Capgemini

REIMAGINING DATA SOURCING AND CONSUMPTION TO MAXIMIZE BUSINESS VALUE

– Mapping the data opportunity beyond enterprise walls

890 by Capgemini

REALIZE A FUTURE WHERE DATA EMPOWERS YOUR DECISIONS

MUKESH JAIN
Global Head of Innovation & Strategic Assets- 890 by Capgemini

Across all sectors, companies must transform and keep pace with digital innovation – using data to unlock greater efficiency and powering the business ideas that will define the future. Right now, only 39% of organizations are successfully turning data-driven insights into sustained competitive advantage. 43% are able to monetize their external data and insights through products and services. These figures will rise over the next 3-5 years, exacerbated by the upward trend of firms leveraging ecosystems. Now, more than ever, those tasked with driving their industry forward, need to look beyond internal data and embrace the trend of leveraging external data, no matter how uncomfortable that may feel.

This Everest Group report examines the state of external data adoption across a range of sectors, the booming ecosystem network, and how organizations need to reimagine data sourcing and consumption.

At Capgemini, we've developed 890 by Capgemini which is a plug and play technology, available on any cloud, and ready to simplify data sharing across the enterprise. Trusted, robust and curated to your needs – 890 by Capgemini gives you access to the data that's right for you, quickly and easily. It combines an extensive ecosystem of industry-leading, open and exclusive sources, with your own data.

This means that you can hone specific insights that will allow you to make key decisions with confidence.
If you would like to continue the discussion, please reach out to mukesh.jain@capgemini.com.

Everest Group®

Reimagining Data Sourcing and Consumption to Maximize Business Value

Mapping the Data Opportunity Beyond Enterprise Walls

Vishal Gupta, Vice President

Contents

Introduction

Data is accepted as a business-critical asset and the practice of data-driven insights is ubiquitous across enterprises, irrespective of size, scale, industry, and geography. Today's organizations increasingly recognize the value that data investments can return, and are, therefore, more open to sourcing/buying data from external data providers. As enterprises increasingly operate in networks (integration across stakeholders along the value chain) and ecosystems (collaboration and data sharing with competitors and other industry participants) the need to look beyond internal data will only rise. The trend is evident across industries – while Data and Analytics (D&A) front runners such as the Banking, Financial Services, and Insurance (BFSI) and retail industries are expanding the number of ways they use external data, adoption among other industries such as healthcare, life sciences, and manufacturing is also rising.

External data maximizes business value by producing better insights; it enables enterprise stakeholders to drive better data-driven decisions related to evolving market dynamics, especially in instances in which historical data might have limited use. In line with increasing external data adoption, the supply ecosystem – both data providers and tech enablers – has expanded to capture the opportunity. Though this booming supply-side ecosystem is simplifying access to external data (both in terms of data marketplaces simplifying access to data from the expanding provider ecosystem and tech enablers simplifying data sharing across enterprise stakeholders), enterprises still face significant consumption challenges. These challenges primarily result from a mismatch between enterprise data sourcing and consumption practices to the current scale and speed of external data consumption.

In this report, we look at the current state of external data adoption, the booming supply-side ecosystem, and – as the future looks promising for increased adoption –how enterprises need to reimagine data **sourcing** and consumption. Specifically, we address:

- The role of external data in enterprise analytics and penetration across industries
- Key drivers of external data adoption
- The booming supply side and the rich ecosystem of external data providers
- Adoption challenges
- Enterprise considerations to maximize value from external data investments

The role of external data in enterprise analytics

While data is universally accepted as a key business asset, data management practices across companies vary widely based on the extent to which they are data driven. Today, most companies leverage internal data such as sales processes data, employee information, and financial information to drive decision-making. Blending external and internal data allows enterprises to remain aware of emerging market trends and derive accurate insights to drive better decision-making. This helps enterprises maximize value from data investments and gain direct business outcomes . While the practice of augmenting internal data with data from external sources (any data that is sourced from outside the enterprise's processes and systems) is not new to customer-centric industries such as banking and retail, it is expanding into other industries and to use cases beyond customer and marketing analytics, as Exhibit 1 illustrates.

EXHIBIT 1

Penetration of external data adoption across industries

Source: Everest Group (2022)

External data adoption maturity: Low → High

Industry	Example of external data types	
Banking	Financial markets data to identify trading opportunities and make projections; sourced from providers such as Bloomberg and Factset	■■■■□
Retail	Pricing and promotion data to track the market and competitors; sourced from vendors such as IRI and GFK	■■■■□
Consumer Goods	Point of Sale (POS) data from retailers to analyze shopper buy rate, brand loyalty, payment preference, and similar insights; sourced from players such as Nielsen and IRI	■■■□□
Telecom and Media	Network coverage data is an essential data type used in infrastructure planning and addressing network strength; sourced from vendors such as S&P Global and GSMA	■■□□□
Healthcare	Healthcare providers use de-identified patient data, sourced from vendors such as IBM and Definitive Healthcare, to offer personalized treatments or patient-centric care	■■□□□
Life sciences	Medical product developers use Real World Data (RWD) to support clinical trial designs to generate innovative treatment approaches; sourced from vendors such as Parexel	■■□□□
Energy	Energy plants source regulatory and plant schedule data from data providers such as S&P Global	■□□□□
Utilities	Large enterprises in utilities and power generation source site assessment data from providers such as Enerdata and IHS Markit	■□□□□

Key drivers of external data adoption

Augment internal data to create a 360-degree view

While enterprises generate a lot of data as part of operations, some of the critical data related to suppliers, competitors, or customers (for example) must be sourced externally to form a holistic view.

Use case: customer targeting

When a customer adds an item to their shopping cart on an e-tail website and leaves the site without purchasing the item, e-tailers typically target the customer with relevant ads on the sites the customer frequently visits. This process requires integration of data that the e-tail firms possess (such as email ID, phone number, and IP address) which is then mapped to third-party advertising data (sites the customer frequents).

Account for factors outside enterprise operations

External factors such as weather data, geographical trends, local crime rates, and macroeconomic changes can affect business operations. These data sets need to be externally sourced to ensure robust operations.

Use case: claims management

In countries with severe weather incidents that generally cause damage to consumers' assets such as vehicles and property, insurers deliver alerts on potential bad weather to reduce claims by helping customers protect their property.

Enhance model accuracy

There is a limit to how much useful data an organization can generate. In such cases, to create robust models, blending internal data with external sources can create rich data sets that can improve the accuracy of machine learning algorithms.

Use case: look-alike modelling

An enterprise that has limited customer data internally to predict customer behavior can source data about other similar customers based on customer attributes such as age, affluence, and interests. This blended data can be used to predict behavior more accurately (such as what products a segment of customer is most likely to buy).

Prepare for agility during market disruption

The COVID-19 crisis was a key driver of external data adoption. In only a few months there was significant change in consumer preferences and the use of digital technologies, among other things. Preexisting internal data on customers, operations, etc. proved to be of limited use. Practices such as data sharing to accelerate learning is gaining adoption and acceptance.

Use case: data sharing for clinical trials

The pandemic increased the use of anonymized trial data shared by public organizations such as the World Health Organization (WHO) and pharmaceutical companies to accelerate the drug discovery process and contain costs.

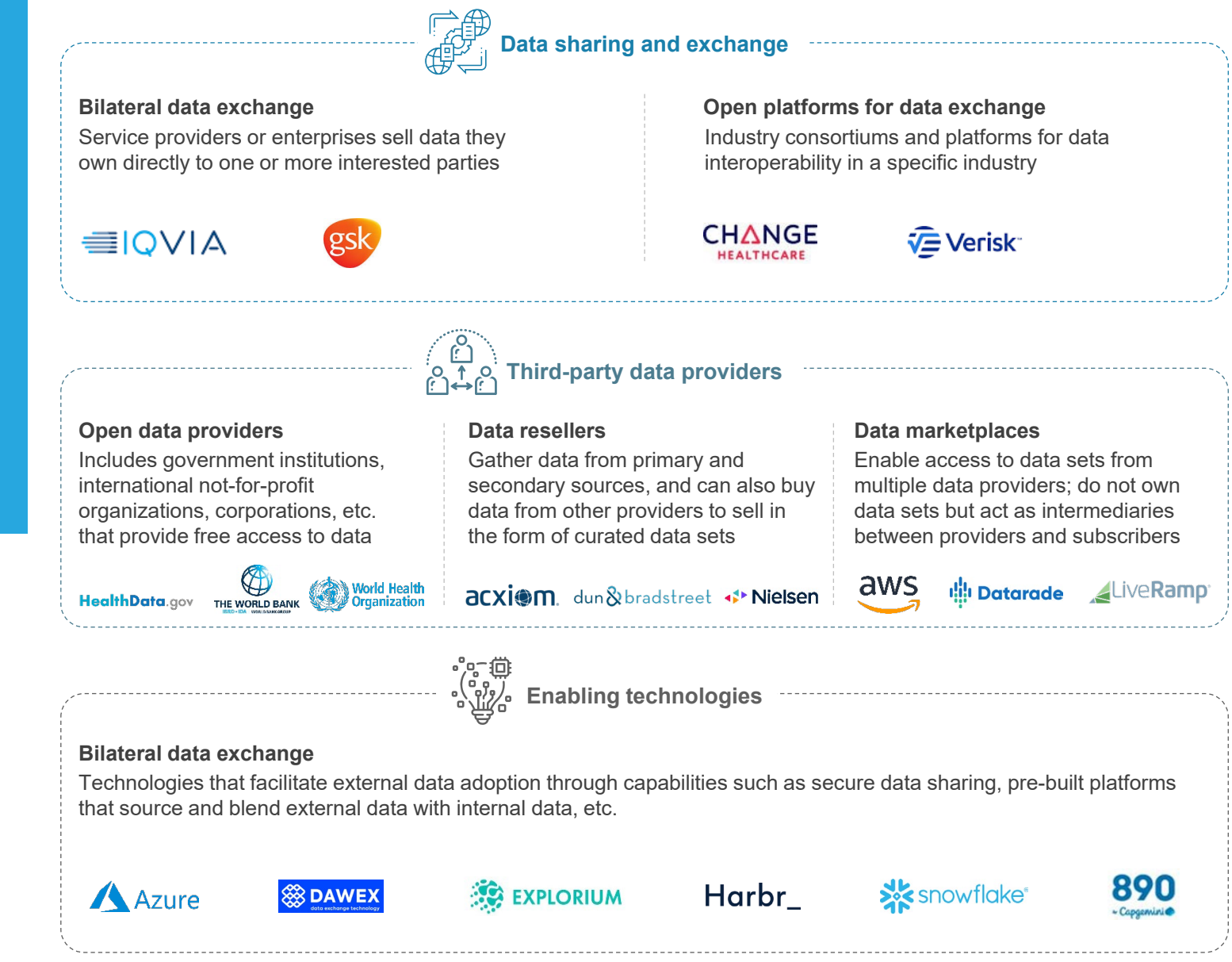
Booming supply side: rich ecosystem of external data providers

Traditionally, data resellers/aggregators such as Nielsen and Acxiom have been the primary providers of external data. As demand for external data has risen, the external data provider ecosystem has expanded exponentially. For instance, in the past a banking firm would have sourced data from resellers such as Blomberg, Refinitiv, and S&P Global. With the phenomenal growth of external data adoption over the last decade, the data provider ecosystem has expanded to include other types of business models with propositions such as open access to data (open data providers) and ease of data discoverability and sourcing (data marketplaces). Exhibit 2 depicts the external data ecosystem.

EXHIBIT 2

External data ecosystem

Source: Everest Group (2022)



Further, the growth of the digital economy has paved the way for data sharing among industry value chain participants to provide inter-organizational benefits. Though there is a long way to go before enterprises fully adopt secure data sharing, unforeseen situations like the COVID-19 pandemic have demonstrated the leverage an ecosystem-based approach can provide to effectively innovate in times of crisis. For example, during the pandemic, data access and sharing among healthcare ecosystem participants proved to be paramount to understanding the effect of comorbidities, optimization of treatment in various comorbidity settings, and even for public health surveillance.

While the pandemic has demonstrated the benefits of external data adoption and data sharing, emerging technology solutions that address concerns such as handling huge volumes of ecosystem data, secure sharing, and cataloging for ease of access are also steps in the right direction that can lead to increased adoption of external data. For example, Snowflake's data exchange solution helps organizations securely share data within the organizations or with external stakeholders such as suppliers and customers or even with other enterprises. Dawex, a data exchange platform, helps enterprises build their own data marketplaces to circulate accessible data with both internal and external participants considering security, compliance, and traceability of transactions. With such advances, we expect enterprise trust in data sharing to increase, further increasing adoption of external data initiatives.

Adoption challenges persist despite the positive outlook and rich supply ecosystem

While adoption of external data is on the rise, and there is increasing recognition of the value of the superior insights it delivers, adoption challenges persist. Even modern solutions like data marketplaces solve only part of the puzzle – they provide a platform that eases data purchase from multiple vendors but do not resolve bigger issues related to the types of data sets enterprises need, the impact these data sets can deliver, and the quality of available data, all of which create major roadblocks in external data consumption Exhibit 3 shows that, while a significant proportion of organizations engage external data vendors, less than a quarter are confident that they are likely to generate insights from external data sets in the next one to three years.

While a significant proportion of organizations engage external data vendors, less than a quarter are confident that they are likely to generate insights from external data sets in the next one to three years.

EXHIBIT 3

State of external data adoption

Source: Everest Group (2022)

81% of enterprises engage external data vendors

Yet only **27%** are confident they will be able to generate insights from external data sets in the next one to three years

Below we describe the five key challenges enterprises face in adopting external data.

Partial view into enterprise data gaps: Data siloes within the enterprise, in addition to the absence of an industry data blueprint to understand key data types by use case, make it difficult for enterprises to identify their data gaps. Further, a disjointed approach to Artificial Intelligence / Machine Learning (AI/ML) initiatives leads to duplicate purchases and deprives enterprises of a view into the external data sets that have the highest potential across the organization

Sourcing challenges: To start with, selecting the right vendor from an already complex and expanding external data provider ecosystem is daunting. Often enterprises work with more than one vendor to procure the necessary data sets. Different vendors have different approaches to pricing and data tagging and labelling. Further, each type of data set might need a different refresh frequency – for example, a data set used to determine real-time pricing may need to be refreshed every half hour, while a data set being used to estimate market size may need to be refreshed only once every year. These data differences result in different terms with different vendors for different data sets, all of which adds to the administrative burden

Governance, security, and compliance: Enterprises struggle to validate the quality of data provided by multiple vendors. Challenges also exist in defining user-based access and control over the data sets sourced. Further, with stringent data privacy and personal data protection regulations, including local laws such as the EU's General Data Protection Regulation (GDPR) and the UAE's Data Protection Regulations (DPR), enterprises need to be cautious regarding vendor compliance

Consumption and RoI: Often enterprises do not have a single interface to readily refer to data sets sourced externally. Users find it difficult to locate the right data sets leading to either duplicate purchases or low adoption of existing data sets. Either of these scenarios results in low RoI from external data spend

Homogenizing external data with internal data: While procuring the right data sets is an ordeal, the challenges do not end there – overlaying this external data with internal data to glean insights is also cumbersome. There can be practical challenges across the different phases of data management such as integration and preparation. For example, a retail enterprise plans to leverage external data for competitor analysis. It has existing data sets and procures additional data from two vendors. Vendor 1 calls a retailer "Amazon.com," and Vendor 2 calls the same retailer "Amazon." While resolving this issue seems simple at a small scale, a data science team is required to resolve the issue when the number of such unique vendors goes to the thousands.

Another issue is granularity. Internal data sets can have level-2 data, while a particular external data set can have level-3 data. The buyer is paying for level-3 data but might not be able to use it since the data sets are not homogenous.

To effectively handle these challenges, enterprises need to rethink their approaches to external data sourcing and management.

Enterprise considerations to maximize value from external data investments

A future-looking approach to data sourcing and consumption needs to be more streamlined. The complexity of vendor management within a crowded and expanding supply side, compounded by internal data integration and consumption challenges, requires organizations to build competencies on several fronts – such as those listed below – to be able to optimize external data adoption.

- **Build an enterprise-wide external data strategy:** While enterprises are interested in sourcing more external data, they can only show value from these investments by generating incremental insights. These incremental insights support cost justifications and can be used to promote enterprise-wide adoption and consumption. An enterprise-wide strategy that identifies the most critical data sets that can deliver value across the business, how to source them, and which strategic partnerships make the most sense is paramount to achieving these incremental benefits. For example, you can build a blueprint of external data types that are relevant to your business and prioritize the data sets that are likely to have more impact and deliver the highest RoI. Based on the data requirements, you can then evaluate the type of sourcing partner that best meets your needs – individual vendors (such as, if you need only financial data to start, and most of these data sets are available with a limited number of data vendors) or data marketplaces
- **Establish a single interface for all external data needs:** Enterprises need to establish a single interface for supply and consumption to limit the chaos of the external data adoption process. We expect this kind of unique interface/platform to become a key component of the enterprise data stack as external data adoption continues to rise. There are several key features to consider as you build or buy this platform:
  - It should be a single interface to manage all vendor relationships – pricing, data sets included in the contract, refresh rate, etc.
  - It should monitor usage to track highly used data sets and identify teams / use cases with high adoption of external data

  - It should be a central repository for all external data sets sourced and act as a portal to raise requests for new data sets; it should be easy to use and access and include features such as cataloged data sets for easier search and download and prebuilt connectors to export data to popular destinations such as BI tools and other downstream business applications
  - It should manage governance and compliance such as role-based access, vendor monitoring, and due-diligence
- **Dedicate roles/teams to external data adoption:** Onboarding new data sets is an ongoing process and needs dedicated roles such as external data curator to serve as a link between business units for their data requirements and the data procurement team, ETL engineers to cleanse the data before it is overlayed with internal data, and ML engineers to match and enrich these huge data sets
- **Actively address data trust issues:** Internal data already suffers from quality concerns, lack of trust, and difficulty in finding the right data sets to address a problem; overlaying external data only exacerbates the problem. Actively addressing these challenges through actions such as those noted below, is critical to increasing effective use of external data:
  - Ensure that the platform includes features such as user endorsement for quality and RoI of data sets
  - Create a playbook of the most impactful use cases and incremental benefits generated across factors such as model accuracy and ultimate impact on business outcomes

Conclusion

As enterprises begin to realize the potential of augmenting internal data with external data, we expect the adoption of external data to rise moving forward. In response, the external data provider ecosystem has expanded beyond traditional providers with some vendors adopting new business models such as open access to data and data marketplaces. Despite the prospects of adoption being stronger than ever, enterprises face serious challenges in adopting external data. To combat these challenges effectively, enterprises need to reconsider their external data sourcing and consumption strategies. Enterprises that can effectively map these roadblocks and strategically invest in external data adoption initiatives will be able to optimize their journey to becoming data-driven organizations, unleashing the true potential and business value that data can deliver.

Everest Group®
With you on the journey

Everest Group is a research firm focused on strategic IT, business services, engineering services, and sourcing. Our research also covers the technologies that power those processes and functions and the related talent trends and strategies. Our clients include leading global companies, service and technology providers, and investors. Clients use our services to guide their journeys to maximize operational and financial performance, transform experiences, and realize high-impact business outcomes. Details and in-depth content are available at **www.everestgrp.com**.

This study was funded, in part, by Capgemini

For more information about Everest Group, please contact:

+1-214-451-3000
info@everestgrp.com

For more information about this topic please contact the author(s):

Vishal Gupta, Vice President
vishal.gupta@everestgrp.com