FCW

Reframing big data

To understand how big data can benefit your organization, refocus how you use information and let your imagination be your guide.

As government agencies and private businesses alike continue to grapple with defining, storing and managing big data, questions about the massive quantities of information that are generated daily also proliferate. The first step to understanding big data is to reframe how we think about it, said David Robinson, vice president and chief innovation officer at SAP Public Services during a panel discussion titled "TechAmerica Briefing: Demystifying Big Data – A Practical Guide to Transforming the Business of Government" at the FCW Executive Briefing "Is Big Data a Big Deal for Government?" on Dec. 12, 2012.

"I'm still waiting for the time – hopefully it's very soon – when we stop talking about it as big data and start talking about it as the data," Robinson said. "We're trying to use data historically to get insight, to see something across different systems or processes, and I think what's different now under the term big data is the velocity of the data."

Robinson noted that it is believed that today as much data is generated every 12 hours as was done from the start of humanity until about 2003.

"The issue there is not about how do I manage the data or make sure it's secure or efficiently used, but what are new outcomes, what are things that I could not do before because I did not have the infrastructure, the performance, the security and all of those capabilities in place to do something that just never could be done before," he said. "Forget about could not be done efficiently, just could not be done."

Where to Begin

What makes big data so daunting is in its name: the sheer size. Having a clear starting point for how big data can help an organization meet its mission can help data scientists at those places determine where to begin. To do that, inventory what you have today and what you can do with it, Robinson said, and then reorient how you use it.

"What else can I do as I learn, build and grow out the use or application of this data, augmenting it over time incrementally and expanding the use case applications?" Robinson said.

Use cases can run the gamut from weather and traffic to waste, fraud and abuse, said Dante Ricci, director of SAP Federal Innovation during the conference's "Technology Insights II: Extracting Actionable Information from Agency Data – What's Possible?"

segment. That's why understanding the data's context and pinpointing mission-critical needs are so important, he said.

"You really do have to take the [information technology] organization and the business organization together, put them together in a room, understand from – I would call it – a design-thinking approach as to who you're customer's customer is," Ricci said.

For example, what is the persona of the end customer? Is it someone

> "Imagination, crowd sourcing and obstacles are really all coming together to create a whole new range of use cases for data."
>
> – David Robinson, vice president and chief innovation officer at SAP Public Services

in field services doing inspections at restaurants? What type of information would that person need to do his or her daily work? What is possible now that was not possible a couple years ago that they could do with mobile devices combined with big data combined with cloud architecture?

"A lot of organizations are looking at combining big batch processing with predictive real-time data platforms to garner new insights," Ricci said.

For example, SAP has been working with universities on how to retain students by looking in real time at the

Usage Trends

With big data comes new ways to analyze information. Here are a few examples of what's possible.

- Examining real-time information on successful university students to determine their characteristics vs. those of students who drop out before graduation.
- Applying similar algorithms to ascertain the best placement for children in foster homes.
- Making information on medical conditions public so that health care providers worldwide can comment on what works and what doesn't for a given condition, such as cancer, in a person with a certain lifestyle, such as smoking.

behavior patterns of successful students vs. those of pupils who leave before graduating. Another client extrapolated that for use in child services, monitoring how children are placed in foster homes.

The result is a new business approach – a partnership between SAP and its customers, Robinson said.

"What we're starting to see in the last two years is that the best innovators are our customers," he said. "The challenge for us is to not focus on what cannot be done and to not try to control our imagination or look at the down side necessarily but really make sure all this data is made available – open data, open-source tools – and let our imaginations be the co-innovators."

What's Next

Agencies claim they use about 5 percent to 8 percent of their data for active purposes, Robinson said. Innovation is driven by securely opening it up to allow internal stakeholders, partners and the public to view and use the information.

"That highlights the difference between open data and big data," he said.

Another difference is that about 80 percent of the world's data is now unstructured, and the majority of solutions are built around structured data, said Robert Ames, senior vice president for information and communications technologies at In-Q-Tel, who delivered the opening keynote, "Big Data: Capturing the Information Opportunity in Government."

Companies that are working to change that are opting to make their innovations open sourced.

"These companies are making strategic decisions, in many cases, that their value proposition is their data, it's not the technology they create," Ames said. "That is incredibly helpful to the government, because now we have these technologies that are out in open source, which have a very powerful and active development community and they're open and they're not proprietary and they avoid the lock-in."

Moreover, the amount of data and the speed with which it's produced can change how we analyze information, Ricci said. "We're going from looking behind to looking forward, but not just looking forward in a predictive manner but also maybe in a prescriptive manner."

For example, he said, if the information about a cancer patient with a specific pathology type and lifestyle was made available, people and physicians worldwide could add to the file about what treatments have been tried rather than that person going to a single specialist at a single hospital.

"Imagination, crowd sourcing and obstacles are really all coming together to create a whole new range of use cases for data," Robinson said. "Data's not just about being in the back office or being a data scientist and really doing deep analytics. All of us are basically swimming in big data -- using, consuming and contributing to big data. I think the difference is the outcomes and the engagement for each and every one of us for scenarios we never thought of before." ●