

Forever Data: Tips for Managing Today's Most Vital Commodity

By: Jenna Sindle, Managing Editor



Just a few years ago, data was something an agency used once and then consigned to a warehouse to wait out the time that it had to be retained. But today, as we prepare for the AI revolution, data is no longer exiled to remote storage never to be used again. Instead it has become ‘forever data’- a vital commodity that will be used many times over in different agency processes and projects to provide mission critical insight.

This advancement into the age of AI and forever data represents a seismic shift for the federal agencies and their data management solutions may not be ready for this new approach to data. “Current data management environments are

built for a different generation,” explained ViON’s Michael Lamb. “Data becomes siloed, which creates three significant problems. Firstly, it is difficult to access data and put it to work. Secondly, it’s difficult to tier data between these silos, and finally, it’s very expensive. It was a great solution for ‘then’ when data wasn’t really touched again after about six months, but now, the way we use data is changing, so the way we store data should change too.”

To store data in a way that makes sense for on-going reuse in AI-driven activities, data management must be simple, scalable, cost-effective, secure, and searchable. “That’s a fairly tall order for a data management solution,”

noted Lamb. “But it can definitely be achieved with today’s technology and the right partner to help build an environment that is flexible enough to achieve all these requirements.”

In Lamb’s experience agencies become focused on the initial price to push data to the cloud. But when that data needs to be recalled agencies the costs of moving data out of the cloud add up quickly and data exfiltration is a slow process, particularly when an agency is trying to restore a large amount of data. To avoid this situation, federal agencies need to get their data management strategy AI-ready, said Lamb. “If you start from the perspective being ready recall all data and building a data management infrastructure accordingly, your organization will be ready for anything.

According to Lamb there have traditionally been two categories of data:

- 1.** Active data that will be used again. This includes, data like COVID-19 data at the National Institutes of Health as research continues into the virus, the disease, vaccines, and therapies.
- 2.** Temporarily inactive data, which is data that is generally useful and may be able to deliver additional value as applications develop. This could include data that needs to be accessed for BCDR and data that must be retained to comply

with federal laws.

However, today in the age of AI both types of data are classified as forever data and need to be readily accessible and inexpensive to store.

“Since the pandemic digital transformation has really accelerated, which increases the likelihood that data will find new applications more quickly than expected,” Lamb added. “I look at all the possibilities for NOAA’s data as more sensors are being deployed into the oceans and atmosphere and how quickly the NAM, GFS, and European models are learning from that information and improving the accuracy of weather forecasting, it’s like nothing we’ve ever seen before.”

As we enter the next phase of our data-driven future, the ability to manage data cost effectively will become a major asset and a critical advantage for federal agencies. Understanding how data is used today and how forever data could be put to work in the future is the first step to making AI mission-ready.