

## REPORT REPRINT

# DataOps: a passing buzzword, or the engine of the data-driven enterprise?

**MATT ASLETT**

**14 AUG 2018**

While mainstream adoption and understanding of the term DataOps is still in its infancy, we are seeing wider adoption of the underlying trends as enterprises move toward a more agile and automated approach to data management.

---

THIS REPORT, LICENSED TO DELPHIX, DEVELOPED AND AS PROVIDED BY 451 RESEARCH, LLC, WAS PUBLISHED AS PART OF OUR SYNDICATED MARKET INSIGHT SUBSCRIPTION SERVICE. IT SHALL BE OWNED IN ITS ENTIRETY BY 451 RESEARCH, LLC. THIS REPORT IS SOLELY INTENDED FOR USE BY THE RECIPIENT AND MAY NOT BE REPRODUCED OR RE-POSTED, IN WHOLE OR IN PART, BY THE RECIPIENT WITHOUT EXPRESS PERMISSION FROM 451 RESEARCH.



©2018 451 Research, LLC | [WWW.451RESEARCH.COM](http://WWW.451RESEARCH.COM)

DataOps is the latest data management buzzword, and while mainstream adoption and understanding of the term is still in its infancy, we are seeing wider adoption of the underlying trends as enterprises move toward a more agile and automated approach to data management.

---

## THE 451 TAKE

Traditionally, data management and data governance have been considered functions in their own right, and as such have had the potential to become isolated from business goals and even seen as barriers to the delivery of value. In comparison, DataOps highlights that data management and data governance are more accurately viewed as enablers of realizing business value through more agile data processing, analytics projects, or compliance initiatives. While the term DataOps itself has not yet become mainstream, it is clear that the associated concepts and technologies are being widely adopted by enterprises as they seek to become more data-driven. We expect this trend to continue, and the term DataOps to become more widely adopted as a result.

---

DevOps, DevSecOps, AIOps, MLOps, DataOps: Just as any mildly controversial political incident since the early 1970s is instantly branded with the suffix ‘-gate,’ ‘-Ops’ has become a widely used suffix in the IT industry in recent years related to attempts to bring agility and automation to existing IT management practices.

One of the more recent is DataOps, a term coined by Tamr cofounder and CEO Andy Palmer in 2015 to describe the need for more agile approaches to data management required for enterprises to realize the potential benefits of becoming more data-driven amid a growing volume of data, as well as the increased number of data sources and use cases. The use of the term DataOps has grown since then, but has yet to enter the mainstream IT vocabulary. As such, it might be easily dismissed as a passing buzzword. However, further investigation highlights that many of the key concepts behind DataOps are being widely adopted, even if the term itself is currently alien to most mainstream enterprises.

## WHAT IS DATAOPS?

DataOps is about delivering more agile and automated approaches to data management: enabling self-service access to data to accelerate the development of database-driven applications and data-driven decision-making, as well as support business agility in response to rapidly changing business requirements. It is designed to reduce data friction, which arises when the demands of data consumers (such as data analysts, developers and senior decision-makers) are not met by data operators (e.g., data management and IT professionals). This is a perennial problem that has been exacerbated in recent years by the growing volume of data and increased demand for data and analytics projects.

IT has traditionally been the gatekeeper to analytics insight – creating reports, dashboards and data models and provisioning data warehouses and associated infrastructure in response to demand from business users. This traditional approach is entirely unsuitable for driving real-time decision-making on large volumes of live data.

In contrast, we have witnessed a shift toward self-service enablement that bypasses the need to depend on IT via the democratization of data and associated infrastructure. This encompasses several key trends that have increased the influence of data consumers over the purchasing and use of data-related technologies, including but not limited to:

- Self-service analytics
- Self-service data preparation
- Self-service infrastructure provisioning (i.e., cloud computing)
- Self-service data transformation (i.e., ETL as a service)
- Self-service data pipeline orchestration

In addition, our research in recent years has highlighted changing attitudes toward data governance and data management, such as:

- Data governance as an enabler of self-service
- Collaborative data governance
- Automation of data management best practices via machine learning

We have also highlighted the rise of the data bazaar (or data marketplace) as an enterprise environment that enables self-service access to data in the context of regulatory or security requirements, addressing, among other things:

- Managed self-service data preparation and analytics
- Policy-based data access, data lineage and data management

These shifts are in part related to the rise of DevOps and the role of developers – as well as data scientists and data engineers – in determining data access and usage requirements, as well as influencing the choice of more agile data management products and services. As such, we are noting the emergence of a growing list of vendors targeting data consumers (such as data analysts and senior decision-makers), rather than data operators (e.g., data management and IT professionals).

In addition to DevOps, DataOps incorporates lessons learned from agile development, continuous development and delivery, and lean manufacturing (such as statistical process controls). It is also related to the need for new cultural and organizational approaches to data management associated with the shift toward being more data-driven. This includes the formation of cross-functional collaborative teams that combine data scientists, data engineers, data analysts and data operators.

It should be clear from the above description that DataOps is defined by new approaches to data management, rather than specific data management technologies. However, key vendors specifically addressing these trends while also driving the use of the term DataOps include:

- DataKitchen
- Delphix
- Nexla
- Tamr
- StreamSets

## MLOPS AND DATA SCIENCE MANAGEMENT

In attempting to become more data driven, many organizations are investing in machine learning (ML) tools and developing ML-driven applications. The success of these projects depends on the ability of the organization to operationalize experimental data science projects through model deployment and management.

MLOps is an associated term that has also emerged in recent years that describes the operationalization of machine learning. While MLOps has its own unique characteristics and challenges, it can be considered part of the overall DataOps movement.

Key emerging vendors in the field of machine learning operations and data science management, which we recently added to our Data Management and Analytics Market Map, include:

- Metis Machine
- ParallelM
- Hydrosphere.io
- Seldon
- Datatron

Also highly relevant is the move toward API-driven integration and microservices, which enables more flexible access to both internal and external data sources, as well as the move to eradicate traditional data silos via the use of abstracted semantic data layers and enterprise knowledge graphs.

## COUNTERPOINT

451 Research has been writing and talking about all of these topics in recent years, although without specific reference to the term DataOps. It is easy to be skeptical – and indeed we have been.

One definition of DataOps we have seen is that it is designed to ‘deliver data to the person, system or application that can turn it into business value.’ One might reasonably ask what it is people think data management staff have been doing for the past 50 years, if not precisely that. Additionally, just as the skeptical could point out that many statistical analysts have rebranded themselves as data scientists in recent years, so it could be argued that DataOps is rebranded data management carried out by data management professionals rebranded as data engineers and data operators.

While it is undeniable that there is an element of rebranding going on and that the functionality on offer from many of the vendors above overlaps with that of established incumbents, it is also true to say that there are real changes in the attitudes to data management that are encapsulated and reflected in DataOps. And as noted, while the growth of DataOps is driven by a specific group of products and vendors, it is also true to say that DataOps is very much a concept or philosophy that is equally enabled by organizational and cultural approaches rather than specific technologies.