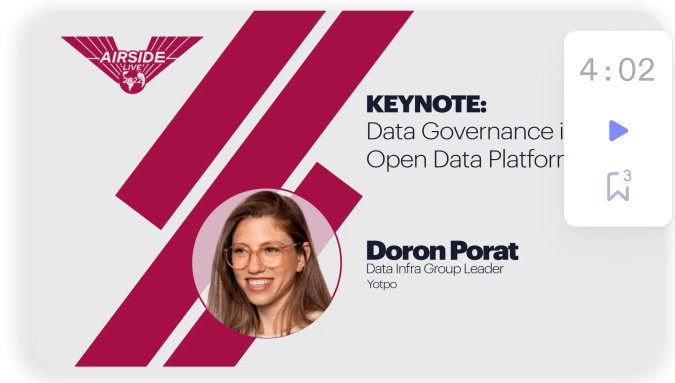


BLOGS

How Yotpo Governs Its Open Data Platform for a Secure Data Democracy



JUN 22, 2022

by **Brent Danis**

Learn more

Doron Porat firmly believes in letting the people who know the data most intimately own it.

As Data Infrastructure Group Leader at Yotpo, a leading marketing platform that drives growth for e-commerce businesses, Doron knows opening access to data is the key to scaling their operations. She also knows that working with an open data platform allows her team the flexibility to

scale infrastructure as Yotpo grows. “An open data platform is built kind of like a puzzle,” she said during her recent session at Airside Live. “There are all these pieces and components you can extract and replace, you can enhance and extend, and the way you can keep growing and evolving. And the way you build systems to last for years.” Simply put, an open data platform comprises software from many vendors versus one provider (a la Snowflake or AWS).

4:02



But while Doron is convinced an open data platform is the best way to future-proof your data stack, she admits it isn't always easy. Working with many vendors creates complexity – around compliance, privacy, integration, maintenance, cost management, observability, and access management. Without proper tooling, organizations are left scrambling trying to manage these challenges.



Yotpo's Open Data Platform

Solving Challenges of the Open Data Platform

Doron's answer to the complexities of managing an open data platform is strong data governance, which she defines as "whatever means you can think of that allows you to control and manage data." For this, Doron relies on leading data governance tools – but not just any solution will do. She has a firm set of criteria her team uses during the selection process:

- **Seamless.** Software changes or additions do not equate to interrupting people's work.
- **Hermetic.** Systems – however complex – require management tools that cover the entire platform.
- **Scalable.** Data governance solutions must scale right along with the platform.
- **Dynamic.** Your open data platform is bound to change and grow over time; your data governance stack will need to move right along with it.

Dissecting the Data Governance Stack

Yotpo has four unique solutions in its data governance stack: data catalogs (metadata management tools), operational data monitoring, data quality monitoring, and access management. Each of these areas poses challenges in an open data platform environment, which Doron covers in depth. To get the complete story, you can [watch the session](#), but here are a few highlights:



Data Catalogs / Metadata Management Tools:

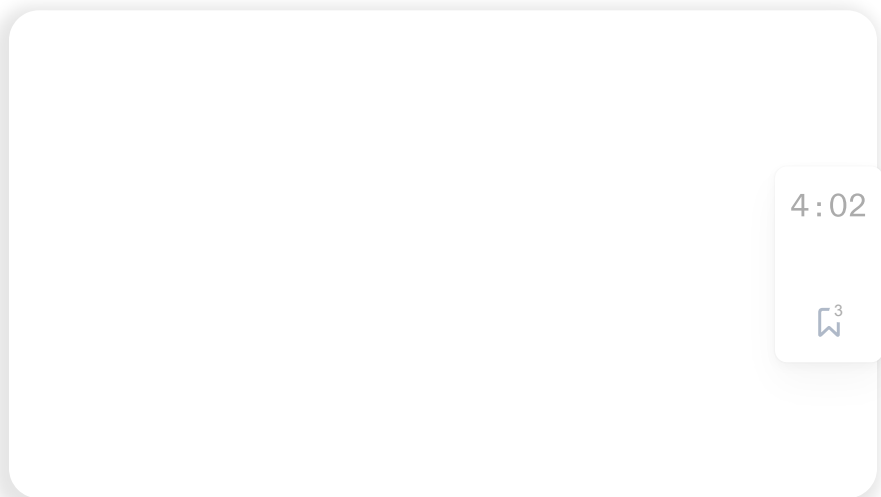
With tons of data formats and no single store, metadata collection and management can be difficult. Yotpo chose Hive Metasearch as the tool for the job.

4:02



Operational Data Monitoring: If it's difficult to understand where data and metadata live, try understanding the behavioral and usage patterns for those data assets within the data ecosystem. This, says Doron, is operational data monitoring. Exposing metrics, tracking the data life cycle and costs, and optimizing across the various tools of the open data platform can be "very, very hard and time-consuming."

Doron says it's also difficult to find dedicated operational data monitoring solutions for an open data stack. "Usually, there are these big solutions, and they have some sort of a model for managing big data. But nothing that I know of that's really designed to support big data ecosystems," she says. "What we are doing [at Yotpo], we rely on Prometheus to monitor the metrics from around the system and have it flow into Grafana where we have all these alerts and monitoring."



Yotpo's Data Governance Stack

Data Quality Monitoring: There's potential for a lot of noise and outliers with so many systems, which "can prevent finding the core issues where the data is not okay." Doron also notes, "a lot of these solutions only alert you once the data has landed in your data platform, which could be too late." Yotpo uses Monte Carlo and Great Expectations for different levels of quality assurance.

And now to our favorite topic, and where Yotpo uses Okera.

The Future of Data Governance for the Open Data Platform

"I think that data democratization leads us to a point where self-service is at the essence of everything. The data stack is [taking on] more and more responsibility as time passes. That means more streaming [data] and more technologies [supporting data as a product]," Doron says. And

with all of this comes more need for governance, like **Access Management**.

The more systems you have, the more critical the need for auditing and control over who has access to what data. In the open data platform, there are many, many endpoints to protect. In addition, Doron says, “We have the physical layer as well as a logical layer, which also makes this hard because even if you protect your data via your metastore or your catalog, it still can be accessed via the actual files, and solutions are reliant on having some sort of centralized user management.”

4:02

3

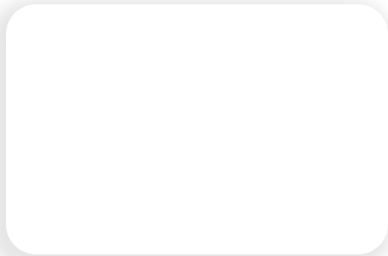
Doron says, “The principle of having this one abstraction layer to protect [your data] all around, this is a hard task to achieve.” It probably won’t come as a surprise then that Okera was the first to provide Yotpo with a viable solution that met all their needs.

To get the complete picture of what Doron sees as the future of governing data in the open data platform and more insight on all the topics covered above, [watch her session](#).

Share on Social

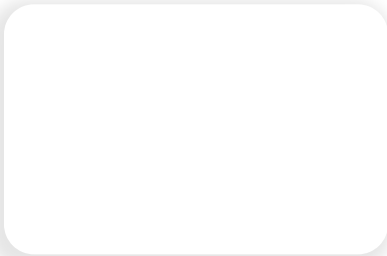


Recent Posts



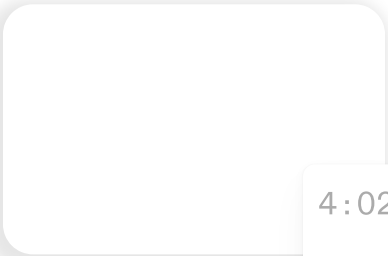
BLOGS

A New Chapter in Our
Journey: Okera is
Joining the Databricks
Family



BLOGS

Distributed
Stewardship Now
Easier with Role-based
Templates



BLOGS

Easy Onboard
Experience for
Dynamic Data Access
Management

4:02

3

[Privacy Policy](#) | [Terms of Use](#)

©2023 Okera, All Rights Reserved. | Site by Go Fish.

