AIMS Guidance

Centralized Public Health Data Exchanges

A Toolkit for Platform Designers, Technologists and Public Health Decision Makers

AIMS

APHL ASSOCIATION OF PUBLIC HEALTH LABORATORIES®

May 2025

Executive Summary

Harnessing the Power of Public Health Data

Reliable, secure exchange of health data is critical to public health, enabling public health practitioners, government officials and other decision makers to monitor infectious diseases, detect emerging threats and plan responses to outbreak events. Timely exchange of laboratory test results, immunization data, health records and other data across healthcare providers, laboratories, public health agencies and governments facilitates situational awareness and collaboration across entities involved in public health responses. A centralized data exchange serves as vital infrastructure in any country's or region's public health management.

The AIMS Experience

In the United States, the APHL Informatics Messaging Services (AIMS) Platform serves as the backbone of public health data exchange. The AIMS Platform connects federal agencies, state and territorial public health agencies, laboratories, healthcare providers, Health Information Exchanges (HIE) and more in a hub-and-spoke model that exchanges over a million messages each day. By providing shared services to data exchange partners for interoperability, security, hosting and visualization of electronic data and applications, the AIMS Platform facilitates rapid, reliable and scalable national data exchange.

Sharing the AIMS Platform Model

Over the past 15 years, the AIMS Platform has scaled to facilitate transport of billions of messages containing laboratory, immunization, case report and vital records data to inform decision making by public health leaders. APHL and its technical partner Ruvos have created this resource to document and distill the lessons learned through the creation, scaling and ongoing operation of the AIMS Platform. This document offers a blueprint to countries and organizations wishing to implement country- and region-wide data exchange. It provides recommendations and best practices that any jurisdiction can use in designing and building a data exchange platform to meet their specific requirements. Topics covered include:

- Architecture and infrastructure
- Standards, security, and compliance
- Data governance and policy
- Maintenance and operations
- Technical assistance and onboarding
- Stakeholder engagement
- Strategy and leadership

Extraordinary Benefits of Centralized Data Exchanges

Centralized message processing, transformation and routing reduces the resources needed by any single data exchange partner (including technical capabilities, hardware, and software) to be able to exchange health and surveillance data. A well-designed platform is secure, reliable and allows for rapid scaling. It may be progressively developed to support monitoring and auditing systems, data visualization dashboards and even hosting of third-party applications that benefit public health initiatives.

Enable Timely Data Exchange

A centralized platform can enable the timely exchange of many types of vital public health data, including:

- Pathogen test results, including influenza and SARS-CoV-2
- Vaccine-preventable disease reports
- Biological threat data
- Immunizations
- Genomic sequencing
- Birth and death records (vital events)
- Electronic test orders
- Electronic laboratory reports
- Electronic case reports

Create Efficiencies

Routing messages through a centralized platform greatly enhances users' ability to manage data exchange routes and data exchange partners. Once a data exchange partner is connected to the platform, they can send and receive with any participating trading partner already on the service. This offers considerable simplicity: rather than establishing and maintaining multiple connections between each sender and requester, a data exchange partner need only maintain a connection with the centralized platform, sending data via their preferred transport protocol.

Standardize, Translate and Transform Data

A central data exchange platform can establish and enforce data standardization, creating high quality data pipelines that can be easily ingested by disparate data systems. Such a platform provides interoperability between a variety of transport protocols, which allows a sender using one protocol to send data securely to a receiver that uses a different one.

Analytics

By building robust data pipelines of standardized data, a centralized data exchange platform provides the foundation for advanced analytics. Through dashboards and integrations with analytic tools, a centralized data platform can leverage cutting edge data science techniques to empower leaders with near real-time data to monitor trends and detect outbreaks.

A Comprehensive Resource

Informatics departments, ministries of health and public health decision makers can refer to information in this document to design a road map for establishing a data exchange platform that meets their circumstances and needs—whether hosting locally or in the cloud—and respecting applicable rules governing data location, use and vendor preference.

Contents

For more Global Health Informatics resources,
visit APHL's Global Health Informatics webpage or
contact globalhealth.informatics@aphl.org

Introduction

Document Purpose and Use

The purpose of this document is to provide a comprehensive and detailed description of the AIMS Platform, its history and use cases, the conceptual and technical foundations upon which its model is based and lastly, design considerations that can serve as a reference framework for other countries in conceptualizing, planning, and implementing a centralized public health data exchange system.

By detailing the functionality and structure of the AIMS Platform and sharing best practices and lessons learned in its growth and development over the past 15 years, this document offers a launching point that can accelerate the implementation of centralized data exchange platforms that will bolster public health globally.

This document is intended as a resource and guide for stakeholders involved in the design, planning, implementation and operation of a centralized data exchange system in their local contexts. It is divided into three sections that build on each other and can be referenced separately as a reader works through the various stages of conceptualizing a centralized data exchange platform, building the case for investment in such a system, making technical and architectural decisions and planning for the system's ongoing operation. The three sections are as follows:

- **The AIMS Platform Story:** The first section details the major functionality and use cases of the AIMS Platform, and chronicles the evolution of the Platform's architecture over the past 15 years. By detailing the AIMS Platform as an exemplar, this section intends to cast vision for centralized public health data exchange, as well as demonstrate how such a system can start simply with limited functionality and be iteratively expanded over time as utility and value are demonstrated.
- **The AIMS Model:** The second section details the foundational components of the AIMS Platform—its technology, data governance, stakeholders and operations—providing readers a view of the technical framework upon which it is built.
- **Implementing a Centralized Data Exchange System:** The final section offers considerations and best practices for designing the basic elements of a data exchange platform. This section walks the reader through detailed decision trees to aid in determining the design best suited to their specific context and needs.

Acknowledgments

This publication has been a multi-year effort, and APHL is grateful to everyone who contributed to its development, especially our partners at Ruvos their help developing this guide and CDC for their review.

Special thanks to:

- **Ruvos:** Kristin Leffel, Eduardo Gonzalez Loumiet, Jeff Couch, Zach Finn, Candice Theron, Daniel Salter
- **CDC:** Xenophon Santas, Eric-Jan Manders

The AIMS Platform Story

Origins, Evolution and Essential Use Cases

This first section describes the major functionality, use cases and major benefits of the AIMS Platform, as well as its architectural evolution over the past 15 years. This section can help cast a vision for the benefits of a centralized data exchange system and also demonstrate how such a system can start simply with limited functionality and be iteratively expanded over time as data exchange partners and stakeholders realize value from its services.

AIMS Platform Introduction

The AIMS Platform is a cloud-hosted secure data exchange platform that is compliant with United States health privacy and security regulations to facilitate the exchange of health- and public health-related data across verified data exchange partners. Owned and operated by the Association of Public Health Laboratories (APHL), a nonprofit membership organization that represents the interests of state, territorial and local public health laboratories across the US, The AIMS Platform is the US Centers for Disease Control's (CDC's) largest data sharing intermediary. Data exchange partners who send and receive data via the AIMS Platform include:

- Laboratories (both public and private)
- Federal agencies
- State, local and territorial government health departments
- Public health agencies
- Healthcare providers and health systems (i.e., hospitals, clinics)
- Military and civilian uniformed services (including the United States Navy)

Initially built to facilitate the exchange of influenza test results, the AIMS Platform has expanded to facilitate data exchange for many and varied use cases. These include:

- Influenza test results
- Immunization data
- Electronic laboratory reports (ELR)
- Electronic test orders and results (ETOR)
- Electronic case reporting (eCR) between health providers' electronic health records systems (EHRs) and public health agencies
- Genomic sequencing data
- Biological threat data
- Vaccine-preventable disease reports
- Antimicrobial resistance lab tests
- Birth and death (vital event) records

Origins

In the first decade of the 21st century many of the vital communications between public health laboratory systems in the US were still being transmitted on paper via fax or courier services.

The AIMS Platform was created to enable swift and secure communication of critical health information between public health laboratories and federal agencies in the US, with the goal of greatly reinforcing the ability of public health agencies to identify and coordinate their response to emerging threats.

The first use case of the new platform was to make it possible for the disparate health systems across the country to exchange influenza test results, starting with the need to persuade state authorities to participate in the effort.

The challenge was considerable. The US has a highly fragmented public health system, with system design and implementation decision making happening at the state rather than federal level. This fragmentation results in highly divergent systems that require considerable technical input to enable data sharing, coordination, and interoperability.

The AIMS Platform transmitted its first influenza surveillance message to CDC in 2008, and the platform has since grown to handle an extensive array of public health use cases, recently playing a key role informing the response to the COVID-19 pandemic. Today, the AIMS Platform's history and evolution provide a trove of information for burgeoning data exchange efforts.

Core Benefits

By centralizing message processing, transformation and routing, the AIMS Platform reduces the resources needed by any single trading partner (including technical capabilities, hardware and software) to be able to exchange data. In addition to message processing, it offers secure, highly available and reliable infrastructure that can be rapidly scaled to meet message volume fluctuations, as well as monitoring and auditing systems, data visualization dashboards, and hosting of third-party applications that benefit public health initiatives.

The AIMS Platform provides critical infrastructure to the US for the daily exchange of public health messaging. Because of the ongoing investment in and development of the AIMS Platform since 2008, it could be rapidly scaled to accommodate surging volumes of COVID-19 lab results and electronic case reports once the pandemic struck, providing health and government leaders with vital information to inform the public health response. To date, it has processed over 2.25 billion COVID messages.

To convey a sense of the scale of operations, in 2024 the AIMS Platform:

- **Processed 1.2 billion interface messages per month**
- **Handled 23 million messages per month between more than 8,000 data exchange partners**
- **Serviced over 7,800 eCR providers**
- **Delivered ELR submitted by several data exchange partners on behalf of hundreds to thousands of aggregated testing facilities**

For an overview of the architecture that enables the AIMS Platform to operate efficiently at scale, see "AIMS's Architectural Evolution" (page 11).

Steadily Expanding Functionality

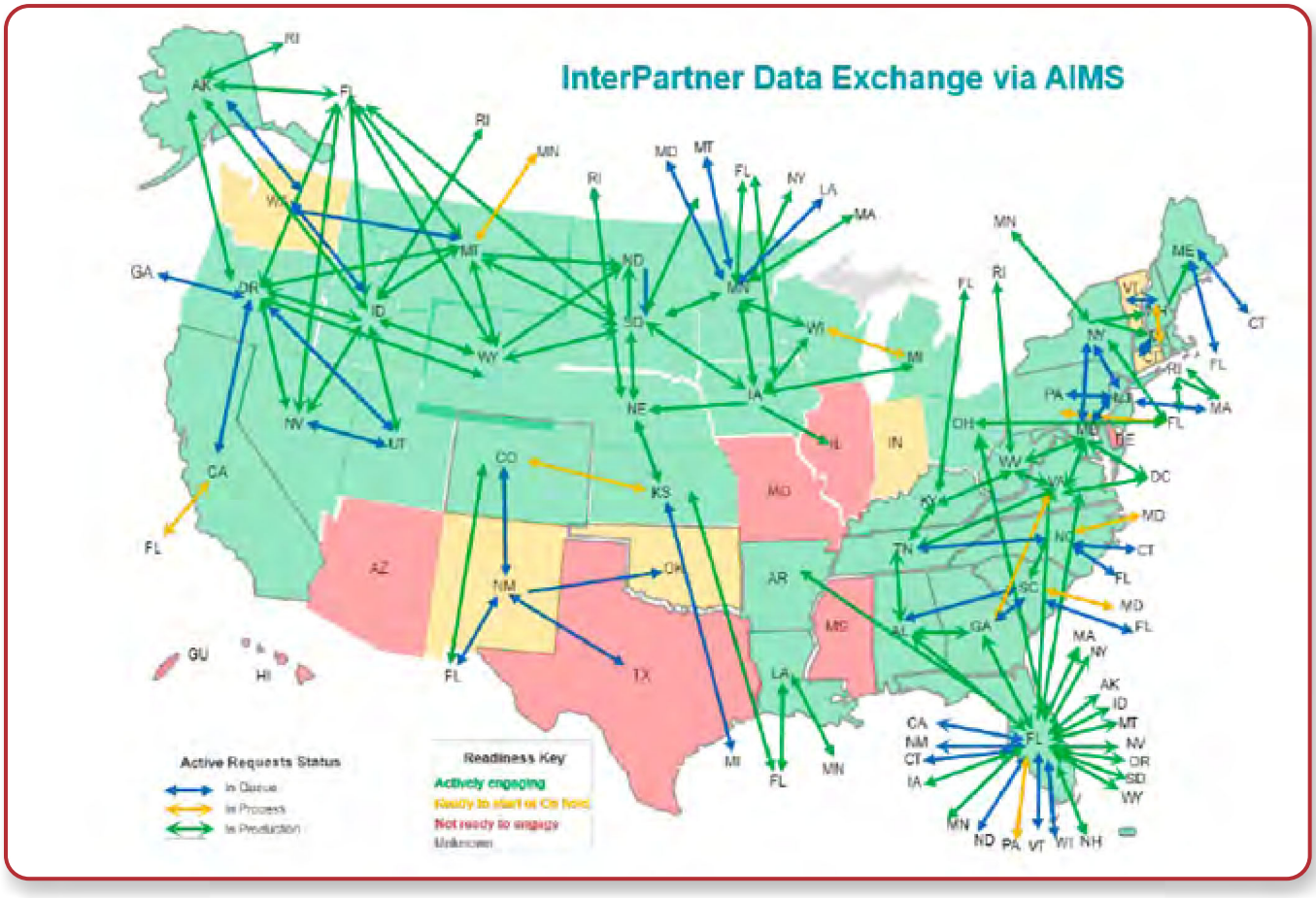
The AIMS Platform has developed and delivered new capabilities in response to the evolving needs of its stakeholders, building on a core set of hosting, transport and data enhancement functionality, complemented by services provided to data exchange partners by APHL.

Data Transport

The AIMS Platform facilitates reliable, scalable and secure data transmission across all data exchange partners by leveraging messaging standards including HL7, Fast Healthcare Interoperability Resources (FHIR), Clinical Document Architecture (CDA), Continuity of Care Document (CCD), ASC X12 and Integrating the Health Enterprise (IHE).

The Platform's InterPartner Data Exchange (**Figure 1**), which debuted in 2019, offered an easy-to-implement messaging structure that greatly facilitated data exchange between providers, public health, HIEs and non-traditional senders. It relies on a strict file naming convention that allows intelligent routing of data without inspecting the content. Once onboarded, partners were able to establish interconnections with minimal support from the AIMS Platform and add mutually agreed use cases as needed. Adoption of InterPartner led to an explosion of data exchanges between states.

Figure 1. InterPartner Data Exchange



Data Enhancement

The AIMS Platform offers centralized data translation and transformation services to parse and map data to public health vocabulary and terminology standards including LOINC, SNOMED, and ICD-10, and to translate messages from one type to another. AIMS Platform validation services ensure messages and payloads conform with message structure and content standards. The AIMS Validator offers a self-service online validation tool for data senders to test message structure and content for various protocols including eCR and ELR prior to sending messages through the platform. The AIMS Platform also performs inline validation of messages transported through the Platform to ensure recipients receive quality payloads.

Data Hosting

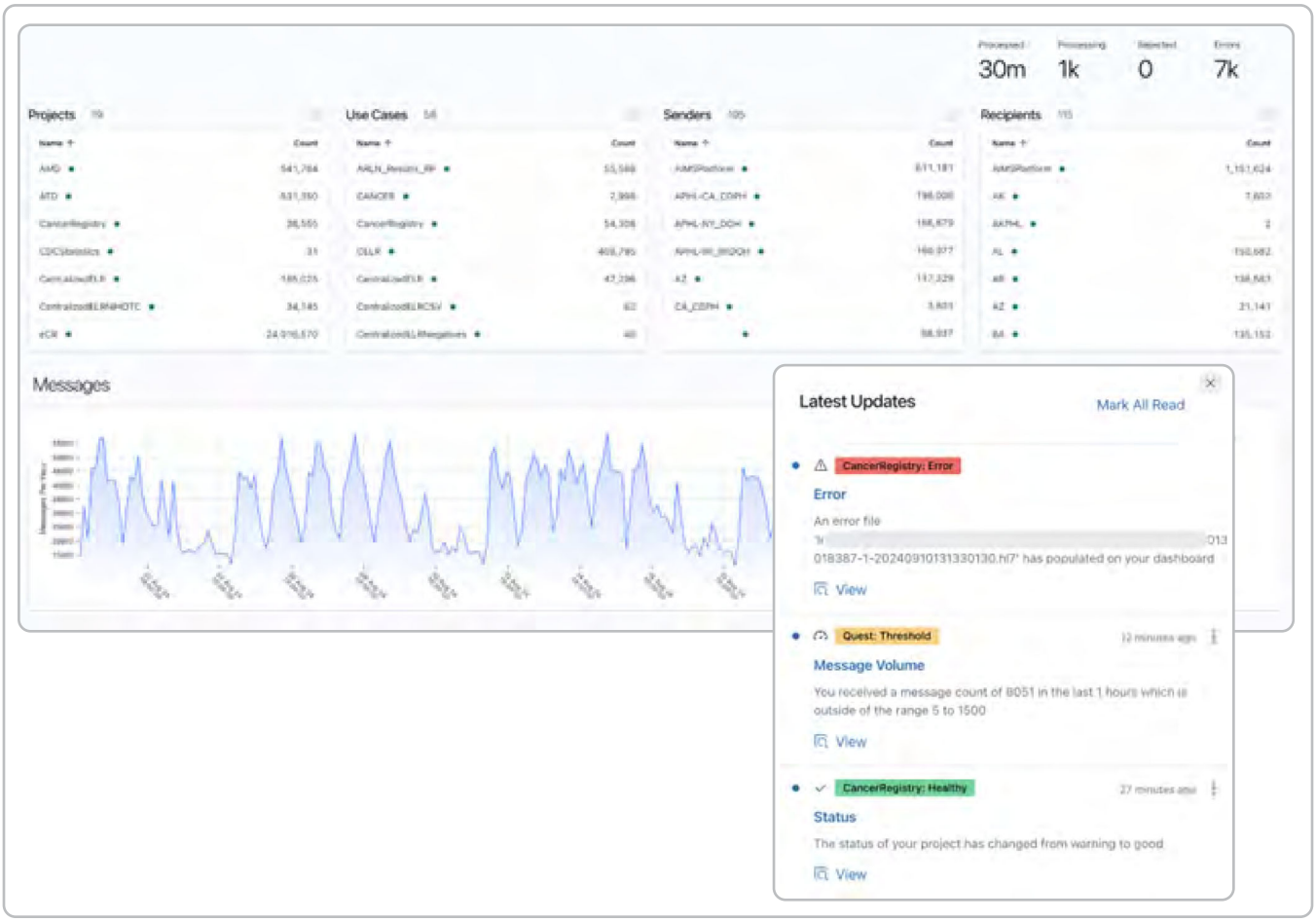
The AIMS Platform also offers secure and compliant hosting and storage services for third party applications. This includes on-demand data storage services to archive, backup, download, and export data and secure development, test and production environments that are accessible via virtual desktops. Leveraging cloud services, the Platform offers high performance computing (HPC) to enable users to quickly and easily build and manage HPC clusters for data-intensive use cases. All data on the Platform are managed using best practices for hosting, accessing and destroying data in compliance with health data regulations.

Data Exchange Partner Services

Data exchange partners can monitor message flow in real-time via the interactive AIMS Platform Dashboard (**Figure 2**), which offers views of message volumes that are searchable and filterable by project type, date/time, sender and recipient across all environments (test, staging, and production). Users can program notifications (via email, SMS, and web) based on specific events, triggers or message volume thresholds, and can also receive alerts when the status of specific interfaces or message routes become slowed or "unhealthy." This observability enables users to monitor trends in their data flows and detect and troubleshoot issues as they arise.

In addition, APHL offers technical assistance from subject matter experts in validation, vocabulary and data services to all data exchange partners to ensure the breadth of AIMS services are accessible to all senders and recipients.

Figure 2. AIMS Platform Dashboard Examples



AIMS's Architectural Evolution

Initially developed in 2008 to eliminate paper-based reporting for influenza, the AIMS Platform has scaled up dramatically, now securely transmitting on average 23 million messages per month. The Platform's scalability is central to its critical utility in public health data exchange in the US, enabling it to rapidly adjust to meet the needs of emerging threats. This section describes the architectural evolution of the AIMS Platform over the past 15 years to illustrate the core features, advantages and disadvantages of various designs, as well as demonstrate how a robust, multifaceted platform can be iteratively built over time, starting with a single use case.

Prior to the creation of the AIMS Platform, all public health laboratory data exchange between jurisdictions and federal agencies happened through point-to-point connections (**Figure 3**) via the Public Health Information Networking Messaging System (PHINMS). This point-to-point connection model required establishing and maintaining a server and client for each node to connect, which involved considerable and ongoing maintenance to ensure current configurations with valid certificates for each server and client. Each node that was exchanging data had to maintain hundreds of connections, requiring a substantial investment in ongoing day-to-day maintenance. A message intended for multiple recipients had to be sent to each recipient individually, so that a message going to five recipients was sent five separate times.

This model required significant investment from all parties involved in trading data, and the quality and reliability of data exchange was highly variable across data exchange partners due to its reliance on resources and maintenance at the level of each node. Because in this model the number of connections to maintain across all nodes (n) is $n^2+ 1$, each additional trading partner considerably increases the complexity of the system, making the point-to-point model difficult and resource-intensive to scale.

The complexity and cost involved in the point-to-point data exchange model necessitated a transition to a clearinghouse model, wherein all nodes connect to a central partner to exchange data with the other parties. Recognizing its trusted position with public health laboratories, CDC engaged APHL to develop two hub-and-spoke platforms (**Figure 4**) to facilitate centralized data exchange, using influenza data as the first use case. These platforms (which later were unified to become the AIMS Platform) replaced paper-based reporting of influenza results from public health laboratories to CDC and reduced the number of point-to-point connections CDC had to maintain from 50 to two. By implementing a standard HL7 message for influenza surveillance, this initiative improved data quality and CDC's ability to process results. The hub-and-spoke model substantially reduces the number of connections that need to be maintained across the system, driving down the n^2+1 connections of the point-to-point model to n+1, and thus drastically increasing the system's efficiency.

Figure 3. Point-to-point model

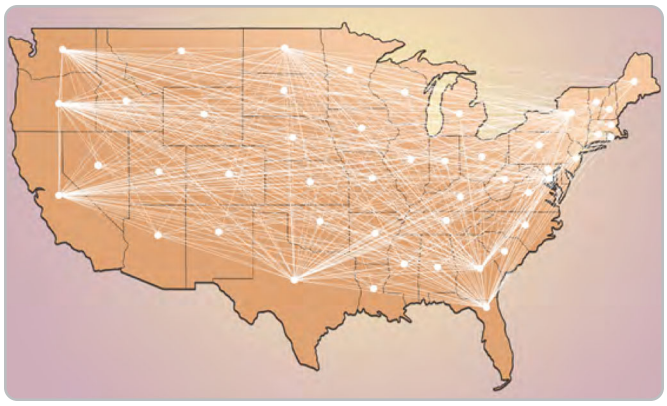
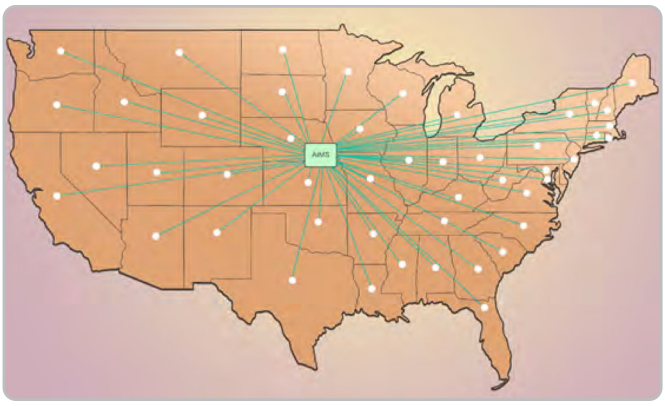


Figure 4. Hub-and-spoke model, with AIMS as central hub



The AIMS Platform was initially implemented as a route-not-read (RnR) hub, with end-to-end encryption so a message can be read only by the intended recipient. Because the intermediary is unable to read any messages sent through an RnR hub, this model is particularly useful in "no trust" environments. In this design for the AIMS Platform, "end points" (i.e., data exchange partners) need only to maintain a client—rather than both a client and a server—to send data to the hub. A trading partner sends a message to the hub and receives a message in return to acknowledge receipt. Included in the receipt message are any pending messages sent to the trading partner via the hub. Data exchange partners can connect to AIMS via their preferred protocol (PHINMS, SFTP, AWS S3), allowing them to exchange data with partners using different protocols than their own.

By initiating data exchange through an RnR model, the AIMS Platform had a lower threshold of trust to establish with each trading partner, which sped up onboarding of data senders. Over time, as the Platform onboarded more data exchange partners and provided consistent service and hands-on technical assistance, data exchange partners' trust increased, and the AIMS Platform was able to offer centralized data services that require the ability to read message content. The AIMS Platform began to offer data services such as translation, transformation and validation, reducing the need for each trading partner to invest in costly integration and transformation personnel and increasing the value of participation. In this centralized services model, data exchange partners encrypt their messages to the Platform, which then decrypts, performs the service function, re-encrypts, and sends the message to the intended recipient. When the intermediary is enabled to decrypt and read message content, it can send a single message to multiple recipients, rather than requiring a sender to send the same message multiple times for multiple end parties. This reduces burden on data senders and increases the overall efficiency of the system.

With the advent of eCR, the AIMS Platform added a publish-and-subscribe data exchange pattern. In this model, data senders "publish: messages to the Platform and recipients create "subscription" rules for the types of messages that they want to receive. Publish-and-subscribe gives control to data recipients to determine the types of data they want to receive, rather than simply receiving anything sent to them.

Today, the AIMS Platform uses all of the above data exchange patterns (route-not-read, centralized service hub-and-spoke, and publish-and-subscribe) across its various use cases on the platform. Each model has advantages and rationale for being employed; readers considering building nationwide or region-wide data exchange systems will need to consider levels of trust in the centralized party, resourcing, capabilities of data exchange partners and anticipated message volumes when selecting their data exchange model(s). See "Table 1. Data Exchange Models" (page 13) for a comparison of the advantages and disadvantages of each model.

For all data exchange patterns, it is critical to design a system's architecture for scalability. The AIMS Platform's migration to the cloud enabled event-based message processing that leverages queuing to avoid system failures when volumes increase rapidly. The AIMS Platform's event-based design leverages object stores and queuing to achieve its stability. A message object sent to the Platform is stored in an object store and triggers an "event," which is then routed to the intended recipient, who then retrieves a copy of the object from the store. This drastically reduces the size of messages processed through the system—each event is a few KB, compared with an entire message file, which can be quite large (many gigabytes, in the case of eCR). With an event-based design, a message that is intended for multiple recipients will be sent once to an object store, from where each recipient will retrieve it, based on the events each receives.

The AIMS Platform's event-based processing leverages queues to ensure messages continue to be processed in the midst of sudden surges in volumes so that data is never lost. It leverages parallel processing within single channels in its interface engine as well as across multiple channels, scaling horizontally as volumes increase. It can also scale horizontally by working across multiple interface engine instances—see Appendix 1, "Parallel Message Processing on the AIMS Platform" (page 43), for further illustration.

Table 1. Data Exchange Models

Model	Connection Architecture	Characteristics
Point-to-point *ex. PHINMS*	• Requires establishment of a server (to receive) and a client (to send) at each node • Each new node added to the network requires new two-way connections at all nodes (n2+1)	• Reliability of data exchange highly variable depending on level of resources and maintenance at each node • Laborious and costly to maintain • Resource-intensive to scale • Certificate management becomes extremely burdensome as the number of data exchange partners increases • Payload size is frequently an issue • Any message that needs to be sent to multiple recipients must be sent multiple times
Hub-and-spoke also known as clearinghouse model *ex. AIMS influenza surveillance*	• All nodes connect to other nodes via a central partner • Only one connection required to add a new node (n+1)	• Vastly easier to maintain and scale than point-to-point
Route-not-read (RnR) variant of hub-and-spoke	• Messages are encrypted end-to-end, i.E., Content is not readable at the central hub	• Useful in no-trust environments or where content is highly sensitive • Could be a good starting point for a new exchange platform
Centralized services variant of hub-and-spoke	• Messages are encrypted to the central partner which decrypts, performs services, re-encrypts, and delivers message to recipients	• A trusted central partner can transform, translate, validate, and route messages to multiple recipients as desired by sender • Allows powerful efficiencies for interoperability and message distribution
Publish-and-subscribe variant of hub-and-spoke *ex. eCR*	• Senders publish messages to central hub, recipients subscribe to messages they want (and are entitled) to receive	• Allows recipients to manage what data they ingest • Allows powerful efficiencies for interoperability and message distribution

The AIMS Platform was originally hosted on the servers of two jurisdictional health departments, and was later consolidated onto a third-party data center already secured for federal use. The reliance on physical provisioning of hardware eventually became a hindrance, however, when presented with a second use case that required considerably more scale and security. In order to facilitate ELR for Quest Diagnostics, one of the largest commercial laboratories in the US, the AIMS Platform needed to process messages containing protected health information at high volumes, necessitating scalability of infrastructure that was compliant with federal standards for health information privacy and security. Cloud hosting offered considerable efficiencies to achieve this, both in cost and operations. The AIMS Platform first built a proof of concept using Amazon Web Services (AWS) cloud services to test and measure efficacy, and by 2014 the Platform was running fully in the cloud.

In addition to cloud-based secure and redundant hosting, the AIMS Platform leverages specialized cloud services and tools for transport, data storage and extraction, analytics and visualization, enabling robust functionality without requiring installation and management of other software.

See "Implementing a Centralized Data Exchange System" (page 29) for a detailed discussion of infrastructure.

Platform Expansion Through Third-party Applications

Cloud services also enable the security, compliance and system redundancy necessary to host and support third-party applications on the AIMS Platform, and in 2017 it began hosting the State and Territorial Exchange of Vital Events (STEVE) software, facilitating nationwide exchange of birth and death record data. With AWS multi-availability zones and managed services, the AIMS Platform could offer a high level of reliability and system availability necessary to run critical data pipelines. Today, in addition to STEVE, the Platform hosts many third-party applications that are essential to the US health infrastructure, including:

- Reportable Conditions Knowledge Management System (RCKMS)
- Immunization Gateway
- Global Action in Healthcare Network—Healthcare-Associated Infections (GAIN-HAI) module
- Data for Action on Antimicrobial Resistance Threats
- Lab Web Portal
- NIST validators

Essential Use Cases on AIMS

Primed and Ready: The Case of COVID-19

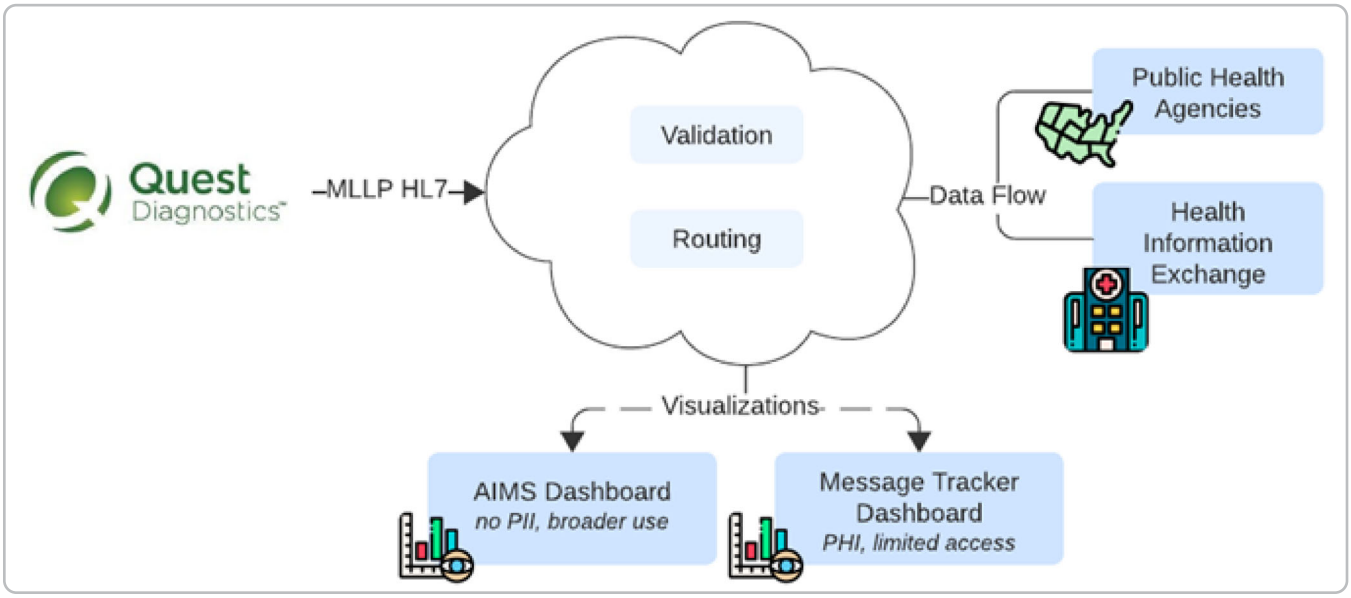
The architectural and hosting decisions traced above put the AIMS Platform in a position to be rapidly scaled. While the COVID-19 outbreak in 2020 strained every nerve of the public health response, the national data messaging platform connecting public health entities was already solidly established and able to deal with the sudden vast increase in the quantity of data that needed to be processed and transmitted. The route initially created in 2008 to transmit influenza data to the CDC was modified to accept new data and within one month laboratories and agencies in every state were reporting COVID-19 data to the CDC and the White House. This contributed substantially to understanding incidence of the disease and the status of response efforts, testing coverage, and lab supply chain issues.

The essential use cases developed on the AIMS Platform played key roles in the COVID-19 response and are very much in use today. They are described individually in this section.

Electronic Laboratory Reporting (ELR)

The AIMS Platform facilitates the transmission of digital laboratory reports from laboratories to state and local public health departments, healthcare systems, and CDC. As a notable example, APHL partnered with Quest Diagnostics, one of the largest commercial laboratories in the US, to facilitate the secure routing of ELR from all Quest Diagnostics facilities to public health agencies across the United States. The AIMS Platform enabled Quest to send all reportable results to the appropriate jurisdiction's public health agency via a single connection, rather than Quest needing to create and maintain connections with each jurisdiction individually. The Platform also offers dashboard, validation, and tracking services to Quest to enable monitoring of message flow and investigation into issues in message content that may arise (**Figure 5**).

Figure 5. Quest Diagnostics ELR



Electronic Case Reporting (eCR)

Public health agencies depend on timely case reporting regarding specific conditions in order to take appropriate action to protect public health. Most jurisdictions require healthcare providers to report on conditions of interest, which may vary considerably from agency to agency.

The automated generation and transmission of initial and follow up case reports from electronic health records to public health agencies for review and action reduces the burden on healthcare providers to accomplish their legal reporting requirements and improves the timeliness and completeness of case reports at the local, state and national levels. This efficiency in data collection greatly enhances the ability of public health agencies to manage case investigations and respond to emerging outbreaks (**Figure 6**).

At the time of the COVID-19 outbreak, a pilot version of eCR already operating on the AIMS Platform was quickly moved into production, making this functionality available across the country. Today, over 7,800 hospitals and healthcare providers send electronic case reports via the AIMS Platform (**Figure 7**).

Figure 6. eCR Workflow

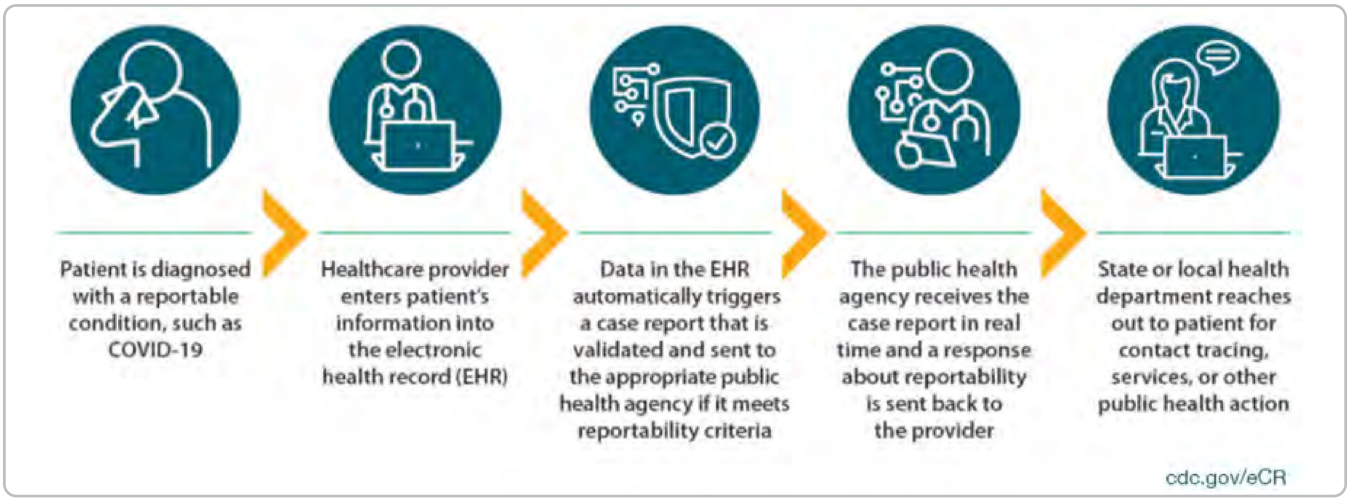
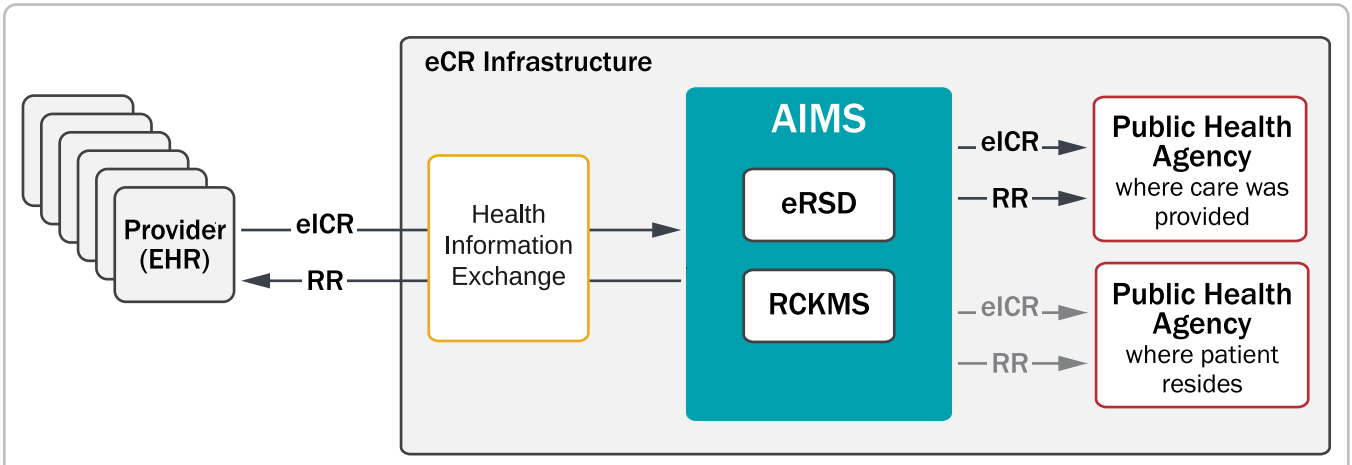


Figure 7. eCR Architecture (simplified)



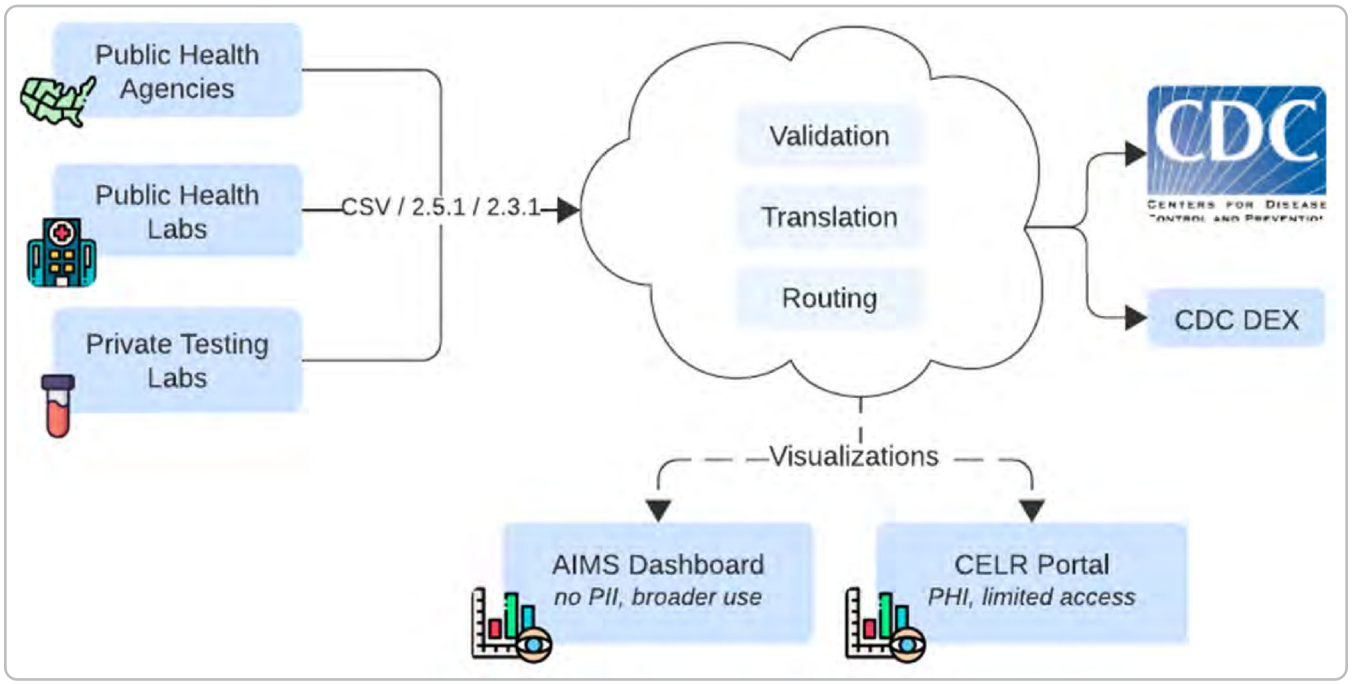
eICR – Electronic Initial Case Report
RR – Reportability Response
eRSD – Electronic Reporting and Surveillance Distribution System
RCKMS – Reportable Condition Knowledge Management System

COVID-19 Electronic Laboratory Reporting (CELR) Data Lake

The public health response to COVID-19 depended on comprehensive laboratory testing data. These data contributed to understanding COVID-19's impact and testing coverage and to the identification of supply chain issues for reagents and other materials.

CDC was authorized to obtain COVID-19 laboratory data from all public health laboratories, public health agencies, and some private laboratories to understand the number of tests ordered and performed, the results of each test, and the percent positive. The project allowed public health departments to receive technical assistance to send laboratory results in HL7 (2.5.1, 2.3.1, or 2.3.z) and/or CSV formats to the AIMS Platform for access by CDC staff via the CELR Portal (**Figure 8**).

Figure 8. CELR Data Lake



Electronic Test Orders and Results (ETOR)

The AIMS Platform facilitates the exchange of electronic test orders and results between healthcare organizations and public health laboratories. Using universal newborn screening tests (federally-mandated universal screening for congenital, genetic, and metabolic conditions not otherwise detectable at birth) as its first use case, DETOR, the electronic test order and result solution on the AIMS platform, facilitates exchange of test orders and results between healthcare organizations and public health laboratories, as well as mandatory reporting to health authorities. ETOR resolves coding and terminology challenges across many disparate systems to enable standards-based data exchange across all parties (**Figures 9** and **10**).

Antimicrobial Resistance Laboratory Network (AR Lab Network)

The AIMS Platform enhances the capacity of public health laboratories in the US for rapid identification of emerging and novel antimicrobial resistance threats. Public health laboratories in the AR Lab Network use the AIMS Platform's Intelligent Message Routing to send antimicrobial resistance results to the CDC for analysis (**Figure 11**).

Figure 9. APHL ETOR Program, Phase 1

Figure 10. APHL ETOR Program, Phase 2

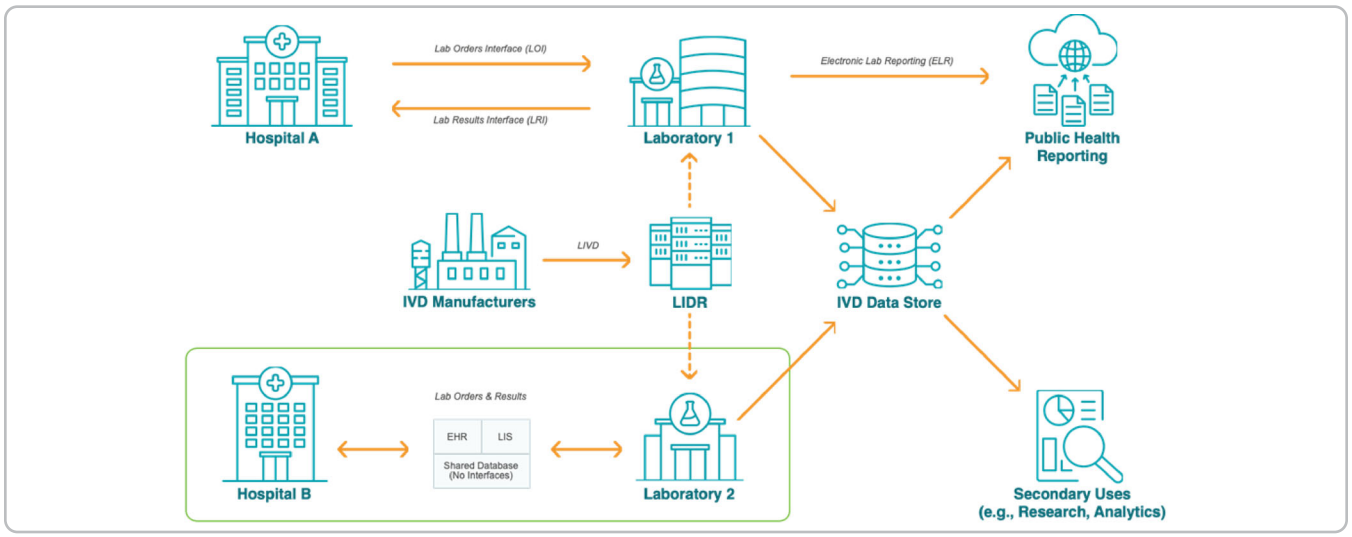
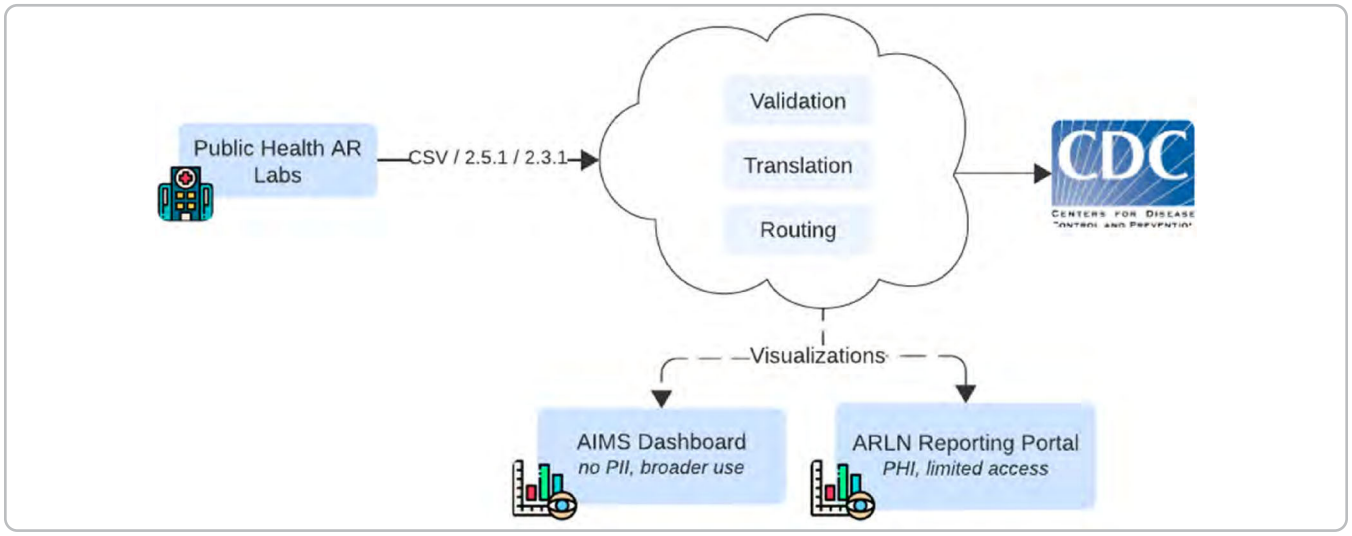


Figure 11. AR Lab Network



The AIMS Model

After discussing the core functionality, benefit, use cases and evolution of the AIMS Platform in the "The AIMS Platform Story" (page 7), this section details the foundational technical components that deliver the Platform's functions. The following subsections represent how the AIMS Platform is aligned to industry best practices for large scale health IT systems, divided into three foundational areas—technology, data governance, and stakeholders and operations (**Figure 12**). For any centralized data exchange system to be successful, it is paramount to ensure these areas are aligned and focused on the needs of stakeholders and their operations.

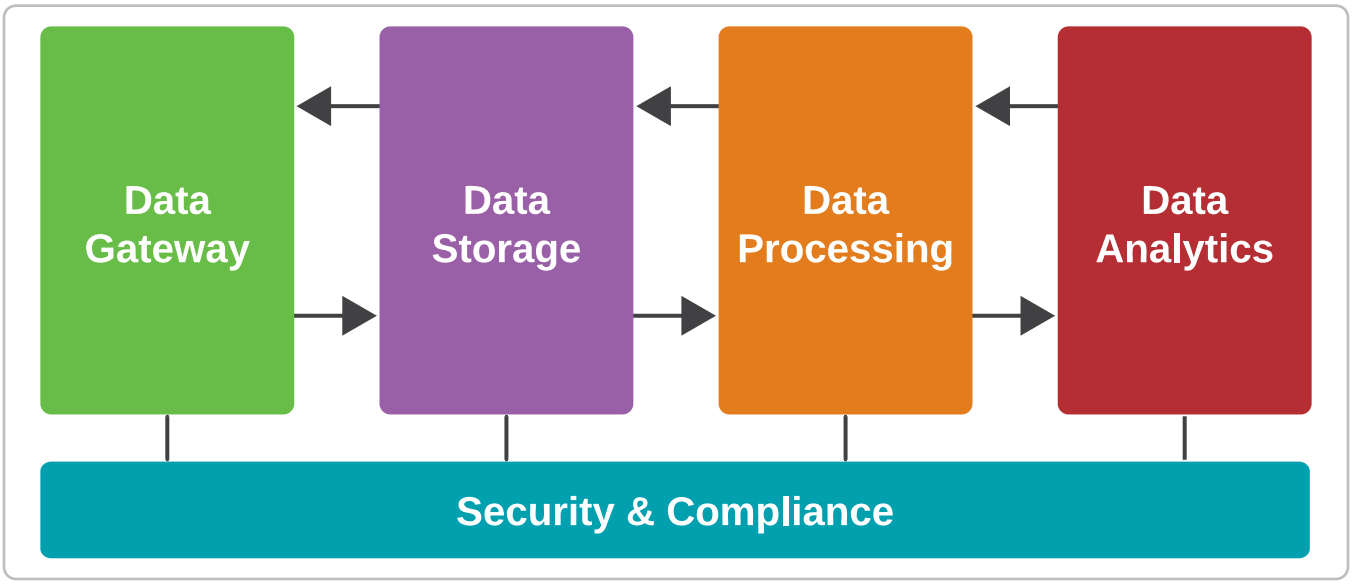
This section describes each of the foundational areas, with particular focus on the technology and its functional requirements. The last section, "Implementing a Centralized Data Exchange System" (page 29), offers specific guidance on how to implement the technology.

Figure 12. Foundational Areas of Health IT Systems

Technology

The technology used for the AIMS Platform is a sophisticated cloud-based infrastructure designed to facilitate the seamless exchange of public health data across a diverse network of healthcare and laboratory systems. Leveraging cutting-edge technology and industry-standard data exchange protocols, the AIMS Platform ensures robust, secure, scalable, and efficient real-time communication and interoperability among public health entities. This section condenses 15 years of history for a complex health IT platform into core "must have" components with a focus on the functional capabilities within the technology. This "model" is intentionally high-level and designed to frame the AIMS Platform into consumable chunks such that it can be successfully replicated. The following high-level conceptual architecture covers the technical components required to deliver the core functionality of the AIMS Platform (**Figure 13**).

Figure 13. High-level Technical Model of the AIMS Platform



Security and Compliance

Security and compliance are essential for protecting sensitive health data and ensuring it is handled properly. Security measures, like encryption and access controls, make sure that only the right people can see the data, keeping it safe from unauthorized access. Compliance means following specific rules and regulations, like HIPAA in the US, to ensure the platform meets legal requirements for data privacy. Together, security and compliance help prevent data breaches, protect individual privacy, and build trust between public health organizations by ensuring that the data is managed safely and responsibly. The AIMS Platform meets these functional requirements through the following:

- **Encryption:** The AIMS Platform implements robust encryption mechanisms for data both at rest and in transit. This includes the use of transport layer security (TLS) for secure communication channels and advanced encryption standard (AES) for encrypting stored data.
- **Identity and Access Management (IAM):** The platform leverages IAM services (e.g., AWS IAM) to control access to resources based on roles and policies. Role-based access control (RBAC) ensures that only authorized users can access sensitive data or perform specific actions within the platform.
- **Compliance Frameworks:** The AIMS Platform is designed to comply with various regulatory frameworks, including NIST* and HIPAA in the US. This involves implementing security controls, conducting regular audits, and maintaining detailed logs and audit trails to demonstrate compliance with legal and regulatory requirements.
- **User Authentication and Federation:** The AIMS Platform supports single sign-on (SSO) and federated authentication, allowing users from different organizations to securely access the platform using their existing credentials. This is often managed through integration with identity providers (e.g., Active Directory, SAML-based SSO).

Infrastructure

Infrastructure provides the backbone for the platform, ensuring reliable data flow and system stability. It includes servers, networks and cloud systems that scale to meet demands and recover quickly from issues. By integrating development operations (DevOps) practices, software code deployment and infrastructure management are automated, allowing for efficient updates and responsive system performance. The AIMS Platform meets these functional requirements through the following:

Cloud Hosted

- **Platform as a Service (PaaS):** The AIMS Platform operates on a cloud-based platform as a service (PaaS) model, leveraging cloud providers like Amazon Web Services (AWS) for its underlying infrastructure. This model abstracts much of the underlying hardware management, allowing the platform to focus on application-level concerns such as data processing, integration, and security.
- **Elastic Scaling:** The use of cloud infrastructure enables elastic scaling, allowing the platform to automatically adjust compute and storage resources based on demand. This is particularly important for handling large, variable volumes of public health data, such as during disease outbreaks or other public health emergencies.
- **High Availability and Redundancy:** The platform is designed for high availability with failover mechanisms and redundancy across multiple geographic regions. This ensures that even if one region experiences an outage, the platform can continue to operate from another, minimizing downtime.

* NIST Privacy Framework and Cybersecurity Framework

Hybrid Cloud and On-premises Integration

- **Legacy System Integration:** On-premises data gateways and middleware solutions bridge legacy systems with the AIMS Platform cloud-based services, enabling secure and efficient data synchronization and management.
- **Data Residency and Compliance:** A hybrid cloud approach stores sensitive data on-premises to comply with local regulations while leveraging cloud scalability for less sensitive operations.
- **Enhanced Security and Performance:** Critical components can operate on-premises for low-latency or enhanced security needs, integrated seamlessly into the cloud infrastructure for broader capabilities.
- **Cost Management:** On-premises processing for data-intensive tasks reduces data transfer costs and can help substitute high compute demands before moving data to the cloud for further analysis.
- **Container Deployment:** Utilizing container technologies such as Docker and orchestrators like Kubernetes allows applications to be packaged and deployed consistently across both cloud and on-premises environments. This approach supports seamless scalability, improves deployment speeds and maintains operational consistency across the hybrid infrastructure, enhancing the Platform's ability to manage workloads flexibly and securely regardless of their deployment location.

Monitoring and Observability

- **Centralized Logging:** The platform employs centralized logging systems (e.g., ELK Stack—Elasticsearch, Logstash, Kibana) to aggregate logs from various services and components. This enables real-time monitoring and facilitates troubleshooting by providing a comprehensive view of system activity and data flows.
- **Monitoring and Alerts:** AIMS uses monitoring tools (e.g., Prometheus, Grafana, AWS CloudWatch) to track the health and performance of its infrastructure and services. Alerts are configured to notify administrators of any anomalies, such as service disruptions or performance degradation, allowing for quick response and resolution.
- **Health Checks and SLA Monitoring:** Automated health checks are performed on critical services to ensure they are functioning correctly. Service level agreements (SLAs) are monitored to ensure the platform meets its uptime and performance commitments.

Automation and DevOps

- **Infrastructure as Code (IaC):** The AIMS Platform employs infrastructure as code (IaC) tools (e.g., Terraform, AWS CloudFormation) to automate the deployment and management of its infrastructure. This ensures consistency, repeatability and easier scaling of the Platform's resources.
- **CI/CD Pipelines:** The Platform uses continuous integration/continuous deployment (CI/CD) pipelines to automate the testing, integration and deployment of new code. This reduces the time to deploy updates and ensures that changes are thoroughly tested before going live.
- **Containerization and Orchestration:** The AIMS Platform utilizes containerization (e.g., Docker) and orchestration (e.g., Kubernetes) to manage its services in a microservices architecture. This approach allows for modular, scalable and resilient service deployment.

Data Gateway

The data gateways component acts as a central access point ensuring that only authorized data enters and exits the platform, applying encryption and access controls to maintain security. The data gateway handles the secure connection of data exchange partners and allows their messages to flow through the system—secure pipes and moving data. Each data exchange partner has an address and front door in the platform to manage the flow of messages through their connected pipe. By managing these connections efficiently and centrally, they support reliable and secure data exchanges across various systems. The AIMS Platform meets these functional requirements through the following:

- **Partner Onboarding:** The platform includes tools and processes for onboarding new partners, ensuring that their systems can securely connect to the AIMS Platform and comply with required data standards. This may involve configuring APIs, setting up secure communication channels, and validating data formats.
- **Enterprise Service Bus (ESB):** The AIMS platform utilizes an enterprise service bus (ESB) architecture, which acts as a central hub for integrating various public health systems and applications. The ESB facilitates the routing, transformation and orchestration of messages between systems, ensuring that data is delivered in the correct format and to the appropriate destinations.
- **HL7, FHIR and Other Standards:** The platform supports multiple healthcare data exchange standards, including HL7 and FHIR. This enables seamless communication between different healthcare systems, laboratories and public health agencies.
- **API Gateway:** The AIMS Platform includes an API gateway that manages the exposure of RESTful and SOAP-based APIs, providing secure access to the platform's services. The API gateway handles API traffic, implements rate limiting, and provides authentication and authorization mechanisms to ensure secure and controlled access to data and services.

Data Storage

The purpose of the data storage component is to securely hold, manage and backup data for future use, ensuring it is readily accessible when needed. Raw message data is temporarily stored to ensure reliable transfer and once fully processed, it is purged from the system. Metadata related to message processing is stored for longer durations to provide lifecycle tracking of the data exchange to support auditability and compliance. Redundancy and backup measures are in place to protect data during transit and prevent loss in case of system failures. The AIMS Platform meets these functional requirements through the following:

- **Database Systems:** The AIMS Platform utilizes a combination of relational (e.g., PostgreSQL, MySQL) and NoSQL databases (e.g., MongoDB, DynamoDB) to store structured and unstructured data. The choice of database depends on the nature of the data, with relational databases typically used for transactional data and NoSQL databases for large-scale, flexible data storage.
- **Data Lakes and Warehousing:** The platform may integrate with data lakes (e.g., AWS S3) or data warehouses (e.g., Amazon Redshift) for long-term storage and analytics. This allows for the aggregation and analysis of large volumes of public health data, supporting research and decision making.
- **Disaster Recovery and Business Continuity:** The AIMS Platform includes disaster recovery and business continuity planning, with redundant systems and regular backups to ensure that data is not lost in the event of an outage or disaster. This ensures that public health functions can continue uninterrupted, even in adverse situations.

Data Processing

The purpose of the data processing component is to clean, organize and transform raw data into a standardized and usable format. This process ensures that the data is accurate, consistent and free from errors, making it reliable for routing to other data exchange partners. This includes everything from, normalization, enrichment, transformation, intelligent routing and more. Data processing is the critical step around data cleanup and translation once a message is received by the Platform so that it can be effectively routed and consumed by other data exchange partners. The AIMS Platform meets these functional requirements through the following:

- **Message Queues and Event-driven Architecture:** The Platform employs message queues (e.g., AWS SQS or Apache Kafka) to decouple systems and enable asynchronous communication. This event-driven architecture allows for real-time processing of data as it flows through the platform, ensuring timely delivery and processing of public health information.
- **Interface Engines and Transformation Tools:** The AIMS Platform utilizes interface engines like Mirth to manage data transformations and integrations, providing robust, configurable pipelines that handle complex data flows and ensure data integrity. These tools are complemented by data transformation services within ESB or by external processors, facilitating conversions between various data formats (e.g., HL7 v2 to FHIR), crucial for system interoperability.
- **Orchestration:** The platform uses orchestration engines (e.g., Apache Camel or AWS Step Functions) to manage complex workflows involving multiple systems and data sources. Orchestration ensures that data is processed in the correct order and that all necessary steps are completed before it reaches its destination.

Data Analytics

The data analytics component is responsible for examining and interpreting processed data to uncover patterns, trends, and insights that help inform decisions. It includes statistical tools, algorithms and sometimes machine learning to analyze data and generate reports, insights or predictions into data exchange activities and system performance. These tools allow users to monitor message traffic, identify trends and track system health in real-time. Visual dashboards and custom reports offer a clear view of key metrics, such as message volume, processing times, and error rates, enabling better decision making and proactive issue resolution. The AIMS Platform meets these functional requirements through the following:

- **Real-time Analytics and Message Tracking Visualization:** The AIMS Platform supports real-time data analytics through the use of streaming data platforms (e.g., Apache Kafka Streams, AWS Kinesis, OpenSearch, Elasticsearch). This allows public health officials to gain immediate insights from incoming data, enabling timely decision making.
- **Reporting Tools:** The Platform integrates with reporting tools (e.g., Tableau, Power BI) that allow users to create and view customized reports. These tools are connected to the data sources within the AIMS Platform, providing up-to-date information and visualizations for analysis.

Data Governance

Data governance is the second of the three foundational areas. The purpose of data governance is to build trust around the quality of data. Data governance plays a crucial role in the AIMS Platform, ensuring that the data exchanged and managed within the platform are accurate, secure and compliant with regulatory requirements. Best practices related to data governance as a methodology have been codified by DAMA International in the Data Management Body of Knowledge (DMBOK2).* The alignment of these governance practices with the DMBOK framework enhances the platform's capability to manage data as a strategic asset. Data governance impacts the AIMS Platform in the following areas:

Data Quality Management

- **Accuracy and Consistency:** Data governance frameworks establish standards for data quality, ensuring that the data exchanged through AIMS are accurate, consistent and reliable. This is critical for public health decision making, where data errors can lead to incorrect conclusions or delayed responses.
- **Data Validation:** Governance policies ensure that data are validated before being exchanged, reducing the likelihood of errors and ensuring that the information is usable by all stakeholders.

Data Security and Privacy

- **Access Control:** Data governance policies define who can access specific data within the AIMS Platform. This includes role-based access controls that ensure only authorized personnel can view or manipulate sensitive data.
- **Compliance with Regulations:** The AIMS Platform must comply with various regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the US. Data governance ensures that all data handling processes meet these regulatory requirements, protecting patient privacy and sensitive information.
- **Encryption and Security Protocols:** Governance frameworks mandate the use of encryption and other security protocols to protect data during transmission and storage, ensuring that data remain secure from unauthorized access.

Data Stewardship

- **Ownership and Accountability:** Data governance assigns clear ownership of data, ensuring that there are designated stewards responsible for managing, maintaining, and ensuring the quality of data within the AIMS Platform. This helps in maintaining accountability and ensuring that data is handled properly.
- **Automated Lineage Tracking:** Tools and technologies that automatically capture and document data lineage. This reduces the manual effort required from data stewards and ensures that lineage information is consistently captured.
- **Lifecycle Management:** Governance policies define how data are managed throughout the lifecycle—from creation and transmission to archiving and deletion. This ensures that data are available when needed and disposed of properly when no longer required.

* DAMA International, The DAMA Guide to the Data Management Body of Knowledge (DMBOK2), 2nd ed., Technics Publications, 2017, Chapter 3: Data Governance, pp. 45-66.

Interoperability and Standardization

- **Standards Compliance:** Data governance ensures that the data exchanged through AIMS adheres to industry standards (e.g., HL7). This standardization is crucial for interoperability, allowing different public health systems to communicate effectively.
- **Metadata Management:** Governance frameworks also manage metadata, providing context and meaning to the data. This helps in ensuring that data are interpreted correctly across different systems and stakeholders.

Risk Management and Incident Response

- **Risk Assessment:** Data governance includes ongoing risk assessments to identify potential vulnerabilities in data handling processes. This proactive approach helps in mitigating risks before they become significant issues.
- **Incident Response:** Governance policies define how data breaches or other incidents should be handled. This includes clear protocols for responding to security incidents, notifying affected parties and restoring data integrity.

Compliance and Auditing

- **Audit Trails:** Data governance mandates the creation of audit trails to track who accessed or modified data and when. This is crucial for accountability and for identifying the source of any issues that arise.
- **Regular Audits:** Regular audits are conducted to ensure that data governance policies are being followed and that the AIMS platform remains compliant with all applicable laws and regulations.

Ethical Use of Data

- **Transparency:** Data governance promotes transparency in how data are used, ensuring that all stakeholders understand the purpose and scope of data collection and usage.
- **Ethical Considerations:** Governance frameworks ensure that data are used ethically, protecting individuals' rights and ensuring that data are not misused.

Data governance within the AIMS Platform is closely aligned with the DMBOK framework, supporting the integrity, security and usability of public health data. This alignment ensures that data management practices in the AIMS Platform are robust, consistent and compliant with industry best practices, thereby enabling effective public health interventions and policy making.

Interested Groups and Operations

Interested groups and operations comprise the third and final foundational area. Once built, a centralized data exchange platform requires considerable ongoing support to onboard users to the system, ensure the system's proper functioning, and provide support to the system's interested groups including end users, regulators, vendors and data exchange partners. The following subsections first outline a service delivery framework used by the AIMS Platform to ensure high-quality operations followed by the core services provided.

Service Delivery Management Framework

Service delivery management (SDM) is the process of overseeing and coordinating the delivery of services to customers. The goal of SDM is to consistently meet customer expectations in terms of timeline, cost, quality and performance.

For the AIMS Platform, this means the service that data exchange partners rely on for sending and receiving data is delivered reliably. If there are disruptions in data flow, the service is restored within an acceptable amount of time, the root cause is analyzed, documentation for the issue is published to interested parties, and mitigating preventative actions are planned and executed.

Service delivery via information technology infrastructure library (ITIL) and IT service management (ITSM) is a structured approach to managing IT services and ensuring they meet the needs of the business. It involves a set of practices for designing, delivering, managing, and improving IT services within an organization, aimed at aligning IT services with business goals.

Key Components of ITIL/ITSM Service Delivery

- **Service Strategy:** Defines the business needs for IT services and how IT will contribute to business objectives. It involves service portfolio management, demand management and financial management.
- **Service Design:** Focuses on designing new IT services or modifying existing ones to meet business requirements. It covers aspects like capacity management, availability management, service level management and IT security.
- **Service Transition:** Ensures that new or changed services are smoothly transitioned into the live environment without disruption. It includes change management, release and deployment management, and service asset and configuration management.
- **Service Operation:** Concerned with the day-to-day delivery of IT services, ensuring they meet agreed-upon service levels. This includes incident management, problem management, event management, request fulfillment and access management.
- **Continual Service Improvement (CSI):** Focuses on improving the effectiveness and efficiency of IT services based on feedback, service metrics and performance data. CSI looks for opportunities to enhance service quality over time.

Key Processes in ITIL/ITSM Service Delivery

- **Incident Management:** Deals with managing and resolving incidents (e.g., unplanned interruptions or reductions in quality of service) to restore normal service as quickly as possible.
- **Problem Management:** Involves identifying the root causes of incidents and preventing their recurrence.
- **Change Management:** Governs the process of implementing changes to IT services while minimizing disruption.
- **Service Level Management (SLM):** Ensures that all services are delivered within agreed-upon performance standards (e.g., service level agreements, or SLAs).

- **Configuration Management:** Keeps track of the configurations of all IT assets and ensures accurate information is available to support other ITSM processes.
- **Capacity Management:** Ensures IT infrastructure is capable of meeting current and future business demands.
- **Availability Management:** Ensures that IT services are available at the times they are needed according to SLAs.
- **Financial Management for IT Services:** Involves budgeting, accounting, and charging for IT services.
- **IT Service Continuity Management:** Focuses on managing risks to ensure that IT services can be resumed after a disaster or major incident.

Benefits of ITIL/ITSM Service Delivery

- **Alignment with Business Needs:** Ensures that IT services directly support business goals.
- **Increased Efficiency:** Streamlined processes reduce duplication of effort and increase productivity.
- **Improved Service Quality:** Through structured processes, IT services meet agreed-upon standards of quality.
- **Better Risk Management:** Helps anticipate potential disruptions and minimize their impact.
- **Cost Control:** Provides a clearer understanding of IT service costs and helps manage budgets more effectively.
- **Enhanced Customer Satisfaction:** Consistent service delivery leads to higher levels of satisfaction from users and customers.

Core Services

Platform Maintenance and Operations

The AIMS Platform has an interdisciplinary team of engineers, technicians, project managers and analysts supporting its ongoing operations. Maintenance and operation of the AIMS Platform includes operating system (OS) patching, vulnerability management, configuration and ongoing tuning of alerts and alarms, infrastructure operations, certificate monitoring, production operations process automation, identity and access management, infrastructure disaster recovery activities, Mirth integration engine interface monitoring, configuration security, database security, creation and maintenance of lifecycle policies, intrusion detection management, application bug resolution and change management.

The team responsible for maintaining the ongoing operations of the exchange platform should define and document which individuals/teams are responsible for identifying which events/issues, which are responsible for resolving them, which are accountable to their resolution, who needs to be consulted on the manner of their identification and resolution, and who should be informed. These assignments can be displayed in a RACI matrix.

End User Support

The AIMS Platform's Production Support team provides assistance to its data exchange partners to bolster their ongoing use of the Platform. The Production Support team monitors a ticketing system to respond to customer requests and communicates planned maintenance downtime windows, service outages, and other important information via email. A ticketing system to standardize, manage and track the processing and resolution of support requests is an essential tool for the ongoing support of a robust data exchange system, as is a tool to automate mass communications to end users.

AIMS Platform users first accessed the ticketing system via a portal or via email submission. Email submission is a nice, lightweight option for less complex service desks. As the complexity of the AIMS Platform grew, however, there became a need to direct users primarily to the service desk portal to standardize inputs (**Figure 14**).

Portal submission allows the AIMS Platform team to require a baseline set of information for the submission, and also allows users to classify their submission as either an incident (e.g., downtime, issues requiring a fix) or a request (e.g., inquiry/question, less urgent service desired) (**Figure 15**). These categorizations enable triage to different teams based on urgency and subject matter expertise required and result in more efficient resolution.

Figure 14. AIMS Platform Service Desk Menu

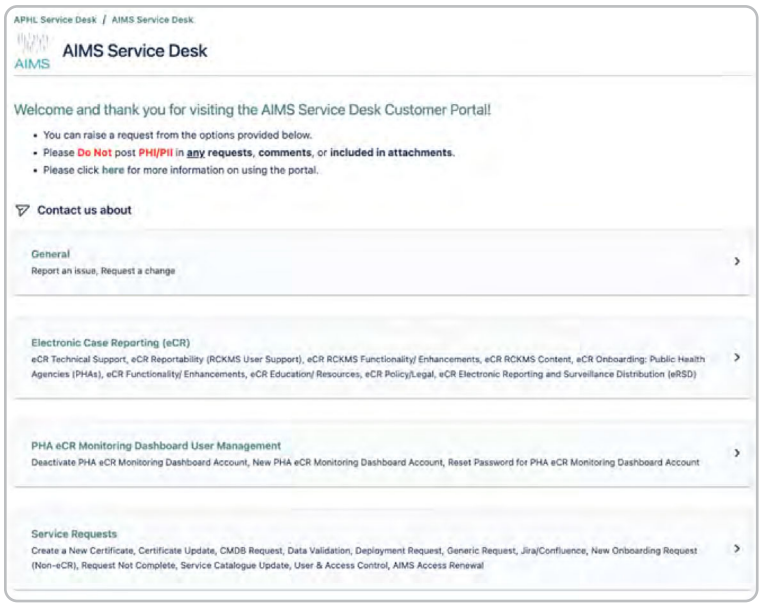
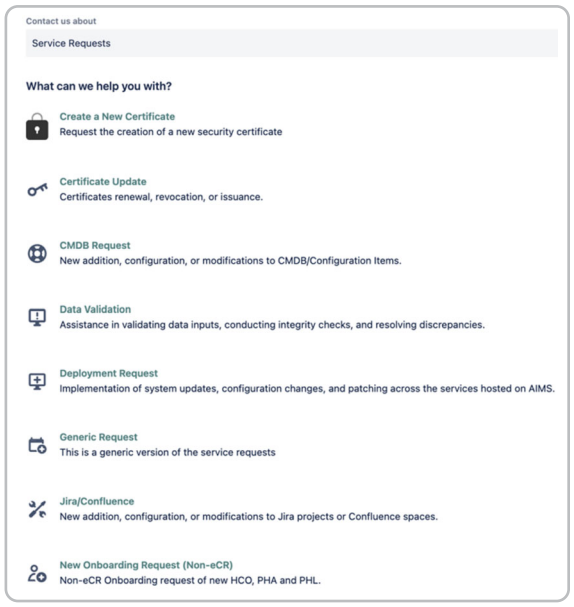


Figure 15. AIMS Service Requests menu



Technical Assistance

APHL collaborates with CDC to offer technical assistance to public health agencies and laboratories to connect laboratory information management systems (LIMS) and data management systems to the AIMS Platform. The Technical Assistance program provides temporary support from technical subject matter experts to public health entities for whom permanently placing staff is not financially feasible or sustainable. Technical assistance areas include:

- **Data Extraction and Migration:** assistance to extract data from local LIMS and ensure it is in a compatible format for migration to the data exchange platform. Technical Assistance subject matter experts offer specialized tools and techniques to facilitate seamless data migration with minimal disruption to ongoing operations.
- **Data Standardization and Integration:** assistance to ensure that data from various local LIMS are standardized and integrated into the data exchange platform, maintaining data integrity and consistency. Technical Assistance subject matter experts advise on data mapping and transformation processes to align different data structures and formats.
- **Customized Solutions and Support:** assistance to offer tailored solutions to address specific challenges faced by local LIMS in different regions. Technical Assistance subject matter experts provide hands-on support and troubleshooting during the data extraction and migration process.
- **Training and Capacity Building:** assistance to train local staff on new processes and tools for data extraction and management in the new environment. Technical Assistance subject matter experts support data trading partners to build local capacity to manage and maintain the new system effectively and sustainably.
- **Ensuring Data Security and Privacy:** assistance to implement robust data security measures during the extraction and migration process to protect sensitive public health information, ensuring compliance with relevant regulations and standards throughout the data handling process.

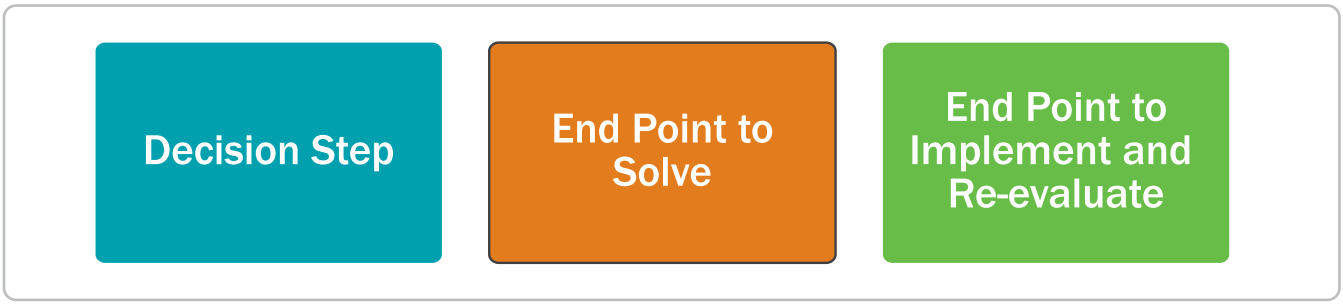
Implementing a Centralized Data Exchange System

The goal of this section is to offer the reader design considerations and best practices as they plan centralized public health data exchanges to meet their own needs. Organization of this section mirrors that of "The AIMS Model" (page 19), providing insights for each foundational area—technology, data governance, and interested groups and operations—and detailed decision trees to guide the reader's decision making processes. Specific attention is given to the technology area to assist the reader in navigating the needs of their stakeholders and standing up their own public health data exchange.

Technology

This section covers each of the components introduced in the high-level technical model of the AIMS Platform including decision trees plus considerations and guidance. Decision trees are used to document thought processes the reader can walk through to help determine the non-functional needs around implementation of each of the necessary technology components in your country. Non-functional requirements focus on enumerating how this work should be done so that the system operates in a way that allows it to be broadly usable and sustainable over time. The color-coded legend in **Figure 16** is used by each of the decision trees.

Figure 16. Decision Tree Legend

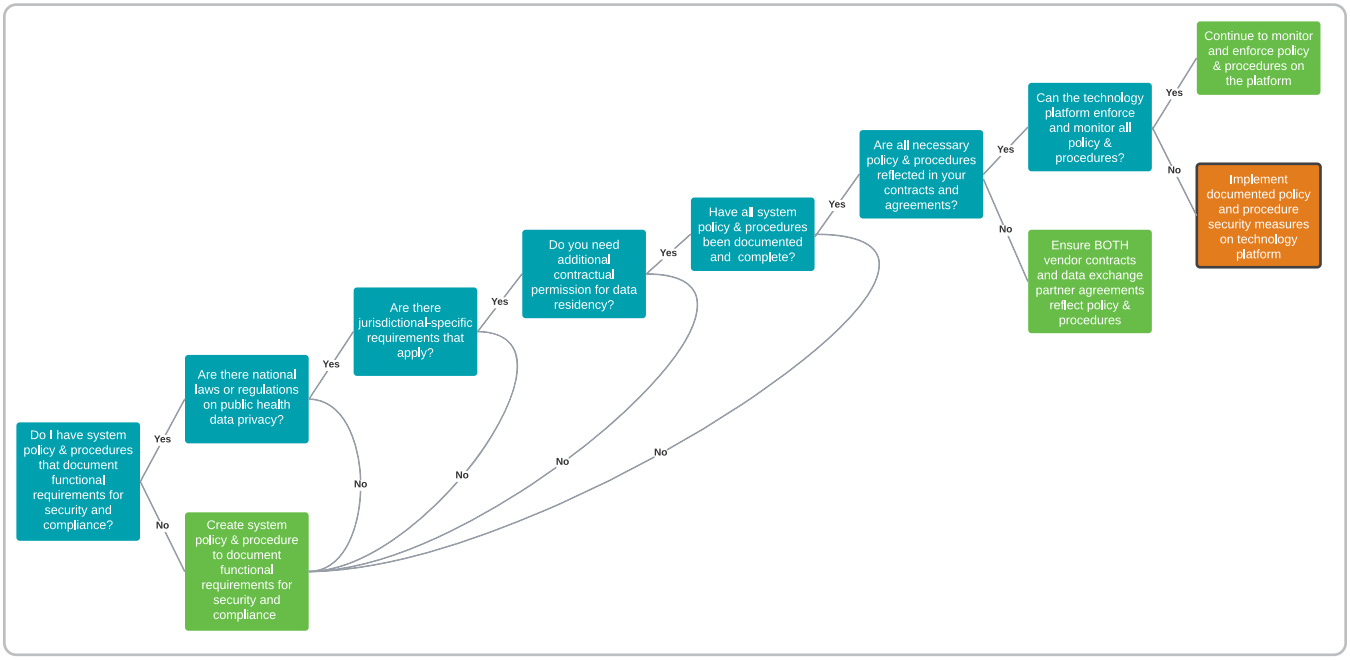


Security and Compliance

Security and compliance, as discussed in "The AIMS Model" (page 19), are essential for protecting sensitive health data and ensuring it is handled properly. Security measures, like encryption, make sure that only the right people can see the data, keeping it safe from unauthorized access. Compliance means following specific rules and regulations, like HIPAA, to make sure the platform meets legal requirements for data privacy (**Figure 17**).

Decision Tree

Figure 17. Decision Tree: Implementing Security and Compliance Controls for the Public Health Data Exchange



Considerations

The following list provides key considerations that will be useful in helping the reader align security and compliance needs with the non-functional needs of their public health data exchange. Compliance with data residency and localization laws is crucial:

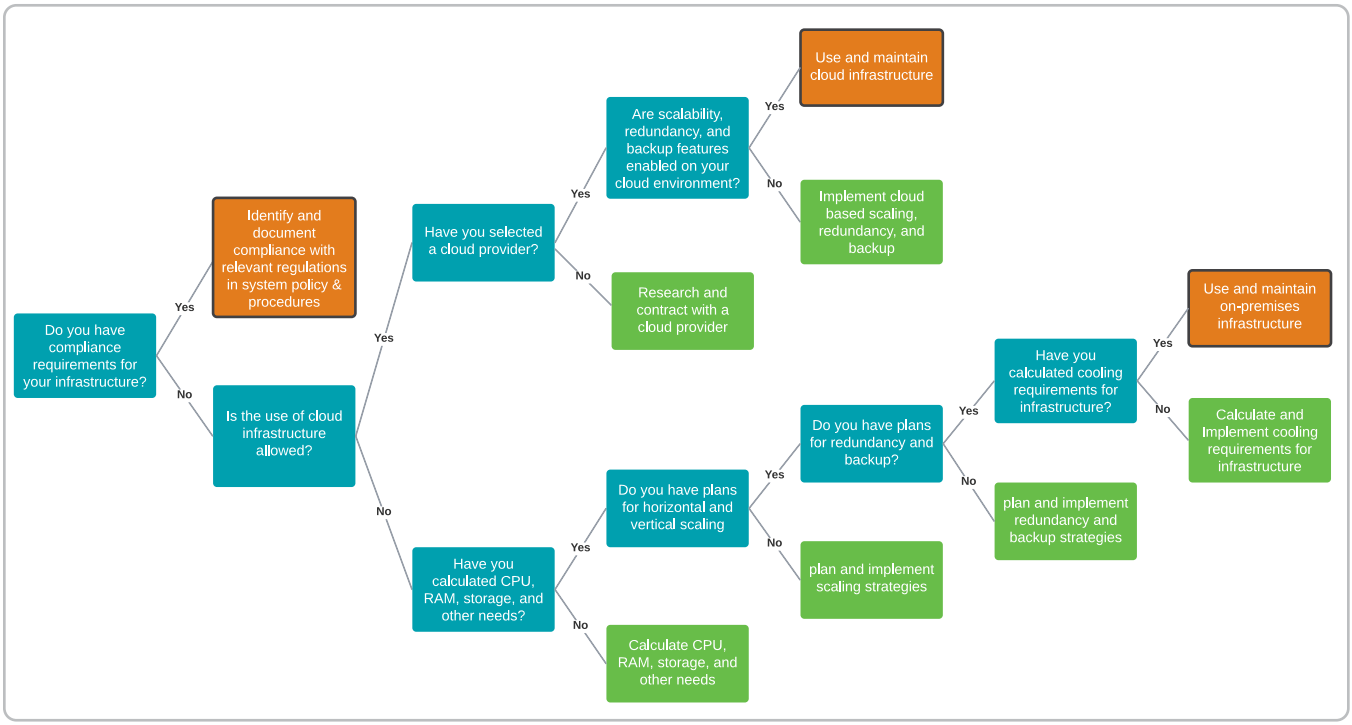
- **Jurisdiction-specific Regulations:** Understanding and adhering to laws that require data to be stored within specific geographic boundaries.
- **Privacy Requirements:** Ensuring that cloud providers comply with regulations such as GDPR, HIPAA, and other data protection laws.
- **Contractual Agreements:** Including data residency requirements in contracts with cloud providers to ensure compliance.

Infrastructure

As covered in the "The AIMS Model" (page 19), infrastructure is the backbone of the public health data exchange, providing reliable data flow and system stability. It includes the use of servers, networks and cloud systems that scale to meet demands and recover quickly from issues. It must integrate DevOps practices where software code deployment and infrastructure management are automated, allowing for efficient updates and responsive system performance (**Figure 18**).

Decision Tree

Figure 18. **Decision Tree: Implementing Infrastructure for the Public Health Data Exchange**



Considerations

Organizations need to consider several factors when deciding between on-premises, private cloud, and cloud solutions.

- **Compliance:** Ensuring compliance with relevant regulations and standards for data privacy and security.
- **Security:** Assessing the security measures offered by commercial cloud, private cloud, or on-premises environments.
- **Cost:** Comparing the total cost of ownership, including upfront investments, ongoing maintenance, and operational expenses.
- **Scalability:** Evaluating the ability to scale resources up or down based on demand. The ability to scale upward to respond to heavy traffic in emergent situations, and downward for economy once such situations have abated. In the cloud, this may be as easy as ticking a few boxes in the administration console. If not in the cloud, rely on architecture that enables containerization (e.g., Kubernetes).
- **Flexibility:** Considering the flexibility to access data remotely and support collaboration.

Guidance for On-premises/Private Cloud

Assuming an on-premises or private cloud implementation, the architecture will end up including a combination of virtual machines (VMs) (i.e., servers), along with a container strategy for the component capabilities. An implementation of VMs and containers can be considered using the following capacity calculations steps:

1. **Estimate Compute Requirements**
   - **CPU requirements:**
     - Containerization overhead: Estimate the CPU overhead for container orchestration (e.g., Kubernetes) and virtualization (e.g., KVM, VMware).
     - Workload requirements: Calculate the average and peak CPU usage for the applications you will be running. This includes the baseline CPU required per container, multiplied by the number of containers.
     - Buffer capacity: Add a buffer (typically 20-30%) to accommodate unexpected spikes in usage and for future scaling.
     - Total CPU cores: The total number of required CPU cores = (CPU required per container x Number of containers) + Overhead + Buffer.
   - **RAM requirements:**
     - Application memory needs: Determine the RAM requirements for each application/container based on historical data or benchmarking.
     - Overhead: Include memory overhead for the container orchestration platform and virtualization.
     - Buffer capacity: Similar to CPU, add a buffer for RAM as well.
     - Total RAM: The total RAM required = (RAM required per container x Number of containers) + Overhead + Buffer.

2. **Estimate Storage Requirements**
   - **Storage volume for data:**
     - Calculate the total amount of data the applications will generate, store, and process. Include structured (databases) and unstructured data.
     - Include storage needs for logs, backups and any data retention policies.
     - Consider data redundancy and replication factors (e.g., RAID configuration, replication factor in distributed storage).
   - **IOPS (Input/Output operations per second):**
     - Determine the IOPS requirements based on the read/write operations of your workloads.
     - Consider the speed requirements for the disks (e.g., SSD vs. HDD) and storage tiers.
   - **Storage type:**
     - Evaluate the need for block storage, file storage and object storage based on your application's requirements.
     - Open-source solutions like Ceph, GlusterFS or others might be appropriate for a private cloud.

3. **Network Requirements**
   - **Bandwidth:**
     - Estimate the network bandwidth required based on the data transfer rate between containers, virtual machines and external systems.
     - Include bandwidth needed for backup, replication and data migration activities.
   - **Latency:**
     - Calculate the acceptable latency for different types of traffic (e.g., internal cluster communication, user requests).
     - Consider the impact of network latency on application performance.
   - **Network topology:** Determine if you will use software-defined networking (SDN) or traditional network configurations. SDN may require additional overhead.

Guidance for On-premises/Private Cloud, cont.

4. **Virtualization and Containerization Overhead**
   - **Hypervisor overhead:** Calculate the resource overhead associated with the hypervisor.
   - **Container orchestration overhead:** Determine the additional resources required for the container orchestration system (e.g., control plane resources in Kubernetes).
   - **Performance degradation:** Estimate potential performance degradation when running containers within virtual machines compared to running on bare metal.

5. **Scalability and High Availability**
   - **Scaling:** Provide guidance on how to calculate resources for horizontal and vertical scaling based on expected growth.
   - **Redundancy:** Estimate additional resources required for high availability (e.g., active-passive setups, load balancing).

6. **Resource Utilization Efficiency**
   - **Resource allocation:** Recommend methods for optimizing resource allocation, such as overcommitment ratios for CPU and memory in virtual environments.
   - **Autoscaling:** Consider the resource requirements for implementing autoscaling based on load.

7. **Power and Cooling Requirements**

   If on-premises resources will be heavily utilized, you should calculate the power consumption and cooling requirements of the data center based on the estimated compute and storage requirements.

8. **Example Calculation Model**
   - **CPU and RAM:**
     - Assume each container needs 2 vCPUs and 4 GB of RAM.
     - The orchestration layer requires 10% of total compute and memory resources.
     - You plan to run 20 containers.
     - Buffer is set at 25%.
     - Calculation:
       - Total CPU cores = (2 vCPUs x 20 containers) x 1.10 (overhead) x 1.25 (buffer) = 55 vCPUs.
       - Total RAM = (4 GB RAM x 20 containers) x 1.10 (overhead) x 1.25 (buffer) = 110 GB RAM.
   - **Storage:**
     - Assume each container requires 50 GB of storage and you require a replication factor of three.
     - Log and backup storage add an additional 20% overhead.
     - Calculation:
       - Total Storage = (50 GB x 20 containers) x 3 (replication) x 1.20 (overhead) = 3,600 GB (3.6 TB).

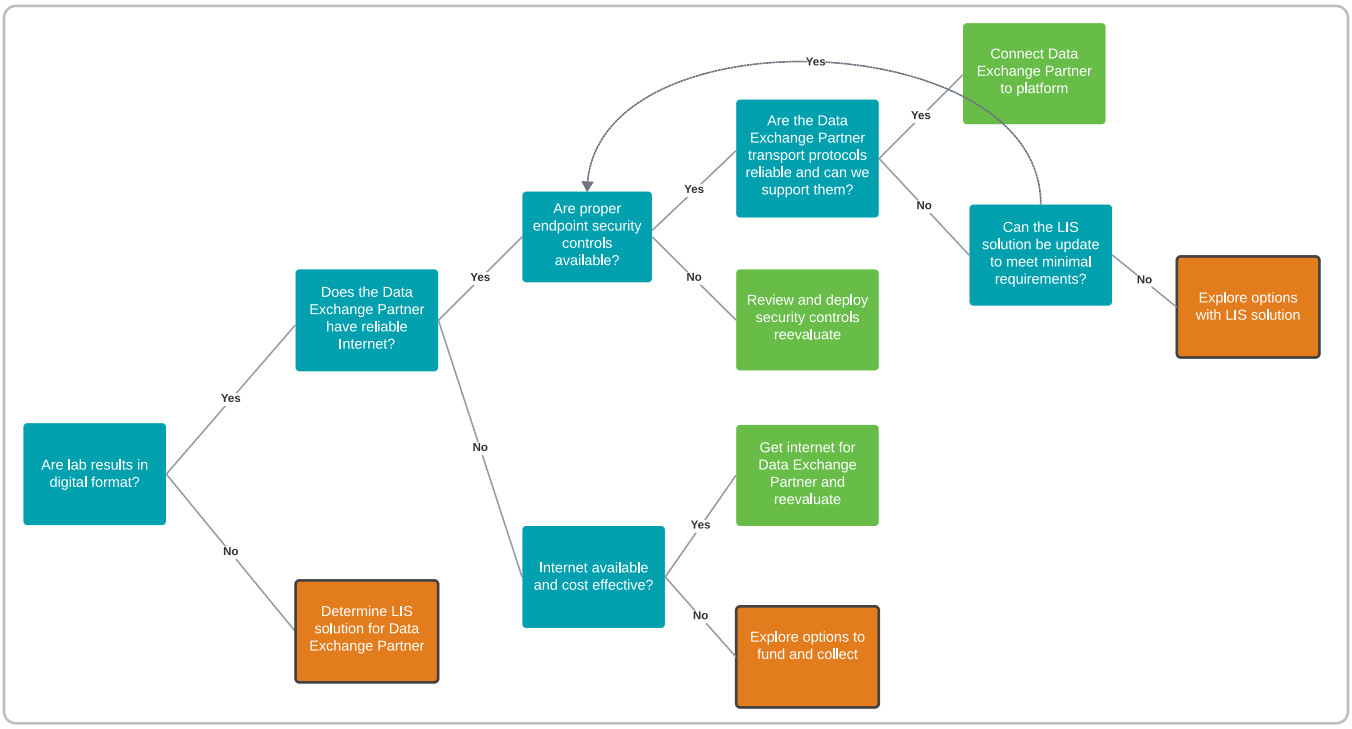
9. **Recommendations for Tools and Platforms**
   - **Container orchestration:** Kubernetes, OpenShift.
   - **Virtualization:** KVM, Xen, VMware (if not strictly open-source).
   - **Storage:** Ceph, GlusterFS for distributed storage.
   - **Monitoring and management:** Prometheus, Grafana for monitoring; Ansible, Terraform for infrastructure management.

Data Gateway

As covered in the "The AIMS Model" (page 19) the data gateway component acts as a central access point ensuring that only authorized data enters and exits the public health data exchange by applying encryption and access controls to maintain security. The data gateway handles the secure connection of data exchange partners and allows their messages to flow through the system (**Figure 19**).

Decision Tree

Figure 19. Decision Tree: Connecting a Data Exchange Partner to the Public Health Data Exchange



Considerations

The following list provides key considerations that will be useful in helping the reader align the functionality of the data gateway component with the non-functional needs of their public health data exchange:

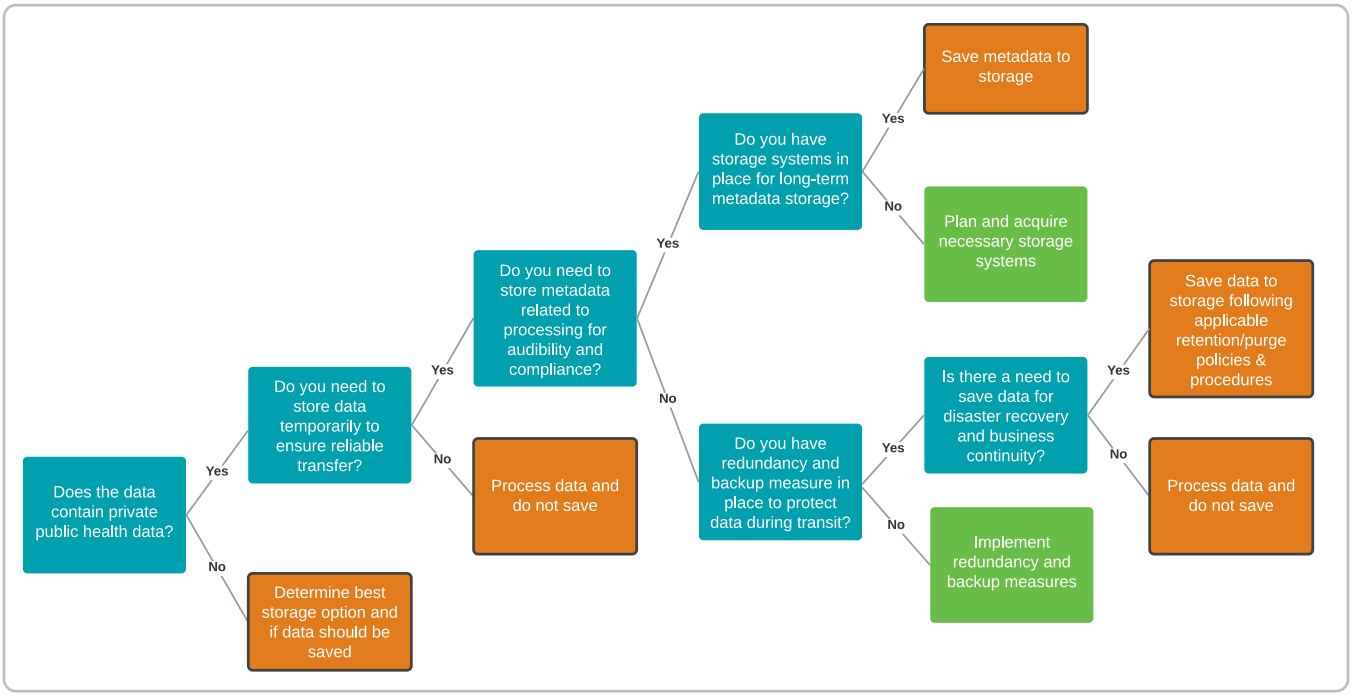
- The Mirth interface engine can be used whether you are building on-premises or in the cloud to facilitate performing the data manipulations that make interoperability possible. Mirth provides a base set of functionalities that can be used to build an interface between two systems with as few as ten lines of code, as compared with thousands of lines using Java or another standard programming language. If deploying in the cloud, compare Mirth with integration options that may be available directly from the cloud service provider.
- Data formats
  - **Consistency:** Standardized data formats ensure that data is consistently structured, allowing for seamless integration and exchange between different systems and organizations. Consistent data formats enable automated processing and analysis, reducing the risk of errors and improving efficiency.
  - **Interoperability:** With standardized formats, data from various sources can be easily combined and compared. This is essential for public health agencies to aggregate data from multiple jurisdictions, enabling comprehensive surveillance and analysis of public health trends.
  - **Scalability:** Standardized formats facilitate the scaling of public health information systems. As more states and organizations adopt common formats, it becomes easier to expand the network and incorporate additional data sources without significant reconfiguration.

Data Storage

The data storage component, as outlined in the "The AIMS Model" (page 19), must securely hold, manage and backup data for future use, ensuring it is readily accessible when needed. Raw message data is temporarily stored to ensure reliable transfer and processing. Additionally, metadata related to message processing is stored for longer durations to provide lifecycle tracking of the data exchange to support auditability and compliance. Finally, redundancy and backup measures must be in place to protect data during transit and prevent loss in case of system failures.

Decision Tree

Figure 20. **Decision Tree: Storing and managing data in the public health data exchange**



Considerations

The following list provides key considerations that will be useful in helping the reader align the functionality of the data storage component with the non-functional needs of their public health data exchange:

- **Data lifecycle management:**
  - Implement policies for data retention, archiving and deletion to manage storage effectively and comply with legal requirements.
  - Use automated tools to enforce data lifecycle policies and optimize storage usage.
- **Performance:**
  - Choose storage solutions that offer the required IOPS (Input/Output Operations Per Second) for your applications, balancing cost and performance.
  - Consider tiered storage options (e.g., SSDs for high-performance needs, HDDs for archival) to optimize both cost and speed.
- **Scalability:**
  - Utilize scalable cloud storage solutions (e.g., Amazon S3) that allow you to expand storage capacity without significant upfront investment.
  - For on-premises storage, plan for future expansion with modular storage units or clustered storage systems.

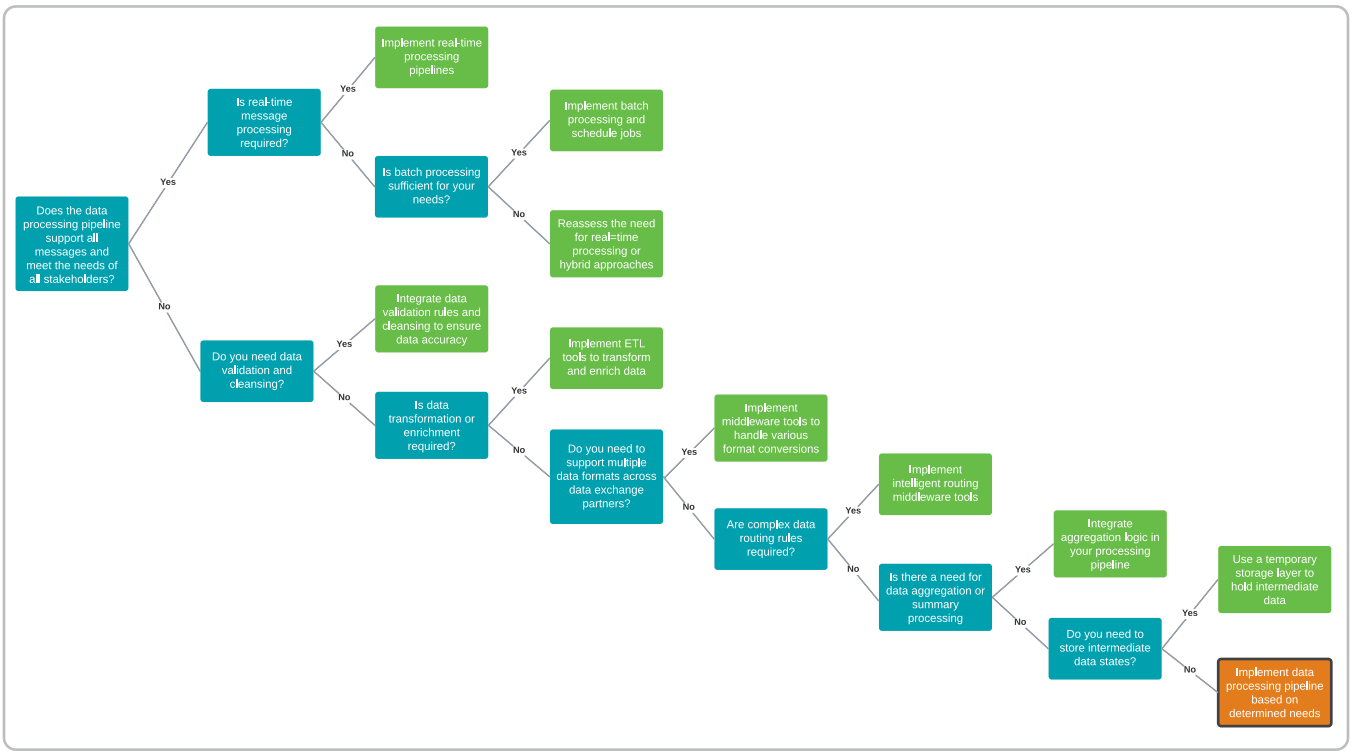
- **Data Security:**
  - Encrypt all stored data using industry-standard encryption algorithms to protect sensitive information.
  - Implement access controls to ensure that only authorized personnel can access specific data sets.
- **Redundancy and Backup:**
  - Design redundant storage systems with automatic failover to ensure data availability in case of hardware failures.
  - Regularly back up critical data and store backups in geographically diverse locations to prevent data loss.

Data Processing

As outlined in the "The AIMS Model" (page 19), the data processing component has the responsibility to clean, organize, and transform raw data into a standardized and usable format. This includes everything from normalization, enrichment, transformation, to intelligent routing and more. This component ensures that the data is accurate, consistent and free from errors, making it reliable for routing and consumption by other data exchange partners on the platform.

Decision Tree

Figure 21. Decision Tree: Facilitating sharing of messages between data exchange partners across potential different systems and environments



Considerations

The following list provides key considerations that will be useful in helping the reader align the functionality of the data processing component with the non-functional needs of their public health data exchange:

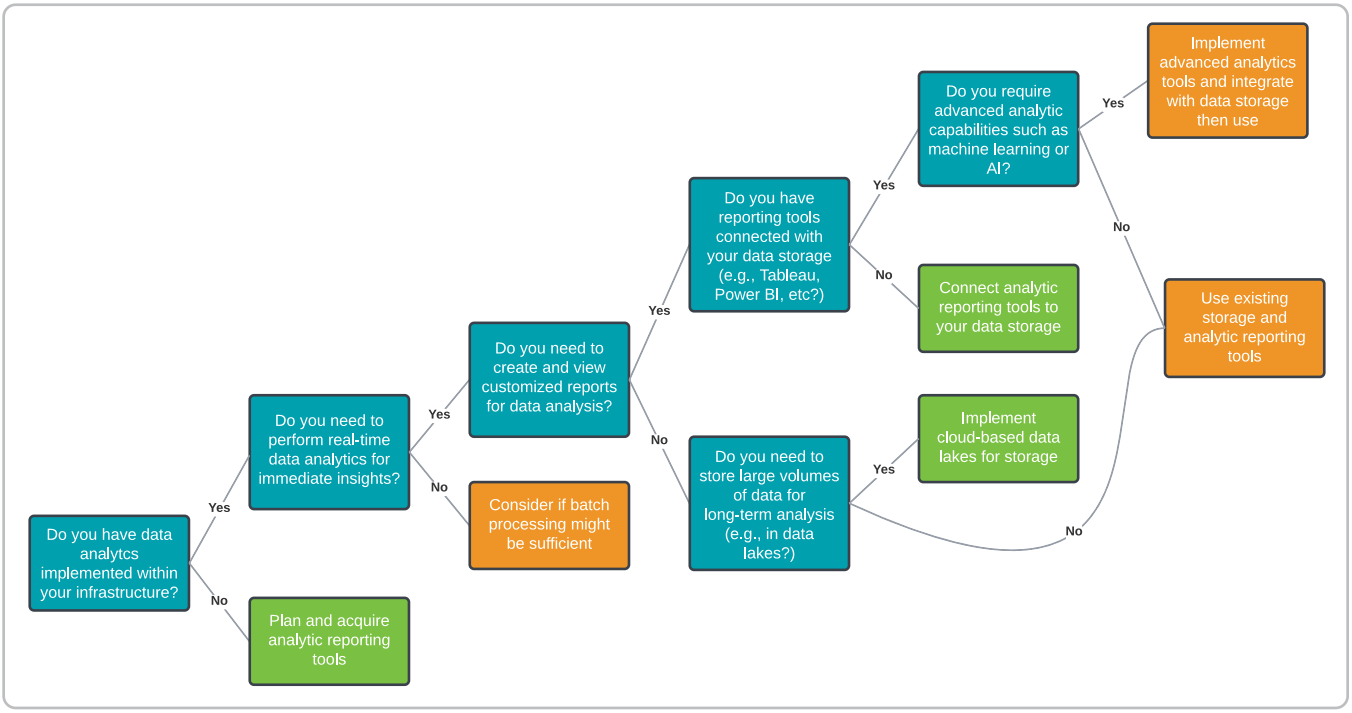
- **Realtime and Batch Message Processing:** To implement real-time processing pipelines use technologies like Apache Kafka, Apache Flink or AWS Kinesis for stream processing. If real-time transfer of data is not necessary, batch processing can be implemented using tools like Apache Hadoop, Apache Spark or traditional extract, transform, load (ETL) processes. Schedule jobs during off-peak hours to optimize resource usage. Reassess the need for real-time processing or hybrid approaches.
- **Data Validation and Cleansing:** Integrate data validation rules and cleansing processes early in the data processing pipeline. Use data quality tools like Talend, Informatica, or custom scripts to ensure data accuracy.
- **Data Transformation or Enrichment:** Utilize ETL processes to transform and enrich data. Consider using platforms like Mirth Connect, Apache Nifi, Talend or custom data transformation tools.
- **Support Multiple Data Formats:** Implement middleware or conversion tools to handle various formats (e.g., JSON, XML, HL7). Use a flexible processing engine that can adapt to different data types.
- **Complex Data Routing Rules:** Intelligent routing using middleware like Mirth Connect, Apache Camel or custom routing. Ensure that routing is flexible and can adapt to changing business requirements.
- **Data Aggregation and Summary Processing:** Implement aggregation logic in your processing pipeline. Use tools like Apache Spark, Flink or custom scripts to aggregate and summarize data.
- **Store Intermediate Data States:** Use a temporary storage layer (e.g., Redis, in-memory databases) to hold intermediate data. Ensure that this data is encrypted and managed according to compliance requirements.
- **Security and Compliance During Processing:** Ensure that all data processing components are compliant with relevant regulations (e.g., HIPAA, GDPR). Implement end-to-end encryption and maintain detailed audit logs.
- **Implement an Event-based Design:** Use queuing to make your system less vulnerable to failure during surges (see the AIMS Model "Data Processing" on page 23)
- **Implement OpenSearch/Elastic Search:** A key component of data intelligence and data governance. Where infrastructure is adequate to permit setting this up at an early stage of development, it will allow creation of dashboards that enable real-time and historical monitoring of data transactions, providing tangible evidence of what the platform is doing, and can establish how funds allocated to the project have been used.
- **Vocabulary:**
  - **Uniform terminology:** Standardized vocabularies ensure that the same terms are used consistently across different systems and regions. This uniformity is critical for accurate data interpretation and analysis. For instance, consistent use of disease names, symptoms and treatment protocols allows for precise tracking and comparison.
  - **Clear communication:** Standardized vocabularies enhance communication among public health professionals. When everyone uses the same terms, it reduces misunderstandings and ensures that data conveyed across different jurisdictions is interpreted correctly.
  - **Data integration:** With common vocabularies, integrating data from different systems becomes more straightforward. This is particularly important for combining clinical data, laboratory results and epidemiological information, enabling a more comprehensive understanding of public health issues.

Data Analytics

As discussed in the "The AIMS Model" (page 19), the data analytics component is responsible for examining and interpreting processed data to uncover patterns, trends and insights that help inform decisions (**Figure 22**). It should include statistical tools, algorithms and perhaps machine learning to analyze data and generate reports, insights or predictions into data exchange activities and system performance. These tools should allow stakeholders to monitor message traffic, identify trends and track system health in real-time for the public health data exchange.

Decision Tree

Figure 22. **Decision Tree: Implementing and using data analytics in the public health data exchange**



Considerations

The following list provides key considerations that will be useful in helping the reader align the functionality of the data analytics component with the non-functional needs of their public health data exchange:

- **Data Lakes:** Data lakes can store large volumes of structured and unstructured data, supporting diverse analytics use cases.
- **Advanced Analytics:** Advanced analytics capabilities, including machine learning and artificial intelligence, can offer meaningful insights on data flow and data content.
- **Integration:** Interoperability of stored data, and advanced analytics with ancillary applications can provide insights at an programmatic/application level.
- **Integration with Other Systems:** Ensure seamless connectivity with data sources, storage systems, and reporting tools for smooth data flow.
    - **Automation:** Automate repetitive tasks, such as report generation and alerts, to enhance efficiency and reduce manual work.
    - **User Access Control:** Define roles and permissions to restrict access to sensitive analytics or reports based on user roles.

Data Governance

Data governance is best represented by enterprise data management operations and are designed to enhance regulatory compliance, mitigate risks, improve operational efficiency, elevate data quality and maximize the value of data assets. Best practices related to enterprise data management have been codified by DAMA International in the Data Management Body of Knowledge (DMBOK2).*

Goals and Considerations

Data Governance Operations: Establishment of a data governance office (DGO) is highly advisable, for the purpose of overseeing data governance policies, standards, and compliance, and ensuring alignment with organizational goals.

- **Establish Internal Policies:** From the perspective of establishing internal policies for how data is gathered, stored, processed, and disposed of, the AIMS Platform establishes rules for how long data is allowed to persist in the systems that transmit it based on business rules and relevant HIPAA data storage laws. For example, eCR messages delivered to the Platform will be purged by AIMS after seven days if they are not picked-up and processed. Various ELR use cases require and allow us to maintain data for 30 days.

- **Automate Those Policies:** From the perspective of limiting access to data and controlling what people can do with it, the AIMS Platform employs a cloud service called Cloud Custodian to deploy and manage policies as code in an AWS environment. This tool offers comprehensive benefits in terms of security, compliance, cost management, and operational efficiency. It supports a proactive approach to cloud governance that is scalable, flexible and aligned with modern data governance practices.

  For institutions who host on premise, following the best practices that Cloud Custodian offers is advisable; however, there would be manual operations incurred. The main benefit of automated, repeatable and version-controlled policy-as-code practices is removal of human interaction which can lead to error—data being incorrectly deleted for example—and missed code runs.

  Best practices that the institution should follow, whether manual or automated, are:

  - **Garbage collection:** Identifying and deleting inactive users/keys, stopped instances, detached volumes, inactive secrets, and generally any unused resources. These policies should be layered to identify, alert, disable, alert again, and eventually delete based on prescribed timelines.
  - **Logging enforcement:** Monitoring and enforcing certain logging requirements such as VPC logs, access logs, etc to prevent drift and ensure visibility.
  - **Tagging:** Auto-tagging or manual tagging and enforcing the inclusion of others to standardize tagging across accounts and environments to better be able to breakdown and analyze costs, ownership, and purpose of resources.

- **Training and Awareness:** Implements regular training and awareness programs to educate stakeholders on data governance best practices, compliance requirements and data management roles.

- **Funding and Sustainability:** Secures the necessary funding for data management initiatives and ensures their long-term sustainability by aligning resources with strategic priorities and demonstrating ROI.

- **Engagement with Interested Groups:** Promotes collaboration among all interested group, including business units, IT, data owners and external partners, to align data management efforts with organizational needs and build a data-driven culture.

* DAMA International, The DAMA Guide to the Data Management Body of Knowledge (DMBOK2), 2nd ed., Technics Publications, 2017, Chapter 14: Data Management Operations, pp. 373-389.

Key considerations that may factor into data governance decisions and prioritizations include:

- **Regulatory Compliance:** Ensures that data management practices adhere to relevant laws and regulations (e.g., GDPR, HIPAA) to avoid legal penalties and maintain trust.
- **Risk Management:** Mitigates risks associated with data breaches, data loss and misuse by implementing data governance, security, and privacy measures.
- **Operational Efficiency:** Streamlines data-related processes, reducing redundancy and improving access to high-quality data, enhancing overall business performance.
- **Data Quality Improvement:** Focuses on maintaining high-quality data that is accurate, consistent and reliable to support effective decision making and operational activities.
- **Data Asset Value and Analytics:** Maximizes the value of data assets by enabling advanced analytics, intelligence and insights that drive innovation, strategy, and improved public health.

Interested Groups and Operations

Essential Business Elements to Support Any New Data Sharing Platform

- **Consider forming an internal integration team:** This is a key interoperability resource. Integration engineers have played a central role on the AIMS Platform.
- **Plan on providing extensive technical assistance (TA):** This is an indispensable element of successful deployment and adoption, and an ongoing support requirement. Every new data exchange partner will require TA to get connected. Partners will require TA for every new feature rolled out over the life of the platform.
- **Plan on instituting a service delivery management (SDM) team:** The SDM team will play a vital role in ensuring that the services provided by the platform are delivered to exchange partners in a reliable, consistent fashion, resolving disruptions to data flow expeditiously and taking appropriate corrective measures to prevent recurrence.
- **Offer production support to end users:** Envisage establishing a dedicated support team and optimizing their efforts by deploying a ticketing system to responsively manage issue resolution and maintain communication with data exchange partners.
- **Cross-functional technical team for maintenance and operations:** An interdisciplinary team comprising the skills of project management, engineering, technical support and more, will be needed to successfully maintain and operate a data exchange platform.

Advocacy and Persuasion

As with any complex undertaking requiring buy-in and support from a large number of interested groups—many of whom will likely have misgivings about the impact of unfamiliar systems and methods on their own workload—establishing a data exchange platform will be greatly accelerated if championed by a person or persons who are able to assume the role of a strong advocate or sponsor.

This initial advocacy may be required for an extended period. In the early days of the AIMS Platform, leaders of state public health laboratories had to be personally contacted and convinced, one by one, of the benefits of the new platform. Once leaders started to come on board, the next hurdle was to persuade their own skeptical security personnel of the safety of the proposed systems. Those new systems then had to be built and tested, and a vast array of technical challenges had to be overcome (e.g., getting LIMS data into HL7 messages).

The process of expansion to additional state public health laboratories gained momentum as the AIMS Platform was seen to comprehensively address problems that laboratories experienced, as the benefits of the platform became more tangible, resulting in trust building over time. When the Platform began offering formal technical assistance to jurisdictions, growth accelerated tremendously.

Thoughtful communication is critical in this campaign to cultivate buy in and participation from stakeholders at all levels. In particular, clearly articulate the benefits of inter-jurisdictional data exchange, including:

- **Improved Data Quality:** Standardized data formats and vocabularies enhance the quality of data collected and shared. This leads to more reliable public health surveillance and decision making, as data from different jurisdictions can be confidently aggregated and analyzed.
- **Enhanced Response Coordination:** In public health emergencies, such as disease outbreaks or natural disasters, quick and coordinated responses are vital. Standardized data allows for rapid sharing and interpretation of critical information across jurisdictional lines, facilitating more effective and timely interventions.
- **Regulatory Compliance:** Many public health initiatives require compliance with federal regulations and reporting standards. Standardized data formats and vocabularies ensure that data shared between jurisdictions meets these regulatory requirements, avoiding legal and procedural complications.
- **Resource Optimization:** With standardized data, jurisdictions can share resources more effectively. For instance, data analytics tools and platforms developed in one jurisdiction can be easily adopted by others, reducing duplication of effort and fostering collaboration.
- **Public Health Research:** Standardized data enables multi-jurisdictional research initiatives, allowing for larger and more diverse data sets. This enhances the validity and generalizability of research findings, ultimately contributing to more effective public health strategies and interventions.

Cultivating Buy-in With Interested Groups

Securing buy-in and approvals for hosting a centralized data exchange platform typically requires several critical steps.

- **Engagement:** Involve key interested groups early in the planning process to address concerns and gather input.
- **Demonstrating Benefits:** Clearly communicate the advantages of a centralized data exchange platform over legacy systems, including efficiency, cost savings, scalability and improved security.
- **Risk Assessment:** Conduct thorough risk assessments and address potential security and compliance issues.
- **Pilot Projects:** Implement pilot projects to demonstrate the feasibility and benefits of hosting a data exchange platform on a small scale.
- **Continuous Communication:** Maintain open lines of communication with stakeholders throughout the process to build trust and address any emerging concerns.

Cultivating Trust with Data Exchange Partners

Cultivating trust with data exchange partners is essential to their ongoing participation in the system. The centralized data exchange team can build trust with data exchange partners over time by showcasing a strong background and proven track record in the field. Demonstrating the potential benefits and successes of hosting a data exchange platform in similar settings can build confidence. Highlighting case studies or examples of successful implementations in other jurisdictions or regions can reinforce credibility. Strategies for cultivating trust with data exchange partners include:

- **Clear Communication of Benefits:** Effectively communicating the advantages of a data exchange platform—such as enhanced data accessibility, scalability, cost savings and improved public health outcomes—is crucial. Sharing a clear, compelling vision and how it aligns with the jurisdictions' goals for public health can foster buy-in.
- **Transparency and Engagement:** Engaging stakeholders in open dialogues, addressing their concerns and providing transparent information about the process can build trust. Involving key stakeholders from the outset and making them feel part of the decision making process is essential.
- **Demonstrating Security and Compliance:** Emphasize the robust security measures and compliance with relevant regulations (e.g., HIPAA) that such a platform can offer. Providing detailed information on how data privacy and security are maintained can alleviate concerns and build trust.

Critical Buy-in: IT Security Teams

It is essential to have the understanding and commitment to ongoing participation from the IT security teams of each data exchange partner. Strategies to cultivate this buy in include:

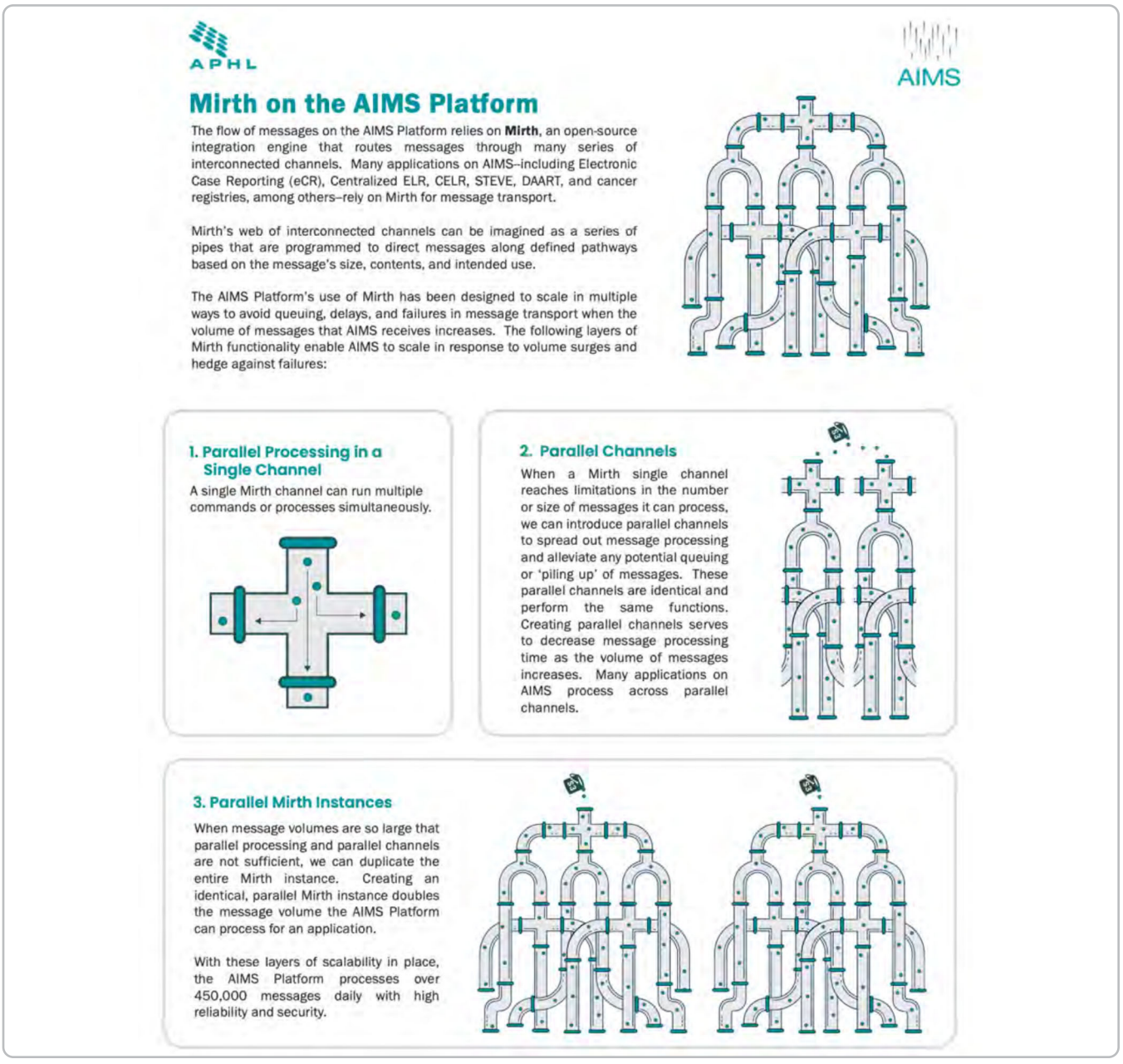
- **Involve IT Security Early:** IT security teams should be involved from the beginning of the project to ensure their concerns and requirements are addressed. Engaging them early helps in identifying potential security challenges and designing solutions to mitigate risks.
- **Conduct Comprehensive Security Assessments:** Conduct thorough security assessments of the platform's planned infrastructure and services.
- **Offer Customization and Control:** Ensure that IT security teams will have control over security configurations, such as access controls, encryption and monitoring. Provide options for customization to meet specific security policies and requirements of the organization.
- **Provide Ongoing Training and Support:** Offer training sessions for IT security teams to familiarize them with security best practices relevant to the selected hosting solution. Provide continuous support and updates to address new security threats and vulnerabilities.

Appendices

1. Parallel Message Processing on the AIMS Platform

An integration engine is essential for orchestrating the transmission of messages from sender to receiver across systems. When designed properly to queue messages ahead of the integration engine, the integration engine can provide parallel processing across channels and instances, enabling reliable scalability for the platform. There are many integration engines that can be utilized for a centralized data exchange platform. For example, the AIMS Platform uses Mirth, an open source integration engine that also offers licensed managed services. **Figure 23** describes how AIMS facilitates parallel message processing using Mirth.

Figure 23. **Diagram of AIMS Platform Using Mirth**



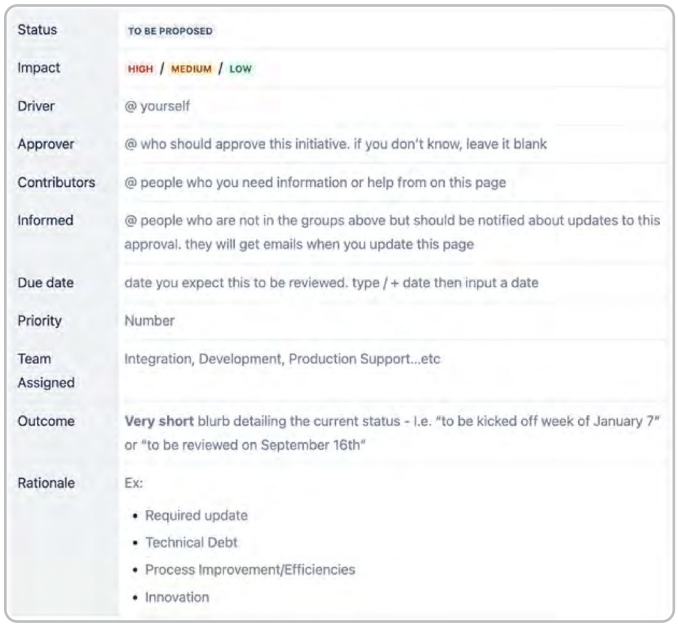
2. Illustrative Policies and Procedures

As with any consequential endeavor, project management will be indispensable to successfully shepherd a nascent data exchange throughout all the stages of its conceptualization and design, secure buy-in and approval, development and deployment, technical assistance and support, and ongoing development phases. Naturally, documentation is a key tool for managing such complex efforts, ensuring that necessary measures are understood and implemented in a correct and timely manner. Beyond that, because of the sensitive nature of public health data being stored and exchanged, establishment of security policies and maintenance of corresponding documentation will likely be significant legal requirements of operation. This section provides a representative selection of some of the types of policies, procedures and documentation implemented for the AIMS Platform.

Architecture and Infrastructure

- **Data visualizations and dashboards:**
  - Maintain history of application(s) (major decisions made, changes in approach, etc.)
  - Documentation on tool(s) used; if third party, include contract information, renewal date(s), cost and contact information for liaison
  - Upgrade and maintenance process
  - User guide(s) for app(s)
- **Road map:** Outline what projects are desired by leadership to maintain and improve the platform and a change control process for discussing:
  - Risks
  - Benefits
  - Costs
  - Interested parties
  - Priority against other initiatives
- **Documentation** of all services and/or locations of servers; redundancy, throttling, failover, etc.

Figure 24. Sample decision matrix for vetting and approving projects



Status	TO BE PROPOSED
Impact	HIGH / MEDIUM / LOW
Driver	@ yourself
Approver	@ who should approve this initiative. if you don't know, leave it blank
Contributors	@ people who you need information or help from on this page
Informed	@ people who are not in the groups above but should be notified about updates to this approval. they will get emails when you update this page
Due date	date you expect this to be reviewed. type / + date then input a date
Priority	Number
Team Assigned	Integration, Development, Production Support...etc
Outcome	**Very short** blurb detailing the current status - i.e. "to be kicked off week of January 7" or "to be reviewed on September 16th"
Rationale	Ex: • Required update • Technical Debt • Process Improvement/Efficiencies • Innovation

Standards, Security and Compliance

Security and compliance of data exchange infrastructure needs to include:

- **Audit documentation**
- **Agreements:** Contractual agreements (e.g., service level agreements (SLAs), data use agreements (DUAs), master service agreements (MSAs), business associate agreements (BAAs), etc.)
- **Compliance schedule of activities**
- **Artifacts of service providers**

Documentation should include:

- **Public-facing Audit Documentation:** Outputs of an audit might include:
  - **Security Assessment Plan:** a document consisting of a test plan to test the security controls for the AIMS Platform. This plan is completed by a third-party assessment organization for the benefit of the hosts/owners of the AIMS Platform
  - **Executive Summary report:** a document developed for senior management, providing a concise summary of the overall review
  - **Full Report:** including all the detailed criteria and responses, along with specific recommendations identified to be addressed prior to the next accreditation cycle
- **AIMS Platform Policy Documents**
  - AWS Resource Tagging Procedure*
  - Contingency planning documents and exercise outcomes
    - Contingency plan exercise document
    - Application failover checklist
    - Incident response tabletop exercise
    - Process and checklist for system and vendor selection, i.e.:
      - Identity service provider vendor assessment
      - Security information and event management (SIEM) vendor assessment
      - Electronic data capture
      - Identity platform solution assessment
      - Security platform / Threat hunting service assessment
      - Open security controls assessment language (OSCAL) vendor assessment
      - Security and awareness training evaluation and course
    - Team member onboarding procedure
    - Production readiness reviews; Platform procedure documents developed for the AIMS Platform

* Cloud-specific process or procedure (an on-premises equivalent may be relevant)

- National Institute of Standards and Technology (NIST) 800-53 Control Family Procedure
- Access control procedure; Mobile access control policy and procedure
- Audit and accountability procedure
- Awareness and training procedure
- Configuration management procedure
- Contingency planning procedure
- Identification and authentication procedure
- Incident response procedure
- Maintenance procedure—Cloud service provider (CSP) inherited
- Media protection procedure (CSP inherited)
- Planning policy
- Personnel security procedure; Position risk designation procedure
- Physical and environmental protection procedure (CSP inherited)
- Risk assessment procedure
- Security assessment and authorization procedure; Continuous monitoring strategy and procedure
- System and communications protection procedure; Boundary protection procedure
- System and information integrity procedure
- System and services acquisition procedure
- Electronic Healthcare Network and Accreditation Commission (EHNAC)

- **Audit Activities Schedule**

Data Governance and Policy

Processes to establish data management procedures that keep data secure, private, accurate and usable throughout the data life cycle. With data exchange partners:

- **Hold discussions and open forums, such as:**
  - Interoperability Forum — DQA
  - QA Discussion
- **Form governance board:** Select and empower a small group of forward-thinking or cutting-edge partners to form a governance board and set rules or recommendations for the whole trading partner body. Hold monthly or quarterly reviews of standards with that group.
- **Provide posted accessible resources, such as:**
  - CDC Laboratory Quality Assurance and Standardization Programs (cdc.gov)
  - Working Together to Improve Citizen Science Data Quality: A Guide for Government Agencies
  - Trusted Exchange Framework and Common Agreement (TEFCA) | HealthIT.gov
  - Quality Assurance Community Discussion (aphl.org)
  - Implementation Guide Schematic or process map
  - Data Quality Schematron; example:
    Public health representatives have expressed a desire to improve the quality of the EHR data made available through eCR. The Implementation Guide (IG) Schematrons are not enough to assess and ensure data quality. To address these concerns the Quality Assurance (QA) Workgroup convened over 31 PHAs to develop data quality rules related to the validity and completeness of eICRs. Before the QA Workgroup, efforts focused on checking incoming XML files for conformance to CDA standards and on supporting RCKMS reportability determination needs.
  - Data mapper tool; Release notes
  - Feed splitter
  - Document or data standard template examples
- **Host a repository of internally accessible documentation; example:** *Data dictionary for indices or DBs.*

Maintenance and Operations

A knowledge base, publicly accessible by any engineers who work on the platform, with documentation describing processes and procedures such as the following:

Processes

- **Project coordinator procedures**
  - Process for initiating a new maintenance initiative
  - Deployment process documentation for all applications hosted on the platform
- **Service desk or customer services procedures**
  - Identity and access management
    - Access request review procedure and approval workflow for any identity and access management tools
    - Network Access Request Form template and approval workflow
    - Any application specific access workflow procedures

- Service desk (AIMS Platform) standard operating procedure — Triage
- Service desk standard operating procedure — Triage
- “Three-Contact Rule” process
- “Three-Contact Rule” external email process and template
- Extended hours
- Daily dashboard checks — process
- Splunk on call alarms production support (9-5)
- Certificate updates
- Including contacts and response — Triage
- Distributing weekly updates for tickets — Process
- Creating a ticket for SQS alerts on Jira/ by Email — Process*
- Creating APHL customer portal account — Process
- Requesting technical assistance
- Change request submission

Technical Guides

- FREERadius MFA for AWS Workspaces Troubleshooting-Process*
- Reprocessing mismatched timing/errored batched messages in Opensearch — Process
- SQS queue Issue — Deletes Failing*
- CheckMK Monitoring Configuration — Process
- S3 Connection Mirth Setup Guide*
- SFTP channels in Mirth
- Resetting AD Password
- Elasticsearch Reindexing
- Datadog Installation and Deployment — Process
- Active Directory and User Access Provisioning
- VPN troubleshooting
- AIMS Troubleshooting S3 Transport*
- Post Mirth restart prod4, prod5 or prod6 process
- Rhapsody self signed cert update

Service Delivery Management

A space designed to meticulously track and manage all facets related to the delivery of services on the platform. The primary aim of this space is to create a centralized hub where team members can access updates, share insights and collaborate on solving challenges that pertain to service delivery. This will not only enhance transparency but also foster accountability and improve efficiency across the entire service delivery chain.

These processes should follow ITIL v4 / ITSM best practices

* Cloud-specific process or procedure (an on-premises equivalent may be relevant)

- Service Level Compliance: this page should house data pertaining to the availability, incident time to report, and change request time to resolution SLAs. Best practice includes automation of the calculation of SLAs, ensuring accurate and up-to-date reporting, transparency and efficiency.
- Define and document the targets and minimum values for time to notify, time to resolution and time to report SLA
- Define and document availability, incident time to report, incident time to notify, incident time to resolution, change request time to notify, change request time to resolution, administrative request time to notify, administrative request time to resolution
- Provide and document monthly service reviews to rigorously assess the quality, efficiency and effectiveness of our service delivery mechanisms in order to:
    - Streamline communication among team members
    - Enhance accountability and transparency
    - Enable data-driven decision making
    - Foster continuous improvement in service delivery
- Service delivery coordinator procedures
    - Detail the standard operating procedure for preparing an incident report for an incident occurring involving an application hosted on the AIMS Platform
    - Detail the standard operating procedure for preparing for a software release for an application hosted on the AIMS Platform
    - Detail the standard operating procedure for preparing a problem investigation for an application hosted on the AIMS Platform

Technical Assistance and Onboarding

Provide:

- A master technical assistance (TA) resources folder where users can access what they need at every stage of onboarding:
    - Pre-implementation
    - Planning
    - Validation and testing
    - Production go live
    - Evaluation
    - Maintenance
- FAQs, including:
    - What is outlined in TA?
    - What is offered?
    - Who qualifies?
    - How do qualified users request it?
    - Expectations for speed of TA
    - How to check status (tickets, website, other)
- Tools and resources, such as:
    - Code packages
    - Custom snippets
    - A place where users can post tools and resources for their colleagues (an open-source marketplace of sorts)

Interested Partner Engagement

- Give users access to a public-facing content wiki, such as Confluence, with folders dedicated to them
- Weekly call with open forum for all 55 state, local and territorial public health laboratories and agencies (PHAs and PHLs) where questions or issues are submitted ahead of time and discussed, webinar-style as a group. Notes and recordings are published and accessible async by all participants

Strategy and Leadership

- Change control board (CCB)
- Empower a small but diverse group of leaders to maintain a strategic vision for the platform. All decisions about projects, infrastructure, and spending should connect back to that vision.

Communications and Marketing

- Communications to existing users and data exchange partners:
  - Service desk standard operating procedure — Triage
  - "Three-contact Rule" process
  - "Three-contact Rule" external email process and template
  - Two week, two day and post-completion notice SOPs on maintenance and downtime
  - Branded materials and email templates for notices
- Communications to subscribed or opted in contacts (e.g., monthly newsletter)

3. Cloud Hosting Considerations

Attractions of the Cloud

In keeping with practices widely adopted by governments and businesses of all kinds and sizes, many public health organizations have made the move to the cloud from local physical servers. These servers were typically located on-premises within the organizations' facilities or in hosted data centers. This kind of setup often involves disadvantages like the following:

- **High Maintenance Requirements:** Regular updates, backups and hardware replacements are necessary.
- **Limited Flexibility:** Scaling up requires purchasing and installing new hardware, which is both time-consuming and costly.
- **Vulnerability to Physical Threats:** Local servers are susceptible to damage from natural disasters, power outages and other physical threats.

Challenges of Maintaining On-premises Resources

Managing local physical servers presents several significant challenges:

- **Labor-intensive Management:** IT staff spend considerable time on routine maintenance tasks, detracting from their ability to focus on strategic initiatives.
- **Scalability Issues:** Increasing storage or processing capacity requires significant capital investment and lead time for procurement and installation.
- **Limited Accessibility:** Access to data is often restricted to on-premises locations, hindering remote work and collaboration.
- **High Risk of Data Loss:** Physical server failures or disasters could result in significant data loss if not properly backed up.

Impetus for Migrating

Several factors motivate the decision to migrate to the cloud.

- **Need for Scalability:** Public health organizations require a more flexible and scalable solution to handle increasing data volumes and growing demands.
- **Cost Considerations:** Cloud solutions often offer more cost-effective pricing models, reducing the need for large upfront investments in hardware.
- **Desire for Enhanced Collaboration:** Cloud platforms facilitate better data sharing and collaboration among public health stakeholders.
- **Focus on Innovation:** Migrating to the cloud allows IT staff to focus on innovative projects rather than routine maintenance.

Benefits from Being in the Cloud

Hosting in the cloud brings numerous benefits.

- **Scalability:** Cloud solutions provide the ability to easily scale up or down based on demand, ensuring that resources match needs.
- **Cost Efficiency:** The pay-as-you-go model reduces capital expenditures and optimizes operational costs.
- **Accessibility:** Cloud platforms enable remote access to data, supporting flexible work arrangements and improving collaboration.
- **Reliability:** Cloud providers offer robust backup and disaster recovery solutions, reducing the risk of data loss.
- **Security:** Leading cloud providers have implemented advanced security measures, often surpassing those of on-premises solutions.
- **Innovation:** Cloud platforms support advanced analytics, machine learning, and other innovative technologies, enabling more effective public health interventions.

Selecting a Cloud Service Provider

Choosing the right cloud vendor involves appropriate research.

- **Evaluate Vendor Reputation:** Considering the track record and reliability of potential vendors.
- **Assess Service Offerings:** Ensuring the vendor's services match the organization's needs, including support for specific data formats and analytics tools.
- **Review Security Measures:** Verifying the vendor's security certifications and compliance with relevant standards.
- **Cost Analysis:** Comparing pricing models and total cost of ownership across different vendors.

Table 3a. Considerations for Cloud vs. Locally Hosted Solutions

Consideration	Cloud	On premises
Compliance	• Jurisdiction-specific regulations may require data to be stored within specific geographic boundaries • Adherence to applicable local privacy regulations such as GDPR, HIPAA, and other data protection laws may be required	
Security	• Number one priority of major cloud providers • Security standards attested by independent audits	Ensuring security may account for 50% of the workload
Cost Total cost of ownership, including upfront investments, ongoing maintenance, and operational expenses	• Pay only for the services used, bypassing fixed expenses (e.g., hardware, site) • With large cloud service providers, economies of scale result in lowered costs	Pay up front for the physical infrastructure, hardware, software, maintenance
Scalability Ability to scale resources up or down based on variable demand	• Allows provisioning only the amount of resources actually needed. Scale resources up or down to instantly grow and shrink capacity as needs change. • Scalable cloud storage options resolve many interoperability problems relating to file size and volume limitations	• Estimating resources needed ahead of time may result in insufficient or excess capacity • Scaling up may be costly and labor intensive, entailing purchasing and provisioning additional physical hardware, and installing software on site
Agility Ability to easily add / remove infrastructure or applications	• Easy access to broad range of technologies that can be deployed in minutes supports innovation • Ability to rapidly add / remove infrastructure facilitates testing	To test new applications, may have to purchase and provision additional physical hardware, install software on site
Support Technical support for problem resolution	Large support staff available 24/7 to respond to issues	Limited support staff to respond to issues
Maintenance	Infrastructure maintenance is performed by cloud service provider	Regular updates, backups, and hardware replacements are necessary
Risk of service interruption or data loss	Cloud service providers rely on redundant geographically distributed infrastructure to minimize risk	• Susceptible to damage from natural disasters, power outages, and other physical threats • Physical server failures may result in significant data loss if not properly backed up
Support for Advanced Functionality Data Lakes, Advanced Analytics, Integration	• Cloud-based data lakes allow storage of large volumes of structured and unstructured data, supporting diverse analytics use cases • Cloud platforms' advanced analytics capabilities include machine learning and artificial intelligence to derive insights from data • Permits seamless integration with existing tools and systems to maximize the value of cloud-based analytics	

4. Terminology

APHL — Association of Public Health Laboratories: A membership organization in the United States representing the laboratories that protect the health and safety of the public. In collaboration with members, APHL advances laboratory systems and practices, and promotes policies that support healthy communities, serving as a liaison between laboratories and federal and international agencies, and ensuring that the network of laboratories has current and consistent scientific information to be ready for outbreaks and other public health emergencies

Cloud computing: The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server

HIS — Health Information Systems: Any system that captures, stores, manages or transmits information related to the health of individuals or the activities of organizations that work within the health sector

Hybrid cloud: A cloud computing environment which uses a mix of on-premises, private cloud and third-party, public cloud services with orchestration between the two platforms

Integration: The combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information. A complete data integration solution delivers trusted data from a variety of sources.

LIS — Laboratory information systems: A software system that records, manages, and stores data for clinical laboratories. A LIS has traditionally been most adept at sending laboratory test orders to lab instruments, tracking those orders, and then recording the results, typically to a searchable database

Mirth: Mirth Connect is a cross-platform integration engine used in the healthcare industry that enables the management of information using bi-directional sending of many types of messages.

Private cloud: A type of cloud computing that delivers similar advantages to public cloud, including scalability and self-service, but through a proprietary architecture. Unlike public clouds, which deliver services to multiple organizations, a private cloud is dedicated to a single organization

Cloud / public cloud: A type of cloud based on the standard cloud computing model, in which a service provider makes resources, such as servers, applications and storage, available to the general public over the Internet.

Ruvos: APHL's technical partner for the AIMS Platform. Ruvos' technical subject matter experts were integral in the initial design, architecture, and build of the AIMS Platform, as well as its migration to the cloud and ongoing scaling and development.

TEFCA — Trusted Exchange Framework and Common Agreement: Created by the Office of the National Coordinator for Health IT (ONC) to become a system of interconnected organizations called Qualified Health Information Networks (QHINs) that improve nationwide health-data exchange. In short, this is essentially a network of networks, designed to improve healthcare interoperability.

Association of Public Health Laboratories

The Association of Public Health Laboratories (APHL) works to strengthen laboratory systems serving the public's health in the US and globally. APHL's member laboratories protect the public's health by monitoring and detecting infectious and foodborne diseases, environmental contaminants, terrorist agents, genetic disorders in newborns and other diverse health threats.

7700 Wisconsin Avenue, Suite 1000 Bethesda, MD 20814 | 240.485.2745 | www.aphl.org

This document was prepared by Ruvos, LLC, the long-term technical partner of the Association of Public Health Laboratories (APHL), as part of Work Order Agreement #56401-700-800-24-04.

Funding for this publication was provided under Cooperative Agreement #NU2HGH000080 from the US Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of, nor an endorsement by, CDC the US Department of Health and Human Services or the US Government.