

## Executive Viewpoint

# DATA DRIVES DOT MISSION

Harnessing massive volumes of disparate data is a significant but valuable challenge.

Nearly every government agency is awash in huge volumes of increasingly varied data. Everyone acknowledges that data can help any agency fulfill its mission, but deriving value from an increasingly disparate and dynamic dataset is a challenge. GCN caught up with **Daniel Morgan, Chief Data Officer at the Department of Transportation**, to hear his thoughts on data analytics and how the DOT can best harness data to drive its mission forward.

### What does data mean for the DOT?

We have a varied mission at the DOT. We invest in, monitor, and regulate the national transportation system: we use data to manage the safety and operations of the national airspace; and we also operate the St.

Lawrence Seaway with our partners in Canada.

Unlike the corporate sector, our data is truly varied. Some of it is tabular; but it also includes images, video, and spatial data. That makes for an interesting

data-management challenge, but we have to use data every day in our decision-making. It informs how we approach delivering our programs to ensure we have a safe and efficient transportation system that serves everybody.

### What do you think about such things as predictive analytics and how that can help?

It depends on what you're trying to predict. One thing I'm interested in is measuring risk in the transportation system. The number of roadway fatalities over the last two years went up eight percent – each year. We went from losing around 32,000 lives just a couple of years ago, to 37,000 lives in 2016.

That census by itself isn't good enough for us to understand whether risk is changing in the environment. Are people driving more? Why are they driving more? Are their jobs in different places? Are there new vulnerabilities in the system? Or have travel patterns changed to the extent that we're seeing old vulnerabilities just now being revealed? It turns out we need more and different data to measure those kinds of risk factors, and not all of it resides inside the DOT.

So when I think about predictive analytics, sometimes it's a matter of measuring risk, sometimes it's a matter of understanding some challenges with regulated entities. Sometimes there are even fraud schemes. For instance, we need to estimate the risk that a new registrant is not just a reincarnation of a previous registrant.

### When it comes to data analytics, are you trying to get more information from the data you already have, or to get a better idea of other data you need?

There are a couple of things happening at the DOT. Do we want to make better use of the data we have? Absolutely.

Transportation happens within both an economic and a social

“It turns out we need more and different data to measure those kinds of risk factors, and not all of it lives inside the DOT.”

context. It's based on how we use the environment, how we zone housing, and how all of these and other land use characteristics drive how people interact with the environment. So I have to bring in other datasets to build context around the things I'm observing in my own transportation data. I have to start building correlations worth exploring to help us understand the extent to which our programs are working and where we might better invest our time.

### Has the new generation of analytical tools given you any idea of new applications and services you can use to further the DOT's mission?

It's not really about new tools. There are decades-old ideas about what drives data analysis that are still applicable today. One is the increasing desire for quantification in a wider range of fields. Turning text into numbers means you can use that for data analytics. Similarly, with sound and video, there are a bunch of things you can do when you turn that data into numbers.

Then there is the ever increasing power of computing. We no longer need a supercomputer or mainframe to do some of

the things we need to do. This is driving a liberation of analysts who have been stuck with subpar tools. We now have an opportunity to reinvigorate the analytical workforce within the agency by equipping them with tools that work faster.

#### **What does that mean for the kind of resources in which you need to invest?**

It's about the capacity of our people to do the work, and ensuring we equip them with the right tools to do this kind of work safely, securely, and ethically. There needs to be a data management and governance process underlying all of that with the aim of providing those analysts with the best possible data. That means having programs and practices in place not only to improve the quality of the data where it lives, but also to democratize the availability of that data.

#### **A big concern is having access to data that is both clean and reliable. What improvements can and need to be made?**

This is indeed a challenge, and it breaks down into two kinds of dynamics. Take the census of fatal car crashes. That begins with a police crash report, and that's a very human endeavor. It's open to a great many challenges as it relates to accurately recording data and ensuring the data fields are complete and consistently filled out. Still, when there are inconsistencies and things missing in the data, we have to deal with that. Our systems and our analytics must be resilient enough to deal with human frailties and the needs of the real world. That's never going to go away.

#### **How do you factor this less-than-perfect data reliability into the analyses you have to do?**

There are 60,000 people in this department. Our analysis and analytical expertise are distributed throughout 10 operating administrations and across the country. My job is to give them all the tools they need to do their analyses. I don't decide whether any given data is fit-for-purpose.

“We now have an opportunity to reinvigorate the analytical workforce within the agency by equipping them with tools that work faster.”

And that's OK. What I want to do is build community, to share code, and make it easier for people to find folks working on similar problems so they can share their practices and technology. We need to continue to build that culture into the agency so we can overcome the challenges. To that extent, I'm more of a trust broker than anything else.

#### **What does all of this mean for system performance and security?**

There's a bunch of data coming off those systems that we could use to tune their performance and improve security. Then there's the way we can architect our systems to better feed the analytical needs of the agency. That starts to point to ways we need to work through our system architectures to distribute the processing workload. That means thinking about cloud architectures that we can use to help us offload some of the processing from our network and servers.

#### **What are some emerging technologies that could make a difference to how you manage and deploy data?**

There's not much that is truly brand new. And it's not just about the technologies themselves, but the user experience around them that is important. For example, I'm looking for ways to make it easier and less expensive to build an analytics platform using a server-less architecture, and I'm starting to see cloud providers understand some of that.

For the most part, I believe it's about the evolution of things already in place. Technologies that make it simpler, easier, safer, and more secure to do things with data are definitely where I'm looking for the most advancement.