WHITEPAPER

The National Public Health Data System:

The case for an aggregated national healthcare database

Introduction

The COVID-19 pandemic exposed major gaps in U.S. public health data systems – interfering with the government's ability to effectively conduct surveillance, outbreak modeling, and research. The lack of complete federal data on chronic diseases has also hindered the ability to achieve major public health goals, such as eliminating health disparities. The challenge centers around the nation's public health surveillance system requiring standards, organization, and streamlining capable of providing complete, centralized data to inform public health responses. In response to these issues, the federal government has expressed interest in constructing a centralized, cloud-based National Public Health Data System (NPHDS).

The NPHDS from Eagle Technologies, Inc., and Pegasystems, Inc., (Eagle-Pega) will help ensure that the U.S. is prepared for the next pandemic and can detect other emerging health threats, to minimize the loss of life. It will also provide federal agencies conducting budgeting, planning, and policymaking with complete U.S. health data as they make crucial decisions. The Eagle-Pega NPHDS will be an invaluable resource for government agencies such as the National Institutes of Health (NIH), the Centers for Disease Control and Prevention (CDC), and the Assistant Secretary for Planning and Evaluation (ASPE) as they conduct research that may lead to disease prevention and cures, and improvements in public health policy and practice.

This whitepaper presents the concept for the Eagle-Pega NPHDS, including the federal and commercial data sets to integrate and provide necessary elements needed to aggregate health data on the entire U.S. population. It reviews the technologies and strategies that could be used for aggregation to achieve a complete database, and discusses data storage and security. It addresses additional services and capabilities that can be gained by leveraging intelligent automation and case management, once a central NPHDS has been achieved. Finally, this whitepaper presents a related use case with the 2020 Decennial Census – a project of similar scale, scope, and complexity to the NPHDS – where the federal government successfully used industry technologies to collect data on every household in America. The potential consequences for U.S. public health if a NPHDS is not created are painfully apparent. The time to consolidate public health data is now.

Whitepaper Co-Authors

David Dastvar serves as Chief Growth Officer with Eagle Technologies, Inc. In his 29 years with public sector and Fortune 1000 companies (including CSC/GDIT, Infosys, CDI, and Northrop Grumman), he has developed and managed professional services and solutions for enterprise-level projects requiring a high degree of program management and technical expertise. David has an engineering degree and management training from George Washington University. He can be contacted at david.dastvar@eagletechva.com or (202) 497-8848.

Douglas Averill is Global Industry Market Leader for Pegasystem's Government business. With much of his career in the global public sector space, Doug now leads Pega's government go to market team as it supports agencies engaged in modernizing their enterprise. He is responsible for government-specific strategic initiatives, compliance, content, applications, and government to government collaboration. During his two decades in the private sector he co-founded a start-up, managed large-scale commercial projects, and lead technical sales teams. Doug holds a BS in Biology from Dickinson College in Carlisle, Pennsylvania and an MBA from Babson College in Wellesley, Massachusetts.

Centralizing siloed data: The case for the National Public Health Data System

The challenge is not a lack of health data, but a lack of centralization of that data. Data is currently being collected independently by various federal, state, and local agencies. The data landscape consists of many different silos. This makes it difficult for the federal government to develop real-time insights on that information.

The Eagle-Pega NPHDS will serve as a centralized, cloud-based data collection system that can pull information from each of these disparate data sets and feed it into a central repository. Once the data has been aggregated and normalized, the federal government could then use the power of decisioning, data analytics, artificial intelligence (AI), and case management to act on this information.

Sourcing the right data

The Eagle-Pega NPHDS will have the capacity to contain data on every U.S. resident – with data coming from federal and commercial databases alongside state and local health agencies, testing entities, and providers. Data about the long-term effects of COVID-19 on population health will be prioritized at the start of the NPHDS's construction. Its initial data will be comprised of COVID-19 cases from public health agencies, COVID-19 vaccinations from the Vaccine Administration Management System (VAMS), and chronic and other conditions in inpatient and outpatient data sets from the Centers for Medicare & Medicaid Services (CMS) and the Children's Health Insurance Program (CHIP). Information will also be collected from the Surveillance, Epidemiology, and End Results (SEER) program cancer database, and the Healthcare Cost and Utilization Project's (HCUP) hospitalization and emergency department use data sets. Finally, information on births and deaths from the National Vital Statistics System will also be included.

Together, these data sets will allow the government to track many consequences of the pandemic – from cases of long-haul COVID-19 to new cases of diabetes, mental illness, and other health conditions triggered by COVID-19 infection. Additionally, missed cancers and heart attacks that occurred during the pandemic, among other sequelae, will be identified and tracked.

Designing the right platform for data collection

The sheer size and magnitude of the NPHDS requires that the ideal platform not only support the collection, aggregation, continual updating, and storage of health data – but also be capable of integrating with existing and future health systems and technologies.

Eagle Technologies, as the systems integrator, will utilize tools like the Pega Government Platform™ – designed to operate within enterprise ecosystems. This platform technology includes a variety of services, connectors, transformations, and robotics that allow it to bi-directionally integrate with third-party products and systems. The Eagle-Pega team's extensive experience in the Healthcare industry means that the platform is equipped with a ready-to-use Fast Healthcare Interoperability Resources (FHIR) API, enabling the exchange of healthcare-related information (i.e., clinical, public health, and research data). This capacity is critical for the NPHDS' ability to collect healthcare data from testing entities and providers.

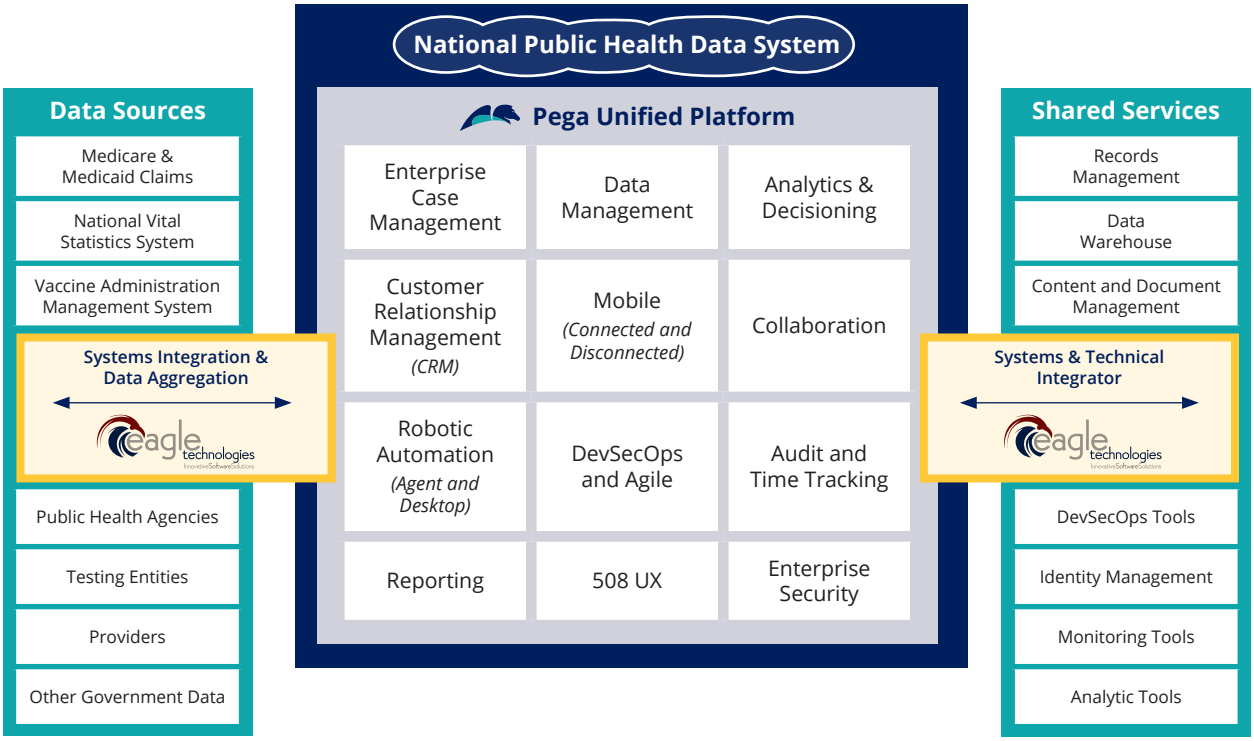


Figure 1. NPHDS data sources, platform, and shared services

Figure 1 presents the data to be ingested by the platform and the platform's shared services. These integrations can be configured within the platform itself and include capabilities to authenticate – ensuring transmitted data remains secure throughout the transaction.

For data storage, Pega uses industry-standard database systems Oracle, DB2, MS SQL, and Postgres. This includes the storage and management of configuration models (Pega repository), as well as runtime instance data, including file attachments (case/process runtime data). Pega automatically supports the separation of runtime data on the persistence level (database),

where the data for different process types (work objects) are stored in separate database tables out-of-the-box. The database mapping of application objects and physical database tables in Pega are configurable.

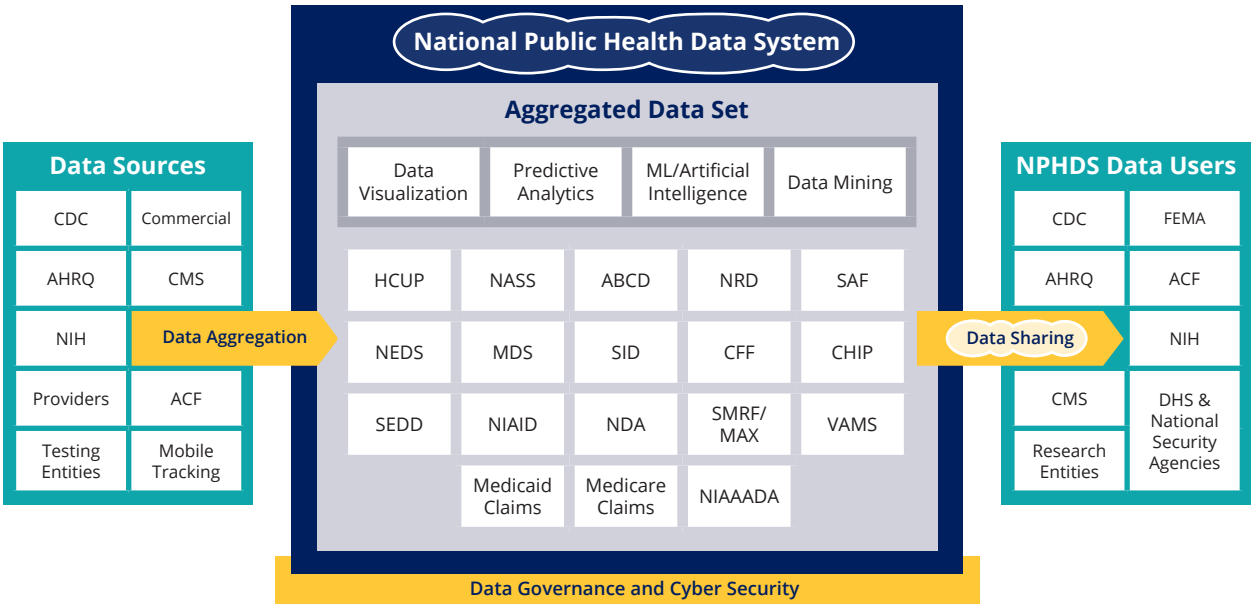
Regardless of which cloud provider is used for the Eagle-Pega NPHDS, each agency or organization can continue to own its core/legacy data. Data normalization would then need to take place – dependent on the individual data sources and formats that feed into the Eagle-Pega NPHDS. Once the data are normalized, the Pega Government Platform would act as an orchestration layer that works with the data to drive meaningful outcomes in the context of cases. As the intelligent automation platform at the center of this solution, the Pega Government Platform would be the system of record for all cases. The case data will be stored in the Pega repository in a non-proprietary, standard relational database.

Once the case data are created, the platform can bi-directionally integrate with a variety of third-party products via a comprehensive suite of technology-level adapters – allowing applications to fit seamlessly into an enterprise complex ecosystem. Additionally, the Pega Government Platform contains an extract, transform, load (ETL) component called Business Information Exchange (BIX). The BIX extracts data from Pega applications and exports it in formats that are suitable for import into popular business intelligence applications. This allows the export of service case data, process execution (data marts and data warehouses) for analysis and reporting, and other records stored in the scope of the platform in a defined data format, such as XML, CSV, or directly to an external database.

The Pega Government Platform is designed for real-time, bi-directional integration with disparate systems of record and third-party applications – so that existing data and processes can be swapped and renewed. This ensures that the back-end complexities of an organization's IT solutions can be masked in the end-user experience. The platform can integrate with and pull/push data from the underlying data stores, ensuring effective preservation and embellishment of the most up-to-date data.

Aggregating, storing, and securing the data

Once the data are in the platform, they will be aggregated, normalized, stored, and secured. The NPHDS's health and other data will be culled from disparate sources, including federal agencies, commercial firms, state and local health agencies, testing entities, and healthcare providers. These data will have been developed independently to serve different purposes. Therefore, they will need to be transformed, normalized, cleaned, integrated, and stored in a unified NPHDS data set. This data set can consist of a combination of data lakes, data warehouses, and other data repositories. The data will later be shared among NPHDS data users, including federal agencies, research entities, and commercial industries, in a unified format. Figure 2 depicts the Eagle-Pega NPHDS's data aggregation and sharing processes.



ABCD: Adolescent Brain Cognitive Development Study
ACF: Administration for Children and Families
AHRQ: Agency for Healthcare Research and Quality
CDC: Centers for Disease Control and Prevention
CFF: Connectome Coordination Facility
CHIP: Children's Health Insurance Program
CMS: Centers for Medicare & Medicaid Services
HCUP: Healthcare Cost and Utilization Project
MDS: Minimum Data Set
NASS: Nationwide Ambulatory Surgery Sample
NDA: National Institute of Mental Health Data Archive
NEDS: Nationwide Emergency Department Sample
NIAID: National Institute of Allergy and Infectious Diseases
NIAAADA: National Institute on Alcohol Abuse and Alcoholism Data Archive
NIH: National Institutes of Health
NRD: Nationwide Readmissions Database
OAI: The Osteoarthritis Initiative
SAF: Hospice Standard Analytical File
SEDD: State Emergency Department Databases
SMRF: State Medicaid Research Files
VAMS: Vaccine Administration Management System

Figure 2. NPHDS data aggregation and sharing process

Building a foundation for data governance and cybersecurity for the data in transit and at rest is paramount. Rather than current storage in legacy systems, we leverage DevSecOps best practices together with cloud technologies to better govern, access, and store data in multi-cloud environments, under identity and access management. The cloud provides flexible security settings toward its service models, e.g., Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Moreover, the cloud secures data, folders, and file activities while maintaining complete visibility to all users. The cloud system proactively identifies and mitigates risks, including security threats, suspicious user behavior, and malware.

The cloud provides high reliability and redundancy – preventing and detecting any data loss or other disruptions. And to maintain high security, the data will be encrypted when shared among NPHDS users and the Eagle-Pega NPHDS will only share the data each individual user needs.

The process of data aggregation will be highly automated by the Eagle-Pega platform, yet will also require close monitoring to ensure data quality and standardization. Every data operation in and out must be properly documented for the purpose of managing data availability, usability, integrity, and security. This provides high quality data while ensuring it will not be misused.

Visualizations, advanced analytics, data mining, machine learning, and AI will be deployed to track and monitor data accessing behavior, as well as gain insights into public health trends – all in real time. If there are anomalies in data accessing, the NPHDS system's analytics will alert system administrators in real time so that protective and restorative actions can be taken as soon as possible.

Supporting technologies, methodologies, and strategies

The creation of the Eagle-Pega NPHDS will incorporate a planned approach to combat typical challenges encountered with large-scale system integrations, such as the Department of Transportation (DOT) integrating financial invoices and processes online for all DOT agencies – including the Federal Aviation Administration (FAA), the National Highway Traffic Safety Administration (NHTSA), the Federal Highway Administration (FHWA), and the Federal Transit Administration (FTA). A unified platform that offers intelligent automation through case management, business processes, and a business rules engine that can operate at massive scale – furthered by a center-out, outcome-driven architecture – are the key components to achieving this vision; a vision that will open doors for the federal government to achieve more with current and ever-advancing technologies.

Leveraging intelligent automation and case management

To improve population health, the federal government must be able to do more than just aggregate the data – they need the tools and technology to act on the data. That is where case management and intelligent automation come in. Case management is a software-based approach to managing a set of processes that collect, track, and consolidate data to achieve a specific outcome. A case management system goes beyond workflow to automate the response and resolution of the work itself. Automation, achieved through the application of policy (business rules), can make decisions based on the situation and its context. For example, case management would trigger an investigation if the number of positive influenza cases in a region, over a given period, was above a certain threshold.

To draw on an existing government example, the U.S. Social Security Administration (SSA) has a mission-critical need to meet the demands of data sharing and collection to ensure proper payment of benefits, reduction of field office traffic, and the overall operation of sharing timely and accurate data. Outside of aggregating large amounts of data, the SSA was able to build

work processes that use the data to support their mission-critical tasks. In fiscal year 2017, the SSA engaged in more than 3,000 data exchange agreements and processed over two billion transactions. This number has continued to rise about 10% each year. In May of 2020, the SSA went live with its first Pega application to support its vision of a centralized, interactive, and dynamic user experience for requesting, sending, receiving, and administering data exchanges. This application shows how Pega's intelligent automation helped SSA provide a fully automated service to various state, local, and federal agencies – improving both user experience and progress toward standardizing information exchange. This standardization and automation of recurring processes will be critical for the services and outcomes that the federal government can achieve with the NPHDS.

Working from the center out

The NPHDS needs to be designed thoughtfully – given that it will have users from over 2,300 state and local public health agencies working with various government agencies across the country. The design and architecture must be scalable, flexible, efficient (maximizing reuse wherever possible), and fully able to adapt to changing public health needs. Rather than making the mistake of building from the bottom up – focusing on the technology and implementation – the federal government should focus on building from the center out. A Center-out™ approach to structuring technology allows agencies to build their business architecture around the heart of the enterprise: the population and outcomes. Agencies achieve this by:

1. **Managing intelligence centrally:** Ensure AI and business rules operate across all channels.
2. **Focusing on outcomes:** Start with the desired outcome and define the stages, steps, and processes that will help achieve that outcome.
3. **Connecting experiences up to channels:** Ensure channels and business logic are in sync using dynamic APIs. User interfaces (for internal and external users) will automatically update with changes, without requiring any code.
4. **Connecting down to data and keeping logic nimble:** Insulate case and decision logic (users and citizens) from the complexity of back-end systems. Use data virtualization to pull in necessary data without worrying about where the data is coming from.
5. **Managing variations to scale:** Future-proof application architecture by building in layers. Capture the unique dimensions of the organization and/or application while reusing common components.

Starting with the right business architecture methodology will help ensure that the technology is future-proof, scalable, secure, and robust enough to handle both the current and future public health needs of the nation. To achieve this, the government must utilize a unified platform that can offer integration without limits, manage end-to-end orchestration of work, automate work and processes, incorporate AI and decisioning, and securely store and/or process health information. One example of how Pega's capabilities were able to support a federal program of a similar scale is with the 2020 Decennial Census.

Unifying operations for the 2020 Census

The Eagle-Pega team has had vast experience collecting population data – saving that data to the cloud, running analytics on it, and making it available for government and commercial use – for the 2020 Decennial Census.

The U.S. Census Bureau serves as the leading source of quality data about the nation's people and economy. Census data is used to inform a variety of social, economic, and political decisions – from congressional seat distribution to community planning to the distribution of more than $675 billion in federal funds. The Bureau uses a series of data collection methods to gather this data, including the Decennial Census of Population and Housing that happens every 10 years. To modernize and streamline its data collection and processing operations, the U.S. Census Bureau launched the Enterprise Censuses and Surveys Enabling (ECaSE) project to create an integrated, enterprise solution for shared data collection and processing. The aim of the solution was to lower complexity and deliver cost savings for all censuses and surveys – with an initial focus on the 2020 Census.

Pega was part of a community of contractors that played a key role in both data collection and partner program initiatives. The Bureau used the Pega Government Platform for its ECaSE operations, including survey operational control, field operational control, and data collection. Specific data collection activities included:

- Survey response management to process and control survey data collected from the census website, paper forms, field enumeration via mobile devices, and by phone via call centers
- Field enumeration for collection through an Apple iPhone application used to perform surveys with residents that didn't respond online, or via phone or paper questionnaire
- Self-response instrument for response collection, used by Census Questionnaire Assistance (CQA) agents in call centers as they interviewed respondents that contacted a call center
- Self-response case and event consolidation for all responses collected from the census website and paper form submission – along with sufficiency checks to determine cases to be enumerated

Operations were highly successful despite various regional rules, hurricanes, wildfires, and the COVID-19 pandemic restrictions, which resulted in a combination of challenges even the census could not anticipate. Despite these challenges, intelligent automation – powered by the Pega Government Platform – provided the ability to dynamically reallocate work – doubling the productivity rates of resources. The Bureau was able to accurately target low-response rate geographies and retarget to drive self-response rates – by using real-time data and analytics. This ability to react and respond quickly was especially critical in instances where weather events, public health concerns, or other circumstances made getting survey responses challenging.

> "This is my third census, and I've never been through a situation where we had so many reasons to think we might not finish. Automation saved us. If we had done the census the way we did it in 2010, we would have been stopped. This is one of the unsung successes of 2020, I think, that the census actually got done."
>
> – *Michael Thieme, Assistant Director for Decennial Census Programs, Systems and Contracts, U.S. Census Bureau (Statement from FedTech)*

Between January and October of 2020, the Pega/ECaSE systems completed data collection for the 2020 decennial operations, processing hundreds of millions of cases and billions of events across field operations and self-response. This resulted in the following accomplishments:

- The 2020 Census was the first to use a primarily digital process for data collection.
- The entire census was completed with zero system downtime.
- 117 million surveys were collected across all self-response channels (93 million online, 2.8 million through the call center, and 20.9 million by paper).
- 92 million households counted (almost 63% of those surveyed) within the first four months.
- Census was 90% complete by mid-September 2020, a full six weeks before the original October 31 deadline.
- 1.15 million cases were created for the operational control systems with 1.98 billion events.
- 220,000 resources in the field completed more than 42 million surveys, despite months of delay due to the COVID-19 pandemic.

All of this was done while safeguarding respondents' data. The team employed physical security techniques and multiple internal and external threat mitigation strategies to ensure the security of respondents' data.

Though not used for the same purpose, both the census and the NPHDS share critical similarities when it comes to dimensions of scale – scale in terms of performance, security, volume of data, development, and system users. This same unified platform that the Bureau used for these activities is also being used by some of the nation's largest healthcare providers. With an existing FHIR API – which uses the HL7 FHIR standard to retrieve data from electronic medical record systems and a diverse set of attestations including ISO 27001, NIST, SOC2, IRAP, FEDRAMP, HITRUST, and HIPAA – these vendor success stories ensure that the technology is ready and capable of helping the federal government address its largest public health data challenges.

Accommodating unique data governance needs

Each federal agency will have unique needs for data in the NPHDS. For example, some agencies like the CDC will only want de-identified data, whereas others will be permitted to obtain data with personally identifiable information (PII). They will differ in the data subsets they want to use from our data repository and may have unique data linkage needs. The variables they want to use will likely be different. In such cases, we will share only what is requested. Finally, some will want the data in SAS, comma-separated value (CSV), or JavaScript Object Notation (JSON) format – whereas others may want one of the other industry-standard formats. We will provide each agency with the data set it needs by extracting it from the repository as a deliverable version or by programming a custom API for them. Additionally, we will fulfill ad-hoc user requests regarding data sharing. The data must be encrypted before they are shared by secure file transfer protocol (SFTP) instead of legacy file transfer protocol (FTP). We adopt SFTP for its high security with authentication via password, secure shell (SSH) key, or both, and due to the availability of cloud-based SFTP SaaS solutions.

Conclusion

As the government recovers from the COVID-19 pandemic and prepares to fill the data gaps that were exposed, using proven technologies that successfully handled the scale and complexity of the 2020 Decennial Census is vital. The overall data governance and systemic management of the National Public Health Data System relies on it. Our centralized, cloud-based NPHDS will act as the central data source enabling federal agencies and the healthcare community to prepare for, track, and respond to the nation's population health needs. This will help the government promote health equity and achieve many other goals, including identifying and proactively managing alterations to public health caused directly or indirectly by pandemics. The federal government can save money, time, and lives with the Eagle-Pega NPHDS solution.

Pega delivers innovative software that crushes business complexity. From maximizing customer lifetime value to streamlining service to boosting efficiency, we help the world's leading brands solve problems fast and transform for tomorrow. Pega clients make better decisions and get work done with real-time AI and intelligent automation. And, since 1983, we've built our scalable architecture and low-code platform to stay ahead of rapid change. Our solutions save people time, so our clients' employees and customers can get back to what matters most.

For more information, please visit us at **pega.com/government** or **pega.com/HHS**

Eagle Technologies, Inc., (Eagle), founded in 2004 and headquartered in Arlington, Virginia, provides consulting services for government agencies with significant expertise supporting the Department of Health and Human Services (HHS). Eagle is a Technical and Systems Integrator delivering effective solutions to complex project requirements for government, large corporations, and business customers nationwide. Since its founding, Eagle has provided enterprise infrastructure, cloud, application development, database, operations, data analytics, and reporting solutions for federal, state and local health agencies nationwide. Eagle leverages its depth of experience managing and delivering end-to-end solutions for its clients that rely on advanced cybersecurity and privacy systems, enterprise architecture, cloud-based services, mobile solutions, and business intelligence services. Eagle effectively utilizes the GSA Multiple Award Schedule (MAS) small businesses, 8(a) STARS II, CIO SP3, CIO CS, Alliant 2, VETS 2, OASIS, HHS PSC BPA, and NASA SEWP contract vehicles supporting and fostering our goal of continued growth in the IT industry.

For additional information regarding Eagle services, please visit **https://eagletechva.com/services.html**