UNDERSTANDING DATA PRODUCTS: MOTIVATIONS, DEFINITION, AND CATEGORIES

Research paper

M Redwan Hasan, Faculty of Business and Economics (HEC), University of Lausanne, Switzerland, mredwan.hasan@unil.ch

Christine Legner, Faculty of Business and Economics (HEC), University of Lausanne, Switzerland, christine.legner@unil.ch

Abstract

As the volume of data exponentially increases, organizations are looking for smarter ways to create the most value from their data. One approach to achieve this is through developing data products. Although the idea was initially presented in the 1990s, the concept remains nascent, leading to different groups forming their own interpretations about data products. Leveraging the literature and multiple case studies, we aim to harmonize the understanding of data products and identify their characteristics. Additionally, our empirical findings shed light on the motivations to develop data products as well as the emerging data product categories. By clarifying the foundations of data products, our study contributes to the ongoing discourse around scaling data and analytics in enterprises to repurpose and consume data efficiently and cost-effectively. For practitioners, our study provides insights into different motivations and priorities associated with data products, which can help them scope their data product initiatives.

Keywords: Data product, Information product, Data analytics, Data mesh

1 Introduction

Recent forecasts show that data will grow at a compound annual growth rate of 21.2% to reach more than 221 zettabytes by 2026 (Burgener and Rydning, 2022). These increasing volumes of data open manifold opportunities to create business value, by accelerating innovation, driving optimization, and improving business performance (Grover *et al.*, 2018). However, organizations struggle to cope with the ever-increasing demand for data and analytics from business users. The current practice of reworking data per use-case has led to fragmented and duplicated data pipelines (Desai *et al.*, 2022). Such an approach can be cost-intensive and slow down analytics-related information delivery (Dinter, 2013). Moreover, data often remains in silos and prevents scattered business units from collaborating toward analytics-based insights (Mikalef *et al.*, 2017). It is rarely reused and shared in cross-functional teams, or even with designated external users, to enable good decision-making and create data-driven innovations (Gelhaar and Otto, 2020). The low adaptability of data-generated insights to the changing business landscape (Mikalef *et al.*, 2020) runs the risk of reducing the overall return on investment on data (Shim *et al.*, 2015).

Against this backdrop, recurring calls have been made to revise the current approach and treat data with a product mindset. Industry experts argue that by managing data as a product, organizations can realize business use-cases 90% faster, reduce data governance risks and standardize data to be aligned with various consumption archetypes (Desai *et al.*, 2022). The data mesh paradigm (Dehghani, 2021) sees data-as-a-product as one of the four principles that help organizations improve data user satisfaction, and decrease the lead time of data consumption. Data products should improve organizations' data operating models (Assur and Rowshankish, 2022), make data accessible to cross-functional teams (Sands, 2018), and offer potential to create new revenue streams (Meierhofer *et al.*, 2019).

Despite these calls and their growing popularity, the concept of data products remains nascent, leading to different groups forming their own interpretations. For instance, data products have been discussed from diverse lenses such as data science (Meierhofer *et al.*, 2019; Stadelmann *et al.*, 2022), service engineering (Meierhofer and Meier, 2017), data mesh (Dehghani, 2019; Machado *et al.*, 2021), and data marketplaces (Fricker and Maksimov, 2017). Organizations find it challenging to define data products, due to end users' differing business priorities (Loukides, 2011). As their requirements vary ongoingly, companies adopt several ways to establish data products to fulfill these needs. Lack of a standard approach results into confusion around data product's purpose, use, accountability and management. Finally, we observe that the recent debate on data products ignores the academic arguments for a product view on data and information that have been discussed since Wang *et al.* (1998)'s seminal article.

We therefore do not have a coherent definition and understanding of the different facets of data products that integrates academics' and practitioners' perspectives. More importantly, a lucid understanding of data product characteristics could help guide how they must be created and who should be accountable for managing them (Dehghani, 2019). With this goal in view, we propose the following research questions:

RQ1: What motivates companies to develop data products?

RQ2: How do companies define and categorize data products?

To study data products in a naturalistic enterprise setting (Benbasat *et al.*, 1987), we conduct case studies with four global firms that represent different industries and have gained experience in building data products. By comparing our empirical findings to literature on data products, we contribute to harmonizing the understanding of data products by outlining characteristics from the literature, and extending them based on the case insights to provide a definition. Our empirical findings shed light on the motivations to develop data products as well as the emerging data product categories. By clarifying the foundations of data products, our study informs the ongoing discourse around scaling data and analytics in enterprises to repurpose and consume data efficiently and cost-effectively. For practitioners, our study provides insights into different motivations and priorities associated with data products that can help them scope their data product initiatives.

In the next section, we discuss the background of this study, followed by a detailed discussion on the five characteristics of data products. Subsequently, we lay out the methodology and the research process. We then present the findings from our within-case and cross-case analysis. We conclude by discussing this study's results and limitations, and providing an outlook on future research.

2 Background

The product view on data and information is not new in academic research, but was initially coined in the 1990s by Wang *et al.* (1998). Since then, the term has been sparringly mentioned, with only a handful number of papers that have truly attempted to explore the subject in details albeit in differing contexts. From our review of the relevant literature[1], we identified 17 papers that go beyond just mentioning data products briefly, but also attempt to offer a definition, elaborate on their characteristics, potential users and discuss their types and examples. Interestingly, earlier publications in the 1990s used the term information product, whereas recent publications prefer the term data product. Based on our examination, we review the development of the concept and identify recurring themes and five characteristics of data products.

---

[1] We retrieved the papers from ACM digital library, EBSCO, AIS eLibrary, ProQuest, and Google Scholar, using the search terms 'data product(s)' or 'information product(s)'. We discarded papers from non-related domains such as remote sensing, chemistry, and earth sciences, only keeping the ones published in top IS outlets or that had significant IS-centric contents.

2.1 From information products to data products

The term information product was initially derived by drawing parallels between information manufacturing and product manufacturing in order to manage information quality, leading to the establishment of the total data quality management method (Wang, 1998). Other authors have expanded the product view of information to various contexts and examples (Table 1). The authors mainly saw an information product as a source to fulfill simple end user needs with regards to data, such as client account data or reports. They frame the concept within a finite scope, similar to a physical form, containing a limited amount of information.

Source	Definition	Examples
Phase 1: Information products (1990s and 2000s)		
(Wang, 1998)	Information products are defined as the result of activities that take place within the information supply chain.	Client account data
(Shankaranarayanan *et al.*, 2000)	Data items that are required to fulfill consumer needs and can range from raw data and semi-processed information to final information products.	Certificates, bills, transcripts, bank statements
(Cai and Ziad, 2003)	An information product is a specific deliverable that aligns with end user requirements.	Invoices, business reports & prescriptions
(Davidson *et al.*, 2004)	Information products are a collection of data elements aimed at a specific purpose.	Birth certificate
(Wang *et al.*, 2005)	An information product is identified in terms of data items that comprise it, and it is the quality of each data item that is of importance to the consumer.	Certificates, mailing labels, sales orders
(Nam and Lamb, 2006)	An information product is any valuable information for which users are willing to pay. for	News products
Phase 2: Data products (since 2010)		
(Loukides, 2011)	Data products are not about data, but about enabling users to do what they want to do. Data products should deliver results rather than data, and data is invisible in the product.	Spreadsheets, recommendations, self-driving cars
(Bengfort and Kim, 2016)	Data products are self-adapting, broadly applicable economic engines that derive their value from data and generate more data by influencing human behavior or by making inferences or predictions upon new data.	Nest thermostat, autonomous vehicles, 'quantified self'
(Davenport and Kudyba, 2016)	Data products (which can mostly be described as services) are not generally sold separately to customers but are used to attract customers for advertising, draw attention to unknown products in large product pools, and enhance revenue through cross-selling and upselling.	Predictive maintenance, property price predictions, matching algorithms
(Meierhofer *et al.*, 2019)	A data product is defined as the application of a unique blend of skills from analytics, engineering and communication aimed at generating value from the data itself to provide benefit to another entity.	Customer analytics insight
(Si *et al.*, 2020)	Data products result from data resources after desensitization, encapsulation, and right identification. They have the dual characteristics of data and product.	Monetizable datasets
(Fruhwirth *et al.*, 2020)	Data products help their users to make better decisions and formulate customer benefit. The users can be internal or external customers	Reports, dashboards, APIs
(Machado *et al.*, 2021)	Data products can be understood as a set of data that instantiate the domain.	Domain sales data, online profit data
(Chen *et al.*, 2022)	Data products may be datasets packaged and designed as products or services by developers for data owners or stakeholders. They have potential applications and values for data buyers or new users to pay.	Personal data, financial data, pharmaceutical data

Table 1. Definitions of information product(s) and data product(s) in the literature

Since the 2010s, the term data product took primacy and was viewed as an artifact that acquires value from data and creates new data using prior data (Loukides, 2011). For instance, the use of iTunes playlists results in additional data being captured from listeners that can be reanalyzed to decide how to optimize the list (Loukides, 2011). Data products can manifest in mainly two forms: *overt* data products, where data itself is the output, and *covert* data products, where data is invisible and works in the background (Loukides, 2011; Meierhofer and Meier, 2017). To leverage the increasing amount of new data that is being created, the approach of combining it with simple business intelligence techniques can lead to valuable data products that facilitate decision support (Davenport and Kudyba, 2016).

Successively, the realm of data science gained precedence due to the widespread adoption of self-learning capabilities and the DataOps concept (Munappy *et al.*, 2020), with the goal to move toward prescriptive knowledge creation (Lepenioti *et al.*, 2020). Data products were viewed as artifacts that were built using ML techniques (Bengfort and Kim, 2016) or approaches like MLOps in conjunction with skills around data management, analytics, art and design, and entrepreneurship (Stadelmann *et al.*, 2022). This highlights advanced supervised and unsupervised techniques that can be used to develop data products (Meierhofer and Meier, 2017). Moreover, visualization capabilities are equally important to communicate patterns in an easy-to-use creative manner (Echeverria *et al.*, 2018). Hence, data products attempt to bridge analytical results from datasets to match information requirements from consumers – a gap that should be addressed to foster higher value creation (Meierhofer *et al.*, 2019).

Most recently, the data mesh paradigm has put data products as one of its core principles (Dehghani, 2021). It emphasizes the domain-oriented creation and management of data in a product-oriented way, led by responsible teams (Machado *et al.*, 2021). Data products created with this mindset would correctly instantiate a domain (Machado *et al.*, 2021), enhance data sharing supported by DATSIS principles (Dehghani, 2021) and help manage data quality in a decentralized manner (Dehghani, 2019), in addition to improving governance by clarifying ownership, access, and control of the data (Joshi *et al.*, 2021).

The data product concept has evidently received significant attention over the years. The continued interest has allowed data products to evolve to types that can address new challenges in organizations. Recent phenomena, such as data repurposing (Zhou *et al.*, 2021) and data sharing (Jussen *et al.*, 2023) have encouraged organizations to reflect beyond the current narrative of viewing data as just a by-product toward acknowledging its strategic important. This leads to vital considerations around data, such as its ownership, quality, or lifecycle – all of which manifest if organizations can manage data as products.

2.2 Key characteristics of data products

From our review of the literature, we derive five characteristics that help us articulate the concept of data products by emphasizing their purpose, intrinsic utility, and tangible properties beyond a particular context. These are discussed below (Table 2).

Data products satisfy recurring information needs: Data products are created in order to fulfill end users' information needs that manifest regularly (Loukides, 2011). If an artifact satisfies only ad hoc or temporary information needs that do not manifest consistently, it cannot be termed a data product. The onus lies on organizations to fulfill their requirements by using the most appropriate, high-quality insights (Salaün and Flores, 2001). It remains a significant challenge to deliberate consumer needs while creating data products due to the different nature of needs (Howard *et al.*, 2012) as well as cognitive biases users may have that impact decisions (Ni *et al.*, 2019). To analyze information needs, several publications suggest methodological approaches, such as the value proposition design framework to map customer pains and gains to product features (Meierhofer *et al.*, 2019) or a dedicated service design process that maps existing customer situations and segments along their journeys (Meierhofer and Meier, 2017).

Source	C1: Satisfies recurring information needs	C2: Has a well-defined consumer base	C3: Creates measurable value for organizations	C4: Produced through collating different data elements	C5: Delivered in a consumable form
(Meierhofer *et al.*, 2019)	Customer pains and gains while using products	External customers or internal users	New revenue stream for insights generated	-	Packaging insights into publishing form
(Meierhofer and Meier, 2017)	Customer product usage behavior; Performance of smart homes	Customer service personnel; Private home owners	Improved firm offerings and revenue	-	Connected homes service; NLP software to support CSR personnel
(Loukides, 2011)	User flexibility to derive information on topics aligned to their needs	-	Improved decision-making through insights from hidden data sources	Join datasets on users, job listings and employers to build LinkedIn's skill database	Music collection spreadsheet; LinkedIn's 'people you may know'
(Stadelmann *et al.*, 2022)	-	Data scientists	-	Various datasets must be used at the core of the data product	Digital service; Physical product; Hybrid
(Davenport and Kudyba, 2016)	-	External consumers	Attracting customers to other products in the portfolio and to drive cross-selling and upselling	Increasing number of data assets are combined as input to analytical operations	Online data products offered through mobile applications with built-in analytics
(Fruhwirth *et al.*, 2020)	User pain points to gain customer intimacy and understanding	Internal company representative; External users	Data-driven business model	Identification of various data sources as input for data products	KPIs, reports, dashboards or APIs as per user needs
(Si *et al.*, 2020)	-	Data product buyers; Data product sellers	Monetizing data products by selling them in the data circulation market	-	Desensitize and encapsulate data to protect privacy
(Bengfort and Kim, 2016)	-	Data scientists; Hadoop developers	Discovery of individual patterns in human activity and drive decisions for business goals	Blend initial and new data generated by influencing human behavior to make inferences and predictions	Fitbit's 'quantified self', Stanford University's autonomous vehicle
(Dehghani, 2021)	Visibility into the different domain data that exists in the organization	Internal domain users; External business partners	Reduced process bottlenecks; High-quality data; Federated governance	Different types of data combined to drive insights for known and unknown use-cases	-

Table 2. Five characteristics of data products based on prior literature

Data products have a well-defined consumer base: Data products are created with (internal or external) consumers in mind, as defined by the organization (Fruhwirth *et al.*, 2020; Meierhofer *et al.*, 2019). If the knowledge around an artifact's consumer base is unclear, it cannot be called a data product. From the perspective of service engineering, both company personnel and customers receiving services were featured as relevant consumers (Meierhofer and Meier, 2017). Other authors only pointed out technical roles such as data scientists and developers as being relevant (Bengfort and Kim, 2016; Stadelmann et al., 2022). In the data mesh concept, the focus is more around the cross-domain users and external business partners with whom data products are shared to provide rich insights (Dehghani, 2021). With the advent of IoT, machine-to-machine interaction implies that smart products and devices can now communicate through data products to share insights with minimal human intervention, such as in a digital twin setup (Gazis, 2017).

Data products create measurable value for organizations: Data products must produce tangible value for companies that may be either monetary and non-monetary (Davis *et al.*, 2020). The value can appear in different forms, such as leveraging hidden trends in human behavior to enhance existing product functionalities (Bengfort and Kim, 2016) or using popular tools such as Google Analytics to build relevant KPIs to measure customer activity around product usage (Davenport and Kudyba, 2016). If concrete value cannot be measured from the use of an artifact, such an artifact cannot be termed a data product. Data product offerings can be further tailored by shaping customer profiles through customer insights research (Meierhofer and Meier, 2017), and convincing them to reuse data products for different purposes while paying a good value for it (Meierhofer *et al.*, 2019). Organizations can also directly trade repackaged data as products in the data circulation market (Si *et al.*, 2020), but may require unique business models to exploit such opportunities (Wixom and Ross, 2017).

Data products are produced through collating different data elements: Data products are generated from various data sources based on their potential application and usage scenarios (Chen *et al.*, 2022). Bringing together data from various sources to build data products is key in creating well-formulated insights (Chen *et al.*, 2022; Fruhwirth *et al.*, 2020) and also helps move beyond data silos (Dehghani, 2019). For instance, data products can be as simple as spreadsheet files with a collection of relevant attributes that provide basic information to end users (Loukides, 2011). Concretizing the analytical needs first and checking if relevant data exists to fulfill them can help to identify key data elements as well (Bengfort and Kim, 2016; Davenport and Kudyba, 2016). However, a data product cannot be built if the key data sources are unknown, or are of poor data quality.

Data products are delivered in a consumable form: Data products must be packaged, delivered, and accessed in a convenient manner to meet specific user conditions (Chen *et al.*, 2022; Meierhofer *et al.*, 2019). If the form does not match consumers' aptitude level, it does not qualify as a data product. It is absolutely critical to provide data products in a format consumers can easily use to digest the information packed within it. This characteristic appears frequently in the literature, with multiple examples. For instance, data products can be different insights packaged together to support decisions (Meierhofer *et al.*, 2019) such as reports, or be developed into organization-wide KPIs and metrices to offer visibility into real-time business situations (Martins de Andrade and Sadaoui, 2017). At an advanced level, data science techniques are applied to produce data products that may either take physical shapes such as autonomous vehicles (Bengfort and Kim, 2016) or be in software form such as recommendations (Kumar and Thakur, 2018). This signifies that design requirements for data products remain a key area to look into (Howard *et al.*, 2012).

3 Methodology

While the idea of data products and their characteristics has been discussed in prior literature, we know relatively little about their implementation in real-life contexts. To address this gap, we chose a case study research design (Yin, 2003) to get practical insight on how data products are formed, organized, and managed. Through case studies, we are able to study the phenomenon of interest in a naturalistic setting (Benbasat *et al.*, 1987) and add empirical insights to the existing literature. Evidence from multiple case studies can lead to more convincing results, reinforce robustness, and assist in analytical generalization (Yin, 2003). A major strength of the case study method is facilitating the use of multiple data sources, which allows data triangulation and eventually enhances research quality (Patton, 2014).

Our research context is a research program in data management that involves data experts from large multinationals and a team of researchers. Initiatilly, a focus group was run with 10 companies to understand their current position around data products and corresponding challenges. Using purposeful sampling, our subsequent activities focused on the four (out of the 10) companies that noted data products to be highly relevant for them and have already developed and deployed data products in their organizations. These cases varied in terms of their approach to data products as well as their industry, date of registration, number of employees, and annual revenue. This allowed us to draw parallels and analyze differences among them. To gather our data, we conducted one-hour long semi-structured interviews with experienced professionals leading these initiatives who were knowledgable about data products, and open to sharing their approach and experiences. We also gained access to and reviewed internal company documents that provided practical insights. These different sources of data collection facilitated the triangulation of information and ensured construct validity (Yin, 2003). Table 3 summarizes the companies selected.

As part of the within-case analysis, we analyzed data from each company to delineate their motivations behind having data products, their categories and examples, allowing the “unique patterns of each case to emerge” (Eisenhardt, 1989, p. 540) and laying the groundwork for richer insights. The first researcher conducted the within-case analysis by coding each case with regards to the motivation and categories of data products. We also compared the firms’ definition of data products to the five characteristics extracted earlier (Table 5). The second researcher undertook an independent and thorough review of the cases and codings. Subsequently, we performed the cross-case analysis “to go beyond initial impressions especially through the use of structured and divserse lenses on the data” (Eisenhardt, 1989, p. 541). We conducted pattern matching to draw parallels between the firms’ data product initiatives. In particular, we categorized the common data product motivations and formulated the emerging data product categories based on the case descriptions and examples. To gather feedback, we organized another focus group with 16 data management and analytics experts from nine global firms, including all the interviewees from the four case companies, to discuss and refine the insights from our case analysis.

Company, year, and industry	Revenue/Number of employees	Goal of the data product initiative	Participants and experience
PackF (1951) Packaging	\$1-50B / ~ 25000	Data product management supports their data monetization strategy	Enterprise data governance manager (30+ years); Service delivery manager (17+ years)
ManufO (1946) Manufacturing	\$1-50B / ~ 80000	The data product approach improves data governance	Data and analytics governance manager (10+ years)
TeleC (1876) Telecommunication	\$1-50B / ~ 100000	Data products expand the data foundation program to enable more data-driven use-cases	Head of data foundations and data management (24+ years)
FoodM (1866) Food & drink	\$50-100B / ~ 250000	Data products are part of the data foundation program to harmonize data pipelines and secure funding	Master data product manager (20+ years); Data and analytics product manager (18+ years)

Table 3. Overview of the case companies

4 Case descriptions

Below we present the data product initiatives in the four companies. The data has been gathered during the interviews, firstly by recording it with the participant’s permission and then transcribing them. The collected data was further improved with the support of internal company documents.

4.1 PackF

Motivation: Courtesy of their data monetization strategy, PackF started to build data products. Through this approach, they aim to scale their analytics activities, sharing valuable insights with different customers in a faster way and find new revenue sources.

Definition and categories: PackF defines data products as products that facilitate an end goal using data, and categorizes them into three categories: data and insight data products that include master data objects such as HR data, accounts and hierarchy data, analytics reports or composite data; value-adding data products such as predictive maintenance algorithms; and data exchange data products such as APIs.

Changes and implications: Various platforms should be able to support the data products, such as SAP ERP to store and manage the master data objects, whereas more consumer-centric data products such as analytics reports and APIs could be hosted in cloud platforms such as Azure and AWS. The company aims to become more proactive, with the goal to cater more toward novel use-cases. To support their data monetization strategy, PackF is implementing their enterprise data management (EDM) framework, which ensures correct data modeling, data cataloging, and data quality. The data product concept plays a key role in this, by defining responsibilities and determing required quality levels for the data products. For instance, PackF appointed a data analytics leader for each development team who manages the data product, the data, and eventually the aggregated data products that are shared across domains. Conversely, the data product development teams are placed under the operations division and include data stewards, business experts, data architects, data scientists, and data engineers. They are also responsible for various platforms that enable data products, such as PowerBI and Azure. This separation has created confusion regarding who is actually responsible for building and managing the lifecycle of data products.

Lessons learnt: A correct product-oriented view around data will allow PackF to overcome challenges and facilitate a consistent approach toward governance. As their enterprise data governance manager pointed out, "*We don't have time to validate data; if we have a data product, by definition it would mean that it has a proper owner, is well-cataloged and has high quality.*" This should eventually enhance reliability and speed up consumption, driving data sharing and better management of the data lifecycle.

4.2 ManufO

Motivation: ManufO has adopted the data product initiative under their digitalization strategy. Through data products, the company is seeking to reduce their time-to-market because it lowers their data usage costs and optimizes business processes. ManufO seeks to improve governance through clarifying ownership, developing relevant documentation and enabling data sharing across domain users.

Definition and categories: ManufO defines data products as autonomous, read-optimized and standardized data units that contain at least one dataset (domain dataset) created to satisfy user needs. They have two main categories: source-aligned data products such as master data and finance data and consumer-aligned data products such as lists/tables of analytical data, analytics algorithm and dashboards, such as HR-related KPIs, metrices, and graphics.

Changes and implications: With around 40 data domains, the firm is aiming to instantiate data products in each of these areas. They have an analytics product owner to own the product, along with a team that includes data owners, data analysts, and data architects. APIs allow end users to access these data products. The data product topic is promoted through smaller and bigger pilots. PowerBI and Mendix are used as tools to create newer dashboards and get people used to it. They are aiming to develop a standard that will guide users regarding which tools to use to build a certain type of data product. This is key because most data domains use SAP Analytics but users often prefer PowerBI. Standardization will be able to help maintain coherence, ensure compliance, and avoid extra work to overload end users.

Lessons learnt: ManufO learnt that working closely with people is the key to unlock data products' potential. The firm's future goal is to clearly understand what type of data they have, where they have it, which data will be heavily used and which tools are available to build data products – all at the intersection of governance and IT delivery. As their data and analytics governance manager stated, "*Most challenges are related to humans; changing their mindset to see data as a product is the issue.*"

4.3 TeleC

Motivation: To accelerate their digital transformation strategy, TeleC launched a data foundation program to provide data at a faster speed to consumers, enhance the data user experience, foster a data-driven culture and improve data security and compliance. Under this initiative, the data product approach was adopted to deal with concrete challenges, such as breaking silos across data domains to foster reusability and packaging of the data to support various analytics use-cases as well as improving governance across domains. The idea is to meet the increasing demand for data smarter, faster, and more cost-efficiently, as opposed to taking a fragmented approach to every use-case.

Definition and categories: TeleC defines data products as packages of data and code at various levels of preparedness and refinement for reusability to support multiple business usage scenarios. They also specify further requirements toward data products that need to be packaged, orderable, provide business value and are managed throughout their lifecycle. TeleC propose three categories: foundational data products that are basic in nature, such as customer data and aggregated data, insight data products that are developed by applying analytics, such as a supply chain dashboard in Tableau, and data delivery data products that are mainly oriented to the consumer, such as algorithms and ML models.

Changes and implications: Their strategy has been to create data products in the customer information domain as a pilot, while gaining maturity over time and eventually replicating the methods to other domains. The possible users are data citizens, analysts, scientists as well as machines and applications. SAP MDG is used as a platform to store all the customer data in one place and push it to other systems when needed. TeleC highlights lifecycle management as a major implication of the data product approach because they have to strike a balance between building standard data products versus use-case-driven data products. They aim to create good data products in limited quantities by correctly translating business priorities to technical designs, having the right metadata available in catalogs, providing product access in marketplaces and managing lifecycles.

Lessons learnt: A data product framework with methods and principles can assist any development team with the required skillset to create their own data products and eventually rally more users. Their head of data foundation and data management stated, "*We want to seamlessly connect data providers and consumers through getting data into the platforms, building data products out of it and then put it on the market.*"

4.4 FoodM

Motivation: As part of their data foundation program, FoodM has adopted the data product initiative to better grasp the consumption mechanisms around end users. Such clarity will allow them to develop and offer relevant products and services instead of building a solution for every need. Having a data product concept will also help them to ensure secured funding to drive important goals around governance and data quality, which is key for the company. As the current challenge lies around creating fragmented pipelines for each use-case, the firm would like to consolidate and optimize this approach through forming data products that can have a standard pipeline and support an array of user requirements.

Definition and categories: FoodM defines data products as a mash of the correct delivery mechanism, compliance, access control, quality, and security of the data to be provided to the organization. They define two main categories of data products: (1) source-aligned, which involves various master data in SAP, transactional data like like order and delivery data and cross-functional data assets such as sustainability and commercial data foundation; and (2) consumer-aligned, which includes PowerBI dashboards in the area of purchase and supply chain and corporate reporting. The firm is has higher maturity in the first category.

Changes and implications: FoodM has around 140 level 3 domains, with level 0 being the company itself. As the documentation of the domains has been completed, the company is looking to launch data products for consumption in some areas. They still lack skills around productizing the data, such as managing metadata and putting the data products into a particular delivery format. Furthermore, having a complex structure, the firm aims to have data owners in different domains who will manage the products and their lifecycles. Their data literacy program is also being utilized to propagate the idea of the productization of data and how it will impact critical quality indicators like consistency, accuracy, and availability – ultimately scaling analytics and enabling faster time to insight.

Lessons learnt: FoodM stresses that the data product concept must be profused carefully through the support of the senior teams and riding on the wave of digital transformations. As their group manager of data and analytics products mentioned, "*We are optimistic because people are slowly understanding* *that data products are a key to addressing critical challenges. We attained that in the master data area, now we have to see how we do beyond it.*"

5 Results

5.1 Motivations to develop data products

Our within-case and cross-case analyses reveal that companies understand data products as a paradigm shift that changes their overall approach to data. While there is consensus about the importance of the concept, the companies have different motivations behind creating data products. Such motivations are associated with goals, but also imply specific focus in their data product initiatives (Table 4).

Data reuse and access have been important drivers for all the companies. A primary reason is that organizations have a lot of data across different functions, teams, and domains. For instance, ManufO has 40 and FoodM has 140 data domains. It is frequently required that end users access and consolidate the data from different areas to create comprehensive insights and drive analytics use-cases. A lack of data reusability will lead to higher cloning – data copies in different systems, leading to multiple sources of truth and reducing data quality through inconsistency. Data products could help overcome this challenge, as it will facilitate harmonizing the source data in individual domains, establishing guidelines for access and making it discoverable to cross-domain users. This makes the data available and reusable in a controlled manner.

Most of the case companies underscore *streamlining governance* as a major push behind adopting data products. In particular, the ownership of data product is seen as a crucial element. For instance, ManufO has appointed analytics product owner and PackF appointed data analytics leader. Without a clear owner of the data products, companies face bottlenecks around improving data quality, remaining compliant and managing lifecycles. As a result, the potential value that could be created through sharing such data remains untapped. This aligns well with the current literature that has highlighted the need for a data product owner who is accountable for the business value of data products (Fadler and Legner, 2021). Clear data product ownership positively impacts data quality, as the owner takes responsibility for fixing the data close to source and improves cross-domain reusability (Dehghani, 2019).

Motivations	Goals	Focus
Data reuse and access	**FAIR**: Data products must be easily findable and accessible through leveraging their metadata.	Data catalogs Data marketplaces
	Transperant access and licensing: Data products must be accessed in a controlled manner and under clear conditions for use and reuse.	Contracts License models
	Uniform use: Have a single source without creating copies in other systems. If absolutely necessary, the owner must be consulted.	Provider-consumer relationships
Governance	**Accountability**: There must be clear accountability on who is responsible for the overall health and success of the data product	Data product ownership
	Life cycle: Each data product must be created and managed through a life cycle approach, from its conceptualization to retirement.	Life cycle processes
	Compliance: Providing basic data products must conform to regulations about what to give, with whom and how much.	GDPR policy
	Data quality: Quality criteria of the data empowering the data product must be clearly defined, communicated and documented.	Data quality by design Data quality criteria
Time to insight	**Lean infrastructure**: Data products must be created by using authorized set of tools, processes and systems.	Data and analytics platforms
	Scalability: Data products must not be created per use-case. Confirm if the same product exists; if not, use an approved procedure to build.	Standards and guidelines

Table 4. Motivations for developing data products

The organizations also emphasized the need to reduce the *time to insight*. There has been a significant increase in the demand for analytics because users aim to exploit the high volume of data to grab value-generating opportunities. Firms like TeleC amalgamate relevant data objects from different sources, package it in an easy-to-use manner and distribute it across the organization in a raw, standardized or prepared format, saving users the manual effort. This allows the possibility of gaining rapid insights, making quick decisions and fostering insight-based learnings (Marjani *et al.*, 2017). Although advanced analytical approaches hold a lot of promise in producing diligent acumen, uncertainties remain due to the changing nature and volume of data and use of complex analytical methods (Hariri *et al.*, 2019).

5.2 Definition and characteristics of data products

Our empirical insights reveal that each case firms define data products differently and use part of the characteristics but the definitions together capture all the elements of the five characteristics of data products identified from the literature (Table 5). We additionally find that there is a key requirement to clarify who owns and is ultimately responsible for the data products. A clear ownership structure will help ensure compliant access, drive the use-case portfolios, and manage the data product life cycle tasks (Fadler and Legner, 2021). Additionally, this characteristic can be grounded in the data mesh approach, where domain-driven design encourages the decentralized ownership of data products (Machado *et al.*, 2021). In turn, this makes the owners accountable for uniting diverse data sources, scaling data product pipelines, and adhering to governance obligations (Dehghani, 2021).

With this, we define data products as follows: *A data product is a managed artifact that satisfies recurring information needs and creates value through transforming and packaging relevant data elements into consumable form*. It is worth pointing out that consumers can be both internal or external to the organization, and be either a human or a machine. TeleC, for instance, has developed data products such as machine learning algorithms that can be consumed directly by other network devices. Similarly, PackF installed predictive maintenance software at the external partner site, while the machine line data used to build it was used and managed internally. This signifies the necessity of interoperability standards that can facilitate multiple data products cooperating with each other (Dehghani, 2019) in specific contexts such as digital twins (van der Valk *et al.*, 2022).

Company	Motivation for data products	Data product definition and characteristics	Categories and examples of data products
PackF	• Scaling analytics • Share insights with different customers • Find new revenue stream	**Definition**: A product that facilitates an end goal using data **Characteristics**: C3, C4	**Data and insights**: HR data, accounts and hierarchy data, analytics report, composite data **Value-add**: Predictive maintenance algorithm **Data exchange**: APIs
ManufO	• Reduce the time to market • Increase data sharing across domains • Improve governance to enhance consumability and data quality	**Definition**: Autonomous, read-optimized and standardized data unit containing at least one dataset (domain dataset) created to satisfy user needs **Characteristics**: C1, C2, C4, C5	**Source-aligned**: Master data objects, finance data **Consumer-aligned**: Lists/tables of analytical data, analytics algorithm, HR dashboards, KPIs, metrices, APIs
TeleC	• Provide data to consumers quickly • Enhance user experience • Develop a data-driven culture • Improve compliance	**Definition**: Packaging of data and code at various level of preparedness and refinement for reusability to support multiple business usage scenarios **Characteristics:** C3, C4, C5	**Foundational**: Customer data, aggregated data **Insight**: Supply chain dashboard **Data delivery**: Algorithms, ML model
FoodM	• Harmonize fragmented data pipelines • Grasp consumption mechanisms of end users • Improve governance and data quality	**Definition:** Mash of the correct delivery mechanism, compliance, access control, quality and security of the data to be provided to the organization **Characteristics**: C2, C5	**Source-aligned**: Master data objects, transactional data, cross-functional data **Consumer-aligned**: BI dashboards, corporate reporting

Table 5. Motivation, definition, and categories of data products in the companies

5.3 Categories and examples of data products

Our cross-case analysis further uncovered that all companies distinguish different categories of data products. ManufO and FoodM inspired their categorization from data mesh (source-aligned and consumer-aligned), PackF created their own tailored categories around the exchange of data and value (data and insight, value-add and data exchange), while TeleC's categories were more oriented toward phases of the information supply chain (foundational, insight, and data delivery). Although firms use different categorizations, we found similarities that allow us to generalize. First, they all distinguish between more foundational data products that are aligned with the source systems, and more analytics-focused data products that are more consumer-oriented. Second, another distinction is made between the complexity of analytics, with one supporting manual decisions and another offering advanced self-learning capabilities. From these considerations, we derive three emerging data product categories.

The first category, *basic data products*, is mainly at a dataset level that can be analyzed and explored to gain foundational knowledge of the concerned domain. Examples are HR data (PackF), customer information (TeleC), finance data (ManufO), and order and delivery data (FoodM). They originate on the operational end in master data systems, such as SAP or business applications like CRMs or ERPs. It can lay the groundwork for further analytics and can also be combined with similar datasets to create more aggregated or composite data products. One example could be the sustainability data asset used at FoodM. Being rudimentary in nature, basic data products can drive a wide array of end user needs such as profiling data attributes, standardizing and validating data, building training and test datasets, performing analytics using self-service platforms and fulfilling operational tasks. End users such as business experts, key decision makers, data analysts, and scientists can consume these data products directly from the source systems or they can be organized, transformed, and stored in data lakes and warehouses for further processing.

The second category, *analytical data products,* are products that result from the application of simple analytics on basic data products. These products offer more dense insights from current and previous trends for both operational and strategic purposes. Self-service analytics platforms such as PowerBI (PackF, ManufO, FoodM), SAP Analytics (ManufO), Tableau (TeleC), and Mendix (ManufO) are commonly used to extract and analyze data from data warehouses. Data visualization has therefore become crucial, as it helps transform the results into an interactive, easy-to-understand view so that correct decisions can be made (Kumar and Belwal, 2017). Examples of this category could be KPIs, metrics, or dashboards used by the case companies to keep track of key parameters. These products could be combined into more comprehensive ones such as reports or dashboards that package insights together for internal or external decision support. Analytical data products can be consumed directly just like basic data products or can become a building block for more complex data products.

On the far end of the spectrum, we have *advanced analytical data products*, which are built by applying sophisticated techniques on both the basic and analytical data products to develop valuable foresights that can recommend actions and aid automated decision-making. Examples are predictive maintenance (PackF), machine learning models (TeleC), analytics algorithms (ManufO) or APIs (ManufO). Self-learning capability is key, as these products learn more by applying themselves in various use-cases, accumulating data and using it to continuously finetune their parameters (Bengfort and Kim, 2016). Data science platforms can access the basic and analytical data products stored in data lakes to apply complex analytical methods. This will help divulge hidden patterns, trends, or insights that can drive specific and critical use-cases. However, the challenge remains in acquiring people with particular skillsets to build advanced data products.

We found that the case companies have launched only a few data products in the analytical and advanced analytical area, as opposed to a large number in basic data products. This can be explained by the firms' prioritization of building a solid foundation with high-quality and well-maintained data in conjunction to adhering to the FAIR principles (findable, accessible, interoperable, and reusable). Such an approach can help acquire quick wins by demonstrating the value of using data products and then creating more complex ones when the adoption barrier has been lowered.

Interestingly, as we move from basic to advanced analytical data products, the use-cases these products can address also become more narrow. Basic data products, being fundamental in nature, are built to drive a broad range of activities and offer the possibility of being manipulated in a manner that suits user needs. However, analytical and advanced analytical data products are fit for comparatively more specific use-cases, with limited room to be reused for other purposes.

6 Summary and discussion

By shedding light on data products, we contribute to the discourse around scaling data and analytics in enterprises to repurpose and consume data efficiently and cost-effectively. From our review of literature and empirical insights, we propose a data product definition and derive six data product characteristics: *Satisfies recurring information needs; Has a well-defined consumer base; Creates measurable value for organizations; Produced through collating different data elements; Delivered in a consumable form; Has clear ownership*. The characteristics uncovered in the literature can be broadly mapped to the data product definitions from the case studies, while also extending them through highlighting the factor of ownership. Compared to previous formulations that are tailored toward specific contexts, our definition offers an overarching view with the critical elements needed to form a successful data product independent of the domain it will serve.

Another important contribution of our research are the three emerging categories of data products: while basic data products are consumed from the source systems and address a wide array of end user needs, analytics and advanced analytics data products address more specific use-cases. All the case companies have products in each of the categories, albeit much more under basic data products. Although this largely depends on the information need and the consumer base as identified in the characteristics (C1 & C2), the maturity in terms of data literacy also impact the development of data products (Sternkopf and Mueller, 2018).

Through this study, we contribute toward a broader understanding of data products. The proposed definition and characteristics, along with the three categories, will create a harmonized understanding across organizations for average users and offer a starting point to deliberate more on data product design, features, practices, and overall management (Chen *et al.*, 2022). Our study further reveals three different motivations to develop data products that can stimulate future research: First, in order to enable *data reuse and access* with data products, we need to develop suitable architecture and concepts for sharing data products with internal and external partners, for instance on marketplaces (Eichler *et al.*, 2022). Second, to make data products an effective *governance* instrument, product management approaches (Davenport *et al.*, 2022), and specifically concepts like ownership and lifecycle, need to be adapted to the different data product categories. Third, to improve *time to insight*, it could be interesting to investigate how supply chain management concepts can be applied in the creation of data products.

Our work underlines the versatility of data products – whether in terms of data products seen as end results that meet consumers' desirability, data products seen as part of the production process that ensures feasibility of the final solution, or data products as a management approach to foster the viability of strategic and operational initiatives and implications. On a practical level, our findings will help managers with enterprise-wide strategic views to identify potential value-generating opportunities with data products, average business users in running their day-to-day activities, and technical professionals to offer efficient solutions to ensure business performance.

Our study also has limitations. We only considered large multinationals as our cases, therefore the findings may not apply to more digital or smaller firms. We interviewed experienced professionals involved in data products, but it could be interesting to extend the study to other data product consumers. As data-savvy firms are inherently more data-driven, a comparative study that draws parallels between traditional and data-savvy companies around data products could be a future research focus. With our study having done the groundwork in motivating, defining, and categorizing data products, the next step could be to understand how data products should be designed and develop principles to manage data products such as around ownership and lifecycle.

References

Assur, N. and Rowshankish, K. (2022), "The data-driven enterprise of 2025", *McKinsey*, available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025 (accessed 24 September 2022).

Benbasat, I., Goldstein, D.K. and Mead, M. (1987), "The case research strategy in studies of information systems", *MIS Quarterly*, Vol. 11 No. 3, pp. 369–386.

Bengfort, B. and Kim, J. (2016), *Data Analytics with Hadoop*, 1st edition., O'Reilly, United States of America.

Burgener, E. and Rydning, J. (2022), "High Data Growth and Modern Applications Drive New Storage Requirements in Digitally Transformed Enterprises", IDC.

Cai, Y. and Ziad, M. (2003), "Evaluating Completeness of an Information Product", *AMCIS 2003 Proceedings*, presented at the AMCIS 2003 Proceedings, pp. 1–9.

Chen, P., Yang, J., Beheshti, A. and Su, J. (2022), "Towards Data Economy: Are Products and Marketplaces Ready?", *48th International Conference on Very Large Data Bases (VLDB)*, presented at the 48th International Conference on Very Large Data Bases (VLDB), Sydney, Australia, pp. 1–4.

Davenport, T.H., Bean, R. and Jain, S. (2022), "Why Your Company Needs Data-Product Managers", *Harvard Business Review*, 13 October.

Davenport, T.H. and Kudyba, S. (2016), "Designing and Developing Analytics-Based Data Products", *MIT Sloan Management Review*, Vol. 58 No. 1, pp. 83–89.

Davidson, B., Lee, Y.W. and Wang, R. (2004), "Developing data production maps: meeting patient discharge data submission requirements", *International Journal of Healthcare Technology and Management*, Vol. 6 No. 2, p. 223, doi: 10.1504/IJHTM.2004.004978.

Davis, J., Nussbaum, D. and Troyanos, K. (2020), "Approach Your Data with a Product Mindset", *Harvard Business Review*, 12 May.

Dehghani, Z. (2019), "How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh", *Martinfowler.Com*, available at: https://martinfowler.com/articles/data-monolith-to-mesh.html (accessed 24 October 2022).

Dehghani, Z. (2021), *Data Mesh - Delivering Data-Driven Value at Scale*, First edition., Vol. 1, O'Reilly, United Stated of America.

Desai, V., Fountaine, T. and Rowshankish, K. (2022), "Manage data like a product to unlock full value", *McKinsey*, Article, , available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/how-to-unlock-the-full-value-of-data-manage-it-like-a-product (accessed 24 September 2022).

Dinter, B. (2013), "Success factors for information logistics strategy — An empirical investigation", *Decision Support Systems*, Vol. 54 No. 3, pp. 1207–1218, doi: 10.1016/j.dss.2012.09.001.

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C. and Buckingham Shum, S. (2018), "Driving data storytelling from learning design", *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, Association for Computing Machinery, New York, NY, USA, pp. 131–140, doi: 10.1145/3170358.3170380.

Eichler, R., Gröger, C., Hoos, E. and Schwarz, H. (2022), "From Data Asset to Data Product – The Role of the Data Provider in the Enterprise Data Marketplace", : *: Proceedings of the 16th Symposium and Summer School On Service-Oriented Computing (SummerSoc 2022)*, presented at the : Proceedings of the 16th Symposium and Summer School On Service-Oriented Computing (SummerSoc 2022), University of Stuttgart, pp. 1–21.

Eisenhardt, K. (1989), “Building Theories from Case Study Research”, *The Academy of Management Review*, Vol. 14 No. 4, pp. 532–550.

Fadler, M. and Legner, C. (2021), “Data ownership revisited: clarifying data accountabilities in times of big data and analytics”, *Journal of Business Analytics*, Taylor & Francis, Vol. 0 No. 0, pp. 1–17.

Fricker, S.A. and Maksimov, Y.V. (2017), “Pricing of Data Products in Data Marketplaces”, in Ojala, A., Holmström Olsson, H. and Werder, K. (Eds.), *Software Business*, Springer International Publishing, Cham, pp. 49–66, doi: 10.1007/978-3-319-69191-6_4.

Fruhwirth, M., Breitfuss, G. and Pammer-Schindler, V. (2020), “The Data Product Canvas: A Visual Collaborative Tool For Designing Data-Driven Business Models”, *33RD Bled EConference Enabling Technology For A Sustainable Society*, presented at the 33RD Bled eConference Enabling Technology For A Sustainable Society, TU Graz, pp. 1–14.

Gazis, V. (2017), “A Survey of Standards for Machine-to-Machine and the Internet of Things”, *IEEE Communications Surveys & Tutorials*, presented at the IEEE Communications Surveys & Tutorials, Vol. 19 No. 1, pp. 482–511, doi: 10.1109/COMST.2016.2592948.

Gelhaar, J. and Otto, B. (2020), “Challenges in the Emergence of Data Ecosystems”, *Twenty-Third Pacific Asia Conference on Information Systems*, presented at the Twenty-Third Pacific Asia Conference on Information Systems, Dubai, p. 14.

Grover, V., Chiang, R.H.L., Liang, T.-P. and Zhang, D. (2018), “Creating Strategic Business Value from Big Data Analytics: A Research Framework”, *Journal of Management Information Systems*, Routledge, Vol. 35 No. 2, pp. 388–423, doi: 10.1080/07421222.2018.1451951.

Hariri, R.H., Fredericks, E.M. and Bowers, K.M. (2019), “Uncertainty in big data analytics: survey, opportunities, and challenges”, *Journal of Big Data*, Vol. 6 No. 1, p. 44, doi: 10.1186/s40537-019-0206-3.

Howard, J., Zwemer, M. and Loukides, M. (2012), *Designing Great Data Products*, Vol. 1, O’Reilly Media, United States of America.

Joshi, D., Pratik, S. and Rao, M. (2021), “Data Governance in Data Mesh Infrastructures: The Saxo Bank Case Study”, *ICEB 2021 Proceedings (Nanjing, China)*.

Jussen, I., Schweihoff, J., Dahms, V., Möller, F. and Otto, B. (2023), “Data Sharing Fundamentals: Characteristics and Definition”, *Hawaii International Conference on System Sciences (HICSS)*, presented at the Hawaii International Conference on System Sciences (HICSS), Hawaii, USA, p. 11.

Kumar, P. and Thakur, R.S. (2018), “Recommendation system techniques and related issues: a survey”, *International Journal of Information Technology*, Vol. 10 No. 4, pp. 495–501, doi: 10.1007/s41870-018-0138-8.

Kumar, S.M. and Belwal, M. (2017), “Performance dashboard: Cutting-edge business intelligence and data visualization”, *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, presented at the 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), pp. 1201–1207, doi: 10.1109/SmartTechCon.2017.8358558.

Lepenioti, K., Bousdekis, A., Apostolou, D. and Mentzas, G. (2020), “Prescriptive analytics: Literature review and research challenges”, *International Journal of Information Management*, Vol. 50, pp. 57–70, doi: 10.1016/j.ijinfomgt.2019.04.003.

Loukides, M. (2011), *The Evolution of Data Products*, Vol. 1, O’Reilly, Sebastopol.

Machado, I., Costa, C. and Santos, M.Y. (2021), “Data-Driven Information Systems: The Data Mesh Paradigm Shift”, *29TH International Conference On Information Systems Development (ISD2021)*, presented at the 29TH International Conference On Information Systems Development (ISD2021), AIS, Valencia, pp. 1–6.

Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I.A.T., Siddiqa, A. and Yaqoob, I. (2017), "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges", *IEEE Access*, presented at the IEEE Access, Vol. 5, pp. 5247–5261, doi: 10.1109/ACCESS.2017.2689040.

Martins de Andrade, P.R. and Sadaoui, S. (2017), "Improving business decision making based on KPI management system", *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, presented at the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1280–1285, doi: 10.1109/SMC.2017.8122789.

Meierhofer, J. and Meier, K. (2017), "From Data Science to Value Creation", in Za, S., Drăgoicea, M. and Cavallari, M. (Eds.), *Exploring Services Science*, Vol. 279, Springer International Publishing, Cham, pp. 173–181, doi: 10.1007/978-3-319-56925-3_14.

Meierhofer, J., Stadelmann, T. and Cieliebak, M. (2019), "Data Products", in Braschler, M., Stadelmann, T. and Stockinger, K. (Eds.), *Applied Data Science: Lessons Learned for the Data-Driven Business*, Springer International Publishing, Cham, pp. 47–61, doi: 10.1007/978-3-030-11821-1_4.

Mikalef, P., Framnes, V.A., Danielsen, F., Krogstie, J. and Olsen, D. (2017), "Big Data Analytics Capability: Antecedents and Business Value", *Pacific Asia Conference on Information Systems (PACIS 2017)*, presented at the Pacific Asia Conference on Information Systems (PACIS 2017), AIS Electronic Library (AISeL), Malaysia, p. 14.

Mikalef, P., Pappas, I.O., Krogstie, J. and Pavlou, P.A. (2020), "Big data and business analytics: A research agenda for realizing business value", *Information & Management*, Vol. 57 No. 1, p. 103237.

Munappy, A.R., Mattos, D.I., Bosch, J., Olsson, H.H. and Dakkak, A. (2020), "From Ad-Hoc Data Analytics to DataOps", *Proceedings of the International Conference on Software and System Processes*, Association for Computing Machinery, New York, NY, USA, pp. 165–174, doi: 10.1145/3379177.3388909.

Nam, J. and Lamb, R. (2006), "The News: Examining Perceptions of Information Product Quality (IPQ).", *AMCIS 2006 Proceedings*, presented at the AMCIS 2006 Proceedings, AIS Electronic Library (AISeL), pp. 425–429.

Ni, F., Arnott, D. and Gao, S. (2019), "The anchoring effect in business intelligence supported decision-making", *Journal of Decision Systems*, Vol. 28, pp. 1–15, doi: 10.1080/12460125.2019.1620573.

Patton, M.Q. (2014), *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*, 4th Edition., SAGE Publications.

Salaün, Y. and Flores, K. (2001), "Information quality:: meeting the needs of the consumer", *International Journal of Information Management*, Vol. 21 No. 1, pp. 21–37, doi: 10.1016/S0268-4012(00)00048-7.

Sands, E.G. (2018), "How to Build Great Data Products", *Harvard Business Review*, 30 October.

Shankaranarayanan, G., Wang, R.Y. and Ziad, M. (2000), "IP-MAP: Representing the Manufacture of an Information Product", *Proceedings of the 2000 Conference on Information Quality*, presented at the Proceedings of the 2000 Conference on Information Quality, pp. 1–16.

Shim, J., French, A., Guo, C. and Jablonski, J. (2015), "Big Data and Analytics: Issues, Solutions, and ROI", *Communications of the Association for Information Systems*, Vol. 37 No. 1, doi: 10.17705/1CAIS.03739.

Si, Y., Xiao, Q., Su, J., Zeng, S. and Hong, X. (2020), "Research on Data Product Quality Evaluation Model Based on AHP and TOPSIS", *Proceedings of the 4th International Conference on Computer*

Science and Application Engineering, presented at the CSAE 2020: The 4th International Conference on Computer Science and Application Engineering, ACM, Sanya China, pp. 1–5, doi: 10.1145/3424978.3425087.

Stadelmann, T., Klamt, T. and Merkt, P.H. (2022), “Data centrism and the core of Data Science as a scientific discipline”, *Archives of Data Science, Series A*, KIT Scientific Publishing, Vol. 8 No. 2, pp. 1–16, doi: 10.5445/IR/1000143637.

Sternkopf, H. and Mueller, R.M. (2018), “Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations”, *Proceedings of the 51st Hawaii International Conference on System Sciences*, presented at the Proceedings of the 51st Hawaii International Conference on System Sciences, Hawaii, USA, p. 10.

van der Valk, H., Haße, H., Möller, F. and Otto, B. (2022), “Archetypes of Digital Twins”, *Business & Information Systems Engineering*, Vol. 64 No. 3, pp. 375–391, doi: 10.1007/s12599-021-00727-7.

Wang, R.Y. (1998), “A product perspective on total data quality management”, *Communications of the ACM*, Vol. 41 No. 2, pp. 58–65, doi: 10.1145/269012.269022.

Wang, R.Y., Lee, Y.W., Pipino, L.L. and Strong, D.M. (1998), “Manage your information as a product”, *Sloan Management Review*.

Wang, R.Y., Pierce, E.M., Wang, R.Y., Madnik, S.E. and Fisher, C.W. (2005), *Information Quality*, M.E. Sharpe.

Wixom, B. and Ross, J. (2017), “How to Monetize Your Data”, *MIT Sloan Management Review*, Vol. 58 No. 3.

Yin, R. (2003), *Case Study Research: Design and Methods, Third Edition, Applied Social Research Methods Series, Vol 5*, Sage Publications, Inc, London, UK.

Zhou, H., Demartini, G., Indulska, M. and Sadiq, S. (2021), “Evaluating the Quality of Repurposed Data – The Role of Metadata”, *ACIS 2021 Proceedings*, presented at the ACIS 2021 Proceedings, AIS, p. 12.