GovCIO MEDIA & RESEARCH

DeepDives

Infrastructure Modernization

INSIDE:

SPONSORED BY

denodo

GovCIO MEDIA & RESEARCH

From the editor's desk

Amy Kluber, Editor-in-Chief

Data Management in the Age of Cloud

In the age of cloud modernization that unlocks better data-computing practices, agencies are faced with different challenges around managing that data. These challenges can include tactics to connect cloud, on-prem and other systems in a hybrid environment, or overseeing change management efforts to recruit and train the workforce properly.

With the current widespread move to zero-trust architecture, getting a handle on these challenges are necessary for agencies to maintain efficient data flow and operations.

Agencies with different infrastructure approaches and missions have varying degrees for how they are treating their cloud environments and how they are encouraging industry support in these areas.

Table of Contents

Sarah Sybert, Staff Writer

Melissa Harris, Senior Staff Writer

State Department, VA are Accelerating the Move to Hybrid Cloud

Cloud leaders address how they're transitioning to hybrid cloud environments.

BY SARAH SYBERT

The departments of State and Veterans Affairs are transforming data management strategies as they adopt hybrid cloud models.

"Hybrid to me means using more than one type of solution. So, whether it be two different cloud service providers (CSPs), or CSP and on-prem, or CSP and a SaaS, or some combination of several," said Dave Catanoso, VA's acting director of application hosting, cloud and edge solutions, during an ATARC event. "Multicloud is leveraging more than one CSP to solve problems, whether it's CSPs in our enterprise cloud or in combination with our SaaS providers."

Agencies are aligning operating structures to better support modernization journeys. Catanoso said VA is developing a new construct to expand the Enterprise Cloud Solutions Office. The unit will combine VA's cloud, on-prem and platform efforts to improve the organizational structure, provide better customer service and enhance service delivery to veterans.

Brian Merrick, director of the State Department's Cloud Program Management Office, said that in a transition to cloud, leaders should focus on three key aspects: people, policy and technology.

Organizational change is key to successful cloud adoption. Since VA began its migration in 2018, the agency has relied heavily on change management and collaboration to ensure its workforce has access to the right training, the agency takes a unified approach to modernization and promote engagement.

"It's not a technology problem, it's a people issue, and it's really a change management effort," Merrick said. "We're figuring out how we provide scalable, shared infrastructure services so that we can abstract away some of the infrastructure pain from the business areas and [the workforce] can get to work faster and leverage the benefits of cloud while still maintaining control over their environments and their delivery to their customers."

As the data ecosystem only gets more complex, the State Department is working to recruit talent that is capable of advanced data analytics, as well as

driving policies that accelerate data sharing to effectively manage datasets. On the technical side, State is leveraging innovative tools like application programming interfaces (APIs) to make data more accessible.

"It starts first with having clean application for data capture with … data tables that you know need to be consistent with automated hard stops to enforce that data entry so that you have clean datasets to begin with," Merrick said. "Then, having a solid API management approach to be able to share that data to a data lake, and then having the right tools available for the actual analytics and the data visualization for decision support has been really critical, and we're continuing to see that evolve across our space."

Currently, VA is at the beginning of its journey toward an enterprise, common data plane, with the goal of enabling the agency to share data across applications more cohesively, Catanoso said. Most of VA's solutions are more mission driven, supporting different priorities like electronic health record modernization, financial management, benefits and memorials. Because of this structure, a lot of VA's data resides in specific mission sets and have different compliance, portability, shareability and retention requirements.

"We're juggling that with how we leverage the cloud technologies that were brought to bear with our on-prem capability to come up with these common tools," Catanoso said. "Now we're starting to look at data-loss prevention across the enterprise … we're still looking at how we're going to solve that across multiple business sets."

As VA and the State Department accelerate their transition to cloud, the agencies will gain greater flexibility, agility, speed of delivery, scalability and access to modern tools.

"What we ended up seeing, the best of all worlds from a hybrid standpoint, is leveraging the data on-prem and holding that core dataset and then using a PaaS and SaaS solution for the data capture and the workflow piece of it. Then,

“It’s not a technology problem, it’s a people issue, and it’s really a change management effort.”

—Brian Merrick, Director, Cloud Program Management Office, State Department

doing the analytics with an IaaS or PaaS type environment, where you can really leverage the speed and scalability and elasticity of cloud to meet those different needs,” Merrick said.

Security and policy also must evolve at the speed of emerging technology and of the mission. Both State and VA are amidst a move to a zero trust architecture to better protect data — aligning with President Biden’s cybersecurity executive order. VA is taking a multi-factor approach, leveraging encryption and implementing TIC 3.0 to “defend data,” Catanoso said.

As for State, “We have a long way to go, but I think that’s going to be the future — really looking at how do we protect [data] from a holistic standpoint, really focusing on where that data is living and providing that defensive depth to the data itself,” Merrick said.

GovCIO MEDIA & RESEARCH

DeepDives

Next FITARA Scorecard Categories

1. CIO direct reporting to agency authority
2. Cybersecurity
3. IT modernization working capital funds
4. Portfolio review savings
5. Risk management and transparency
6. Telecommunications services transition

GovCIO MEDIA & RESEARCH

PARTNER INTERVIEW

denodo

Data Virtualization Platforms Ensure Security, Efficiency

Denodo's data fabric solution simplifies access to data within hybrid/multi-cloud environments.

Paul Moxon, SVP Data Architecture and Chief Evangelist, Denodo

How can a data virtualization platform help agencies best compute upon their data in a multi-cloud environment?

Moxon Government data infrastructure modernization projects are motivating cloud migration, but as different cloud service providers offer different benefits, it would be surprising to see any organization limit its options by committing to a single cloud vendor and in the process face vendor lock-in.

A logical data fabric leverages fundamental data virtualization techniques in which separate instances of the platform or product are placed within each cloud domain (such as AWS, Azure or GCP) and can access and aggregate the data within that cloud provider. This enables the logical data fabric to coordinate access and aggregate the data from across the different clouds to provide a holistic view of data across the hybrid enterprise.

A logical data fabric ultimately simplifies data access across multi-cloud environments. It streamlines data management and analytics efforts, supports access to a wide variety of data types and sources, speeds data delivery and optimizes performance at a fraction of the cost of traditional integration approaches.

(ctd.)

“Data virtualization is emerging as the leading approach to agile data integration and delivery, and it can provide massive productivity returns by unlocking the full value of your organization’s data, no matter where your data or data consumers reside.”

—Paul Moxon, SVP Data Architecture and Chief Evangelist, Denodo

How can a logical data fabric powered by data virtualization prevent and break down silos?

Moxon As our government ecosystem continues to transform from primarily on-premises-only IT architectures to hybrid and multi-cloud architectures, it will be more necessary than ever before to invest in a logical data fabric platform that can tame the complexity and provide uniform access to data. Having flexible, uniform, real-time access to data assets, wherever they are, directly from the sources, can accelerate “time to insight” and the development of new data services. Data virtualization is emerging as the leading approach to agile data integration and delivery, and it can provide massive productivity returns by unlocking the full value of your organization’s data, no matter where your data or data consumers reside.

What is key about Denodo’s platform that federal leaders would find unique in the market?

Moxon Denodo’s data fabric platform is powered by data virtualization. Data virtualization helps agencies in various ways:

- It avoids vendor lock-in when it comes to data access. Through hundreds of out-of-the-box adapters, the Denodo platform offers a single pane of glass from which to access any data set across all enterprise assets no matter what type of data source or application, even across different cloud provider environments.
- It reduces data security risk since data virtualization insulates critical source systems from users and applications, preventing them from unintentionally changing data.
- It provides faster data access at a fraction of the cost of traditional integration approaches, even across the most complex IT environments and cloud infrastructures. Data virtualization avoids the need for data replication, which takes time and costs money. Unlike more traditional

integration approaches like “extract, transform, load” (ETL), data virtualization doesn’t need to “collect” data into a repository to work with data. With a “zero-replication” approach, data consumers receive up-to-the-minute information without investing in additional storage.

How does this tool ensure data security and system protections from external or internal threats?

Moxon In addition to the reduced data security risk since data virtualization insulates critical source systems from users and applications, preventing them from unintentionally changing data, the platform also delivers seamless security and governance. The platform delivers advanced data masking and attribute-based access control (ABAC), and makes it available for all data assets in the organization via a single point of control and administration.

Photo Credit: greenbutterfly

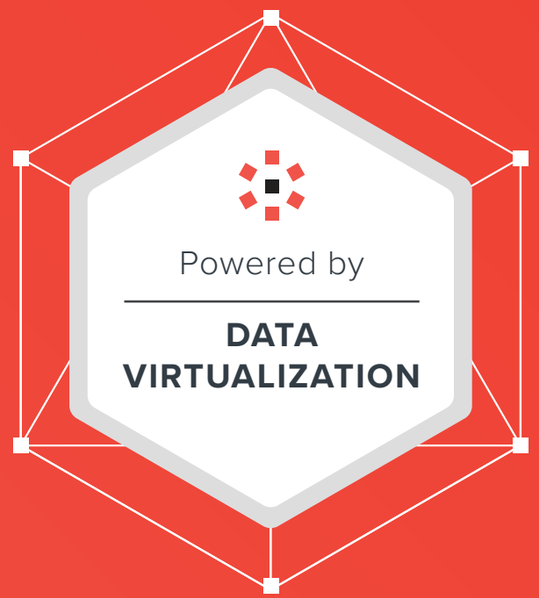
Accelerate the Path to Data-Driven Government with a Logical Architecture

denodo

Hybrid/Multi-Cloud Data Integration

Self-Service BI

Data Mesh



Enterprise Data Services

Data Science

Data Fabric

Logically integrate, manage, and securely share your distributed data

<6 MONTHS PAYBACK

65% FASTER THAN ETL

67% LESS DATA PREPARATION

Source: Forrester Total Economic Impact™ of Data Virtualization, 2021

VISIT DENODO.COM

NIH Gears Up for Data Sharing Policy, Modernized Biomedical Data Ecosystem

Data leader Susan Gregurick discusses how the agency is improving data sharing and usability for biomedical research.

BY SARAH SYBERT

The National Institutes of Health (NIH) is creating infrastructure capabilities and new programs to promote data discovery, use and sharing in alignment with its upcoming data management policy.

"Our goals are to catalyze data science capabilities across all the 27 institutes and centers at NIH," Susan Gregurick, associate director for data science and director of the Office of Data Science Strategy, said at GovCIO Media & Research's Infrastructure: Health IT virtual event. "We do that by working with a very large number of colleagues across NIH — almost over 200 NIH staff work with us on various teams — to help us implement different data science strategy capabilities."

Data in the biomedical space is historically very heterogeneous, which has presented challenges in data integration and big data analytics. NIH has also faced obstacles with data policies and data governance that are structured for a "one dataset at a time approach."

Serving as the foundation of NIH's data efforts is the agency's upcoming policy on data management, which is set to go into effect in January 2023. The policy will strive to make data as widely and freely available as possible while safeguarding privacy and protecting confidential and proprietary data.

In addition, the agency is preparing data repositories and developing guidance to help researchers create data management plans that ultimately drive the concept of "data curation at scale."

"I hope this just gives you a little bit of idea of some of the things that are happening in the biomedical data landscape, computing architecture, to ways in which we're thinking of data as an infrastructure, to finding data and resources across different programs, and then, of course, helping researchers create data management and sharing plans and really enable greater data sharing across all our programs," Gregurick said.

Gregurick noted the agency's STRIDES Initiative that provides cloud computing resources to NIH investigators and also training and workforce

Photo Credit: tiero/iStock

Susan Gregurick

Associate Director for Data Science, NIH

development. Through the program, the agency has experimented with several different infrastructure approaches, including hub-and-spoke, distributed, centralized and federated, to build out its health care delivery and research ecosystem.

"A few promising activities include the National COVID Cohort Collaborative (N3C)," Gregurick said. "It's really an amazing activity in terms of what they're doing with the data. So as a highlight, they developed a way to extract and harmonize clinically related data at a scale that's really quite unprecedented for NIH."

The collaborative worked across more than 72 sites to extract participant data derived from electronic health records (EHR), patient chart data, medical histories, diagnostics, demographics, immunization records and imaging data pathology data to improve access and use of COVID-19 clinical data, which could then inform pandemic-related research questions.

"All this data is harmonized … so that researchers have access to a very large and extensive amount of data sets that can be used at scale," Gregurick said.

Under NIH's Rapid Acceleration of Diagnostics (RADx) Program, a hub-and-spoke approach enabled the agency to speed up development, validation and commercialization of point-of-care home-based testing. Gregurick said the program is spearheading a community-driven, integrated approach for creating data models and common data elements that can be harmonized.

As the agency continues to explore different federated approaches, like its Cloud Platform Interoperability Effort, the goal is to provide researchers with capabilities to find, access and use distributed datasets across NIH's supportive platforms and institutes.

"This is a fairly complicated and complex program, but the idea is that the flexibility of creating an integrated and interoperable and somewhat federated approach allows new programs to be added and new capabilities to be

“The overall goal is to provide a modernizing integrated biomedical data ecosystem. … That sounds easy, but it’s actually quite challenging because of the diversity of science across NIH, and the diversity of needs and capabilities. An overall one-size-fits-all strategy is very challenging.”

Susan Gregurick, Associate Director for Data Science, NIH

developed on demand, so we can start out small and build the program in a really flexible manner,” Gregurick said.

NIH aims to grow its STRIDES Initiative to explore the use of cloud environments through commercial cloud providers. These partnerships will enable access to improved datasets and advanced computational infrastructure, tools and services.

“The overall goal is to provide a modernizing integrated biomedical data ecosystem,” Gregurick said. “That sounds easy, but it’s actually quite challenging because of the diversity of science across NIH, and the diversity of needs and capabilities, an overall one-size-fits-all strategy is very challenging.”

With the promise of emerging technologies like hypercomputing, Gregurick said the agency will eye “data as a product” and “data as an infrastructure” to enable NIH to provide diagnostic functionality and tools and maintain interoperability.

“We have a significant amount of data and computational capabilities in our STRIDES partnerships. We want to leverage these, but we also want to include ways in which we can provide equality for all our research communities,” Gregurick said.

Are Data Literacy Programs the Key to Data-Driven Decision-Making?

The Army is embracing data literacy training to empower personnel to leverage data.

BY MELISSA HARRIS

Federal IT leaders have been aiming to make data a strategic asset across their agencies, and while acquiring tools and infrastructure is a boon in harnessing data, many are focusing on developing data literacy and a data-driven culture.

People, platforms, processes and culture are four critical areas for developing an organization that leverages data effectively, said Army People Analytics Deputy Director Lt. Col. Kristin Saling at GovCIO Media & Research's Cloud Modernization event.

Denodo SVP Data Architecture and Chief Evangelist Paul Moxon and Army People Analytics Deputy Director Lt. Col. Kristin Saling.

Amid these needs, the Army launched a data literacy program to help enable its workforce to work with data, as well as help foster a culture that embraces data in decision-making and business activities.

"We needed a base level of understanding for all soldiers, all civilians, everybody working in this space to be able to speak about data because we don't have a career field that isn't using data in some aspect," Saling said.

"We piloted a number of different programs. I think our most successful one right now is Army Data 101."

The course is four days long, two and a half hours a day, and it aims to help personnel think more critically about data. Saling added that her team is looking to develop more hands-on courses for employees more interested in workshopping data-driven solutions for challenges across their organizations.

Denodo SVP of Data Architecture and Chief Evangelist Paul Moxon also stressed the importance of data literacy training, adding that all agencies should be implementing similar programs.

"If you don't understand the data, you'll be led by the nose by somebody else who does understand data," Moxon said. "You can need the people who can deliver the insights, but you have to be able to understand the insights. … It's necessary because we are becoming a digital-driven community."

Moxon added that it's also important for data literacy training across the

Photo Credit: Rod Lamkey/GovCIO Media & Research

entire organization, not just the senior level, not just to leverage data across different business areas, but also to drive a more data-forward culture.

"People do tend to do things because they know and when you have people who are very experienced, this is the way they've always done it," Moxon said. "This is the way we're going to do it, and it's getting them to think, 'Okay, can I do something differently? Can I look at the data and actually come to a different conclusion, an alternative way of doing something?' … You have to change that, and that's part of what data literacy is."

To build a culture that embraces use of data in business practices, Saling has also been highlighting the value of using data-backed solutions and decision-making. She touched upon the benefit of teaching how data-driven tools like automation can make small but meaningful impacts for different business activities.

"We've done a lot of this work at Human Resources Command with our career managers who are, for the most part, non-technical folks but have to manage a lot of data on their personnel," Saling said. "Getting some tricks to do that a whole lot more easily than when they were doing it, when they were manually copying things over … that word starts spreading and start thinking about, talking about all the other cool things that we can build."

Amid the data literacy and culture efforts, Saling and Moxon both encouraged a culture of iterative innovation and experimentation so that users and analysts can find a starting point in using data.

"You just need to unleash that innovation within the agency itself rather than bringing in consultants and things like that," Moxon said. "The agencies themselves are the most innovative people that you have."

“We needed a base level of understanding for all soldiers, all civilians, everybody working in this space to be able to speak about data because we don’t have a career field that isn’t using data in some aspect.”

Lt. Col. Kristin Saling
Deputy Director, Army People Analytics