IDC PlanScape

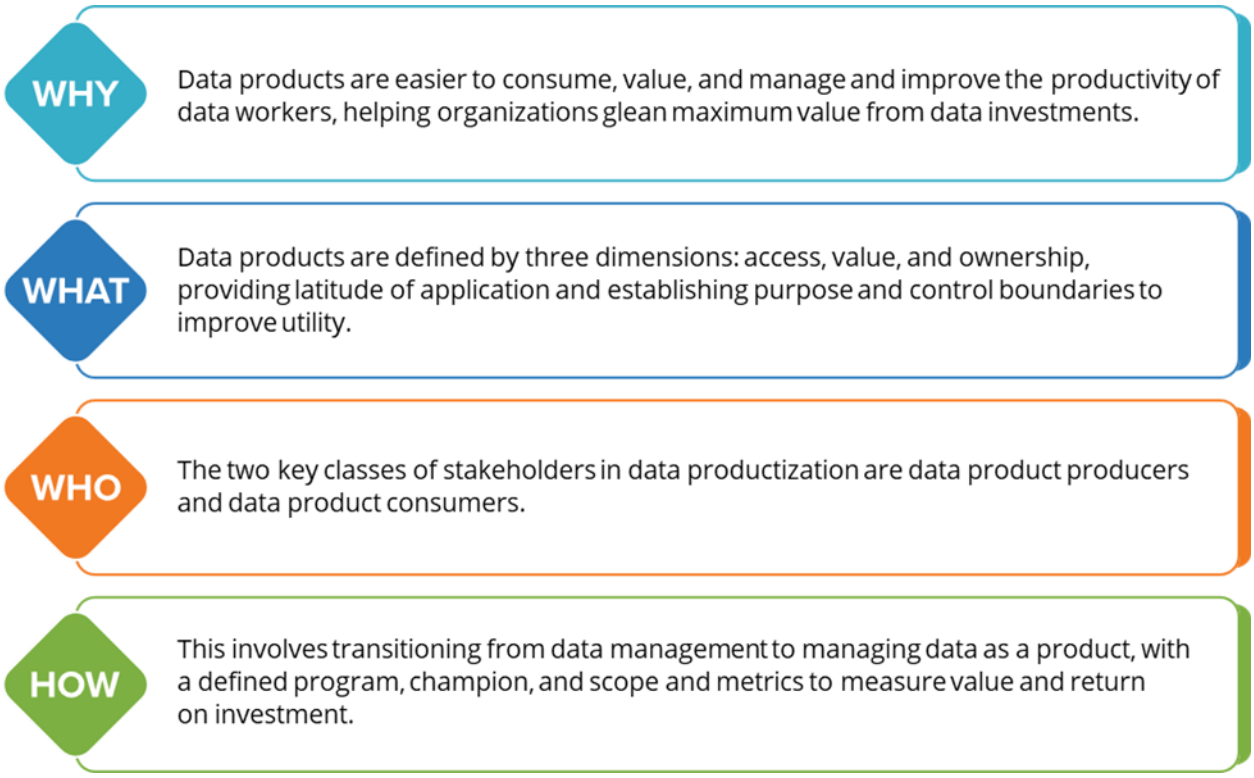
IDC PlanScape: Data as a Product

Stewart Bond Chandana Gopal Lynne Schneider

IDC PLANSCAPE FIGURE

FIGURE 1

IDC PlanScape: Executive Summary of Data as a Product



Source: IDC, 2024

May 2024, IDC #US51712224

EXECUTIVE SUMMARY

Treating data as a product addresses the escalating complexity of data landscapes and the critical need for organizations to harness data more effectively to drive superior business outcomes. With data's scale, diversity, and dynamics expanding rapidly, chief data officers (CDOs) are challenged with maximizing the utility of data assets and enhancing the productivity of data workers.

Treating data as a product simplifies its consumption, management, and valuation, thereby augmenting the productivity of data workers and ensuring organizations derive maximum benefits from their data investments. Several advantages of data products include increased data accessibility, quicker time to value, support for digital business growth, innovation multiplication, new revenue streams, improved governance, and risk reduction. Furthermore, data products play a pivotal role in supporting artificial intelligence (AI) initiatives by providing high-quality, compliant, and relevant data for AI model training, thereby enhancing the accuracy and relevancy of AI outcomes.

The concept of managing data as a product is not without its challenges. The lack of a standard definition and immature management practices are significant hurdles. IDC advocates for a flexible definition of data products across three dimensions — access, value, and ownership — to accommodate various use cases and business domains. Effective data products connect data producers to consumers and are described as sets of data and operational or analytical assets that are accessible by multiple personas within specific business domains, providing clear ownership and accountability.

Recommendations

Transforming from data management processes to managing data as a product requires executive buy-in and support; a clear charter, scope, and plan; and dedicated resources, along with metrics for measuring success. It is essential to include change management within the office of the CDO and IT, as well as with data consumers in the line of business (LOB). Furthermore:

- Start the initiative with a project that is manageable in scope to limit the number of stakeholders and minimize the impact of changes, ensuring the first data product delivers tangible business value to demonstrate the program's value and secure further funding.
- Embrace collaboration among data, business, and IT teams from the outset of the pilot program, and as the organization matures in its use of data products, continue to systematically measure and communicate the value that these data products generate to all the teams involved.
- Implement new processes and procedures for handling data product requests and fulfillment in parallel with the required technology to enable data sharing or marketplace-like functionality for internal data consumers, ensuring the data product approach is effectively integrated into the organization's operations.

Synopsis

This IDC PlanScape on data as a product emphasizes the growing complexity of data landscapes and the necessity for organizations to effectively utilize data assets for improved business outcomes. It advocates for treating data as a product to enhance accessibility, management, and value extraction, highlighting benefits such as increased innovation, new revenue streams, and support for AI initiatives.

The document also outlines the roles of key stakeholders in data productization and offers guidance on implementing a data-as-a-product approach to drive organizational success.

"Data as a product transforms complexity into opportunity, driving innovation, governance, and new revenue streams in the digital economy," says Stewart Bond, vice president of Data Intelligence and Integration Software research at IDC. "The concept of data products is gaining traction among organizations looking to better understand the value and utility of data and enabling connections between data producers and consumers within an organization, and we have seen the market respond with data sharing platforms, also known as internal data marketplaces."

WHY IS DATA AS A PRODUCT IMPORTANT?

Increasing complexity of data landscapes, where the scale of data distribution, diversity, and dynamics continue to skyrocket, is making it imperative for organizations to find ways to make more effective use of data assets and skills of workers required to leverage data and drive better business outcomes. Chief data officers are often hired with the Herculean task of "fixing" data problems, but without understanding the business value of data, they struggle with identifying data ownership and managing appropriate access to improve the quality of data and the productivity of data workers to justify data investments. The tenure of a CDO at 2.5 years is far shorter than other CXOs because they are not able to effect transformation or are frustrated by a lack of appreciation of the value of data investments.

When data is treated as a product, it becomes easier to consume, value, and manage and improves the productivity of data workers, helping organizations glean maximum value from their investments in data. Without data products, data producers and consumers will have less focus and clarity about where data quality issues exist and the connection between improving quality and business outcomes. Data consumers will also spend more time searching for data and working with data producers to transform it, build pipelines to make it accessible, and run bespoke analyses, often duplicating efforts with other teams. Data products provide organizations with the following benefits:

- **Increased consumption through better access:** According to IDC's December 2022 *Data Valuation Survey,* almost 44% of the 1,024 respondents cited that they don't have access to data that they need to do their jobs. Data products connect data producers to data consumers and make data more accessible and discoverable, promoting use and reuse.
- **Time to value:** According to the same *Data Valuation Survey*, 51% of respondents cited that data loses its value in a few hours (or less). Unfortunately, data is often stale by the time consumers find, prepare, and analyze it to support the decision or business outcome being sought. Data products can ensure that data is produced and packaged in a way that data consumers have access to the most up-to-date insights needed for decision-making.
- **Supporting digital business:** According to IDC's 2024 *CEO Survey*, roughly one-third of organizational revenue comes from digital products today, which is expected to grow to half of the revenue within five years. Data is the lifeblood of digital business, and data products will enable better delivery of data to digital products and services to improve business outcomes.
- **Multiplied innovation:** Data products are catalogable, discoverable, and reusable. As organizations mature in the creation and use of data products, benefits are multiplied at scale.
- **New revenue streams:** Internal data products may have an opportunity to be monetized externally, generating new revenue streams for the organization and increasing the value of data.

- **Improved governance and reduced risk:** Data products have defined attributes, data sources, ownership, contracts, and a life cycle. This formalized structure allows data products to be curated, governed, and auditable, reducing risk and noncompliance issues.
- **Supporting artificial intelligence initiatives:** New data value chains are required in disciplines that are emerging in support of enterprise predictive, interpretive, and generative AI solutions. Training and tuning of AI models require safe, compliant, relevant, and quality reference data, both structured and unstructured. The disciplines associated with managing data as a product can improve the overall quality, relevancy, compliance, and safety of data managed by the organization, improving the effectiveness of AI model training overall. Using data products at the time of inference can ensure appropriate context, high quality, timeliness, and proper classification to improve the relevancy and accuracy of AI outcomes while lowering compliance, regulatory, and sensitivity risks.

WHAT IS DATA AS A PRODUCT?

Treating data as a product is not a foreign concept for companies that monetize and sell data in business-to-business data marketplaces. As a seller of data, these companies have product development and management capabilities focused on the data products that are marketed, sold, and fulfilled for customers. The value of the data products can be understood in part by the amount of money that these vendors are asking for in exchange.

The concept of data products is gaining traction among organizations looking to better understand the value and utility of data and enabling connections between data producers and consumers within an organization. One of the challenges these organizations face is that there is no standard definition of what a data product is, with definitions ranging from "it's a table in a relational database, or an object in an object store" to "it's an analytical dashboard and all of the assets in the pipeline that deliver the data and analytics in the dashboard." Another challenge organizations face is that the practices associated with managing data products are immature.

One standard definition of data products would be too constraining for all potential use cases of data across all business domains, but rather, data products are defined along three dimensions: access, value, and ownership. Access is making data products available, discoverable, and reusable. Value involves determining the business value derived from the data products that ultimately drives the data valuation methodology. Ownership determines who is responsible for developing, maintaining, monetizing, and nurturing the data product throughout its life cycle.

Defining data products along these dimensions provides latitude of application and establishes purpose and boundaries of control. This definition allows the data product to be as fine grained as an individual table or object, but this would be impractical as access, value, and ownership would not be scalable. The definition also allows a data product to be large grained such as a dashboard and all assets that go into making that dashboard, but the utility of such a data product is likely minimal and value would be too constrained. Rather than these two extremes, effective data products tend to be a set of data and operational or analytical assets that can be used across multiple solutions, accessible by multiple personas, typically within specific business domains to provide clarity of ownership and accountability, to deliver business value.

Examples of data products include:

- Propensity to churn customer model

- Recommendation engine for related products on an ecommerce site
- Dashboard that offers a consolidated view of an account

As the concept of data products becomes more mature, organizations will develop clear methodologies for data valuation and possibly drive the inclusion of data as an asset on their financial statements. Enterprises and executive leaders are demanding better visibility into the return on investment on digital initiatives, and data is a critical component of a digital strategy. However, although being data driven is a must for success in this digital economy, many organizations struggle with attributing an actual monetary value to their data. There have been books written about assigning value to data, and while there are some products and services available in the market to assist with doing so, there is no standard method to valuing data, and therefore data is most often treated as an intangible asset.

In 2Q23 earnings reports, some of the largest technology companies including Meta, Alphabet, Amazon, Microsoft, and Salesforce all mentioned that data was essential to their success and an asset, but unlike intangible assets such as goodwill and intellectual property, data is not formally listed on financial reports. The chief data officer of Northwest Bank, Michael Capata, in the 17th annual Chief Data Officers Symposium stated that one of his first exercises as CDO was to understand the real value of data to his organization and estimated that it was close to 3% of the asset base of the bank, leveraging a data economics assessment from a consulting firm. Attributing an economic value to data is challenging, and with the lack of accounting standards and the requirement to measure data value, only a few companies with leading enterprise intelligence will formally do so.

WHO ARE THE KEY STAKEHOLDERS?

The two key classes of stakeholders in data productization are data product producers and data product consumers. These two classes are made up of multiple organizational roles, each contributing to the processes and activities involved in data product management and consumption. Data producers can also be data consumers, depending on the context in which the data product is being used.

Data Product Producers

A data product producer is not one person, but typically a team of individuals that assume data product ownership, design, development, and management responsibilities. A data product is designed across the three dimensions that define it — ownership, value, and access:

- **Who owns the data product and is accountable for it?** Data product ownership ideally lies within the business domains of the organization. Line-of-business leaders understand how data is used within operations and what insights are required to support decisions that will impact outcomes. Business analysts, data stewards, application owners, and team leads are examples of roles that may take on or share data product ownership responsibilities. Data product owners are accountable for identifying the need for a data product, working with data and analytical resources to define, design, and deliver the data product. If monetization through internal chargebacks is part of data productization, the owner should be setting the price for the data product and defining the service levels that will be delivered with the product. Data product owners are also responsible for managing data products through their life cycle, including quality, changes, versions, branching, and sunsetting.
- **What business value does the data product deliver, and in which domain, over its lifetime?** Data architects, engineers, analysts, and data scientists work with the data product owners in

the business to design and deliver the data product to meet business requirements across multiple use cases to maximize utility of the data product and within the promised service levels. These data roles will be involved in change management and version control at the technical level and publishing the data product into an internal data marketplace or exchange where consumers can search, discover, and request access to the product.

- **Who, what, and when will the data product need to be accessed?** This dimension is primarily associated with data consumers, but data producers maintain control of access to, and use of, the data product. More specifically, data product owners have the final say in who and what can access a data product and authorize how the data product can be used. Data exchange and marketplace software typically includes workflow capabilities for routing data product access requests to owners for approval.

Data Product Consumers

As with data product producers, consumers are also not one person. The consumers are generally in the line of business, except in situations where the intent is to assemble a new data product from existing data products. In these cases, many of the same data product producer roles are also data product consumers. Furthermore:

- Data owners may also be consumers of the product to achieve the business purpose and value for which the product was created.
- Data architects, engineers, analysts, and data scientists may use already built data products to create new derivative or higher-grained data products.
- Data product consumers search for, discover, and request access to data products for use to deliver the business value for which the products were created.

Governance and Shared Services

Managing data as a product is a distributed concept, as data product ownership, delivery, and use happen within business domains inside of lines of business. This provides more autonomy for data consumption, but there is still a need for centralized governance to maintain control, consistency, and regulatory compliance of data product implementations. Whether federated or centralized, the following roles need to be in place for data productization:

- Chief data officers, to set principles and guidelines for data product definition and design to promote consistency and ensure regulatory compliance
- Data architects, to create reusable frameworks and standards for implementation that promote interoperability and reusability of data products
- Data administrators, to manage databases, repositories, backups, and high availability for supporting data product service levels set by data product owners
- Security administrators, to manage credentials across networking, databases, repositories, marketplaces, and applications
- System administrators, to manage service and application delivery for data product producers and consumers and all related support staff
- Billing and collection administrators, to manage chargebacks if it is part of the data productization design within the organization

HOW CAN MY ORGANIZATION TAKE ADVANTAGE OF DATA AS A PRODUCT?

Managing data as a product is different from data management. Data management is focused on maintaining the consistency and integrity of data held within transactional, analytical, and object repositories, without any one specific business purpose other than perhaps supporting business operations. Typically, data engineers and data stewards address requirements and requests from the business in a one-off manner. Managing data as a product requires a focus on the business value delivered by the product, which may itself be made up of several data and analytical assets. Data management tends to be centralized, whereas data products are distributed throughout the organization based on the business domain in which value is being delivered. Managing data as a product requires product management.

Managing data as a product will require three primary components to be successful:

- Product management
- Data marketplace technology
- Fulfillment

All of these together need to be implemented and governed by the shared services in the organization.

Three Primary Components

Product Management

Managing data as a product is different from data management, as it requires a focus on business value, design, and delivery, within specific data or business domains. Data management is broader and more general in its approach to maintaining consistency and integrity of data across the enterprise. Managing data as a product requires three primary components to be successful: product management, data marketplace or sharing technology, and fulfillment capabilities.

Generating data product ideas proactively, based on current and prior requests, enterprise data landscape, and capabilities, in addition to understanding the data consumer, is part of data product design and development. Product management also requires developing a plan, which includes prioritization of product design, development, and implementation, to meet business and technical stakeholder priorities. Product life-cycle management requires tuning and tweaking over time, monitoring usage, providing levels of backward compatibility as new versions of data products are created, providing upgrade paths when required, and sunsetting of legacy products at end of life.

Data Sharing/Marketplace Technology

Managing data as a product may also require marketing and "sales" efforts, even if the products are only used internally. A new supply of technology in the software market is emerging in data sharing platforms, also known as data marketplaces through which data can be "sold" and "bought" internally. Traditionally, data marketplaces have been focused on business-to-business buying, selling, or trading data, which is why *data sharing* is being used as a term for internal data marketplaces. The value of data should be well known in external marketplaces if money is being exchanged. Increasingly, data sharing/marketplaces that focus on connecting internal data producers to internal data consumers are becoming more prevalent within data intelligence and management software platforms or suites, and some include the infrastructure required to enable internal chargebacks.

An internal data product sharing platform or marketplace is good not just as a focal point for releasing and provisioning data but one that includes process support as the data team works to create, launch, and manage the data products. Today, many organizations try to use data catalogs or data governance tools to do these functions. While those tools can be helpful in definition or managing access, they do not provide the support necessary for data product management. A data product sharing solution or marketplace needs to support each activity in the data product management life cycle and add value into data product creation, monitoring for increased utility and efficient distribution processes in the enterprise.

Fulfillment

How data products are defined is in part a product of data marketplace software. There are software vendors that offer the flexibility to make fine-grained artefacts products in the marketplace, whereas others provide the ability to define a data product as a collection of assets — data and/or analytical. If a data product is a prepackaged, curated set of data and/or analytical assets available in the marketplace, placing the product in the shopping cart and then checking out should make fulfillment as simple as providing access and authorization approval for the buyer. If the products are fine grained, fulfillment may also require additional delivery mechanisms such as extract-transform-load (ETL or ELT) processes, and this should be avoided because it is more akin to data management than managing data as a product.

Implementation and Governance

Implementation needs to be a collaborative effort across the business, IT, and data resources within the organization. Data products cannot be a field of dreams — building the product thinking the people will come is not a recipe for success. Data products need to serve a purpose and deliver real business value. Requests for data can come from many functions, many divisions, and data consumers of varying types. Even if there is a central data area of responsibility, this can be quite overwhelming. These requests are often similar, but tend not to be the same. As a result, data producers may treat each request as a unique project to be fulfilled, resulting in backlogs and repetitive work. As with any solution design, the need to capture functional and nonfunctional requirements is critical to success, and data product development should follow well-known iterative and agile methods, including DevOps and, more importantly, DataOps. Careful consideration also needs to be made as to where and how the data product could be used outside of the request for which it is being built to improve reuse and utility of the data product. Orchestration technology can provide needed capabilities here in ensuring integrity in continuous integration and continuous deployment (CI/CD) pipelines.

When fulfilling data product requests becomes a series of "projects," the data consumers often feel that there are gatekeepers between them and the data. On the other hand, self-service with data sharing platforms or marketplaces and data catalogs alone makes data owners nervous about how data will be used and whether the data consumers are choosing to work with the most applicable data. The execution of data sharing tends to fall short of its promise without proper governance and control mechanisms in place.

Data products need to be controlled with policy and data governance activities, including providing all necessary intelligence about the data products in metadata to help ensure they are being used in the most appropriate context and for the right reasons. Control is being exercised in the form of data product contracts that specify when, where, how, and who can have access to the product. Contracts also provide expected service levels for availability, quality, and reliability of the data product. Data intelligence available in the contracts needs to be inclusive of quality levels, lineage, business context,

security and privacy classifications, and ownership. This will be critical as data products become part of generative AI solutions, perhaps being embedded into prompts at the time of inference to improve the accuracy and relevancy of what gets returned by the model. Data governance steering committees will need to include representation from data product owners, data product stewards, IT, line-of-business representatives, and data leaders.

ADVICE FOR TECHNOLOGY BUYERS

Transforming from data management processes to managing data as a product cannot be approached as a trivial exercise. It will require executive buy-in and support, a charter, a scope, a plan, a set of metrics that will be used to measure success, and dedicated resources. Taking a data product approach will require product owners and management teams and evangelists of data product success. This program will need to include change management within the office of the CDO and IT as the data producers and with data consumers in the LOB. New processes and procedures for product requests and fulfillment need to be put in place, in parallel with new technology to enable data sharing or marketplace-like functionality for internal data consumers.

Scoping out the first initiative is often the most difficult task. Start with something that is small enough to limit the number of stakeholders and the impact of changes on both the producer and consumer side. But the first data product that is created needs to also deliver tangible business value to demonstrate the value of the program and secure expansion funding. Metrics tied to the data product need to demonstrate not only usage and utility but also return on investment. Such metrics may include productivity gains because the time required to search for and prepare data is eliminated. Another metric may be related to the time required to make decisions and improve the impact of decisions, which may in turn improve the top or bottom line of the business.

Every pilot program needs to embrace collaboration among data, business, and IT teams and, as your organization matures in its use of data products, continue to systematically measure the value these data products generate and communicate that value to your teams and to the business stakeholders.

RELATED RESEARCH

- *IDC PlanScape: Generative AI Data Value Chain* (IDC #US51946724, March 2024)
- *IDC PlanScape: Enabling Data Governance for Generative AI* (IDC #US51958724, March 2024)
- *IDC PlanScape: Data Quality Management* (IDC #US51397423, December 2023)
- *Artificial Intelligence in Data Intelligence and Integration Software for the Future of Enterprise Intelligence* (IDC #US50923223, September 2023)
- *Four Planes of Enterprise Intelligence Architecture: A Conceptual View into the Data Plane, Data Control Plane, Data Analysis Plane, and Decisioning Plane* (IDC #US50793023, June 2023)

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology, IT benchmarking and sourcing, and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives. Founded in 1964, IDC is a wholly owned subsidiary of International Data Group (IDG, Inc.).

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com