DataOps – Towards a Definition

Julian Ereth

University of Stuttgart, Stuttgart, Germany

Abstract. Organizations seek to streamline their data and analytics structures in order to meet increasingly demanding business requirements. This can be difficult due to complex and fast-moving data landscapes. DataOps promises a remedy by combining an integrated and process-oriented perspective on data with automation and methods from agile software engineering, like DevOps, to improve quality, speed, and collaboration and promote a culture of continuous improvement. The goal of this on-going research is to elaborate DataOps as a new discipline. For this, it explores the body of knowledge and presents a working definition of DataOps as well as an initial research framework based on an explorative literature review and eight interviews with industry experts.

Keywords: DataOps, Analytics, Agile, DevOps

1 Introduction

Data is a key asset to compete in today's business. Data-driven decision making significantly increases business success [1] and data is essential for many business processes or even entire business models [1, 2]. Consequently, companies seek to streamline their data and analytics processes in order to make them more efficient, provide data faster and in superior quality, and ensure a stable and reliable operation in general. This, however, can be hard due to fractured data landscapes with heterogeneous tools and technologies, a broad scope with various stakeholders, rapidly changing business requirements, and a general lack of standards [3, 4]. To revise these insufficient and inefficient structures, there is a need to shape enterprise data and analytics in a way that enables a stable operation and increases speed, quality, and overall productivity. Discussions on how to achieve this, often includes topics like agile methods, data governance concepts, or the use of automation. Here, many see similarities to challenges in software engineering, where DevOps and continuous integration were introduced to provide high-quality software at an every-increasing pace [5]. However, data analytics is different to software engineering, and consequently the new term DataOps emerged [6].

The goal of this ongoing research is to academically elaborate DataOps as a new discipline. This paper explores the body of knowledge and presents a working definition of DataOps and an initial research framework. We firstly depict related topics and our methodical approach. Then, we discuss first results from a qualitative exploration and derive a working definition and an initial research framework. The paper finally concludes with an outlook for future work.

2 Related Work

The field of DataOps goes hand in hand with a continuous professionalization of data and analytics processes in companies. It combines ideas from information systems research with different areas like agile and lean thinking and modern software engineering [7]. For one thing, DataOps is taking its cue from DevOps which is "an organizational approach that stresses empathy and cross-functional collaboration within and between teams" [5] in order to accelerate delivery of changes and increase quality of software [5, 8]. For this, DevOps promotes integrated and highly-automated engineering pipelines to bridge the gap between development and IT operations. This can be challenging, since development and operations are historically separate tasks with conflicting goals. Software engineers need to quickly respond to changing requirements and seek to rapidly deploy new features. The operations team, in contrast, is interested in providing stable and reliable services and infrastructure and therefore avoids risks and works as predictably as possible [8]. This is why DevOps requires a mind-shift and strives for change in company culture [5, 8]. As DevOps is a rather novel concept, related research is mainly focusing on exploring ways of adaption [9, 10] and its business value [10, 11]. Next to DevOps, other organizational and technological approaches from software engineering like, behavior-driven development [12] or scrum [13], as well as the general philosophy of the agile manifesto [14] appear in DataOps discussions.

From an information systems research perspective, DataOps fits into the continuous stream of work about agility in business intelligence (BI) and data science. Here, agility is seen as "the ability to react to unforeseen or volatile requirements regarding the functionality or the content of a BI solution" [15]. This encompasses the transformation of processes with agile methods like Scrum or Kanban, as well as the adaption of new technologies and architectural concepts to provide flexibility and increase value [3, 16, 17]. There is also an overlap with research about BI and analytics maturity models that tries to assess the level of development of organizational capabilities and resources in organizations [18]. Here, the associated goals are usually to reveal necessary steps to advance BI and analytics in order to increase productivity and business value [19]. Findings in this area shows that mature analytics solutions involve a high-level of collaboration, strategic alignment, and an enterprise-wide integration [20, 21], what matches the goals of DataOps.

3 Methods

In order to examine the field of DataOps, we followed an exploratory research design outlined in Fig. 1. First, we conducted an exploratory literature review for an initial exploration of the body of knowledge and to provide a better general understanding of terms and concepts [22].

Fig. 1. Research design

For this, we followed Levy and Ellis [23] and queried four scientific databases, namely SpringerLink, ACM, AISeL and IEEE, as they cover numerous important journals and conferences in the information systems domain. In the search we used the keyword *DataOps* as well as a combination of the terms *DevOps* and *Data*. We manually screened the results and filtered out articles that seemed to have no relevancy, e.g. focus on another domain or too specific topics. This process resulted in 783 documents of which 243 were selected relevant and were further investigated by screening the abstract and full text. It turned out that only 6 documents contained DataOps in its title or abstract. The temporal distribution of the publications confirmed the novelty and relevancy of the topic, as the first document is dated at 2012 and since then the number of publications increased continuously (c.f. Fig. 2). In addition to the academic literature, we reviewed practitioner literatures, like blog posts or white papers of industry analysts, in order to investigate the understanding of related terms and concepts in practice and increase relevance for practitioners [24].
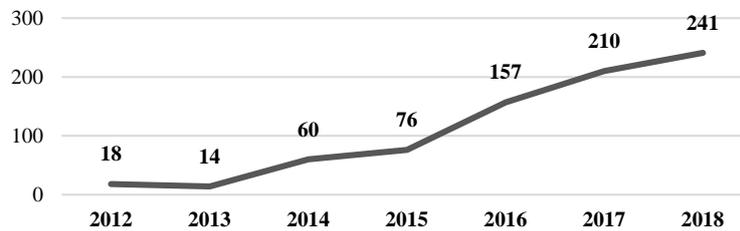

Fig. 2. Temporal distribution of publications

For a qualitative exploration, we then interviewed selected industry experts to gain deeper insights about the scope of DataOps. To select relevant experts, we searched for companies that offer DataOps services (vendors) or use DataOps in practice (users). We conducted eight interviews with international experts from different companies (c.f. Table 1). Most companies were vendors of which five can be denoted as startups (< 10 years or < 50 employees) and three can be assigned to the enterprise level. The interviews were conducted as semi-structured discussions of 30 – 60 minutes by phone. The experts were asked about their understanding of DataOps, how their services fit into this area and about goals and principles related to DataOps. The interviews were transcribed and coded to identify more general components of DataOps. For this, we first labeled statements (e.g. "we allow unit-tests in databases" as *technical tests*) and then grouped similar labels (e.g. *technical tests* and *acceptance tests* as *testing*).

4

Table 1. Background of the interviewed industry experts

Company	Category	Type	Field of Action/Product
Company C1	Vendor	Startup	Data Warehouse Automation
Company C2	Vendor	Startup	DataOps Platform
Company C3	Vendor	Startup	Big Data Automation
Company C4	User	Enterprise	Semiconductor Manufacturer
Company C5	Vendor	Enterprise	Self Service BI
Company C6	Vendor	Enterprise	Data Warehouse Automation
Company C7	Vendor	Startup	Data Operation Platform
Company C8	Vendor	Startup	DataOps Platform

The results of the literature review and the insights of the interviews were then triangulated to (i) formulate an initial working definition for the term DataOps and (ii) to derive a research framework for further work. Moreover, the derived working definition was discussed with two experts of the preceding interviews to test its relevance for practice.

4 Results

The exploration of the domain shows, that there is no general accepted definition of DataOps yet. The term itself was introduced by Andy Palmer who described DataOps as a discipline that "addresses the needs of data professionals on the modern internet and inside the modern enterprise" [6]. Since then, there are different understandings of the term within various scopes, like DataOps is "a hub for collecting and distributing data" [25], "the function within an organization that controls the data journey from source to value" [26], or "a new way of managing data that promotes communication between, and integration of, formerly siloed data, teams, and systems" [7]. Similarly, the conducted interviews also included various understandings like "all activities between the data and operation teams" or "an integrated perspective over the entire data lifecycle".

When describing these vague understandings, most authors emphasize continuous improvement and a culture of collaboration and trust as the underlying goal of DataOps [7, 27, 28] and elaborate their definitions with a set of goals, principles, and components. Moreover, many mentioned the empowerment of citizen-users and an end-to-end thinking as core objectives next to enhanced speed and quality [6, 28, 29]. When it comes to implementation, some use the term data pipeline to describe a process-oriented structure where data is transferred through multiple stages (e.g. extracted, transformed, and visualized) [26, 28]. This concept is often used to support orchestration and automation in complex scenarios. In this context, some advocate the idea that every artifact (e.g. data models or visualizations) can be represented as code (analytics as code) [27]. From an agile perspective, DataOps adopts the strive for short-cycles and incremental change. One interview summarized the DataOps with the goal to "become as fast as a startup, while being as robust as a manufactory". In addition to that, the examined literature and the interviews revealed data-driven improvements, reuse of artifacts, and testing and monitoring as key principles of DataOps [7, 27, 30].

Table 2. A list of DataOps components and the emphasize of the interviewed experts

	C1	C2	C3	C4	C5	C6	C7	C8
Goals:								
Continuous improvement	x	x	x	x	x		x	
Orchestration		x				x	x	x
Empowerment of citizen users		x			x		x	
Agility & speed	x	x	x	x	x	x	x	x
Collaboration & trust		x		x	x		x	
Principles:								
Reuse of artifacts	x	x						x
Automation	x	x	x	x	x	x	x	x
Integrated end-to-end thinking		x	x			x	x	x
Short cycles & incremental change	x	x	x			x		
Analytics as code	x	x	x			x		
Testing	x	x		x		x	x	x
Monitoring	x	x		x	x	x	x	x
Data-driven improvement	x	x				x		
Process-oriented Data pipelines		x	x	x			x	x

Table 2 shows a consolidated list of DataOps goals and principles that were extracted from literature. It also shows emphases of the interviewed experts. Here, it is apparent that different experts have different focus areas and only C2 seems to have a comprehensive DataOps understanding that spans all identified points. Moreover, the results confirm continuous improvement and the strive for agility and speed as valid descriptions of basic DataOps goals and identify automation as a common principle.

5 Working Definition and Research Framework

The results indicate, that DataOps is rather a collection of various practices and technologies, than a particular method or tool. Based on our initial findings, we suggest the following working definition of the term DataOps:

DataOps is a set of practices, processes and technologies that combines an integrated and process-oriented perspective on data with automation and methods from agile software engineering to improve quality, speed, and collaboration and promote a culture of continuous improvement.

This definition is intended to be a starting point for a common understanding of the term DataOps and an early limitation of its scope. It does not claim to be final and should be subject of discussion in future work.

Next to the working definition, we propose the research framework in Fig. 3 to further explore the DataOps space. This framework differentiates between the exploration of *DataOps as a discipline*, which includes methods, technologies and concrete implementations, and the investigation of the *business value of DataOps*.
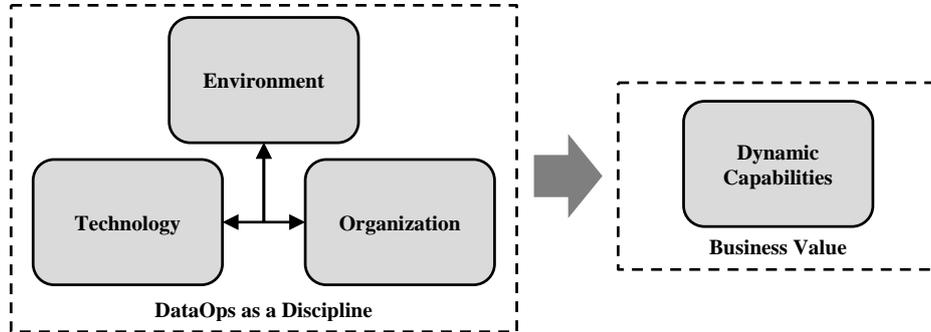
Fig. 3. Initial Research Framework

DataOps as a discipline can be further broken down by the means of the Technology-Organization-Environment (TOE) Framework [31] that helps to explain the relation between technology, organization and external factors in the diffusion of innovations. This framework brings the necessary flexibility to cover the wide scope of DataOps that spans from technological advancements, like automated data testing and continuous deployments, to organizational initiatives, like end-to-end collaboration throughout various business functions. Moreover, the TOE-framework has already been successfully used in similar areas, e.g. cloud computing [32] or big data [33]. Regarding the examination of the business value of DataOps, the concept of dynamic capabilities [34, 35] seems to be a valid candidate for a theoretical foundation, as it focuses "the ability to integrate, build, and reconfigure internal and external competencies to address rapidly-changing environments" [34] and it was already used in prior work to explain the business value of data and analytics [1, 36, 37] or DevOps [38]. This research framework is intended as an initial structure for future research. The prosed methods and theoretical foundations need to be validated and refined in upcoming work.

6   Conclusion and Outlook

This research illustrated the broad character of DataOps and showed that it is not a particular method or tool, but rather a collection of principles and a way of doing things on a cultural, organizational, and technological level. With its exploratory character, the research contributes to the body of knowledge by academically defining DataOps and arranging it in the field of information system research. Moreover, the research can help organizations to understand and adopt DataOps in practice.

DataOps is a rather novel term and there is only little experience and many research challenges. The proposed framework provides a starting point for future work that should, first, elaborate DataOps as a discipline, e.g. by developing blueprints for roles, processes and governance, and further explore related technologies, and, second, investigate the business value proposition of DataOps. As the diffusion of DataOps is still low, we suggest to conduct in-depth case studies to compare traditional approaches with DataOps-like implementations in order to gain further insights and refine this work.

References

1. Seddon, P.B., Constantinidis, D., Dod, H.: How does business analytics contribute to business value?. Information Systems Journal 4(3), 3380–3396 (2012).
2. Wamba, S.F., Akter, S., Edwards, A., Chopin, G., Gnanzou, D.: How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics 165, 234–246 (2015).
3. Baars, H., Ereth, J.: From Data Warehouses to Analytical Atoms – The Internet of Things as a Centrifugal Force in Business Intelligence and Analytics. In 24th European Conference on Information Systems (ECIS) Istanbul, Turkey (2016).
4. LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big data, analytics and the path from insights to value. MIT sloan management review 52(2), 21–32 (2011).
5. Dyck, A., Penners, R., Lichter, H.. Towards Definitions for Release Engineering and DevOps. In 2015 IEEE/ACM 3rd International Workshop on Release Engineering, Florence, p. 3. (2015).
6. Palmer, A.: From DevOps to DataOps, https://www.tamr.com/from-devops-to-dataops-by-andy-palmer/, last accessed 2018/04/21 (2015).
7. Thusoo, A., Sen Sarma, J.: Creating a Data-Driven Enterprise with DataOps - Insights from Facebook, Uber, LinkedIn, Twitter, and eBay. O'Reilly Media Inc., Sebastopol, CA (2017).
8. Hüttermann, M. DevOps for developers. Apress (2012).
9. Wiedemann, A.: IT Governance Mechanisms for DevOps Teams - How Incumbent Companies Achieve Competitive Advantages. In Proceedings of the 51st Hawaii International Conference on System Sciences (2018).
10. König, L., Steffens, A.: Towards a Quality Model for DevOps. In Continuous Software Engineering & Full-scale Software Engineering, p. 37 (2018).
11. Mishra, A., Garbajosa, J., Wang, X., Bosch, J., Abrahamsson, P.: Future directions in Agile research: Alignment and divergence between research and practice. Journal of Software: Evolution and Process 29(6), p. e1884 (2017).
12. Wynne, M., Hellesoy, A., Tooke, S.: The cucumber book: behaviour-driven development for testers and developers. Pragmatic Bookshelf (2017).
13. Schwaber, K., Beedle, M.: Agile software development with Scrum. Prentice Hall, Upper Saddle River (2002).
14. Beck, K., et al.: Manifesto for agile software development. http://agilemanifesto.org/, last accessed 2018/07/29 (2001).
15. Zimmer, M., Baars, H., Kemper, H.-G.: The impact of agility requirements on business intelligence architectures. In: Proceedings of the 45th Hawaii International Conference on System Science (HICSS), pp. 4189–4198. IEEE (2012).
16. Krawatzeck, R., Dinter, B.: Agile Business Intelligence: Collection and Classification of Agile Business Intelligence Actions by Means of a Catalog and a Selection Guide. Information Systems Management 32(3), 177–191 (2015).
17. Larson, D., Chang, V.: A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management 36(5), 700–710 (2016).

8

18. Cosic, R., Shanks G., Maynard, S.: Towards a business analytics capability maturity model. In: Proceedings of the 23rd Australasian Conference on Information Systems. Geelong, Victoria (2012).

19. LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big: The new path to value. MIT Sloan Management Review 52(1), 1–25 (2010).

20. Lismont, J., Vanthienen, J., Baesens, B., Lemahieu, B.: Defining analytics maturity indicators: A survey approach. International Journal of Information Management 37(3), 114–124 (2017).

21. Dinter, B.: The maturing of a business intelligence maturity model. In: Proceedings of the 37th American Conference on Information Systems (2012).

22. Hart, C.: Doing a Literature Review: Releasing the Research Imagination. Sage (2018).

23. Levy, Y., Ellis, T.J.: A systems approach to conduct an effective literature review in support of information systems research. Informing Science 9, 181-212 (2006).

24. Marrone, M., Hammerle, M., An Integrated Literature Review: Establishing Relevance for Practitioners. In: Proceedings of the 37th International Conference on Information Systems (ICIS), Atlanta, Georgia (2016).

25. Gartner Inc: IT Glossary - Data Ops, https://www.gartner.com/it-glossary/data-ops, last accessed 2018/04/19.

26. Euston, J.: What is DataOps?, https://www.nexla.com/define-dataops, last accessed 2018/04/20 (2017).

27. The DataOps Manifesto, http://dataopsmanifesto.org/, last accessed 2018/05/28.

28. Ereth, J.: DevOps in Business Intelligence and Analytics, https://www.eckerson.com/articles/devops-in-business-intelligence-and-analytics, last accessed 2018/04/16 (2017).

29. Rouse, M.: DataOps (data operations), https://searchdatamanagement.techtarget.com/definition/DataOps, last accessed 2018/05/25.

30. DataOps in Seven Steps, https://medium.com/data-ops/dataops-in-7-steps-f72ff2b37812, last accessed 2018/05/15 (2017).

31. DePietro, R., Wiarda, E., Fleischer, M.: The context for change: Organization, technology and environment, In: Tornatzky, L.G., Fleischer, M. (eds.) The processes of technological innovation, Lexington Books: Lexington, MA. p. 151-175 (1990).

32. Low, C., Chen, Y., Wu, M.: Understanding the determinants of cloud computing adoption. Industrial management & data systems 111(7), pp. 1006-1023 (2011).

33. Kang, D., Kim, S.H.: Process of big data analysis adoption: Defining big data as a new IS innovation and examining factors affecting the process. In: Proceedings of the 48th Hawaii International Conference on System Science (HICSS), pp. 4792-4801. IEEE (2015).

34. Teece, D.J., Pisano, G. Shuen, A.: Dynamic capabilities and strategic management. Strategic management journal 18(7), pp. 509-533 (1997).

35. Eisenhardt, K.M., Martin, J.A.: Dynamic capabilities: what are they?. Strategic management journal, 21(10-11), pp. 1105-1121 (2000).

36. Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., Childe, S. J.: Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research 70, pp. 356-365 (2017).

37.   Knabke, T., Olbrich, S.: Capabilities To Achieve Business Intelligence Agility –
       Research Model And Tentative Results (Research-in-progress). In: Proceedings of the
       20th Pacific Asia Conference on Information Systems (PACIS), p. 35 (2016).
38.   Forsgren, N., Humble, J.: The role of continuous delivery in IT and organizational
       performance. In: Proceedings of the Western Decision Sciences Institute (WDSI), Las
       Vegas, NV (2016).