Data Democratization – No, Really!

Modern

Data democratization is one of the data world's favorite buzzwords, but data management platforms have a hard time delivering it––even when they say they will. This article examines some of the goals and roadblocks on the way to data democratization and shows how DataOS makes the process simpler and, above all, genuine.

Goals and Roadblocks of Data Democratization

Data democratization is easy to describe but hard to implement. The goal is to give every user access to the data they need to make business decisions. There are many considerations when realizing this goal. Business users will need data skills that they didn't need under the old "gatekeeper" approach when access to data was strictly limited. Those same users will need self-service BI/Analytics tools to help them extract insights from the data they have. There are wide-ranging changes in business and data culture, such as transferring ownership and maintenance of domain-specific data sets to the departments that use them, rather than IT.

As complex as all of these issues are, though, some of the most intractable considerations have to do with data governance.

Governance problems arise as soon as we formulate the definition of data democratization. "Access for all" can never mean "access to everything for everyone." Matters of privacy, data security, and simple data overload necessitate some form of access controls. Just what data does a given user "need" in their work, and how can they get access to it while maintaining privacy, data security, and regulatory compliance? It may not be practical to restrict or permit access to individual cells in a data table. More often, entire columns must be restricted, or even entire datasets, in order to restrict the information at all. As a result, users lose access to data that they could use and ought to have.

Solutions to these issues include curated "data marts" and selected data catalogs that allow users to find and search domain-specific datasets. One benefit of this approach is that it spares users from searching the entire enterprise data domain in order to find information they want, but it also means that they may miss relevant data that isn't included in their data mart.

Another governance problem is the tendency of users to construct customized databases for their own particular needs. It makes sense in terms of making queries faster and simpler, but it creates new data silos just as the business is getting rid of them in the course of modernizing their data architecture. All the problems of silos—such as stale, duplicated data—come right back.

The Modern Difference

Attribute-Based Access Control (ABAC)
ABAC is a central part of governance with DataOS. DataOS provides a system of tags applied to both users (as permissions) and datasets at the row and column level. ABAC allows the governance policy to become highly granular and to grant access to selected parts of

diverse datasets. In effect, ABAC allows a governance team to create an unlimited number of "virtual datasets" with access granted by privilege level and business unit, without having to create data marts that provide access only to specific curated datasets.

For example, a governance team might create three tags: business_1, business_2, and business_3, offering decreasing access from 1 to 3. A column in a data set might be tagged "business_1" and give access only to users with a business_1 tag on their account. Another column in the same dataset could be tagged "business_3," offering access to users with any of the three access tags. Data is presented to the user with filters and hashes to obscure data that the user has not been cleared to see.

Attributes like these can be changed, added, or removed immediately as business needs and user access change, either from the DataOS GUI or with a few lines of YAML code through a command-line interface. Attributes can also control data retention and many other governance functions quickly and simply.

Virtual Data

DataOS eliminates the problem of data duplication and siloed data sets because users never have to move data out of the datastores in which they reside. The data returned by a query is automatically presented to the user as a virtual dataset, which can be used by analytics and BI tools just as any other dataset would be.

Data is never duplicated or siloed. Each search presents the data as it exists in the most current versions of the source datasets.

DataOS delivers what other platforms only promise: simple, transparent, genuine data democratization. It lets you focus on the cultural changes needed for the new paradigm without worrying about your data management platform sabotaging your efforts.

To find out more about how DataOS can transform your data ecosystem, contact us for a demo today.

[Schedule a demo →](#)

BY PAUL SCOTT