

[Defense Department Has a Data Problem](#)

In order to fix it, the chief data officer is working on improving data quality and making data sets machine-readable.

[Amanda Ziadeh](#)

Thu, 01/17/2019 - 17:18



Photo Credit: Phototreat/iStock

Despite legal requirements for the federal government to share all its non-sensitive data, the Defense Department is falling short, according to his data chief.

The Defense Department is [publishing](#) financial data and other datasets, but it's still challenged with data quality and cleansing, and sharing data in machine-readable formats, rather than PDFs, according to Michael Conlin, the Defense Department's chief data officer.

The Digital Accountability and Transparency Act (DATA) requires the federal government to transform its spending information into open public data, and the

recently passed Open, Public, Electronic and Necessary (OPEN) Government Data Act requires all non-sensitive government data be available in standardized, non-proprietary machine-readable format.

At the DOD, Conlin embraces the idea of citizen data scientists, and making defense-related data available to academia and industry so researchers and developers can train models to propose solutions and uncover insights.

But “almost none of the IP addresses that are accessing Department of Defense data are in the United States,” Conlin noted at the Jan. 16 ACT-IAC Artificial Intelligence and Intelligent Automation Forum in Washington, D.C. Tuesday.

So, though the Pentagon is publishing data, Conlin identified three areas of improvement and challenges that remain.

Identifying the Data Problems

First, policy now requires agencies to publish machine-readable data with application program interfaces, not just raw data sets. “The data that we publish are human accessible, not really machine accessible,” Conlin said. “Maybe we’re a little behind on the API space, so we’re trying to fix that.”

Conlin said the Defense Department has published all its financial data and research and development spending data, whether the research was done by the department or third parties. But these data sets also are published in PDF.

“The more signals we get from citizen data scientists that happen to be in the United States, the easier it is to generate some level of activity around that,” Conlin said, “but we do think more should be accomplished.”

The second challenge is the department’s concern about what it is revealing with that data. “The reality is, probably, we’ve already revealed uncomfortable levels of information without realizing it,” Conlin said.

There was a time data could be hidden in all the disorder and complexity of data sets, but that becomes increasingly more difficult today.

“Now with a credit card, anybody . . . can start using online [artificial intelligence] services from general purpose cloud members like Amazon, and then you can start

connecting dots with the data that's already there," Conlin said. "And there — there's a real internal tension."

He questioned how the Defense Department is to know if it is inadvertently allowing people to access certain data that then allows them to connect certain dots.

Lastly, Conlin discussed the importance of machine-learning algorithms and how pre-trained algorithms could be limited to a specific problem. But the Defense Department can't fully take advantage of transfer learning without clean, usable data sets.

There are pockets where the department is doing transfer learning, but it's not doing so broadly because "our data is of incredibly crap quality," Conlin said. "My biggest challenge right now is improving the quality of the negative of the Department of Defense."

Conlin is referring to the data maturity benchmark. The department calculated this benchmark against the [Federal Government Data Maturity Model](#), a five-stage maturity model from a low to high capability rating, from one to five respectively, based on six categories: analytics capability, data culture, data management, data personnel, systems/technology and data governance.

"We benchmarked ourselves, and we had external parties benchmark us, and we uniformly came in at level one mature. the lowest possible level of maturity," Conlin said. "And when we explored that the reason we were a level one is that there is no negative levels."

So, the department is working to clean its data, one of its biggest challenges right now. There are pockets of excellence, but in order to truly improve department-wide, Conlin said it's important to pick and identify a maturity model.

"If you try to perform at a level beyond where you are, you'll typically fail. And that's true of the spaces where you need to improve step by step in order to be able to perform more effectively," he said.

Fixing the Data Problem Inside Out

Conlin has three big documents that rule his world: the President's Management Agenda, the [National Defense Strategy](#) and the National Defense Authorization Act.

The [NDAA](#) refers to common enterprise data.

"The defense business enterprise shall include enterprise data that may be automatically extracted from the relevant systems to facilitate Department of Defense-wide analysis and management of its business operations," the NDAA says.

It then says the Office of the Chief Management Officer is tasked with extracting that data and using it to help decision-makers to make those decisions that enhance department-wide oversight and management.

"Part of the challenge is that we don't have any department-wide systems," Conlin told GovernmentCIO Media & Research. Each of the department's services and agencies have their own systems, and it's an "incredible sprawl of IT."

And the task of "enhancing department-wide oversight and management" poses its own question of what Defense Department business leaders are trying to accomplish and how they define that management.

The National Defense Strategy discusses improving affordability and performance, but that also poses another question. "How do you know when you're done?" Conlin asked. "You're done when you achieve commercial sector levels of efficiency," a goal outlined in the President's Management Agenda.

So, how does the government reach a public sector-level of efficiency?

Conlin said it starts with business leaders and asking them what data they need to achieve the outcomes they are targeting. This way, Conlin is able to identify critical business questions for department business leaders.

"Once we get the critical business questions . . . we get the experts to make sure we can standardize our data in a way that is accurate and meaningful," Conlin said. "Then we generate the answers to those questions, and we generate the comparison against the commercial sector-levels of efficiency."

This way, the department can stay on track with where it has reached commercial sector-level of performance and where it still needs improvement.

[View printer friendly version](#)

[Data management](#)

[DOD](#)

[data analytics](#)

[data](#)