



Subscribe (<https://www.datanami.com/subscribe/>)



4166980)  (<http://www.datanami.com/feed>

VENDORS

10/07/2017 21:04

 (http://tw
/share?re
%E2%80
%99s%2C
%3A%20I
url=https
%3A%2F
%2Fwww
challengi
big-data-
now-
integrati
privacy%
via=datar

 (http://w
/share?r
%3A%2F
%2Fwww
challengi
big-data-
now-
integrati
privacy%
t=What
%E2%80
%99s%2C
%3A%20I

 (http://w
/share?r
url=https
%3A%2F
%2Fwww
challengi
big-data-
now-
integrati
privacy%
title=Wh
%E2%80
%99s%2C
%3A%20I

 (http://w
/submit?i
%3A%2F
%2Fwww
challengi
big-data-
now-
integrati
privacy%
title=Wh
%E2%80
%99s%2C
%3A%20I

 (http://w
/print?url
%3A%2F
%2Fwww
challengi
big-data-
now-
integrati
privacy%
title=Wh
%E2%80
%99s%2C
%3A%20I

concerns about possible misuse of these kinds of information, such as discrimination in hiring or in purchasing health insurance, if this information is not handled properly. The healthcare community is on the front lines in this area, but, given the complexity of issues involved, progress in addressing these concerns is very slow."

Big data has the potential to help society in massive ways, but it will require solving thorny privacy questions that go right to the heart of the matter, Stonebraker says.

"Privacy is a really good big data question," he says. "Imagine this simple example: you show up at your doctor's office and have an x-ray done and you want the doctor to run a query that shows who else has x-rays that look like yours, what was their diagnosis and what was the morbidity of the patients. That requires

integrating essentially the country's entire online medical databases and presumably would extend to multiple countries as well.



Daphne Koller, adjunct professor of computer science at Stanford University



Vipin Kumar, Regents Professor and William Norris Chair in Large Scale Computing at University of Minnesota

"While that is a daunting data integration challenge, because every hospital chain stores its data with different formats, different encodings for common terms, etc., the social value gained from solving it is just huge," Stonebraker continues. "But that also creates an incredibly difficult privacy problem, one that I believe is not a technical issue. Because by and large, if you're looking for an interesting medical query, you're not looking for common events; you're looking for rare events, and at least to my knowledge, there aren't any technical solutions that will allow access to rare events without indirectly disclosing who in fact the events belong to."

We must find the right balance between making data public and respecting people's privacy. It's a question of "how public is too public," said Professor Blei.

"To me, we get the most bang for our buck if we make everything public; however, that's going to have some serious security and morality issues with it, so we shouldn't do that," the ACM Fellow said. "But making everything completely private and not benefiting from that data also doesn't seem like a great option. I think this is a difficult, thorny issue. It lives at the intersection of policy, philosophy, morality, computer science, data science, and machine learning."

There are no simple answers to questions like these. But considering the societal good that will come from solving them, it's a challenge that must be undertaken.

Related Items:

How to Survive Tech's Chaotic Transformation (<https://www.datanami.com/2017/06/30/survive-techs-chaotic-transformation/>)

Big-Data Backlash: Medical Database Raises Privacy Concerns (https://www.datanami.com/2014/04/17/big-data_backlash_medical_database_raises_privacy_concerns/)

Share this:

[Tweet](#) [Share](#) [f Share](#) [G+1](#)

[+ reddit this!](#) (<http://www.reddit.com/submit?url=https://www.datanami.com/2017/07/05/whats-challenging-big-data-now-integration-privacy/>)

Tags: big data (<https://www.datanami.com/tag/big-data/>), integration (<https://www.datanami.com/tag/integration/>), machine learning (<https://www.datanami.com/tag/machine-learning/>), privacy (<https://www.datanami.com/tag/privacy/>)

Contributors



Alex Woodie
Editor in Chief



George Leopold
Contributing Editor



Tiffany Trader
Contributing Editor



Steve Conway
IDC

Featured Events

Artificial Intelligence Conference San Francisco 2017

(<https://www.datanami.com/event/artificial-intelligence-conference-san-francisco-2017-2/>)



September 17 -
September 20
San Francisco CA United States

Strata Data Conference New York 2017
(<https://www.datanami.com/event/strata-data-conference-new-york-2017/>)



September 25 -
September 28
New York United States

Big Data at Cloud Expo West 2017

(<https://www.datanami.com/event/big-data-cloud-expo-west-2017/>)



October 31 - November 2
Santa Clara CA United States

Chief Data & Analytics Officer Week 2017
(<https://www.datanami.com/event/chief-data-analytics-officer-week-2017/>)



September 25 -
September 28

DCD >Colo + Cloud
(<https://www.datanami.com/event/dcd-colo-cloud/>)



September 26
Dallas TX United States

AI World Conference & Expo 2017

(<https://www.datanami.com/event/ai-world-conference-expo-2017/>)



December 11 - December 13
Boston MA United States

[View More... \(https://www.datanami.com/events/\)](https://www.datanami.com/events/)