

Will moving to the commercial cloud leave some data users behind?

7 February 2019



Georgia Tech assistant professor Mariel Borowitz is shown with satellite communications equipment. Dramatic growth in the generation and collection of data will change the way federal agencies make data available. Credit: Allison Carter, Georgia Tech

As part of their missions, federal agencies generate or collect massive volumes of data from such sources as earth-observing satellites, sensor networks and genomics research. Much of that information is useful to commercial and academic institutions, which now can usually access this publicly-generated data from agency servers at no charge.

But as the volume of [data](#) continues to expand, many agencies are considering the use of commercial cloud services to help store and make it available to users. While agencies may have different strategies, these new partnerships could result in [user fees](#) levied on downloads and analyses performed on the data while it remains in the cloud.

Writing in a policy forum article published February 8 in the journal *Science*, a Georgia Institute of Technology space policy researcher who studies

such data use urges caution about the design of these commercial cloud partnerships and possible imposition of user fees.

"Under the current system, free and open government data is used by scientists to conduct research, by entrepreneurs to create new businesses, and by citizens and other organizations to promote government transparency," said Mariel Borowitz, an assistant professor in Georgia Tech's Sam Nunn School of International Affairs. "If users must pay fees to download or analyze the data, this will decrease the ability of these users to access and work with data. Past experience suggest that the impacts of this decrease in data use could be large—both for individual users and for society as a whole."

Moving data to commercial cloud systems would likely provide broader access and more efficient analysis options, but she cautions those advantages could be offset by the cost, particularly for organizations with small budgets.

"Agencies risk losing some of the benefits of this transition by not budgeting for the costs associated with data downloads and analysis, up to a reasonable level," Borowitz said. "Many who would be interested in using the data may not be able to pay the associated fees. Researchers, nonprofit organizations and others who do not directly profit from the use of this data are most likely to be affected."

Borowitz recently spent two years at NASA and witnessed both the development of systems that will dramatically increase data collection and debates about future data storage. She recently authored a book, *Open Space: The Global Effort for Open Access to Environmental Satellite Data*, published by MIT Press.

She would like to see the agencies that provide data continue to shoulder the costs, up to some

"reasonable level," to ensure that the data continues to be readily available to all users. As an alternative to commercial services, some agencies are considering development of their own, custom-built cloud solutions, and will have to weigh the cost of benefits of the different options. There will also be technical, organizational and policy issues to consider.

"Agencies are taking seriously issues of security and long-term preservation of data," Borowitz added. "When working with commercial providers, some are concerned about the possibility of getting 'locked in' to one provider, due to the large costs of migrating data from one system to another. It is possible that costs and capabilities could change over time. On the other hand, commercial cloud providers have large workforces and extensive infrastructure that allow them to provide services and capabilities well beyond what any one agency would be able to maintain."

Borowitz notes that most agencies have not made final decisions about their cloud-based programs, so there should be adequate time to work through these issues.



Georgia Tech assistant professor Mariel Borowitz is shown with satellite communications equipment. Dramatic growth in the generation and collection of data will change the way federal agencies make data available. Credit: Allison Carter, Georgia Tech

"Most agencies that make data publicly available, particularly science agencies, are already

discussing and/or beginning to make the transition to cloud systems," she said. "However, these programs—at agencies like NSF, NIH, NASA and NOAA—are still in their early phases, and there is still opportunity for feedback to be provided and adjustments to the programs to be made."

The existence of fees for access to government data is not without precedent, but Borowitz argues that past experience suggests that user fees result in significantly less use. Before Landsat data—satellite imagery of Earth—was made freely available in 2008, no more than 25,000 images a year were purchased from the collection. "Within a few years of implementing the free and open data policy, the government was distributing 250,000 images a month," she said.

That number provides a suggestion of what the often cash-strapped agencies are dealing with. According to the paper, the National Oceanic and Atmospheric Administration (NOAA) houses more than 100 petabytes (PB) of data and generates more than 30 PB per year from satellites, radars, computer models and other sources. NASA projects that its archive will grow to 250 PB by 2025. And the amount of genomic data at the National Institutes of Health is growing exponentially.

A petabyte is 1,024 terabytes, or a million gigabytes. A gigabyte is 1,024 megabytes. For scale, an average photograph taken by a high-end cell phone camera can be in the neighborhood of 10 megabytes. Laptop computers may be able to store as much as a few terabytes of data.

Borowitz sees the transition to cloud computing as both an opportunity and a challenge for the future availability of government data. "The decisions being made right now about the structure of these programs have the potential to significantly impact researchers and society as a whole, so it is important to raise awareness and increase engagement on these issues."

More information: "Government data, commercial cloud: Will public access suffer?" *Science* (2019). [science.sciencemag.org/cgi/doi/10.1126/science.aat5474](https://www.sciencemag.org/cgi/doi/10.1126/science.aat5474)

Provided by Georgia Institute of Technology

APA citation: Will moving to the commercial cloud leave some data users behind? (2019, February 7)
retrieved 20 February 2019 from <https://techxplore.com/news/2019-02-commercial-cloud-users.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.