Transformative Data Strategy for Federal Agencies

Highlights from a Roundtable hosted by the Advanced Technology Academic Research Center (ATARC) in partnership with AWS, July 2024

Well-executed data strategies are now a cornerstone of operations in the Federal government. In a recent roundtable discussion, Federal experts explored successful data strategies, shared challenges with developing and executing data strategies, examined use cases enabled by strategic data initiatives, and discussed the importance of a data strategy to advance technologies, such as AI and ML.

"Data should be used to accomplish the mission much more effectively."

Although data strategies are imperative to mission success, many organizations do not have formal data strategies in place, or their strategies are stagnant and not widely distributed.

For some agencies, data in legacy systems was simply inaccessible until recently, and therefore never documented or explored. Now that data has been extracted, agencies are beginning to explore data sets, break silos, remove redundancies, and determine how best to utilize the data.

Panelists discussed the importance of identifying the right type of architecture and technology stack to support data from numerous sources, and establishing strong governance to guide data management. Often, this requires a significant culture shift within the workforce.

For some panelists, the goal of a data strategy is to make data more readily available to users, especially those working in the field. Developing a single source of data allows stakeholders to produce better quality data products and reports.

Other data strategies aim to collect and consolidate data from a myriad of sources, including anecdotal, social sentiment, and socioeconomic data, in order to make more informed decisions. However, panelists underscore the importance of developing data governance to protect sensitive information before utilizing data.

Challenges with Advancing Zero Trust

Panelists discussed challenges with developing and sustaining an effective data strategy. Agencies can sometimes operate in environments where technology is owned by another agency, which makes data management and funding acquisition more challenging.

At the enterprise level, each dataset is typically owned by a different entity. Access controls may make data accessibility and data delivery more challenging. Once data is accessible in business intelligence (BI) tools, challenges arise with determining business ownership, data custodian roles, and security protocols.

What makes a data strategy successful?

Governance

Agencies with strong leadership have been able to build governance structures that enable free flowing discussion through committees and the development of effective data strategies. Other panelists recommend building governance around a few low-risk use cases before growing and scaling. Ultimately, data governance should be based on the needs of an agency.

Interoperability

Agencies often have data services that are inherited, outdated, partially built, or forgotten about, which prevents interoperability and data accessibility. A part of a successful data strategy is selectively choosing and utilizing data services that align with business goals.

Tech Stack

“If you have a data strategy that doesn't align to your tech stack, you're not going to meet your data strategy.”

Panelists note the importance of understanding how the backend technology dictates a data strategy. Agencies should consider the type of databases used, whether data lakes, data warehouses, or data fabric architecture, or whether a relational database is needed. It’s important for agencies to know what type of technologies can support current and future data use and strategy.

Panelists also recommend having purpose built data services where agencies can choose from these services to support diverse data models and specific use cases. AI and ML technologies rely on data from a single source of truth, which enables automation and orchestration of information from hybrid cloud environments. Ultimately, as one panelist noted, tech stacks drive data strategies.

Strong Leadership

Panelists emphasized the important role of strong leadership who understand technology to make informed, strategic decisions about resources. Unfortunately, panelists often have to overly simplify complex technology concepts for leadership.

Data Lakes and Data Fabric Architecture

In addition to utilizing data warehouses for structured data, agencies are often using data lakes to capture semi-structured and unstructured data together. However, data is often trapped in silos, which makes bringing data into a single data lake challenging and expensive. Panelists are now seeing more agencies move towards a data fabric architecture.

AI and ML Use Cases and Challenges

Panelists also spent time discussing artificial intelligence and machine learning. Some of the most common use cases involve anomaly detection, such as with smart home aging-in-place technology to alert caretakers if certain signs of life triggers are detected. Other examples of anomaly detection include fraud detection. AI is also being used for customer support, as well as extracting data from images.

Roundtable participants openly discussed the challenges with these technologies, specifically generative AI. Because generative AI creates new information, agencies will likely face challenges with records management and data storage capacity.

Panelists also discussed challenges surrounding the accountability of AI outputs. In certain scenarios, it's unclear whether the technology provider or the user should be held accountable for inaccurate, false, or misleading outputs, or outputs that lead to legal challenges. There are many factors to consider, including the point in time where responsibility shifts, how accountability is defined, and how enforcement should work. Panelists shared encounters with vendors claiming to solve complex problems with AI, yet have not considered the legal ramifications of doing so.

"We've got to get people to not trust their computer. That's going to be a hard thing to do, especially since we've taught them to trust it."

Panelists highlighted the critical role of humans in AI and ML use. As technology accelerates and generative AI continues to proliferate the market, panelists contend that AI outputs must be audited by humans to reduce risk of liability to the agency. Since AI hallucination is exceedingly common and often goes unnoticed, the potential for liability is high. Unfortunately, there are not enough man hours in the workforce to audit AI outputs at the rate technology is accelerating.

Some agencies on the panel provide multiple avenues of training and development for their workforce, including mentorship programs, coaching programs, shadowing personnel, cross-platform training, and more casual learning workshops. Other agencies are focusing on training everyone in the workforce on the foundational concepts of AI and ML.