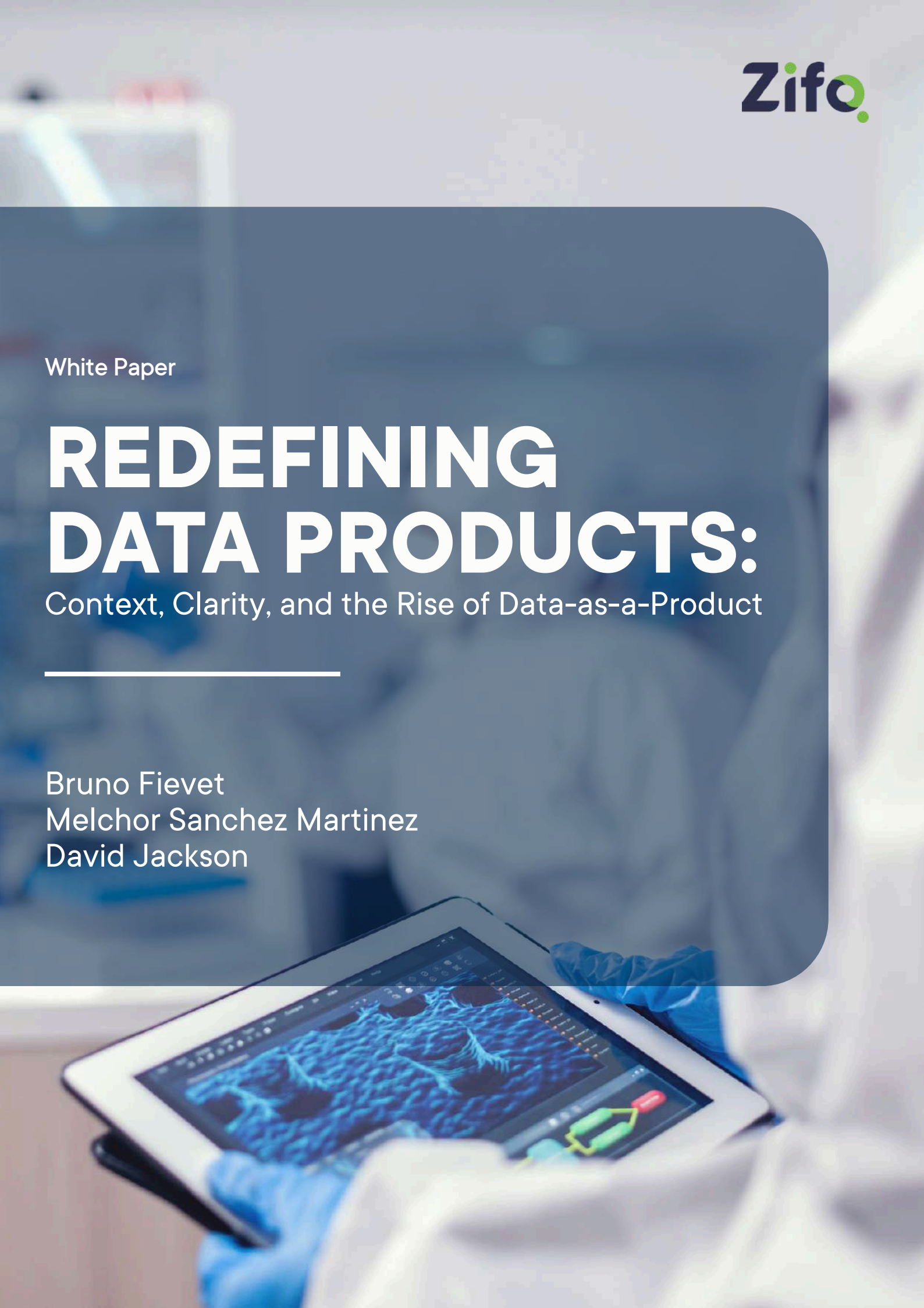
White Paper

REDEFINING DATA PRODUCTS:

Context, Clarity, and the Rise of Data-as-a-Product

Bruno Fievet
Melchor Sanchez Martinez
David Jackson



Abstract

The term “data product” is widely used but inconsistently understood across industries. This paper clarifies the ambiguity by distinguishing between two valid interpretations: curated datasets that enable data reuse, referred to as Lite Data Products, and digital solutions that convert data into measurable business value, referred to as Full Data Products.

Context determines which definition applies, and organizations seeking to operate in value-driven, outcome-oriented models should prioritize the latter. Building on this distinction, the paper introduces the concept of Data-as-a-Product.

This is a mindset and operating model that applies product management principles such as ownership, lifecycle management, and continuous improvement to data itself, transforming it from a passive asset into a managed, living source of sustained value.

PART 01 What Is a Data Product?

The term “data product” has become widely used in data, analytics, and digital transformation circles.
Yet, for many organizations, it remains a source of ambiguity because the same term is used in different ways by different communities. In simple terms, the context of utility matters.

On one hand, there are organizations whose business is data. They collect, curate, package, and sell data to other companies. For them, a data product often means a curated dataset that is ready for marketplace consumption.

On the other hand, there are science-driven organizations whose business is not data per se, but the use of data as an enabler of drug development, manufacturing, regulation, and commercial operations.

In this case, a data product is far more aptly aligned with a digital solution that converts data into business value through insights, decisions, or automation.

At the heart of the debate are two distinct perspectives. The dataset-centric view defines a data product as a curated, ready-to-use dataset, often accompanied by schema, metadata, and documentation.

It emphasizes standardization, quality, and reusability. This perspective emerged primarily from management consulting and enterprise architecture circles, where the goal was to improve consistency and governance across large organizations.

In practice, it aligns closely with the concept of a data contract—a defined agreement between producer and consumer about what data will be delivered, in what structure, and under what quality guarantees.

The digital solution-centric view defines a data product as a managed digital solution that converts data into business or scientific value through analytics, visualization, or automation.

It includes dashboards, models, and digital applications that are designed, owned, and iteratively improved like software products.

This view emerged from technology-driven organizations and frameworks such as data mesh, which emphasize decentralization, domain ownership, and value delivery. Both perspectives use the same term, but they point to different parts of the data value chain.

Understanding this distinction is essential for organizations seeking to invest effectively in data.
Unlike raw data, a product implies additional dimensions.

It is created with a clear user or business goal in mind. It is not just a “data dump” but structured to solve a specific problem. It is packaged with metadata, semantics, and documentation, accessible, trusted, and reusable. It is managed like a product, with an owner or steward, a roadmap of improvements, and performance objectives.

In short, if something is called a data product, it goes beyond mere data; it implies packaging, purpose, and lifecycle.

If you buy a physical product, you expect it to be ready to use, you shouldn't have to screw it together. The same applies to data products."

Stefan Mittelmaier,
Head of Digital Transformation
Global Quality Control at Merck

Two Definitions in Practice

In organizations that sell or monetize data, a data product often takes the form of a packaged dataset. A provider creates a market intelligence dataset, curates and standardizes it, documents it, and publishes it for clients to consume.

The dataset may be reused in many ways by many customers, but the connection between the dataset and a specific downstream business outcome is often indirect or varied.

In organizations whose primary business is not data selling, but rather using data to generate value, a data product much more often means a digital solution.

Examples include a dashboard that uses quality assurance data and analytics to provide decision support, a machine-learning model embedded into manufacturing operations to optimize yield, or a digital twin of a supply-chain process that integrates data, simulation, and user experience for continuous monitoring and improvement.

In this model, the dataset is typically an input, but the product is the solution. The value is directly linked to a business case, such as better quality control, faster product development, reduced deviations, or improved regulatory compliance.

When objectives are transformative, value must be articulated differently: not as incremental savings but as enabling capabilities that were previously impossible. In many circumstances, we see that traditional ROI metrics fail to capture strategic impact, requiring narrative-driven value articulation."

Adam Paton,
Head of Strategic Accounts at Zifo

Analogy: Components and Solutions

Across the life sciences industry, success depends on transforming scientific components into validated, usable solutions.

The same principle applies to data. Scientific components, such as compounds or reagents, are analogous to curated datasets. They are high-quality, standardized, and well-documented building blocks that enable discovery and innovation.

However, on their own, they do not deliver measurable business or scientific outcomes. Their value is realized only when they are combined, interpreted, or integrated into something purposeful.

Every person is interpreting data set their own way. Individual data points only gain meaning through domain knowledge and interpretation."

Jens Stein,
LIMS Product Manager at BASF

Finished solutions, by contrast, represent the application of these components in context, such as a validated diagnostic test or a predictive model for patient outcomes. These solutions integrate data, analytics, infrastructure, and user experience to deliver actionable insights or automation that directly advance research, operations, or decision-making.

Lite Data Products resemble scientific components — dependable, reusable foundations that enable interoperability and experimentation.

Full Data Products resemble finished solutions — integrated, outcome-oriented systems that deliver tangible business or scientific value.

Both are indispensable within a mature data ecosystem, but they serve different purposes. Components enable consistency and reuse; solutions enable decision-making and impact.

A Two-Tiered Definition for Clarity

We propose the following classification: Lite Data Products are curated, standardized datasets packaged for reuse or resale, where the business model centers on distributing the data itself. Full Data Products are managed digital solutions that use data, analytics, and user experience to convert data into specific business value. They are characterized by design for purpose, ownership, lifecycle, and clear business linkage.

This two-tiered definition is critical because, right now, **'data product'** has become jargon that means everything and nothing. In science organizations, that confusion leads to wasted time and money spent.

Just to reiterate: **Lite Data Products** are your curated datasets — ready to use but waiting for purpose.
Full Data Products are the complete digital solutions that deliver results. When you separate these two, leaders can finally prioritize correctly, and it gives everyone a shared understanding.

And **for science-driven organizations, the Full Data Product definition is usually the most useful and strategic**, especially when it comes to investment decision-making and use case prioritization.

PART 02 Data-as-a-Product

Having distinguished between Lite and Full Data Products, it is worth considering the philosophy that underpins them both — the idea of managing data-as-a-product.

Whereas a data product is a tangible artifact or solution, data-as-a-product refers to the way of working that makes those solutions valuable, sustainable, and aligned with business needs.

It brings together the disciplines of data management and product management to ensure that every piece of data, whether it powers a Lite or Full Data Product, is designed, maintained, and evolved with intention and accountability.

When an organization adopts the data-as-a-product mindset, every data initiative embodies five key principles: purpose and audience, ownership and accountability, lifecycle management, quality and discoverability, and user-centric value creation.

Governance cannot be an afterthought; it must be integrated into the design and implementation phase to ensure compliance and trust.

“Governance cannot be an afterthought. It has to be part of the product itself. And these data products are not static, they need to evolve continuously, like living entities.”

Marc De Luca,
Director, Global IT Business Development Lead at Bayer

Managing a Lite Data Product requires the definition of a data contract, a formal agreement that describes how the dataset should be consumed, including ontology, data model, metadata, and quality rules.

Managing a Full Data Product requires the creation of a specification artifact that references datasets and includes all additional information needed to design, build, and operate the product. Both practices ensure clarity, trust, and reusability over time.

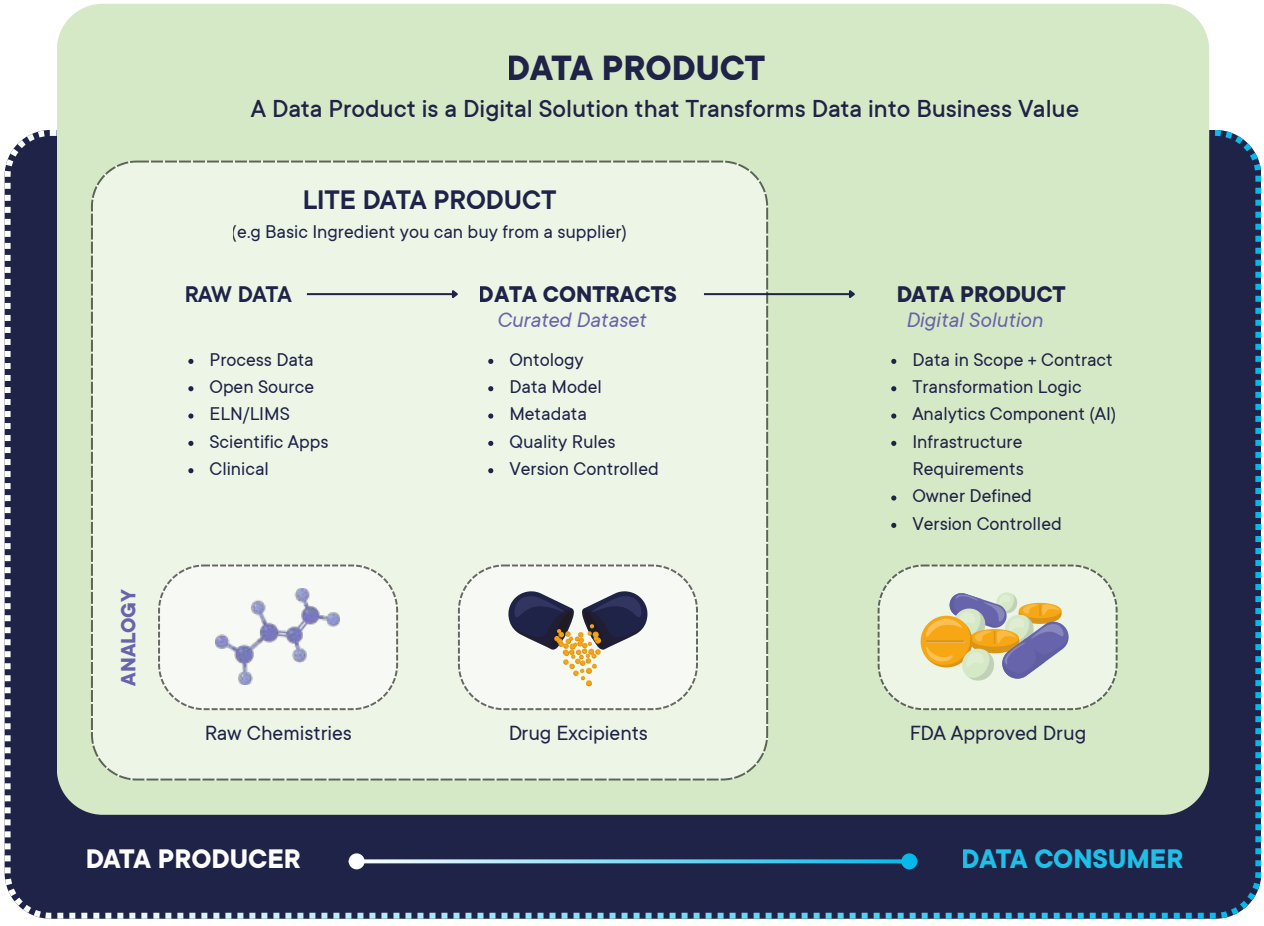


Figure 1: *Lite and Full Data Products – what's required to unlock value*

Conclusion

The modern conversation about data has moved beyond technology to one of value. This whitepaper has clarified that a data product can take two valid forms: Lite Data Products, which provide high-quality, reusable datasets, and Full Data Products, which integrate data, analytics, and user experience to deliver measurable business or scientific outcomes.

Crucially, both rely on the mindset of Data-as-a-Product, which introduces product management discipline into data work. This approach ensures clear ownership, continuous improvement, and alignment between data initiatives and business goals. It shifts the emphasis from delivering datasets to managing evolving products that generate lasting value for the business.

Lite Data Products enable trust and reuse. Full Data Products enable insight and impact. Data-as-a-Product ensures both are managed intentionally, with accountability and purpose.

Authored by

Bruno Fievet
Team Lead,
Data and Architecture

Melchor Sanchez Martinez
Head of Data Product Analytics

David Jackson
Head of the Scientific
Data Foundation Business Unit

About Zifo

Zifo is the leading global enabler of AI and data driven enterprise informatics for science driven organizations. With extensive solutions and services expertise spanning research, development, manufacturing, and clinical domains, we serve a diverse range of industries, including Pharma, Biotech, Chemicals, Food and Beverage, Oil & Gas, and FMCG. Trusted by over 190 science-focused organizations worldwide, Zifo is the partner of choice for advancing digital scientific innovation.

info@zifornd.com **www.zifornd.com**