Data Sourcing: Turning Data into Dollars

Businesses spend billions of dollars every year on data-driven projects, typically to win or retain customers or otherwise improve performance. Many of these initiatives fail to achieve their objectives. Others fall short of their potential, often without the managers in charge even knowing why or how. Still others result in reputational, legal, or financial harm.

The modern data economy is driven by an explosion in data quantity, types, and sources, as well as methods for managing and operationalizing it. This offers businesses rich possibilities that were unthinkable in the past, and that are becoming more expansive by the month. Yet this new and ever-changing ecosystem is highly complex, presenting a vast array of considerations and a minefield of pitfalls and potential missed opportunities.

To use data to expand profits, businesses need to have a mastery of:

- What data is available
- Where to find it
- How to evaluate it and compare it to alternatives
- How to contract for data with proper protections and advantageous business terms
- How to avoid mistakes that cause expensive setbacks, legal challenges, and/or reputational harm

This document provides an overview of how Think Data Group believes companies should approach sourcing, evaluating, and contracting for data, as well as tips on hazards to avoid.

Understanding the Data Universe

One of the most common shortcomings of corporate data strategies is a failure to fully — or even adequately — capitalize on the data opportunity. Companies need to understand the range of data available to them, and they must grasp the associated cost, management, quality, and risk issues. Companies have access to three types of data, which can be deployed together:

- **First-party data** is collected when people interact with you through any proprietary channel, including stores, websites, call centers, events, and social media. First party data is the least expensive and has the fewest strings attached but is frequently insufficient on its own.
- **Second-party data** is essentially first-party data that a trusted partner shares with you. Common sources include suppliers, distributors, and firms with complimentary needs. While second-party data is generally free or low-cost, challenges include negotiating with partners and managing the data.
- **Third-party data** is purchased from the commercial market and is often the most useful source. There are a wide variety of data providers that offer very detailed information. Types include business firmographics (including business type, locations, employee counts, and sales) and consumer demographics (age, gender, education level). A wide and growing range of sophisticated, higher value data is also available.

Higher value third-party data includes attitudinal, behavioral, transactional, and interest attributes, typically linked to individual firms or consumers. These can indicate buying "signals" useful in customer loyalty, retention, and marketing initiatives. These signals can be identified through geolocation data collected from mobile devices, attendance at shows or conferences, search activity, downloads, social media posts, and a wide variety of other detection methods. For instance, a fast-food company may market to a competitor's customers by purchasing geolocation data indicating who is frequenting the competitor's restaurants.

The following are examples of high-value third-party data:

- **Identity graphs** offer a single view of an individual across their devices and identifiers, such as phone numbers, email, and physical addresses. Some providers can associate a consumer to their professional life (B2C2B) to create even more advanced digital audiences
- **Specialty and vertical market-specific data** provide businesses with a deeper understanding of trends, behaviors, and external market factors that can inform business decisions
- **Job-listing and social-media data** can identify potential new employees in a tight labor market
- **Credit and other business-risk data** can help determine the health of supply chains
- **Weather data** is useful in forecasting demand changes or risks in producing or shipping goods
- **Satellite and drone imagery** can be analyzed to gain competitive intelligence or to address risks that could impact infrastructure

Regardless of your sector or geographic reach, the opportunities to gain an edge using such data are nearly unlimited. This is particularly true given the new platforms, collection methods, and emerging technologies that further enhance how data can inform your business decisions. You can gain powerful new levels of insight by using single or multiple data sources or data lakes to aggregate and integrate data from disparate firms and marketplaces. IoT, OTT/CTV, machine learning, AI, blockchain, NFT's, crypto currency, and other innovations are all relatively recent advancements that will continue to shape the data ecosystem.

See Think Data Group's proprietary Periodic Table of Data Elements to learn how we think about the data ecosystem and use it to help clients solve their most challenging business needs. We will soon be publishing a paper to describe in detail how you can use the Periodic Table of Data Elements to make more informed decisions about your data strategies and needs.

Evaluating Data

Once you clearly understand the available data types, the next step is to evaluate data providers and their assets to determine the best fit for your needs. This is a crucial step as the outcome of your efforts will depend on the quality and suitability of the data you use. We recommend a disciplined, thorough analysis of potential providers and their assets as detailed below.

First, you will need to clearly define, and document use cases and data specifications, to ensure you understand what you seek to accomplish, what data you will need, and how it will be used. Experience shows the entire data sourcing and evaluation process runs most efficiently when internal teams, decision makers, and potential data providers are aligned on use cases and data specifications.

The detailed specifications can be used to collect information from multiple sources to develop a deep understanding of the data that is available from each provider. We typically seek the following in a request for information (RFI) or request for proposal (RFP) format:

- Listing of each individual data attribute or a data dictionary
- File layout
- Volume (total number of records available)
- Fill rate (% of each data attribute available in the total number of records)
- Geographic coverage
- Frequency of update cycles
- Amount and length of time historic data is available, if required

In addition to information about the data itself, the evaluation process includes:

- Learning how the data is collected
- Ownership details
- Permissible usage rights and restrictions
- Compliance with privacy and legal requirements
- Data costs and licensing models

The next step is to narrow the list of potential providers and request sample data and records to further understand data structures and file layouts. Final pricing can also be requested, enabling you to move forward with one or more providers. For larger or more complex initiatives, the data's value may need to be further assessed before you license it or sign final contracts. In these cases, we recommend proof of concept studies using live data in real-world applications.

Contracting with Providers

There are several important and sequential contracting steps to consider throughout the data-sourcing process. Skipping any of these steps can lead to unnecessary legal complications, reputational harm, or worse.

Mutual Non-Disclosure Agreements

Early in the sourcing process, data buyers (licensees) and data providers (licensors) should establish a mutual non-disclosure agreement (MNDA) to protect the commercial and competitive interests of each party. Common terms contained in MNDAs cover the following:

- Description, allowable use, and protection of confidential information
- Disclosure limitations and requirements
- Term of the agreement
- Termination clause, including return or destruction of confidential information
- Other obligations and remedies
- Governing law

<u>Data Evaluation Agreements</u>

Evaluation or test agreements allow data buyers and providers to exchange data for the purpose of evaluating quality, performance, and the data transfer process. Most test agreements include many of the non-disclosure and legal terms and conditions described above. Additionally, test agreements should cover the terms and conditions of the actual data test, including but not limited to the following:

- Purpose of the test
- Data attributes to be included
- Length of time of the test
- Permissible use of results
- Test criteria – what a successful test looks like and anticipated next steps if successful

<u>Data Licensing Agreements</u>

Once you have decided to procure data, data license agreements must explicitly state all terms of the contractual relationship between the sellers and buyers. At a minimum, the terms should include the following:

- Description of the actual data attributes that are being purchased
- Time period the data covers, including update cycles if part of the purchase agreement
- Geography the data covers
- Format, mechanism, and timing of data delivery
- Length of time the data can be used
- Permissible uses of the data, enterprise level access
- Any restrictions on data usage
- Cost of the data
- Payment terms
- Responsibilities that each party has for all legal and privacy compliance issues related to data collection and usage
- Renewal options for multiple years

Hazards to Avoid

Throughout the data sourcing process there are several hazards to avoid ensuring a successful data sourcing outcome. Below is a list of some of the more common missteps that can lead to wasted time, effort, expense, and/or even more serious legal, privacy, and compliance issues:

- **Not having clearly defined business requirements or use cases** that are easily understood and can be communicated to internal teams, decision makers, and potential data providers. Without everyone having a clear understanding of what data is needed and how it will be used, the data sourcing effort can be easily derailed or create unnecessary friction.
- **Not having the necessary resources to ingest, analyze, and operationalize the data**. This includes technology systems and platforms, analytical capabilities to extract the most value from

the data, visualization and other tools to communicate findings, and governance processes to ensure proper data management and compliance practices.

- **Not having clearly understood and documented data usage terms.** We have seen product development and marketing communications initiatives fail late in the development process because of misunderstandings about how the data can be used to create derivative products or to communicate with desired marketing prospects.
- **Acquiring data that is not sourced from reliable or credible sources.** Data that is inaccurate, unreliable, unusable, or illegal can lead to poor business decisions and expensive mistakes.
- **Not understanding current privacy and compliance legislation or industry association guidelines and best practices.** There are a variety of laws and best practices that vary across states, countries, and industry associations that can invalidate a product or kill a launch or marketing program if not clearly understood. Anticipating potential future changes to legislation and best practices is also important to ensure the long-term viability of initiatives.
- **Acquiring, storing, and paying for data that is not needed** or does not contribute to desired business outcomes.
- **Acquiring the wrong data, overpaying for it, or improperly contracting and licensing** data due to lack of familiarity or expertise.

Summary and Key Takeaways

Everyone knows that data underpins the modern economy, but harnessing it successfully takes expertise. As we have described, there is a vast and expanding universe of data available to businesses, and capitalizing on data takes knowing:

- The business needs you are trying to solve for and the capabilities you have to apply data solutions to meet those needs
- What data is available
- Where to find it
- How to evaluate it and compare it to alternatives
- How to contract for data with proper protections and advantageous business terms
- How to avoid mistakes that can cause expensive setbacks, legal challenges, and/or reputational harm

Think Data Group is a boutique consulting firm composed of leading veterans of the data and information-services industry. We offer clients state-of-the-art consulting and implementation support – a skill set that few companies develop in-house and few competitors can match. Whether you are a start-up or a billion-dollar enterprise, we can help you maximize the value of your data and data initiatives.

We would love to start a data related conversation with you. Please email us at barry.gold@thinkdatagroup.com.