NTT DATA

# Data marketplace: unlocking data value

# Content



# Section 1: From Big Data to data democratization

## Data as a source of added value

There was a time a decade ago when business reports were filled with provocative statements about data, such as: Data is the New Oil! Headlines designed to focus attention on one undeniable fact, which is that most organizations have historically underestimated the importance of the data they already possess.

As enterprises became more aware of the value that existing data could offer, they started to create data warehouses and, later still, data lakes, which could be continuously "mined" for insights of value, which had previously been overlooked or were simply not available to those best placed to use such data to create value. This was often known as the Big Data approach.

In the Age of generative AI, some of these concepts and techniques seem old fashioned, even quaint. There was a time when copying data from inside applications and repositories to a data lake for easier searching seemed truly ground-breaking. Today, it seems no more than an early first step on a long journey to optimized data usage.

**Data markets represent the essential evolution of data utilization, and marks an important step towards data democratization. So what do we mean by democratization?**

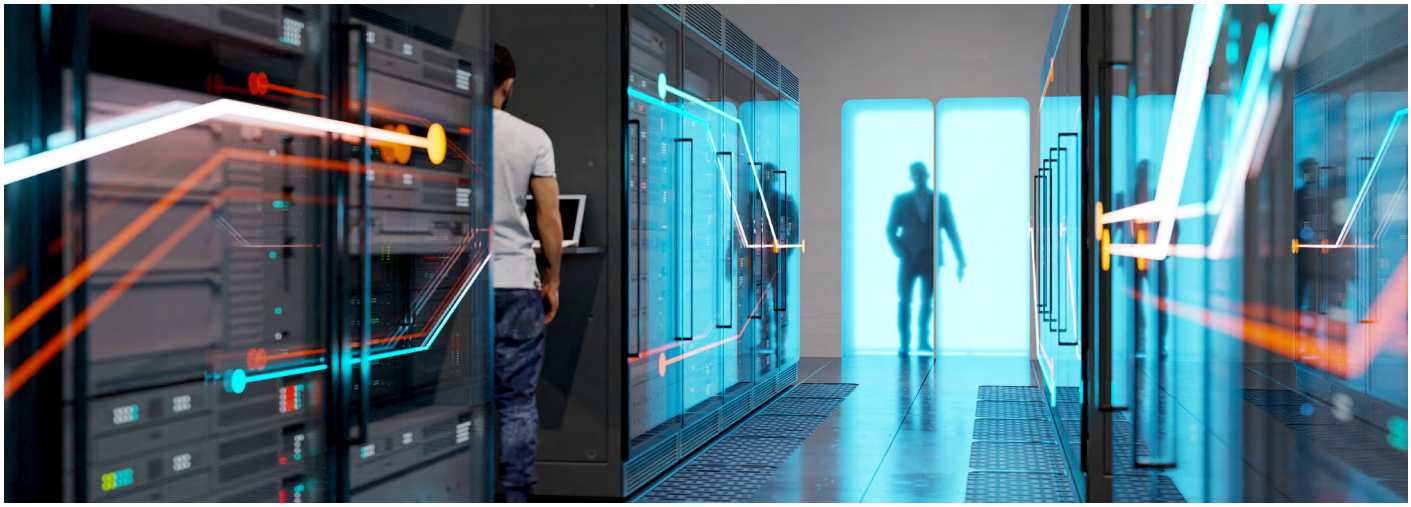
## Why does democratization matter?

Data starts to deliver its full value when it becomes freely searchable, using simple and intuitive methods, by every team and individual within an organization and (sometimes) its authorized ecosystem. This has never been the case in the past, when data (produced by every transaction within and between every application and process) largely remained enclosed within its originating systems and their own repositories.

The creation of data lakes was designed to address this issue by separating data from applications and establishing a single repository that both **feeds into** and is **fed by** each transaction, theoretically making the totality of data within an organization more widely available to a much larger pool of users.

In practice, these existing methods still do not amount to full democratization, which makes it possible for every person in an organization (and potentially its ecosystem) to access and make full use of all data produced, owned or acquired by the organization. **This enables valuable new insights and also informs and enhances future transactions.**

The moment we start to define what a democratic data environment might be like, we can see how far away from this most of our enterprises currently are.



We envisage **development of a combined business, cultural and technological construct to enable true data democratization**, and this is becoming known as **the Data Marketplace**. To move from data warehouses, repositories and data lakes, with specialized search tools, to a Data Marketplace, we need to establish:data warehouses, repositories and data lakes, with specialized search tools, to a data marketplace, we need to establish:

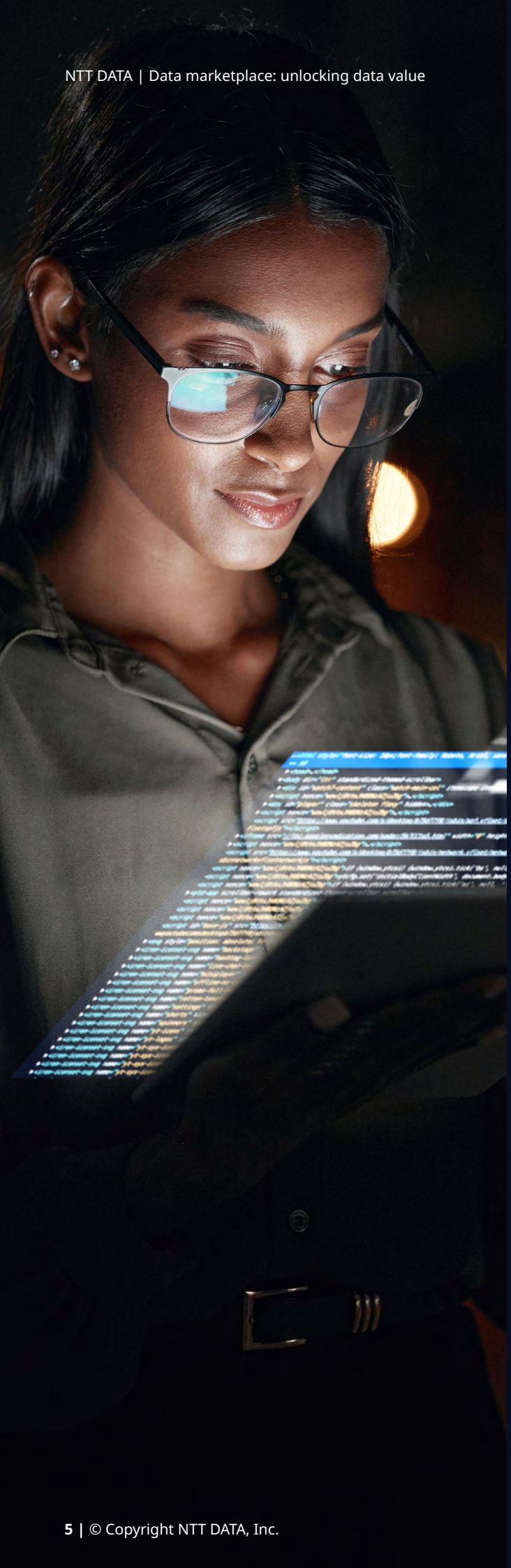
**Data culture**, in which procedures, rules and methods for use of data are consistently understood and applied across the entire user base. There should be no exceptions, irrespective of status or position in the business: every user needs to have the same cultural approach.

**Governance regime**, ensuring that all activities related to data management and usage is given effective oversight. That means putting a clear strategy in place, ensuring that all players have clear roles, rights and duties, with regular reporting and full management control. Effective governance also ensures that regulations are understood and complied with at all times, including data sovereignty, respect for IP, copyright and privacy.

**Intuitive interfaces**, including metadata that has enough clarity and depth to help users explore with the greatest available speed and simplicity.

**Centralization of data access**, putting technologies and procedures in place that ensure all data related to the corporation and its ecosystem can be made available to all users (subject to role and access privilege). This will ensure that all search is comprehensive and cannot lead to gaps in knowledge due to the need for updating and work-arounds. This enables access to users, irrespective of the underlying logical structures, which we will briefly explain later in this document.

Increasingly, we also see the need for inclusion of **AI and Machine Learning algorithms**, both because these are becoming an integral part of all search mechanisms as a matter of course, and because, when properly applied, they will **deliver essential operational efficiency gains.**



## The difference between internal and external data marketplaces

In simple terms, an **Internal Data Marketplace** exists to enable fast, **easy access** to all data **within an organisation** by its authorized users. There is no access for external players except in some cases by partners- who may be designated as “internal users” for this purpose.

An **External Data Marketplace** is set up to **enable monetization of corporate data** (which in many cases means data originating from customer transactions), and is **based on the knowledge that real-world data is a resource of great value and importance.**

For the purpose of this paper, we are going to leave External Data Marketplaces largely outside the scope of discussion. We note, however, this is an area of great interest to regulatory authorities, with recent court cases (most- but certainly not all originating in the European Union) asking serious questions about current and emerging future practices.

These are as diverse as gen AI training (potential IP, copyright and privacy concerns) and data sovereignty (under what circumstances can data be allowed to reside on servers in different jurisdictions?).

An **Internal Data Marketplace** makes it easier for **enterprises to remain fully compliant** but still requires major technical issues to be managed. This can be illustrated by one example of **data value mining** that is of particular interest to NTT DATA, as it covers the introduction and development of a Data Marketplace inside our own business.

## Proving the data marketplace concept in NTT DATA

As a business that is highly dependent on collection, analysis, extensive use and reuse of data, it is perhaps not surprising that NTT DATA took the decision in late 2020/early 2021 to develop and rollout, step by step, a fully functional Data Marketplace across or own business. The goals, as expressed by the company's Chief Data Officer, included:

- Providing a central data repository of all the company's data products and assets.
- Improving levels of democratized access to data across the business.
- Enhancing the quality of data in use, while also improving access to always current, up to date information.
- Cutting risks due to non-compliance or use of inappropriate data.
- Improving data literacy and competence, while also opening up business innovations as a result of better synergies.

This was always seen as a major project, which is why introduction has been carefully staged. By mid 2022, the Data Marketplace included access to all corporate reports, data models and a range of curated external assets, sourced form partners and other third parties. Introduction of the project has been intensively supported through online tools (glossaries, FAQs, user manuals) and hands-on training, backed by fast access to specialist guidance and support.

By mid 2023, the Data Marketplace already had thousands of daily users, while measurable benefits range from better governance and security (improved access is matched by lower risk and high levels of oversight), improved user experience (leading to higher levels of satisfaction and better output). Most important, perhaps, operational efficiency has improved, as personnel spend less time looking for data and more time working with it.

As the **NTT DATA experience proves, true data marketplace is based on simplicity, universal access, intuitive interfaces and democratization.** Our proof of concept also shows that the idea works in practice. Standard components are normally used, but the way we build and manage the platform are different. Let's take a closer look at that.

# Section 2: Creating a data marketplace

## Definitions and long-term objectives

### Data fabric & data mesh

For context, and because it will help us understand the value and positioning of a Data Marketplace, let's introduce the concepts of **Data Fabric and Data Mesh.** These architectural frameworks, or paradigms, have become a centrepiece of data architecture designs because they help us to escape from the monolithic architectures and centralized data platforms that have made extracting value from data so cumbersome for the past decades.

Conceptually, a **Data Fabric** is a **metadata-driven way of connecting a disparate collection of data tools or repositories, in a cohesive and self-service manner.** More specifically, data fabrics deliver capabilities in the areas of data access, discovery, transformation, integration, security, governance, lineage and orchestration. In this way, a data fabric addresses the growing complexity of data management, by intelligently integrating and connecting enterprise data, making reusable data assets more easily available for consumption. It is important to note that only data management is unified, while the actual data remains distributed.

While the data fabric aims to build a single virtual management layer on top of distributed data, a Data Mesh advocates **for distributed, domain-based ownership and custodianship of data, focusing on data products that are self-described and atomic.** These data products are shareable with other domains and can be combined with other data products that form the **data mesh** to create new data products.

In this framework, data is managed as a distributed network of self-describing data products. This decentralized approach is balanced, from a data management point of view, by the principle of Federated Data Governance, in which a set of centrally defined data governance standards is applied, giving data domains the freedom to apply these standards autonomously in the most appropriate manner for their particular environment.

In our view, these two paradigms are not mutually exclusive: **they actually complement each other**, and organisations should adopt those principles from each that fit best with their particular situation. A key distinction between the two is that data fabric describes a **more technology-centric approach**, while adopting a data mesh framework usually involves a deeper **cultural and organizational change.**

Among other similarities, both approaches involve a distributed approach to data, together with a certain level of centralization in its management, which also includes how it is consumed. Each needs a layer where data can be searched for and found, then accessed on demand, while complying with all data governance policies.

**This is the purpose of the Data Marketplace, the single point of access to enterprise data products or assets, leveraging the capabilities of these architectural frameworks.**

## Development and operations

We have already looked at some key characteristics of Data Marketplaces. Let's now consider how to develop them, together with the operational characteristics they must deliver to their users. Simple definitions first. An Internal Data Marketplace must become:

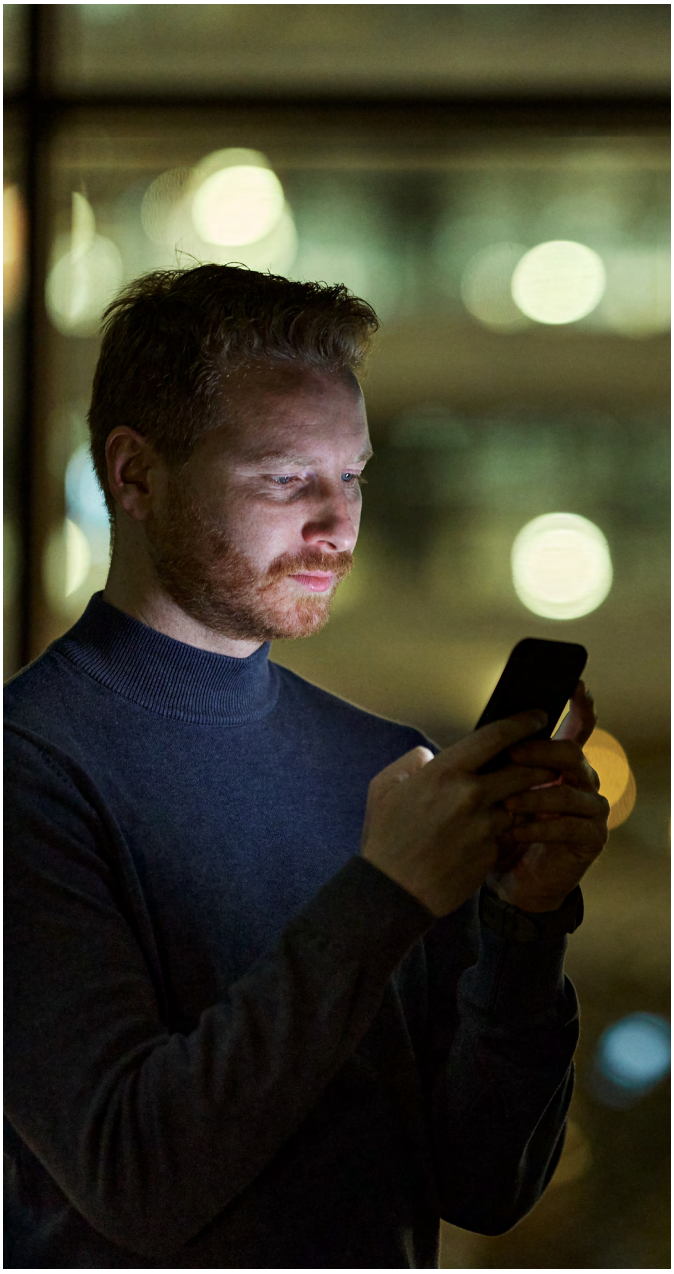
- **A single point of access to all relevant data**, which means either a central repository or a mesh-based solution that enables comprehensive access. In this case, all relevant data means everything produced and owned by the organisation concerned, including real-time transaction-based data and all information used through subscriptions and other methods.
- It must also **provide actionable data on demand**. This means enabling users to carry out simple, intuitive searches, based on requirements linked to their own work, with a high probability that such searches will produce usable insights, not just raw data. The goal is to transform information into a source of business action and ultimate value. This is where introduction of AI and ML based tools can potentially add significant value.
- **Allow and ensure all aspects of enterprise-ready data management**. That means being clear about who owns data and the work owners need to do in keeping their data clean, up to date and fully available. It also covers usability and accessibility, ensuring that data can be found and used as needed, where and when business users require it. We also need to make sure that data can be made ready for reporting at high speed: internally (enabling better and more timely decisions) and externally (demonstrating compliance through accurate, detailed and timely reports). Finally, data management includes security, managing access criteria and privileges, while safeguarding against accidental loss, internal bad actors and external attacks.

In other words, we need tools that not only mines data for value but structures it into the most useful and usable formats in order to meet enterprise goals. This requires,not just a repository, but effective tools to manage the data.

The net result of this approach is to set up a Data Marketplace that is based on a scalable, well-governed platform that includes all enterprise data in a curated, catalogued, searchable and accessible form. This approach will encourage and enable self-service access and usage of both enterprise and authorized external data, leading to much more effective and productive collaboration between data analysts, business users and decision makers.

The end goal of a Data Marketplace is therefore to **ensure that all business users possess a better understanding of the data assets available in the organization, ensuring that the most relevant data finds its way to those who need it, quickly easily and intuitively**. Business need is always the key driver here, while this approach reduces dependence on IT departments, reducing the burden on IT professionals by cutting the number of requests for help in locating or interpreting information.

Data Marketplaces, therefore, can be seen as **Business Improvement Tools**, with the aim of delivering competitive advantage through efficiency, speed and insights.



## Benefits and opportunities

In practical terms, building a Data Marketplace will bring the benefits of Data Democratization to an enterprise at speed and with a minimum of problems. We have already touched on the importance of democratization, but let's just analyse the benefits from a user perspective.

The **goal of democratization is to place power in the hands of business users**. They now decide what data is most important to them right now, for their most important business tasks. They will then have the ability to select the data they want to work with and access it rapidly, on demand, avoiding bottlenecks.

This transforms the relationship between data analysts and data consumers (normally business users), making them far more productive than at present. Teams will need to spend less time than before in searching for useful data, requesting access, waiting to see if the data in question **really is** usable and important, and less time again in downloading, cleaning it and preparing it for use. Within a Data Marketplace, users should be able to search, locate, access and start using data by themselves, without expert help. The end result will be access **to better insights, faster than before, and that should feed into better decision making and competitive advantage.**

**Faster time to value is the key**, and that is what a Data Marketplace delivers. This also reduces or eliminates a great many extremely common challenges, which currently waste huge amounts of time and resource, cutting productivity inside most enterprises. These even apply within organizations that have well managed and mature data lakes. In these enterprises, data analysts are employed to create pipelines for certain kinds of analysis and transformation. Experience shows this often leads to a lot of time investment for little value created. For example:

- In some businesses, data scientists spend more time searching for information, asking for access permission and preparing information for use in their own models than they spend on creating the model for a business use case.
- The same report, containing highly relevant data, maybe be located in two places, and there may be discrepancies between them. How can you accurately decide which report is the most current, and also ensure that it matches the data sent to executive decision makers?
- Business users may have an urgent need for data that are currently stored in legacy systems. They may then be uncertain about how to gain access, either for reasons of permission or technical viability or both.
- Data may be visible in the data lake, and some of it may appear to be highly relevant. But it is not clear who created it, owns it or possesses (and can grant) access rights. It is also not clear, until deeper analysis exactly what insights are contained within the data, nor even what fields are relevant.

Many enterprises have invested heavily in data lakes and related technologies, only to find that the results are disappointing. These are precisely the kind of challenges Data Marketplaces are designed to overcome.

**So what the stages we need to follow in building them?**

## Building the capabilities

Before looking at the technology structure and reference architecture of the Data Marketplace, let's consider the **core capabilities** that must be designed into the solution from the start.

**Data Management and Governance.** This is fundamental to an effective internal Marketplace. The first step is to abstract the data from the technologies on which they are stored, and which normally have a technology focus, providing a strong business focus instead. This will involve adding metadata to provide clear guidance, direction and context for business users.

By doing this we start to build a Business Glossary that enables business users to orient themselves, find their way through the repository quickly and efficiently, making it easier for users to make use of the data. This approach also provides the mechanisms and tools needed to perform essential governance functions, including metadata management, security, data privacy, workflow permissions and management, together with other policies as directed by changing regulations and business priorities.

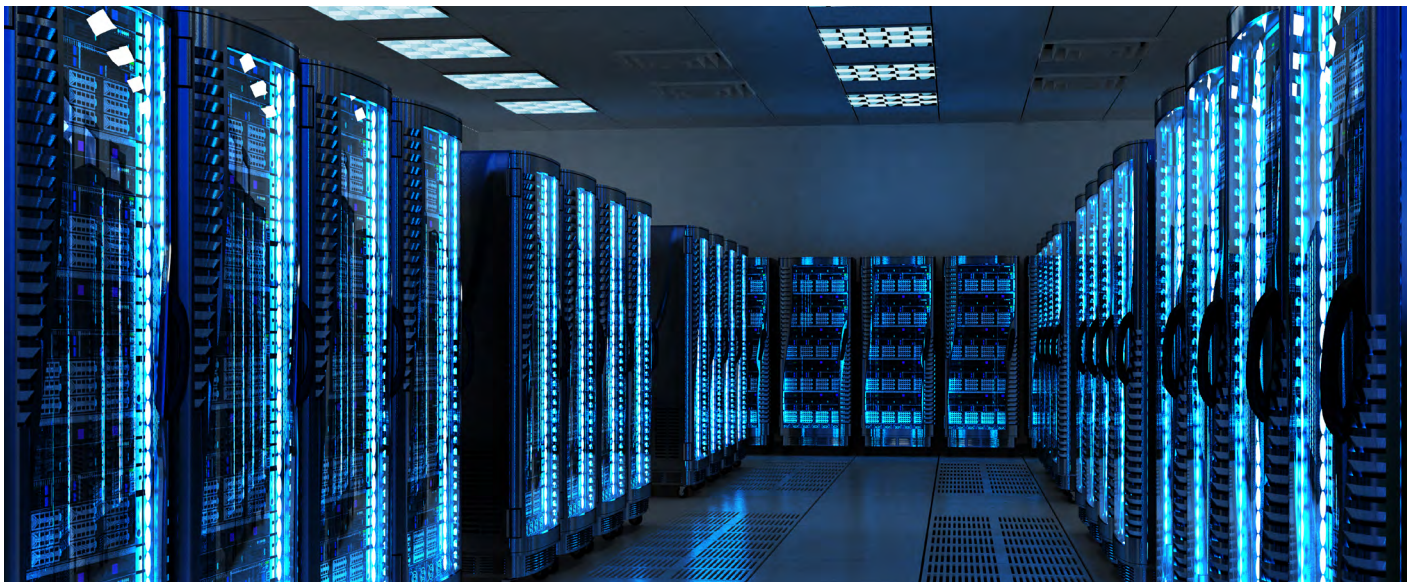
**Single, Simplified, Comprehensive View**. All data products must be made available at a single point of access, supported by the information needed for data consumers to identify the data product they require. Data products in this context means ready to consume, encapsulated data assets, ranging from dashboards to tables to any other form of data curated and published by its owners.

This means that, in the Data Marketplace, information suppliers have the requirement and duty to publish their own data product, while including a template that allows the data to be discoverable, understandable, trustworthy, addressable (natively accessible), secure, interoperable and valuable.

Wherever a single data product is comprised of information that comes from very diverse sources, data virtualization tools may be used to unify these sources more easily. Publication and subscription events can also be used for this purpose.

**Data Self-Service.** This is the essence of Democratization and defines the user experience by ensuring that users or consumers of information can search for information through different filters; obtain the information they require from a data product; preview the information; and then request access so they can view it in full. They must then be able to download it as needed or even have a direct connection with the source in order to consume the information as a service.



## Addressing the challenges

In every solution that aims to simplify and extend access to data, there will be a number of **key challenges** which must **be understood and fully managed**. These include:

**Privacy.** This is a legal imperative and highlights the need for effective, legally compliant data privacy controls. This involves offering mechanisms for data classification, correctly managing access based on different roles and privileges.

**Integration with data lifecycle management**. This is a basic organizational requirement and focuses on the need to integrate data consumption into the data lifecycle. This requires automation of data processes (data ops) and is covered later in this document.

**People management**. Whenever a new solution or platform is launched, we remind our clients that technical requirements form one part only of the change process. People and organization are at least as important- often more so. A Data Marketplace drives changes to the roles of some people in any organization, and this is not always easy. There is a natural resistance to change, and it is quite common for an enterprise to underestimate the requirement, and also its likely costs.

**Standardization and Interoperability**. To ensure consistency and ease of use, we need to standardize the data products published in the marketplace. This is achieved through a standard Data Product Sheet, a template that will be transformed into metadata in a consistent format, ensuring that all Data Products are displayed with the same, unified form to enable customized search filters.

**Volume, Velocity Quality, Veracity**. Business users will expect and demand near real time response times, no matter how large the volume of data to be processed. This means the platform itself must operate to the highest possible standards, while it is equally important to incorporate a quality index to validate and assure the reliability of the data being accessed and used. This applies both to fully internal data and to data purchased from an external source.

**Secure Data Storage and Transmission**. Security is perhaps the greatest concern of all in any enterprise, and in any industry sector. A Data Marketplace has to guarantee the security of all data in storage (data at rest) and of all data as it is accessed and downloaded (data in transit). Rules are required to define who is to have access to what, and the permissions needed to view and access such data. Identity and Access Management (IAM) solutions will need to be continuously updated and supported by cyber-security tools of the highest quality. This is especially the case when sensitive data is to be used. Leaks are unacceptable and the battle to keep data secure is never ending.

**Access, Workflow and Usage Controls**. The Marketplace must include a series of workflows, used by consumers to request information access. This is where first stage security management will be implemented, so that IAM and role-based regulations can be enforced at once. This has a major implication for the entire organization, which must have the infrastructure and tools in place to enable near instant access through pre-approvals.

Dashboards will also be required to display available products, which will be customized to individual users and their roles.

The following figure describes the Data Marketplace lifecycle, including actors, components and workflows involved.

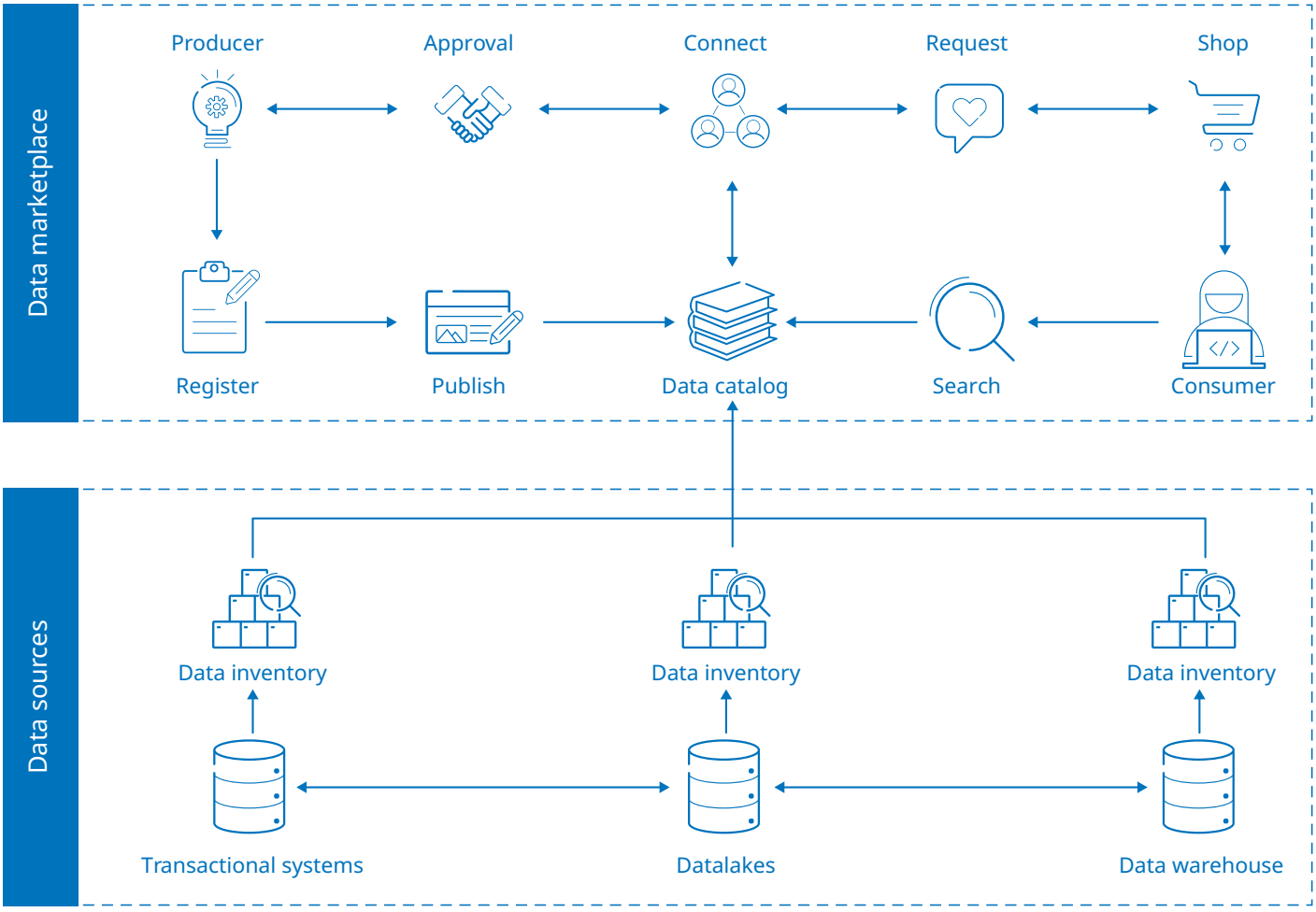



Figure 1. Data Marketplace lifecycle.

Finally for this section, though we have defined each of these items as “challenges”, in reality they are simply natural requirements for efficient operation. They are not “problems” as such but basic operational building blocks, complementing the technical structures, which we cover below.

## Technical set up

The Data Marketplace is **built on an architecture** that includes **Governance and quality management, with audit, metadata and catalogue** as the foundation. It incorporates data flow from ingestion to consumption, wrapped in security and access management.

Figure 2 below shows a very simplified, top-level logical view of every block or layer that fits into this architecture.

The specific components in each block vary according to the chosen architectural frameworks, use cases and vendors. This diagram shows the basic building blocks that will always be required to support the operation of a Data Marketplace. The structure of the Marketplace, itself, is explored in more detail in the Figure 3.

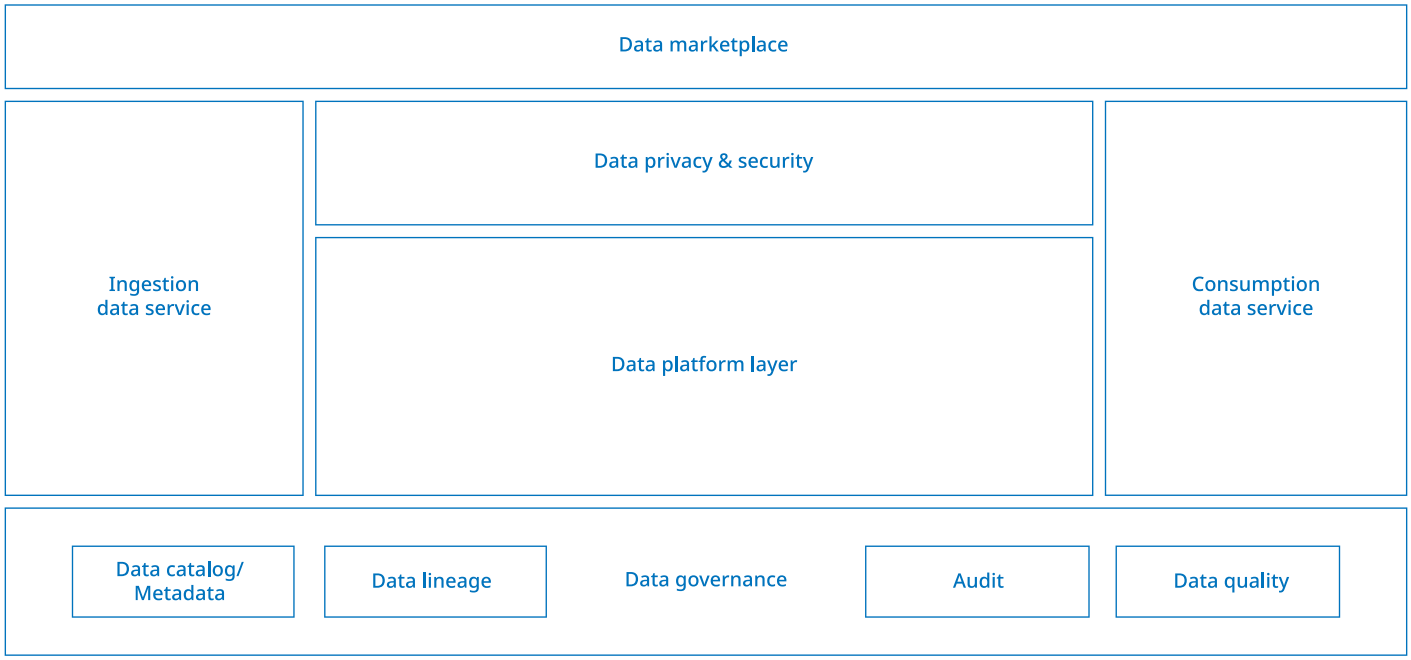
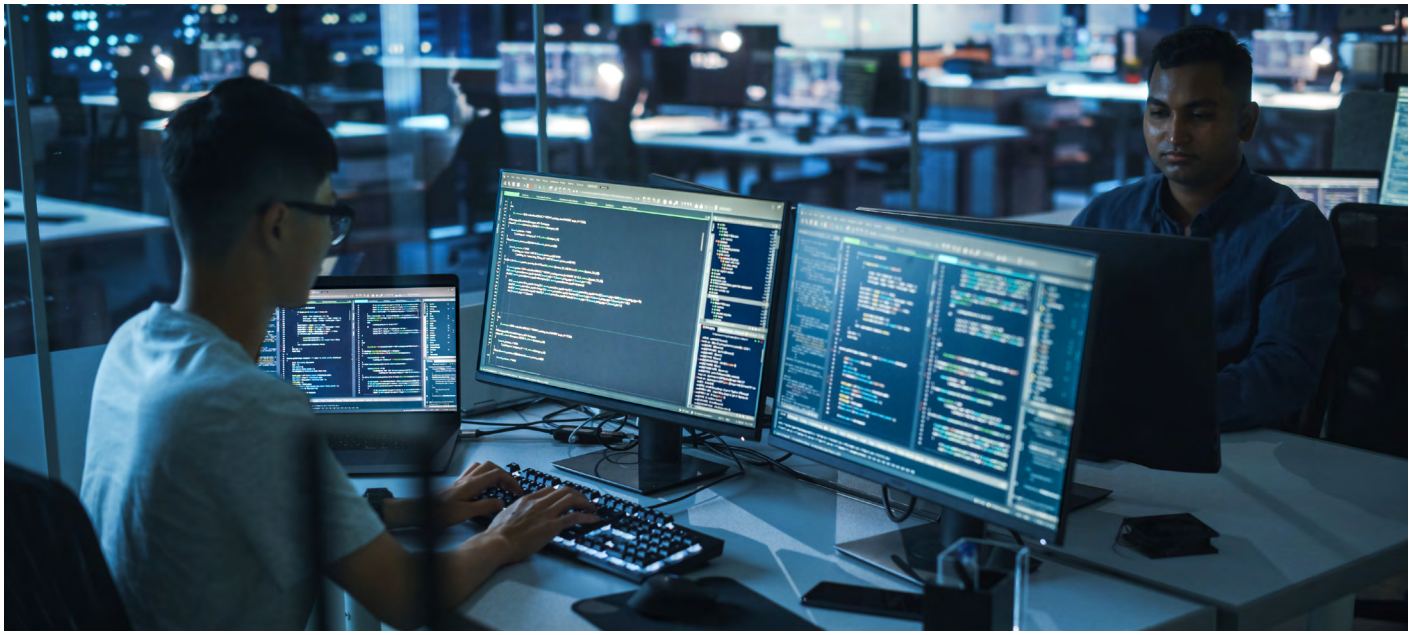



Figure 2. Top-level view of logical architecture.



The Data Marketplace, itself, consists of several interconnected layers, with a logical flow of information between layers. Using virtualized components is a natural way to build a Data Marketplace, but is not mandatory. Figure 3 below shows how to build a logical architecture using virtualization in a data mesh framework. We expect this to be one of the preferred options for building a successful Data Marketplace.

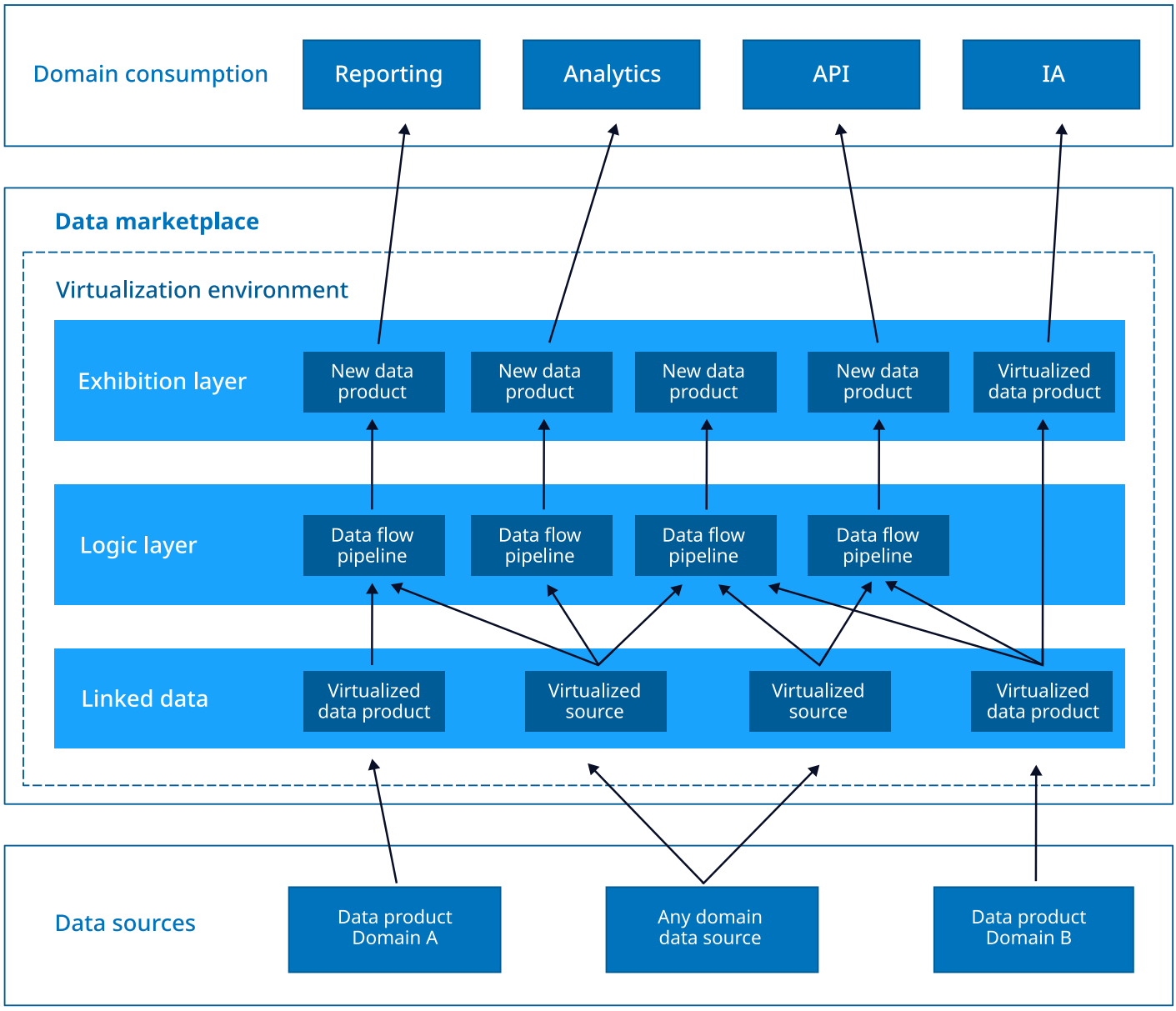



Figure 3. The key layers within the Data Marketplace, top level view.

**01. Data Sources**. This is the layer where all data products potentially available to business users are accessed. In this layer we have repositories related to existing and legacy applications, transaction data, data lakes and other repositories, together with other data that is made available to users in the Marketplace but is owned and stored outside the corporate firewalls (data from partners, clients, specialist research data and much more). This implies the need for governance and rules for accessing such repositories.

**02. Linked Data**. As described earlier in this paper, the key requirement for an operationally efficient Data Marketplace is the ability to access data wherever it is and however it is stored. In this layer, we ensure that users can see and access the data they need in various forms. Some will be stored in repositories and can be accessed there. Some data will need to be virtualized as this is the most efficient way to make it available, despite known issues (scaling, speed of access...).

In some cases, it will be possible for users (subject to security clearance) to work with data inside their applications, that is, by taking the user to the data rather than the other way round). There is no single "right" way to manage this: we ensure that data and the relevant metadata needed for context and background are available in the most appropriate form at all times.

**03. Logic Layer**. This is where business users can work with virtualized data to define and fulfil their own exact business requirements. This could involve creating new data products combining them, transforming, cleaning and filtering them, then saving the output for future use and further development. This layer is absolutely critical for effective value creation, as it gives users a flexible, intuitive toolset than enables them to extract the insights they need from data in any format they desire.

**04. Exhibition Layer**. This provides a display and entry point for business users, enabling them to see the data products they have accessed and developed within their own work, together with linked data products from a wide range of other sources. This layer is also where we provide native interfaces to applications, specialized toolsets, project environments and development platforms, enabling data (which has been prepared in the underlying layers) to be applied quickly and effectively to ongoing work. We have interfaces with APIs, AI platforms, Analytics platforms and Reporting, among other options.

Users can therefore define needs and establish a "straight through" channel that ultimately feeds data from physical repositories, in the lowest layer, right up to the point where, for example, a management report can be automatically produced, reviewed and updated. All with complete confidence as to the security, accuracy and integrity of the data used.

## Summary

This overview is of necessity top level and only skin deep. There is a great deal of technical detail to be covered in order to provide a comprehensive overview. Above all, we need to ensure that establishing a data marketplace does not happen in isolation. We see this as an **integral part of an Intelligent data platform**, which provides a **framework and integration tools** that enable an entire enterprise (and ecosystem) to become **more data driven, agile and creative through more effective access to and deployment of data**. For more insights to this broader topic, please see our separate white paper:

### The Intelligent Platform

# Section 3: NTT DATA vision and roadmap

## Data in the cloud

Data Marketplaces can be set up in existing, on-premise environments, but when enterprises move to Cloud, we expect them to become an integral part of cloud transformation, with deployment across individual hyperscaler or hybrid platforms an inevitable fact of life. This means that the basic architecture, as set out in figure 1, will need to include interfaces and tools that are designed to optimize the capabilities and resources offered by cloud providers.

Data Marketplaces will not be built on “greenfield” sites as a general rule. NTT DATA has spent decades working with, and helping in transformation of large enterprises in virtually every industry sector and geography. We know that each enterprise will have existing investments that need to be repurposed and reused wherever possible, together with complex technology landscapes, created as a result of previous investments and long-term evolution.

Our job is to work with enterprises as they really are. That means we need to manage complexity in the fastest, most flexible and low cost, low risk, low disruption way possible. In the move to cloud, that makes the ability to work with hyperscalers, large application vendors and other solution providers in creative, expert ways to deliver the best possible outcomes for our clients.

NTT DATA has partnering agreements with all the major hyperscale providers and with key application vendors (such as SAP, Oracle, Salesforce and many others). We have “glide paths” to transformation that will take each client enterprise from where they are today to wherever they need to be in the future, safely, smoothly and efficiently.

We can move enterprises directly to AWS or Azure, for example, using native tools and ensuring they receive maximum value from their long-term investment in this primary business partner. We can use snowflake-based methods and protocols to deliver safe transit to hybrid environments, and we ensure secure functionality at every step.

We recognise this subject is complex and requires detailed analysis. We will publish a dedicated paper on the subject in the near future.

## External marketplaces

Our focus in this paper has been on internal Marketplaces, but we recognise that the same structures and frameworks can potentially be adapted for external provision. Under the right circumstances, this can enable an enterprise to monetize their own data, while also adding new service options to their portfolio.

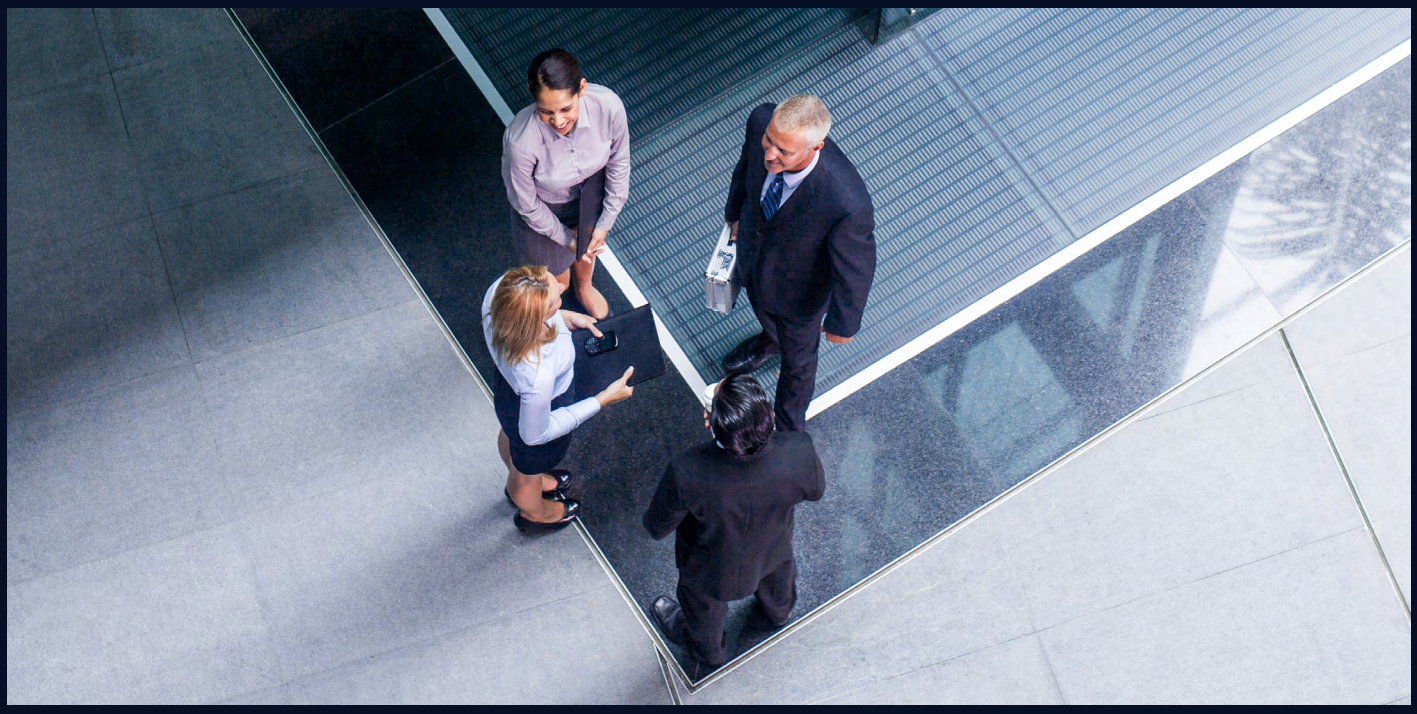
An **External Data Marketplace** is built on the same principles and with the same tools as an Internal Marketplace, but with the addition of Data Services that include as a minimum:

- Dashboard, to enable easy interaction with the service, together with subscription and billing functions.
- Service Availability Management (ANS). This will enable data to be provided securely to authorized subscribers while always enforcing the current rules, legal restraints and regulations.
- Audit, ensuring that a record is kept of all data sales and provision. This is used to provide proof of compliance (by the supplier) and also to ensure that rules are correctly enforced at point of use.
- Billing, with workflow in place to manage subscriptions, provide cost information and manage payments enforced at point of use.

These functions are modular and can be added to the existing internal structures relatively simply. All External Data Marketplaces, however, need to deal with a range of clear requirements and challenges to avoid serious risk. These include:

- Regulatory compliance, proven through audit and accurate, timely reporting.
- Privacy, by guaranteeing protection of personal data, while also meeting GDPR and other legal requirements at all times.
- Trust, which is a major issue for many enterprises- and understandably so. There is a natural reluctance to share internal data with third parties, and that requires enterprises to set up their own internal rules, and then make sure their own employees have confidence in them.
- Service levels, which will define the quality of service to be offered and the payments to be made. Once again, this is a complex matter that will need to be carefully designed and agreed in the same way as for any other service.

An externally-focused data service can be offered to users from other organizations in the same way as any other service offer. Once the Internal Data Marketplace has been built and is fully operational, offering data as a service is an option that remains available at all times. Using data in this way is complex, however, and requires the most carefully preparation and management. NTT DATA is ready to help with this.



## The key to transformation

This is no “right” or “wrong” path to the future. Enterprises must start their journeys to their target future state from where they are now, and that requires a special combination of great flexibility and broad expertise from their change management partners.

**NTT DATA is a leader in all aspects of intelligent networking and cloud technologies**, from Multi-access Edge Computing (MEC) to 5G to AI and beyond. We have exceptional transformational expertise, based on “change factories”, focused on fast transition of complete application families, through to people, culture and organizational change methodologies.

Each enterprise is unique and their path to implementation of a Data Marketplace must reflect their own complexity and the specific context of their organization. By working with the organization and planning a transformation program that complements existing structures, we can help you move quickly, at low risk, without disruption, to the point where your organization has secure data democracy and is able to unlock the full value of your data quickly and easily.


NTT DATA