# Introduction: Dataspaces: A Foundation for Enterprise Digital Sovereignty

To realize a viable data economy, companies and organizations of all types and sizes around the world need to have full control over their data combined with the ability to share their data in a controllable and compliant way. They need an enterprise grade architecture and solution that puts them in control of their data while sharing with other stakeholders such as customers, supply chain, the public, and in certain scenarios even with their competitors.

This architecture begins with the concept of ***dataspaces***. A given dataspace consists of one or more datasets, large or small, static, or streaming, relational, or not, that are configurated for sharing. The dataspace also defines the syntax and semantics of the dataset it represents, the associated policies for sharing of that dataset, the type of identity management used (e.g. whether it is centralized or self-sovereign/decentralized), the security mechanism used to protect that dataset, and last not but least, the laws/regulations and implicit or explicit contracts and policies in effect for (sharing) that dataset.

A dataspace can be formed by any number of parties interested in sharing data in a controlled way. This can be partnerships of multiple companies, or private – public partnerships. Dataspaces can be accessible to a closed set of members or be open and accessible to the public.

Participants in a dataspace usually organize around a central purpose, such as a company and its partners and supply chain, a specific industry, a geographically limited use case, or simply a limited interest group, which then forms the basis of rules and policies for the dataspace. Individual members can be participants of multiple dataspaces and thus also expose their data and services in multiple ways, with rules and policies depending on the requirements of each individual dataspace.

Dataspaces are not bound to a specific technology or processing location. They allow for data exchange across multiple clouds, on-premises, and edge. A dataspace isn’t limited to a single data protocol. Depending on the scenario and technology used it can accommodate blob data, unstructured data, relational data, streaming data and a wide array of protocols and storage technologies.

A possible way to look at dataspaces is as the control plane for data sharing: controlling the flow, filtering, and application of policy on data transfers and data operations that happen in the data plane, but not being prescriptive about the infrastructure and processing layer. Any storage technology or provider can be incorporated in a dataspace.

## Enterprise Digital Sovereignty and its Elements

Any organization would want to exert control over its data and other digital resources, to the extent possible, given the applicable laws/regulations, their organizational governance model and associated risk/reward analysis. This desired control could be about protection of the digital assets as well as the goal of generating value from it (by sharing them with others, for example). This means that the organization needs to assess its risk for loss of control/access to its data by actions of various jurisdictions or other stakeholders. The control over data also requires the digital mechanism for enforcing their policies for sharing of their data only with the explicit consent of the organization (e.g., as part of a contract) or based on proceedings under the applicable jurisdiction in which the organization operates in.

A key difference between political digital sovereignty and enterprise digital sovereignty is that with the former the jurisdiction typically applies to a geo-political boundary (which could be a province/state, a nation or political union such as the EU). The key differentiator is that an enterprise can be subject to many jurisdictions based on where they operate and with whom they trade (for example a multi-national company with manufacturing and sales operations across several continents). A government by contrast is only subject to the laws and treaties it agreed to. The ability to identify and comply with the applicable laws and regulations are particularly important when crossing geo-political boundaries, because this may impact jurisdictional obligations which may need to be transferred to organizations with which data is shared.

At a more technical level, enterprise digital sovereignty requires an organization to have control over its digital identities, control over which digital identities (internal or external to the organization) have access to data, and to store and process data on desired platforms or infrastructures. Enterprise digital sovereignty also requires the ability to move or copy data between different platforms or infrastructures. In other words, enterprise digital sovereignty requires **identity control** (centralized, or decentralized for maximum sovereignty), **access control**, and **usage control**.

One of the important elements of enterprise digital sovereignty is avoidance of ***lock-in***. Dataspaces could help address lock-in concerns of digital platform users in a couple of ways. When implemented with inherently portable, cloud native technologies such as containers and the associated orchestration engine, they provide a technological foundation for data sharing that is inherently platform neutral, addressing ***platform lock-in*** concerns. And because they allow data to be shared using connectors that communicate with each other while running on the desired, but optionally different platforms on each end of the wire, they demonstrate ***interoperability***. In addition, given the data is exchanged across the wire, dataspaces need to address ***data portability*** facets such as syntactic and semantic portability, as well data policy portability defined as “the ability to transfer data between source and destination while complying with the legal, organizational and policy frameworks, including applicable data regulations in areas such as security and privacy[[1]](#footnote-2).”

# Dataspace Conceptual Model

Dataspaces can be an instrument for achieving enterprise digital sovereignty, and if used by governments to maintain political sovereignty. Dataspaces can leverage centralized or decentralized identity to provide desired level of control over identity and build access policies around that. Usage control is done by implementing policies and attaching those inseparable to the data. In addition, dataspaces that are built on top of cloud native infrastructure offer code portability. By nature, dataspaces provide interoperability of data as well as metadata describing data sources and their access policies. Optional semantic models can help build a joint understanding of the meaning of data within a dataspace.

A dataspace is a cornerstone of enterprise digital sovereignty. It consists of one or more datasets, large or small, static, or streaming, relational or not, that are designated for sharing. The dataspace also defines the syntax and semantics of the dataset it represents, the associated policies for sharing of that dataset, the identity management system used (e.g., whether it is centralized or sovereign/decentralized), the security mechanism used to protect that dataset, and last not but least, the laws/regulations and implicit or explicit contract in effect for sharing that dataset.

A dataspace is both an agreement and a supporting technical infrastructure (hardware and software) that enables data sharing between two or more participants.

A dataspace is unique with respect to other data sharing arrangements such as a B2B data exchange because it adheres to the following principles:

1. Participants maintain control (agency) over their identities
2. Participants maintain control (agency) over which other participants they trust
3. Participants maintain control (agency) over their data
4. All data sharing transactions are observable and verifiable

The sharing of the data in a dataspace is accomplished by ***dataspace connectors***. Each dataspace is represented by one or more such connectors that facilitate the actual sharing of the data at runtime, while enforcing the policies and requirements put in place by the data controller in the dataspace. A connector includes executable code and other configuration and metadata artifacts that can be run on any cloud infrastructure, on premises or on an edge device.

There will be complementary documents describing how dataspaces and connectors work, and how various data controllers and data users as well as other stakeholders can come together to share data in a secure and sovereign fashion that satisfies their requirements and addresses the concerns of all stakeholders.

# Dataspace Value Generation: Decentralized Data Applications and Collective Participants

Dataspaces can adopt different levels of maturity. A basic dataspace may exist to exchange data in a peer-to-peer fashion. In this architecture, value is derived from sharing data on a 1:1 basis in an automated fashion.

A dataspace can evolve to provide value beyond peer-to-peer data exchanges by building collective data services and applications. A collective data service aggregates data from multiple participants to create a new offering. Consider an industrial dataspace with multiple supply-chain networks. The dataspace may collectively introduce a parts tracing service that aggregates data from participants to provide new processing capabilities. Each participant shares data with the parts tracing service and attaches specific usage requirements. One requirement could be that participants higher up in the supply chain do not have access to detailed parts data deriving from lower levels in the chain.

In traditional data exchange systems, the “Parts Tracing” application would typically be built using a centralized database. A major drawback of this approach is that a central organization running the database would have access to all sensitive supply chain data. By contrast, a dataspace is fundamentally distributed. In a dataspace, the parts chain application would be partitioned and hosted by participants on disparate computing infrastructure. The dataspace provides the connective fabric that ties the data partitions into a coherent whole.

Sometimes it will be necessary to expose the results from applying an algorithm or data application on data shared within the dataspace to the members of the dataspace as a common data asset or service of the dataspace. In this case the dataspace has a need for one or more collective participant(s) that represent data and services that are forming collective value of the dataspace and are not controlled by an individual member alone. Control and rights to the collective results is usually governed by legal structures and policies put in place for the dataspace. Technically those collective participants act like any other participant, following policies and exposing an identity.

1. ISO/IEC 19941: Free download link is [here](https://standards.iso.org/ittf/PubliclyAvailableStandards/c079573_ISO_IEC_19944-1_2020(E).zip). [↑](#footnote-ref-2)