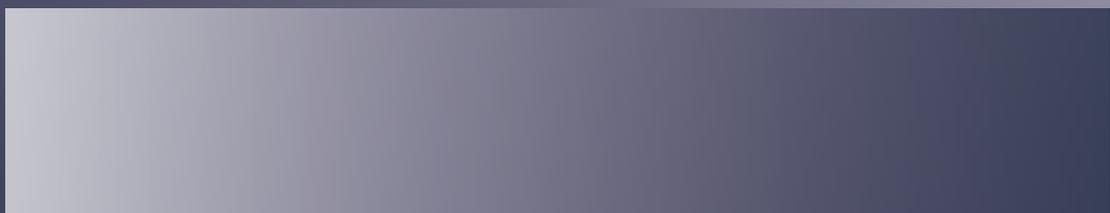
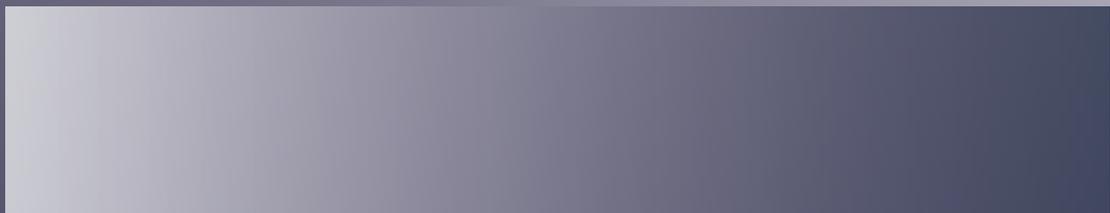
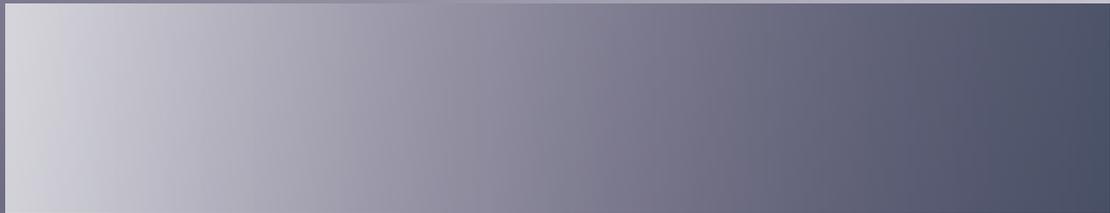


Harbr's Data Exchange Glossary



Data Exchange

A [data exchange](#) is the transfer of data between two parties, typically a data owner and a data consumer. An [exchange](#) can involve any type of data and can include data access or taking a copy of the data. Exchanging data supports many use cases that require access to data that has been created or modified in a different environment to where it needs to be used. This includes internal and external data sharing, external data acquisition and even data monetization.

Data Exchange Platform

A Data Exchange Platform is software that supports the exchange of data between a supplier and a consumer. You can either operate or participate in a data exchange platform. AWS Data Exchange is an example of a data exchange platform you participate in, which supports a data marketplace use case featuring multiple suppliers and consumers that can all interact with each other. Harbr is an example of a data exchange platform you operate, which can be structured to support multiple use cases by customizing the interaction between suppliers and consumers. If you participate in a data exchange you are bound by the rules set by the operator, usually including revenue sharing arrangements, and you have no control over other participants. If you operate a data exchange you can create a branded experience and set the rules for all participants. You can also access platform-level reporting to understand, manage and optimize the platform.

Data Product

A [data product](#) is a container for one or more digital assets or services to support controlled and scalable transactions between data product owners and data consumers, and the ongoing consumption of the assets/services. [Data products](#) can contain any combination of assets in any digital format, can be static or updating, and can be of any size or volume and services are typically provided via APIs. The aim of a [data product](#) is to reduce the time to value and cost of ownership for the data consumer while providing control, auditability and feedback loops to the data product owner.

Data Monetization

[Data monetization](#) is generating measurable economic benefits from data assets and services that have typically been 'productized' to ensure scalability and reliability.

Data Cleanroom

A Data Cleanroom is a virtual environment where data consumers can access data but cannot remove it. A Data Cleanroom is typically used to provide data access without the risks associated with losing custody of proprietary, confidential or legally restricted data.

Collaborative Data Cleanroom

A [Collaborative Data Cleanroom](#) is a virtual environment where multiple parties can collaboratively work on data but no party can remove it. Typically, collaborative clean rooms are used to support use cases where the data, models and expertise from multiple parties is required to achieve an outcome. A Collaborative Data Cleanroom delivers the benefits of data sharing and data collaboration without the risks associated with losing custody exposing proprietary, confidential or legally restricted information.

Data Analytics Cleanroom

A [Data Analytics Cleanroom](#) is a virtual environment with specific data processing capabilities that delivers predetermined outputs to data consumers without giving either party access to the underlying data or model. A Data Analytics Cleanroom is typically used to support the delivery of analytical outcomes without the risks associated with providing access to proprietary, confidential or legally restricted data.

3rd-Party Data

[3rd-Party Data](#) (or '[Third-Party Data](#)') is information provided by a separate, external organization.

Data Marketplace

A [Data Marketplace](#) is an environment for buying and selling third-party data. [Data marketplaces](#) facilitate the external exchange of data via financial transactions and may be built on a Data Exchange Platform.

Data Catalog

A [Data Catalog](#) is an organized inventory of the data within an organization built upon extensive metadata. Data catalogs help data owners and consumers understand their internal data landscape including what is available, where it is stored and who owns it.

Data Democratization

[Data democratization](#) is a phrase used to articulate a future where data is easier to discover, access, transform and use, reducing the barriers to entry for a wider range of people to be involved in data-related activities and enable better business outcomes.

Data Lake

A [Data Lake](#) is a repository that stores and processes large amounts of structured and unstructured data via an ELT (Extract, Load and Transform) process. It provides greater flexibility than a Data Warehouse, because it does not require the data to be transformed to a preconceived schema in advance of it being used so the most appropriate transformation can be applied depending on the use case. However, this concept can lead to scenarios where the data stored is not well-understood and a lack of governance means it can become a 'data swamp' that becomes expensive and fails to deliver the anticipated value.

Data Warehouse

A Data Warehouse a type of data management system that centrally stores and processes large amounts of data that has been converted into a structured, SQL-queryable format via an ETL (Extract, Transform and Load). A data warehouse provides less flexibility than a Data Lake because it structured and unstructured data. It provides greater flexibility than a data warehouse, because it requires the data to be transformed to a preconceived schema in advance of it being used, which can also become very expensive. However, it does mean the data is relatively well-understood in comparison to a Data Lake and is ready-to-use for the range of use cases it supports.

Data Mesh

A [Data Mesh](#) is a distributed data network that connects localised storage and processing technologies to facilitate access to data products that are managed and maintained by data owners. It provides an alternative approach to the monolithic Data Warehouse and Data Lake architectures and also foregoes the data virtualisation where virtual copies are made to support remote querying. Instead the focus is on creating consumable data products that can be delivered and maintained at a given source, which may be a lake, or a warehouse and may also utilise virtualisation, but could also be other types of storage and processing technologies. A Data Mesh can be enabled by the Harbr Data Exchange Platform that operates as the system of record for data products and offers both a centralised storage and processing layer and the ability to coordinate with distributed storage and processing layers.

