From Silos to Superhighways

A Framework for Public Health Data Modernization

About the Authors

Zach Finn
Senior Technical Architect – Ruvos
Founder & President – Data Finn

Zach is the founder and President of Data Finn, a boutique data consulting firm that specializes in studying the U.S. healthcare system and developing innovative ways to improve it. In addition to leading Data Finn, he serves as a senior technical architect at Ruvos, a national leader in public health informatics and health information technology. His work in Health IT began in 2005 when he led the creation of one of the nation's first Health Information Exchanges (HIE) in northern Florida, The Big Bend Regional Healthcare Information Organization.

A nationally recognized advisor, Zach has provided expertise to HIEs, state public health agencies, and hospitals across the country. His recent focus has been on Data Modernization Initiatives, working with states including Vermont, Idaho, Missouri, New Jersey, and Florida. He also shares his knowledge as an adjunct professor at Florida State University, where he teaches Advanced Health Informatics when time allows.

Jeff Couch
Managing Partner – Ruvos

Jeff is a co-founder and Managing Partner at Ruvos, leading the company's strategic direction in public health data modernization and healthcare technology solutions. He plays a hands-on role in shaping Ruvos' approach to data interoperability, security, and infrastructure modernization, ensuring that public health agencies and healthcare institutions can effectively harness the power of data.

With extensive experience in complex data ecosystems, Jeff works closely with partners to design scalable, secure, and forward-thinking solutions that support national and state-level modernization initiatives. He provides direct guidance on cloud strategy, data governance, and interoperability, helping organizations navigate evolving technical landscapes. Under his leadership, Ruvos has expanded its capabilities in cloud computing, cybersecurity, and data science, reinforcing its position as a trusted partner in public health innovation.

Welcome!

Whether you are contemplating embarking on a data modernization project for your public health jurisdiction, or are already deep in the weeds of this process, this document is for you.

It lays out an ambitious but attainable vision of what a modern public health data system should optimally look like and provides time-tested goals and guidelines for getting there.

This includes advice on undertaking an assessment of the current state of your data systems and what to look for in understanding the baseline upon which you will be building.

It provides a framework for analyzing and prioritizing among the many options available, so that your plans are well thought out, and you have the best assurance that your efforts will provide optimal value and alignment.

The authors are happy to discuss any questions you may have. Feel free to reach out at hello@ruvos.com.

A practical, field-tested **toolkit** is available, free of charge, as a supplement to this document to assist in your data modernization planning. It provides a working spreadsheet that includes a register for recording the current state of your data systems and a matrix to facilitate scoring and prioritizing your options in accordance with the modernization framework, with instructions and examples. Reach out to our team at hello@ruvos.com to request a copy.

In This Document...

Introduction

A Modern Architecture for Public Health

Goals of the Modern Architecture Model

1 – Adopt a Data-Centric Approach

2 – Support Medicaid Enterprise Systems

3 – Adhere to an Advanced Data Maturity Model

4 – Integrate Operational and Reporting Data Lifecycles

5 – Leverage National Networks

6 – Integrate Modern Products and Tools

Components and Technical Capabilities of the Model

Data Lakehouse and Pipelines Layer

Data Enrichment Layer

Data Analytics and Visualization Layer

Data Operations and Presentation Layer

Baseline Systems Assessment

A Comprehensive System Definition

System Registry

In-scope / In-scope-ancillary / Out-of-scope

Dataset Category

Hosting Location

Software Type

System Profile

Modernization Analysis and Prioritization Framework

Prioritization Matrix

Scoring Rubric – Value, Effort, and Alignment

Wildcard Factor

Value and Effort

Cross-Program Insight Value

Workforce Automation Value

Data Quality Value

Technical Architecture Effort

Data Sharing Effort

Dataset Dependency Effort

Modern Architecture Alignment

Modernization Area 1: Scalable Data Pipelines

Modernization Area 2: Data Lakehouse

Modernization Area 3: Secure Data APIs

Modernization Area 4: Shared Data Quality and Analytics

Modernization Area 5: Applications in the Cloud

Additional Support – Toolkit

Introduction

Today's public health data systems face significant challenges in managing the increasing complexity, volume, and diversity of health data. Many existing data systems are outdated, siloed, and system-centric, leading to inefficiencies in data collection, cleaning, and analysis. These limitations hinder the ability to generate real-time, actionable insights critical for responding to public health emergencies, supporting Medicaid Enterprise Systems, and meeting national reporting standards such as those mandated by the Centers for Disease Control and Prevention (CDC). Furthermore, the lack of interoperability across systems exacerbates delays, duplication of effort, and data inconsistencies, ultimately impacting the quality and timeliness of public health decision-making and service delivery.

To address these challenges, public health agencies at every level acknowledge the need for data modernization, but experienced guidance for planning such initiatives is in short supply. Moreover, the circumstances of each jurisdiction are unique, requiring unique solutions. Allocating limited funding amid uncertainty about which investments will align to achieve the intended progress is a challenge for many jurisdictions.

There is consequently a pressing need for an analytical framework that can guide and empower jurisdictions to plan and implement modernization efforts with maximum impact. This paper presents such a framework, based on extensive experience assisting state agencies, large and small, in assessing their status and determining the most effective avenues for successfully advancing their modernization initiatives. It provides a tested methodology that will assist you assess, analyze, and prioritize your jurisdiction's public health modernization efforts in line with industry standards and regulatory requirements.

There are three key sections to the document:

- **A Modern Architecture for Public Health:** This section presents an advanced conceptual architecture detailing the essential technology and software components necessary for modern public health solutions.
- **Baseline Systems Assessment:** Provides a comprehensive evaluation process for existing systems, collecting insights into their current state and relevance within the context of the modernization initiative.
- **Modernization Analysis and Prioritization Framework**: Outlines a systematic approach for analyzing and prioritizing systems based on their alignment with modernization goals, effort required, and their overall value to the public health enterprise.

A Modern Architecture for Public Health

The modern architecture model for public health is a robust enterprise-wide conceptual architecture that outlines a comprehensive overview of the technology and software components required for public health modernization. This model aligns with industry best practices, modular Medicaid Enterprise Systems (MES), national network guidelines such as the Trusted Exchange Framework and Common Agreement (TEFCA), and the reporting requirements set by the Centers for Disease Control and Prevention (CDC). In the era of advanced cloud platforms, the complexity of technical system architectures has significantly increased. This paper aims to clarify this complexity by defining the model's goals, offering a diagram that visually organizes the components into logical layers and illustrates their interrelationships, and provides detailed definitions and capabilities for each component.

The model represents a highly desirable future state of modernization and provides a robust framework for building towards it; needless to say, it does not encompass all possible options.

Goals of the Modern Architecture Model

Public health data often arrives from multiple sources, lacks standardization, and is often incomplete. Processing this data requires a great deal of time and resources — including spending a significant amount of time cleaning data to support program functions such as case investigations or emergency responses. Historically, individual program areas manage their data from input to final output, often building their own system to support this process. However, when this process must be completed across multiple systems, it can result in unnecessary duplication of work, delays, and data inconsistencies across the public health jurisdiction.

The goals of a modern architecture model are to reduce the amount of effort needed to collect, clean, and analyze data and streamline that process across multiple programs. Within this model, cloud technology and other modern tools serve as modular components that allow for increased data processing speed and assurance that incoming and outgoing data is accurate, consistent, and secure.

The remainder of this section highlights important concepts that differentiate this modern architecture model from systems that are currently prevalent.

1 – Adopt a Data-Centric Approach

What is your program's Data Experience (DX)? Is the role of Chief Data Officer a recent addition in your organization? It is data's time to step into the spotlight!

New enterprise-wide public health data systems are increasingly becoming data-centric rather than system-centric, changing the way health data is managed and utilized. In a data-centric approach, the focus is on the dataset itself, rather than the individual systems that store and process it. This means that data is standardized, integrated, and made accessible across various platforms and applications, ensuring interoperability and seamless data flow. Such systems prioritize the creation of a unified data model that supports diverse data types and sources, enabling comprehensive data analysis and reporting. This approach facilitates better decision-making by allowing health professionals to access real-time, accurate, and complete data, regardless of the original system it resides in. In contrast to traditional system-centric models, which often result in data silos and fragmented information, data-centric public health systems enhance data sharing, collaboration, and transparency. This shift not only improves the efficiency and effectiveness of public health responses but also supports advanced analytics, predictive modeling, and research, ultimately leading to better health outcomes and more informed policy decisions.

2 – Support Medicaid Enterprise Systems

Public health modernization initiatives play a critical role in supporting Medicaid Enterprise Systems (MES) and ensuring the effective delivery of Medicaid benefits in states. MES are the evolution of Medicaid Management Information Systems (MMIS) and seek to leverage modern technologies to enhance the overall efficiency and effectiveness of Medicaid administration. These initiatives involve updating and integrating public health data systems, enhancing data interoperability, and adopting advanced analytics to provide real-time insights into population health trends. By modernizing public health infrastructure, states can improve the accuracy and timeliness of data used by MES, enabling more efficient eligibility determination, claims processing, and care management. Enhanced data sharing and interoperability facilitate better coordination of care, particularly for vulnerable populations, ensuring that Medicaid dollars are used more effectively to improve health outcomes. Moreover, the integration of advanced analytics helps identify and address health disparities, optimize resource allocation, and support preventive care initiatives, ultimately leading to a more efficient and responsive Medicaid program that can adapt to evolving public health challenges.

3 – Adhere to an Advanced Data Maturity Model

In the era of digital transformation, public health systems are increasingly reliant on robust data management practices to enhance decision-making, improve patient outcomes, and streamline operations. An advanced data maturity model offers a structured approach to elevate data quality and usability, vital for public health modernization initiatives. Inspired by the Data Management Association (DAMA) International's Data Management Body of Knowledge (DMBOK)[1], we propose a simplified model that outlines progression through four key stages: raw, cleaned, enriched, and consumable.

- **Raw Data:** Unprocessed and unrefined data collected from diverse sources such as hospitals, clinics, labs, and electronic health records (EHRs), often including inconsistencies, duplicates, and errors.
- **Cleaned Data:** Data processed to remove inaccuracies, inconsistencies, and duplicates, ensuring data quality and integrity crucial for public health analyses.
- **Enriched Data:** Data augmented with additional context, metadata, or external data sources, providing more insights and making it more meaningful for public health initiatives.

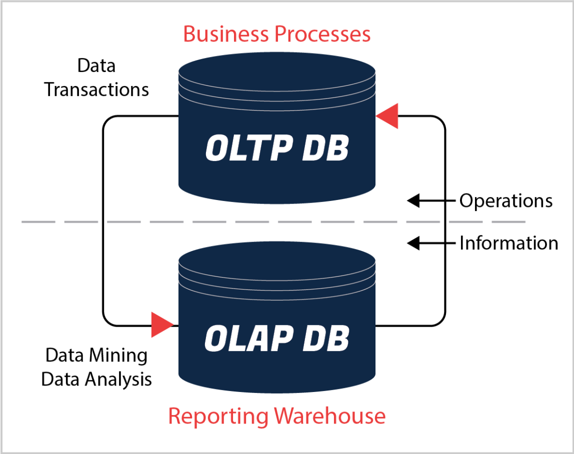
[1] https://www.dama.org/cpages/body-of-knowledge

- **Consumable Data:** Fully prepared for analysis, reporting, and decision-making, characterized by high usability, accessibility, security, and thorough documentation, making it accessible and understandable to public health officials, policymakers, and analysts.

An advanced data maturity model is essential for public health systems aiming to maximize the value of their data assets. By progressing through the stages of Raw, Cleaned, Enriched, and Consumable data, public health organizations can transform raw data into powerful insights, driving informed decision-making. As public health data grows in volume and complexity, adopting a structured data maturity model will be critical. Public health modernization initiatives that embrace this model will be better equipped to respond to health crises, improve patient outcomes, and optimize resource allocation.

4 – Integrate Operational and Reporting Data Lifecycles

When leveraging a data maturity model, recent advances in data storage and computing power allow modern cloud architectures to simultaneously handle both operational and reporting capabilities.

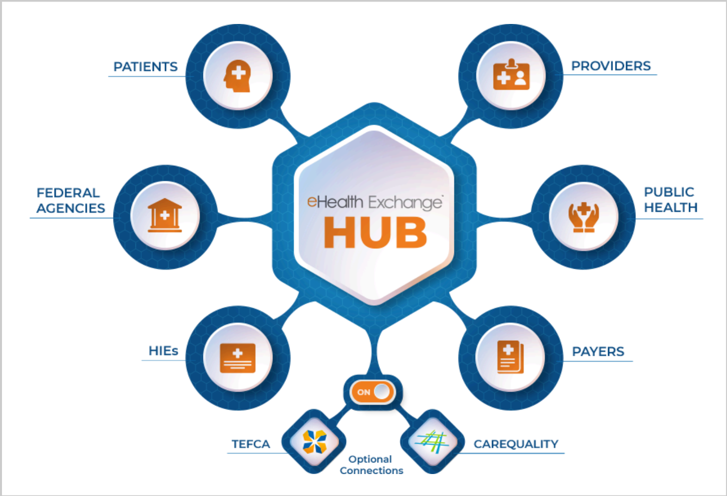


This represents the integration of historically separate Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) data lifecycles. In this new DX architecture, raw transactional data from OLTP applications is ingested into the data lakehouse in real-time or near-real-time, ensuring up-to-date operational data is readily available. This data is stored in its native format, allowing for flexible schema-on-read processing, which facilitates quick adaptation to changing data structures. Simultaneously, the data lakehouse supports OLAP through event-driven architecture that rapidly transforms, batch-processes, and aggregates data needed for complex analytical queries.

Advanced analytics, machine learning models, and BI tools can access the cleansed and enriched data directly from the data lake, promoting comprehensive reporting and insights generation. This convergence of OLTP and OLAP within a unified cloud-based data lake architecture ensures streamlined data management, reduces latency in data availability, and provides a holistic view of business operations, supporting both real-time decision-making and long-term strategic planning.

5 – Leverage National Networks

Public health data modernization should strategically leverage existing national networks like the eHealth Exchange[2] and the Association of Public Health Laboratories (APHL) AIMS Platform[3] to enhance data sharing, interoperability, and real-time public health responses. These networks, already integrated with many public health systems, complement the national infrastructure provided by the CDC, which serves as a critical hub for data collection, analysis, and dissemination. The eHealth Exchange connects healthcare providers, government agencies, and non-profit organizations, facilitating the seamless exchange of critical health information across jurisdictions. By integrating with this network, public health systems can ensure comprehensive, up-to-date patient information, improving disease tracking and management. Similarly, the AIMS Platform links state and local public health laboratories which provides the necessary infrastructure for the rapid exchange of laboratory data, enabling timely identification and response to emerging public health threats. Leveraging these established networks and existing CDC infrastructure not only enhances the efficiency of data exchange but also strengthens the capacity for coordinated, nationwide public health initiatives.



6 – Integrate Modern Products and Tools

The conceptual model presented here outlines the essential technical capabilities required to implement an advanced, data-centric public health solution. It is vendor agnostic; preparing for its practical execution will involve evaluating various vendors whose software solutions overlay and intersect with the data lakehouse model in different ways, providing a diverse array of functionalities. Each vendor may focus on different aspects, such as data ingestion, storage, processing, analytics, master data management (MDM), or visualization, contributing to the overall ecosystem. This diversity means that public health organizations will need to carefully select and integrate products to ensure compatibility and coherence within their data lakehouse architecture. The success of such a system depends on the seamless integration of these disparate technologies, each offering unique strengths, to create a unified, efficient, and comprehensive public health data solution.

[2] https://ehealthexchange.org/
[3] https://aimsplatform.com/

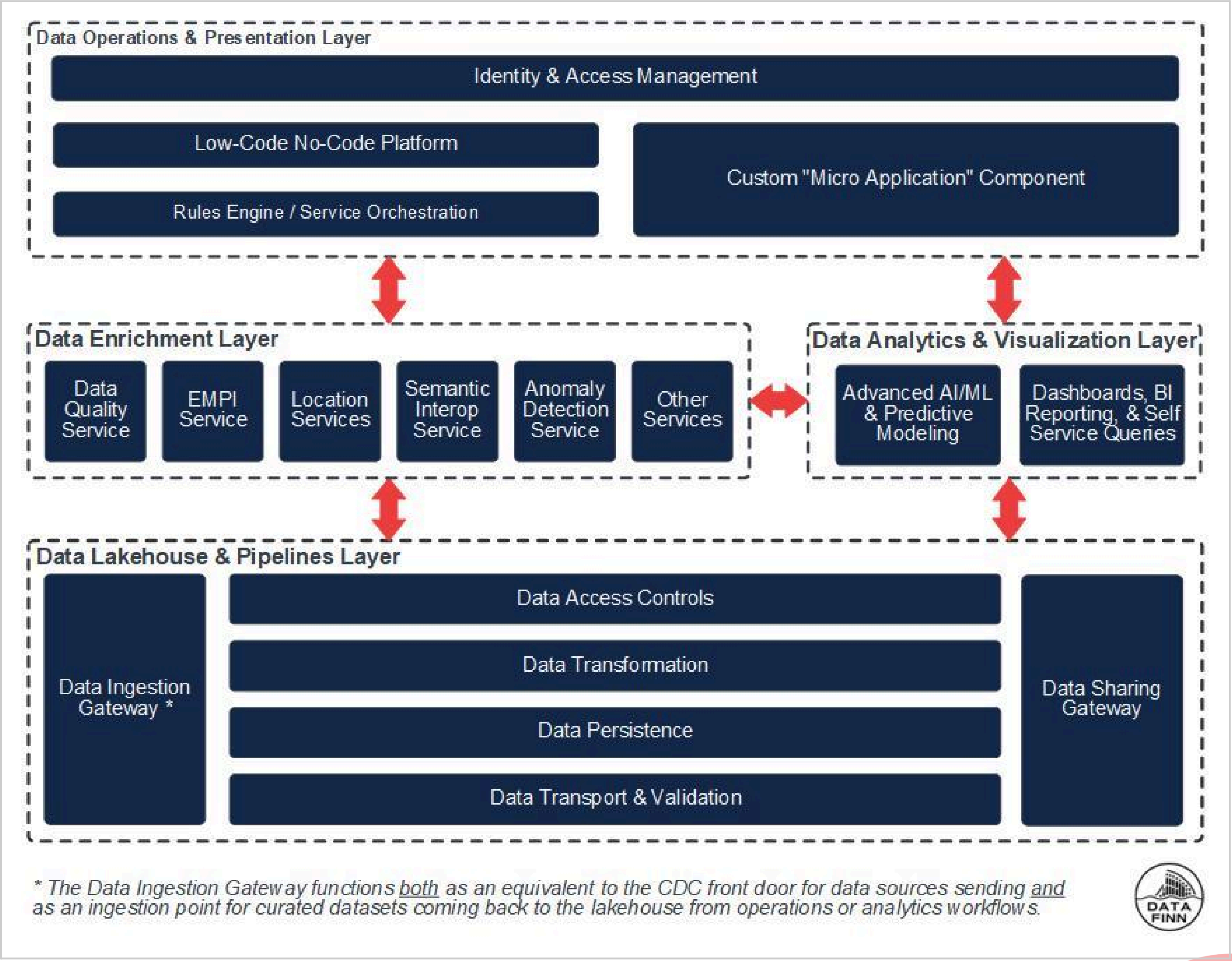
Components and Technical Capabilities of the Model

This modern architecture model is organized as an enterprise-wide stack with four layers:

- **Data Lakehouse and Pipelines**
- **Data Enrichment**
- **Data Analytics and Visualization**
- **Data Operation and Presentation**

The structure of the model is depicted in Figure 1.

Figure 1: Public Health Modern Architecture Model



The model's layers offer a logical organization of the components required for the bi-directional, and oftentimes iterative, journey of data through the modern architecture model as it moves from a raw state to more consumable forms that are utilized in specific program areas. This provides more timely and accurate data for system operations, as well as consumable data for enterprise-wide analytics and reporting.

The following sections offer a comprehensive exploration of the four layers of the modern architecture model, noting the distinct characteristics and technical functionality of each. These components serve as the tools and services necessary for delivering optimal functionality within the architectural framework.

Data Lakehouse and Pipelines Layer

A data lakehouse blends the features of a data lake and a data warehouse, allowing for storage of vast amounts of unstructured, semi-structured, and structured data that can be accessed and leveraged for multiple purposes. In the data lakehouse, Artificial Intelligence/Machine Learning (AI/ML) and other cloud-services can be leveraged for data quality and data enhancement. Data pipelines and ingestion gateways are responsible for transporting data to and from the data lakehouse.

Components of the data lakehouse and pipelines layer
(*Note that the order of the components aligns with how data generally flows through the architecture layer*):

Component	Definition and Technical Capabilities
Data Ingestion Gateway	• Facilitates the process of collecting, importing, and transferring data from various sources into a target data storage or processing system. • Simplifies and streamlines data ingestion, enabling organizations to efficiently receive data from diverse origins and formats. • Unifies data collection interfaces for enhanced efficiency.
Data Transport and Validation	• Ensures efficient data movement between systems and compliance with quality and integrity standards. • Applicable in domains such as database management, ETL/ELT workflows, and data integration for seamless data transfer.

Component	Definition and Technical Capabilities
Data Persistence	• Responsible for storing, retrieving, and managing data in a persistent manner (i.e., not power dependant — it's still there after restarting). • Stores and manages information over time, like databases, data warehouses, cloud file stores, or file systems — with attention to auditing and security requirements.
Data Transformation	• Converts data from its original format or structure into a standards-based format suitable for consumption, analysis, reporting, and decision-making. • Streamlines the restructuring of diverse health datasets for enhanced interoperability among different health data systems.
Data Access Controls	• Authorizes the retrieval, manipulation, and management of data from various sources. • Provides a standardized interface or set of application programming interfaces (APIs) for interacting with and accessing data.
Data Sharing Gateway	• Establishes data governance approved availability of consumable data to partners in its original format into a standards-based format suitable for analysis, reporting, and decision-making. • Streamlines the sharing of diverse health datasets for enhanced interoperability among different health data systems.

Data Enrichment Layer

The data enrichment layer integrates cloud and third-party service components to significantly enhance data quality, making it versatile for use across various program areas. Key considerations for understanding and utilizing shared data enrichment services include:

- **Efficiency Gains**: Leveraging shared data enrichment services reduces the workload for each program area by centralizing tasks related to data cleaning and preparation. Automated processes within these services further streamline data enhancement, promoting consistency and accuracy across systems.
- **Selective Usage**: Not all program areas will adopt shared data enhancement services, depending on their specific needs and compatibility with existing systems.
- **Privacy and Compliance**: Certain datasets may be inaccessible to some enhancement services due to strict privacy regulations. Compliance with laws such as the Health

Insurance Portability and Accountability Act (HIPAA), Health Information Technology for Economic and Clinical Health (HITECH) Act, Trusted Exchange Framework and Common Agreement (TEFCA), and state-specific legislation is crucial in determining data availability and usage.

Components of the data enrichment layer

Component	Definition and Technical Capabilities
Data Quality Service	A suite of tools, processes, and methodologies aimed at enhancing the quality of data within an enterprise system.
Enterprise Master Person Index (EMPI) Service	A system designed to manage and maintain a centralized index of person level information across multiple domains. It ensures data accuracy, links, and reconciles records across different systems and databases.
Location Services	Technologies that provide address standardization, validation, and geocoding, ensuring accurate and consistent location data.
Semantic Interoperability Service	Technologies, standards, and processes that ensure different information systems can accurately exchange data with standardized meanings of data elements, formats, and terminologies.
Anomaly Detection Service	Techniques in data analysis and machine learning used to identify patterns, behaviors, or data points that significantly deviate from expected norms within a dataset. Commonly employed in public health for early detection systems, system monitoring, and quality control.
Other Services	Modular shared data analytics components that can be integrated into the modern architecture model over time, providing additional value across various data domains within the enterprise.

Data Analytics and Visualization Layer

The data analytics and visualization layer incorporates a diverse array of tools, platforms, and solutions designed to empower users in creating insightful visual representations of data. These visualizations facilitate the identification of patterns and effective communication of information, empowering informed decision-making. This layer encompasses stand-alone tools as well as integrated platforms that support comprehensive data sources, analytics capabilities, and collaborative features. Leading examples of data visualization solutions include Tableau, Power BI, and AWS QuickSight, which streamline the process of transforming complex data into actionable insights.

Additionally, AI/ML services are integral components within this layer, facilitating the integration of advanced artificial intelligence and machine learning capabilities. These services enable developers and enterprises to leverage AI algorithms for tasks such as predictive analytics, anomaly detection, and natural language processing. By harnessing AI/ML, organizations can improve data quality, extract deeper insights from data, automate decision-making processes, and enhance operational efficiency across various domains.

Components of the data analytics and visualization layer

Component	Definition and Technical Capabilities
Advanced AI/ML and Predictive Modeling	• Tools and platforms that facilitate the integration of AI/ML capabilities into applications, products, or processes. This includes model training datasets and pre-built models to streamline development for enterprises and developers. • AI algorithms are utilized to generate insights, predictions, and data-driven solutions for public health challenges. These services can synthesize large and diverse datasets to identify trends, forecast public health scenarios, and personalize health interventions.
Dashboards, BI Reporting, and Self-Service Queries	Modular shared data analytics components that can be integrated into the modern architecture model over time, providing additional value across various data domains within the enterprise.

Data Operations and Presentation Layer

The data operations and presentation layer encompass both the user interface/user experience (UI/UX) as well as the business logic and operational functionality within the modern architecture model. This layer is integral to the transactional data processing pathway (i.e., OLTP) within the enterprise-wide architecture. It can be implemented using a custom micro-application component approach, a low-code/no-code product approach, or a hybrid of both. These approaches utilize data exposed by the underlying layers and should consistently leverage the same identity and access management component to ensure seamless access control and security.

Moreover, refined and processed data generated through operational business rules can and should be fed back into the data lakehouse. This integration supports OLAP reporting and greatly enhances the data analytics and visualization layer of the architecture, enabling comprehensive data analysis and richer insights. By continually enhancing the data lakehouse with operational and outcomes data, this model fosters a robust and dynamic analytical environment, improving overall user experience, operational efficiency, and adaptability of the system.

Components of the data operation and presentation layer

Component	Definition and Technical Capabilities
Identity and Access Management (IAM)	Also referred to as user authorization and management, encompasses technologies for user authentication, authorization, access control, user management, password policies, single sign-on (SSO), and role-based access control. Ensures secure and efficient management of user identities and permissions.
Low-Code No-Code Platform	● Platforms that simplify the application development process by minimizing traditional coding requirements. ● Designed to enable rapid creation of functional software applications. ● Often include a range of pre-built components, templates, and workflows that users can assemble to create applications efficiently.
Rules Engine / Service Orchestration	● Software component designed to manage and execute business rules in an automated and consistent manner. ● Commonly used to separate business logic from application code, promoting flexibility and maintainability. ● Refers to the coordination and arrangement of multiple, often distributed, services to work together seamlessly to achieve specific business processes or workflows. ● Involves managing the flow of data and timing of execution across different services to deliver desired functionality efficiently.
Custom "Micro Application" Component	Refers to a custom-built micro application layer such as a .NET C#, Python, or Node.js Express web application. These are small business tier and front-end pieces of software that sit on top of the lower layers, tailored to meet specific business requirements and provide bespoke functionalities.

Baseline Systems Assessment

Conducting a current state systems assessment serves as a necessary entry point for public health jurisdictions to catalog the systems that support their public health initiatives. Establishing and documenting a comprehensive baseline is crucial for guiding future improvements and transformations. This process includes collecting quantitative data points to provide understanding of the existing systems — it is these insights that will be used for analyzing then prioritizing systems and datasets for migration to the modern architecture.

The subsequent sections provide a comprehensive system definition and introduce the concepts of a system registry and system profile which can be used to effectively document the current state of systems.

A Comprehensive System Definition

What exactly constitutes a “system” in public health? Based on our experience in both public health and general health IT, the model relies on the following definition of what should be considered a system.

A system encompasses the combination of front-end applications, databases, services, websites, and other components that run on computers or smart devices to support program areas in their business functions. Often, a system also relies on cumulative knowledge of program staff as the “core database” and incorporates "manual tools" such as post-it notes, fax machines, mail, phone calls, and similar methods. A comprehensive system includes inputs, outputs, data, actors, and workflow processes (i.e., a series of actions or steps taken to achieve a particular end) necessary for the operation of the program area.

Understanding the current state of systems, especially those that are less technically advanced, is crucial for several reasons. First, transitioning these systems to a modern architecture can offer a high value-to-effort ratio. Second, important datasets from an enterprise-wide public health perspective may be overlooked. Historically underfunded program areas that never had budget for their own stand-alone systems can thrive in a modern architecture model by leveraging shared components and relatively inexpensive, secure data storage. Furthermore, incorporating the datasets from these programs enhances the value of the data lakehouse and enriches public health reporting insights across the jurisdiction.

System Registry

The system registry provides a comprehensive organized list of the public health systems within the project. This register includes crucial information such as if a system is in scope, its associated dataset category, the systems hosting location and software type classifications, along with other relevant details. A working template of the system registry can be found in the Toolkit.

In-scope / In-scope-ancillary / Out-of-scope

Within the registry, systems should be marked as "**in-scope**," "**in-scope-ancillary**," or "**out-of-scope**," based on their alignment with project objectives and dependencies.

In-scope systems are directly essential to core program area functions such as vital record statistics, disease surveillance, or emergency medical services (EMS).

In-scope-ancillary systems play a supportive role and are generally national systems used by an in-scope system. Typically, in-scope-ancillary systems cannot be changed as they support many public health agencies, and the in-scope systems must connect to them. Examples of in-scope-ancillary systems would be Electronic Case Reporting (eCR) on the APHL Informatics Messaging Services platform (AIMS)[4] or the State and Territorial Exchange of Vital Events (STEVE) hosted by the National Association for Public Health Statistics and Information Systems (NAPHSIS.)[5]

Out-of-scope systems are deemed irrelevant or not critical to public health modernization goals.

This classification system enables clear understanding and strategic allocation of resources, ensuring efficient management and alignment of technology investments with overarching public health priorities.

Dataset Category

Systems are used to create and manage datasets — the **system dataset category** is designed to identify the type of public health data a system works with. Datasets in public health are determined by their public health function and then by program area and funding source. Determining and associating a dataset category will be beneficial to scoring value and prioritization grouping of systems during analysis.

The following dataset categories are typical for data in public health and can be refined by your jurisdiction as needed:

- Advance Directives
- Cancer Registry
- Children's Special Health Program
- Early Childhood Screening Programs
- Emergency Medical Services (EMS)
- Emergency Preparedness and Response
- Environmental Health Programs
- Healthcare-Associated Infections Program
- HIV Care / Ryan White Program
- Immunization Registry
- Integrated Disease Surveillance System
- Prescription Drug Monitoring Program (PDMP)
- Prevention Programs
- Public Health Lab
- Rural Health Programs
- Vital Records
- Women Infant Child (WIC)

[4] https://ecr.aimsplatform.org/about-us
[5] https://www.naphsis.org/get-vital-records/for-work/automated-reports

Hosting Location

The **system hosting location** is a classification assigned to a system that defines where data is physically located and must be accessed. Each system hosting location classification comes with its own set of factors influencing how users interact with the system and will be useful during analysis and scoring. State-hosted data may be more accessible, but integration capabilities might be limited by the availability of state resources. CDC-hosted data typically comes with its own set of restrictions on data usage. Third-party hosted systems often necessitate contract amendments, potentially incurring significant charges from the host for integration, if data sharing is permitted at all. End-User Developed Applications (EUDA) systems are frequently undocumented and rarely have disaster recovery plans, posing additional challenges.

System hosting classifications

Hosting Location Classification	Description
Third-Party	Systems that are proprietary, hosted, and managed by the vendor.
CDC	Systems housed and managed at the CDC, generally within the Amazon Web Services (AWS) environment.
State Data Center	Systems housed and managed at a state-run datacenter.
End-User Developed Applications (EUDA)	Systems run by state staff on computers within their program areas.

Software Type

The **system software type** classification identifies the origin of the source code, providing insight into the programming language used, who manages the software, and how it may be maintained. Similar to hosting location, having a system's software type classification will be helpful during the analysis and scoring process.

System software classifications

Software Type Classification	Description
Third-Party Vendor	Proprietary software that must be managed and maintained by the vendor.
Mobile App	A specific type of third-party software designed to run on mobile devices (though users can also access the system via a web browser). For this analysis, these are generally public-facing applications.
CDC Software	Microsoft stack application software provided by the CDC, utilizing MSSQL and the .NET framework.

Software Type Classification	Description
In-House	Software created by the custom application development team at the state.
End-User Developed Applications (EUDA)	Software developed by state staff within each program area using various available tools.

System Profile

In addition to the system registry, each public health system should have an overview including insights into its value and specifics on where and how it is used. To facilitate this process and account for similarities within program areas, a "system profile" should be created as part of the assessment. Content for the system profile should be pulled from existing system documentation, industry best practices, subject matter expertise, and insights gathered from stakeholder interviews.

The following list offers a baseline of the areas you should document for each system in its profile:

- **Description and Purpose:** Describes the business function of the system.
- **System Category**: Identifies the area of public health the system and dataset are in (e.g., immunization registry, emergency medical services, environmental health, prevention and outreach, disease surveillance, vital statistics, etc.)
- **Stakeholders**: Identifies the system owners and program area users.
- **Key Infrastructure Points**: Lists the national data frameworks and standards used by the system, the actors and vendors involved, and relevant system resource links.
- **Datasets and Data Flows**: Provides logical data flow diagrams that show the actors involved (i.e., people), inputs, outputs, and workflow processes around the data.
- **User Feedback**: Documents user feedback and experiences with the system.
- **Resource Links**: A list of web resources (URLs) that can be referenced for additional information regarding this system.

Modernization Analysis and Prioritization Framework

This section presents a systematic methodology to analyze and prioritize the in-scope systems identified in the current state assessment. The analysis framework comprises three key components: a **prioritization matrix**, qualitative **value and effort** factors, and **modern architecture alignment**. The prioritization matrix provides a structured approach to rank systems for modernization initiatives based on scores derived from these qualitative factors and their alignment with the modern architectural model.

The following *Prioritization Matrix* section introduces the scoring structure and point values while the Value and Effort and the Modern Architecture Alignment sections describe each item used in the matrix and how to rate a system for that item by assigning points.

Prioritization Matrix

The prioritization matrix is a spreadsheet tool designed to assist in evaluating a system's value relative to the effort required to modernize it, as well as its alignment with the modern architecture model. The matrix begins with the in-scope systems identified in the system registry (collected during the current state assessment) and calculates a total score for each system based on qualitative factors and modernization area alignment. This process results in a ranked list of systems, highlighting their potential for modernization and their perceived benefit to agency functions. The higher the score, the higher the system's prioritization. Both a working template of the prioritization matrix and an example scoring for a hypothetical system can be found in the toolkit section.

Scoring Rubric – Value, Effort, and Alignment

The scoring structure of the prioritization matrix consists of five modernization areas to score a systems alignment to the modern architecture model, three value factors to score public health value, and three effort factors to score effort to modernize. Factors that add value are "positive" and increase score while factors that increase effort are "negative" and decrease score. The scoring rubric outlined in Figure 2 provides point values and associated ratings.

Figure 2: Prioritization Matrix Scoring Rubric

	Modern Architecture Alignment *(MA-1 to MA-5)*		**Value Factors *(V-1 to V-3)***		**Effort Factors *(E-1 to E-3)***	
Scoring Rubric	**10**	Aligns with the modernization area	**10**	High value to public health	**-10**	High effort required to work on the system
	5	Partially aligns with the modernization area	**5**	Medium value to public health	**-5**	Medium effort required to work on the system
	0	Does not align with the modernization area	**0**	No value to public health	**0**	No effort required to work on the system

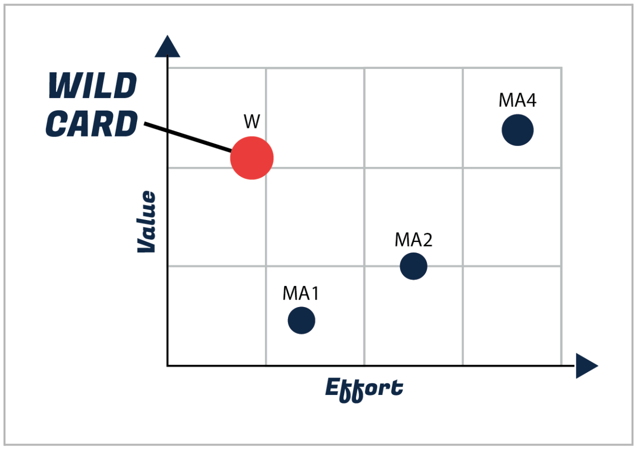
Guidelines for scoring systems within the Prioritization Matrix as part of the Public Health Data Modernization Framework.

Follow these guidelines when scoring systems in the prioritization matrix:

- Within the prioritization matrix, a system's alignment to each of the suggested modernization areas (MA), identified as MA1 through MA5, is rated and summed to calculate the total MA alignment.
- The total MA alignment score is combined with the rating from the three value factors identified as V1 through V3 and the three effort factors identified as E1 through E3 to create an overall score.
- The overall score is increased through alignment points from total MA alignment and higher ratings in value factors (positive points), reflective of an increase in overall value and modernization ability.
- The overall score is decreased by higher ratings in effort factors adding negative points and reflecting anticipated challenges to the modernization efforts.
- A higher overall score reflects a higher priority to enterprise-wide modernization efforts.

Wildcard Factor

The wildcard factor is a flexible component designed to adjust the overall score by adding or subtracting points based on specific, contextual considerations that may not be fully captured by standard factors. This allows for nuanced scoring that accounts for unique attributes or exceptional circumstances related to a system's modernization potential.



How to use the wildcard factor:

- **Identification**: Determine if there are unique circumstances, strategic priorities, or other exceptional factors that warrant an adjustment to the standard scoring. This could include anticipated regulatory requirements or significant operational impacts not covered by existing factors. For example, a system may have the highest overall score, but does not fully align with MA1. The authors would advise beginning with a few systems that fully align with the first two foundational modernization areas. Additionally, you may find a system that has a low effort level and high value that should be considered in the early phases of modernization ahead of systems that align more fully for a quick win. Jurisdictions must also pay close attention to the public health needs of its population as well as the cultural and political desires of their region.
- **Adjustment**: Decide whether the wildcard factor should add or subtract points from the overall score. Positive or negative adjustments may be warranted to alter a system's prioritization.

- **Application**: Apply the wildcard factor adjustment after calculating the Total MA Alignment and qualitative factor scores. Ensure that the adjustments are documented, providing a clear rationale for the changes made.
- **Consideration**: Use the wildcard factor sparingly and ensure it aligns with the overall objectives of the modernization initiative. Avoid over-reliance on this factor to ensure that the prioritization remains objective and data driven.

The wildcard factor helps to incorporate insights and strategic considerations into the prioritization process, ensuring a more comprehensive evaluation of each system's modernization potential.

Value and Effort

The prioritization matrix relies on three value factors (positive points) and three effort factors (negative points). This section focuses on defining these factors with consideration to the scoring rubric covered in the *Prioritization Matrix* section earlier. Factors are tailored to public health and designed to provide valuable insights into a system and dataset. They are derived from a national perspective based on subject matter experts and industry expertise. Note that these factors can be adjusted by your jurisdiction to incorporate other priorities, such as costs, legislative requirements, or permissions, ensuring the assessment is aligned with local needs and objectives.

Cross-Program Insight Value

The cross-program insight value factor is meant to capture if a system's dataset has broad public health value across program areas. Consider if integrating the dataset into the modern architecture would enable data consumption by multiple program areas/agencies or third parties (e.g., payers, researchers, judges, other parties). How many public health functions depend on data captured or generated by the system? Is there a high overlap between program recipients and state Medicaid or other benefit programs? Dataset value can be based on stakeholder function and perspective. Different stakeholders have different views of dataset value based on their area of interest (e.g., internal staff, public, legislature). The higher the counts, the higher the score.

Workforce Automation Value

The workforce automation value factor is designed to help score if modernizing the system helps to clearly eliminate manual tasks. Would it reduce data requests? Would it result in time savings for public health staff? Would it improve decision making or improve outreach and communication with the public. Are there manual steps for eligibility processing with state Medicaid or other benefit programs? The more "yes" responses compared to other systems, the higher the score.

Data Quality Value

The data quality value factor is meant to grade the quality of a system's dataset – the higher the quality, the higher the score. A score should consider data quality standards such as those outlined in the Data Management Association (DAMA) International – Data Management Body of Knowledge (DMBOK)[6]. Additionally, a score should consider stakeholder's input provided during the current state assessment regarding data quality (e.g., low manual data entry, well documented and automated processes, deduplication and other data quality processes in place).

Technical Architecture Effort

The technical architecture effort factor is designed to evaluate the technical expertise required to work with the existing system during modernization. The hosting location and software type classifications captured in the System Registry during the current state assessment can be useful in helping to determine a score for this factor. If a system's existing infrastructure results in challenges or increased scope (e.g., Access databases, legacy software, poor technical documentation), then it should have a higher score.

Data Sharing Effort

The data sharing effort factor is meant to capture a score based on the complexity and volume of rules required to share a system's data as part of the modernization initiative. This may include state statutes around the dataset in addition to current data governance and data sharing agreements or the need for additional agreements. The more to do, the higher the score.

Dataset Dependency Effort

The dataset dependency effort factor is meant to grade if a dataset is dependent on coordination with a third-party vendor, or another dataset being implemented before or at the same time it is modernized. Third party hosted systems may require contract amendments for integration often result in a significant charge from the vendor and in some cases the vendor may not be willing to share data.

---

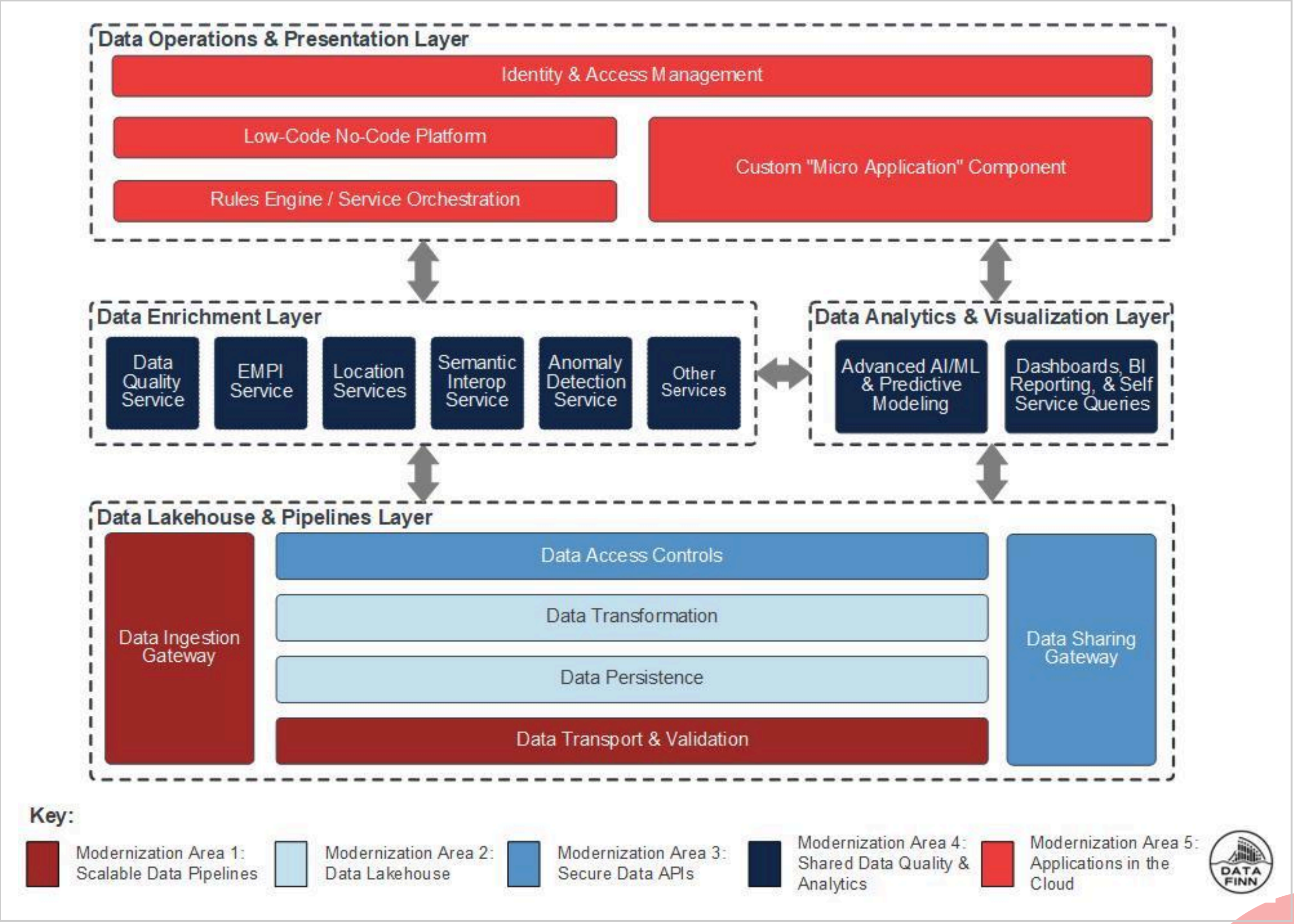
[6] https://www.dama.org/cpages/body-of-knowledge

Modern Architecture Alignment

This section focuses on technical mechanics of software and hardware technology outlining five key modernization areas essential for transitioning systems to modern architecture. Each area highlights a crucial concept of the modern architecture model and underscores its significance for the future state of public health systems. The modernization areas provide a methodology to score the alignment level of in-scope systems within the prioritization matrix. This alignment analysis facilitates a systematic process for prioritizing systems, based on their modernization potential and perceived benefit to agency functions.

The modernization areas are strategically ordered to reflect their sequential relationships. The technical components within each of the modernization areas are identified by color in Figure 3. Although many of these conceptual technical components are complementary and not easily separated, the visualization aims to clarify the sequential relationships of the modernization areas, and how they overlay on the modern architecture model.

Figure 3: Modernization Areas for Modern Architecture Alignment



The following sections describe the five modernization areas and includes some considerations for scoring a system's alignment level to each.

MA1 Modernization Area 1: Scalable Data Pipelines

Modular data ingress and data quality to organize interfaces and transfer to the cloud

Scalable data pipelines enable public health agencies to efficiently ingest large volumes of data into a cloud-based data lakehouse, supporting performance monitoring and near real-time data quality improvements. This modernization area encompasses the data ingestion gateway and data transport and validation services defined earlier in this document. The data ingestion gateway simplifies and streamlines the ingestion of data into the data lakehouse, regardless of its format or origin. It allows for the integration of data from multiple sources, including public health departments, tribal agencies, healthcare providers, facilities, and more. The data transport and validation services ensure that the data meets quality and integrity standards, removing inconsistencies and minimizing the risk of errors. Over time, these pipelines will reduce the need for point-to-point connections via interface engines like Rhapsody, centrally absorbing related workloads (e.g., data cleaning) and providing a more scalable and manageable technology solution. Data pipelines are foundational to utilizing a data lakehouse, as described in Modernization Area 2: Data Lakehouse.

Factors to help determine a systems alignment level to Modernization Area 1: Scalable Data Pipelines include:

- The system receives large amounts of data from diverse sources and formats.
- The system currently relies on point-to-point connections that could benefit from centralization.
- The system requires real-time or near real-time data processing.
- ***Conversely,*** systems requiring significant manual intervention or frequent updates, which could negate the efficiency gains from automated data pipelines may not be a good fit.
- Systems can be labeled as partial if they don't directly receive data via scalable pipelines but benefit from data enhancements within the lakehouse through datasets derived from other systems.

MA2 Modernization Area 2: Data Lakehouse

Easily and securely store raw unstructured and structured data enterprise-wide

The data lakehouse provides long-term storage and consolidation of vast amounts of data from multiple sources within a cloud platform, supporting data cleaning, enhancement, reporting, and visualization across the enterprise. This modernization area includes the data persistence and data transformation services defined earlier. The data persistence service manages the storage and retrieval of data over time, while the data transformation service converts data into usable formats for various program areas, reducing duplicative efforts for analysis and reporting through secure, scalable data APIs, as discussed in Modernization Area 3: Secure, Scalable Data APIs. The lakehouse enables centralized storage accessible to all authorized program areas, saving time by allowing reuse of data enhancements.

Before implementing a data lakehouse, it is crucial to document the datasets of systems under consideration for modernization, considering how and when the data is stored throughout their operational and reporting life cycles. For datasets that cannot be collected via scalable data pipelines as described in Modernization Area 1: Scalable Data Pipelines, a methodology should be developed to integrate these datasets into the data lakehouse. This methodology may include adding an API to allow data to remain within the source system or implementing batch data ingestion into the lakehouse via the data ingestion gateway.

The majority of public health systems will benefit from the features of the data lakehouse, however the following factors can be considered as lowering the score of certain systems:

- Legacy Constraints and Limited API Support: Lower alignment level for systems with outdated technologies or proprietary vendor formats lacking API support, complicating integration.
- Data Sensitivity and Security Concerns: Systems handling highly sensitive data that may face regulatory or security challenges should be scored lower.
- Systems that have datasets which are tied to a physical on-premise device or have datasets that are managed outside of public health should reflect a no alignment score.
- Systems should be scored lower if their datasets are created via manual program workflows that are not conducive to cloud architectures.

MA3 Modernization Area 3: Secure Data APIs

Modular data egress between the Data Lakehouse and program systems and national networks

Secure Data APIs enable public health jurisdictions to share data from the data lakehouse across multiple program areas or national networks, reducing the need for separate data collection and processing systems. This modernization area includes the data access service and data sharing gateway. The data access service provides standardized APIs for authorized data retrieval and management. The data sharing gateway ensures secure, compliant data sharing between sources and consumers.

Public health departments often depend on data received and supplied to other programs (e.g., Medicaid Enterprise Systems (MES), Integrated Disease Surveillance System, Cancer Registries). Secure APIs can streamline data access and reduce manual processes, allowing for direct, secure access to centrally governed, analysis-ready datasets. APIs and the data sharing gateway also simplify reporting to national networks like the CDC. For datasets hosted by third parties, they can be imported into the data lakehouse for baseline cleaning and enhancement. Cloud platforms offer tools for API creation, but jurisdictions must define available datasets and enforce appropriate data governance.

Factors to consider when scoring systems alignment level with Modernization Area 3: Secure Data APIs include:

- **High Cross-Program and External Value**: The system handles datasets that are valuable across multiple programs and frequently shared with external entities, required for state Medicaid claims processing, or national networks, necessitating unified and secure access methods.
- **Need for Standardized and Secure Access**: The system requires standardized protocols for secure data retrieval and management, ensuring compliance with data privacy and security regulations.
- **Integration from Multiple Data Sources**: The system integrates data from diverse sources, making APIs essential for providing consistent and efficient access across these sources.
- **Streamlined National Reporting and Real-Time Access**: The system must support efficient national reporting and real-time or near real-time data access, where APIs simplify and accelerate the data retrieval and reporting processes.

MA4 Modernization Area 4: Shared Data Quality and Analytics

Modular data enrichment and linking services to harness analytics and advanced AI/ML services enterprise-wide

Data within the data lakehouse undergoes baseline cleaning to eliminate variances at the recorded level, resulting in more actionable data for program areas and population health analytics. In Modernization Area 4: Shared Data Quality and Analytics, both cloud-native and third-party services can further enhance data for improved linking and analysis. Enhanced data is then stored back in the data lakehouse for future analytics and reporting. For instance, baseline cleaning in the data lakehouse addresses issues like swapped first and last names, ensuring accurate records. The Enterprise Master Person Index (EMPI) service centralizes patient information across multiple entities, enabling data consumers to link and reconcile patient records. Location services normalize and augment geographic metadata, improving data consistency across program areas. Enriched data can then be utilized by the Data Analytics and Visualization layer for AI algorithms and advanced machine learning insights.

Note that datasets not migrated to the data lakehouse will not be immediately available for shared analytics. However, third-party solutions may enable the integration of these datasets through secure APIs for use in data enrichment or analytics layers.

Factors to consider when scoring a systems alignment level to Modernization Area 4: Shared Data Quality and Analytics include:

- **Data Linking and Reconciliation**: The system requires linking and reconciling data across multiple sources, benefiting from centralized indexing services like the Enterprise Master Patient Index (EMPI).
- **Use of Enriched Data**: The system utilizes enriched data for advanced analytics, including AI and machine learning, needing capabilities for data augmentation and normalization.
- **Geographic Metadata Utilization**: The system requires normalized and augmented geographic metadata to improve consistency and support spatial analysis across program areas.
- **Population Health Analytics**: The system contributes to population health analytics with integrated, enhanced data ready for in-depth analysis and reporting.
- **External Data Integration**: The system needs to integrate datasets from third-party sources via secure APIs to leverage shared data quality and analytics capabilities.

MA5

Modernization Area 5: Applications in the Cloud

Refactor legacy software to leverage modern architecture features and improve control of user access to data

Modernization area 5: Applications in the Cloud builds upon the first four modernization areas and, unlike previous areas, fully encompasses a single software layer of the modern architecture model — data operations and presentation. Representing the "top" of the software stack, this layer relies on data from the data lakehouse and pipelines layer while benefiting from data enrichment and data analytics and visualization tools. Additionally, as the area that exposes data to the users, it handles the critical functionality of user management via the cloud identity and access management (IAM) component. Cloud IAM solutions should be aligned to data governance and added to the tools of the data reporting and visualization layer. IAM provides an enterprise-wide structure that can control user authorization and management of data within the system, eliminating the need for each program area to manage user accounts in their own system, and improving access by program areas and external users. It is often overlooked how much this component may impact a public health jurisdiction when applied to the transactional operational systems and simplifying user management enterprise-wide across the department.

As the modern architecture model is deployed, existing application front ends must also be transitioned to the cloud. This can be achieved through a custom application component approach, a low-code/no-code product approach, or a combination of the two approaches. Low-code/no-code platforms simplify and accelerate the development process by providing pre-built components and templates that users with less “code-proficiency” can assemble to create applications. However, low-code/no-code platforms are typically associated with higher upfront costs and sacrifice flexibility in design and management because of their reliance on a third-party vendor and sometimes proprietary processes[7].

Factors to consider when scoring system alignment with Modernization Area 5: Applications in the Cloud include:

- **System Complexity**: Evaluate the intricacy of the system's business rules and processes. Systems with complex logic may benefit significantly from modernization.
- **User Interface Condition**: Assess the current state of the user interface. Prioritize systems with outdated or non-existent UIs for modernization to achieve substantial improvements with minimal effort.
- **User Management Needs**: Determine the need for centralized IAM for secure user authentication, authorization, and access control. Systems requiring robust user management will align well with cloud-based solutions.
- **Operational Efficiency Gains**: Consider the potential for enhancing operational efficiency by integrating with the cloud. Systems that can streamline processes and improve user experience should be prioritized.

[7] https://hbr.org/2021/06/when-low-code-no-code-development-works-and-when-it-doesnt

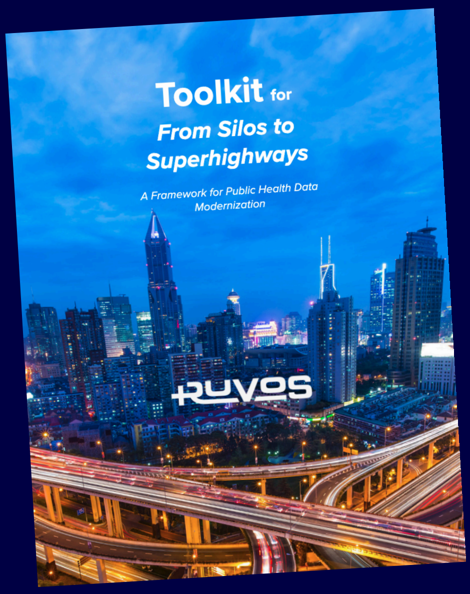
Additional Support – Toolkit

As a supplement to this document the authors have created a handy toolkit to facilitate some of the tasks prescribed in preceding sections.

It contains a working spreadsheet file that contains a system registry and prioritization matrix, essential tools for assessing and prioritizing systems during a modernization initiative. It is accompanied by detailed explanations and examples.

The system registry facilitates creation of a comprehensive inventory of systems, capturing critical information to facilitate informed decision-making. It allows you to document key details about each system, including its functionality, significance, and current state. This registry serves as the foundation for a thorough analysis, enabling stakeholders to understand the landscape of existing systems and identify those that require modernization.

The prioritization matrix provides a structured framework for scoring and ranking systems based on various qualitative factors and alignment with modern architectural principles. The tool facilitates a systematic approach to evaluating and prioritizing systems. By capturing scores for modern architecture alignment as well as value and effort factors, this matrix helps ensure that modernization efforts are focused on the most critical and impactful systems. The matrix provides a clear, transparent method to rank each system, facilitating strategic decision-making and efficient resource allocation.



Together, the system registry and prioritization matrix exemplify best practices in system assessment and prioritization, offering a practical guide for jurisdictions undertaking modernization initiatives. These tools can be customized to meet the specific needs and priorities of your organization, ensuring a tailored and effective approach to system modernization.

The toolkit is available at no charge. To request a copy, or to discuss any of the preceding content with the authors, please reach out to hello@ruvos.com.