

Open-Source Data Brings Challenge and Opportunity



Megan Dane, director of plans and programs in the Office of Naval Intelligence, makes a point during a panel discussion on open-source data. *LISA NIPP*

NATIONAL HARBOR, Md. – Open-source data is a “fascinating,” if vexing, issue that has transformed how information is disseminated and consumed, according IT professionals in an April 5 panel discussion at Sea-Air-Space 2022.

“When we say open-source intelligence and open-source information, it could be literally anything you see on the internet,” said moderator Shane Harris, a senior national security writer at the Washington Post. “It could be things that are produced by the press. It is tweets, it is YouTube videos. It is an overwhelming amount of information.”

Panelist Joseph Obernberger, a software engineer in Space & Intel for Peraton, said his interest is in “big data.” Peraton assists government agencies with global national security, enterprise IT and cyber solutions and supports missions that include cyber, digital, cloud, operations and engineering. Obernberger said the problem with scale and managing information is a priority for him. Furthermore, open source is a challenge because there is so much data – “a lot of stuff” – in which the intelligence community is not interested to have as open source.

“[Open source data] is huge problem,” said Obernberger. “The number of Tweets per day, the number of YouTube videos per day. Seven hundred and twenty thousand hours of YouTube videos are uploaded per day. If you were to watch that, it would take 82 years. So, how can we build systems that would scale to that level? If you consider just a billion records. If it takes a computer one millisecond to process a billion records,

that is 11 and a half days for one system to do that. We need to deal with trillions of records.”

Panelist Megan Dane, director of plans and programs in the Office of Naval Intelligence, said, “We are really concerned with what types of information we are looking at and what we’re not looking at. We try to really leverage the commercial industry and what you are able to create through big data analysis and things of that nature, and then really pinpoint through requirements what information sources and streams we need to ingest, and then really clear the way for our analysts so that they don’t have to ingest or syphon through all the rest of it. That is really the most important part for us in that front-end proces.”

Panelist Andy Henson, a senior vice president for artificial intelligence at SAIC, said it has “gotten harder to know what matters.” He suggested a method for handling so much data involves knowing what to look for.

“My simple filter is, what question do we want to ask with the data?” Henson said. “That gets rid of a lot of noise. What question do we want to ask of the data, and then we can get to a real subset of the data and start getting at some of those challenges.”