Bridging Organizational Boundaries with Data Sharing Platforms

Eckerson Group

Data Pain Points

Data, data, everywhere… As data volume and complexity grows unabated, organizations are faced with the Ancient Mariner's paradox of thirsting for the very thing they're drowning in.

In many organizations, business users can't easily find or evaluate data they need to build solutions. They spend too much time navigating and mashing together data from various source systems, or they wait days or weeks for permission from busy data owners to access data files.

Conversely, data owners are often besieged with data requests from both internal and external partners that they must individually evaluate to ensure conformance with data security and privacy policies. This situation is inefficient at best and fraught with risk at worst.
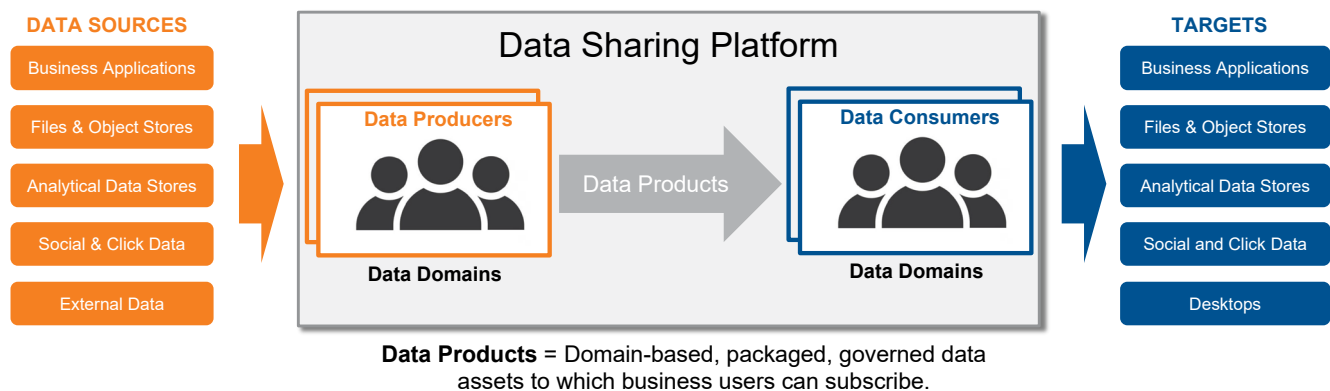
What is a Data Sharing Platform

A data sharing platform (DSP) addresses these pain points and more. It facilitates the creation and consumption of data assets and data products across data domains. A data asset could be a table, a PDF, a dashboard, a model, etc., while a data product consists of one or more data assets that have been packaged and permissioned for self-service consumption at scale. It fosters a culture of analytics by simplifying data access while governing critical data assets.

> *"A DSP fosters a culture of analytics by simplifying data access while governing critical data assets."*

A DSP takes the friction out of sharing and consuming data products. It bridges organizational boundaries and silos. It is the missing ingredient in a decentralized data mesh approach, and it can work alongside centralized approaches, such as data warehousing.

A DSP Connects Data Producers and Consumers Across Domains

DATA SOURCES
- Business Applications
- Files & Object Stores
- Analytical Data Stores
- Social & Click Data
- External Data

Data Sharing Platform

Data Producers — Data Domains

Data Products

Data Consumers — Data Domains

TARGETS
- Business Applications
- Files & Object Stores
- Analytical Data Stores
- Social and Click Data
- Desktops

Data Products = Domain-based, packaged, governed data assets to which business users can subscribe.

Benefits

A DSP makes it easy for:

> Data producers to create, publish, and control access to data products.

> Data consumers to find, evaluate, subscribe, and use data products.

> Data product managers to track consumer interests, needs, and data flows.

> Business users to launch private workspaces to collaborate around shared data.

A DSP automates data sharing in a safe, secure way and alleviates the pain and risks involved in manually sharing data assets. It also makes it easier for organizations to adopt a data product paradigm. Ultimately, DSPs accelerate data usage, improve data literacy, and foster a data-driven organization.

A DSP is a … requirement for self-service analytics and a product development paradigm.

The Role of a DSP in an Enterprise Data Architecture

Data Assets
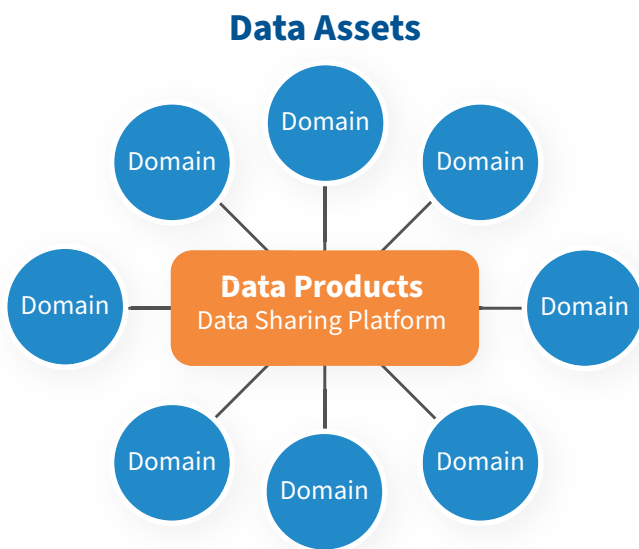
Data Architecture. A DSP is a critical component in an enterprise data architecture and a requirement for self-service analytics and a product development paradigm. A DSP integrates data from diverse domains and systems by providing a central mechanism for publishing and navigating data products.

Versus a Data Catalog. A data sharing platform is not a data catalog although it shares many of its characteristics. A DSP enables business users to discover data products, like a data catalog, but also to access and subscribe to those data products. A DSP can incorporate metadata from a data catalog when publishing data products.

However, unlike a data catalog, DSPs also enable data developers to query, filter, and aggregate data assets and create data pipelines that can be turned into data products. A DSP also enables business users to access and consume data products directly, either in their own environment or on the DSP itself using platform-embedded tools of their choice.

Packaging Data Products

In a DSP, data producers turn data assets into data products through a packaging process. Here, they define:

1. **Product content** by selecting files or constructing queries or pipelines from data assets.

2. **Business metadata,** including description, definitions, date created, users, related products, etc.

3. **Technical metadata,** including schema, lineage, attributes, tags, statistics, creator, etc.

4. **Social configuration,** including whether to accept ratings, reviews, and comments.

5. **Access rights.** Who can browse and access the data product in the catalog.

6. **Subscription options.** The content, frequency, and duration of consumption.

7. **Delivery options.** The channel, format, and target where the data product is delivered.

8. **Terms of use.** How the product may be used, e.g., in commercial products or not.

9. **Certification.** A mark or seal indicating that the product has gone through a governance review.

10. **Pricing or cost.** If a product is for sale or subject to chargebacks.

11. **Track and monitor** product usage and subscriptions.

Consuming Data Products

A DSP replicates the experience of using an online retail marketplace, such as Amazon.com. It enables data consumers to:

1. **Browse and search data products** in an online catalog.

2. **Evaluate data products** by reading business and technical metadata and social reviews.

3. **Sample data products** by viewing a subset of data or an entire data set on a limited basis.

4. **Launch a sandbox** by inviting peers to jointly view and manipulate a data product.

5. **Create derivative products** by combining data products into new data products.

6. **Select subscription and delivery options** to obtain a data product.

7. **Define a team** of departmental users authorized to view and/or subscribe to data products.

8. **Purchase a data product** or register a chargeback, if required.

9. **Track** and monitor subscriptions and orders.

Managing Data Products and their Ecosystem

A DSP enables:

1. **Platform admins** to manage a multi-tenant ecosystem with tailored setups for each client.

2. **Product admins** to configure options for subscriptions, terms, delivery, etc. for their products.

3. **Product managers** to track subscriptions, deliveries, and browsing activity.

4. **Data producers** to evaluate consumption activity and interact with prospective subscribers.

5. **Business managers** to create a team of users who can browse and subscribe to products.

6. **Data consumers** who can track their individual orders and subscriptions.

Challenges

Awareness. Data sharing platforms are an emerging technology that few organizations know about. Most don't know there is a solution to their vexing data sharing pains. Even the ascendant data mesh methodology doesn't mention a DSP even though sharing data products is fundamental to the approach.

Habit. In addition, data sharing is scary for data owners who are used to managing data with an iron grip to eliminate security risks or wield power within an organization.

Governance. Finally, a data product, by definition, is governed. This requires organizations to implement governance and product management processes, such as review boards. Not all organizations have the knowledge or appetite to implement governance controls for data products.

Ecosystem. Govern whole ecosystem – set up indiv marketplaces which can be different for every user in an org. usage stats

Recommendations

Readiness. To assess your readiness for a data sharing platform, we recommend that you:

1. Ascertain the extent of manual data sharing occurring in your organization.

2. Evaluate your organization's interest in implementing data products using data mesh or another approach.

3. Evaluate your organization's appetite for governing and curating data products.

4. Quantify external demands for internal data products and solutions.

5. Assess which business domains are comfortable creating and consuming data products and which need education and training.

Platform Selection. To select a data sharing platform, consider a product that:

1. Supports both internal and external data sharing.

2. Doesn't require storing data assets or data products on its platform.

3. Enables data producers to create virtual data products via federated queries.

4. Enables data consumers to manipulate data products on the DSP using embedded tools.

5. Provides robust reporting so administrators, producers, and consumers can track activity.

Eckerson Group

RESEARCH · CONSULTING · ADVISORY

G E T · M O R E · V A L U E · F R O M · Y O U R · D A T A

About Eckerson Group

Eckerson Group is a global research, consulting, and advisory firm that helps organizations get more value from data. Our experts think critically, write clearly, and present persuasively about data analytics. They specialize in data strategy, data architecture, self-service analytics, master data management, data governance, and data science. Organizations rely on us to demystify data and analytics and develop business-driven strategies that harness the power of data.

Learn what Eckerson Group can do for you!

Harbr_

About Harbr_

Harbr's customers rely on their data sharing platforms to deliver award-winning, world-class data experiences. Harbr's market-proven technology unlocks immense value for its customers and the full spectrum of their data users. Through deep relationships and high levels of service, Harbr empowers its customers to become truly data-driven organizations.