interoperability.

Marital status may present a unique perspective related to workflow, source of data, and associated metadata as NVSS intentionally defines the source of each data element, marital status being one of them. Currently, marital status is obtained from the mother using the natality worksheet and does not originate from EHR data or EHR users. The natality worksheet is the preferred source of marital status data, as the source of the marital status data element is unknown if the data element is to be obtained from the EHR.

While this analysis presents opportunities for harmonization, it also highlights the importance of data capture in places where provenance is critical because reliability and accuracy may vary by data source. Vital statistics and other datasets may wish to offer preference based on source.

The project team analyzed five vendors' API/FHIR documentation to identify code and value sets and FHIR IG profiles used by each vendor for the "marital status" field. This process resulted in the project team reviewing five data standards. Then, the four additional source documents were reviewed to analyze the marital status data element.

Documents Reviewed:

- U.S. Standard Certificate of Birth (2003 Revision)
- U.S. Standard Certificate of Death (2003 Revision)
- 2022 Mortality Multiple Cause-of-Death Public Use Record
- 2022 Natality Public Use File

Standards Reviewed:

- IJE
- FHIR (VRDR v2.2.0)
- FHIR (BFDR v1.1.0)
- HL7 v2
- HL7 v3

EHRs FHIR Specifications Reviewed:

- Epic
- Cerner
- CPSI
- MEDITECH
- NextGen

Modernization Opportunity: Value Set Alignment

Using the marital status data element as a case study, the project team encourages the NCHS and the DVS to conduct additional value set mapping activities to better understand the numerous value sets within the NVSS data ecosystem. By mapping the various value sets prescribed by governing data standards, NCHS and DVS will have a better understanding and a clearer picture of the potential amount of data granularity that could be preserved in aggregated datasets. This could improve the overall quality of data or flexibility in the NVSS data ecosystem.

Additionally, NCHS mapping would increase metadata-based insights, enabling NVSS data users to understand the potential values they are not receiving within NVSS datasets. These value set mapping activities could provide findings and insights that may be critical for the next version of the U.S. Standard Certificate of Birth (2003 Revision) and the U.S. Standard Certificate of Death (2003 Revision). These data standard and value set mappings could be made publicly available to serve as additional data documentation materials and could be offered as a service in a potential future NSDS.

Preliminary review of these value set mapping recommendations with this project's sponsors revealed that data standard and value set mapping activities are of interest and planned within NCHS and DVS. Future revision versions of the U.S. Standard Certificate of Birth (2003 Revision) and U.S. Standard Certificate of Death (2003 Revision) can be designed with data standards and value sets alignment taken into consideration to leverage interoperable data to reduce data entry or data custodian mapping burden and improve the consistency and quality of data received by NCHS.

NVSS Modernization Takeaways

Insights garnered from NVSS modernization efforts can illuminate best practices and key success factors and facilitators for a potential future NSDS and the overarching data and evidence ecosystem. NVSS is a well-established ecosystem and stands as a testament to effective intergovernmental data sharing in public health. It has decades of experience applying statistical methodologies to generate vital record datasets for public access, while also ensuring the security of confidential data through secure enclaves.

NVSS has functioned as an innovation hub, acting as a dynamic testing ground for the experimentation and implementation of novel methods and cutting-edge technology aimed at enhancing data interoperability. Like NVSS, a potential future NSDS will support data users who have obtained source data from different stakeholders and will generally collect data locally, though it may require intermediaries to route and process data for analysis and dataset generation at the national level. To shed light on best practices and learned lessons, our team offers the following highlights of NVSS modernization efforts to better model potential NSDS core functions:

1. **Communication:** By collaborating with industry groups and associations, NCHS has been able to identify and assist with stakeholder specific interventions to improve data collection for vital records. For example, in collaboration with the National Association of Medical Examiners (NAME), NCHS has published guidelines and recommendations for opioid-related death investigations for medical examiners and coroners. To support these efforts, NCHS also developed educational resources for jurisdictions and their technical partners (e.g., NVSS Modernization Toolkit[xviii]) and implemented trainings for medical examiners and coroners. These efforts have helped to standardize data collection, improve data quality, and drug overdose reporting.

[xviii] NVSS Modernization Toolkit: https://www.cdc.gov/nchs/nvss/modernization/tool-kit.htm

Key Success Factors and Enablers

- Federal (NCHS) outreach and engagement with professional organizations that serve individuals that enter data to be processed for statistical use.
- Federal (NCHS) commitment to evaluating and enhancing data quality at the source.
- Federal (NCHS) assistance in the development of communication material to educate local entities and data enterers.
- Segmented communities and convenings by aligned roles and responsibilities for effective collaboration.

2. **Data standardization.** NCHS has engaged in standards development and implementation for birth and death data. These efforts are ongoing, but projects to date have benefitted from two-way data exchange between stakeholders. Learning communities have played a role in advancing data standardization efforts. The use of HL7 FHIR standards has improved the timeliness, accuracy, and completeness of vital records data.

Key Success Factors and Enablers

- Standard certificates and reports have been standardized since 1989 and the latest version (2003) has been implemented by all jurisdictions since 2018.
- Identification and convergence on a standard (HL7 FHIR) with established communities, processes, and procedures.
- Federal (NCHS) oversight and management of operational processes.
- Federal (e.g. Coronavirus Aid, Relief, and Economic Security Act (CARES) Act[xix]) funding to jurisdictional partners to progress data standardization.
- Federal (NCHS) leadership in launching and organizing communities.
- Public-private collaboration with data enterers, data custodians, data users, and technology developers.

3. **Coordination.** NVSS has encouraged the development of communities of practice such as the Vital Statistics Modernization Community of Practice (VSMCOP). An implementer's community was initially supported by funding through the Patient Centered Outcomes Research Trust Fund (PCOR-TF), which evolved to the current community of practice. Focusing specifically on modernizing death data, this community focused on developing, testing, and piloting new approaches to exchanging mortality information using HL7 FHIR standards. Expanding on the work from the Implementer's Community, VSMCOP emerged as an expanded effort to modernize birth and death data using HL7 FHIR standards. Each community brought together a diverse set of stakeholders to share ideas, technical tools, and resources. These learning communities have facilitated the adoption of standards leading to improved timeliness and quality.

Key Success Factors and Enablers

- Federal (NCHS) leadership in launching and organizing communities.
- Federal (PCOR-TF) funding to support state-level testing and pilots.
- Public-private collaboration with data enterers, data custodians, data users, and technology developers.
- Identification and convergence on a standard (HL7 FHIR) with established communities, processes, and procedures.

[xix] CDC COVID-19 State, Tribal, Local, and Territorial Funding: https://www.cdc.gov/budget/fact-sheets/covid-19/funding/index.html https://www.cdc.gov/budget/fact-sheets/covid-19/funding/index.html

4. **Research and Development.** NVSS has served as an innovative sandbox for the integration of modern tools and technologies. From the modernization efforts we cataloged, our team identified fourteen activities that explored the use of coding systems (seven activities) or HL7 FHIR (seven activities) technology to advance data interoperability and data processing. For example, NCHS developed a system to automatically code the cause of death on death records. The integration of an automatic coding system increased the proportion of causes being auto-coded on death records from 75 percent to 90 percent, thus helping to improve the timeliness of death data. In addition to responding to emerging data needs from the COVID-19 pandemic, NCHS adjusted internal data processing systems to track COVID-19 related deaths.

Key Success Factors and Enablers

- Federal (NCHS) commitment to utilize modern technology to advance data analytics and capabilities.
- Public-private collaboration to support urgent or emerging data needs.
- Identification and convergence on a standard (HL7 FHIR) with established communities, processes, and procedures.

NVSS Modernization Projects Listing

Table 7 below provides a listing and brief description of the NVSS modernization projects reviewed and indicates the impact to NVSS data quality and/or timeliness.

Table 7. Reviewed NVSS Modernization Projects Listing

Short Description	Description	Impact Type
Opioid Related Guidelines for ME/Cs	Updated 2013 guidelines on death investigations, evaluations, and certifications for opioid related death investigations.	Data quality
Fetal Death Reporting	Redesigned the 2003 US Standard Report of Fetal Death.	Data quality
Trainings on Opioid Related Deaths for ME/Cs	Developed training materials and methodology for educating ME/C on guidelines for opioid related death investigations.	Data quality
Criteria for ME/Cs on Drug Overdose Cases	Developed criteria for identifying suspected drug overdose cases.	Data quality
Recommendations for Forensic Toxicology Documentation	Provided recommendations for forensic toxicology practices to document opioid related drug investigation.	Data quality
Toolkit for Disaster Related Death Scene Investigation	Developed a disaster related death scene investigation toolkit for collecting data at death scenes during and after natural disasters.	Data quality
Disaster Related Death Certification Reference Guide	Developed and published the 2017 disaster related death certification reference guide.	Data quality
Trainings for Disaster Related Death Certification and Registration	Conduct trainings to support efforts to strengthen disaster-related death certification and registration processes in PR, TX, USVI.	Data quality
Front-end validation systems (VIEWS)	Improve and incorporate consistent validations for data entered into EDRS.	Data quality
Certification Guidance on COVID-19 Related Deaths	Developed certification guidance for COVID-19 related deaths.	Data quality
Fetal Death Processing System	Managed a new fetal death processing system.	Data quality Timeliness
Machine Learning and Natural Language Integration	Incorporated natural language and machine learning techniques to death data in NVSS.	Timeliness
Individual Record Level Reporting of Fetal Death	Enabled reports on individual record level transactions from states by creating supplemental drug data beyond ICD-10.	Data quality
Automatic Coding for Cause-of-Death Information	Develop a system to automatically code the cause of death on death records.	Timeliness
COVID-19 Related Deaths Processing System	Adjusted internal data processing systems to track COVID-19 related deaths.	Data quality
Vital Statistics Modernization Community of Practice (VSMCOP)	A learning collaborative to share ideas, technical tools, resources, and practices to improve birth and death data using FHIR standards.	Data quality Timeliness
e-Vital Standards	Developed vital record standards to enable interoperable electronic data exchanges.	Data quality Timeliness
Designed Puerto Rico's EDRS	Supported PR as they designed and implemented their first EDRS.	Data quality Timeliness
Automating Birth Information from Hospitals	Develop national standards for the automatic transfer of medical and health birth certificate data from EHRs to VROs.	Data quality
Implementer's Community for Modernizing Death Data	A learning collaborative to develop, test, pilot, and track new and more interoperable approaches to exchange mortality data using FHIR standards.	Data quality Timeliness
Standards Development for Vital Records Death Reporting	Established a nationally approved HL7 FHIR standards for vital records death reporting.	Timeliness
Standards Implementation of Vital Records Death Reporting	Created applications to test FHIR data standards.	Timeliness
Linked Birth and Infant Death Data Files	Link period and cohort data files to study the relationship between infant death and risk factors present at birth.	Timeliness
Improved Drug Overdose Reporting	Redesigned the vital statistics rapid release program by adding additional demographic information and drug type information.	Data quality
COVID-19 Related Deaths Surveillance System	Developed a surveillance system to release daily updates of COVID-19 related deaths.	Timeliness

Legend

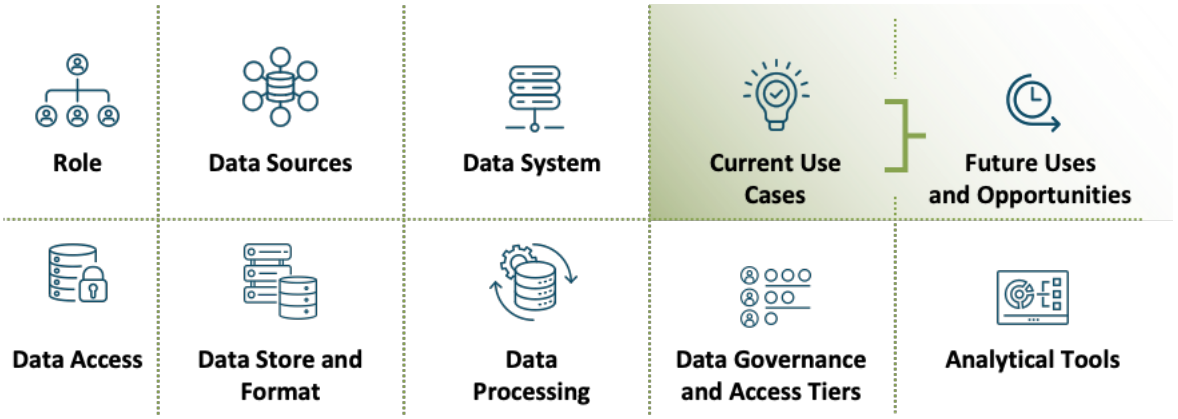
- ME/C: Medical Examiner and Coroner
- PR: Puerto Rico
- TX: Texas
- EDRS: Electronic Death Registration System
- USVI: U.S. Virgin Islands

APPENDIX D - STAKEHOLDER WORKSHOP PERSPECTIVES

Clinovations hosted three-hour stakeholder workshops with four sets of stakeholders that support the NVSS ecosystem: data entry personnel, data custodians, data users, and analytics and innovator vendors. Stakeholder workshops received OMB approval and consisted of 90 minutes with the entire stakeholder group and 90 minutes in two smaller breakout groups for all groups except the analytics and innovator vendors who remained together as a full group in the latter 90 minutes. There were 30-minute breaks between each 90-minute session.

General group discussions were led by a facilitator who is a subject matter expert (SME) on the discussion topics and questions. Invitees were provided with the list of discussion topics depicted in Figure 10.

Figure 10. Stakeholder Workshop Discussion Topics



During general group discussion (first 90 minutes), the project team asked participants to share their experiences in collecting, storing, accessing, processing, and using vital statistics data and provide perspectives on future needs and modernization opportunities. The second 90-minute session offered a smaller group environment of similar participants (i.e., by role or function) to foster conversation amongst the three stakeholder workshops with a higher number of total participants. Participants were not responsible for synthesizing any breakout group documents (i.e., notes or report out presentation).

Online polling was used to gather insights that the facilitator then used to prompt discussion amongst participants. Poll questions were used at the beginning of each discussion topic to engage participants and lead into discussion topics and questions. Poll results were presented live to participants and were not intended for quantitative data collection, but to foster and prompt discussion amongst participants.

Figure 11 offers a profile of invitees by stakeholder workshop.

Figure 11. Stakeholder Workshop Invitees

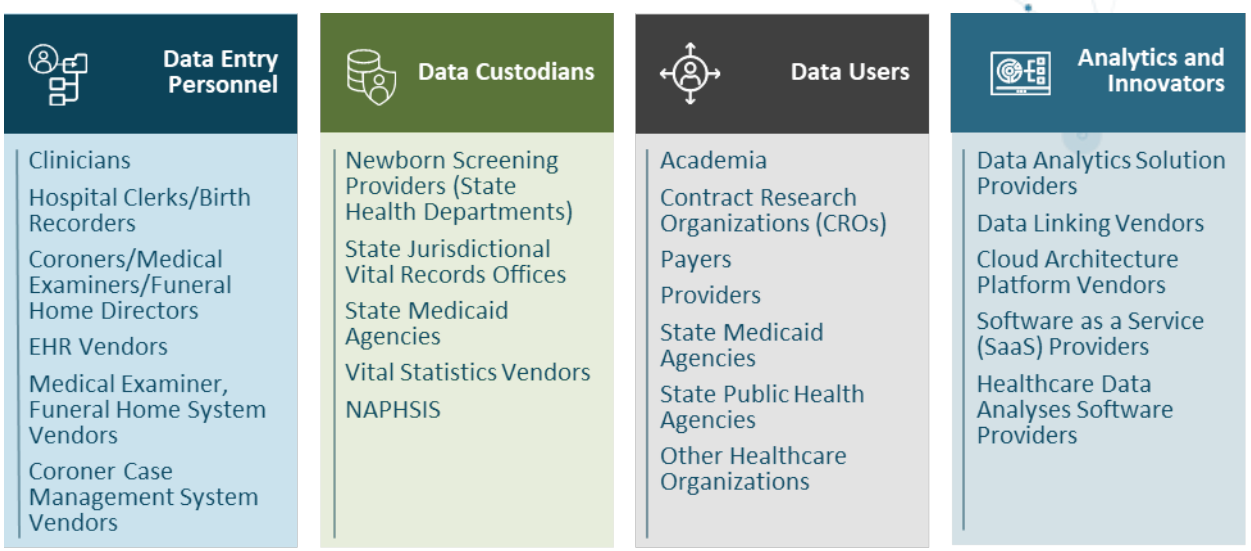
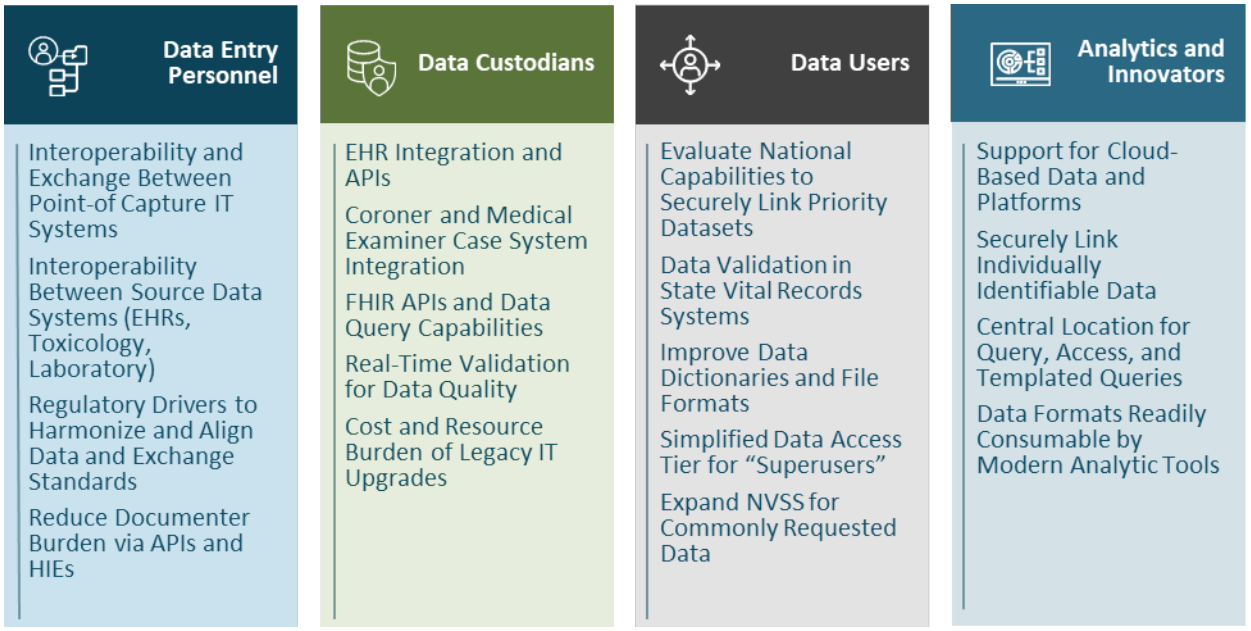


Figure 12 offers a summary of modernization opportunities for consideration provided by stakeholder workshop participants.

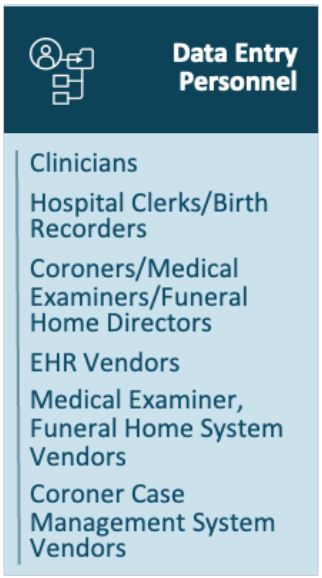
Figure 12. Modernization Considerations Identified by Stakeholder Workshop Participants



Brief summaries for each stakeholder workshop are provided in the following sections of this appendix.

clinovations
GOVERNMENT + HEALTH

Data Entry Perspectives



Participants in this stakeholder workshop were invited from a broad group of "data entry" personnel, which was comprised of hospital stakeholders, medical examiners, and coroners.

A total of 14 data entry stakeholders participated in this stakeholder workshop. Of the 14, six individuals represented hospital stakeholders (e.g., clinicians, safety/quality leaders, and EHR vendors), and eight represented medical examiners and coroners in state or local jurisdictions.

NVSS Modernization Opportunities – Data Entry

Increase Interoperability and Exchange Between Point of Capture IT Systems

Stakeholder workshop participants highlighted that there is limited interoperability between point-of-capture systems that contain birth or death information and state/national reporting systems. This presents many opportunities to:

- Reduce duplicate data entry in point of care/capture and Electronic Death Registration (EDRS)/state Vital Records Office (VRO) systems,
- Update and amend processes that are delayed and burdensome, and
- Align EHR certification requirements with national certificate standards.

Investigate Interoperability Between Source Data Systems

There is also very limited interoperability between other health IT systems (e.g., toxicology laboratory, state electronic death registration system) that contain critical information to complete birth or death certificates that can delay completion. Stakeholder workshop participants discussed opportunities to:

- Increase real-time access to toxicology, laboratory, medical records, mental health records, and
- Encourage state HIEs to support improved access to historical and clinical records (e.g., ambulatory visit notes).

Consider Regulatory Drivers to Harmonize and Align Data and Exchange Standards

Stakeholder workshop participants commented that regulatory requirements could accelerate technology developer prioritization and the implementation of interoperability standards. EHR vendors indicated that it is technically feasible to develop structured data elements and to create FHIR configuration and mapping capabilities to support vital records. Some medical examiner and coroner case management systems have integrated with state VRO systems, but these modernization efforts are jurisdiction-dependent rather than nationally aligned. There are opportunities to:

- Advance standards, interoperability, and product certification efforts to support inclusion of new or expanded data elements, which will help improve research and general understanding of death, and
- Harmonize metadata regarding data provenance to better support trust-building and insight into origination of data, support decision-making, and increase data availability.

Reduce Documenter Burden via APIs and HIE

Stakeholder workshop participants called for higher levels of systems integration and interoperability via APIs and HIE to support access to interoperable data and facilitate population of birth and death certificates from other systems. There are opportunities to:

- Support best practice workflows and physician education on the downstream impact of accurate death documentation,
- Reduce redundant and inefficient data entry between EHRs and EDRS/Electronic Birth Registration System (EBRS),
- Support data accesstiers and data exchange for physicians, medical examiners, and coroners to easily access historical data or external data via APIs or HIEs, and
- Streamline and automate data entry and national certificate requirements to reduce EDRS death certificate creation from 15-45 minutes to less than 10 minutes per certificate.

Data Custodians Perspectives

Participants in this stakeholder workshop were invited from a broad group of "data custodians" stakeholders, which was comprised of subject matter experts from organizations that store and maintain vital records data or that support organizations that facilitate the exchange of information between vitals data sources and the CDC. Participants represented state vital records offices, public health agencies, technology vendors that support VROs, and NAPHSIS.

A total of 14 data custodian stakeholders participated in this stakeholder workshop. Of the 14, eight individuals represented state VROs or public health offices, and six represented technology vendors or consultants that support state VROs in collecting or exchanging vital data with NCHS.

NVSS Modernization Opportunities – Data Custodians

Increase EHR Integration and Use of APIs

Stakeholder workshop participants highlighted the following opportunities for improvement and modernization by a potential future NVSS for EHR integration and use of FHIR APIs:

- Certify and report U.S. Standard Certificate of Death (2003 Revision) data from providers and EHRs, and
- Increase active participation from EHRs by pulling directly from them for vital records use cases.

Integrate Coroner and Medical Examiner Case Systems with VROs

While some models of coroner and medical examiner system integrations with VROs are emerging, stakeholder workshop participants called for increased use of FHIR and other health data standards that support integration between case management systems and state VRO systems for completion of birth and death certificates. Opportunities for modernization include:

- Integrate EDRS with coroner and medical examiner case management systems for new registrations, amendments, and cremation approval/permits use cases, and
- Configure routine reports for reporters to download directly from their systems.

Expand FHIR APIs and Data Query Capabilities

Stakeholder workshop participants discussed the following modernization opportunities to reduce burden on data reporters and to improve submission timeliness:

- Eliminate the need for data providers to log into state and jurisdiction VRO systems by implementing FHIR APIs with EHRs,
- Develop the capability for data recipients to query data or subscribe to data of interest to their office/program, and
- Allow VRO systems to support amendments via an interface after a record has been completed and certified (vs. only when the record is still in progress).

Implement Real-Time Validation for Data Quality

Stakeholder workshop participants explained that duplicate data files and datasets are often created as the original document is secured and undisturbed. Additionally, duplicate data are also obtained at the state level from hospitals and various health department agencies and divisions. Duplicative processes can result in the delay of vital record statistics data files, especially when source documents (e.g., U.S. Standard Certificate of Birth and U.S. Standard Certificate of Death) are amended. There are modernization opportunities to:

- Implement identity assurance authorization and authentication workflows for data custodians, reduce the level of burden for data users to access the data, and reduce the need to create multiple duplicate data files and datasets,
- Develop identity validation/data sharing APIs with data partners based on use cases,
- Replace the need to manually share data files with data exchange partners (e.g., confirming that a person exists for the Department of Motor Vehicles or verifying the cause of death to determine if someone died of an infectious or reportable condition), and
- Improve real-time validation services to ensure high data quality, including data exchange testing modules and responses.

Reduce Cost and Resource Burden of Legacy IT Upgrades

Organizations and state health divisions and agencies could conduct cost and benefit analyses to better understand whether investment(s) into modern health IT system(s), module(s), and tool(s), would benefit and meet their organizational needs. There are opportunities to:

- Reduce the variability and support equitable budgets across state and jurisdiction health divisions and agencies to implement upgrades,
- Promote implementation of standards and FHIR capabilities in technology solutions and systems that are used by state VROs, and
- Ensure that both the data requester and partner have compatible capabilities for the implementation and adoption of APIs and FHIR.

clinovations
GOVERNMENT + HEALTH

Data Users Perspectives

Participants in this stakeholder workshop were invited from a broad group of data user stakeholders, comprised of users of statistical files, including academic researchers, public health researchers, state public health agencies, and statisticians.

A total of 20 data user stakeholders participated in this stakeholder workshop. Of the total, nine individuals represented academic or university researchers, and 11 represented state public health agencies, including epidemiologists, statisticians, and state Pregnancy Risk Assessment Monitoring System (PRAMS) coordinators.

NVSS Modernization Opportunities – Data Users

Evaluate National Capabilities to Securely Link Priority Datasets

Stakeholder workshop participants acknowledged that they had limited understanding of restrictions and limitations to linking federal statistical and state-level datasets. Participants recommended a review of secure linking and computing capabilities that could support expanded record-level linking of NVSS data. They suggested the following modernization opportunities:

- *NCHS:* Review the common dataset and data linking needs from the academic and public health user communities to identify datasets that could be securely linked, either by the government or authorized data users.
    - This action could expand the potential for research and evidence-building.
- *Researchers:* Seek the ability to link individual and national records data with DUAs that support secure and appropriate use for longitudinal analyses.
    - While NCHS may not have the capacity to support data linking, they could encourage conversations between states to set up common standards or data use agreements.
    - An example of databases that have linked vital records and longitudinal databases include the Massachusetts Pregnancy to Early Life Longitudinal Data System (PELL). PELL is a population based longitudinal reproductive data system that annually links birth and fetal death certificate records with hospital discharge records for both the mother and infant from 1998 to the present.

Improve Data Validation in State VROs

Stakeholder workshop participants discussed data validation and noted that, when state-level data are accessible, there are duplicate data entries because the information has not gone through the same level of processing (e.g., data cleaning and editing) as national vital records data. Participants shared examples of modernization opportunities such as:

- Improved state-level data validation would be particularly helpful for death data. For example, a system could check if a pregnancy check box was mistakenly selected, and
- Incorporating data validation systems within state vital records systems could improve data quality of NVSS data as it is impacted by missing, miscoded, or inaccurate state-level data.

Improve Data Dictionaries and File Formats to Support Users

To help save time and resources for analyzing and preparing data, data users identified modernization opportunities to:

- Create a more complete and updated data dictionary for vital records data, and
- Ensure that data dictionaries and file formats are ready for use by statistical and analytic packages (e.g., .CSV format).

Implement a Simplified Data Access Tier for "Superusers"

Stakeholder workshop participants suggested that a potential future NSDS or NCHS could implement a simplified and rapid process for "super users," who are known entities and who frequently request the same dataset on regular intervals using the same process for each request. Participants indicated that

- They avoid studies that would require restricted use data or only sought collaborations with individuals who commonly work with NVSS restricted use and federal statistical research data centers and
- Certain researchers are "known" as they regularly apply/re-apply to access the same data for the same purpose(s) and inquired whether an access tier could be simplified for users that are well-known that regularly submit applications.

Expand NVSS for Commonly Requested Data Elements and Datasets

Data users sought methods to request additional data elements and ready access to commonly requested datasets. Stakeholder workshop participants noted opportunities to:

- Develop a process to track and monitor common data requests and needs by researchers. Examples of expanded data elements or processes included:
  - Expand NVSS data to include subpopulation capabilities to align with other standards such as the OMB race/ethnicity standard, and
  - Consider collecting state-specific data elements.
- Create ready-use datasets or APIs for common requests and data analyses across states, such as:
  - Frequent (e.g., monthly, quarterly, annual) analyses published by multiple states and regions could be readily available for use.

Data Analytics Vendors and Innovator Perspectives

Participants in this stakeholder workshop were invited from a broad group of "data analytics and innovator vendors" stakeholders, which refers to organizations that may offer data analytics, processing, visualization, linking, and other commercial platform or custom solutions capabilities. Participants represented cloud service providers, data analytic solutions, data linking vendors, and other platform providers. Participants discussed emerging technologies that may be considered as part of a modernized data ecosystem to capture, exchange, and use vital statistics data.

A total of eight stakeholders participated in this stakeholder workshop. All participants in this stakeholder workshop provided solutions using cloud-based data, linking, or analytic platforms. While the participants worked for organizations that provide solutions for commercial purposes, all participants reported customers that included the federal government and academic researchers who may be authorized users of public and restricted-use NVSS data.

NVSS Modernization Opportunities – Analytics and Innovator Vendors

Support for Cloud-Based Data and Platforms

While not all participants stored confidential data in their cloud-based platforms, there are still cloud-based storage components derived from identifiable data. Applying their experience analyzing other federal datasets to NVSS, participants noted that:

- Where cloud-based data platforms are intended specifically to support expanded users, perform analysis in the cloud, and linkages across datasets, the NVSS Conditions of Use for Restricted Data prevents use of cloud-based platforms, prohibits data sharing outside of named individuals in the DUA and does not permit linking to individually identifiable data.
- While current NVSS and other statistical datasets may have use restrictions due to privacy and confidentiality concerns, a potential future NSDS could serve as a national expert or advisor on emerging technologies that could offer expanded capabilities, reduce data access burden, and contribute to further evidence-building.
- If a centralized data lake is not feasible at the federal level, one participant suggested the requirement of having a static, predictable link for each monthly/weekly data file as an alternative model to a centralized data lake. This would be essential for automated data pull modernization opportunities. By having a static, predictable link for each monthly/weekly data file, data users could create an automated program that pulls data from the static, predictable link, even for data files that are released in the future.
- While reference architectures will function most of the time, when a new crisis or use case arises, a new variable may require the complete re-design of the IT infrastructure architecture. Participants suggested that the project team and the federal government identify the necessary IT infrastructure, such as data lakes, analytical tools, and secure computing environments, should another crisis or use case arise. This may allow for development to be conducted ahead of the next crisis or use case.
- Recruitment of highly qualified staff to private vendor and government roles and related challenges around salaries, skillsets, and qualifications is a challenge. As the Evidence Act requires data users to be technically trained, there is an opportunity for a potential future NSDS to offer support and training for skill building in cloud-based data and platforms.
    - Participants indicated that the industry and vendor community should be providing the necessary training and training materials (e.g., user manuals) to use their platform and

analytic tools. They encouraged data users to demand this service from their platform vendors.

Securely Link Individually Identifiable Data

Participants discussed opportunities to address research limitations without linked data and encouraged future modernization and potential NSDS efforts to evaluate options to securely link individually identifiable data. Without linked data across public health reporting systems, participants indicated that challenges persist in obtaining data needed for research or surveillance. A potential future NSDS could consider key questions for which data is being exchanged that can't be answered because the data is not linked or de-identified prior to exchange. Participants provided examples of priority areas for secure linking:

- National Provider Identifier (NPI) information should be captured in electronic health records, a source health IT system for vital records data, as well as the National Plan and Provider Enumeration System (NPPES). Data analysis for death statistics is a challenge because information on the presiding physician/pathologist who coded the cause of death (using ICD-10) is not present in the NDI dataset.
- Participants hoped for an opportunity to support de-identified data linkages across independent public health reporting data streams to CDC (e.g., immunizations, electronic lab reporting, electronic case reporting, vital statistics). Participants noted that each jurisdiction either sends identifiable data for de-identification by CDC (e.g., NVSS) or uses their own process to de-identify data prior to sending reports to CDC.
- ***Exa mple:*** Participants discussed the National COVID-19 Cohort Collaborative (N3C), managed by the NIH as an example of securely linked identifiable data that has been successfully implemented for research. The N3C model acquires and ingests raw datasets, tokenizes personally identifiable information, and links datasets together, using the tokenized personally identifiable information, enabling the enclave to link datasets while maintaining confidentiality. The N3C infrastructure (e.g., data governance and oversight, data standards, and analytical tools) enables data users to request access, access the data, and conduct the computational analysis on the accessed dataset in a secure environment.

Central Location(s) for Query, Access, and Templated Queries

While open-source languages, such as SAS, R, and Python, are commonly used to conduct computational analysis activities on NVSS datasets, the learning curve to learn one of the learning source languages is high. Moreover, there is a dearth of technical individuals who are available for these positions. To lower the curve to using computational analysis, NCHS could consider the development of a codebase library consisting of commonly conducted computational analysis script templates.

- NCHS and other federal statistical agencies could lower the bar to use computational analytics more by exploring the possibility of creating "plug and play" computational analysis modules. These modules would not require any manipulation of a code, while still enabling the data user to conduct computational analysis.
- Other potential data tools may include (but are not limited to) a data visualizer, a dashboard for "fast statistics" (i.e., common queried statistics that are available for download), synthetic datasets to simulate datasets that require higher access permissions, secure computing environments, and cloud-based data analytics.
- Participants recommended scenario-based ethical hacking and "war games" cloud-based capabilities where the community can support preparation for future data governance and risk scenarios. An example offered by a participant who supported related activities for a federal workgroup described a scenario where a cybersecurity attack (e.g., denial of service attacks) occurred during a national pandemic (e.g., COVID-19).
- Another consideration offered by participants was a generative AI bot that trains only on the data that exists within the data user's secure computing environment. This supports preserving confidentiality and ensuring that only data users with all necessary access rights can use the data. The AI bot can generate code to conduct computational analysis on the data user's dataset. Additionally, the AI bot has capabilities to test, debug, and regenerate code for computational analysis. Another key feature of the AI bot is that data users can pull and/or query other datasets into the data user's secure computing environment, reducing the burden for the data user. The AI bot uses natural language to interact with the data in the data user's secure computing environment. Users of this new AI bot tool do not need an account with the participant's organization. However, there would be a cost incurred that may need governance models to offset excess activity or limit AI bot services to select groups.

- One participant suggested transforming commonly requested AI bot actions into templated queries or code as a potential solution to decrease overall cloud computing activity. Data users could then use these templated queries or codes to complete the computational analysis, instead of requesting that the AI bot complete the activity.

Example: Participants offered an example of a virtual research data center that may serve as an area for further investigation and support via a potential future NSDS: Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC).

- The CCW VRDC is a virtual research environment that provides timely access to Medicare and Medicaid program data. Databricks manages half of the analytical capabilities in the CCW VRDC and the native statistical analysis software in the environment is SAS. However, there is a cost to obtaining access to the CCW VRDC.
- Participants noted the CCW VRDC as a prime example of a data enclave that enables data users to conduct multiple different data abstractions in a vendor, coding language, and dataset agnostic environment. Additionally, the participant noted that the CCW VRDC has dramatically increased the amount of accessible data while reducing the burden of obtaining it.
- The process to obtain data from the CCW VRDC involves user identification of the data elements necessary to construct the desired dataset, which may qualify for public-use (e.g., no personally identifiable information and geographical data). Automations exist to assist with the compilation of the selected data elements. The list of data elements is then submitted to CCW VRDC for manual review. CCW VRDC then notifies the participant via email whether the request was approved or denied.

Use Data Formats Readily Consumable by Modern Analytic Tools

When discussing the challenges and barriers to using data elements and datasets obtained from federal agencies, it can be difficult to browse and access government datasets because each federal agency chose a different data platform vendor to host their public use data files. Stakeholder workshop participants offered the following considerations for both NVSS and areas for NSDS to develop guidance and best practices.

- Federal datasets (such as NVSS data) were offered in the Amazon Web Services (AWS) data marketplace because it created one centralized location to query and access publicly available government datasets.
- Data users use a combination of analytic tools native to their platforms but also third-party tooling services. All statistical analysis software (e.g., SAS, R, and Python) are critical tools. Depending on the data user stakeholder type (e.g., clinical, research, billing/finance), one statistical analysis software may be preferrable or is the industry standard (e.g., SAS and Agency for Healthcare Research and Quality (AHRQ) common formats), compared to other statistical analytical software.
- Participants noted preferences for using tools such as Databricks and SAS, moving forward, because they want to align with the government's platform to maximize efficiencies and minimize costs.
  - Hiring costs and available workforce are primary considerations for the migration towards cloud-based analytics tools. Currently, there are a dearth of technical data users who are proficient in one statistical analysis software language, let alone multiple languages.


clinovations
GOVERNMENT + HEALTH