Deep Dive on Data Exchanges: Three Tools to Consider

Different Approaches to the Data Exchange

BY JOE HILLEARY

FEBRUARY 2022

CUSTOM REPRINT PREPARED FOR HARBR

Harbr_

About the Author

Joe Hilleary is a writer and a data enthusiast. He believes that we are living through a pivotal moment in the evolution of data technology and is dedicated to helping organizations find the best ways to leverage their information. With a background in both analytics and the liberal arts, he crafts clear, articulate narratives on technical topics that empower stakeholders to make informed decisions. Hilleary is a Senior Research Analyst at Eckerson Group.

About Eckerson Group

Eckerson Group is a global research and consulting 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 them to demystify data and analytics and develop business-driven strategies that harness the power of data. **Learn what Eckerson Group can do for you!**

About This Report

To conduct research for this report, Eckerson Group interviewed numerous industry experts and practitioners. The report is sponsored by Snowflake, Dawex, and Harbr, who have exclusive permission to syndicate its content.

This is an excerpt from a larger report that profiles three data exchange products. For the full report, go to:
https://www.eckerson.com/deep-dive-on-data-exchanges-three-tools-to-consider.

Table of Contents

This is an excerpt from a larger report that profiles three data exchange products. For the full report, go to:
https://www.eckerson.com/deep-dive-on-data-exchanges-three-tools-to-consider.

Executive Summary

Data is one of the most valuable assets in the portfolio of the modern enterprise. Not only does it drive value internally through the insights it provides, but it can be converted into an additional revenue stream when sold to third parties. Until recently, however, buying and selling data was messy and complicated. Companies struggled both to find relevant data and validate its quality before purchasing it. Methods for integrating external data lacked consistency, demanded significant resource expenditures, and, at times, raised concerns about data security. In the last few years, data exchanges have started to solve these problems.

Data exchanges provide a user-friendly experience within a secure platform to facilitate the sharing and monetization of data. They mirror online marketplaces for other consumer goods by increasing the convenience of locating and purchasing third-party assets while improving trust in the products by ensuring base levels of quality.

Given the exponential growth in both the demand for and supply of data, numerous data exchanges have risen to meet the challenge of connecting buyers and sellers. This report will profile three data exchanges, each with a different approach to solving the need for secure, easy-to-use environments in which to share data. The report will enable data leaders to better understand how data exchanges vary and which are best suited to the requirements of their organization.

This is an excerpt from a larger report that profiles three data exchange products. For the full report, go to:
https://www.eckerson.com/deep-dive-on-data-exchanges-three-tools-to-consider.

Introduction

Once, organizations only had access to data they generated. Any analysis they wished to conduct was limited to the proprietary data they created or collected. Then professional data vendors entered the scene. Companies such as Nielsen, S&P, and Bloomsburg specialized in data collection and sold their data to corporations, which used it to enrich their in-house data. This external data, while useful, was limited in scope, expensive to acquire, and difficult to integrate into internal systems. In the last decade, however, we've entered a golden age of external data thanks to advances in data sharing technology and an explosion in the size of the market for third-party data.

Technological Advancement

In the earliest days of sharing, vendors would send data on floppy disks, which organizations would physically insert into internal machines. Today, data sharing is more seamless and secure. Data providers deliver data through application programming interfaces (APIs) or the cloud, allowing consumers to integrate these sources directly into their environments. Data exchange platforms, the underlying technology that supports data exchanges, also enable a better acquisition experience, allowing data consumers to search for specific data products and vet the data before purchase. Data catalog-like features let consumer see metadata and compare offerings for multiple data products in the same place. Often, they can even view sample data to confirm its utility. At the same time, these platforms make it easier for data providers to distribute their data products by taking away the burden of infrastructure development.

Market Growth

Digital transformation initiatives and new digital-first companies have created an environment in which organizations in every sector, not just data vendors, generate enormous quantities of data. At the same time, attempts to become more data-driven have increased the appetite for external data. Recognizing this rising supply and demand, new companies have sought to disrupt traditional data vendors and ushered in a new paradigm for data providers. Using data exchanges, any business can become a data provider— from start-ups developing their own exchanges on top of a platform to companies simply listing data products on an open exchange, traditional data vendors make up a decreasing share of data providers.

Key Terms for Data Exchanges

A data exchange, at its most basic, is an environment where a data provider can share data with a data consumer. Data exchanges can be public or private, internal or external, free or paywalled. What they have in common is a product-based approach.

Data products. Data exchanges exist to facilitate the sharing of data products. The simplest data product is a static data set in the form of a flat file. More sophisticated products might include multiple data sets, live API connections, or even the results of machine learning models run against proprietary data. Essentially, they can be any data asset that one party wishes to send to another. The only limitations on size, type, format, and freshness are those imposed by the exchange platform. The price or lack thereof of a data product is determined by the provider.

Providers. A data provider distributes data products via the exchange. It can sell the products or provide them for free, but ultimately the provider owns the data it lists on the exchange. Providers often include traditional data vendors, but the ease of use of many exchange platforms has allowed any organization looking to monetize its data to become a data provider. This second group of providers holds the most potential for growth, as companies from different sectors start to make new data available to consumers for the first time.

Not all providers seek to make a profit, however. Some organizations use exchanges internally, in which case the data provider is the line of business that owns a particular data asset. In this context, the provider shares the data for free, but still manages it and makes it available to its internal data consumers.

Consumers. A data consumer acquires data products from the exchange. Traditional data consumers were corporations with the resources to purchase and integrate data from data vendors through large-scale subscriptions. The ability to pick and choose data products and to acquire data on a smaller scale through data exchanges has diversified the profile of data consumers. Now, companies of all sizes purchase à la carte data and can acquire data products on a limited basis. As in the case of data providers, not all data consumers are external. For internal exchanges, data consumers might be other departments or data science or analytics teams.

Operators. The final key persona on a data exchange is the exchange operator. A data exchange operator is the entity responsible for maintaining the data exchange platform. It can either build the platform from scratch or use a pre-built one provided by a data exchange platform vendor. The operator hosts the exchange and incurs the costs of keeping up the platform, although it can pass along those costs to providers and/or consumers in the form of fees to use the exchange. The identity of the operator typically depends on the model of the exchange.

Types of Data Exchanges

Data exchanges come in many forms, but we can group them by their overarching sharing patterns into three general categories.

A distribution hub is a one-to-many data exchange in which the operator is also the primary data provider. The operator might be a dedicated data vendor, a company looking to build a sideline by monetizing its data while cutting out the middleman, or an organization that needs to share data with a large ecosystem of partners.

A consumption hub is a many-to-one data exchange in which the operator is also the primary data consumer. This pattern is most typical of large corporations or government entities that consume data from numerous external sources and that have the heft to force all of those providers onto a single platform for its convenience.

A data marketplace is a many-to-many data exchange in which a variety of providers and consumers exchange data on a platform hosted by a dedicated operator. Strictly speaking, the operator of a data marketplace might also participate in it as a provider and/or consumer, but what distinguishes a data marketplace is the exchange of data between third-parties that are not involved in operating the exchange. Data marketplaces can be broad, general, and open to the public, allowing anyone to acquire data of any type, or they can be more specialized. Some marketplaces focus on a single industry or consist of invited, subscription-paying participants to ensure high-value data doesn't get buried.

This report profiles three approaches to data exchanges—one open data marketplace, Snowflake Data Marketplace, and two data exchange platforms, Dawex and Harbr, that can be used to operate any model of data exchange. Each places emphasis on different features of the data exchange and provides advantages in certain areas. The goal is to demonstrate the breadth of the technological capabilities within the modern data exchange space and to provide a point of comparison for other data exchanges.

Harbr

Founded: 2017
CEO: Gary Butler
Product: Harbr
Initial Product Launch: 2020

Executive Summary

The Harbr platform not only helps organizations create their own internal or external data exchanges, but it also provides a set of tools for building and managing data products. Most Harbr customers are large organizations that serve as the hub for consumption or distribution networks, or companies just starting to build out data business lines. Harbr delivers a secure software-as-a-service (SaaS) offering that allows data to live in the owner's virtual private cloud rather than Harbr's. Its unique "spaces" feature brings users together to facilitate collaboration and encourage the development of novel data products. It recognizes that most data products are on-going services and puts the means to deliver those services in the hands of data owners.

Background

Company

Gary Butler and Anthony Cosgrove co-founded Harbr in 2017. Butler brought expertise with tech start-ups from his previous venture, Evergreen Technologies, which sold to Greenplum in 2009, while Cosgrove drew on his experience leading the big data intelligence team at HSBC. Both felt the explosion of data generation in the wake of widespread digital transformation efforts and cloud migrations warranted a new approach to data. Specifically, data needed to pay for itself. As organizations created and stored more data, the costs of doing so increased. Butler and Cosgrove's solution—Harbr—helps companies develop and distribute data products, in order to realize the value of their data assets. It aims to facilitate collaboration between business and technical stakeholders, so data owners can discover previously unconsidered business opportunities for their data. In addition, it encourages the productization of data and the creation of a consumer-centric experience that appeals to business users.

Butler and Cosgrove felt the explosion of data generation in the wake of widespread digital transformation efforts and cloud migrations warranted a new approach to data. Specifically, data needed to pay for itself.

Harbr emerged from stealth in May of 2020. Its series A funding round, led by Tiger Global Management and Dawn Capital, brought in $38.5 million in November that same year, helping total investment in the company reach $52 million. Headquartered in London, Harbr now employs 80 people and serves about 140 different organizations.

Customers
Harbr customers come from a range of industries, but generally fall into one of three buckets:

Large existing data vendors or consumers. About half of Harbr's customers are large data vendors that use the platform to distribute data products to their clients or large-scale data consumers that rely on Harbr to consolidate and organize data subscriptions coming from multiple providers. Some of these customers manage thousands of data products. For example, one of Harbr's flagship customers is Moody's, which provides data products to its own customers via a one-to-many Harbr-powered exchange.

New data vendors. An emerging segment consists of pure-play data monetization start-ups, typically series B or later, that see collaborating with customers on data products as a way to differentiate their offerings from the data commodity market.

Sideliners. Like other exchange vendors, Harbr also draws companies that already have a significant digital footprint and want to create more value from the data they possess. These organizations generally have a high volume of recent data that's externally valuable to an audience that already engages with them.

Product

The Harbr data exchange platform is a SaaS offering available on Amazon Web Services (AWS), Google Cloud Platform (GCP), or Microsoft Azure. To ensure a high-level of security, Harbr segregates the platform into two different cloud environments. One, a virtual private cloud managed by Harbr, retains the codebase and the core service engines for the software. The other, a virtual private cloud managed by the customer, contains all of the actual data and metadata for the data products. The two clouds communicate via API to execute services without any transfer of data or other intellectual property. This approach also allows the customer to take advantage of any preferential pricing arrangements it has with cloud providers, rather than essentially renting cloud space through Harbr.

Setting Up an Exchange
Once the software is up and running on the cloud, the exchange operator can start to invite participants. These participants could be its customers, data providers, or even different departments within its own organization. To invite new users, the exchange operator must first create a profile for the organization

to which the user belongs. In addition to the name of the organization, the operator can add contact information, designate admins to manage the account going forward, and specify the organization's type.

Types correspond to different roles on the platform and restrict the kinds of interactions members of the organization can have. For instance, the exchange operator might define the following types: provider, customer, and support. It could then decide that provider and support organizations can interact with any other accounts, but customers can't interact with other customers. This approach allows the exchange operator to manage its relationships with a variety of partner organizations in a single platform without the partners necessarily being aware of one another.

> *Harbr's approach to organization types allows the exchange operator to manage its relationships with a variety of partner organizations in a single platform without the partners necessarily being aware of one another.*

Once an organization exists on the platform, admins can add new individual users and specify their roles within the organization. These roles include product manager, technical, ecosystem administrator, administrator, internal, and general, each of which has permissions to perform certain functions in the platform on behalf of the organization. Admins can also specify the endpoints where organizations will deposit or retrieve the data they provide to or acquire from the platform. Today, an endpoint can be either a Secure File Transfer Protocol (SFTP) connection or cloud object store.
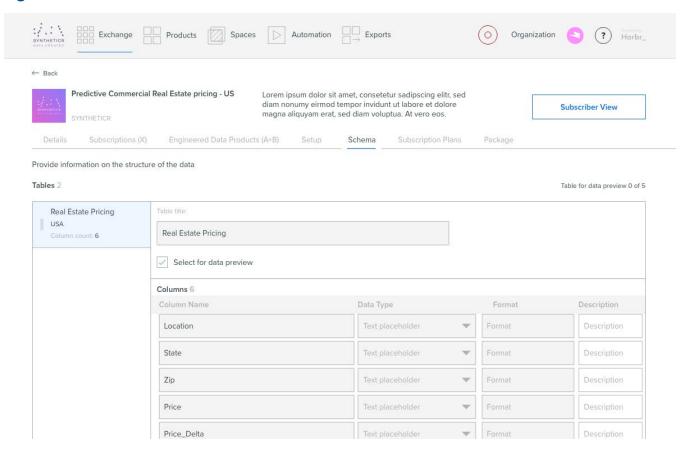
Developing a Data Product

Every product on a Harbr exchange offered by a data provider has a product manager assigned to it. This manager oversees the development and delivery of the data product. The product manager builds a data product by bringing data into the platform via their organization's established endpoint. This data can be any digital asset, whether structured or unstructured. For structured data, Harbr automatically identifies the schema of the data and generates a data dictionary. (See figure 7.)

The product developer can then explore the data's structure, enrich it, and decide what tables to include in the product. They can also edit the data dictionary to provide greater clarity for consumers.

After defining the product on a data-level, the product manager chooses subscription elements. They decide who can subscribe or view their product, how long subscriptions last, and where consumers can use their product—on the platform, off the platform, or both. They can also select a variety of license options for consumers of the product to choose from. Finally, the product manager builds out a profile for their product that contains a detailed description and basic metadata. If they want, this profile can even include visualizations of the data.
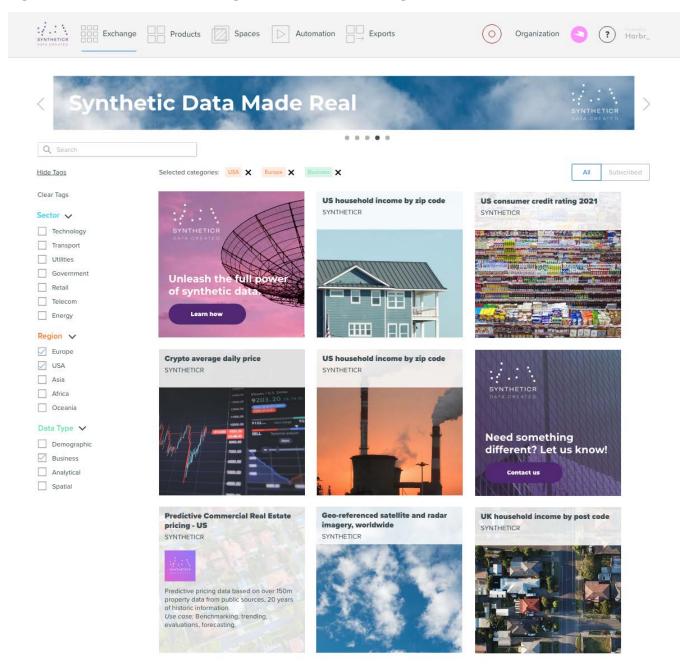
Figure 1. Structured Data in Harbr

Consuming a Data Product

Data consumers join a Harbr-powered data exchange as either existing customers of data providers that have moved their distribution to the platform, or as potential customers the operator thinks would be interested in the exchange offerings. Once added to a data exchange, the data consumer can browse for data products. (See figure 8.)
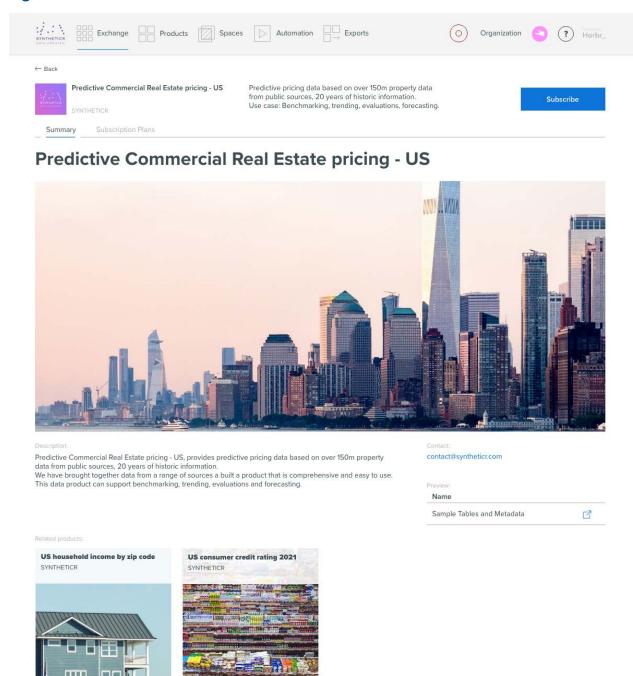
Figure 2. The Harbr Data Exchange Product Overview Page

In addition to any data products to which they already subscribe, consumers can search for new products via keyword search. This search matches the search terms to the titles and descriptions. The platform also suggests data products to the consumer based on the other products they've reviewed or consumed, their role, and their organization. In both cases, the platform limits search results or product suggestions to the data products that providers have invited the specific consumer to view. When a consumer selects a product tile, they can view its complete profile, as defined by the data provider. (See figure 9.)

Figure 3. Harbr Data Product Profile

If a new data product interests the data consumer, they contact the product owner directly using the contact information provided in the platform. From there, the consumer and the provider discuss next steps. The provider might give the consumer trial access, or the consumer might go ahead and purchase the product. Rather than requiring packaged pricing within the platform, Harbr lets providers and consumers negotiate custom pricing agreements and payment methods externally, based on the idea

that most data sharing relationships are formed at the enterprise level. When contracts are signed, the provider simply toggles permissions within Harbr to give access to the consumer.

Once a consumer has access to a data product, they can either have the product delivered to their designated endpoint or use the data in a "space" on the platform. While the first option is a straightforward way to bring data into the consumer's environment, the second is a unique feature of the Harbr platform. A space is a temporary environment hosted on the platform itself. When a consumer spins up a space, they can select a set of tools to use with the data from a stack of open-source technologies including R, Python, Hadoop, Spark, Zeppelin, and Hue. They can also invite other platform users to work with them to analyze or use the data in the platform environment.

The space provides a secure, neutral territory where collaborators can bring together and use data from different organizations. It allows consumers to work directly with providers to troubleshoot or test drive data products and helps providers understand how consumers actually use their data. It even facilitates the creation of new data products, as any space can be converted into a new data product and exported or made available on the exchange, given the appropriate licensing and permissions.

A "space" provides a secure, neutral territory where collaborators can bring together and use data from different organizations.

Pricing

Because the data never leaves the customer's own cloud environment, Harbr does not charge for storage or compute. Instead, it offers a base license that covers the cost of deployment. As the exchange grows, the price of the software increases. The most common metric it uses to determine pricing is the number of organizations on the exchange, but it also provides flexible pricing models based on customers' situations. Ultimately, Harbr works out pricing with customers on an individual basis, finding a framework that meets their needs.

Recommendation

Harbr provides a data exchange platform with a robust set of capabilities that enable in-platform data productization and consumption. Its "spaces" concept allows for iterative development and collaboration, fostering the creation of tailored solutions. Ultimately, Harbr delivers a complete tool set for multiple parties to work together in a product-driven approach to data sharing.

Harbr is ideal for customers that:

> Value collaboratively developing and working with data products within an exchange environment.

> Want to operate a fully-branded exchange while also serving as the primary data provider or consumer on that exchange.

> Need an exchange that supports both structured and unstructured data.

> Would like to keep data in their own cloud environment and negotiate storage and consumption costs with cloud providers directly.

Conclusion

Data exchanges are positioned for explosive growth. Thanks to a decade of digital transformations, many companies are poised to become first-time data providers. Simultaneously, data science projects and efforts to become more data-driven have created an enormous demand for third-party data. Data exchanges sit at the intersection of these two trends, helping organizations distribute and consume data products with unprecedented ease. Whether as a data consumer, provider, or exchange operator, most large companies will engage with a data exchange sometime in the next five years, if they haven't already. When that time comes, it will be important to understand the difference between exchanges. The following guide gives you questions to help you find the best fit, regardless of role you might take on.

Questions for Data Consumers:

> Who operates this exchange?
- What do they want out of it?
- Do I trust them?

> Do I have to pay to participate?

> What formats can I consume?

> Does this exchange contain data relevant to me?

> Am I able to negotiate directly with providers to get better pricing?

> Do I need to manage regulatory compliance questions or will the operator/provider do so for me?

> Can I purchase data on-platform?

Questions for Data Providers:

> Who operates this exchange?
- What do they want out of it?
- Do I trust them?

> Do I have to pay to participate?

> Who will be able to view my data products?
- Do I have any control over it?

> Can I set my own pricing?

- – What are the limitations on pricing models?

- – Can I take payment through the exchange?

> Can I build data products on-platform?

> What metrics can I view for my consumers?

Questions for Exchange Operators:

> Can I brand the exchange?

> How secure is the environment?

> What exchange models does the platform support?

> Can I control access?

> Can I monitor participants?

> Where is the exchange hosted?

- – What is the cost?

  - – Do I pay it directly?

- – Can I leverage existing relationships with cloud providers?

- – Are there compliance issues at stake?

> How do participants find the exchange?

> Can I charge for access within the platform?

> How will my costs increase with:

- – More participants?

- – More data product listings?

- – More transactions?

- – More data volume?

About Eckerson Group

Wayne Eckerson, a globally-known author, speaker, and consultant, formed **Eckerson Group** to help organizations get more value from data and analytics. His goal is to provide organizations with expert guidance during every step of their data and analytics journey.

Eckerson Group helps organizations in three ways:

> **Our thought leaders** publish practical, compelling content that keeps data analytics leaders abreast of the latest trends, techniques, and tools in the field.

> **Our consultants** listen carefully, think deeply, and craft tailored solutions that translate business requirements into compelling strategies and solutions.

> **Our advisors** provide competitive intelligence and market positioning guidance to software vendors to improve their go-to-market strategies.

Eckerson Group is a global research, consulting, and advisory firm that focuses solely on data and analytics. Our experts specialize in data governance, self-service analytics, data architecture, data science, data management, and business intelligence.

Our clients say we are hard-working, insightful, and humble. It all stems from our love of data and our desire to help organizations turn insights into action. We are a family of continuous learners, interpreting the world of data and analytics for you.

Get more value from your data. Put an expert on your side. **Learn what Eckerson Group can do for you!**

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 the Sponsor

The Harbr data commerce platform powers high-margin data businesses and revenue streams. It provides a white-labelled storefront to deliver a refined customer experience and own the customer relationship and enables selection from multiple business models. Collaborating on data products in real-time enables providers to unlock high-value use cases, while customers get the exact outcome they need. That's why high-margin data businesses develop, launch, and scale on Harbr.

Learn more at **www.harbrdata.com.**