Article

Data-Driven Enterprise Architecture for Pharmaceutical R&D

Nailya Uzhakova (née Sabirzyanova) * and Stefan Fischer [ID]

Institute of Telematics, University of Lübeck, 23562 Lübeck, Germany; stefan.fischer@uni-luebeck.de
* Correspondence: nailya.sabirzyanova@student.uni-luebeck.de

Abstract: This paper addresses the research gap in the realm of data-driven transformation by leveraging the Resource-Based View (RBV) theory and the dynamic capabilities concept to the contours of a data-driven enterprise. It confronts the limitations of conventional digital and data transformation programs, which often prioritize technological enhancements over crucial organizational and cultural shifts. Proposing a more holistic perspective, the Data-Driven Enterprise Architecture Framework (DDA) is introduced, emphasizing the domain decomposition and productization of an architecture, distributed ownership, and federated governance, while ensuring the continuous harmonization of data, application, and business architecture. A case study featuring a leading pharmaceutical company illustrates the practical implementation of the DDA framework as a pillar of their Digital Transformation Strategy. By integrating scalable and distributed data architecture into the overarching Enterprise Architecture landscape, the company has initiated their data-driven transformation journey, showcased through their initial and very early results. This research not only offers valuable insights for pharmaceutical organizations navigating the complexities of data-driven transformations, but also addresses a research gap in the field.

Keywords: data-driven enterprise; enterprise architecture; dynamic capabilities; clinical trials; pharmaceutical R&D; data fabric; data mesh; data domains; data products; data ownership model; federated governance

1. Introduction

1.1. Data-Driven Enterprise. Why? What? How?

1.1.1. Why Become a Data-Driven Enterprise?

In today's digital era, disrupted by the emergence of Big Data; AI; Internet of Things; Robotic Process Automation; and Augmented, Virtual, and Mixed Reality, data have become a critical resource and a top priority for many companies [1–4]. The value of data has significantly increased over the years, with nearly 90% of an enterprise's value now attributed to intangible assets, including data [5]. Furthermore, digital companies including Apple, Microsoft, Alphabet, Amazon, Tesla, Meta (Facebook), and NVIDIA were in the top 10 of the world's biggest companies, as of 2022 [6].

The cost of data breaches is also on the rise, with the healthcare industry being the most affected, with an average cost of USD 10.1 million in 2022 [7]. Studies and white papers have also shown a direct correlation between a firm's productivity and the intensity of data leverage [8–12]. In addition, an extensive literature review conducted by Schweikl et al. reveals the prevailing positive influence of IT investments on firms' productivity levels [13].

With data evolving from merely a potential competitive advantage to a vital necessity for seamless integration, sustainable growth, and market survival, enterprises worldwide are increasingly driven to become "data-driven", embedding this into their business strategies. The pharmaceutical industry (Pharma) exemplifies this trend, where the implementation of cutting-edge digital technologies such as AI, robotics, Big Data analytics, cloud computing, embedded systems, adaptive manufacturing, and the Internet of Healthcare

Things (IoHT) is encapsulated within the concept of Pharma 4.0 [14]. This paradigm not only aims to optimize pharmaceutical operations, but also to fundamentally transform the entire pharmaceutical value chain.

A pivotal catalyst for the global surge in the adoption of Pharma 4.0 technologies was the outbreak of COVID-19, making the pharmaceutical industry a realm of rapid transformation in 2020 [15,16]. In response to the urgent need to manage the extensive spread of the pandemic and avert the loss of millions of lives worldwide, pharmaceutical companies swiftly mobilized to accelerate the implementation of digital technologies. For instance, companies engaged in the development of vaccines embarked on innovative approaches, leveraging the decentralization and virtualization of clinical trials to accelerate research efforts, while mitigating the burden on hospitals and minimizing risks to trial participants [17]. Furthermore, a diverse array of statistical models and data analytics tools were deployed to predict and monitor COVID-19 cases across different countries [18]. By harnessing the power of data analytics, pharmaceutical companies could anticipate outbreaks, allocate resources effectively, and inform strategic decision-making in real-time, thereby enhancing the global response to the pandemic.

The potential impact of Pharma 4.0 is underscored by numerous studies, emphasizing its capacity to enhance efficiency, accelerate time-to-market, mitigate risks, elevate product quality and safety standards, foster innovation, and facilitate modernization [14–21]. Moreover, Pharma 4.0 is increasingly recognized as a pivotal enabler for achieving sustainable development objectives within the industry, aligning closely with strategic management agendas [15].

Within pharmaceutical research and development (R&D), data assume a paramount role as the cornerstone of clinical development. The seamless flow of data across the lifecycle—from collection and processing to sharing, analysis, and regulatory submission—is indispensable for establishing the safety and efficacy of pharmaceutical products. Consequently, the effectiveness of pharmaceutical R&D organizations in managing and leveraging data significantly influences their overall business performance. By embracing data-driven approaches, they can unlock numerous potentialities, as follows:

- Accelerated Drug Discovery and Design: Leveraging machine learning algorithms and AI, data-driven R&D methodologies enable pharmaceutical companies to model and analyze vast datasets encompassing genetic, molecular, and clinical information. These approaches streamline the drug discovery process by uncovering hidden patterns, identifying novel drug targets, and predicting therapeutic outcomes, thus reducing time and resource requirements. Thus, several studies have demonstrated promising outcomes in such areas as drug molecular design, retrosynthetic analysis, chemical reaction outcome prediction, adverse events detection, virtual screening, peptide synthesis, biomarker discovery, and others [22–30].
- Precision Medicine and Targeted Therapies: By harnessing comprehensive patient data, including genetic profiles, biomarkers, and clinical histories, pharmaceutical firms can develop personalized treatment regimens tailored to individual patient characteristics. Data-driven insights facilitate the identification of the patient subpopulations likely to respond positively to specific therapies, enabling the development of targeted and more efficacious treatments [29,31–33].
- Optimized Clinical Trials: Data-driven analytics enhance the design and execution of clinical trials, enabling pharmaceutical companies to identify suitable patient cohorts, optimize trial protocols, optimize site-selection process, increase trial participant recruitment, and maintain engagement [24,34].
- Drug Repurposing and Combination Therapies: Data analytics enable pharmaceutical researchers to explore existing datasets and identify opportunities for drug repurposing and combination therapies [35,36]. By leveraging insights from large-scale genomic, transcriptomic, and phenotypic datasets, researchers can uncover novel indications for existing drugs or synergistic combinations with enhanced therapeutic efficacy.

- Pharmacokinetics and Pharmacodynamics: By implementing data-driven approaches including AI methods to pharmacovigilance, pharmacodynamics, and adverse event monitoring, companies can detect and respond to safety concerns faster, more cost-effectively, and with greater accuracy, decreasing regulatory risks and safeguarding patient welfare [37–40].

In essence, the adoption of data-driven approaches holds immense potential for transforming pharmaceutical R&D, driving innovation, enhancing efficiency, and, ultimately, improving patient outcomes.

1.1.2. What Does Data-Driven Enterprise Mean?

Despite a growing firm's aspiration to become a data-driven organization, postulating it in their business strategies [41–43], the term "data-driven enterprise" has gained limited scholarly attention [43–45]. In contrast with a discernible scientific research gap, it has found extensive discourse in the business literature [46–51]. Commonly identified characteristics of a data-driven organization include recognizing the strategic significance of data and fostering a culture that harnesses data for informed decision-making [44,45,52–55]. However, defining a data-driven enterprise is no straightforward task, due to the dynamic evolution of the "data-driven" concept, coupled with rapid transformations in both the business and technological landscapes [52]. What was considered "data-driven" just a few years ago is now deemed outdated, complicating the articulation of clear goals and objectives for transformative programs and contributing to corporate skepticism regarding the outcomes of data and digital transformation initiatives [45,52].

Building upon mainstream studies of digital transformation, we employ the Resource-Based View (RBV), a well-established theory in strategic management, to delineate the contours of a data-driven enterprise and data-driven transformation. The RBV conceptualizes organizations as collections of resources shaping their competitive advantage [56–60]. This perspective categorizes resources into assets and capabilities, with the latter facilitating the utilization of assets to achieve defined outcomes [58,60]. Extending the RBV to the realm of information technologies [60], we focus on applying the RBV framework to data, as a strategic resource of a data-driven enterprise. In this context, data assets encompass various elements such as datasets, metadata, data lineage, data products, data engineering and analytics skills, data platforms, data architecture, infrastructure, standards, and more. Possessing these data assets alone is insufficient for gaining a competitive advantage; an organization must cultivate and evolve its data-driven capabilities. Consequently, we define data-driven capabilities as the ability to deploy data assets, either independently or in conjunction with other organizational resources, to support and enhance business strategies and processes. It is crucial for firms to align these capabilities with their overall business strategies, emphasizing that a transformation in how organizations utilize their data assets is directed to maximize overall business value for a firm.

Thus, within the RBV framework and drawing on the extensive experience in digital and data strategy consultancy projects of the first author, we propose a data-driven enterprise as an organization continually refining and enhancing its data-driven capabilities to achieve and sustain competitive advantages. Examples of such types of data-driven capabilities include data management and governance, analytics, data-driven decision-making, data ingestion and integration, data storage, data cost optimization, development of strategic alliances and partnerships, investment in new data technologies, and others. Our definition highlights that a data-driven enterprise is not a static term or a "snapshot", but rather an organization that continuously evolves its data-driven capabilities to address a rapidly changing business environment. Therefore, a data-driven enterprise is highly dependent on an organizational ability to react and adapt to ecosystem changes efficiently, which is often called "agility" or "strategic renewal of an organization's business model, collaborative approach, and eventually the culture" [52].

1.1.3. How to Become a Data-Driven Enterprise?

The transition toward becoming a data-driven enterprise is an enduring transformative process, commonly identified as a data-driven transformation. This evolution is commonly achieved within the strategic programs of a company, entailing a continuous journey. Digital and data-driven transformations are interconnected realms, both rooted in recent technological disruptions, exploring how enterprises respond to these changes—utilizing digital and data resources, respectively. While the theory of digital transformation is well-established in academic research [42–45,52–54], characterizing it as a reconfiguration of how firms leverage digital technologies to innovate business models and enhance operational processes [52–54], the concept of a data-driven transformation has recently gained prominence, allowing the employment of analogous research methods and concepts to explore how organizations evolve their data-driven capabilities.

To define data-driven transformation, we draw upon the dynamic capabilities theory that originates from the RVB and focuses on continuous adaptation to a changing environment and building a competitive advantage [55]. Thus, the RBV defines two categories of capabilities, as follows: ordinary and dynamic. Ordinary capabilities facilitate a firm's current operational, administrative, and governance processes, while dynamic capabilities reconfigure internal and external competencies to drive transformative changes in future operational modes, influencing the pace of change in ordinary capabilities [52,55]. Being idiosyncratic, dynamic capabilities are challenging to replicate and necessitate intentional development [61]. Drawing a parallel in pharmaceutical R&D, ordinary capabilities encompass tasks such as clinical study capacity planning or the creation of regulatory submission data, focusing on the status quo operations and business model. In contrast, dynamic capabilities involve transformative changes, like the development and implementation of a decentralized clinical trial model or the formation of strategic alliances with research institutions, contributing to gaining a competitive advantage.

Consequently, we define data-driven transformation as the process of establishing dynamic data-driven capabilities, strategically positioned to respond efficiently and exploratively to changes in the business environment. Thus, a company should intentionally deploy data assets to explore, investigate, and comprehend both internal and external changes. This involves assessing potential outcomes and risks associated with various strategies to address these changes, ultimately enabling the organization to make well-informed decisions that align with their desired goals in a timely and qualitative manner. Beyond facilitating efficient decision-making through data resources, data-driven transformation encompasses the initiation and refinement of novel data-driven business models or the enhancement of existing ones. Our definition emphasizes that data-driven transformation is not merely a set of capabilities to be implemented, but the development and evolution of a firm's capability to explore its business environment. This includes the ability to react promptly to changes, identify emerging opportunities, and respond to them with agility, utilizing both data resources and competencies. Consequently, our approach (see Section 3.2) places paramount importance on establishing an organizational framework and a corporate culture that fosters responsiveness. In this context, the evolution of data-driven capabilities becomes a continuous and intrinsic part of the organizational journey, adapting to the ever-changing dynamics of the business landscape.

Delving deeper into the dynamic capabilities theory, Teece categorizes dynamic capabilities into the following three groups: sensing, seizing, and reconfiguring [62]. Sensing involves identifying opportunities and threats in the business environment, seizing focuses on extracting value from these opportunities, and reconfiguring entails the continuous transformation of organizational routines and resources for sustainable growth [53,63]. This classification is widely acknowledged in the literature, with applications extending to the domain of digital transformation [52,53,55–57,64,65]. Proposing an alternative typology for data-driven transformation, we aim to provide companies with a structured approach to progress through their data transformation journey. We define three levels of data-driven capabilities that companies progress along to embrace data.

At the first "Supportive" level, data support strategic and operational decisions, replacing reliance on instincts or outdated practices. If we compare this with the typology of Teece, the supportive level most closely relates to the enhancement of ordinary data-driven capabilities. Companies should leverage data to explore and understand customers and customer journeys, business weaknesses and competitive advantages, and both partner and ecosystem trends, and strategically act to improve customer experience, optimize financials, and position the company in the market. Referring to Pharma, initiatives here can vary from data quality and interoperability improvements, such as harmonizing International Organization for Standardization (ISO) Country Codes across systems, to advanced AI integration for disease analysis and treatment efficacy, e.g., AI analytics of clinical images.

The next level of data-driven capabilities is "Transformational". Here, data assets transform business models, enabling new digital products and services. For example, decentralized clinical trials, enabled by new ways of collecting data from participants, could potentially decrease site burden, improve the participant journey, reduce costs, enhance data accuracy, and result in boosting participant recruitment and retention rates. The transformational level closely relates to the dynamic type of capabilities.

At the third "Accelerative" level of data-driven capabilities, companies participate in digital ecosystems, exchanging data with various sources, such as electronic health records, clinical trial data, wearables, and connected devices. By exchanging data with other organizations within the healthcare digital ecosystem or related ecosystems like insurance providers, hospitals, and research institutions, pharmaceutical companies can accelerate the outcomes achieved in the previous data-driven levels. Promising examples of digital ecosystems include tokenization engine platforms like DATAVANT or HealthVerity, which offer healthcare data marketplaces that allow consumers to connect their patient data with third-party data in a privacy-preserving and secure manner. As more companies join these platforms and exchange data, the pool of available data grows, making it easier to identify patterns and insights that were previously hidden. This, in turn, can lead to more effective treatments and better patient outcomes.

In addition to digital ecosystems, generative AI models like Chat GPT also play a significant role in the "Accelerative" level of data-driven capabilities. These models have the potential to revolutionize enterprise decision-making processes and accelerate numerous business operations. For example, McKinsey estimated that generative AI could contribute roughly up to USD 110 billion in additional value for the Pharma industry, with the highest impact on the Pharma R&D function [66]. While many existing generative AI models are open source and trained on publicly available datasets, their true value is realized when they are exposed to internal enterprise data or other non-public digital ecosystem data.

For instance, in the context of study protocol development, open-source AI models can provide consolidated knowledge about clinical trial regulatory requirements, phases, terms, best practices, and so on. By integrating internal data or Real World Data (RWD) from non-public sources (e.g., Health Data Marketplaces), these models can generate detailed responses tailored to specific requests, such as: "Create a Phase 2 trial protocol for a Molecule XYZ for the treatment of Sjogren's Syndrome. Please use the outcomes of pre-clinical and 0–1 phases of the trials as well as all planning and forecasting, feasibility studies outcomes and preparation materials". While human experts still play a crucial role in reviewing, refining, and adding domain-specific expertise to the generated protocol, the AI-powered approach significantly reduces the time and effort involved in protocol development.

Further evidence of the acceleration of a drug design process is a generative AI framework for retrosynthesis prediction, named G2Retro. Retrosynthesis involves transforming a target molecule into potential reactants, to identify synthesis routes. Studies have shown that compared to manual estimation methods, G2Retro improves the accuracy and speed of identifying reactions that work best for creating a given drug molecule [67].

These examples highlight how data-driven decision-making facilitated by generative AI models can accelerate various processes within clinical drug development. This acceleration has profound implications, potentially reducing the time patients wait for life-saving treatments and alleviating the burden on physicians, researchers, and trial teams.

By embracing the transformative potential of AI and data-driven capabilities and placing a strong emphasis on upholding ethical standards, data privacy, and security, organizations can unlock new opportunities, drive innovation, and make a positive impact in the healthcare industry and beyond. The shift towards a data-driven future is not only advantageous but also inevitable in today's digital landscape. Therefore, it is essential for companies to proactively invest time and resources in understanding the implications, developing the necessary skills, and establishing robust frameworks to ensure a smooth and responsible transition into the data-driven future. This proactive approach will potentially position the company as a leader in the industry. On the other hand, failing to embrace this transformation will leave the company as a follower, subject to the disruptive changes driven by the industry.

Notably, the specific order and prioritization in which a company progresses through the levels may vary depending on the synergy between data-driven capabilities, goals, resources, ecosystem readiness, and industry context. For example, some pharmaceutical companies may prioritize the "Acceleration" level by extending their presence in digital ecosystems early on, as the benefits of a connected data ecosystem can be particularly impactful in the Pharma industry. On the other hand, some Pharma companies may slow down in external digital collaboration, due to regulatory constraints and cultural acceptance in specific markets (such as in Germany, France, and Poland) [68].

Three data-driven levels are not necessarily sequential stages of development, but rather reflect differing levels of effort and expected value. At the Supportive level, data are leveraged to automate or improve existing business processes, which often require less investment compared to the Transformational level. Transformational changes entail reshaping the company's business model or introducing a new digital product to the market, which requires more significant investments in terms of both time and resources.

It is crucial to emphasize that the successful implementation of data-driven capabilities at each level necessitates a solid foundation, which is not solely dependent on technology, but rather driven by a data-driven culture and organizational change [69]. Establishing this foundation requires a substantial amount of effort, commitment, time, and readiness to change, which is comparatively greater than the immediate outcomes.

The true value of data-driven transformation may only be realized over time and, as a company evolves through the data-driven capability levels, it will generate a higher return on investments (ROIs) compared to initial outcomes. This is why many companies that embark on their digital transformation journey may not be content with the outcomes and have become increasingly skeptical in recent years [4]. However, until this foundation is established, the promising benefits of leveraging data cannot be fully realized. The next chapter will detail how modern data-driven Enterprise Architecture looks and how it can establish a solid foundation for successful data-driven transformation.

1.1.4. Challenges on the Journey towards Becoming a Data-Driven Enterprise

Despite the growing recognition of the potential benefits of leveraging data to improve clinical trial operations, facilitate decentralization of clinical trials, reduce the burden on participants and sites, and foster collaboration within and outside the healthcare digital ecosystem, many organizations have struggled to achieve success in their data-driven transformation initiatives.

Thus, a McKinsey survey found that most organizations could not achieve even one-third of the expected outcomes from their data-driven initiatives [70]. Similarly, a report from the BSC revealed that only 30% of digital transformation programs were considered successful, while 44% were reported as having a neutral or mixed result, and 26% were reported as failures [71]. Moreover, a study conducted by NewVantage Partners, starting

in 2019, has revealed that the percentage of companies that have been identified as "data-driven" is decreasing, from 37.8% in 2020 to 23.9% in 2023, and the establishment of a data culture within organizations has also been declining, from 26.8% in 2020 to 20.6% in 2023 [4].

It can be argued that the decreasing statistics reflect the rapid evolution of the expectations of data-driven transformation, which have outpaced the outcomes of investments in this area. Nevertheless, the findings of studies highlight significant challenges that organizations face in becoming truly data-driven. The literature mentions common barriers such as a lack of clarity around goals and objectives, cultural resistance to change, insufficient data quality, inadequate data governance and security, and a lack of data literacy and expertise among staff [72,73].

However, these challenges are only the tip of the iceberg. This paper argues that a data-driven Enterprise Architecture is an essential prerequisite for any digital transformation journey and could be compared with a nervous system, with data as nervous impulses circulating within the whole enterprise organism and accelerating its performance. Therefore, any fragmentary initiatives within an organization to improve data integrations, enhance data management and decision-making, eliminate data silos, advance data mastering, or implement data fabric toolkits will have limited outcomes without a comprehensive data-driven architecture in place. Data-driven Enterprise Architecture (DDA) is a comprehensive framework that encompasses how enterprise data are to be utilized and perceived, for example as a sub product of systems and business processes, or, conversely, as a "first-class citizen", being a critical asset or product, and so on. It defines how the data are to be organized, managed, and governed to achieve specific business goals. This strategic foundation then leads to the selection of target-oriented resources, such as technology, tools, infrastructure, and skills, that support the framework, automate processes, and augment its capabilities.

A data-driven architecture is a nervous system of the entire digital transformation, with data as nervous impulses circulating within the whole enterprise organism and accelerating its performance.

Revisiting the lens of dynamic capabilities theory, numerous studies have delved into the impact of Enterprise Architecture (EA) on a firm's dynamic capabilities, showcasing that EA enhances efficiency in managing digital technologies and IT investments and significantly influences business value [74–78]. Additionally, these studies observe that EA-driven capabilities serve as crucial enablers for operational or ordinary capabilities. Some scholars argue that EA-driven capabilities contribute significantly to agility, adaptability, and innovativeness, aligning enterprise resources with changing business needs and the environment [79–81]. Consequently, these capabilities should be positioned at the core of digital transformation.

Despite the existence of several empirical studies estimating the positive impact of EA and EA-driven capabilities on digital and data-driven competencies, dynamic capabilities, innovativeness, operational and IT efficiency, and overall agility, the realization of their value remains a highly debated topic [74,75,77,78,81]. Our extensive consulting experience with digital and data transformation programs shows that the value realization of the DDA is not binary, but exists on a continuum, making it exceedingly difficult to distinguish the specific benefits attributable to each element of the DDA. The benefits can be quite opaque in comparison to a system implementation or process improvement that has immediate tangible effects. This, in turn, complicates not only convincing stakeholders to commit, but also defining metrics and tracking the success of the transformation on this foundational phase. Nevertheless, a DDA is supposed to build a basis for data-driven transformation, accelerating, harmonizing, and streamlining the outcomes of all subsequent data-driven investments.

The rest of the paper is structured as follows: Section 2 presents the DDA framework alongside the evolution of the Enterprise Architecture concepts. Section 3 provides insights of implementation of the framework for pharmaceutical R&D organization at a

Pharma company that includes their background and motivation, applied methodology of design and implementation phases, as well as preliminary results, limitations, and recommendations. Section 4 concludes the work with key findings.

2. Materials and Methods

2.1. Enterprise Architecture: Definition and Evolution of the Concept

> *Why is centralized and layered Enterprise Architecture outdated?*

Enterprise architecture (EA) is a discipline that has evolved over time to align an organization's IT systems with its business strategies and objectives [82]. The origins of EA can be traced back to the 1960s, when the field of information systems was emerging as a discipline. The need for a holistic approach to manage and integrate IT systems within organizations led to the development of the first generation of EA frameworks. These frameworks, including the Zachman framework [83], the Open Group Architecture Framework (TOGAF), the NIST Enterprise Architecture Model, and the Federal Enterprise Architecture Framework (FEAF) [84], provided a structured approach to design, manage, and align IT capabilities with business needs via IT centralization and a layered EA approach. Figure 1 presents an original depiction of a Centralized Layered Enterprise Architecture, providing an aggregated view of the first generation of EA frameworks. It illustrates centralized layers such as Business Architecture, Application Architecture, Data Architecture, and Technical Architecture, as well as core functions or teams that support or own these layers.
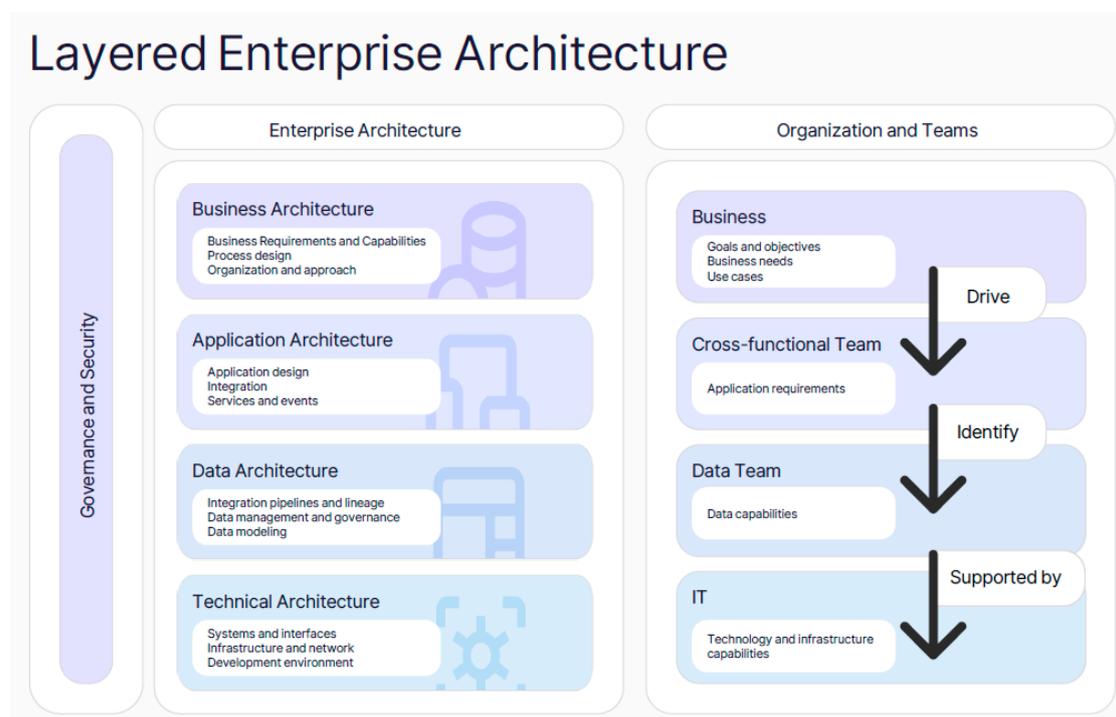

Figure 1. Centralized Layered Enterprise Architecture (EA)—aggregated view of the first generation of EA frameworks, created by the authors.

As EA has always been evolving alongside demanding business requirements and the evolution of technologies, a radical change of the technological landscape and business environment over the last decades has emphasized the limitations of these frameworks. First, the centralized and layered architecture is often inflexible and does not allow for the rapid implementation of new digital capabilities, like cloud computing technologies and the Internet of Things (IoT), which are becoming increasingly important for organizations to remain competitive. Second, it may not be able to support the complex, dynamic, and distributed IT requirements of modern organizations, especially in handling the three

main data challenges—volume, velocity, and variety. Third, the traditional approach often results in siloed systems that do not effectively communicate and integrate with each other, leading to redundancy and inefficiency, as well as inhibiting the transformation to a data-driven organization.

Moreover, we argue that information silos are a consequence of inefficient EA. Silos originate when the organizational environment does not provide the required capabilities for efficient collaboration and information sharing between domain areas. As a result, domains that have a focus on achieving their goals may not be motivated to put in extra effort and resources to ensure discoverability and accessibility of their domain information across an organization, or they do not realize if other domains could potentially benefit from the data that they hold. For instance, someone working in participant recruitment may not realize that making the data available for consumption can improve the work of a feasibility selection or vice versa. This leads to a tendency for projects and domains to optimize locally, ignoring the objectives of unlocking their silos (e.g., by building point-to-point integrations instead of public or private APIs). Another fundamental reason for enterprise silos could be the lack of the necessary cultural, organizational, and technological baselines coming from the top.

To successfully navigate the complexities of data-driven transformation, organizations need to re-evaluate their traditional EA approach and adopt a more flexible and adaptable framework. This necessitates a shift away from the traditional centralized and layered approach to a more modular and decentralized architecture that can quickly integrate modern technologies and capabilities.

2.2. The Data-Driven Enterprise Architecture framework

In this chapter, we introduce the Data-Driven Enterprise Architecture (DDA) framework, which we have developed as a solution to harmonize emerging trends in Enterprise Architecture (EA), such as decentralization, modularization, democratization, socialization, product thinking, domain-driven design (DDD), and agility, while emphasizing the critical role of data as a strategic asset. DDA presents a paradigm shift from traditional centralized layered architectures (see Figure 1) to distributed and federalized EAs (see Figure 2). It focuses on creating a flexible, agile, and scalable architecture that aligns with the digital capability needs of individual domains within the organization, while emphasizing a cross-domain consistency and interoperability.
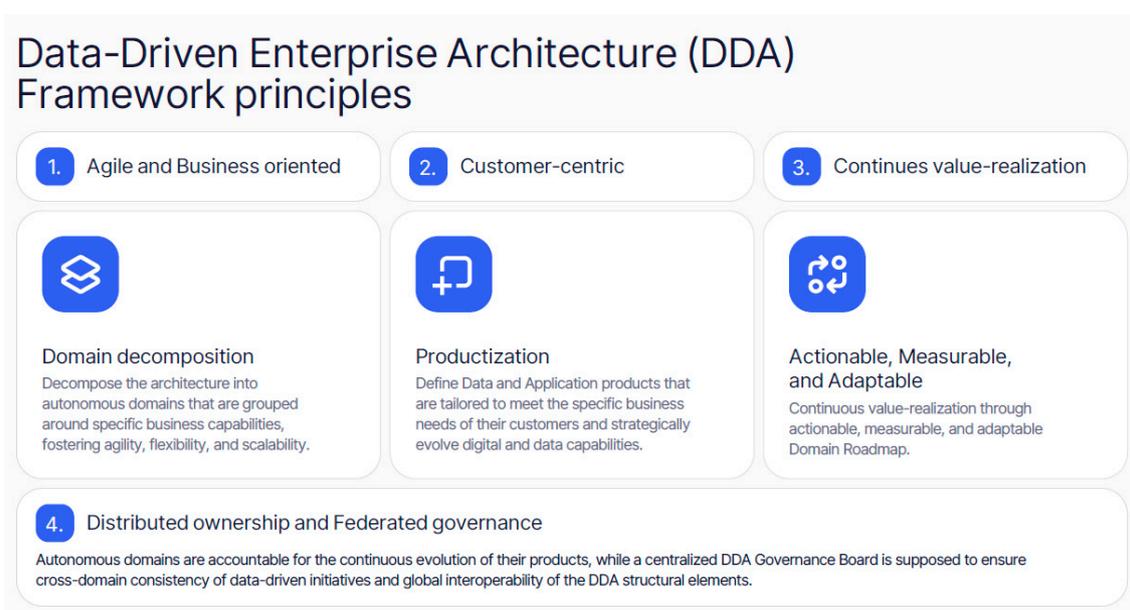

Figure 2. Principles of the Data-Driven Enterprise Architecture.

Notably, the DDA distinguishes from mainstream scientific concepts and business practices due to its holistic approach. Unlike many existing approaches that tend to focus on the transformation of application, business, and data architecture independently, DDA emphasizes the continuous harmonization of these crucial architectural components toward common domain business goals. Business practices have shown that decoupling architectural components over time may inadvertently lead to issues of inconsistency and interoperability. Addressing these challenges at a later stage can often require significantly more effort and resources compared to considering harmonization, continuous alignment, and interoperability from the outset.

The DDA represents a pioneering framework for distributed and federated architecture at a scale that breaks down architecture into domain-specific components, including application, business, and data architectures, while keeping continuous harmonization across these components and domains.

While the DDA was initially designed with pharmaceutical R&D in mind, it is a generic framework that offers fundamental architectural guidelines and principles for companies expediting data-driven transformation. As such, it is adaptable to various industries and can be customized and refined to align with the unique characteristics of each enterprise, including internal and external factors such as industry, geographic location, strategic goals, corporate culture, existing architecture, and more. In the following chapter, we will demonstrate how the DDA was tailored to meet the specific requirements of a Pharma company's R&D division.

DDA Framework Principles

1. Domain decomposition

DDA promotes the decomposition of EA into autonomous domains (see Figure 3), each logically grouped around specific business capabilities with well-defined boundaries. This decentralized approach empowers each domain with its own autonomous Domain Architecture (DA) that is decomposed further into Domain Application Architecture, Domain Business Architecture, and Domain Data Architecture. While this distributed character of an EA is aimed to foster agility, flexibility, and scalability in adapting to changing business needs, an accelerative synergy effect is going to be achieved due to harmonization of all components of the DA towards a common goal—to serve, enable, and accelerate Domain Business Capabilities.

For instance, within the pharmaceutical R&D context, the Clinical Development Domain could incorporate the following capabilities that define its boundaries: Investigative Staff Engagement, Study Management, Participant Data Capture, and more. Thus, the grouping of the capabilities within one domain could be based on a common system landscape or stakeholders, connected business processes, and identifiable common domain goals.

Decomposing EA into manageable domains requires strong leadership support, commitment from domain experts, well-defined goals, and an understanding of the unique value each domain contributes to the organization. It involves analyzing core business processes, business users, system landscapes, and data flows to establish optimal domain boundaries.

Importantly, while domain-specific business, application, and data architectures are decentralized, the Infrastructure and Technology Layer often remains centralized to provide a foundational backbone supporting the entire architecture. This centralized infrastructure includes underlying technology, integration capabilities, and their governance processes, but its extent may vary based on the organization's specific needs.
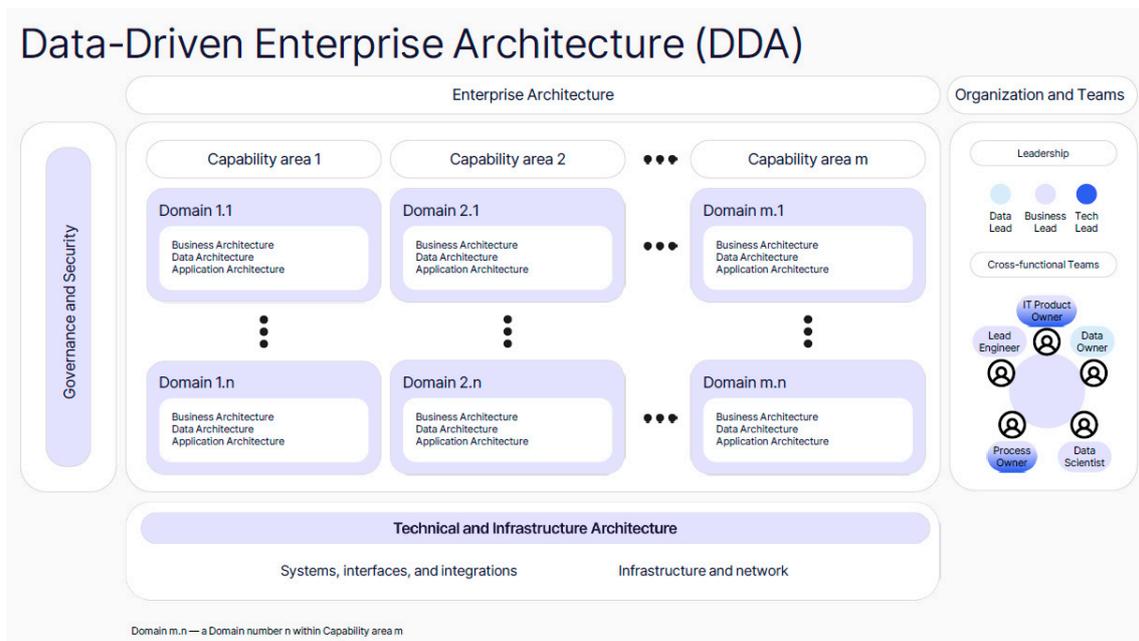
Figure 3. Data-Driven Enterprise Architecture.

2. Productization and customer centricity

DDA places a strong emphasis on productization, transforming assets into products, characterized by several core attributes, as follows:

- **Definition and Boundaries:** Products are well-defined entities with clear boundaries. This means knowing exactly what the product is, what it encompasses, and what it does not.
- **Identifiable Purpose:** Each product serves a distinct purpose, aligned with the organization's ecosystem and value delivery.
- **Domain Ownership:** Each product is owned and proactively managed by the respective Domain.
- **Customers or Potential Customers:** Products revolve around customers, whether internal or external, ensuring alignment with their specific needs.
- **Tangible or Measurable Value:** Products are designed to generate tangible or measurable value, with well-defined value propositions for their customers or stakeholders.

The critical factor that sets products apart from mere assets is this focus on customers. While the first two characteristics are also applicable to assets, the presence of customers, whether external (such as end-customers, clients, partners, or regulatory bodies) or internal (including other departments, teams, or domains), and the emphasis on delivering value to them is what truly differentiates a product. Notably, not all assets are suitable for productization, and organizations should carefully evaluate their transformation into products.

Productization within DDA primarily pertains to Application and Data Architecture, while Business Architecture guides the evolution of digital and data capabilities within data and application products to meet business needs. This focus on productization is vital for organizations seeking excellence in data-driven strategies for several reasons, as follows:

- **Tailored Solutions:** Application architecture and data products are tailored to meet specific business needs, enhancing operational efficiency.
- **Continuous Evolution:** Products are designed for ongoing evolution, fostering long-term collaboration among cross-functional domain teams throughout the product lifecycle that differentiates from a project approach with a predefined timeline, scope, and budget.
- **Data-Driven Decision-Making:** Data products empower teams with actionable insights, facilitating data-driven decision-making.

- **Resource Optimization:** Concentrating resources on products that directly contribute to business capabilities improves resource allocation.
- **Competitive Advantage:** Organizations leveraging application architecture and data products gain a competitive edge by aligning products with customer needs and industry trends.
- **Scalability and Innovation:** Products are inherently scalable and drive innovation by encouraging creative thinking to enhance functionality and user experiences.
- **Flexibility:** Application architecture and data products are agile and adaptable, ensuring long-term relevance.

In the context of DDA, Data and Application Products are structural components of Domain Architectures (see examples in Table 1), providing tailored solutions that align with the domain objectives and customer business needs, where:

- Application Product (AP)—a group of software components that provides specified digital capabilities.
- Data Product (DP)—data entities that have a distinct purpose and value for the organization and are logically grouped for effective management and improvement. Besides general product attributes (defined above), the additional data-specific attributes of DPs encompass reliability (reflecting business accuracy that directly depends on the quality of data); discoverability and understandability (encompassing self-describing semantics, syntax, and usage metadata); composability of integral components such as metadata, data lineage, access, and governance; and semantics.

Table 1. Productization of Domain Application and Data Architecture.

Characteristics	Application Product Example	Data Product Example
Definition and boundaries (What is it? What is it not? What are the components?)	Investigative Staff Training Application	Phase 2 Clinical Study Data for COVID-19 Vaccine ABC
	Feasibility Assessment Tool for Clinical Studies (as software components delivering this digital capability)	Feasibility Assessment Report for Clinical Studies (as a combination of data components)
Identifiable purpose (What it serves for?)	Distributing training resources for site staff involved in clinical studies and recording the completion	Providing evidence of safety and efficiency of the vaccine
	A repository for a collaborative site selection, based on their capabilities and previous experiences with them	Support decision-making about the feasibility of a study and incorporated risks. Keep transparency of the decision and ability to use it for future assessments.
Customer(s) (Whom it serves?)	Site staff, PII, and other study team members	Internal: Study team, submission team, and other R&D departments. External: regulatory agencies, research organizations, etc.
	Study planning team and other R&D departments and teams	Study planning team and other R&D departments and teams, as well as cross-organizational departments
Tangible or measurable value (What value does it generate? How could it be measured?)	Improved site satisfaction and clinical trial operations (e.g., yy FTE savings, zz points increase in the "Sponsor of choice" assessment)	A prerequisite for an approval to market the vaccine and receive a return on investments. In addition, the data would help to further investigate the disease
	Increases efficiency and accuracy of the feasibility assessment by digitalization (e.g., xy FTE savings)	Increase in accuracy and efficiency of decision-making about the feasibility of the study and incorporated risks

These products serve as the foundational building blocks of Domain Roadmaps, guiding strategic data-driven investments.

3. Actionable, Measurable, and Adaptable

The DDA framework is strategically oriented towards the continuous realization of value through ongoing product improvement and the expansion of capabilities. To achieve this, it is imperative to ensure that the framework is both actionable, measurable, and adaptable.

Actionable: Actionability is driven by Domain Roadmaps, which consist of prioritized product initiatives outlined on a Domain's Roadmap. These product initiatives are typically rooted in identified challenges or improvement opportunities and they come with clearly defined, measurable value upon their implementation. For instance, within pharmaceutical R&D, an Application Product initiative could involve the introduction of a "Feasibility Decision Support" feature, enhancing the "Feasibility Operation" domain. This feature might focus on predicting whether a study site in a particular country can meet enrollment targets within specified timeframes.

Measurable: The implementation of product initiatives should yield measurable outcomes, predefined prior to their execution. This approach enhances the efficiency and transparency of investment decisions, aiding in the prioritization of initiatives within Domain Roadmaps. Moreover, it enables the monitoring of value realization and provides the flexibility to adapt when actual outcomes deviate from predefined expectations.

Adaptable: Defined actions and strategies related to product initiatives should undergo regular review, based on preliminary outcomes and the resources utilized during each realization phase. This ongoing monitoring and proactive adaptation are designed to maximize the overall value achieved by product initiatives and to mitigate investment risks.

For large-scale enterprises, we recommend the establishment of standardized sets of measurable outcomes, often referred to as a "benefit realization matrix". This matrix serves to enhance transparency and comparability, facilitating informed decision-making and alignment with the organization's overarching goals.

4. Domain ownership and Federated Governance

DDA recognizes the importance of striking a balance between centralization and decentralization through Federated Governance. To emphasize this principle, the DAA approaches autonomous domains to take over accountability of Data and Application Products, while a centralized DDA Governance Board plays a pivotal role in harmonizing and aligning various aspects of data-driven architecture across different domains. Its mission is to ensure that data-driven initiatives contribute to the overall business strategy and drive successful digital transformation.

The DDA Governance Board acts as a collaborative forum, uniting representatives from diverse domains, and could be extended via the introduction of IT, data management, business units, and strategy teams. Together, they make informed decisions about prioritizing initiatives, resource allocation, and aligning Domain Roadmaps, focusing on strategic data and application projects that impact on digital transformation goals. Furthermore, the DDA Governance Board monitors progress, evaluates performance against value-realization metrics, and provides guidance to overcome challenges. It fosters knowledge sharing, implementation of best practices, promotes data-driven decision-making, and advocates for cross-functional consistency and interoperability.

3. Results of a Use Case of the Data-Driven Enterprise Architecture at a Pharma Company

3.1. Background of Pharma Company

Our case study focuses on a global R&D organization of one of the top five pharmaceutical companies (Global R&D). Given the highly regulated nature of the pharmaceutical industry, the Pharma company is required to adhere to stringent data privacy and IT secu-

rity measures to navigate the complex regulatory landscape and fast-growing healthcare digital ecosystem successfully.

In addition, the Pharma company faces the challenges posed by the rapid expansion of the digital environment. This includes constantly exploring and assessing the feasibility of the implementation of modern technologies such as generative AI, cloud computing, and IoT, effectively managing the increasing volume, velocity, and variety of data, meeting the heightened expectations of patients, sites, regulators, and other stakeholders for technological innovations and fostering a culture of innovation and agility. Moreover, as the design and architecture of each clinical trial are individual, the Pharma company must navigate these challenges within an existing complex and dynamic Enterprise Architecture, while focusing on regulatory compliance and high quality as leading principles. In addition, the dynamic and complex character of the system landscape of such a large-scale global company makes even capturing an existing architecture an enormous challenge, as an architectural view of yesterday will not be actual today. This, in turn, creates an additional barrier for a complex consideration of the Pharma company's digitization initiatives.

To address these challenges, the Pharma company developed a Digital Transformation Strategy (DTS), aiming to explore and implement strategies that leverage digital technologies to transform and accelerate clinical trial operations, bringing the best products and outcomes for patients, sites, and the entire healthcare industry.

3.2. Methodology

The developed DTS methodology consists of four phases, as follows: Input and Analysis, Future State Design, Build the Roadmap, and Execute the Roadmap (see Figure 4).
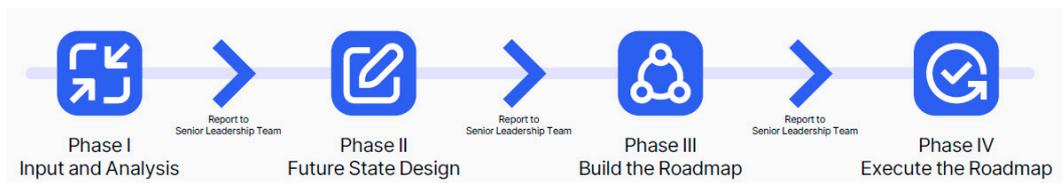

Figure 4. Digital Transformation Strategy approach.

3.2.1. Phase I—Gain Input and Analysis

During Phase I, "Gain Input and Analysis", all employees of the Global R&D organization, comprising approximately 10,000 individuals, were sent a short email survey to respond to the question "What should be in a strategy for digital transformation, and how do we enable it?", which received over 200 responses, generating over 150 pages of ideas. The responses underwent thorough analysis and were clustered based on relevant systems and categories, including Technology, People, Processes, Metrics, and Capabilities. These delineated categories served as the foundational framework for the Future State Design, as outlined in the Section 3.2.2. To complement the survey, interviews with 37 stakeholders, representing the organization's leadership, therapeutic areas, and key functions and programs, were conducted to further delve into this question.

Notably, the DTS, including strategic initiatives, their structure, prioritization, metrics, and realization roadmap, was defined in close collaboration with core business representatives and a company leadership team. This approach facilitated a strong alignment between the strategy and core business needs. Moreover, as the strategy was "bottom-up", sourcing from the stakeholders' digitalization demands, it ensured stakeholders' commitment not only to the DTS definition, but also to its execution. In addition, collaborative effort to explore digital capability needs across different departments created transparency about common challenges, and a large volume of responses reduced risks of bias.

By grouping and categorizing the insights derived from the survey and interviews, six key strategic pillars were derived to form the core structure of the DTS. Data Fabric was identified as one of these key strategic pillars, receiving more than 90 out of around

200 responses, highlighting the weight of the data architecture transformation as a baseline to enable the overall data-driven transformation of the Pharma company. Figure 5 highlights the groups of the most frequent business needs addressed towards Data Fabric in this Phase.
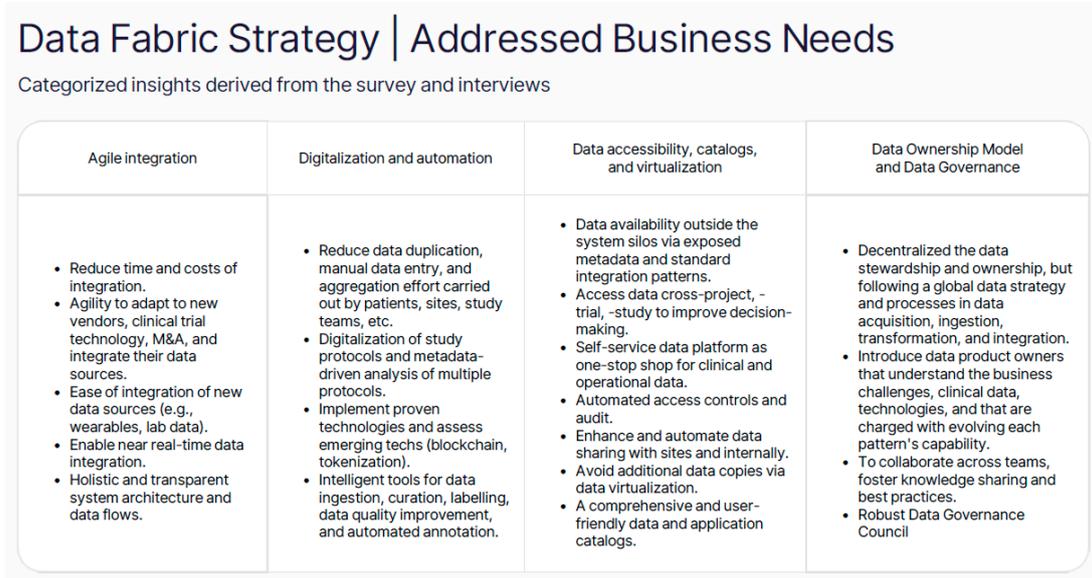

Figure 5. Business needs addressed towards Data Fabric.

3.2.2. Phase II—Future State Design

Following the presentation of the output of Phase I of the DTS to the Senior Leadership Team and with their alignment to proceed, work on Phase II commenced. The objective of Phase II was to develop a future state vision for Data Fabric. Whilst the objective of Phase I was to gain input on *What should be enabled*, Phase II focused on *How to achieve it with means of Data Fabric*. Therefore, the approach of Phase II narrowed the core contributors to internal and external technology and data expertise.

Firstly, several workshops focusing on Data Fabric were conducted with internal experts for the following purposes:

- to brainstorm and ideate the future state of Data Fabric around the following five categories: Technology, People, Processes, Metrics, and Capabilities
- to detail the five categories by defining four core components for each category that correspond to the grouped and categorized needs derived from Phase I (see Figure A1 in Appendix A)

The collaborative assessment, categorization, and prioritization of the insights by internal experts was a preliminary step to define a target vision of Data Fabric at the R&D organization of the Pharma company. During the next step, the project team assessed a Data Fabric concept outside the Pharma company's ecosystem including its capabilities, industry experience, core vendors, and external experts' recommendations, as well as alternative approaches such as Data Mesh and their compatibility. This brought the business needs of the Pharma company and Data Fabric capabilities closer, including the implementation of best practices.

The analysis resulted in the recognition of the fact that trying to fulfill the defined business needs as fragmented capabilities of Data Fabric will have an extremely limited outcome and a more fundamental and future-oriented data architecture framework, and that incorporating organizational and mindset changes was required. This framework should lay the foundations for steadily enabling and progressively addressing the evolving business needs in the data and digitalization area. One of the reasons for this is a rapid transformation of the technological and business environment that impacts the business

requirements and expectations. Thus, the business needs addressed today may be less relevant in the future, whilst their implementation could require years, especially in a large-scale organization with a complex and dynamic EA, similar to that at the Pharma company.

As a result, the Pharma company's Data Fabric strategy was primarily aimed at establishing the foundation of Data Fabric, enabling iterative definition, implementation, and progressive evolution of its capabilities, rather than solidly focusing on implementing identified Data Fabric business needs. For that purpose, we created a Federated Leadership Model (FLM) that grounded the base for the Data Fabric approach (see Figure 6).
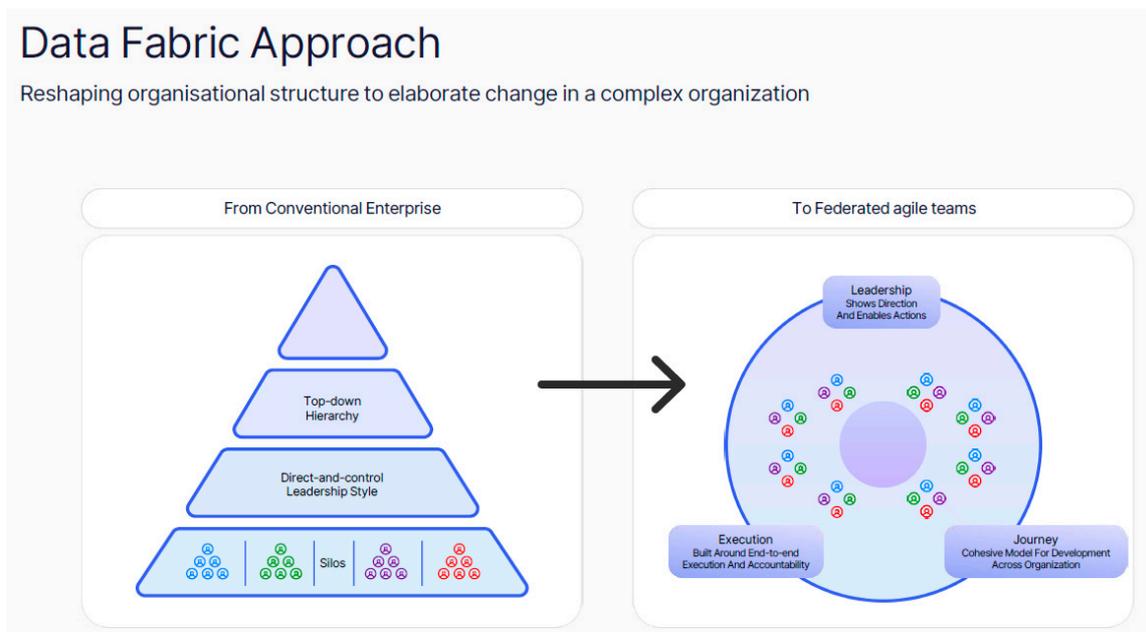
Figure 6. Data Fabric approach.

The FLM inherits a lot from modern organizational structures (e.g., matrix, flat, network, and others). It significantly differentiates from traditional organizational models that have functional or divisional pyramidal structures, impacting decision-making, communication, flexibility, and teamwork. Traditional organizational models are primarily based on a top-down hierarchy and have a direct-and-control leadership style. They, and their combinations, are the most common models for large-scale companies because they provide stability and foster functional specialization of employees, boosting accountability and efficiency. However, the traditional models are prone to generate silos between the teams, especially below the leadership levels. As information silos became one of the major barriers in the way of data-driven transformation, companies started to rethink their organizations towards more cross-functional teams, flexible and decentralized structures, more flat hierarchy, and a less directive leadership style that promotes more autonomy and responsibility of the teams.

The Global R&D function has evolved from standard hierarchical structures incorporating elements of modern agile organizational models over time. The FLM did not aim to reshape the official organizational structure of the Global R&D organization, but rather to generate an additional virtual federated layer upon the existing one. This approach reduced the size of organizational change and, consequently, the risk of resistance. However, as the employees were assigned additional tasks according to their FLM roles (on top of their existing responsibilities), convincing them of the added value of their extra effort not only for the whole organization, but also for their individual functions was required. The second key success factor was a leadership support that prioritized the topic and promoted its significance for the Pharma company.

Thus, the FLM embedded agility and cross-functional collaboration between silos within the current organizational structure to effectively drive data-driven initiatives. This allows key knowledge brokers to share and push requirements and solutions for the benefit of both Global R&D's stakeholders, such as therapeutic areas, sites, and study participants, and internal users at a higher pace, with increased oversight and adaptability.

As depicted in Figure 6, the FLM defines the following three core elements:

1. **Leadership:** Data Domain (DD) leadership organizations responsible for defining directions of DD evolvement, representing DD, and driving cross-domain alignments.
2. **Execution:** Cross-functional and autonomous DD teams that have end-to-end accountability for DD Data Products (DPs) and drive the execution of DD strategies and roadmaps. In addition to DD Teams, the FLM defines Data Fabric High Performing Teams (HPTs) that are responsible for implementing integration, data visualization, a shared semantic layer, and other data engineering tasks, as defined by the DD Teams.
3. **Journey:** The Data Fabric Maturity Journey is represented by the iterative implementation and steady evolution of Data Fabric capabilities that are driven by DD business needs and derived from DD roadmaps.

3.2.3. Phase III—Roadmap

The objective of Phase III is to define a roadmap from the current state to the envisioned future state of the Data Fabric at the Global R&D. At the onset of this phase, an up-to-date overview of the system landscape and data architecture was not available, making it challenging to establish an architectural starting point for the roadmap. Creating such an overview would require the implementation of tools and processes for EA management, which was not feasible within the defined time horizon. Nonetheless, valuable insights about the AS-IS application and data architecture, as well as forthcoming changes, were gleaned from several related ongoing projects. This supports one of our key findings about the need of continuous alignment with correlated ongoing initiatives to maximize the overall benefits of any transformational program.

In addition, the outcomes of Phases I and II provided a broad business view of the current state and identified pain points. By aligning these architectural and business inputs with the defined future vision, we have defined a Data Fabric roadmap (see Figure 7).
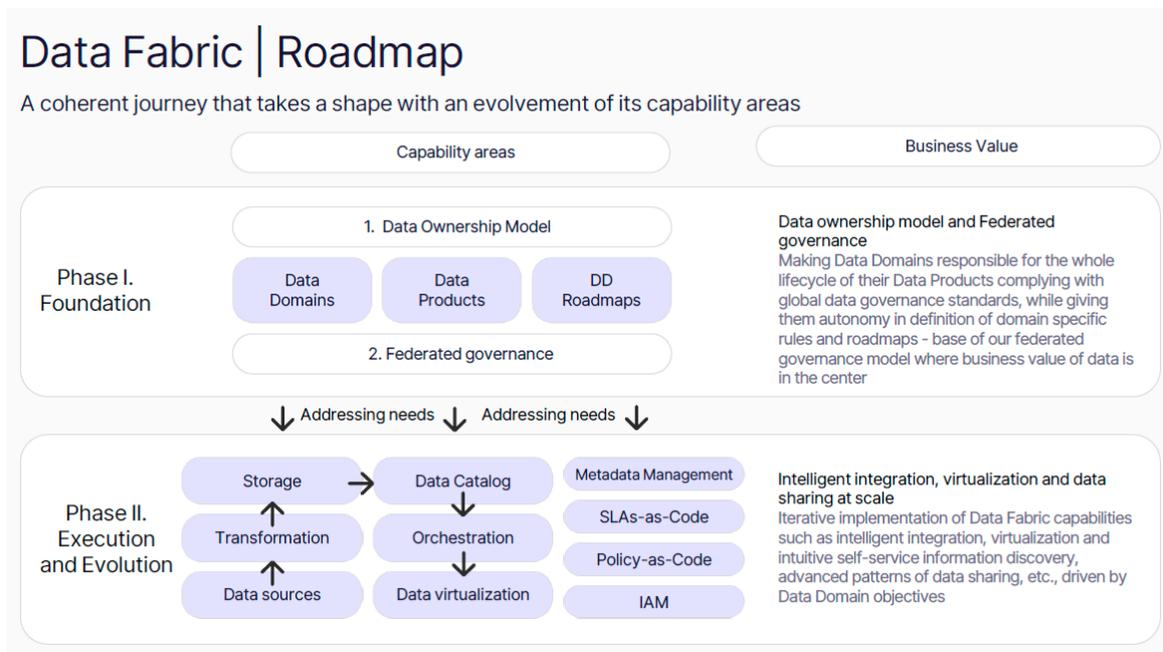

Figure 7. Data Fabric roadmap.

Phase I: Foundation of Data Fabric

Phase I is anchored in the FLM, where data leaders take ownership of the data vision and roadmap within their designated DDs. These DD leads orchestrate key stakeholders, define objectives, and drive execution. The core components of Phase I, as depicted in Figure 7, are as follows:

1. Data Ownership Model that, in turn, consists of the following:
    a. DDs and their leadership organizations
    b. DPs are the most valuable data for the organization that is grouped to manage it effectively. They are defined and owned by DDs.
    c. DD roadmaps represent an actionable way to address, align, implement, and track the continuous improvement of the DPs.

2. Federated governance framework that defines roles, responsibilities, decision-making, and escalation pathways, enabling the efficient handling of data in a federated way.

Phase I necessitates significant organizational, functional, and cultural transformations in how the Global R&D data are organized and governed. It involves defining which teams are accountable for data, clarifying their roles and responsibilities and establishing the role of data and their value within the organization. However, it can be challenging to immediately discern the tangible benefits associated with each element of the foundational phase, as their value may be realized over time, during the subsequent execution phase. Consequently, a comprehensive change management effort was required to socialize the framework. This effort involved conducting pre-assessments and presenting the draft framework to stakeholders to integrate their feedback and refine the roadmap. This approach contributed to a smooth and steady acceptance of the change and fostered stakeholder commitment.

Phase II: Execution and Maturing of Data Fabric Capabilities

Phase II is dedicated to the execution of DD roadmaps, leading to the maturation of Data Fabric capabilities. A core principle of this framework is its business orientation, whereby the evolution of Data Fabric capabilities is primarily driven by DD business needs, as defined in their roadmaps. DD teams are empowered with Data Fabric HPTs to drive the execution of their objectives, ensuring conceptual, architectural, and technological consistency in data engineering and integration initiatives.

Phase II consists of the following three core components:

*   DDs execute their strategies and roadmaps, empowered by Data Fabric HPTs.
*   Data Fabric HPTs implement integration, data visualization, and a shared semantic layer.
*   The Data Fabric Maturity Journey, which represents the progression of Data Fabric capabilities through structured levels, with an increasing ability to deliver.

Despite being named Data Fabric, Phase I of the strategy incorporates most elements from the Data Mesh concept, showing that Data Mesh and Data Fabric can complement each other effectively. Whilst Data Mesh lays the organizational and mindset foundation for how enterprise data are organized and managed, Data Fabric establishes the technological basis for the data architecture. By carefully blending the most suitable components of these two concepts and customizing them to fit the specific needs of the enterprise, we developed a comprehensive roadmap for implementation and evolution.

The ongoing debate between proponents of Data Mesh and Data Fabric often revolves around the perceived superiority of one approach over the other. However, our case study has demonstrated the effectiveness of a complementary approach. This approach has garnered high acceptance and commitment from stakeholders and leadership, validating its efficacy thus far.

By embracing the strengths of both paradigms, the Pharma company has fostered an environment where organizational and technical aspects converge harmoniously. The FLM adopted during Phase I enabled data leaders of distributed and autonomous DDs to steer the data vision and roadmaps, facilitating the seamless orchestration of key stakeholders and efficient execution of objectives.

As the Pharma company advances into Phase II, the focus shifts towards execution and the maturation of Data Fabric capabilities. Establishing Data Fabric HPTs to implement integration, data visualization, and a shared semantic layer, ensuring consistency and cohesiveness across data engineering and integration initiatives, will empower DDs in the execution of their designated DD roadmaps. The business orientation at the core of the framework should emphasize that DD roadmaps, driven by business needs, take precedence in guiding the evolution of Data Fabric capabilities. This approach allows the organization to respond dynamically to emerging challenges and to capitalize on new opportunities, while continuously improving its data architecture.

> *Data Mesh and Data Fabric can complement each other effectively. While Data Mesh lays the organizational and mindset foundation for how enterprise data are organized and managed, Data Fabric establishes the technological basis for the data architecture.*

3.2.4. Phase IV—Execute the roadmap

The execution of the Data Fabric Roadmap commenced with the Phase I "Data Fabric Foundation" components such as the Data Ownership Model (DOM) and the Federated Governance Framework (see Figure 8).
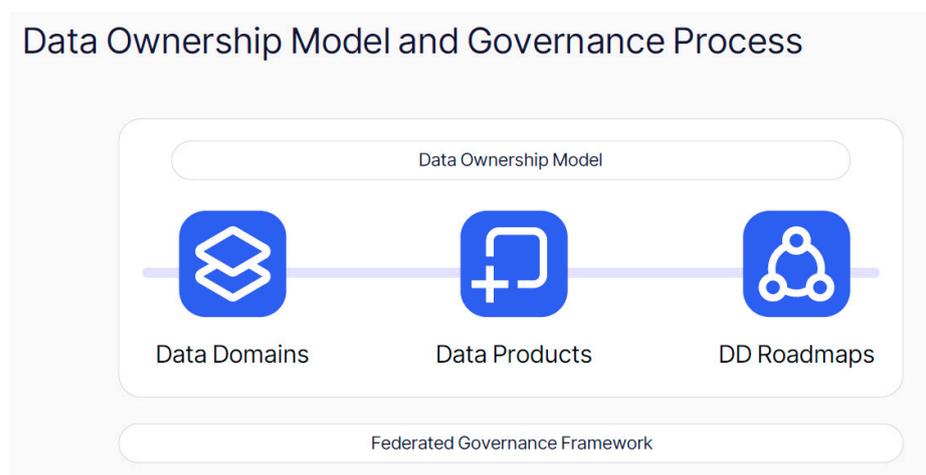
Figure 8. Data Ownership Model and Governance Framework.

Data Ownership Model (DOM)

The DOM operates as a catalyst to amplify visibility, quality, discoverability, and reusability of the Global R&D data, thus expediting the cultivation of a data-driven culture. The key objectives of the DOM are as follows:

- **Transparency of Global R&D Data**: By structuring DDs and DPs, the DOM enhances the transparency of data, ensuring its clear organization and accessibility.
- **Effective Data Organization**: The DOM defines the most valuable data for the organization, efficiently organizing and managing it to maximize its utility.
- **Selective Data Improvement**: The DOM facilitates targeted improvements in the organization's data and their governance, guided by the DD roadmap.
- **Enhanced Decision-Making**: Business-oriented data ownership is identified, promoting informed decision-making and data quality enhancement.
- **Autonomous DDs**: Autonomous DDs are established, reducing reliance on central data teams and enabling end-to-end accountability for DPs within each domain.

Whilst the cornerstone of the DOM comprises the following three core constituents: Data Domains (DDs), Data Products (DPs), and DD Roadmaps, the Federated Governance Framework creates an underlying foundation for its efficient operation and governance.

1. Data Domains

In our case study, DDs signify logically clustered data elements, organized around the organization's specific R&D business capabilities, and is characterized by well-defined contextual boundaries and dedicated ownership.

Defining the optimal boundaries for DDs necessitated a comprehensive assessment of the business, data, and systems architecture of the Global R&D organization. This assessment encompassed the analysis of the related Business Operations Process Frameworks and the Organization Breakdown Structure, representing the business architecture, as well as data flows and classifications that pertain to data architecture. A novel Application Architecture Framework (AAF) was also integrated, including its Domain and Product Structures, as well as the business capabilities they support.

This comprehensive approach allowed us to develop a DD structure, as illustrated in Figure 9.
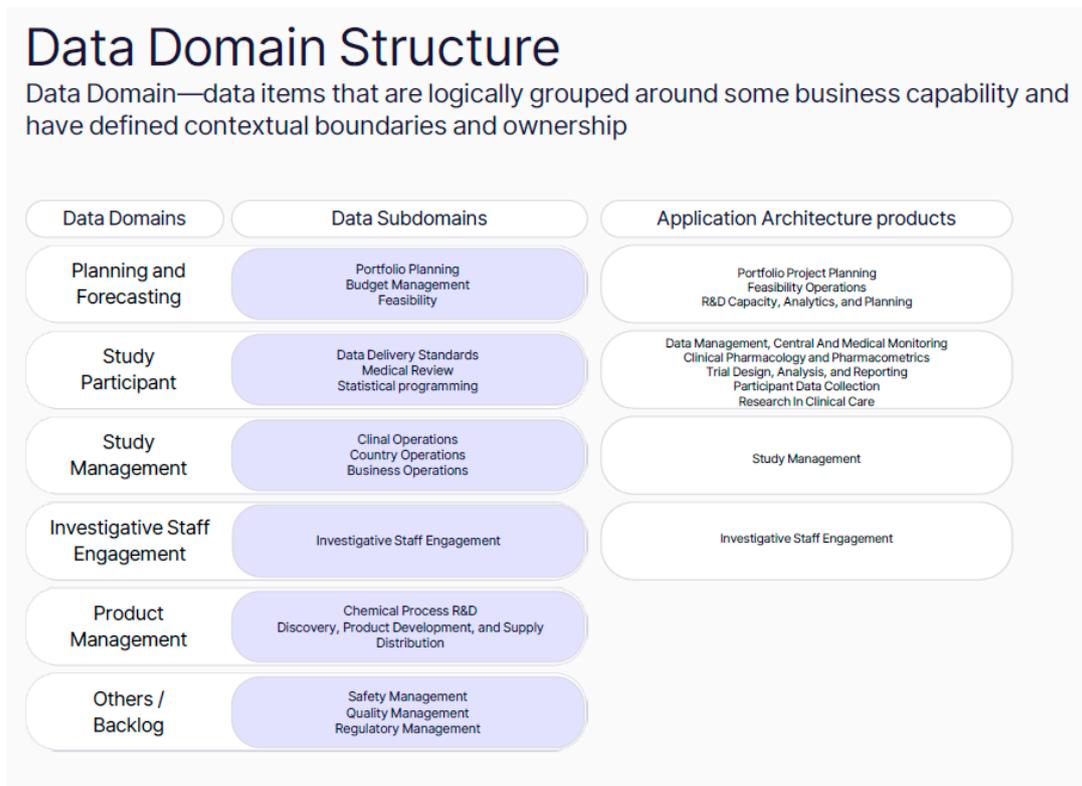
Figure 9. Data Domain structure.

The first column of Figure 9 depicts the initial six DDs where the first four DDs encompass the logically grouped Global R&D data, while certain DDs and Subdomains like Product, Safety, Quality, and Regulatory Management lie outside the functional purview of the Global R&D, yet contain data that are managed within the organization or holds close relevance. Despite the apparent low granularity of the DD structure, these broad DD areas are further divided into distinct subdomains, each inheriting the characteristics of its parent domain. The last column on the right of Figure 9 displays the mapping of DDs to APs (the Application Architecture Products defined within the AAF).

A striking parallel was identified between the DOM and AAF objectives during the DD structure definition. Both frameworks share the common goal of transforming the EA towards a more adaptable, business-oriented model, in alignment with architectural trends like Product Thinking, Domain-Driven Design (DDD), modularization, and decentralization. The fundamental distinction lies in their focus whereby AAF targets the application layer transformation, while DOM addresses data architecture optimization. Since data and applications are inherently interdependent, separating their architectures

into disparate structures would introduce unnecessary complexity and hinder potential synergies. Moreover, maintaining separate data and application architecture structures could introduce the risk of conflicts between these structures and their respective objectives, necessitating constant alignment efforts and consequently increasing operational costs.

Although the AAF had commenced earlier than the DOM and had already defined a new Application Domain (AD) structure with defined APs and cross-functional teams "Squads" collaborating to evolve these APs, both frameworks exhibited strong interdependence. Recognizing this dependency and the importance of synchronization and harmonization, the DD structure was intentionally designed to align with the AD structure.

In Figure 10, the mapping of the AAF to the DD architecture is depicted. It shows that the Pharma company's DDs closely mirrored the AD structure, with specific exceptions that are driven by logical data flow groupings and classifications. It is crucial to acknowledge that a failure to consider data flows and their lifecycle when defining DD boundaries, and a mere blind integration into the AD structure, holds the potential to inadvertently foster the creation of new data silos. Such silos would counteract the fundamental objectives of the DOM, thereby jeopardizing the intended fluidity of data movement and accessibility across the organizational landscape.
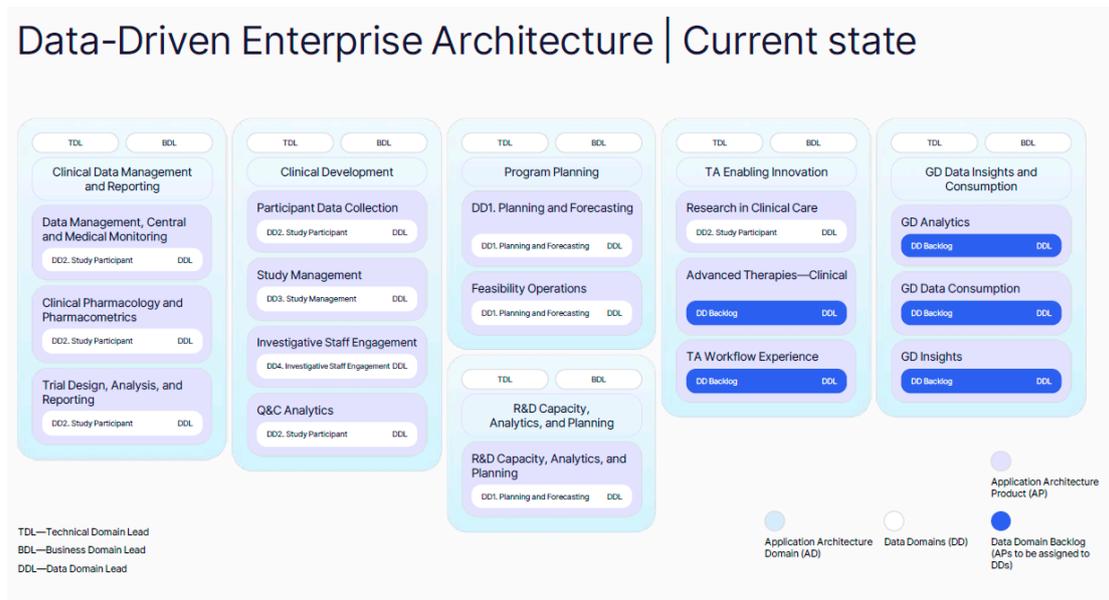

Figure 10. Data-Driven Enterprise Architecture at the Pharma company's R&D organization—current state.

Thus, certain DDs converge or diverge from corresponding ADs based on shared or distinct data flows. For instance, the DD "Planning and Forecasting" effectively merges two ADs, "Program Planning" and "R&D Capacity, Analytics, and Planning", due to the commonality of data feeds to these groups. Conversely, the DD "Investigative Staff Engagement" maintains a direct one-to-one mapping with an AP of the AD "Clinical Development", reflecting specific data-related considerations. These exceptions illustrate the flexibility and adaptability of the framework, reflecting the dynamic relationship between application and data architecture.

This strategic alignment is further affirmed by the assignment of DD Lead roles to the corresponding Business Domain Leads. This move underscores the ongoing integration of the AAF and DD frameworks, solidifying their synergy.

Drawing from classic strategic management methodologies and the principles of effective change management, securing stakeholder commitment emerges as a key for the success of transformative endeavors [85–87]. This necessitated a concerted effort to garner approval and alignment from DD Leads and Teams, ensuring organic assimilation

and enduring utility of the framework. The added complexity of DD Teams and Leads concurrently taking on new roles alongside their existing responsibilities underscored the need for comprehensively articulating the change's value proposition—both for the organization as a whole and for each specific function. This involved a series of thorough reviews, assessments, and iterative refinements of the proposed structure and approach, conducted collaboratively with each DD team.

Furthermore, this aligned yet adaptable structure garnered approval and acceptance from the Senior Leadership Team, as well as from AAF Business and Technical Domain Leads, marking the commencement of the DOM implementation (see implementation approach details in Figure A2).

This intentional alignment, coupled with strategic exceptions, underscores the value of continuous collaboration and steady integration of the AAF and DOM frameworks. This approach should foster the future organic growth of the DDA, contributing to seamless coordination between application capabilities and data assets, as well as the organic evolution of the EA through well-coordinated efforts.

2. Data Products

Within the architecture of DDs, the foundational building blocks are DPs that are defined as data entities that are important for the global R&D function's business. These are grouped and organized for effective management and improvement. The key attributes of DPs encompass the following:

- **Value on its own:** DPs are selective combinations of data entities that actively support the organization's business capabilities and subsequently furnish tangible business value. This inherent value is characterized by self-sufficiency, rendering each DP meaningful in its autonomous context or "on its own". Thus, for instance, metadata could not be classified as a DP because it has no meaning without a related dataset.
- **Domain Ownership:** The DOM framework embodies a principle of entrusting the stewardship, accountability, and progressive improvements of DPs to respective DDs. While individual DP ownership is not mandated initially (only DD ownership), it remains a dynamic prerogative for DD teams, with their Leads, to implement it, if needed. This approach provided more flexibility and autonomy to DDs and reduced the framework implementation's complexity.
- **Reliability:** A hallmark of DPs lies in their capacity to faithfully mirror business accuracy. To do so, a DP is adhering to quality benchmarks tailored to its distinctive purpose. For instance, clinical data for regulatory submissions adhere to stringent quality standards in terms of consistency, completeness, and accuracy, but could be less stringent in terms of timeliness because it might imply lags between the trial event occurring and its capture in a Clinical Trial Management (CTM) system. However, timeliness of data captured from IoT devices could be critical to generate a valid insight. Consequently, a DP should have a clearly defined purpose that determines its quality requirements, which satisfy it becoming reliable. However, together with the evolution of a DP and extension of use cases, the purposes of DPs will also be extended. For example, a data model that consumes, among others, a participant recruitment DP could generate some valuable insights for a reduction in retention rates and could be incorporated into recruitment operations. In this case, based on the new consumption purposes, the DP could have extended quality standards (e.g., additional metadata could be recorded). It is notable that although each DP has its unique purpose, all of them should contribute to the common vision defined for a DD.
- **Discoverability and Understandability:** DPs are seamlessly discoverable through standard tools like data catalogs, ensuring their accessibility within the system landscape. Additionally, their inherent structure promotes self-describing semantics, syntax, usage, and inter-relationships.
- **Composability:** DPs consist of one or more constituent datasets, serving as their fundamental building blocks. Beyond datasets, other integral components such as

Metadata, Data Lineage, Data Access and Governance, and Semantics contribute to the holistic identity of a DP. Comparable to the assembly of Lego blocks or a Rubik's cube, where each individual block has a limited meaning, while their logical composition into a DP generates a value [88].

The conceptualization of DP characteristics was underpinned by findability, accessibility, interoperability, and reusability (FAIR) principles [89] and the foundational tenets of Data Mesh [90] such as product thinking for data, embodying the individual needs and organizational and cultural aspects of the Global R&D.

Table A1 summarizes DPs that were defined by corresponding DD Leads and their teams. However, it is now only the initial overview that depicts which data are produced and consumed for operational purposes. In future, we expect that with the evolution of the framework and the Global R&D's maturity along the digital transformation journey, the defined DPs will also evolve, by expanding their use cases, and new DPs will emerge.

Our DP methodology employs a DP card system that consolidates essential details about the DP, including DD, Name, Purpose, Stakeholders, Data Management and Governance specifics, Systems and Access information, and Business Process insights, in a standardized and structured manner (see Figure A3). Notably, the DP card design strikes a balance between two contrasting aims—reducing the number of DP card fields to prevent resistance from DD teams and avoiding overlap with other tools (e.g., data catalog)—while simultaneously broadening the scope of DP card fields to provide comprehensive information about DPs.

By implementing DP cards, the Global R&D aims to offer a practical solution to the challenge of data discoverability and to dismantle data silos. Previously, users encountered difficulties when searching for and accessing data, even though several data catalogs were available. Furthermore, identifying the data's origins, lineage, associated business processes, system sources, and relevant contact person or owner were complex tasks. The DP cards will be consolidated into a DP catalog, functioning as a centralized "one-point-shop" for pivotal data. This approach should enhance transparency, discoverability, understandability, and the reusability of the organization's data throughout the entire company. Notably, the current project phase was more focused to define the DP card structure and gather the DP information in a simple Excel template, while still exploring a specific tool for its user-friendly maintenance and discoverability in the future. This approach steadily defines the business requirements towards the required tool, whilst continuing to fine-tune the approach.

Another aspect of the DP cards involves the consolidation of DP initiatives and enhancement needs (see Figure 11). This segment encompasses investment plans and challenges related to specific DPs, along with the anticipated value upon realization. Once these initiatives gain approval from the relevant stakeholders and the Data Governance Council (see the decision-making workflow of the Federated Governance), they are integrated into the DD Roadmap. Subsequently, core implementation activities and tracking metrics for each initiative are documented within the DP card. This structured approach establishes a feasible and controllable method for the ongoing enhancement of the organization's data, aligned with their business objectives.

If we draw a parallel to the AAF, a similar product-oriented approach is employed. While the DOM defines DPs, the AAF refers to APs. Interestingly, the alignment between these two concepts is evident through the parallel structuring of DP and AP cards. The DP cards encapsulate information about APs, while AP cards establish a link to the corresponding DPs. This strategic linkage underscores the ongoing synchronization and harmonization of the AAF and DOM frameworks. This alignment is maintained, while still accommodating the requisite flexibility to prioritize objectives specific to Data and Application Architecture within each respective framework.
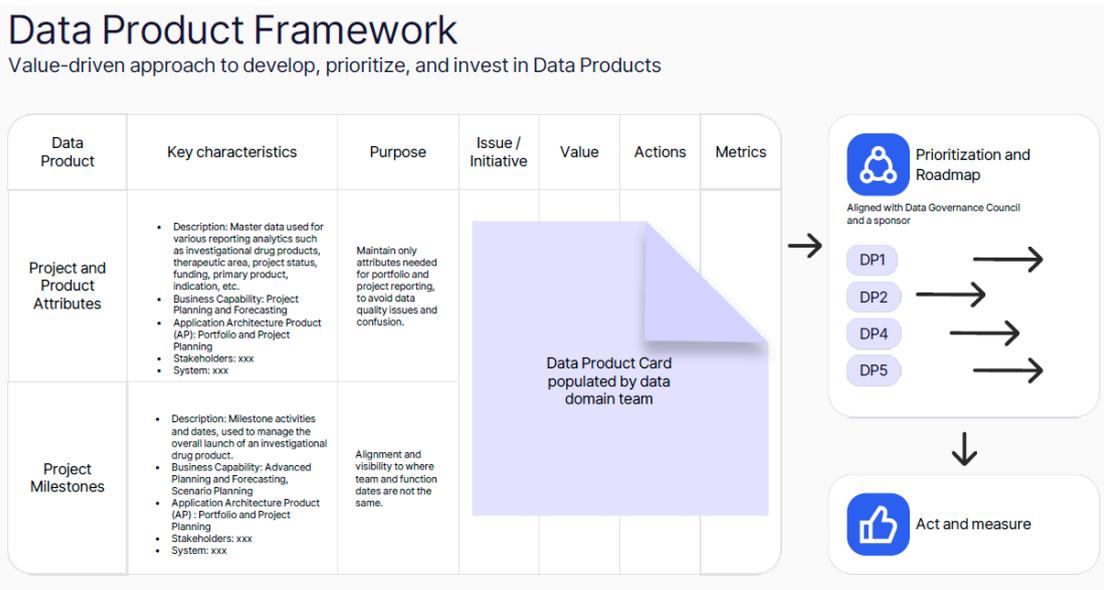
Figure 11. Data Product framework.

3. Data Domain Roadmaps

Integral to the DOM framework, the DD Roadmap stands as its third foundational component. This dynamic roadmap plays a pivotal role by aggregating prioritized DP initiatives and illustrating them along a timeline. In essence, the DD Roadmap functions as a strategic compass, steering the course of the evolution of the Global R&D's data capabilities. The essence of the DD Roadmap lies not only in its temporal portrayal, but also in its profound alignment with the Data Governance Council (DGC). This alignment contributes to conceptual coherence and cross-domain consistency for data-driven initiatives within the realm of the Global R&D. As these initiatives span diverse areas of the enterprise, the DGC is an essential structure to uphold a cross-functional alignment and overarching perspective on data-driven transformation.

In the intricate landscape of the Pharma company, the realization of complex integration initiatives is going to be handed over to the Data Fabric HPTs, entrusted with the crucial roles of maintaining architectural and engineering consistency, while driving effective implementation. This integration cycle completes the transformation journey from strategic envisioning to on-ground deployment, orchestrated with precision and coherence.

The strategic visualization offered by DD roadmaps empowers the Global R&D to judiciously allocate resources and synchronize projects with the overarching business objectives. As the landscape of data-driven initiatives evolves, the DD Roadmap remains a living and adaptable document, mirroring the agile nature of the organization's data evolution journey. Thus, as new initiatives emerge and existing ones are accomplished, the DD Roadmap maintains its relevance, steering the enterprise on a synchronized and progressive path towards improved data accessibility, quality enhancement, and heightened utility.

Notably, mirroring the DOM, the AAF also entails the formulation of an AD roadmap, which serves to aggregate AP investment projects and initiatives. Moreover, the strategic alignment between DD and AD roadmaps is distinctly apparent. Significant DD initiatives, with implications for APs, find their place within the corresponding AD roadmap. Conversely, items within the AD roadmap can foster data requirements that are addressed in the DD roadmap. This interconnection between DD and AD roadmaps underscores the strategic alignment of these frameworks.

At the time of writing this paper, the Pharma company's DDs had not yet defined their DD roadmaps. However, this crucial step was scheduled to be undertaken as part of the maturity journey, once DPs were defined and key information, including existing challenges and improvement initiatives, was populated in DP cards.

4. Federated Governance Framework

The Federated Governance Framework lies at the backbone of the DOM, presenting an intricate balance between centralization and decentralization. This framework serves as the connective tissue between strategic vision of newly defined structures, such as DDs and DPs, and operational execution, ensuring that the data-driven evolution within the Global R&D is both guided and agile.

- Centralized Facets of Governance

Centralization within the framework manifests through the Data Governance Council (DGC), a dynamic forum encompassing DD and AD Leads, Subject Matter Experts (SME), and stakeholders. The DGC functions as a nexus, fostering consistent communication channels to align and enhance the quality of data practices. This Council should ensure the conceptual consistency and synchronization of the Global R&D's data management strategies. Moreover, it expedites decision-making, ensuring that the value of data is maximized both in the present and the future, thereby catalyzing data-driven transformation.

For this purpose, the DGC introduces the following:

- A DGC Charter that outlines the scope and objectives for creating and managing data governance across the Global R&D, including accountabilities for functions supporting data management efforts across the company.
- An operational model underpinned by distinct roles, responsibilities, and a robust decision-making framework.
- Integration with the DOM Framework to harmonize and streamline data initiatives.

Notably, the AAF also has a centralized organ—the AAF Acceleration Squad—responsible for the continuous alignment of AD topics, supporting cross-functional decision-making and ensuring conceptual consistency of application architecture evolution. However, even if we imply that AAF and DD structures tend to merge over time through continuous integration and harmonization, the DGC and AAF Acceleration Squad will exist separately, keeping their focus on data and applications correspondingly.

- Decentralized Empowerment

Decentralization materializes through the autonomy of DD teams that are the driving force behind their respective DD objectives, orchestrate the evolution of DPs in alignment with their individual DD roadmaps. This decentralized approach empowers DDs to act as stewards of their data assets, fueling continuous growth and adaptability within their designated domains.

By entrusting DDs with ownership, the Federated Governance Framework brings data accountability closer to the business, where core business value exists, and emboldens agile decision-making, tailored to the specific needs and contexts of each domain. This approach mitigates bureaucracy and accelerates responsiveness, while contributing to efficient and aligned decision-making of the Global R&D data. By harmonizing, prioritizing, and aligning data initiatives and DD roadmaps with the strategic direction of the Pharma company, the framework optimizes resource allocation, allowing for the efficient achievement of data-driven capabilities.

Moreover, recognizing that data transcends domain boundaries, the framework encourages collaborative efforts between DD Leads. This cross-domain collaboration should ensure that data initiatives are not siloed, but rather orchestrated to create synergistic effects that amplify their impact.

One of the core objectives of the framework was to establish an efficient process of raising, aligning, and accomplishing data decision requests, which is depicted in Figure 12.
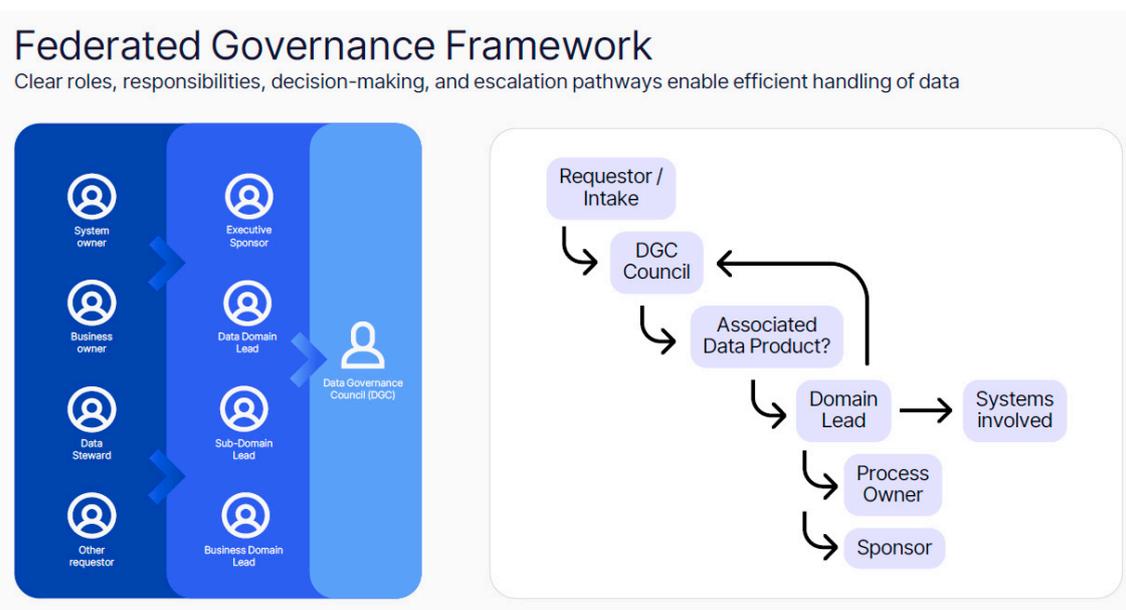
Figure 12. Federated Governance Framework.

Decision-Making Workflow within the Federated Governance Framework

1. Raising a Data Request or Initiative: Data requests or initiatives primarily originate at the domain level, where any DD Team member, such as a Business Owner, System Owner, Data Steward, or Data Domain/Subdomain Lead, can identify challenges, risks, inconsistencies, or improvement opportunities. They can raise a request using a specific request tool or directly during the DGC meeting by nominating the topic in advance. The Request Tool ensures that every data challenge, proposal, or initiative is captured and assessed.

2. Request Intake by the DGC: The DGC reviews and takes in the request for further assessment.

3. Assessment and Alignment: The DGC assesses the request, identifying the impacted data scope or DPs and the scale of change using the "T-shirt approach". This leads to the definition of stakeholders to drive collaborative decision-making. If a relevant DD is identified, decision-making authority shifts to the DD Lead. The DD Lead aligns the change with stakeholders, such as system owners, business owners, data stewards and sponsors, and presents the suggested decision to the DGC in the regular DGC forum. For major changes or data investment initiatives, the DGC orchestrates a validation of the suggested decision within the context of the overarching business strategy, technical feasibility, and alignment with data governance and architecture principles to ensure strategic and technological viability.

4. DGC Decision: The request is approved if the assessment and alignment confirm its validity. Otherwise, it is either rejected or postponed, pending further clarification.

5. Actions and Metrics: The DGC supports the DD Lead to define actions and metrics for the approved request or initiative that is incorporated into DP cards.

6. Alignment with DD Roadmap: If an approved initiative or major change request (T-size bigger than M) impacts DP or DD scope, it is integrated into the DD Roadmap. This integration provides a clear timeline for execution, milestones, and metrics to monitor outcomes. The DD Roadmap serves as the guiding blueprint for the evolution of the Global R&D data capabilities, ensuring synchronized and harmonized initiatives.

7. Execution and Iteration: Data initiatives are executed within the scope of the DD Roadmap, led by DD Leads. As initiatives are implemented, the framework supports an iterative approach, allowing for continuous improvement and adaptation based on emerging needs and insights. The Data Fabric HPTs, once established (see Section

"Phase II: Execution and Maturing of Data Fabric Capabilities"), will drive the implementation of complex data engineering and integration tasks, effectively bridging the gap between DDs and the Integration Competence Center.

The Federated Governance Framework emerges as an exquisite synthesis of the centralized leadership of the DGC and the decentralized autonomy of DDs, encapsulating the diverse facets of a modern data ecosystem. By distributing decision-making authority, promoting cross-domain collaboration, and aligning data initiatives with strategic goals, the framework should empower the enterprise to harness the full potential of its data assets.

As of the time of writing this paper, the Federated Governance Framework, in conjunction with the DGC and its operational model, was in the process of establishment. Consequently, we cannot furnish specific evidence of its acceptance or the realization of its value, at this point. However, it is anticipated that this framework will persistently evolve and adapt, drawing from the insights garnered through its iterative implementation.

3.3. Preliminary Results

The current Data Domain (DD) structure at the Global R&D, though aligned with the Application Domain (AD) structure, also reveals deviations, arising from the logical grouping of data flows. This underscores the reality that achieving a harmonized Data-Driven Architecture (DDA) necessitates an incremental approach, especially within the complexities of a large-scale organization, featuring a multitude of stakeholders and intricate system and business process landscapes. Consequently, the present DD structure stands as an interim consensus towards the harmonized DDA vision where data, applications, and business structures converge towards common domain objectives.

However, the alignment and integration extend beyond the DD and AD structures. As is summarized in Table A2, the journey towards harmonization encompasses various facets of the DOM and the AAF framework such as Products, Roadmaps, and Federated Governance mechanisms. It is crucial to acknowledge that the Pharma company's DDA framework is not aimed at realizing an ideal, fully integrated state, but rather an optimal balance of harmonization and flexibility of the DOM and AAF. This optimization implies incorporating strategic distinctions that preserve the necessary flexibility to prioritize objectives that are specific to Data and Application Architecture within each corresponding framework.

The essence of such a phased approach cannot be underestimated, as it contributes to the gradual assimilation of change and mitigates resistance of transformative shifts, enhancing stakeholder commitment. Despite the potential inefficiencies related to entering and exiting interim phases in terms of effort, communication, and time, this approach yields significant socio-psychological impacts and overall positive outcomes. Empirical validation of this approach emerged from the Global R&D case study. While evaluating potential DD structures initially favored the full integration of the AAF and DOM, the complexity of the existing EA and the size of the incorporated change perceived risks of internal resistance and associated costs outweighed the potential cost-effectiveness of the immediate, complete harmonization of the frameworks.

This experience gives rise to a vital methodological insight, whereby significant changes should incorporate one or more incremental interim phases. The number of required interim phases depends on a range of factors, including the size of the change, corporate culture, organizational structure, complexity, and the organization's scale.

Each big change should incorporate one or more interim phases.

Moreover, if the project faces social resistance to the change, a step back to the previous phase and a re-evaluation of the maturity journey towards the target view could be needed. This incremental approach is an effective instrument to manage the overall success of the digital transformation being a socio-technological transformation with a strong dominance of social aspects.

In essence, the current alignment with deviations reflects the strategic approach that seeks to create a flexible yet coherent framework, fostering a balance between the ideal DDA vision and the realities of implementation within a dynamic organization like the Pharma company. This strategic evolution, guided by the principles of incremental change and informed decision-making, positions the Pharma company to seamlessly adapt its architecture over time, while remaining responsive to the shifting demands of its industry and stakeholders.

Although continuous value-realization is a core principle of the DDA, it is a subject of extensive debates [77,78,80,81,84] and often extends over the long term. The success of this phase primarily manifests in its acceptance and commitment within the organization, as the DDA progresses along its maturity journey.

Evidence of this progress is drawn from the Stakeholders Feedback Survey, directed at the Domain Leads, featuring the following three key questions:

1.   In the journey to embed DDA, what has helped you the most? (a free-text question)

Responses provided insights, with the majority highlighting collaboration between Domains and tight partnership as instrumental in the transformation journey.

2.   Which aspects of support have been the most useful to date? (a dropdown question)

Eight out of ten respondents identified regular Acceleration Team Meetings as the most supportive aspect, showcasing the importance of regular meetings between Domain Team Leads and the Project Team.

3.   In your opinion, how self-sustaining are the Domains? e.g., meeting on a regular cadence, acting autonomously, tracking the performance of their OKRs? (with 1 being the lowest self-sustaining and 5 being the most self-sustaining)

The analysis showed a high level of framework acceptance and sustainability for this foundational phase of the project, reflected in an average score of 3.2 (with a range of 1 to 5). As the framework matures, this score is anticipated to rise, implying a reduced need for extensive change management efforts. Currently, however, there is a recognized necessity for an intensified socialization effort to foster a deeper understanding and broader adoption of the DDA.

This initial phase indicates a high degree of stakeholder acceptance, with collaborative efforts, knowledge sharing, and a community of practice emerging as significant instruments for the sustainable evolution of the framework. Looking forward, an acceleration effect on benefit realization for each DP and AP initiative is anticipated in the long term, facilitated by the foundational principles of the DDA. However, this effect is cumulative and may require time to manifest as the organization undergoes a transformative process.

3.4. Limitations and Recommendations

- Diverse Transformation Approaches

The process of defining and implementing the DOM at the Global R&D unveiled that similar transformational programs were concurrently in progress across other divisions of the Pharma company. Divisions such as Data Science and Commercial Operations were pursuing comparable objectives, albeit with distinctive strategies. For instance, in the realm of Commercial Operations, the distinct decision was made to omit the DD Layer. However, this was accompanied by a heightened emphasis on DPs and their individual ownership. This contrasts with the global R&D organization's current approach, which delineates DD ownership, while not explicitly accounting for individual accountability.

The existence of parallel transformational tracks beyond the Global R&D illuminates the dynamic nature of architectural transformation within the Pharma company, where diverse divisions adapt methodologies in alignment with their distinct contexts and requirements. This observation underscores a significant proposition, whereby the pursuit of harmonization extends beyond the AAF and DOM frameworks. It is evident that a broader harmonization of Data Architecture frameworks across the Pharma company must

be needed. This could potentially entail the formulation of a comprehensive global Data strategy for the Pharma company, serving as the overarching business strategy. Such a strategy should align not only diverse architectural frameworks, but also define common visions, principles, and definitions, while enabling efficient and flexible localization of the strategy across the divisions.

Strategic agility, an essential factor for navigating the success of digital transformation, underscores the imperative of adapting to evolving market needs and technological advancements [52,53,91]. Balancing centralization and decentralization in a data strategy emerges as a pivotal approach, empowering organizations to maintain agility. Centralization of core data standards, principles, and governance components at a global organizational level provides a solid foundation, ensuring consistency and alignment with overarching strategic objectives. Simultaneously, decentralizing the customization and tailoring of data strategy elements to align with the specifics of business units or divisions within the organization fosters adaptability and responsiveness. This balanced approach enables organizations to swiftly adapt to evolving data needs, while upholding global consistency and governance standards. It also contributes to agility in responding to changing market dynamics and technological opportunities, thereby enhancing competitiveness in the digital landscape.

This localization-versus-centralization balance of the data strategy is akin to defining a common language, while allowing for local jargons and lexicons or professional terminologies used by subgroups. In this manner, diverse groups can communicate using a common language, while retaining the flexibility to utilize a local language or professional lexicon for specific purposes. Striking a similar balance between centralization and decentralization of the Data strategy is crucial for the Pharma company. A corporate strategy aims to establish a common "framework language" to ensure interoperability, while providing the necessary agility and flexibility to adapt to division-specific needs.

Deviation from this balance towards centralization could lead to non-acceptance and a lack of commitment, particularly if the unique needs and intricacies of individual divisions are not adequately considered. It might also entail an unreasonable amount of alignment effort to develop a global approach that can effectively accommodate the diverse requirements of a large-scale organization. On the other hand, ignoring the need for a global Data strategy and maintaining decentralized local strategies could result in "misunderstandings" or a lack of interoperability between the divisions. For example, if each division has a different definition of DPs and uses different technologies to store, search, and share DPs, this approach might improve data transparency, discoverability, and reusability within one division, but could create even more siloed data ecosystems within the entire organization.

This balance is highly individual for each organization, being dependent on the corporate culture, organizational structure, industry specifics, maturity of digital and data capabilities, and many other internal and external determinants. Moreover, this balance may evolve with the organization's digital and data maturity. Companies may initially adopt a highly centralized data strategy and gradually move towards decentralization, or vice versa. In our case, elements of the data strategy were initially pursued separately in a decentralized manner across different organizations within the pharmaceutical company. However, as the need for a global data strategy became apparent at the global management level, the initiative for a global Data Product strategy was launched.

Notably, as of the time of publishing this paper, a global Data Product strategy was initiated in the Pharma company that confirms that the expressed necessity was acknowledged on the global management level.

> *Striking a balance between centralization and decentralization of the data strategy is crucial. A corporate strategy should aim to establish a common "framework language" to ensure interoperability, while providing the necessary agility and flexibility to localize the framework towards division-specific needs.*

- DDA: one size does not fit all.

While the results of this case study provide evidence of the iterative implementation of the DDA framework within the Global R&D, it is important to note that the applicability of DDA may vary across different organizations within the pharmaceutical R&D sector. The unique design and implementation decisions made within the Pharma company were deeply influenced by the organization's internal and external factors, including corporate culture, existing architecture, strategic objectives, ongoing initiatives, market conditions, financial considerations, and more. Therefore, while the case study serves as a guiding reference, its direct replication without careful adaptation may not yield the same results in other organizational contexts. For instance, the Global R&D had already fostered a culture of agility, robust cross-functional collaboration, and a readiness for continuous improvement and change. This pre-existing disposition facilitated the acceptance and commitment to transformative frameworks like DDA. Conversely, organizations with more conservative or less mature enterprise ecosystems in these aspects might face greater challenges in achieving a similar level of commitment, necessitating the consideration of more extensive change management measures or a more iterative approach to bridge the gap effectively.

Additionally, the fact that the AAF was initiated prior to the DOM framework allowed the latter to benefit from several AAF elements, implementation learnings, and an established acceptance of the transformation. Both frameworks share common core principles, reinforcing their alignment. The iterative evolution of the DDA framework within the Pharma company further underscores its context-specific nature. The journey towards the current state of DDA was marked by continuous learning, adaptation, and refinement, with each iterative step contributing to its maturation. It is worth noting that the current comprehensive view of DDA was not entirely foreseeable at the project's outset. The insights gained from each phase played a pivotal role in shaping the subsequent phases.

However, the case study serves as a valuable resource for other pharmaceutical R&D companies aiming to embark on a similar transformational journey. It provides a foundational framework that offers reference points and a starting point for analysis. Yet, the successful implementation of DDA requires a meticulous feasibility assessment and careful iterative execution, tailored to the specific circumstances of each organization. The deep integration of various elements, approaches, tactical decisions, and strategic visions must be considered in light of the organization's unique landscape.

- Productization and Customer Orientation

One of the primary objectives of the DOM was to shift data ownership towards the teams responsible for its production. Given that data production primarily occurs within the sphere of business operations, the Pharma company's DDs and DPs were meticulously designed with a primary focus on the existing operational usage of data. This operational-centric approach was indispensable for addressing current business needs and securing commitment from DD teams. However, it is essential to recognize that it represents just the starting point of a more comprehensive journey.

While this initial approach was pivotal, it is equally crucial to explore the untapped potential for deriving additional value from data by examining secondary data usages and accommodating new use cases and data customers. Such an approach contributes to gradually dismantling data silos that have naturally formed around operational data usage, thereby fostering data reuse throughout the entire organization. This shift is instrumental in maturing towards becoming a truly data-driven enterprise.

Consequently, the forthcoming phases of this transformation must expand their focus beyond operational usage. It is imperative to delve into the potential use cases of new DP customers, extending beyond the Global R&D. These may encompass Therapeutic Areas, Medical Affairs, Commercial Operations, and Data Science, among others. Gaining a profound understanding of their unique data needs and discerning how the global R&D function's data assets can be instrumental for them is paramount. This understanding should serve as a foundation for initiating Proof of Concepts (POCs) and feasibility analyses.

Subsequently, new DPs can be conceptualized and developed, promoting cross-functional and cross-company collaboration and facilitating the creation of value on a global scale.

It is crucial to acknowledge that expanding the scope of data usage represents a vital step in the journey. However, it should not diminish the significance of the initial phases. The foundation laid during these initial stages remains invaluable in establishing a robust data architecture that can efficiently support and accelerate the subsequent phases of the data-driven transformation.

- Replication of the Use Case

Acknowledging a potential concern about the use case replicability for another Pharma company and, therefore, a risk of research bias, we would like to highlight the expected usage of the results of this case study to mitigate these concerns. Firstly, we provide a very high-level conceptualization of the DDA framework (introduced in Section 2.2) and exemplify its individual tailoring for a Big Pharma company in Section 3. Our objective is not to present a universally applicable use case, but rather to offer transparency regarding the specificities of how DDA could be tailored, refined, and uniquely adjusted to meet the specific needs of the Pharma company. While we understand the unique nature of our case study, we believe that the approach and methods employed can serve as a valuable guide for other companies in the pharma industry R&D seeking to define their unique Data-Driven Enterprise Architecture.

4. Conclusions

In this study, we have addressed significant research gaps within the realm of enterprise data-driven transformation. Leveraging the Resource-Based View (RBV) and dynamic capabilities theories, we have defined a data-driven enterprise and transformation, as well as proposing three levels of data-driven capabilities, as follows: supportive, transformational, and accelerative. As follows, we developed the concept of Data-Driven Enterprise Architecture (DDA), which forms the foundation for data-driven transformation, accelerating the development of data-driven dynamic capabilities. The DDA stands out from mainstream scientific concepts and enterprise practices due to its holistic approach. Unlike many existing approaches that tend to focus on the transformation of application, business, and data architecture separately, DDA emphasizes the continuous harmonization of these crucial architectural components within each domain. In essence, DDA represents a pioneering framework for distributed and federated architecture at a scale that decomposes architecture into domain-specific components, while considering their continuous harmonization and alignment toward common domain business goals. This contributes to a consistent and interoperable ecosystem across domains, making it a valuable approach for organizations seeking a comprehensive architectural solution.

Although the DDA is a generic framework and could be theoretically applied for any industry after corresponding adjustments, we have empirically validated it within the pharmaceutical R&D ecosystem of the Pharma company. Herein, we summarize the key conceptual and empirical findings that have emerged from the Pharma company's digital transformative journey and implementation of the DDA structural elements such as Data Fabric strategy, the AAF and DOM, as follows:

1. **Leadership and Stakeholder Commitment:** Unwavering leadership support and stakeholder commitment are crucial for the success of any transformative efforts. The use case evidenced that intensive stakeholder involvement in defining the framework, its objectives, and implementation strategy fostered ownership, broad acceptance, and commitment to the change.
2. **Actionable, Measurable, and Adaptable Strategy:** A data-driven transformation strategy must not only be conceptual, but also actionable, measurable, and adaptable to succeed. The Pharma company's Data Fabric strategy, the AAF, DOM, and other structural elements of DDA addressed these objectives. The success of this phase is primarily reflected in the acceptance and commitment garnered within the organiza-

tion. In the long term, an accelerative effect on the socio-economic outcomes of each DD and AD initiative is anticipated that is facilitated by the DDA foundations (see the Chapter "Value-realization").

3. **Prioritizing Organizational and Mindset Change:** Successful socio-technological transformation prioritizes organizational and mindset change before technology adoption. DDA provides the foundation for this cultural and organizational shift that must accelerate data-driven transformation in an agile and scalable manner during the subsequent phases. While the concept of Data Fabric is inherently technology-centric, it became evident during the exploration phase that merely implementing Data Fabric technological advancements (such as knowledge graphs and active metadata) would not move the Global R&D closer to its goal of fostering a data-driven culture and empowering digital transformation for innovative medicine development. The critical factor influencing the value realization of any technological advancement is its adoption rate, which hinges on having a solid organizational and mindset foundation.

Therefore, the Global R&D within the Foundation Phase of Data Fabric established organizational structures such as DDs, cross-functional DD Teams, and their Operating Models. These structures were put in place to effectively address business requirements concerning Data Fabric technology, define implementation strategies, and closely monitor business value realization (see the Chapter "Phase I: Foundation of Data Fabric").

4. **Federalization:** DDA introduces the concept of federalization as a strategic balance between centralization and decentralization. This approach emphasizes the development of modularized domain architectures, marking a departure from traditional monolithic methods. While encouraging autonomy within DDs and ADs, DDA maintains interoperability and contextual consistency on a global scale.

Autonomous DDs and ADs can independently address domain-specific needs, fostering innovation and agility. Simultaneously, central governance organizations like the Data Governance Council (DGC) and the AAF Acceleration Squad harmonize and align these autonomous entities, ensuring that the broader organizational objectives and strategic directions are maintained. This balanced approach empowers the R&D ecosystem of the Pharma company to leverage the benefits of both centralization and decentralization, while achieving a cohesive and harmonized data-driven ecosystem.

5. **Navigating Complexity Through Decomposition:** DDA's approach of decomposing EA into smaller, cohesive domains such as DD and AD addresses the challenge of managing complexity. Structuring these domains around core business capabilities allows the Pharma company to focus on specific business functions and data requirements within each domain, while empowering DD Teams with required resources and mechanisms to develop targeted solutions.

6. **Assess Different Baseline Concepts for Optimal Data Architecture:** The Pharma company's Data Fabric approach evidenced the importance of the exploratory assessment of various baseline concepts related to data architecture, insights from industry experiences, internal expertise, subject matter experts' opinions, and the unique internal business needs of the enterprise to design an optimal Data Architecture. It is crucial not to be confined to a single mainstream model, such as Data Mesh or Data Fabric, as there is no one-size-fits-all solution in the realm of Data Architecture. Instead, this process should be envisioned as constructing a customized model, akin to assembling a set of Lego blocks. This model is constructed by selectively integrating elements from various baseline concepts, tailored precisely to meet the specific needs and goals of the enterprise.

Furthermore, this construction process does not conclude with a static solution but evolves iteratively, adapting to changing requirements and insights. Throughout its maturity journey, the model undergoes refinements and adjustments to ensure that it remains optimally aligned with the enterprise's evolving data landscape and strategic objectives.

This iterative approach contributes to the agility and responsiveness of the data architecture, accommodating the enterprise's dynamic needs effectively.

7. **Synergy of Data Mesh and Data Fabric.** The Data Fabric Strategy evidenced efficient integration of the Data Mesh and Data Fabric principles. While Data Mesh laid the foundational base of organizational structure and mindset for data management, Data Fabric formed the foundation of the technological underpinnings. This harmonious blend should facilitate structural realization of data-driven objectives through robust and scalable technological solutions.

8. **Collaboration and Alignment:** Continuous collaboration and alignment with ongoing projects within the Global R&D, such as AAF, as well as related initiatives across the whole organization, along with the iterative and adaptable implementation of DDA, are essential to mitigate the risk of creating "framework silos" and to emphasize that all dependent elements of various frameworks work in harmony. Like Lego blocks, these frameworks need not be identical in size or shape, but they must fit together cohesively. Thus, the focus lies on ensuring the necessary interoperability and compatibility between related elements of different frameworks. Even in cases where DD and AD structures exhibit deviations, these frameworks incorporate mechanisms like an integration of AD Leads into the DGC, mapping DPs with APs, and aligning DDs with AD Leads to guarantee close collaboration, interoperability, and alignment.

This approach is supposed not only to enhance short-term success, but also make the framework sustainable in the long run, as it fosters interoperability and maintains conceptual consistency in strategic objectives (see the Chapter "Diverse Transformation Approaches").

9. **Phased Approach:** Breaking down transformation into manageable chunks, such as the data-driven capabilities, domains, with their initiatives and implementation roadmaps, improve efficiency of the program management by delivering transparency, accountability, and contributing to stakeholder acceptance. In addition, transformational changes should incorporate one or more incremental interim phases. The number of required interim phases depends on numerous factors, including the size of the change, corporate culture, organizational structure, complexity, and the organization's scale.

In summary, this paper outlines a comprehensive DDA framework and early preliminary insights of its implementation within the pharmaceutical R&D ecosystem of the Pharma company embarking on the data-driven transformation journey. It underscores the importance of leadership commitment, stakeholder integration, actionable and measurable strategies, collaboration, cultural change, and business-centric approaches. Moreover, it emphasizes the significance of interoperability between frameworks and the continuous nature of value realization in data-driven transformations. We believe these key findings provide valuable insights for organizations seeking to embark on similar transformative journeys, offering guidance for creating an architectural foundation, leveraging data as a strategic asset in the ever-evolving landscape in pharmaceutical R&D.

Appendix A

Figure A1. Categorized Data Fabric components.
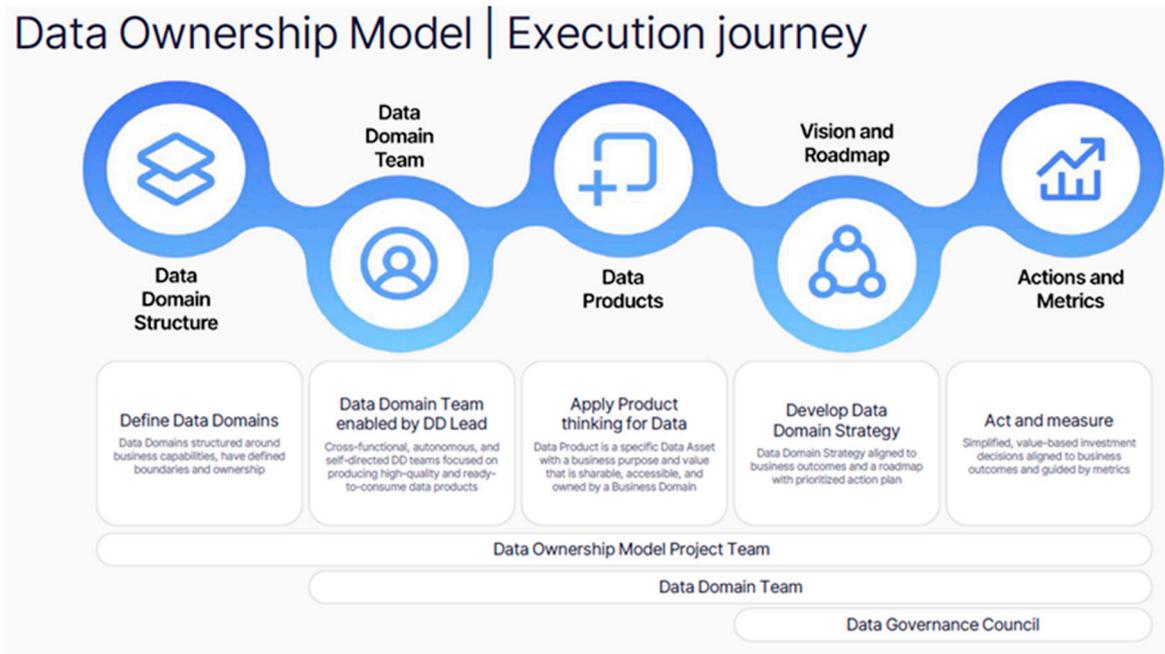
Figure A2. Execution journey of the Data Ownership Model.

Figure A3. Data product card.

Table A1. Data Domains/Subdomains and Data Products.

Data Domains	Planning and Forecasting	Study Participant	Study Management	Investigative Staff Engagement	Product Management	Others/Backlog
Subdomains	Portfolio Planning Budget Management Feasibility	Data Delivery Standards Medical Review Statistical programming RWE	Clinal Operations Country Operations Business Operations	Investigative Staff Engagement	Chemical Process R&D Discovery, Product Development and Supply Distribution	Safety Management Quality Management Regulatory Management
Data Products	Resource Forecasting OOP Forecasting Project and Product Attributes Project Milestones Forecasting Drivers Resource Capacity Project Accounting (Valuations, Actuals, etc.) Time reported Site information PI information Study information Site feasibility	Pharmacokinetics/ Pharmacodynamics datasets and analysis Phase 1 trials patient data (Case Report Form) Phase 1 trials source data ADaM datasets and metadata Study intake data DH Early Assessment data EHR data used in a study Lab data used in a study External real-world databases Tokens DH Technology Enablement board intake data DH Capability Master requirements data DH Technology/Supplier Qualification data Training data eConsent study data eCOA study data	In progress	Study site data Site identity and access management data Training data Contracts Enrollment Data Document collaboration Document repository Managed Access Requests Site relations management Site communication Site feedback collection	In progress	-

Table A2. Data Ownership Model (DOM) and Application Architecture Framework (AAF)—comparison of key elements.

Elements	DOM	AAF	Harmonization	Distinctions
Structure	Data Domains (DDs)	Application Domains (ADs)	Iterative harmonization towards optimal view (not full integration)	Exceptional deviation based on logical grouping of data flows and data categories
Products	Data Products (DPs)	Application Products (APs)	Alignment and Mapping	DP and AP are distinct elements (although interdependent)
Product cards	DP cards	AP cards	Interlinked within DP and AP cards	Cards have a different structure specific to their focus area
Roadmaps	DD roadmaps	AD roadmaps	Strategic alignment between DD and AD roadmaps and cross-reference of the selective items	Distinct focus on data and applications correspondingly
Federated Governance				
Centralized Facets	Data Governance Council (DGC)	Acceleration Squad	Strategic alignment of interrelated topics	Distinct focus on data and applications correspondingly
Decentralized Facets	Autonomous cross-functional DD Teams	Autonomous Squads of cross-functional AD teams	Harmonization through close correlation between DD and AD Leads. Thus, a Business AD Lead is always a member of DD team	DD and AD Teams are autonomous in their distinct focus areas such as data and applications

References

1. Hannila, H.; Silvola, R.; Härkönen, J.; Haapasalo, H. Data-Driven Begins with DATA; Potential of Data Assets. *J. Comput. Inf. Syst.* **2022**, *62*, 29–38. [CrossRef]
2. Coyle, D.; Manley, A. *University of Cambridge. What Is the Value of Data? A Review of Empirical Methods*; Bennett Institute for Public Policy of the University of Cambridge: Cambridge, UK, 2022.
3. Assur, N.; Rowshankish, K. The Data-Driven Enterprise of 2025. McKinsey & Company. 28 January 2022. Available online: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025 (accessed on 12 November 2023).
4. NewVantage Partners. Data and Analytics Leadership Annual Executive Survey 2023: Executive Summary of Findings. 2023. Available online: https://static1.squarespace.com/static/62adf3ca029a6808a6c5be30/t/63a477e23d7f9162b7771d6b/1671722982363/WAVESTONE+NVP--+Data+&+Analytics+Executive+Leadership+2023--Survey+Findings.pdf (accessed on 12 November 2023).
5. Neely, A. Why You Should Be Treating Your Data as an Asset. Anmut. Available online: https://www.anmut.co.uk/why-you-should-be-treating-your-data-as-an-asset/ (accessed on 12 November 2023).
6. Johnston, M. Biggest Companies in the World by Market Cap. Investopedia. Available online: https://www.investopedia.com/biggest-companies-in-the-world-by-market-cap-5212784 (accessed on 12 November 2023).
7. Ponemon Institute; IBM. *Cost of a Data Breach Report 2022*; IBM Corporation: Armonk, NY, USA, 2022. Available online: https://www.ibm.com/downloads/cas/3R8N1DZJ (accessed on 12 November 2023).
8. Brynjolfsson, E.; Wang, J.; McElheran, K. The Power of Prediction: Predictive Analytics, Workplace Complements, and Business Performance. *Bus. Econ.* **2021**, *56*, 217–239. [CrossRef]
9. Ertz, M.; Sun, S.; Latrous, I. The Impact of Big Data on Firm Performance. In *Advances in Intelligent Systems and Computing*; Springer: Cham, Switzerland, 2021. [CrossRef]
10. Müller, O.; Fay, M.; Brocke, J.V. The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *J. Manag. Inf. Syst.* **2018**, *35*, 488–509. [CrossRef]
11. Mikalef, P.; Boura, M.; Lekakos, G.; Krogstie, J. Big Data Analytics and Firm Performance: Findings from a Mixed-Method Approach. *J. Bus. Res.* **2019**, *98*, 261–276. [CrossRef]
12. Gul, R.; Ellahi, N. The Nexus between Data Analytics and Firm Performance. *Cogent Bus. Manag.* **2021**, *8*, 1923360. [CrossRef]
13. Schweikl, S.; Obermaier, R. Lessons from Three Decades of IT Productivity Research: Towards a Better Understanding of IT-Induced Productivity Effects. *Manag. Rev. Q.* **2019**, *70*, 461–507. [CrossRef]
14. Hariry, R.E.; Barenji, R.V.; Paradkar, A. Towards Pharma 4.0 in clinical trials: A future-orientated perspective. *Drug Discov. Today* **2022**, *27*, 315–325. [CrossRef] [PubMed]

15. Sharma, D.K.; Patel, P.K.; Shah, M. A comprehensive study on Industry 4.0 in the pharmaceutical industry for sustainable development. *Environ. Sci. Pollut. Res.* **2023**, *30*, 90088–90098. [CrossRef] [PubMed]

16. Tetteh, M.G.; Jagtap, S.; Salonitis, K. Pharma 4.0: Revealing drivers of the digital transformation in the pharma sector. In *Lecture Notes in Mechanical Engineering*; Springer: Cham, Switzerland, 2023; pp. 528–535. [CrossRef]

17. Dulko, D.; Kwong, M.; Palm, M.E.; Trinquart, L.; Selker, H.P. From a decentralized clinical trial to a decentralized and clinical-trial-in-a-box platform: Towards patient-centric and equitable trials. *J. Clin. Transl. Sci.* **2023**, *7*, e236. [CrossRef]

18. Eltoukhy, A.E.E.; Shaban, I.A.; Chan, F.T.S.; Abdel-Aal, M.a.M. Data Analytics for Predicting COVID-19 Cases in Top Affected Countries: Observations and Recommendations. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7080. [CrossRef]

19. Hariry, R.E.; Barenji, R.V.; Paradkar, A. From industry 4.0 to pharma 4.0. In *Handbook of Smart Materials, Technologies, and Devices*; Springer: Cham, Switzerland, 2021; pp. 1–22. [CrossRef]

20. Inuwa, H.M.; Raja, A.R.; Kumar, A.; Singh, B.; Singh, S. Status of Industry 4.0 applications in healthcare 4.0 and Pharma 4.0. *Mater. Today Proc.* **2022**, *62*, 3593–3598. [CrossRef]

21. Malheiro, V.; Duarte, J.M.; Veiga, F.; Mascarenhas-Melo, F. Exploiting Pharma 4.0 Technologies in the Non-Biological Complex Drugs Manufacturing: Innovations and Implications. *Pharmaceutics* **2023**, *15*, 2545. [CrossRef] [PubMed]

22. Dong, Y.; Wang, W.; Xing, Y.; Du, J.; Meng, Q. Data-Driven modeling methods and techniques for pharmaceutical processes. *Processes* **2023**, *11*, 2096. [CrossRef]

23. Kumar, S.A.; Kumar, T.D.A.; Beeraka, N.M.; Pujar, G.V.; Singh, M.; Akshatha, H.S.; Bhagyalalitha, M. Machine learning and deep learning in data-driven decision making of drug discovery and challenges in high-quality data acquisition in the pharmaceutical industry. *Future Med. Chem.* **2022**, *14*, 245–270. [CrossRef] [PubMed]

24. Vora, L.K.; Gholap, A.D.; Jetha, K.; Thakur, R.R.S.; Solanki, H.K.; Chavda, V.P. Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics* **2023**, *15*, 1916. [CrossRef] [PubMed]

25. Dara, S.; Dhamercherla, S.; Jadav, S.S.; Babu, C.M.; Ahsan, M.J. Machine Learning in Drug Discovery: A review. *Artif. Intell. Rev.* **2021**, *55*, 1947–1999. [CrossRef] [PubMed]

26. Koutroumpa, N.-M.; Papavasileiou, K.D.; Papadiamantis, A.G.; Melagraki, G.; Afantitis, A. A Systematic Review of Deep Learning Methodologies Used in the Drug Discovery Process with Emphasis on In Vivo Validation. *Int. J. Mol. Sci.* **2023**, *24*, 6573. [CrossRef]

27. Tang, M.; Li, B.; Chen, H. Application of message passing neural networks for molecular property prediction. *Curr. Opin. Struct. Biol.* **2023**, *81*, 102616. [CrossRef]

28. Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R.K. Artificial intelligence in drug discovery and development. *Drug Discov. Today* **2021**, *26*, 80–93. [CrossRef]

29. Bhattamisra, S.K.; Banerjee, P.; Gupta, P.; Mayuren, J.; Patra, S.; Candasamy, M. Artificial intelligence in pharmaceutical and healthcare research. *Big Data Cogn. Comput.* **2023**, *7*, 10. [CrossRef]

30. Nag, S.; Baidya, A.T.; Mandal, A.; Mathew, A.T.; Das, B.; Devi, B.; Kumar, R. Deep learning tools for advancing drug discovery and development. *3 Biotech* **2022**, *12*, 110. [CrossRef] [PubMed]

31. Huo, L.; Tang, Y. Multi-Objective Deep Reinforcement learning for personalized dose optimization based on Multi-Indicator Experience replay. *Appl. Sci.* **2022**, *13*, 325. [CrossRef]

32. Mhatre, S.; Shukla, S.; Chavda, V.P.; Gandikota, L.; Patravale, V. AI and ML for Development of Cell and Gene Therapy for Personalized Treatment. In *Bioinformatics Tools for Pharmaceutical Drug Product Development*; Scrivener Publishing LLC: Beverly, MA, USA, 2023; pp. 371–400. [CrossRef]

33. Quazi, S. Artificial intelligence and machine learning in precision and genomic medicine. *Med. Oncol.* **2022**, *39*, 120. [CrossRef]

34. Elkin, M.E.; Zhu, X. Predictive modeling of clinical trial terminations using feature engineering and embedding learning. *Sci. Rep.* **2021**, *11*, 3446. [CrossRef]

35. Yang, F.; Zhang, Q.; Ji, X.; Zhang, Y.; Li, W.; Peng, S.; Xue, F. Machine learning applications in drug repurposing. *Interdiscip. Sci. Comput. Life Sci.* **2022**, *14*, 15–21. [CrossRef]

36. Krishnamurthy, N.; Grimshaw, A.; Axson, S.A.; Choe, S.H.; Miller, J.E. Drug repurposing: A systematic review on root causes, barriers and facilitators. *BMC Health Serv. Res.* **2022**, *22*, 970. [CrossRef] [PubMed]

37. Kumar, Y.; Koul, A.; Singla, R.; Ijaz, M.F. Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda. *J. Ambient Intell. Humaniz. Comput.* **2022**, *14*, 8459–8486. [CrossRef]

38. Obrezanova, O. Artificial intelligence for compound pharmacokinetics prediction. *Curr. Opin. Struct. Biol.* **2023**, *79*, 102546. [CrossRef] [PubMed]

39. Chou, W.-C.; Lin, Z. Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling. *Toxicol. Sci.* **2022**, *191*, 1–14. [CrossRef]

40. Chapman, A.; Peterson, K.; Alba, P.R.; DuVall, S.L.; Patterson, O.V. Detecting Adverse Drug Events with Rapidly Trained Classification Models. *Drug Saf.* **2019**, *42*, 147–156. [CrossRef]

41. Teubner, R.A.; Stockhinger, J. Literature Review: Understanding Information Systems Strategy in the Digital Age. *J. Strateg. Inf. Syst.* **2020**, *29*, 101642. [CrossRef]

42. Riedl, R.; Stieninger, M.; Muehlburger, M.; Koch, S.; Heß, T. What Is Digital Transformation? A Survey on the Perceptions of Decision-Makers in Business. *Inf. Syst. E-Bus. Manag.* **2023**. [CrossRef]

43. Hanelt, A.; Bohnsack, R.; Marz, D.; Marante, C. A Systematic Review of the Literature on Digital Transformation: Insights and Implications for Strategy and Organizational Change. *J. Manag. Stud.* **2020**, *58*, 1159–1197. [CrossRef]

44. Mahboub, H.; Sadok, H. Contribution to a Better Digital Transformation Implementation: An Integrative Approach. In *Machine Learning and Artificial Intelligence*; Frontiers in Artificial Intelligence and Applications; IOS Press: Amsterdam, The Netherlands, 2023. [CrossRef]

45. Schneider, M.; Kanbach, D.K.; Kraus, S.; Dabić, M. Transform Me If You Can: Leveraging Dynamic Capabilities to Manage Digital Transformation. *IEEE Trans. Eng. Manag.* **2023**, 1–15. [CrossRef]

46. Rashedi, J. *The Data-Driven Organization: Using Data for the Success of Your Company*; Springer: Cham, Switzerland, 2023. [CrossRef]

47. Hupperz, M.; Gür, I.; Möller, F.; Otto, B. What Is a Data-Driven Organization? In Proceedings of the Conference AMCIS 2021. Digital Innovation and Entrepreneurship, Virtual, 13–19 August 2021.

48. Hebbar, P. *10 Indicators of a Truly Data-Driven Organisation*; Analytics India Magazine: Bengaluru, India, 2019. Available online: https://analyticsindiamag.com/10-indicators-of-a-truly-data-driven-organisation/ (accessed on 10 April 2024).

49. Janmeskens. The Road to a Data-Driven Organization: Navigating Data Dilemmas. Medium. 3 December 2023. Available online: https://medium.com/@meskensjan/the-road-to-a-data-driven-organization-navigating-data-dilemmas-898c612895af (accessed on 12 November 2023).

50. Thusoo, A.; Sarma, J.S. *Creating a Data-Driven Enterprise with DataOps*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017.

51. Yang, H.-L.; Morgan, S.L. Business Strategy and Corporate Governance: Theoretical and Empirical Perspectives. In *Business Strategy and Corporate Governance in the Chinese Consumer Electronics Sector*; Elsevier: Amsterdam, The Netherlands, 2011; pp. 23–46. [CrossRef]

52. Warner, K.; Wäger, M. Building Dynamic Capabilities for Digital Transformation: An Ongoing Process of Strategic Renewal. *Long Range Plan.* **2019**, *52*, 326–349. [CrossRef]

53. Ellström, D.; Holtström, J.; Berg, E.; Josefsson, C. Dynamic Capabilities for Digital Transformation. *J. Strategy Manag.* **2021**, *15*, 272–286. [CrossRef]

54. Verhoef, P.C.; Broekhuizen, T.; Bart, Y.; Bhattacharya, A.; Dong, J.Q.; Fabian, N.E.; Haenlein, M. Digital Transformation: A Multidisciplinary Reflection and Research Agenda. *J. Bus. Res.* **2021**, *122*, 889–901. [CrossRef]

55. Teece, D.J. The Evolution of the Dynamic Capabilities Framework. In *FGF Studies in Small Business and Entrepreneurship*; Springer: Cham, Switzerland, 2022; pp. 113–129. [CrossRef]

56. Cannas, R. Exploring Digital Transformation and Dynamic Capabilities in Agrifood SMEs. *J. Small Bus. Manag.* **2021**, *61*, 1611–1637. [CrossRef]

57. Chirumalla, K. Building Digitally-Enabled Process Innovation in the Process Industries: A Dynamic Capabilities Approach. *Technovation* **2021**, *105*, 102256. [CrossRef]

58. Helfat, C.E.; Peteraf, M.A. The Dynamic Resource-based View: Capability Lifecycles. *Strateg. Manag. J.* **2003**, *24*, 997–1010. [CrossRef]

59. Curry, E.; Tuikka, T. An Organizational Maturity Model for Data Spaces: A Data Sharing Wheel Approach. In *Data Spaces*; Springer: Cham, Switzerland, 2022; pp. 21–42. [CrossRef]

60. Wade, M.; Hulland, J. Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research. *Manag. Inf. Syst. Q.* **2004**, *28*, 107. [CrossRef]

61. Teece, D.J.; Leih, S. Uncertainty, Innovation, and Dynamic Capabilities: An Introduction. *Calif. Manag. Rev.* **2016**, *58*, 5–12. [CrossRef]

62. Teece, D.J. Explicating Dynamic Capabilities: The Nature and Microfoundations of (Sustainable) Enterprise Performance. *Strateg. Manag. J.* **2007**, *28*, 1319–1350. [CrossRef]

63. Teece, D.J. Dynamic Capabilities and (Digital) Platform Lifecycles. In *Advances in Strategic Management*; Emerald Publishing Limited: Leeds, UK, 2017; pp. 211–225. [CrossRef]

64. Soluk, J.; Kammerlander, N. Digital Transformation in Family-Owned Mittelstand Firms: A Dynamic Capabilities Perspective. *Eur. J. Inf. Syst.* **2021**, *30*, 676–711. [CrossRef]

65. Witschel, D.; Döhla, A.; Kaiser, M.; Voigt, K.; Pfletschinger, T. Riding on the Wave of Digitization: Insights How and under What Settings Dynamic Capabilities Facilitate Digital-Driven Business Model Change. *J. Bus. Econ.* **2019**, *89*, 1023–1095. [CrossRef]

66. Chui, M.; Hazan, E.; Roberts, R.; Singla, A.; Smaje, K.; Sukharevsky, A.; Yee, L.; Zemmel, R. *The Economic Potential of Generative AI: The next Productivity Frontier*; McKinsey: New York, NY, USA, 2023. Available online: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#industry-impacts (accessed on 18 January 2024).

67. Chen, Z.; Ayinde, O.R.; Fuchs, J.R.; Sun, H.; Ning, X. G2Retro as a Two-Step Graph Generative Models for Retrosynthesis Prediction. *Commun. Chem.* **2023**, *6*, 102. [CrossRef] [PubMed]

68. Bertelsmann Stiftung; Thiel, R.; Deimel, L.; Schmidtmann, D.; Piesche, K.; Hüsing, T.; Rennoch, J.; Stroetmann, V.; Stroetmann, K. #SmartHealthSystems Digitalisierungsstrategien im internationalen Vergleich. Available online: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/smarthealthsystems (accessed on 12 November 2023).

69. Dykes, B. The Four Key Pillars to Fostering a Data-Driven Culture. Forbes. 28 March 2019. Available online: https://www.forbes.com/sites/brentdykes/2019/03/28/the-four-key-pillars-to-fostering-a-data-driven-culture/?sh=6cf7bc367d90 (accessed on 12 November 2023).

70. LaBerge, L.; Smaje, K.; Zemmel, R.W. *Three New Mandates for Capturing a Digital Transformation's Full Value*; McKinsey & Company: New York, NY, USA, 2022. Available online: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value#/ (accessed on 12 November 2023).

71. Forth, P.; Reichert, T.; De Laubier, R.; Chakraborty, S. Flipping the Odds of Digital Transformation Success. BCG Global. Available online: https://www.bcg.com/publications/2020/increasing-odds-of-success-in-digital-transformation (accessed on 12 November 2023).

72. Shein, E. 8 Reasons Why Digital Transformations Fail. CIO. Available online: https://www.cio.com/article/228268/12-reasons-why-digital-transformations-fail.html (accessed on 14 November 2023).

73. Lawton, G.; Roy, M. 11 Reasons Why Digital Transformations Fail, Explained by Pros. TechTarget. 20 October 2023. Available online: https://www.techtarget.com/searchcio/tip/Top-6-reasons-why-digital-transformation-failures-happen (accessed on 14 November 2023).

74. Van De Wetering, R. The Role of Enterprise Architecture-Driven Dynamic Capabilities and Operational Digital Ambidexterity in Driving Business Value under the COVID-19 Shock. *Heliyon* **2022**, *8*, e11484. [CrossRef]

75. Van De Wetering, R. Enterprise Architecture Resources, Dynamic Capabilities, and Their Pathways to Operational Value. In Proceedings of the Conference: International Conference on Information Systems (ICIS), Munich, Germany, 15–18 December 2019.

76. Van Den Berg, M.; Slot, R.; Van Steenbergen, M.; Faasse, P.; Van Vliet, H. How Enterprise Architecture Improves the Quality of IT Investment Decisions. *J. Syst. Softw.* **2019**, *152*, 134–150. [CrossRef]

77. Ahlemann, F.; Legner, C.; Lux, J. A Resource-Based Perspective of Value Generation through Enterprise Architecture Management. *Inf. Manag.* **2021**, *58*, 103266. [CrossRef]

78. Pattij, M.; Van De Wetering, R.; Kusters, R.J. Enhanced Digital Transformation Supporting Capabilities through Enterprise Architecture Management: A fsQCA Perspective. *Digit. Bus.* **2022**, *2*, 100036. [CrossRef]

79. Van De Wetering, R. Understanding the Impact of Enterprise Architecture Driven Dynamic Capabilities on Agility: A Variance and FSQCA Study. *Pac. Asia J. Assoc. Inf. Syst.* **2021**, *13*, 32–68. [CrossRef]

80. Törmer, R.L. Dynamic Capability Building In The Lego Group—Prospective Activities Vs. Reflective Learning In Preparation for a Turbulent Digital Future. AIS Electronic Library (AISeL). Available online: https://aisel.aisnet.org/ecis2019_rp/136/ (accessed on 12 November 2023).

81. Enterprise Architecture Artifacts Facilitating The Strategy Planning Process for Digital Transformations: A Systematic Literature Review and Multiple Case Study. *IADIS Int. J. Comput. Sci. Inf. Syst.* **2021**, *16*, 46–62. [CrossRef]

82. Kotusev, S.; Kurnia, S.; Dilnutt, R. The Practical Roles of Enterprise Architecture Artifacts: A Classification and Relationship. *Inf. Softw. Technol.* **2022**, *147*, 106897. [CrossRef]

83. Zachman, J.A. Business Systems Planning and Business Information Control Study: A Comparison. *Ibm Syst. J.* **1982**, *21*, 31–53. [CrossRef]

84. Buckl, S.; Schweda, C.M. *On the State-of-the-Art in Enterprise Architecture Management Literature*, Technical Report; Technical University of Munich: München, Germany, 2011.

85. Herscovitch, L.; Meyer, J.P. Commitment to Organizational Change: Extension of a Three-Component Model. *J. Appl. Psychol.* **2002**, *87*, 474–487. [CrossRef] [PubMed]

86. Armenakis, A.A.; Harris, S.G. Crafting a Change Message to Create Transformational Readiness. *J. Organ. Chang. Manag.* **2002**, *15*, 169–183. [CrossRef]

87. Meyer, J.P.; Herscovitch, L. Commitment in the Workplace: Toward a General Model. *Hum. Resour. Manag. Rev.* **2001**, *11*, 299–326. [CrossRef]

88. Broda, E. Managing a Federated Data Product Ecosystem—Towards Data Science. Medium. 16 January 2023. Available online: https://towardsdatascience.com/managing-a-federated-data-product-ecosystem-3c6bff94c728 (accessed on 12 November 2023).

89. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]

90. Dehghani, Z. What Is Data Mesh? In *Data Mesh: Delivering Data-Driven Value at Scale*, 1st ed.; O'Reilly Media: Sebastopol, CA, USA, 2022; pp. 49–50.

91. Teece, D.J.; Peteraf, M.A.; Leih, S. Dynamic capabilities and organizational agility: Risk, uncertainty, and strategy in the innovation economy. *Calif. Manag. Rev.* **2016**, *58*, 13–35. [CrossRef]