HHS CTO Reveals New Data Science Platform to Address Sharing Fears

The agency's new initiative incorporates AI and machine-learning capabilities to advance health research.

Faith Ryan

Tue, 12/10/2019 - 09:57

Ed Simcox, chief technology officer of the Department of Health and Human Services, discusses automation's role in effective data sharing. Photo by Rodney Lamkey.

As the largest federal civilian agency in the U.S., the Department of Health and Human Services is rich with data. However, it has lacked ways to scale and use the data to its full potential, in part due to considerable fears people have about sharing research data with others and with the involvement of artificial intelligence.

Focusing on the need to advance health science as well as a human-centered

design practice to alleviate data-sharing concerns, the HHS Office of the CTO launched an internal data platform in support of its Data Insights Initiative that aims to address these issues, its chief Ed Simcox told attendees of the GovernmentCIO Media & Research AI and RPA CXO Tech Forum last week.

Prior to establishing the platform, Simcox and his team went on a federal-listening tour — speaking with scientists, researchers, actuaries, administrators, and other professionals across HHS — to understand the primary fears and desires toward sharing data.

"We spent many months gathering requirements at the people level … to figure out how we could create a platform that would serve as much of the need from an enterprise-level as possible," Simcox said.

Simcox and his team ultimately identified two recurrent themes that people felt prevented them from sharing data. The first was that HHS and its affiliated agencies are starving for data insights, but that they hold a great reluctance to sharing data due to individual privacy regulations, such as HIPAA.

"There is a multitude of lesser-known restrictions and regulatory issues associated with a lot of our data sets," Simcox explained. "Sometimes, you layer on three or four of those and it makes it supremely difficult to try and surface that data and put it to better use."

The second was that people want a sense of stewardship over the data they collect. Simcox raised the point that researchers who have gathered years of data do not necessarily want to block others from using it; some are concerned that their data can harm others based on how it is analyzed, and consequently result in poor policy decisions.

"We wanted to build into the platform a way to engender trust with data stewards so that they felt at all times they had visibility into that data and the way that it was being used," Simcox added.

Data stewards will be able to access the real-time computing environment to see how their data is being used, and will also be able to view a historical record dashboard to see who accessed their data, at what time their data was accessed by a user, and what algorithms the data was run on so they can run it themselves, Simcox said.

Another critical component for data stewards using the platform will be automated data use agreements (DUAs).

Completing DUAs is a necessary administrative requirement for researchers who want authorized access to federal data sets, which can take months for a set of professionals to manually process and review, Simcox said.

"One of the things that we're building into the platform is a way to automate that process — a document of routing and creation type of a mechanism — where we can template these things and adjudicate these things one time," rather than requiring personnel complete multiple, time-consuming steps, he said.

By automating this process, the goal is that a significant portion (80%) of the data-use agreement will be reused, eliminating timely administrative hurdles for researchers and allowing them to view and run analyses on data sets that have been approved on the platform.

"The reason this is important to AI, I think, is twofold," said Simcox. "Going back to the stewardship of the data, we want people to feel comfortable that they have visibility into how their data is being used. That is supremely important."

"It's very difficult to get folks to share data in the first place," Simcox said in closing. "But if they feel like they're part of the team that is doing the work, and they feel like it's a collaborative environment where you have auditing of the entire process … and in the corner of the screen, they have the ability to look at the data use agreement, they can then make sure that the data is being accessed and used within the four corners of that data use agreement."

[View printer friendly version](#)
[CXO Tech Forum AI RPA 2019](#)
[data](#)
[Health IT](#)
[research](#)

[HHS](#)
[AI](#)
[automation](#)
[Standard](#)