

EXPERT EDITION

Closing the Gap Between Mission and Data

Insights on Data from:

- General Services Administration
- NASA
- Air Force
- Federal Energy Regulatory Commission

BROUGHT TO YOU BY:



Setting Data in Motion for Government Missions

Data in motion allows an agency to connect its systems and applications to supply ready access in real time to the data it needs to make better decisions and improve efficiencies.

confluent.io

TABLE OF CONTENTS

Data-in-motion optimizes management of GSA's vehicle fleet...**2**

Air Mobility Command learns to provide support 'at the speed of data'...**4**

NASA adopting Apache Kafka to enable real-time data from Mars...**6**

FERC stands up cloud data science, analytics platform to enable collaboration...**8**

Agencies moving to a data-in-motion paradigm creates a 'central nervous system' for their missions...**10**



Over the last 15 years, agencies have come to realize the value data can bring to the mission areas. But it's only been in the last few years that they've started to make the connection between information and decision-makers.

The Foundations of Evidence Based Policymaking Act of 2018 pushed agencies to create the infrastructure, particularly by naming a chief data officer.

Now, the CDOs are leading efforts to construct data super highways that go from sensor to decision maker.

Lt. Gen. Brian Robinson, deputy commander of the Air Mobility Command, said the K-46 taker will use the Advanced Battle Management System (ABMS) to become flying "hotspots" so pilots can offload new data to F-22 and F-35 fighters at the same time they're receiving fuel.

Other agencies like NASA, the General Services Administration and the Federal Energy Regulatory Commission (FERC) are attracted to similar potential benefits of creating platforms to drive data to edge and back to the decision makers.

FERC CDO Kirsten Dalboe said her office is building a cloud data science and analytics platform, and taking steps to modernize its data footprint.

GSA has been able to get its data not only on the same page but a lot more quickly, accessing data while it's still flowing through the system, rather than waiting to send individual queries to separate databases.

Will LaForest, public sector chief technology officer at Confluent, said these and other agencies are starting to think about how to create a central connective tissue that conducts their data across their organizations, and to begin handling their data-in-motion as it is created and flows to those who need it.

This e-book highlights some examples of how agencies are benefitting from that concept of data-in-motion to improve mission outcomes.

Jason Miller
Executive Editor
Federal News Network

Data-in-motion optimizes management of GSA's vehicle fleet

BY KEVIN MCCANEY

The General Services Administration is reaping the benefits of data-in-motion in managing the expansive fleet of vehicles the agency either owns or leases, using the vast volumes of data it gathers to improve the efficiency of the fleet.



requirements under the web-based Federal Automotive Statistical Tool (FAST) and GSA's Asset Level Data (ALD) systems for reporting to FAST.

The key to making this work is data-in-motion, which allows the processing and use of data as it's being created or when it's in transit, in order to better manage vehicle use, in line with the Federal Data Strategy and Executive Order 13834 on efficient federal operations.

"While the data needs are diverse across the organization and federal government, Fleet requires reliable and robust data systems and tools to manage vehicles across their entire lifecycle," a GSA spokesperson said.

That data includes information from a range of sources, as well as from locations around the world, including original equipment manufacturers, Fleet's payment card provider and GSA's telematics program.

GSA's fleet is among the largest in the world, supporting more than 210,000 vehicles (more than FedEx has) for both civilian and military customers in domestic and overseas locations. Managing those vehicles efficiently involves a lot of data from multiple sources that needs to be collected, sorted, analyzed and put to use.

The data can, for instance, help GSA Fleet Management lower expenses through better management of contracts and identify opportunities for sharing vehicles among groups rather than over-allocating cars that may spend most of their time sitting unused. And it helps agencies meet reporting

Data in motion allows an agency to connect its systems and applications to supply ready access in real time to the data it needs to make better decisions and improve efficiencies. In this case, it has resulted in the GSA getting better mileage out of its FedFMS.

Getting data in gear

The traditional approach to data collection is to dump it into a storage system that might be siloed from other systems and then send queries to that data when necessary. It wasn't a particularly fast system for making use of data and had a few other limitations, such as having to clean up the data to accommodate complex business rules, according to GSA. It also had difficulty interoperating with other GSA systems, such as the fleet drive-thru safety program for leased vehicles. The different data sources and requirements could result in confusion and errors when reporting from the two systems was combined.

With a data-in-motion platform from Confluent, which provides data-in-motion services to a variety of federal agencies, GSA has been able to get its data not only on the same page but a lot more quickly, accessing data while it's still flowing through the system, rather than waiting to send individual queries to separate databases. Confluent is a data-in-motion platform with Apache Kafka at its core built for event stream processing. It's designed to collect and process events as they happen, allowing customers to make immediate use of the information.

"These interfaces enable GSA to provide safe and reliable vehicles to our customers," the GSA spokesperson said.

The agency is better able to track compliance with federal regulations, replace vehicles according to schedules that maximize value to the taxpayer, perform electronic recall management and analyze data to decrease maintenance and repair costs while reducing vehicle down time, the spokesperson said.

The agency has created the Federal Fleet Inventory Tool, which pulls data from FAST and provides robust data visualization into the inventory and operating costs of federal vehicles. Fleet managers and users can access the data in an effort to make better management decisions.

GSA Fleet also incorporates telematics, which collects and analyzes data such as maintenance, driver behavior, reporting, and fuel use and idling, with an eye toward lowering costs. The agency said telematics has helped it reduce risky driving behaviors, finds opportunities for car-sharing, saves time conducting market research and avoids unnecessary contracting expenses. It's also helped it meet ALD and FAST reporting requirements.

By increasing the speed and accessibility of data-in-motion, agencies like GSA and those working with the Federal Fleet Management System (FedFMS) can respond more quickly to the available information and make better, faster decisions.

"We provide customer agencies with the data they require to effectively and efficiently manage their fleets," the agency spokesperson said.

Data-in-motion allows an agency to connect its systems and applications to supply ready access in real time to the data it needs to make better decisions and improve efficiencies. In this case, it has resulted in the GSA getting better mileage out of its FedFMS. 🚗

"While the data needs are diverse across the organization and federal government, Fleet requires reliable and robust data systems and tools to manage vehicles across their entire lifecycle."

— GSA SPOKESPERSON

Air Mobility Command learns to provide support 'at the speed of data'

BY DAVID THORNTON

The Air Force recently announced that it will modify its KC-46 aerial refueling tankers to host its Advanced Battle Management System (ABMS) to become flying "hotspots" that can offload new data to F-22 and F-35 fighters at the same time they're receiving fuel.



situational awareness with mobility aircraft is gleaned through what we call voice wrap or voice reports. So it's human beings in the command and control nodes, and aircrew amongst aircrew relaying, via push to talk radio, the situation. And then you

have to write it down on a piece of paper or a chart, that you have to build a visual depiction of a situational awareness they're dealing with so we learned that that's actually value added, it actually reduces the workload for all involved."

Lt. Gen. Brian Robinson, deputy commander of the Air Mobility Command, said pilots during the recent exercise Mobility Guardian called the increased situational awareness from ABMS "game changing."

ABMS is the Air Force's main contribution to DoD's Joint All-Domain Command and Control (JADC2) vision of interconnected military systems.

"We did learn for one, what it takes to field the right equipment to be part of that architecture, so that it's useful," Robinson said during a recent Defense News webcast. "We also had commentary from some of the receiver aircraft, A-10s in this case, and having the tanker on Link 16 was 'game changing' from the receiver perspective, because today, a lot of that

He said it also gives mission commanders on the ground a near-real time situational awareness of what's going on around them. And that situational awareness extends out as far as a feed can be broadcast. Even warfighters who can't maintain direct connectivity to the command and control node still get enough information to make better decisions to keep the mission on track.

Robinson said placing the ABMS system on the KC-46 tankers is a marriage of functionality and convenience. The tanker's presence in operations like this is required, he said, to allow those fighters to reach their mission. So it's

conveniently already on site during the mission. But it also has the perfect size, weight and power to carry the hardware required to link the ABMS system to its space-based components while providing an alternate path to connectivity.

"That tanker can also, because of its size, carry what we're talking about: Edge computing. So it has enough processing power on board, if it gets disconnected, the processing power onboard can store that data."

— LT. GEN. BRIAN ROBINSON, DEPUTY COMMANDER, AIR MOBILITY COMMAND

"When you're in a position like we are where we support 10 other combatant commands and their components, we have to have a robust and persistent understanding of what's most important to whom, when and why, and how that links back to the overall strategy or campaign, if you will."

— LT. GEN. BRIAN ROBINSON, DEPUTY COMMANDER, AIR MOBILITY COMMAND

"That tanker can also, because of its size, carry what we're talking about: Edge computing. So it has enough processing power on board, if it gets disconnected, the processing power onboard can store that data," Robinson said. "So as a fighter pulls up for receiving fuel from the tanker for its mission, it can pass information that it has, if the link is broken, to the tanker that can then store it or transmit it through the capabilities that it has on ABS construct. As well as anything received to the tanker, in between refuelings, it can also pass information back to the fighters as well, that they can take onboard and with their systems figure out the most optimal way to prosecute their mission."

He said ideally, he wants every one of Air Mobility Command's aircraft equipped to handle this system. Due to fiscal realities, that probably isn't possible, but Robinson said he intends to buy as many as he can. His target for installing this equipment is the fourth quarter of fiscal 2022.

Air Mobility Command has been angling to become the earliest adopters of this system since late last year.

"Air Mobility Command ... is one of most forward-leaning commands we have in JADC2 thinking, and they are ready to go put [ABMS capabilities] on mobility platforms so they can act as data relays," Will Roper, the then-Air Force acquisition chief said in November. "We've got tankers that top you up with gas, the vision of topping you up with data makes a lot of sense: You're going to be there anyway to get fuel. And then that tanker standing off also can act as a battlefield relay and a network node. So they've got the right thinking."

Robinson said this is a natural extension of the supportive mission of Air Mobility Command.

"When you're in a position like we are where we support 10 other combatant commands and their components, we have to have a robust and persistent understanding of what's most important to whom, when and why, and how that links back to the overall strategy or campaign, if you will," he said. "And that's what we do pretty well, right now, but we're going to have to get better at it. We're going to have to be able to do that at the speed of data. Because right now, we don't necessarily do that. We use data, but not in a near real time fashion."

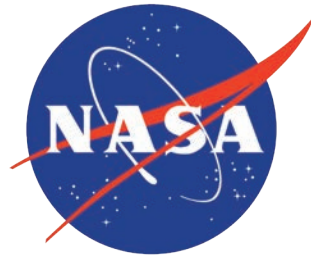
The KC-46 project is known as "Capability Release 1," and the initial work is part of \$170 million the Air Force has allocated for ABMS this year. 🚀

NASA adopting Apache Kafka to enable real-time data from Mars

BY KEVIN MCCANEY

Federal government agencies are seeing an increased need to make quick use of real-time data, in areas ranging from healthcare and processing benefits requests to managing fleets of vehicles and supporting military missions. For NASA, real-time data also extends into the far frontiers via its Deep Space Network (DSN). It's another example of how agencies are using Apache Kafka to set their data in motion.

The DSN is rooted in NASA's array of three massive radio antennas spotted equidistantly around the globe. One of these antennas is located in Gladstone, California. The other two are located in Madrid, Spain and Canberra, Australia. But its reach extends far into space in support of the space agency's interplanetary missions (as well as a few missions in orbit around Earth). Operated by the Jet Propulsion Laboratory



(JPL) in Pasadena, California, DSN collects the feeds from missions such as the ones working on the exploration of Mars.

Although some commands and data can be transmitted directly between Earth and NASA's rovers, the mass of information being generated first needs to go through the Mars Relay Network. That network is comprised of a constellation of orbiters collecting scientific data from the Perseverance rover, which touched down in February, as well as the Curiosity, which has been exploring the red planet since 2012, and the InSight lander, which has been operating there since 2018.

That network is comprised of a constellation of orbiters collecting scientific data from the Perseverance rover, which touched down in February, as well as the Curiosity, which has been exploring the red planet since 2012, and the InSight lander, which has been operating there since 2018.

"We're going to be getting higher data rates from spacecraft, more information into the network, more information that needs to be monitored, and we're going to need a way to deal with this information, and parcel it and make sense of it very quickly."

— RISHI VERMA, DATA ARCHITECT, NASA-JPL

For most of the agency's space exploration history, NASA's workflows from space were slow, even accounting for the transmission times required across tens of millions of miles. The building and launching process for satellites and probes happens over a long lifecycle, so missions often went into space with some technology that was already out of date. The DSN has always been reliable, but it couldn't accommodate the massive growth of data or the real-time use of it.

That's changing now because of JPL's use of Apache Kafka, which enables the sharing and use of data as it's created.

As NASA's space mission continues to expand, the amount of data it collects grows exponentially.

"We're going to be getting higher data rates from spacecraft, more information into the network, more information that needs to be monitored, and we're going to need a way to deal with this information, and parcel it and make sense of it very quickly," Rishi Verma, a NASA-JPL data architect, said at a recent Kafka Summit in San Francisco.

In addition to notable missions such as the exploration of Mars, the moon and the outer solar system, NASA is also collecting information from a growing number of other, smaller efforts, including CubeSats. JPL has been aiming to put all that data into a single picture. 🚀

FERC stands up cloud data science, analytics platform to enable collaboration

BY JORY HECKMAN



Chief data officers are a relatively recent addition at most agencies, but they all face an enormous task – taking inventory of all of their agency’s data and figuring out how to get the most use out of it.

Large agencies have a harder time moving ahead on that inventory, but smaller agencies like the Federal Energy Regulatory Commission are getting started by focusing on building out the data community, and making sure it has data stewards up maintain data quality.

That groundwork, however, is allowing the agency to stand up a modernized data infrastructure. FERC CDO Kirsten Dalboe said in an interview that her office is building a cloud data science and analytics platform, and taking steps to modernize its data footprint.

“We want to make sure that we’re offering capabilities across that full data value chain where we have storage tools, we have modeling tools, we have analytics tools, we have collaboration tools, or we’re building out that whole roadmap of capabilities into our cloud based analytics platform,” Dalboe said.

In order to modernize the data footprint and get that data platform running, FERC has spent the past two years on building out the team of officials focused on maintaining data quality.

“Our big focus really has been on building our data governance community. I often talk about building the muscle memory of data governance because those conversations are often new to people. The first year of our data governance body was just building the community. Figuring out what we needed to talk about, how we needed to talk about it,” Dalboe said. “Some of that included the foundational things that needed to occur before we can really start doing the inventory. While none of those things actually had us building [the] inventory, they were incredibly important, because without those conversations starting early on, we would have come out with an inventory that was potentially not very meaningful.”

To facilitate FERC’s ongoing data inventory work, Dalboe said she works closely with agency chief information officer Mittal Desai, who also sits on the agency’s data governance council.

“We do have to rely on each other. Unless a CDO is largely more in kind of the policy and compliance space, you’re going to have to have a very strong partnership with the CIO ... because at the end of the day, data does live in and sit on technology, so making sure there’s that that’s being managed in a way that facilitates the data strategy is important, she said.

Beyond building out a data community at FERC, Dalboe said the agency’s new data platform will give rise to greater sharing and collaboration among employees.

“The first step is making sure we have a strong kind of enterprise data thinking strategy into our analytics environment, so that we have that live access of data available. This also improves our sharing ability, in terms of analytics being able to build out their data

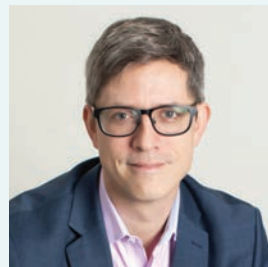
products through the analytics platform, and easily share that with their peers,” Dalboe said. “Pre-modern data platform, people were building off of static data sets on their desktops or their laptops ... so if the person they wanted to share their data product with didn’t have the same tools on their desktop, then they’d be kind of dead in the water in the ability to share. So some of the first steps in having that modern data platform are we want to make sure that we’re bringing our people, our data and our capabilities as close together as possible.”

“The first step is making sure we have a strong kind of enterprise data thinking strategy into our analytics environment, so that we have that live access of data available. This also improves our sharing ability, in terms of analytics being able to build out their data products through the analytics platform, and easily share that with their peers.”

— KIRSTEN DALBOE, CHIEF DATA OFFICER, FEDERAL ENERGY REGULATORY

Agencies moving to a data-in-motion paradigm creates a 'central nervous system' for their missions

THIS CONTENT HAS BEEN PROVIDED BY CONFLUENT



Will LaForest,
public sector chief
technology officer,
Confluent

The federal government is coming around to the importance of data, and is in the process of making a significant amount of policy around how to fully leverage it. Not least among this new policy is the Defense Department's five new "data decrees," set out in a memo signed by Deputy Defense Secretary Kathleen Hicks in

May. The memo, among other things, creates a data council for all DoD components to coordinate their activities, and to increase data sharing among the components.

Hicks wrote that the changes are "critical to improving performance and creating decision advantage at all echelons from the battlespace to the board room, ensuring U.S. competitive advantage. To accelerate the department's efforts, leaders must ensure all DoD data is visible, accessible, understandable, linked, trustworthy, interoperable, and secure."

There is an acute need to avoid the same mistakes of past data modernizations efforts. Attempting to create yet another centralized data lake with all an agency's data would just be more of the same and adding to data sprawl as existing databases aren't likely to be going

anywhere. More importantly, data at rest in a datastore isn't useful when it isn't being queried, and isn't driving decision making. Instead, DoD and other federal agencies need to start thinking about creating a central connective tissue that conducts their data across their organizations, and begin handling their data-in-motion as it is created and flows to those who need it.

"Data-in-motion is really about inverting that dynamic: rather than storing the data away in these silos where it's static and asking retroactive questions, what you want to do is publish the data as a stream and constantly deliver it to the questions, or analysis," said Will LaForest, public sector chief technology officer at Confluent. "Suppose I care about supply chain optimization, and I'm in the military. I want to constantly optimize whenever inventory arrives or leaves, when forces deploy, when production capability changes, updates are made to equipment capabilities, and hundreds of other events. All missions are really made up of thousands, millions or even billions, when you are leveraging sensors,

"Data-in-motion is really about inverting that dynamic: rather than storing the data away in these silos where it's static and asking retroactive questions, what you want to do is publish the data as a stream and constantly deliver it to the questions, or analysis."

— WILL LAFOREST, PUBLIC SECTOR CHIEF TECHNOLOGY OFFICER, CONFLUENT

of these small events, things that are changing constantly throughout the day. So data-in-motion is about handling it as it happens, rather than just sending it to a database, and then after the fact, asking questions."

It's a matter of how data gets from one place to another within, and even between, organizations. Rather than storing it in one place and then having to go and find it to make use of it, Confluent creates a kind of central nervous system to automatically conduct data to where it's needed. After all, if you touch a hot stove, the pain data isn't stored in your hand until your brain goes looking for it to use it. That information is conducted to the brain automatically in order to drive action – namely, moving your hand. In this instance, why should artificial intelligences, to name just one use case, function any differently from organic ones?

Another use case where data needs to be handled in motion in order to trigger actions as quickly as possible is edge computing. Many federal agencies, especially ones with highly distributed missions that don't always have reliable or constant connectivity, simply can't wait on data collected in the field to travel back to a centralized storage and compute, be processed, and then have the results transferred back out to them. For warfighters on the battlefield, or Customs and Border Protection agents screening vehicles at a port of entry, the opportunity to use that data is already passed by the time it's processed. And that's if they have connectivity. Handling the data-in-motion can deliver results and trigger actions at the speed of conflict.

And that's where the central nervous system comes into play. It determines what data is anomalous, useful and relevant in the moment, conducts that data from the sensors to the edge compute for rapid analysis and action, and sends the rest back to central compute to be stored and analyzed later.

"The idea is that one taps into some source of data and as changes occur, they are published as a sequence of events, and then any number of downstream consumers can receive it anywhere they happen to be. And this is really key, you are decoupling the producers of the data from the consumers. It's what makes it operationally scale."

— WILL LAFOREST, PUBLIC SECTOR CHIEF TECHNOLOGY OFFICER, CONFLUENT

"The idea is that one taps into some source of data and as changes occur, they are published as a sequence of events, and then any number of downstream consumers can receive it anywhere they happen to be. And this is really key, you are decoupling the producers of the data from the consumers. It's what makes it operationally scale," LaForest said. "That connective tissue of decoupling all the different actors within an organization, so they can all produce and consume independently is the only way to make it really scalable across a large organization when you are sharing data."

Another relevant use case is cybersecurity. President Joe Biden's recent executive order encouraged agencies to adopt a zero trust approach to cyber. But LaForest said the problem there is again the scale; a key component of zero trust is the monitoring and threat detection in observability data, but there are massive amounts being produced, and agencies have to figure out what to use and which tools to send it to. That's where Confluent comes in.

"Doing the same thing we've always done and just calling it something different is not a solution," LaForest said. "Making another data lake doesn't constitute a data strategy. That's one technique. But really, you need more than that."

