# Decoding Data Products through the Lens of Work System Theory

*Completed Research Paper*

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

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

## Abstract

*Data products are seen as important levers in repurposing the large volumes of data that are captured in organizations and in satisfying the increasing demand for analytics. Despite the increasing popularity of the data product concept, it remains unclear how data can be productized. This article proposes work system theory to study the implications data products have for how organizations manage their data. Adopting a multimethod approach involving case studies and focus groups, we identify three types of data products and analyze how organizations coordinate their resources to build work systems around each data product type. Our findings contribute to the ongoing discourse on scaling data and analytics capabilities to repurpose and consume data effectively.*

**Keywords:** Data products, Work System Theory, Data management, Data culture

## Introduction

Recent forecasts predict the volume of data will expand at an annual compound growth rate of 21.2% to reach the 221 zettabyte mark by 2026 (Burgener and Rydning 2022). Such a high volume of data provides numerous opportunities through optimizing processes, accelerating innovation, and creating business value (Grover et al. 2018). However, data is predominantly stored and managed in silos which "restricts visibility across different verticals" (Patel 2019, p. 2). This approach allows for realizing function-specific, isolated use-cases and eventually leads to duplication of data per use-case scenario (Desai et al. 2022). This is not only a cost intensive practice which slows down analytics delivery (Dinter 2013) but also limits opportunities to collaborate toward developing vital analytical insight (Mikalef et al. 2017). Cross-functional teams rarely share insights, whereby decision making is hampered (Gelhaar and Otto 2020) and firms are poorly equipped to adapt to the rapidly changing business landscape (Mikalef et al. 2020).

One way of addressing these challenges is to introduce data products (Chen et al. 2022). In a seminal paper, Wang (1998) argues that organizations need to shift from the predominant data-as-a-by-product centric view toward a data-as-a-product centric view. This is enabled by a focus on data consumers, managing data with a product mindset, following lifecycle phases, and ensuring clear ownership. Data products significantly impact organizations' data operating models, which has clear implications for how data is acquired, prepared, delivered, and managed. More precisely, this shift accentuates the importance of data as a strategic resource (Legner et al. 2020) which should be managed with a product-mindset. Data products combine relevant data from various sources to support consumers in their information needs (Loukides 2011). Consequently, industry experts find that data products not only enable 90% faster use-case implementation, but also allow scaling analytics delivery while addressing various consumption patterns in organizations (Desai et al. 2022). Recently, data products have emerged as one of the four principles underpinning the data mesh concept (Dehghani 2021). This encourages decentralized data management practices which brings improved data quality and user satisfaction (Machado et al. 2021).

Therefore, data products not only improve customer gratification but also allow organizations to explore and create new revenue streams (Meierhofer et al. 2019).

The debate on data products is relatively new, so that currently the body of knowledge mainly elucidates data products using specific lenses such as data mesh (Machado et al. 2021), data science (Meierhofer and Meier 2017), and service engineering (Meierhofer et al. 2019). Despite the increasing popularity of data products, it is still not clear how data can be productized. Akin to any physical product, to create a data product requires technology and processing (Stadelmann et al. 2022), while roles, responsibilities, and a lifecycle need to be managed (Dehghani 2021) and governance is necessary to own and secure the product (Joshi et al. 2021). However, we lack a broader understanding, first, of how data products are actually built and, second, how these activities impact the way in which enterprises handle data (Chen et al. 2022). Data moving from its source to eventually become a consumable data product, is an outcome of the productization of the data. This is facilitated by blending together multiple resources – human, technical, and organizational – which, when operationalized, can affect existing data management approaches, thus demanding adaptation to key organizational strategies. By factoring in the changes data products bring we can help improve the way data within organizations could generate value (Grover et al. 2018), thereby ultimately enhancing organizational capabilities. However, very little literature testifies to empirical studies of the data productization phenomenon (Chen et al. 2022). Further, we find no proper theoretical framing to establish the academic arguments on the product-view of data. Hence, we propose the following research question:

*RQ: How do data products change the way organizations manage and use data?*

To address this question, we adopt the work system theory (WST) (Alter 2013) as our study's main theoretical framework. The WST lens is fitting due to its primary goal of creating valuable products/services for consumers through the orchestration of various resources. It departs from the technical perspective on IT systems toward a genuine system view for IT-reliant systems in organizations (Alter 2013). Therefore, it offers an established framework to explicate the productization of data by blending in key organizational elements. It covers the human aspect through *customers* and *participants*, the technical aspect through *process and activities*, *information* and *technology,* and organizational aspects through the *environment*, *strategy,* and *infrastructure* – finally instantiating data products from the interaction of these elements in the productization process through *product and service*. Based on multimethod approach (Mingers 2001) consisting of five case studies and six focus groups with representatives from large multinational organizations, we have inferred three emerging data product types in organizations: basic, analytical, and advanced analytical. Applying the WST, we first analyze how organizations coordinate various resources to build work systems around each of the data product types. By analyzing the dynamics of interaction between work system elements and identifying commonalities across the three work systems, we could derive four key effects that data products have on the way data is used and managed in organizations. Our study responds to a recent call for establishing new approaches to manage and govern data to optimize the data production, recombination, and use (Aaltonen et al. 2023). Defining different data products as work systems complements existing research on building big data and analytics capabilities (Grover et al. 2018) and provides insight into scaling these capabilities to repurpose and consume data effectively. For practitioners, our findings offer guidelines regarding the participants, processes, and technologies required to build data products in organizations and they underscore the key areas of impact.

In the remainder of the paper, we provide background on the evolution of the data product concept and the WST. Next, we outline our research methodology, following it up by applying the WST to the identified data product types found in our field data. We continue to highlight the various implications data products have for organizations, and finally, we present our conclusion and point to future research.

# Background

## *From data products to productizing data*

Data products can be defined as managed artifacts that fulfill different types of consumers' information needs by transforming, packaging, and delivering relevant data elements in a consumable form (Hasan and Legner 2023). Although earlier studies have viewed data products from a purely technical angle, seeing them, e.g., as a data science pipeline's output (Bengfort and Kim 2016) or as part of a service design process

(Meierhofer et al. 2019), recent studies have shifted toward a more customer-centric view that drives data product requirements, enables re-use of data for unknown-unknown use cases (Dehghani 2021) and further packages the products with governance elements (Joshi et al. 2021). This indicates a growing interest in how data is actually "productized", which if done well, can lead to the creation of valuable data products.

As mentioned, the newness of the debate on data products means we still know little about what "productizing data" means. Considering the literature, we distinguish four themes (Table 1): (i) production process focusing on the interaction between various resources while creating data products, (ii) the tools and platforms that support the storage and delivery of data products, (iii) data packaging that underscores key social and organizational elements, and (iv) value generation that focuses on customer satisfaction.

| Key themes | Description | Sources |
|---|---|---|
| **Production process:** Focus on interaction between the data, systems, and users to make data products | Managing information products like physical products, such as processing raw data within systems to create final information | (Wang 1998) |
| | Creating a process map to exhibit the data journey from raw state to information product for patient records | (Shankaranarayanan et al. 2000) |
| | Establishing a data production map to control information quality and process improvement for patient discharge data | (Davidson et al. 2004) |
| | Improving analytical model prediction by combining newly created data with prior data in the data science pipeline | (Bengfort and Kim 2016) |
| | Designing data products systematically to facilitate service value for customers using a service design process | (Meierhofer and Meier 2017) |
| **Tools and platforms:** Provide storage, dissemination, and management of data products | Developing a completeness metric in ERP systems to assess quality level of inventory data | (Cai and Ziad 2003; Davidson et al. 2004) |
| | Storing key metadata for order data in an information product catalog to support organizational users | (Wang et al. 2005) |
| | Proposing an architecture model to facilitate the right set of tools, such as mesh catalogs and platforms for building data products | (Machado et al. 2021) |
| | Designing platforms for creating and delivering data products to enable data sharing supported by data warehouses and lakes | (Krystek et al. 2023) |
| **Data packaging:** Encase data by providing governance, quality, and management components to foster data product usage | Enabling formulation of useful new data products by combining key data and delivering in required format | (Loukides 2011) |
| | Ensuring data products are discoverable, addressable, trustworthy, self-describing, interoperable, and secure | (Dehghani 2021) |
| | Building a comprehensive ecosystem to orchestrate design, transformation, delivery, and management of data products | (Chen et al. 2022) |
| | Managing the data product as an artifact that contains proper ownership, access, and control, and has a dedicated lifecycle | (Hasan and Legner 2023) |
| **Value generation:** Fulfill consumer needs with tailored data products | Blending skills in analytics, engineering, and communication to extract value from data products | (Meierhofer et al. 2019) |
| | Optimizing data product lifecycle model to capture the correct requirements and ensure data product's market-fit | (Davenport and Kudyba 2016) |
| | Designing a cross-functional data product canvas to conceptualize innovative business models to serve consumers | (Fruhwirth et al. 2020) |
| | Creating good data products using various data quality parameters to exhibit value in the data circulation market | (Si et al. 2020) |

**Table 1. Key themes related to data productization from prior literature**

Certain studies attending to the data product concept focused on the production process involved in transforming data into a product. Only a few papers (Davidson et al. 2004; Shankaranarayanan et al. 2000), took the traditional route of identifying and connecting the processes, storages, decisions, evaluations, boundaries, and people required to transform data into a final product. Such linear visualization allowed dedicated teams to highlight the risk areas where data quality could be diluted so that they could work back in the lineage to disclose the reasons for a data product's poor quality. However, such mapping of the data flow focused largely on the structured data objects, such as master data (Davidson et al. 2004). As the volume and variety of data exponentially increased over time, sophisticated analytical techniques appeared as mechanisms for acquiring, transforming, and delivering data as data products. Frameworks such as Hadoop (Bengfort and Kim 2016) were proposed as approaches to help create data products efficiently. However, such approaches have a strict data science focus which disregards data products' social and organizational elements.

Various studies investigated central elements in the production process, i.e., the tools and platforms needed to store, deliver, and manage data products. For instance, Wang et al. (2005) emphasized the importance of centralized catalogs that store relevant metadata to inform users on key features and characteristics of data products. Such metadata could include data quality indicators, lineage data, as well as ownership and maintenance details. However, as organizations gradually shift toward a decentralized setting, the latest concepts, such as data mesh, offer guidelines to support appropriate technological stack formation. Machado et al. (2021), for instance, highlighted emerging tools, such as self-service platforms, mesh catalogs, and mesh metadata management systems to improve the data products' discoverability and re-usability. These platforms further enable strategic goals such as data sharing between entities (Krystek et al. 2023). However, orchestrating these tools and platforms impacts the enterprise architecture's agility and often implies the need for standardization (Sukur and Lind 2022).

Building information or data products goes beyond just using tools and technologies. In fact, a blend of technical, managerial, and entrepreneurial skill is needed to turn data into useful data products (Meierhofer et al. 2019; Stadelmann et al. 2022). Analogous to physical products, data products require clear ownership and governance, quality assurance and lifecycle management (Hasan and Legner 2023). A product-centric view emphasized several characteristics, such as being discoverable, addressable, trustworthy, self-describing, interoperable, and secure, as design guidelines to productize a set of data into a data product (Dehghani 2021). Also, the governance aspect remains a central element guiding how data should be used, owned, accessed, and secured (Joshi et al. 2021) – comparable to instructions dictating how physical products should best be utilized to generate the most value.

Along similar lines of argumentation, a few other studies have focused on the value generation of data products. Besides adapting a product's lifecycle (Davenport and Kudyba 2016), authors have proposed combining multiple data elements and offering them in either overt (data is the main output) or covert (data working in the background) data products to realize user desires (Loukides 2011). Fruhwirth et al. (2020) went further to design a data product canvas to support formulating an analytics-driven business model that would build increased consumer value by addressing user concerns. Additionally, value generation can arise from directly monetizing data products by selling them through various channels (Davenport and Kudyba 2016; Si et al. 2020) or from improving internal business processes that support key enterprise objectives (Meierhofer et al. 2019).

## *Research gap*

Based on prior literature, we noted that earlier studies adopted specific lenses in trying to understand data productization. From a production-centric view that accentuates how data transforms into a product during its flow from source to destination, the focus has shifted toward a more value-oriented view concerned about end consumers. We ascribe this to the positive impact on key organizational capabilities fostered by the effective analysis and management of increasing amounts of data (Grover et al. 2018) – highlighting the growing perception of data as a strategic resource (Legner et al. 2020). Therefore, if built and managed as a product, data quality can be augmented by packaging it with governance, lifecycle, and ownership components (Machado et al. 2021). However, prior research has mainly adopted isolated lenses to understand the various data product contexts despite a growing need to study "combined methodologies, processes, enablers and innovative technologies for transforming data resources, under the effective data governance, into data products with quality assurance" (Chen et al. 2022, p. 4). Additionally, we ascertained that the existing literature lacks a clear theoretical framing of the product view on data.

The WST appears as a promising lens for elucidating how data products impact the way in which organizations manage and use data. According to Alter (2013), a work system is defined as "a system in which human participants and/or machines perform work (processes and activities) using information, technology, and other resources to produce specific products or services for specific internal and/or external customers” (2013, p. 75). Data products can be built by structuring, coordinating, and deploying various types of resources to create expected returns. Thus, the work system approach appears as a logical fit to build data products (*products/services*) to fulfill needs of end users (*customers*) by analyzing the intricacies of how human, data, and technology resources (*participants, information, and technologies*) are mobilized (through *processes and activities*). Further, the interactions between the work system elements have “direct or indirect impacts on the performance results, aspiration levels, goals, and requirements for change” (Alter 2013, p. 81) – indicating possible changes that organizations may have to undergo. WST has previously been applied to systematically analyze how enterprises organize their tangible and intangible resources and build analytical capabilities. For instance, Alter (2004) applied the work system idea to shift from an artifact-centric view on analytical systems to a broader decision support view. Heart et al. (2018) investigated how it can help design an analytics tool for clinical use, Marjanovic (2016) applied the WST lens to understand analytics-supported knowledge-intensive business processes, whereas Fadler and Legner (2020) studied how companies build capabilities for business intelligence and analytics. Compared to the latter, our interest is not only to understand how work systems build analytics capabilities, but also how the data products that address a wide range of stakeholders and their information needs are formed.

# Methodology

Our research is embedded in an industry-research collaboration on data products, where we adopt a multimethod approach (Mingers 2001) to investigate our research question. This allows us to validate data and triangulate results by using a range of methods, be creative in discovering fresh array of views and draw strong inferences within a qualitative frame (Mingers 2001). As part of the approach, we combined case studies and focus group discussions as data collection methods.

## *Data collection*

We started with an initial pool of 10 companies that were highly interested in the topic of data products. After a preliminary discussion with each firm, we did purposeful sampling (Yin 2003) in selecting five companies to be used as cases. The selected firms had inaugurated different data product initiatives and successfully created and launched multiple data products across their organizations. They were represented by 11 highly experienced data management experts who had been chosen based on their deep strategic and technical knowledge and their willingness to share details of their data product journey with us. We rejected the other five companies because they did not have any active data product program. Additionally, our final sample varied, not only in terms of data product approaches but also of the industry, size, employees, and revenues, allowing us to draw parallels and analyze differences between them. Table 2 provides an overview of the companies, participants, and data product initiatives.

| **Company** (revenue, employees) | **Participant designation & experience (in years)** | **Data product initiative goal** |
|---|---|---|
| PackageA ($1-50B, ~ 25000) | Enterprise data governance manager (30+); Service delivery manager (17+) | To support data monetization strategy |
| ManufactureB ($1-50B, ~ 80000) | Data and analytics governance manager (10+) | To streamline data governance |
| TelecomC ($1-50B, ~ 100000) | Head of data foundations and data management (24+); Lead data architect (30+) | To reduce the time-to-insight |
| FoodD ($50-100B, ~ 250000) | Master data product manager (20+); Data and analytics product manager (18+); Global analytics senior manager (17+); Business analyst manager (19+) | To harmonize data pipelines for advanced analytics |
| PharmaE ($1-50B, ~ 80000 | Operations IT lead (24+); Senior data business analyst (27+) | To facilitate data democratization |

**Table 2. Overview of the case companies**

For the case study, we performed two-hour long semi-structured interviews with the participants online through MS Teams. Before commencing, we briefed the interviewees on the purpose of the research, data confidentiality, and the interview plan. Also, all the participants gave us permission to record the interviews. The sessions were fully conducted in English and the recorded interview data was transcribed. This allowed us to document the original richness of the data for further analysis. Additionally, we gained access to internal company documents that offered practical insights on each company's data product initiative – allowing us to triangulate the data and ensure construct validity (Yin 2003). The first half of the interview explored the companies' motivation for developing data products, categories and examples, changes and their implications, as well as lessons learnt. The second half was designed to deep-dive into one successful data product in each company, to outline relevant details on its desirability (roles, responsibilities, use-cases, support structure, and delivery mechanism), feasibility (key data attributes, tasks and processes, and key resources), and viability (costs and benefits). This exercise allowed us to enrich our data with very specific and tangible input.

As our second data collection method, we conducted six focus group discussions. Table 3 provides an overview of the topics discussed. The focus groups involved a mix of the interview participants and other experts, giving us the opportunity to discuss and refine our interim findings. Two meetings were run on-site at a conference hall and four meetings were online through MS Teams. The data was audio and video recorded, respectively, with the participants' permission. Each focus group meeting was guided by a specific area chosen for discussion, allowing us to attain granular details and enhance our understanding of data products and their wider implications on the organizations. Specifically, we started with companies' challenges and priorities regarding data products. The discussion revealed persisting issues related to data silos, fragmented pipelines, high cost of data ownership, lack of data access, and slow analytics delivery. Although data products were perceived as a way of addressing these challenges, we came across significant divergence in terms of understanding what is and what is not a data product. Focus group #2 and #3 were dedicated to clarifying these concepts. Subsequently, the focus shifted to the wider enterprise architecture and the various tools and platforms required to build and deliver these data products. Eventually, the discussion shifted further, toward identifying the various issues impacting data management tasks and the technological, social, and strategic implications that can result.

| **Number** | **Date, participant # (firm #)** | **Duration (mode)** | **Discussion topics for data products** |
|---|---|---|---|
| 1 | May 2022, 18 experts (10 firms) | 2.5 hours (on-site) | Current status, challenges, and priorities |
| 2 | Aug 2022, 15 experts (8 firms) | 1.5 hours (online) | Characteristics and examples |
| 3 | Sep 2022, 12 experts (6 firms) | 2.5 hours (on-site) | Definition, categories, and lifecycle |
| 4 | Nov 2022, 20 experts (12 firms) | 1.5 hours (online) | Architecture and technologies |
| 5 | Feb 2023, 25 experts (17 firms) | 3 hours (online) | Organizational changes and implications |
| 6 | Mar 2023, 17 experts (11 firms) | 1.5 hours (online) | Management and governance |

**Table 3. Focus group details**

The first researcher undertook the main coding of the data and developed potential themes. The themes were related to the different data product types and their implications for the organizations when they build, implement, and maintain them. A second researcher undertook the independent reviewing of the coding. We then discussed the different types of data products uncovered with the participants during the focus group #4 discussion and ensured confirmation. Additionally, we mapped the case data against the work system elements and followed up by consolidating the work system elements per data product type.

## *Case analysis and theoretical integration*

To integrate our empirical insights, we used the WST in analyzing our data according to the work systems framework's and lifecycle model's components (Figure 1). Using WST as analytical framework allows us not only to accentuate how various tangible or intangible resources are orchestrated to build data products, but also to emphasize the implications they have regarding mindset change, value creation and appropriation, management techniques, process and people, and overall change management.

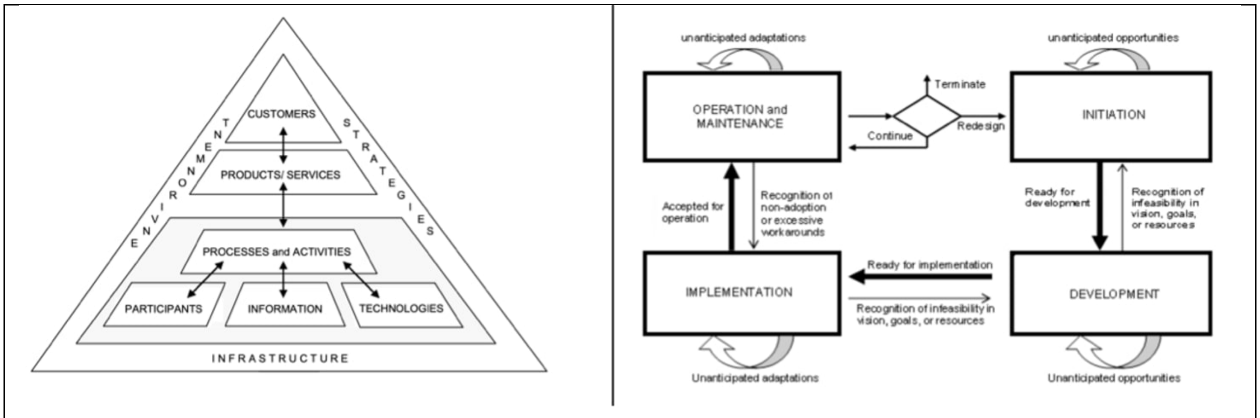


**Figure 1. Work system framework (left) and lifecycle (right)**

The work system framework helps describe and analyze IT-reliant work systems in organizations using the nine elements that must be aligned (depicted by the arrows). Alter (2013) describes the elements as follows: *customers* receive the work system's product or services for use in context other than the work system's activities; *products or services* are the information, physical things or actions a work system produces to fulfill customers' needs; *processes and activities* enable building the products or services and can be guided by clearly defined steps or human judgement; *participants* are "people who perform work within the work system, including both users and non-users of IT" (p. 80); *information* is "expressed as informational entities that are used, created, captured, transmitted, stored, retrieved, manipulated, updated, displayed, and/or deleted by processes and activities" (p. 80); *technologies* are the tools participants use, as well as the hardware or software automated agents use. These elements are the inherent components that make up the work system. However, the remaining three elements impact work systems from an external perspective. Alter (2013) describes them as follows: *environment* "includes the relevant organizational, cultural, competitive, technical, regulatory, and demographic environment within which the work system operates, and that affects the work system's effectiveness and efficiency" (p. 81); *infrastructure* constitutes all the technical, human, and informational resources that are used and shared between various work systems; and *strategies* are the alignment between the enterprise, department, and work system strategy.

The work system lifecycle represents an iterative process through which the work system evolves over time through different planned and unplanned changes (Alter 2013). The author describes the four stages of the lifecycle as follows: *initiation* addresses tasks which specify the need for a work system; *development* involves accumulating and transforming resources to build the work system; *implementation* refers to implementing the work system in the organization; and *operation and maintenance* is the stage in which the work system is monitored and sustained.

## Decoding data products with Work System Theory

### *Emerging types of data products*

We analyzed the data collected during the case studies and identified three emerging types of data product prevalent in the organizations, namely basic data products, analytical data products, and advanced analytical data products. Table 4 gives a mapping of the data products with examples from the case studies. Additionally, we highlight similar conceptualizations found in the academic literature and emphasize some interesting characteristics of these data products. Applying the WST lens, we then expand on the three data product types considering each as a work system on its own. This allows us to illuminate the resources that need to be orchestrated, showing how and by whom, to create data products that serve customer needs.

| | Basic data product | Analytical data product | Advanced analytical data product |
|---|---|---|---|
| **Characteristics** | | | |
| Description | Ready-to-use datasets to address multiple use-cases | KPIs or metrices visualized in dashboards/reports | Advanced models for prediction/forecasting |
| Conceptualization based on prior literature | Allows users to link various data attributes or datasets to gain foundational insight in the area(s) instantiated by the data and to drive multiple use-cases (Loukides 2011) | Provides users with visualized insight to facilitate decision making by applying analytical techniques to data arising from relevant business events (Davenport and Kudyba 2016) | Offers valuable predictions and foresights based on large amounts of basic and analytical data while facilitating self-learning (Meierhofer and Meier 2017) |
| Characteristics | Re-usability: high<br>Time to develop: low<br>Ease of development: high<br>Strategic orientation: low | Re-usability: medium<br>Time to develop: medium<br>Ease of development: medium<br>Strategic orientation: medium | Re-usability: low<br>Time to develop: high<br>Ease of development: low<br>Strategic orientation: high |
| **Examples** | | | |
| Data product literature | iTunes music spreadsheet (Loukides 2011); Health dataset (Chen et al. 2022) | Google Analytics; Mobile app with built-in analytics (Davenport and Kudyba 2016) | Connected homes; NLP software (Meierhofer and Meier 2017) |
| PackageA | HR data; Accounts & Hierarchy data | Composite data; Analytics report | Predictive maintenance |
| ManufactureB | Master data; Finance data; Production floor data | Tables/lists of analytical data; KPIs; Metrices | Analytics algorithm |
| TelecomC | Customer price item; Customer information | Supply chain dashboard | Algorithms; Machine learning models |
| FoodD | Master data; Commercial data foundation; Transaction data | Corporate reporting | Recommendation engines; AI models |
| PharmaE | Vendor data; Material data; Order data | Operations insight dashboard; Clinical insight | Recommendation engines; Predictive modelling |

**Table 4. Types of data products**

All the case companies have a significant number of *basic data products* due to their consumers' prior familiarity and re-use potential and the processes and workflows established to create them. The basic data product allows the users to gain a foundational understanding of the domain from which the data originates. It supports multiple use-cases based on the goals and priorities of the business without getting into the details it might contain (Loukides 2011). This reduces the need to rework and retrain and facilitates products' faster deployment to solve business problems. TelecomC, for instance, calls these basic products foundational data products but ManufactureB calls them source-aligned data products due to their data mesh initiative. The *analytical data product* enables users to monitor, control, and support decision making based on data arising from business events (Davenport and Kudyba 2016). It helps consumers gain descriptive knowledge based on historical and current data and allows them to spot relevant trends. TelecomC uses the term insight data product, PackageA the term data and insights data product, whereas PharmaE uses the term digital product. The *advanced analytical data product* offers automated decisions for consumers. It makes predictions and forecasts to enhance companies' foresights (Meierhofer and Meier 2017). FoodD refers to advanced data products, whereas ManufactureB categorizes both analytical and advanced analytical data products as consumer-aligned.

## *Basic data product work system*

This work system delivers basic data products that help users to gain an in-depth understanding and knowledge of the domain that the data represents. It facilitates diverse data exploration opportunities to fulfill users' different informational needs. To successfully address users' concerns, the data for this system, which can be structured, semi-structured, or unstructured, is gathered from multiple internal or external sources. This work system not only allows the combination and re-use of data for novel purposes; it also lays the groundwork for empowering customers through self-service analytics. Table 5 represents this work system using the work system framework and lifecycle. We have adopted the template used in Alter (2013).

<table>
<tr><th>Customers</th><th colspan="3">Product/Services</th></tr>
<tr><td>Business users<br>Analytical systems<br>Automated platforms</td><td colspan="3">Ready-to-use dataset providing foundational insight and knowledge of the domain(s) represented by the data</td></tr>
<tr><th colspan="4">Processes and Activities</th></tr>
<tr><td colspan="4">• Initiation<br>- Express need to re-use multiple datasets to drive novel use-cases: BU, BA<br>- Gather and translate business requirements into technical requisites: BU, BA, DPO [DC]<br>- Check and improve existing data models to build the right dataset: DE, DO, DS [DMT]<br>• Development and testing<br>- Source the right data from internal or external sources to meet requirements: DO, DS [OS]<br>- Provide access to respective data sources based on governance policies: DO, DE, BA [OS]<br>- Ingest required data onto staging area for final processing: DE [DOT]<br>- Profile, assess, and improve the quality of the data: DQM, DS [DQT]<br>- Transform and curate the data into a dataset and perform testing for consumption: BA, DPO [DW, DL]<br>• Deployment<br>- Create and maintain lineage and other metadata for the dataset: DS, DPO, BU [DC]<br>- Create the necessary documentations and communicate: DPO, BA [DC, DM]<br>- Provide access to the dataset to enable enterprise-wide use: DPO, BU [DM]<br>• Operation and maintenance<br>- Gather feedback, complaints, and suggestions on the dataset: BU, DPO<br>- Monitor usage, quality, and evolution of the product and continuously improve: DPO, BA, DE, BU [MT]<br>- Ensure adherence to governance policies: DPO</td></tr>
<tr><th colspan="2">Participant</th><th>Information</th><th>Technologies</th></tr>
<tr><td colspan="2">Business users (BU)<br>Business analyst (BA)<br>Data engineer (DE)<br>Data product owner (DPO)<br>Data steward (DS)<br>Data owner (DO)<br>Data quality manager (DQM)</td><td>Domain data<br>Customer requirements<br>Use-case prioritization<br>Data governance rules<br>Data contracts/SLA<br>Data lifecycle</td><td>Operational systems [OS]<br>Monitoring tools [MT]<br>Data orchestration tools [DOT]<br>Data modelling tools [DMT]<br>Data catalog [DC]<br>Data marketplace [DM]<br>Data warehouses [DW]<br>Data lakes [DL]<br>Data quality tools [DQT]</td></tr>
<tr><td colspan="4">Table 5. Basic data product work system</td></tr>
</table>

We exemplify this work system using the commercial data foundation (CDF) data product from FoodD. The CDF is an enriched dataset that serves the sales function and combines data for all sales activities conducted by both the salesforce and the retailers. The *initiation* phase began by business users outlining the key use-cases, i.e., to feed the sales recommendation engine, to do product forecasting, and to support customer investment management. The business analyst identified data required to fulfill the user needs, such as sales and financial sell-in, trade promotion planning, syndicated data, retailer and distributor data, and basic master data as mandatory. The data engineer reviewed and performed the functional and cross-functional data modeling to ensure data objects were linked. In the *development and testing* phase, the data engineer got data access from respective data owners and built pipelines to source the data using orchestration tools. Having loaded the data onto a staging area for processing, the business analyst prepared the CDF, checked for errors, and packaged it with quality assurance. During the *deployment* phase the data product owner deployed the CDF on the data and product catalog for access, entering all the required metadata, and building a communication plan to inform the business users. Together with the data steward and business analyst, the documentation necessary to meet governance obligations, was completed. In the *operation and maintenance* phase, through surveys or 1-on-1 meetings, the business users provided feedback on the CDF to the data product owner, who worked with the data engineer and business analyst to discuss improvements and plan new releases.

### *Analytical data product work system*

This work system delivers analytical data products to support decision making based on the current and past business performance. It facilitates real-time monitoring of KPIs and metrices that offer transparency at both strategic and operational levels. Data storytelling with appropriate visualization is vital in this work system because of the diverse user base with their different backgrounds, experiences, and biases. Table 6 represents this work system using the template given in Alter (2013).

<table>
<tr><th>Customers</th><th colspan="2">Product/Services</th></tr>
<tr><td>Business users</td><td colspan="2">Real-time KPIs and metrices visualized in dashboard/report for decision-support</td></tr>
<tr><th colspan="3">Processes and Activities</th></tr>
<tr><td colspan="3">• Initiation<br>- Gain consolidated views on the historical and current situation of the company: BU, BA<br>- Document the artifact specification required to support user needs: BU, BA, DPO [DC]<br>- Check whether similar data products already exist: DPO [DC, DM, DW, DL]<br>• Development and testing<br>- Identify the data sources required to fulfill analytical needs: DS, BA [OS]<br>- Establish an ETL pipeline to onboard the data into the analytics platform: DE, DA [DW, DL, BIT]<br>- Build dashboard/report prototype and gather feedback for testing and improvement: DA, BU, DPO [BIT]<br>- Incorporate all input to build the final dashboard/report: DA, BU, DPO [BIT]<br>• Deployment<br>- Create and maintain lineage and other metadata for the dashboard/report: DS, DPO, BU [DC]<br>- Create training materials to onboard users: DPO, DA, BU [DC, DM, BIT]<br>- Publish the dashboard/report to enable enterprise-wide discoverability and usage: DE, DPO, BU [DM]<br>• Operation and maintenance<br>- Gather feedback, complaints, and suggestions on the dashboard/report: BU, DPO [MT]<br>- Monitor ETL connections, usage, and quality; continuously improve: DPO, DA, DE, BU [MT]<br>- Ensure adherence to governance policies: DPO</td></tr>
<tr><th>Participant</th><th>Information</th><th>Technologies</th></tr>
<tr><td>Business users (BU)<br>Business analyst (BA)<br>Data analyst (DA)<br>Data engineer (DE)<br>Data product owner (DPO)<br>Data steward (DS)</td><td>Current and historical data<br>Customer requirements<br>Data governance rules<br>Standardization (tools, process)<br>Data contracts/SLA<br>Lifecycle (status, versions)</td><td>Operational systems [OS]<br>Business intelligence tools [BIT]<br>Monitoring tools [MT]<br>Data warehouse [DW]<br>Data lake [DL]<br>Data catalog [DC]<br>Data marketplace [DM]</td></tr>
<tr><td colspan="3">Table 6. Analytical data product work system</td></tr>
</table>

We exemplify this work system using the customer information dashboard (CID) data product from TelecomC. The CID is a dashboard within the customer domain that is driven by the changes and evolution of TelecomC customers over a certain period of time. In the *initiation* phase, the business users expressed the need to understand the evolution of customer acquisitions. The business analyst supported by the data product owner collected the user requirements in terms of the exact insights required, time horizon, drill-down/drill-up functions, and the preferred format. To avoid redundancy, the data product owner checked whether similar data products had been implemented anywhere else in the enterprise. In the *development and testing* phase, the data engineer worked to build an ETL pipeline to source required data, such as data on customers, products, proposals, and sales, from the warehouse to enter into PowerBI, which is the standardized tool at TelecomC to build dashboards. The data analyst structured, aggregated, and transformed the data into desired KPIs/metrices to be tracked on the dashboard. An initial prototype of the dashboard was evaluated with the business users and feedback was incorporated to design a final version while keeping the data product owner in the loop. The *deployment* phase involved the data engineer deploying the CID on key platforms, particularly in the cloud, to enable internal and, in some cases, external usage for business partners as well. The required level of access and licensing was determined based on governance aspects and the data steward stored all required metadata in a data catalog. In the *operation and maintenance* phase, the data product owner oversaw the dashboard and its performance, working closely with the data analyst, business users, and data engineer to improve it over time.

## *Advanced analytical data product work system*

This work system delivers advanced analytical data products that combines all the basic and analytical data to apply sophisticated machine learning or deep learning techniques that help discover hidden patterns and knowledge. This not only supports specific, business critical use-cases but also fosters automatic decisions for customers. These work systems can constantly attune the parameters to improve the models, eventually building self-learning capabilities. Table 7 represents this work system using the Alter (2013) template.

| **Customers** | **Product/Services** |
|---|---|
| Business users<br>Automated platforms | Advanced models offering prescriptive foresight and knowledge that facilitates automated decisions |

| **Processes and Activities** |
|---|
| • Initiation<br>- Highlight the need for automated decision support in complex scenarios: BU, BA<br>- Collect the requirements and prioritize them based on business value generation: BU, BA, DPO [*DC*]<br>- Check whether similar data products already exist: DPO [*DC, DM, DW*]<br>• Development and testing<br>- Identify the right data sources required to fulfill advanced analytical needs: DS, BA [*OS*]<br>- Build pipeline to onboard the data from source systems to data science platforms: DE, DSC [*DL, DW*]<br>- Create training/test sets and build a prototype for testing and feedback: DSC, BU, DPO [*CR, DT, SEN*]<br>- Finetune parameters and finalize the model for deployment: DSC, DPO, SE [*SEN*, *DT*]<br>• Deployment<br>- Create and store lineage and other metadata for the model: DS, DPO, BU [*DC*]<br>- Develop training materials for business users: DPO, DSC, BU [*DC, DL*]<br>- Design the architecture to support the deployment of the model: DAR, DE, SE<br>- Publish the model in production systems or in marketplaces for access and use: DAR, SE, DPO [*DM, DP*]<br>• Operation and maintenance<br>- Gather feedback, complaints, and suggestions on the model: BU, DPO [*MT*]<br>- Maintain and improve the model parameters to retain predictive power: DSC, DPO [*DT, MT*]<br>- Ensure adherence to governance policies: DPO |

| **Participant** | **Information** | **Technologies** |
|---|---|---|
| Business users (BU) | Current and historical data | Development tools [*DT*] |
| Business analyst (BA) | Customer requirements | Code repositories [*CR*] |
| Data scientist (DSC) | Data governance rules | Monitoring tools [*MT*] |
| Data engineer (DE) | Standardization (tools, coding) | Data lake [*DL*] |
| Data product owner (DPO) | Data contracts/SLA | Data catalog [*DC*] |
| Data steward (DS) | Enterprise architecture | Data marketplace [*DM*] |
| System engineer (SE) | Lifecycle (status, versions) | Delivery platforms [*DP*] |
| Data architect (DAR) | | Sandbox environment [*SEN*] |

**Table 7. Advanced analytical data product work system**

We exemplify this work system using the predictive maintenance (PM) data product from PackageA. The PM is deployed within PackageA production sites to enable condition-based monitoring of machines based on sensor devices. In the *initiation* phase, business users expressed the need to pre-emptively identify and fix potential faults in their production line. The business analyst and data product owner collected the requirements and prioritized them based on how much value they could generate. The selected requirements became input for the model development in the next phase. The data product owner checked whether similar models had been in live production elsewhere. In the *development and testing* phase, the data engineer created the pipelines to source and then feed the data into the lake and built a sandbox environment around it. The data scientists then took over the training/test sets' development, building and testing the model, and gathering feedback from business users. Further iterations were conducted to achieve a balance between the model accuracy and number of parameters to ensure a lean and efficient performance. In the *deployment* phase, the system engineer deployed the model in production and the data architect designed the architecture based on standard guidelines. Due to the PM complexity, the data product owner and data scientist developed training materials, and to facilitate clarity and use all the relevant metadata was stored in the catalogs. In the *operation and maintenance* phase, the model was regularly monitored using relevant tools. Data scientists constantly updated it based on new parameters which appeared as important for the PM.

From the above analysis, we realized that these work systems can function independently but also be building blocks for one another. To a large extent, they share similar participants, information, technologies and processes, as well as activities. For instance, datasets (basic data product work system) support ad-hoc data exploration activities but can also be combined and transformed into key metrices (analytical data product work system) that feed into dashboards; further, they can become parameters for the advanced models in predictive maintenance software (advanced analytical data product work system). Therefore, in many cases, one work system supports the goal and success of another work system. In fact, dedicated smaller work systems, put together, can help us perceive an entire organization as one big work system (Alter 2013). Hence, companies might want to develop enterprise-level strategies to create and manage these work systems in an integrated way instead of adopting an individualistic approach. Additional to improving resource orchestration, a holistic management approach would expedite coordinated efforts to assist organizational capability building in key areas (Grover et al. 2018). Nevertheless, building data product work systems also has implications for how organizations manage and use data.

## Key changes and implications

Building, implementing, and managing data product work systems changes the way in which companies organize themselves and handle their data. This is due to the interaction between the work system elements having "direct or indirect impacts on the performance results, aspiration levels, goals, and requirements for change" (Alter 2013, p. 81) in the organizational environment. We discuss the changes below and spell out further implications in Table 8.

**Consumer-provider relationship**: One of the central elements that emphasize the creation, use and management of data is the relationship between data producers and data consumers. Sahri and Moussa (2021) point out that data producers have extremely limited ideas of the type and criticality of use-cases in which their data could be used. Similarly, data consumers lack clarity regarding the source of the data, its transformation history, and fitness for use. Such poor vision causes misalignment of the intended purpose and actual application of the data, leading to costly data reworks and fragmented data pipelines (Desai et al. 2022). Data products help to overcome this challenge by facilitating the re-use of data. As data from producers are combined, prepared, and packaged into data products through different processes, a constant flow of reliable data is key. Hence, to support this, companies establish data contracts as an agreement between the two main parties. Data contracts guarantee a certain level of performance and availability of data from the producers so that consumers can use for various purpose including building data products. For instance, data contracts at ManufactureB guarantee a response time of a few milliseconds for an API call of their production floor data; at FoodD the data contract specifies 99.99% uptime for their commercial foundation data which is specifically meant to support business critical use-cases; and, at PharmaE data is guaranteed to be 80% FAIR (findable, accessible, interoperable, and re-usable). As the operations IT lead of PharmaE states, "*we want to automate, consolidate and democratize data products and our data FAIR rate plays an important role in communicating this to the business.*" Hence, such an approach enables building and deploying data products effectively by ensuring a desired level of performance – eventually improving the trust and interaction between the data producers and consumers (Truong et al. 2012).

**Standardized data production**: With increasing volumes of data becoming available, business users have the opportunity to explore novel use-cases by building their own data products using ad-hoc data pipelines. As Raj et al. (2020) indicated, this demonstrates the diversity of business users' consumption goals and the plethora of tools, platforms, and storages organizations have available to fulfill those goals. Such divergent approaches create burdens for the enterprise architecture by complicating the data flow, slowing down analytics delivery, and allowing disorganized maintenance. To streamline such issues, organizations aim to establish standards that offer guidelines regarding how data should be created, used, and managed to build successful and lasting data products. For instance, TelecomC developed a standard framework directing how data should be handled during onboarding, data product development, and in the operations phase. In total they cover 22 sub-areas ranging from data access to data product backlog management. FoodD is phasing out the SAP Business Objects which focuses on structured data, while making Snowflake the new standard environment for accommodating newer data types to support building and storing basic and analytical data products. ManufactureB has decided to keep their Developers Portal as the standard gateway to access data products using only API calls. As part of their enterprise data management, PackageA aims to establish data product as a standard concept in itself, which covers various critical dimensions. As their enterprise data governance manager pointed out, "*We don't have time to validate data for every need; if we have a data product, by definition it should mean that it has a proper owner, is well-cataloged and is built on high quality data*." As recent literature has highlighted, such standardization eventually enhances IT efficiency and overall organizational agility (Sukur and Lind 2022).

**Team reorganization**: Organizations have traditionally managed and disseminated data using a centralized approach. As Fadler and Legner (2021) explained, despite business users creating data and executing processes, IT departments have long been perceived as solely responsible for the data. Such a setting gives rise to bottlenecks due to the growing number of both trivial and complex requests from business users, which restricts IT teams' availability in supporting more enterprise-wide, value generating activities. Although accumulating and translating customer information needs remain paramount for building data products (Davenport and Kudyba 2016), central IT teams lack the resources needed to address these tasks. Hence, firms attempt to restructure teams, while also adapting roles and responsibilities to complement their respective data product journeys. For instance, ManfactureB established a dedicated governance team that receives and prioritizes incoming requests for data product usage and builds APIs to offer controlled access. PackageA has placed a data and analytics leader in every line of business and has split their platform and analytics team. The leader works closely with business to discern information requirements, and the platform team arranges the data for the analytics team to build the necessary data product, and eventually oversees its delivery. TelecomC took a different approach by establishing roles alongside their data product lifecycle. Their global data product lead works with a use-case squad to detect and qualify requirements in the initiation phase. The digital product owner oversees the development and provisioning of the data product, the data domain manager supports data-related tasks such as metadata management, and an information asset custodian provides access. Overall, we could disclose that no firms truly embrace a decentralizing approach. As TelecomC's data foundation head stated, "*our team structure helps us balance the creation and delivery of data products centrally with the new ideas and suggestions coming from the business regarding other data products*." Such team formations based on dedicated roles and responsibilities play a key role in creating value from data and analytics (Fadler and Legner 2021b).

**Product-oriented mindset**: Users have predominantly viewed data as a by-product of the various business events taking place across the organization. Legner et al. (2020) points out that it is mainly driven by the process-centric view which manifests due to the advent of integrated information systems. Successful completion of known business use-cases takes precedence over the proper management of their underlying data. Despite paying dividends in the short run, such a mindset will lower the trust in data due to declining quality over the years and thus impact analytics scalability to fulfill unknown-unknown use-cases. Because data gains strategic importance due to emerging phenomena such as inter-organizational data sharing (Jussen et al. 2023), viewing and managing data as equivalent to a physical product will help highlight and improve key data management areas such as quality, lifecycle, governance, or ownership (Dehghani 2021). However, nurturing this mindset among business users remains a challenge. Several initiatives have been adopted to address this. For instance, PharmaE founded a data office to educate employees on the value proposition of managing data like a product and guiding its creation using standardized processes. Further, the office is tasked with developing strategies for data product adoption across the organization and directly reports to the C-suite. TelecomC adopted a 'one-at-a-time' approach where they selected the domain with the highest data maturity to build data products there. Through securing quick wins, they aim to rally more people to the data-as-a-product concept and replicate the approach in other less mature domains. FoodD is currently seeking proposals from vendors to launch a data product literacy program to train average data users with structured training modules. Buoyed by the support from their chief data officer, FoodD's data and analytics product manager stated, "*good thing is that our CDO understand the value of product-mindset on data and our analytics team also gets it; now we want to expand this view to our business colleagues who work in the 100+ domains that we have.*" Overall, we realize that various educational programs play a key role in diffusing such challenging concepts (Sternkopf and Mueller 2018), while top management is also strongly committed to addressing these mindset changes.

| **Firms** | **Consumer-provider relationship** | **Standardized data production** | **Team reorganization** | **Product-oriented mindset** |
|---|---|---|---|---|
| Work system elements | Participants; Product/services; Customers | Technologies; Process and activities; Information | Participants; Strategies | Environment; Strategies; Infrastructure |
| Key themes in literature | Value generation | Production process; Tools and platforms; Data packaging | Data packaging; Production process | Value generation |
| PackageA | Offer 100% complete metadata for discovery | Data product-as-a-standard approach | Split of platform and analytics team | Data product policy for users and partners |
| ManufactureB | Responds in milli-seconds to internal and external users | Developers' Portal as only gateway to data products | API governance team made of users and developers | Pilot project with dashboard to track cost variations |
| TelecomC | Provide data in raw, standardized, and prepared format | Data product framework to support 22 key areas | Roles established alongside the data product lifecycle | Measure success in customer domain and rally more users |
| FoodD | Ensures 99.99% uptime for vital analytical data | Snowflake to build and manage data products | Two roles moved from LATAM office to EU analytics team | Literacy program to train 36 product managers and owners |
| PharmaE | Promises 80% FAIR data for business users and partners | Nine key metadata fields must exist for all data products | Data stewards take proxy data product ownership role | Data office to execute and track data product adoption |

**Table 8. Changes and their implications mapped to work system elements**

We mapped the changes to the themes highlighted in the literature review (Table 1). An improved consumer-provider relationship is a key driving force behind high value generation because that would foster consumers' trust and their reliance on producers' provisioning of the data and its quality. This will enable more re-use of the data and allow users to address unknown-unknown use-cases (Dehghani 2021). Similarly, higher diffusion of the product-oriented mindset could help consumers perceive the strategic importance of data and the benefits of managing it like a physical product, thereby encouraging consumers to take some responsibility for the well-being of the data in their work context. Generating such value can benefit both internal consumers, such as business users running daily tasks, or external consumers, such as business partners whose various value chain activities would be supported (Fruhwirth et al. 2020). Standardizations could drive decisions regarding which tools and platforms are to be used to manage data products, what process steps should be followed to build the products, and what metadata components would be needed to complete packaging the data product for consumption. Reforming the team could help streamline accountability at various organizational levels and reinforce the productization of data by ensuring its good governance.

Interestingly, our empirical data uncovers another key theme, namely *metrics and measures*. We observed that companies aim to establish tangible measures for determining the success or failure of data products. Such measures play a vital role in data product-related investment decisions. For instance, the data contracts in all the case companies have certain data-related KPIs that need to be tracked and updated regularly to guarantee reliable data product quality. Further, we find that concrete tracking of the diffusion of the product mindset can be done in terms of the numbers of training events completed or number of data products adopted (FoodD), whereas subjective tracking can be done using surveys, interviews, feedback, or rating (PharmaE). Hence, these findings could have implications for how teams monitor and report organizational performance, which could be driven directly or indirectly by data products.

## Conclusion and future research

By investigating data products with a work system lens, we provide insight into the productization of data. Based on our empirical insights, we identify three emerging data product types: basic, analytical, and advanced analytical and we highlight the capabilities they create for the organization. For each of these data products, we shed light on the intricacies underlying the orchestration of resources and analyze how the different work system elements interact. By analyzing commonalities across the three work systems, we derive four key implications that data products have on the way organizations manage and use data: the changing consumer-provider relationships, the standardized production processes, the reorganization of teams, and a product-oriented mindset.

Academically, to the best of our knowledge, this is one of the first attempts to elaborate on the product view of data using an established theoretical lens. Our findings contribute to the emerging academic discourse on data products, which is currently still scattered. We draw attention to the changes that data products imply from an organizational and resource perspective. By outlining different data products as work systems, our study complements existing research on building big data and analytics capabilities (Grover et al. 2018) and it provides insight into scaling data and analytics capabilities to repurpose and consume data effectively. In fact, our results complement Fadler and Legner's (2020) work that identified four broad enterprise analytics capabilities using the work system lens. However, their approach lacked the product-centric view. Hence, to connect our findings in this paper to their work, we argue that basic data products can enable *data exploration* capabilities, analytical data products can facilitate *reporting* capabilities, and advanced analytical data products can facilitate the *analytics experimentation* and *analytics production* capabilities. As an implication of our research, organizations need to rethink the building, deployment, and management of data products by combining the right resources. Further, we disclose the opportunity to create metrics and objectively measure the performance of data products. One possible way of achieving this could be to leverage the six work system elements as measurement framework and to build concrete metrices for each element mapped onto every data product type. A simple example for basic data products could be 'percentage of new users per month' (*customers*), 'net promoter score of the dataset' (*products*), 'percentage change in lead time of preparing the dataset' (*process and activities*), 'hours saved by reducing manual data preparation effort' (*participants*), 'completeness percentage of the attributes' (*information*), and 'percentage of compliant platforms used in the dataset creation' (*technologies*).

For practitioners, our study offers support in managing data as a product which enables new ways of working with data. It further identifies the resources – human, technical, and organizational – required to build data products. Additionally, the lifecycle offers a consolidated view of all the tasks and activities required to productize the data. As businesses experience dynamic changes, such an iterative approach to managing data products will ensure the concept's alignment with broader company goals and improve value realization. Further, practitioners can improve the *consumer-provider relationship* by establishing simple, non-technical data contracts in platforms such as Confluence, with easy-to-measure parameters such as quality, uptime or downtime and accurate metadata. They can ensure *standardized data production* by scanning the vendor landscape, acquiring, and formalizing the tools that best align with the approved data product strategy, and facilitate *team reorganization* by hiring new talents or developing unique career paths leading to key roles such as those of data product manager. They can also nurture the cultivation of *product-oriented mindset* by appointing a dedicated team led by a senior business leader to evangelize the importance and value of managing data as a product.

Our study does not come without limitations. First, our case companies are large, traditional multinationals and hence our findings might not be generalizable to start-ups or other innovative companies. One avenue to address this might be to include digital native firms and conduct a comparative study to draw parallels and recognize differences regarding how they create and manage data products. Second, we engaged with a limited sample of participants who are highly experienced professionals specializing in data products. This could present us with biased views that articulate senior managers' fixed perspectives. As data products are meant for wider organizational use, such limitations can be addressed in future studies by incorporating insights from operational teams, IT professionals and senior business managers who could offer process-oriented, technology focused, and strategically relevant views on data products respectively. For future research, it could be interesting to empirically measure the interactions between the work system elements in the context of data products and illuminate on which interactions play an influential role for its success as well as developing a concrete framework to guide the productization journey of data.

## References


Aaltonen, A., Alaimo, C., Parmiggiani, E., Stelmaszak, M., Jarvenpaa, S., Kallinikos, J., and Monteiro, E. 2023. "What Is Missing from Research on Data in Information Systems? Insights from the Inaugural Workshop on Data Research," *Communications of the Association for Information Systems* (53:1), pp. 4–19.

Alter, S. 2004. "A Work System View of DSS in Its Fourth Decade," *Decision Support Systems* (38:3), pp. 319–327.

Alter, S. 2013. "Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future," *Journal of the Association for Information Systems* (14:2), pp. 72–121.

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

Cai, Y., and Ziad, M. 2003. "Evaluating Completeness of an Information Product," in *Proceedings of Americas Conference on Information Systems (AMCIS 2003)*, AIS Electronic Library (AISeL), pp. 1–9.

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

Davenport, T. H., and Kudyba, S. 2016. "Designing and Developing Analytics-Based Data Products," *MIT Sloan Management Review* (58:1), 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* (6:2), p. 223. (https://doi.org/10.1504/IJHTM.2004.004978).

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

Desai, V., Fountaine, T., and Rowshankish, K. 2022. "A Better Way to Put Your Data to Work," *Harvard Business Review*. (https://hbr.org/2022/07/a-better-way-to-put-your-data-to-work).

Fadler, M., and Legner, C. 2020. "Building Business Intelligence & Analytics Capabilities - A Work System Perspective," in *Proceedings of the 41st International Conference on Information Systems (ICIS)*, Hyderabad, India, December 13, p. 2615. (https://aisel.aisnet.org/icis2020/governance_is/governance_is/14).

Fadler, M., and Legner, C. 2021a. "Data Ownership Revisited: Clarifying Data Accountabilities in Times of Big Data and Analytics," *Journal of Business Analytics* (5:1), Taylor & Francis, pp. 123–139.

Fadler, M., and Legner, C. 2021b. "Toward Big Data and Analytics Governance: Redefining Structural Governance Mechanisms," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, Hawaii, USA.

Fruhwirth, M., Breitfuss, G., and Pammer-Schindler, V. 2020. "The Data Product Canvas: A Visual Collaborative Tool For Designing Data-Driven Business Models," in *Proceedings of the 33rd Bled EConference Enabling Technology For A Sustainable Society*, TU Graz, pp. 1–14.

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* (35:2), Routledge, pp. 388–423. (https://doi.org/10.1080/07421222.2018.1451951).

Hasan, M. R., and Legner, C. 2023. "Understanding Data Products: Motivation, Definition, And Categories," in *Proceedings of the Thirty-First European Conference on Information Systems (ECIS 2023)*, Kristiansand, Norway.

Heart, T., Ben-Assuli, O., and Shlomo, N. 2018. "Using the Work System Theory to Bring Big Data Analytics to the Inpatient Point of Care," in *Proceedings of the 26th International Conference on Information Systems (ICIS)*, San Francisco, CA, pp. 1–9.

Joshi, D., Pratik, S., and Rao, M. 2021. "Data Governance in Data Mesh Infrastructures: The Saxo Bank Case Study," *ICEB 2021 Proceedings (Nanjing, China)*. (https://aisel.aisnet.org/iceb2021/52).

Jussen, I., Schweihoff, J., Dahms, V., Möller, F., and Otto, B. 2023. "Data Sharing Fundamentals: Characteristics and Definition," in *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, USA, January 1.

Krystek, M., Morzy, M., Mazurek, C., and Pukacki, J. 2023. "Introducing Data Mesh Paradigm for Smart City Platforms Design," in *Proceedings of the 56th Hawaii International Conference on System Sciences*, Hawaii: AIS, pp. 1–8.

Legner, C., Pentek, T., and Otto, B. 2020. "Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management," *Journal of the Association for Information Systems* (21:3), pp. 735–770.

Loukides, M. 2011. *The Evolution of Data Products*, (Vol. 1), United States of America: O'Reilly Media.

Machado, I., Costa, C., and Santos, M. Y. 2021. "Data-Driven Information Systems: The Data Mesh Paradigm Shift," in *Proceedings of the 29th International Conference On Information Systems Development*, Valencia, p. 6.

Marjanovic, O. 2016. "Improvement of Knowledge-Intensive Business Processes Through Analytics and Knowledge Sharing," in *Proceedings of the 24th International Conference on Information Systems (ICIS)*, Dublin, Ireland, December 11, pp. 1–19.

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

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

Mingers, J. 2001. "Combining IS Research Methods: Towards a Pluralist Methodology," *Information Systems Research* (12:3), pp. 240–259. (https://doi.org/10.1287/isre.12.3.240.9709).

Raj, A., Bosch, J., Olsson, H. H., and Wang, T. J. 2020. "Modelling Data Pipelines," in *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, , August, pp. 13–20. (https://doi.org/10.1109/SEAA51224.2020.00014).

Sahri, S., and Moussa, R. 2021. "Customized Eager-Lazy Data Cleansing for Satisfactory Big Data Veracity," in *25th International Database Engineering & Applications Symposium*, Montreal QC Canada: ACM, July 14, pp. 157–165. (https://doi.org/10.1145/3472163.3472195).

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

Si, Y., Xiao, Q., Su, J., Zeng, S., and Hong, X. 2020. "Research on Data Product Quality Evaluation Model Based on AHP and TOPSIS," in *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, Sanya China: ACM, October 20, pp. 1–5. (https://doi.org/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* (8:2), KIT Scientific Publishing, pp. 1–16. (https://doi.org/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," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, Hawaii, USA.

Sukur, A., and Lind, M. L. 2022. "Enterprise Architecture to Achieve Information Technology Flexibility and Enterprise Agility," *International Journal of Information Systems and Social Change (IJISSC)* (13:2), IGI Global, pp. 1–20. (https://doi.org/10.4018/IJISSC.298337).

Truong, H.-L., Comerio, M., De Paoli, F., Gangadharan, G. r., and Dustdar, S. 2012. "Data Contracts for Cloud-Based Data Marketplaces," *International Journal of Computational Science and Engineering* (7:4), Inderscience Publishers, pp. 280–295. (https://doi.org/10.1504/IJCSE.2012.049749).

Wang, R. Y. 1998. "A Product Perspective on Total Data Quality Management," *Communications of the ACM* (41:2), pp. 58–65. (https://doi.org/10.1145/269012.269022).

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

Yin, R. K. 2003. *Case Study Research: Design and Methods*, (4th ed., Vol. 5), London, UK: SAGE Publications.