10-9-2023

DataOps as a Prerequisite for the Next Level of Self-Service Analytics – Balancing User Agency and Central Control

Henning Baars
Universität Stuttgart, Germany, henning.baars@bwi.uni-stuttgart.de

DataOps as a Prerequisite for the Next Level of Self-Service Analytics – Balancing User Agency and Central Control

Research Paper

Henning Baars[1]

[1] Universität Stuttgart, Chair of Information Systems I, Stuttgart, Germany
henning.baars@bwi.uni-stuttgart.de

Abstract. The area of Business Intelligence and Analytics (BIA) has repeatedly oscillated between more central, efficiency-oriented, professionalized approaches and decentral, agility-oriented, user-driven ones. This paper investigates whether and how to alleviate that tradeoff by combining an agility-oriented self-service analytics approach with the professionalization-driven DataOps concept: DataOps aims at transferring ideas from DevOps to the realm of analytics, namely a mutual integration of Development and Operations and a high degree of professionalization and automation. A case study with a series of interviews and a workshop led to insights into the viability of such a combination. These inspire a theoretical concept for capturing the economics behind the approaches that is considering the (opportunity) costs of the components "user agency" and "central control". The concept has been evaluated with representatives from the case study. Based on the results, I argue that the discussed combination can push BIA solutions towards fine-tuned federated environments.

Keywords: DataOps, self-service analytics, DevOps, analytical capabilities.

1 Introduction

The 2010s have radically transformed corporate data landscapes and therefore approaches to Business Intelligence and Analytics (BIA) (Marjanovic & Dinter, 2018): The decade was marked by the entrance of Big Data technologies (Chen & Zhang, 2014), the emergence of the data lake as a new type of analytical data repository (Fang, 2015), the rise of cloud-based BIA platforms (Thomson & van der Walt, 2010), as well as significant progress in the field of machine and deep learning (Janiesch et al., 2021). These developments have shifted the attention away from centrally managed, curated data warehouse-(DW-)based Business Intelligence (BI) infrastructures towards distributed, more flexible, and often cloud-enabled data management approaches (Baars et al., 2014). The new concepts and technologies come with the promise of a "democratization" of data analytics with empowered end users (Lefebvre et al., 2021) and a federated "mesh" of data resources (Dehghani, 2022). A cornerstone here is *self-service analytics*, the idea to equip end users with user-friendly tools to autonomously gather, transform, and analyze data (Michalczyk et al., 2020). In its most far-reaching forms, this allows users to join heterogeneous data sources and even build advanced analytical models (Alpar & Schulz, 2016), thereby maximizing *user agency*. This is seemingly at odds with older recommendations for centralized and tightly governed BIA solutions to achieve professionalization and efficiency – approaches that stress *central control*.

Since the second half of the 2010s, centrality regained traction, e. g. with the introduction of structured data lakes (LaPlante & Sharma, 2018), data lakehouses (Oreščanin & Hlupić, 2021), the increased role of data catalogs for central metadata management (Labadie et al., 2020), as well as a general resurgence of governance aspects (Fadler & Legner, 2021). It needs to be emphasized that similar swings between central and decentral approaches have taken place since the 1990s under various names.

This paper investigates the possibility of complementing self-service analytics with the novel concept of "DataOps" to alleviate the tradeoff between central control and user agency and move beyond a fruitless back-and-forth. *DataOps* translates the ideas of "DevOps" to the analytics world (Ereth, 2018). DevOps is an established approach to IT systems development that combines system development ("Dev") with system operations ("Ops") in an agile and software-supported fashion. It is primarily driven by a rationale of professionalization and is heavily geared towards software support and an automation of the development processes (Ebert et al., 2016).

For our study the discussed issues are condensed into the following research questions:

RQ1: How can DataOps complement a self-service analytics approach?

RQ2: What impact does a combination of DataOps and self-service analytics have on the tradeoff between user agency and central control and how can this relationship be captured economically?

The findings contribute to research by furthering the body of knowledge on BIA by addressing new options for the organization of BIA (RQ1). Additionally, the derived economics-driven rationale of RQ2 can act as a new conceptual tool for capturing and evaluating combinations like the one discussed.

As DataOps is a rather recent approach and its combination with self-service analytics is particularly novel, this study pursues a qualitative and explorative approach that aims at hypothesis generation rather than hypothesis testing (Miles et al., 2014). More concretely, the research is based on an interpretative (single) research case study (Yin, 2018; Benbasat et al., 1987; Walsham, 1995) at a mid-sized automotive original equipment manufacturer (OEM). The approach was explorative and qualitative, as I strived for hypothesis generation rather than hypothesis testing.

The course of the paper is as follows: After a short overview on foundational concepts, section 2 presents a literature review for the identification of related work. Section 3 gives an overview on the applied methodology and the core findings regarding RQ1. The results motivate the presented concept that abstracts the cost rationales behind self-service and DataOps (RQ2). It was evaluated together with key case participants (section 4). The paper concludes with a discussion and an outlook (section 5).

2 Conceptual Foundation and State of the Research

The term *(data) analytics* is used rather broadly and refers to all kinds of activities relating to the collection, transformation, and analysis of data, as well as their use in reports and analysis models (Davenport & Harris, 2007). *Analytics* is now often combined with *BI* to ***B**usiness **I**ntelligence and **A**nalytics* (BIA) in order to stress that the subject in discussion goes beyond basic reporting-driven solutions and to highlight that it subsumes all components of an integrated approach to management and decision support (Lim et al., 2013; Sharda et al., 2014, pp. 18-21; Marjanovic & Dinter, 2018). Economically, BIA solutions are understood to provide organizations with a certain type of dynamic capabilities, namely “analytics capabilities” (Seddon et al., 2012). An *analytics capability* indirectly contributes to business value by enabling data-related activities for the provision of insights with a defined information quality for a certain type of business task that is itself embedded in a technological and organizational environment (Ereth & Baars, 2020).

An established approach for an efficient and consistent provision of analytical capabilities is to combine centralized data stores (esp. DWs) with defined governance structures for BIA-related roles and responsibilities, processes, solution portfolios, as well as for rules and regulations. Such solutions stress *central control*. Their main downside is a potential lack of *BIA agility*, i.e., the ability to react to unforeseen requirements swiftly and efficiently (Baars & Zimmer, 2013). While an arsenal of remedies has been proposed to tackle this issue (Krawatzeck & Dinter, 2015), a core ingredient for BIA agility was and is the inclusion of end users into the design and the implementation of BIA applications as it is done in self-service analytics. Current self-service approaches allow users to design data extraction, transformation, and analysis routines (Alpar & Schulz, 2016) – and potentially even to develop advanced machine learning (ML) models (“self-service ML”) (Jacquin et al., 2020; Ooms & Spruit, 2020). On the solution side, this means to grant the user *agency* over large parts of the analytics process.

“DataOps” appears to bridge fully uncoordinated individual solutions (with maximal user agency) and inflexible “classic” central DW concepts (with maximal central control). In essence, *DataOps* transfers the ideas from the software development approach of DevOps to the realm of data analytics (Ereth, 2018): DevOps aims at integrating **D**evelopment and **O**perations. It emphasizes the complete software lifecycle, a customer-oriented perspective, a transparent cooperation of the responsible units, and a culture of cooperation (Sánchez-Gordón & Colomo-Palacios, 2018). This is also supported by defined “pipelines” for created artifacts. DevOps also includes highly professionalized and tool-supported processes for software testing, for the integration of software artifacts into the system landscape (*continuous integration*), for their deployment (*continuous deployment*), as well as for the version management for its infrastructures (*infrastructure as code*) (Ebert et al., 2016).

The term “DataOps” still lacks some clarity (Munappy et al., 2020; Ereth, 2018) and this is exacerbated in self-service environments, where the end user is the developer of the report or the model (resp. of parts of it). But while DevOps cannot be applied in a one-to-one fashion to the realm of analytics, a tight cooperation between the units that are providing infrastructure, models, and reports (or templates therefore) and the users consuming them (or building up on them) is of central importance here as well.

To identify related work, a structured literature review (vom Brocke et al, 2009; Webster & Watson, 2002) was conducted with a search for peer-reviewed papers on the AIS library (112 hits, 3 relevant), IEEE-Explore (1 hit, 0 relevant), the ACM digital library (24 hits, 4 relevant), and the Web of Science (1 hit, redundant) on all available search fields:

- "Self-Service" AND ("analytics" OR "business intelligence" OR "machine learning" OR "artificial intelligence") AND ("DataOps" OR "MLOps" OR "DevOps"),
- "DataOps" AND "agility",
- "DataOps" AND "democratization" AND ("analytics" OR "machine learning" OR "artificial intelligence"), and
- ("DataOps" OR "MLOps") AND "user empowerment"

I filtered out research in progress papers, solely technical papers, papers that mentioned the search terms without really addressing them, papers that were not connected to the user side as well as papers that were not linked to the operations and support side (e. g. that were merely discussing the local development of a model).

The low number of relevant hits despite the rather wide net of search strings indicates that the subject has not yet been adequately addressed and thus forms a research gap. A closer look into the papers further bolsters that conclusion. While the papers acknowledge the interplay of self-service analytics and DataOps or related concepts, they do not delve further into this. Nevertheless, the identified literature provides additional relevant insights:

Relation to the subject of the Internet of Things: The paper "DataOps for Cyber-Physical Systems Governance" by Garriga et al. (2021) focusses on the development of a single ML solution for the prediction of passenger flows at an airport in a highly distributed environment under the utilization of selected DataOps concepts (esp. by building end-to-end data pipelines). The authors particularly demonstrate how DataOps techniques facilitate the exploitation of complex Internet-of-Things data sources.

Maturity levels of DataOps approaches: In their paper "from ad-hoc data analytics to DataOps", Munappy et al. (2020) develop an evolutionary model based on a case study at Ericsson: They distinguish between the steps "ad-hoc data analysis", "semi-automated data analysis", "agile data science", "continuous testing and monitoring" and "DataOps" (as the goal).

DataOps as a driver of agility: In "Scaling Enterprise Recommender Systems for Decentralization" van der Goes (2021) discusses an MLOps/DataOps based platform for building and operating recommender systems in the federated organizational environment of Heineken. They observed a higher agility of the distributed data science teams.

DataOps as a conduit for the digital transformation: Informed by a literature review, the authors Xu et al. (2021) link DataOps to aspects of digital transformation. Their paper argues that the components of DataOps bring stakeholders together in a dynamic and creative fashion and thus strengthen organizational performance.

Relevance of socioeconomical factors: Mucha et al. (2022) stress the importance of the socioeconomical dimension of DataOps/MLOps and derive a multi-cyclical development model that highlights non-technical components.
Relevance of a central support for self-service analytics: Both Namyslo & Baars (2022) and Iyer et al. (2021) discuss environments for self-service analytics and the prerequisites for successfully running them. While Namyslo and Baars discuss how end users can be empowered by self-service ML to conduct forecasts of warranty costs in the automotive industry, Iyer et al. present a solution for self-service analytics with complex geospatial data. Both illustrate the need to underpin self-service approaches with supporting tools and structures – on the technology (templates, feature, model catalogs etc.) and the organizational side (defined processes, support, quality assurance etc.). These tools and structures can be directly mapped to the core DataOps ideas.

3 Case Study: Methodology and Findings

The following section presents the case study and its immediate results which primarily feed into RQ1. They also lay the groundwork for answers to RQ2, which will be addressed by the concept design in the subsequent chapter 4.

3.1 Methodology

The company in discussion presented the opportunity to collect relevant and rich insights into the subject matter: The mid-sized German automotive OEM had already successfully adopted a self-service analytics approach, which is now used widely across all business units. After perusing the relevant documentation and reviewing relevant artifacts (esp. the transformation pipelines, analysis models and reports created by the end users), a series of qualitative, semi-structured interviews with participants from seven different business units was conducted. Methodologically, the interviews followed the recommendations of Meyers and Newman (2007).

The derivation of the interview questionnaires, as well as the coding were based on a conceptual framework (Ravitch and Riggan, 2012) that comprises of the following components: The business context of the self-service use case (including the background and IT/analytics competences of the users), the concrete analytical tasks tackled with self-service analytics, the state of automation and the DataOps potential in data preparation, data integration, testing, and data management (as potential parts of data pipelines), the cooperation with the central BIA unit, as well as outcomes and trends.

The seven interviews took on average 60 minutes and were conducted in January-April 2022 (each with two interviewers). The BIA unit identified and contacted the potential interviewees (either business or engineering experts, all self-service adopters) and – encouraged by the researcher – made sure that all relevant business units of the company were covered. All interviews were recorded and transcribed. The analysis of the transcripts followed the recommendations of Miles et al. (2014) and Mayring (2014) and was based on a stepwise category formation process (Mayring, 2014, pp. 79-88). After that, a workshop with the IT department was conducted with the chief information officer (CIO) of the company and members of the BIA unit responsible for data analysis, systems operations, and system development, as well as the BIA unit's leader (5 participants in total). In the workshop, I presented, discussed, and enriched the findings in an open discussion (responses were documented in field notes). The focus of the workshop were the processual and technical roles of the IT in the self-service approach, as well as their take regarding on an introduction of DataOps.

The automotive OEM had already implemented a sophisticated BIA architecture backed by a strong analytics team. According to the model of Munappy et al. (2020), I would place them in between the development steps "semi-automated data analysis" and "agile data science" with a set of state-of the art tools, professionalized processes, and already selected established pipelines in place.

The seven interviews were done in the business units and addressed the core use cases shown in Table 1.

Table 1. List of interviews

Business unit (BU)	Main use case for self-service analytics
BU1: Managerial control 1	Data preparation and analysis for sales controlling
BU2: Managerial control 2	Data preparation and managerial-control tasks as well as accounting-related analyses
BU3: After sales	Clustering and analysis of complaints
BU4: Research & development (R&D) for powertrain components	Statistical analysis of data from vehicles for the continuous improvement of vehicle design
BU5: Production	Constant monitoring of the quality of assembly processes and descriptive analysis of faults
BU6: Logistic	Inventory management analysis
BU7: Marketing and sales	Preparation and analysis of data on inventory and retail stock levels for different markets

BU1: For sales control, data from heterogenous systems from the areas of finance, retail, and marketing are provided in a specialized sales-oriented data mart by the central BIA unit. These data are already transformed and prepared (assisted by the business unit); data cleansing and integration are therefore subject to a high degree of central control. Report generation, however, rests completely in the hand of the users.

BU2: The topic of BU2 also revolves around managerial control. Here, indicators and reports are generated based on data from mostly file-based source systems. Unlike in BU1, the main challenges for the users lie in data extraction and transformation.

BU3: The data for the user of BU3 (after sales) primarily originate from a single system. Since the analyzed complaints are entered as free text, their analysis is not straightforward and data quality is a problem. The user applies cluster analyses techniques to group complaints. This scenario goes beyond simple descriptive statistics and requires the domain knowledge of the user and therefore high degrees of user agency.

BU4: Like in BU3, the relevant data in this environment are also extracted from a single system. The data structures in this case are both complex and heterogeneous (vehicles measurement data). In addition, data volumes are high, and each analysis is individual and relies on the specialized know-how of the user as a domain expert. This reduces both the potential to automate data analysis and the possible contribution of the central BIA unit beyond a technical support after the data provision.

BU5: In production, a variety of (often embedded) IT source systems from different suppliers are accessed, which leads to considerable challenges during data extraction and integration. Currently, each analysis is done by the users who prepare the source data in a mostly (semi-)manual fashion. The setting bears similarities to the ones presented by Garriga et al. (2021), that have already shown potential for DataOps.

BU6: All necessary data for inventory management come from a single and centrally managed SAP™ system with a guaranteed high degree of data quality. The main challenge in this use case is to access the necessary data from the system in time.

BU7: Marketing and sales have access to a DW provided by the IT department. Similar to BU1, business users develop their own dashboards. The user frequently incorporates additional data sources, esp. file-based ones, and crafts their own ETL-tasks.

3.2 Findings

It is noteworthy that the interview partners were all regular employees of the business units and by no means professional data analysts or data scientists – even those who conducted more complex data transformations (BU2, BU4, BU5, BU7) or advanced analytics (BU3, BU4). It was also striking that the users were not using any form of standardized process or relying on a fixed toolset. They employed heterogeneous software products such as MS Excel™, Tableau™, Tableau Prep™, KNIME™, specific tools for analyzing engineering data, as well as VBA or Python programming. Some of the users were working with isolated data from file systems or different transactional systems, while others were building their solutions on top of a centralized DW, data mart, or a Big Data cluster in the cloud.

Despite an established centralized DW, a Big Data cluster, and a standardized reporting platform, the required BIA capabilities cannot be fully covered by the BIA unit. The high degree of user autonomy is correlated to the variety of use cases that spans both the data and the analysis side. The self-service approach allows the company to react to the heterogeneity of the required analytical capabilities and gives the users as much autonomy as they demand, while keeping central control for infrastructure provision. The users translate the granted degrees of freedom into BIA agility.

Finding 1: *The need for heterogeneous BIA capabilities can be countered by a high degree of user autonomy and supported by a self-service analytics approach.*

The benefits of the self-service approach do not come without caveats: Two-thirds of the interviewees cited data availability and data provision as their biggest concerns. This is reflected in statements from the interviewees like *"Data can only be downloaded manually in the form of Excel files"* (BU6). *"Data are not always consistent across time and between different analyses, so every time I have to look for the right data"* (BU5). *"Data from different sources must first be integrated"* (BU7).

For the questioned users, a lack of automation in data provision seems to be the greatest pain point and binds a considerable amount of time and costs. The need to pour self-service resources into data preparation was also stated as a reason for the focus on descriptive analysis – despite the explicit wish to try more advanced ML techniques. Corroborating this is the fact that the use cases with more advanced models were also the ones with a single source system, namely BU3 and BU6. The following typical statement illustrates those points: *"I wish, I could concentrate on my business analytics tasks and would not have to spend so much time on data provision and data preparation".* Such issues came alongside requests for a more centrally managed data integration and storage pipeline and for increased support.

The issue of data preparation was also discussed during the workshop with the IT department: The BIA unit stated to have already begun to address it, e. g. by building special data marts with integrated data for specific business units (as in BU1), standardizing REST APIs, etc. These technical solutions were developed in close cooperation with the business units and incorporated ideas and routines that the users came up with in their self-service routines (esp. in BU2 and BU3). Incorporating DataOps would mean doubling down on this approach by automating data pipelines that are specified and maintained by a joint team of business users and IT. The aim would be an enhanced central control without sacrificing perceived user agency – maybe even enhancing it by facilitating additional analyses.

Finding 2: *A professionalized data provision (e. g., based on a DataOps approach) can support a more efficient BIA application and potentially free resources for enhanced or new BIA capabilities. This does not necessarily imply a reduction of user agency – as long as the processes remain adequately cooperative and agile.*

The BIA team also pointed out that the issues of data provision run deeper than just fixing the IT side of BIA. A problem was seen in the variety of suppliers of the data sources – some of which are often not even subsumed under the umbrella of IT, e. g. smart machines. As the IT is usually not involved in the selection and implementation of such components, aspects of data interfaces, data formats, and data quality are often neglected during the specification by the business units. The problems often only surface at the time of data analysis. A stricter data governance that promotes a "analytics by design" thinking could alleviate those issues. This would entail reduced user agency and stronger central control but facilitate self-service.

Finding 3: *A strict, analytics-oriented data governance that encompasses the complete data landscape of the organization can lay the groundwork for a self-service approach, particularly when it comes to potential source systems – which are ideally understood as stubs of future analytics pipelines. This entails curtailing the user agency of the units responsible for acquiring or developing those systems.*

Also problematic were data sources with manually or semi-manually collected data (BU1, BU2, BU3 and BU7), e. g. error reports in after sales (as in BU3). Additionally, an increased number of data sources also correlates with issues of consistency and completeness (BU4, BU5). Some sources also lack timeliness (BU1, BU2, BU7).

This is a major difference to a centralized approach with a strong governing unit at its core. In self-service analytics with individual solutions built upon individual data transformations, there is an inherent risk of insufficient oversight, of redundant developments of data transformation pipelines, and of a lack of power to enforce changes at the side of the source systems. This would be an anchor for a DataOps approach in which the developing user is working with the BIA unit that operates the infrastructure. The BIA unit could also be made responsible for the continuous integration of the user-developed solutions into the wider analytics environment of the organization.

Finding 4: *Efficiency gains can be unlocked if a unit with central control is overseeing, guiding, and leveraging the development of the users' data pipelines.*

The consensus in the interviews was that the documentation of data sources, data models and formats, data lineage, and analyses was inadequate. In the cases with one or two stable source systems, this does not pose a problem. In the other cases, however, it is seen as one, esp. because of the time needed to find the right data for a given analysis. Most users consider a central metadata management as necessary for self-service analytics. In the workshop, the BIA unit acknowledged this problem and stated that this was one of the reasons for the ongoing development of a company-wide data catalog. There remains some skepticism regarding the sustainability of a typical data catalog, as most of the important metadata are documented manually. This is thought to lead to a creeping loss of topicality of the metadata. Ideally, these aspects are integral parts of the design of data pipelines with high degrees of automation.

Finding 5: *Metadata management is a precondition for a comprehensive usage of self-service analytics in an organization – and should be part of a DataOps approach.*

Since automated testing plays a significant role in both the DevOps and the DataOps literature, it was also addressed it in the study – with rather meagre results: None of the BUs that were encountered in the study has implemented automated testing in their use cases. There were two reasons given for that: Firstly, most of the self-service analytics tasks are by definition ad-hoc and individual, which makes it difficult to systematically define test cases or implement automated testing routines. Secondly, many of the applied data sources are not suited, esp. the ones with manually entered data. The approach to quality assurance chosen by the users were expert assessments, which were stated to identify errors quickly and thoroughly.

Finding 6: *Due to the unique nature of the analyses, self-service analytics is rarely compatible with automated testing routines. However, to capture the spirit behind testing, workflows with expert assessments are a valid alternative.*

One point that came up in the workshop was that issues of compliance and data security require an oversight of the central BIA unit and form a natural limitation of user agency.

Finding 7: *Data Governance should be integrated into the DataOps process based on an oversight by a central unit.*

In the workshop, it was intensively discussed among the group of participants in how far the company would benefit from a combination of a DataOps approach with self-service analytics. The CIO in particular questioned if the stated problem and the presented solutions were novel at all. The final consensus of the group, however, was that the unprecedented breadth and the depth of the self-service approach was indeed new. On the one hand, this is fueled both by the competencies of the users and the user-friendliness of the self-service tools. On the other hand, it is driven by the massively increased awareness of both management and users for the business potential of BIA and of relevant use cases – even in units that have had relatively few contacts to the BIA domain before such as logistics, manufacturing, or development. The potential of the combination of self-service analytics with DataOps was seen as an option to further unlock the business potential of an enhanced user agency. Moreover, a stronger and professionalized cooperation between the IT/BIA unit and the end users for both infrastructure operation and the provision of central components (data marts, data catalog) was deemed critical – as was its role for overarching questions of governance.

4 Concept Design and Evaluation

A core theoretical insight I gleaned from the case study was how the nature of centrality of BIA is changed with the introduction of self-service analytics and DataOps. This particularly encompasses the aspects of *central control* and *user agency* as opposing poles. Economically, a reduced user-agency incurs opportunity costs resulting from insufficient alignment of BIA applications with the users' requirements and possibly a lack of acceptance. For example, it became clear that the highly domain-driven, complex solutions of BU3, BU4, or BU5 would incur prohibitive costs without the active involvement of the users. Decreased central control, however, comes at the costs of reduced efficiency and missed overarching benefits. In the case, this can be seen across the use cases in inefficiencies with data provision and transformation, and in the lack of metadata that led to tedious and repeated information seeking activities.

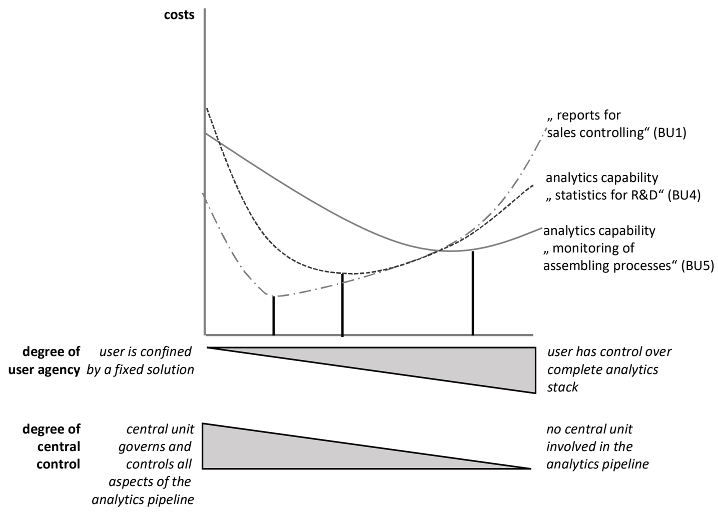

Figure 1. DataOps and centralization of analytics

Figure 1 captures these ideas for the examples found in BU1, BU4, and BU5 (that represent different types of cases with respect to the subject at hand): In each case, the costs of a purely central solution with minimal user agency and a maximum of central control can be lowered by the inclusion of users. However, a complete handover of the responsibility would also raise costs significantly as this would correspond with the typical inefficiencies of de-central analytics solutions (duplication of data provision tasks, tool variety etc.). This leads to the conclusion that the minimal costs are somewhere along the spectrum.

The BU1 case (reports for sales controlling) is a base case as it aligns well with established BI cases that have been around for decades. Here, the (limited) contribution of the self-service approach primarily comes from a slightly more agile and individual report design and some indicator calculations. A full-fledged self-service, that also includes data provision, bears the risk of tremendous redundancies.
Providing the analytics capability "monitoring of assembly processes" (BU5) however, heavily depends on the domain knowledge of the users. A full centralization would lead to significant coordination requirements. This moves the point of minimal costs to the right of the axes. In that case, inefficiencies became also visible, which would normally demand a partial recentralization of those tasks, esp. the provision and preparation of machine data. Similar, the complex statistics in R&D (BU4), requires the expert hand of the users. Note that the lines in all illustrations are so far only meant to be qualitative illustrations rather than measured costs and are in reality comprised of a variety of discrete cost components for different technical and organizational measures.

Figure 2 visualizes the potential effects of (successfully) implementing ideas from DataOps on those cost curves. For BU5, an active involvement of the central BIA unit with control over the central data provision pipeline (including the specification of data interfaces and standards for non-IT-components), as well as a stronger and professionalized automation of the data transformation, could lower the costs of the current self-service approach. The minimal cost point would move to the *left* towards the central unit. Contrarily, for BU4 (R&D), the point would move to the *right,* as a stronger support makes room for advanced analytics.

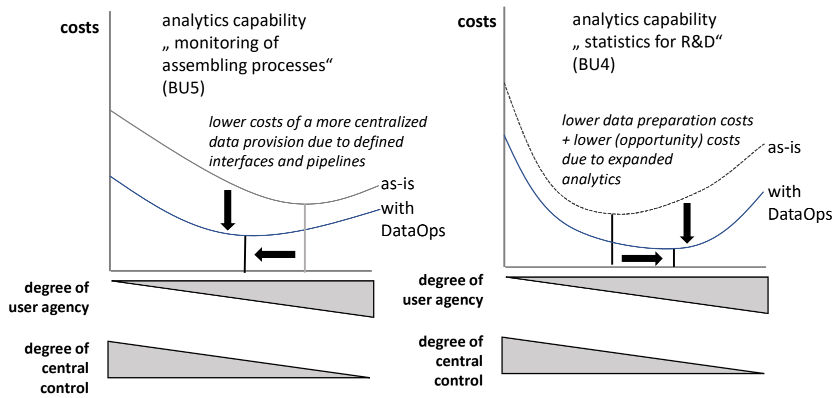

Figure 2. Potential of introducing DataOps to BU5 and BU4

These ideas were evaluated in two interviews in March 2023 with the head of the BIA unit and a representative of the R&D team (BU4), because the cost curve of that unit was the most atypical. We discussed the understandability and plausibility of the approach and of the selected cost curves, its comprehensiveness, its practical usefulness, as well as the need for further research.

The interviewees acknowledged all those criteria as fulfilled, apart from smaller corrections on the cost curves of BU4 (which became flatter). Both considered it relevant to include the competencies of the users as a relevant factor. The BIA head also stressed the factor time (esp. one-time reports vs. development of an operationalized model). As for practicability, the head of analytics plans to apply the concept for BIA portfolio management, in individual investment decisions, and for policy design. The company currently sets up on a follow-up study for further operationalizing the approach.

5 Discussion and Outlook

The results of the presented study highlight the increased awareness of BIA among users across a variety of domains, their acceptance of BIA solutions and their willingness to engage in a self-service approach. This is undeniably fostered by technical advances and can be expected to be unspecific for the given company. Here, however, the approach was additionally fostered by the BIA unit with its close ties to the user departments and its eye for innovations. The importance of such a unit was stated in all interviews. This is a first step towards DataOps, that will be implemented in the future.

The case also demonstrates the viability of a self-service analytics approach that is rolled out across heterogeneous business units. It highlights how a DataOps-based professionalization of data pipelines (including those designed by the users) and a tightly controlled data and metadata management can unlock further business potential, esp. by freeing resources. While some DataOps concepts are by definition rather hard to employ in a self-service environment (esp. automatized testing), adapting the ideas behind DataOps can indeed propel self-service towards higher degrees of efficiency – either by enabling more central control where needed or by enabling more self-service. Prerequisites are a motivated and capable user base as much as an active and professional IT provision and support. All these insights contribute to RQ1.

This paper proposes to capture the impact of the topic in discussion with a cost-based rationale that is built up on the components “central control” and “user agency” and the (opportunity) costs introduced by them. The concept underwent a preliminary evaluation with representatives from the case study. All this feeds into answers for RQ2.

The main contributions to research lie in the insights on self-service analytics and its DataOps-based professionalization, as well as in the cost-based concept. The evaluation indicates that the concept is also relevant to practice as it provides a conceptual toolset for discussing the benefits, prerequisites, and impacts of different degrees of central control, user agency and of implementing professionalization measures like DataOps.

A core limitation is that the results are based on a single case only. Although this is already addressed by disseminating the results, further studies are recommended. Also, the derived concept needs further fleshing out, esp. regarding the measurement of the dimensions and the cost calculation. Given the oscillations in BIA for or against centrality, I regard the approach as a first suggestion on how to settle the debate for good one day.

References

Alpar, P. & Schulz, M. (2016), 'Self-Service Business Intelligence', *Business & Information Systems Engineering* **58**(2), 151-155.

Baars, H. & Zimmer, M. (2013), A Classification for Business Intelligence Agility Indicators, *in* 'Proceedings of the 21st European Conference on Information Systems', Utrecht, Netherlands.

Baars, H., Felden, C., Gluchowski, P., Hilbert, A., Kemper, H.G. & Olbrich, S. (2014), 'Shaping the next incarnation of business intelligence', *Business & Information Systems Engineering* **6**(1), pp. 11-16.

Benbasat, I., Goldstein, D.K. & Mead, M. (1987), 'The Case Research Strategy in Studies of Information Systems', *MIS Quarterly* **11**(3), 369–386.

Chen, C. & Zhang, C. (2014), 'Data-intensive applications, challenges, techniques and technologies – A survey on Big Data', *Information Sciences* **275**, 314-347.

Davenport, T.H. & Harries, H.G. (2007), *Competing on Analytics: The New Science of Winning*. Boston, Massachusetts: Harvard Business School Press.

Dehghani, Z. (2022), *Data Mesh: Delivering Data-Driven Value at Scale*. Beijing, Boston, Farnham, Sebastopol, Tokyo: O'Reilly.

Ebert, C., Gallardo, G., Hernantes, J. & Serrano, N. (2016), 'DevOps', *IEEE Software*, **33**(3), 94-100.

Ereth, J. (2018), DataOps – Towards a Definition, *in* 'Proceedings of the Lernen, Wissen, Daten, Analysen – LWDA', CEUR Workshop Proceedings **2191**, Paper 13, pp. 104-112.

Ereth, J. & Baars, H. (2020), A Capability Approach for Designing Business Intelligence and Analytics Architectures, *in* 'Proceedings of the 53rd Hawaii International Conference on System Sciences', Maui, USA.

Fadler, M. & Legner, C. (2021), Toward big data and analytics governance: redefining structural governance mechanisms, *in* 'Proceedings of the 534th Hawaii International Conference on System Sciences', Kauai, USA.

Fang, H. (2015), Managing data lakes in big data era – what's a data lake and why has it became popular in data management eco-system?, *in* 'Proceedings of the IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems', Shenyang, China.

Garriga, M., Aarns, K., Tsigkanos, C., Tamburri, D.A. & van den Heuvel, W.J. (2021), 'DataOps for Cyber-Physical Systems Governance: The Airport Passenger Flow Case', *ACM Transactions on Internet Technology* **21**(2), 1-25.

Iyer, C.V.K., Hou, F., Wang, H., Wang, Y., Oh, K., Ganguli, S. & Pandey, V. (2021), Trinity: A No-Code AI platform for complex spatial datasets, *in* 'Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery', Beijing, China.

Jacquin, S., Pawlewitz, J. & Doyle, A. (2020), Democratizing Data with Self-Service Analytics, *in* 'Proceedings of the Offshore Technology Conference', Houston, USA.

Janiesch, C., Zschech, P. & Heinrich, K. (2021), 'Machine learning and deep learning', *Electronic Markets,* **31**, 685-695.

Krawatzeck, R. & Dinter, B. (2015), '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.

Labadie, C., Legner, C., Eurich M. & Fadler, M. (2020), FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs, *in* 'Proceedings of the 2020 22nd IEEE Conference on Business Informatics'. Antwerp, Belgium, pp. 201-210.

LaPlante, A. & Sharma, B. (2018) *Architecting Data Lakes*. 2nd ed., Sebastopol: O'Reilly.

Lefebvre, H., Legner, C. & Fadler, M. (2021), Data democratization: toward a deeper understanding, *in* 'Proceedings of the 42nd International Conference on Information Systems', Austin, USA.

Lim, E.P., Chen, H. & Chen, G. (2013), 'Business intelligence and analytics: Research directions', *ACM Transactions on Management Information Systems,* **3**(4), 1-10.

Marjanovic, O. & Dinter, B. (2018), 'Learning from the History of Business Intelligence and Analytics Research at HICSS: A Semantic Text-mining Approach', *Communications of the Association for Information Systems',* **43**(1).

Mayring, P. (2014), 'Qualitative content analysis: theoretical foundation, basic procedures and software solution', *Social Science Open Access Repository (SSOAR)*.

Myers, M.D. & Newman, M. (2007), 'The qualitative interview in IS research: Examining the craft', *Information and Organization*, **17**(1), 2-26.

Michalczyk, S., Nadj, M., Azarfar, D., Maedche, A. & Gröger, C. (2020), A State-Of-The-Art Overview and Future Research Avenues of Self-Service Business Intelligence and Analytics, *in* 'Proceedings of the 28th European Conference on Information Systems', Marrakech, Morocco.

Miles, M.B., Huberman, A.M. & Saldaña, J. (2014), *Qualitative data analysis: A methods sourcebook*. 3rd ed., Los Angeles, London, New Dehli, Singapore, Washington DC: Sage.

Mucha, T.M., Ma, S. & Kaveh, A. (2022), Beyond MLOps: The Lifecycle of Machine Learning-based Solutions, *in* 'Proceedings of the 2022 Americas Conference on Information Systems', Minneapolis, USA.

Munappy, A.R., Mattos, D.I., Bosch, J. Olsson, H.H. & Dakkak, A. (2020) 'From Ad-Hoc Data Analytics to DataOps', *in* 'Proceedings International Conference on Software and System Processes', Republic of Korea, pp. 165-174.

Namyslo, N. & Baars, H. (2022) Harnessing the Business Potential of Self-Service Machine Learning for Forecasting Warranty Costs – Insights from a Case Study in the Automotive Sector, *in* 'Proceedings of the Pacific Asia Conference on Information Systems', Taipei, Taiwan and Sydney, Australia.

Ooms, R. & Spruit, M. (2020), 'Self-service data science in healthcare with automated machine learning', *Applied Sciences,* **10**(9).

Oreščanin, D. & Hlupić, T. (2021), Data Lakehouse – a Novel Step in Analytics Architecture, *in* 'Proceedings 44th International Convention on Information, Communication and Electronic Technology', Opatija, Croatia, pp. 1242-1246.

Ravitch, S.M. & Riggan, M. (2012) *Reason & Rigor: How conceptual frameworks guide research*. Los Angeles, London, New Dehli, Singapore, Washington DC: Sage.

Sánchez-Gordón, M. & Colomo-Palacios, R. (2018), Characterizing DevOps culture: a systematic literature review', *in* 'Proceedings of the International conference on software process improvement and capability determination'. Thessaloniki, Greece, pp. 3-15.

Seddon, P.B., Constantinidis, D. & Dod, H. (2012), How does business analytics contribute to business value?, *in* 'Proceedings of the 33rd International Conference on Information Systems', Orlando, USA.

Sharda, R., Delen, D. & Turban, E. (2014), *Business Intelligence and Analytics – Systems for Decision Support*. 10th ed. New Jersey: Pearson.

Thomson, W.J.J. and van der Walt, J.S. (2010), 'Business intelligence in the cloud', *South African Journal of Information Management,* **12**(1).

van der Goes, M. (2021), Scaling Enterprise Recommender Systems for Decentralization, *in* 'Proceedings of the 15th ACM Conference on Recommender Systems', RecSys, Amsterdam, Netherlands, pp. 592-594.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R. & Cleven, A. (2009), Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process, *in* 'Proceedings of the 17th European Conference on Information Systems', Verona, Italy, pp. 1-12.

Walsham, G. (1995), 'Interpretive case studies in IS research: nature and method', *European Journal of Information Systems*, **4**(2), 74-81.

Webster, J. & Watson, R.T. (2002), 'Analyzing the Past to Prepare for the Future: Writing a literature Review', *MIS Quarterly,* **26**(2), pp. xiii-xxiii.

Xu, J., Naseer, H., Maynard, S. & Filippou, J. (2021) 'Leveraging Data and Analytics for Digital Business Transformation through DataOps: An Information Processing Perspective', in *Proceedings of the Australasian Conference on Information Systems*, Sydney, Australia.

Yin, R.K. (2018), *Case study research and applications – design and methods*. 6th ed., Los Angeles, London, New Dehli, Singapore, Washington DC, Melbourne: Sage.