Alexis Olson, Bradley Lai, Karan Sandhu, Nolan Hartwick
MBAi 417
25 January 2021

Reflection Assignment: Data Exchanges and Alternative Data

Data Exchanges

Modern organizations are increasingly turning toward data as the fuel that drives their innovation. As more companies 'go digital' as a means to optimize decision-making and improve operational efficiencies, the demand for data grows exponentially. Most firms already have databases where financial, sales, inventory, HR info, and other required business information is stored. Many of these databases, however, are separate from one another so firms are increasingly trying to pull this information into one accessible data storage location. In addition to internal business data, firms are looking towards other sources of data that they can acquire to drive further business insights.

While data is critical to the digitization and advancement of firms, the quality of this data is critical for reliable outputs. Data quality can vary significantly and firms should increasingly evaluate the quality of their data. Data quality is measured on accuracy, completeness, the time to acquire, consistency, and uniqueness. A 2016 study from IBM states that poor data quality strips about $3.1 trillion from the US economy annually. Firms should aim to have a good understanding of their data quality so that they can increase their reliability in the outputs of the data and, therefore, better inform their decision making. Evaluation of data quality is especially important as firms turn to new data sources such as data exchanges.

Unfortunately, the volume of an individual company's data set is often not enough to create a complete picture of their customers. While Google and Amazon have clear pluralities in terms of search data volume they still lack information, such as details on potential future customers. Therefore, in order to develop a more coherent picture of the entire ecosystem, whether profiling customers or monitoring complex systems with sensors, there is value in knitting together individual sources. A whole economy of cloud storage and cloud storage services is built around this problem. This quest for a complete "quilt" of data has led to the development of Data Exchanges or marketplaces.

The diagram below shows the potential transactions of a data exchange. In this scenario, "Your Organization" (let's say Amazon.com) is the primary source of data and is opting to sell data to Customers A, B, and C. In this case Amazon.com may be selling data to small business that sell on their platform (Customers A and B) as well as allowing it to be accessed internally by their customer service group (Customer C). The data is shared through the use of a third party tool, such as Snowflake as represented in this diagram. Snowflake, Amazon Redshift, and Google BigQuery are all tools that enable this type of data sharing - offering secure data transfer with clear publishing and access pathways. Additionally, as this is a market, these dataset are bought and sold, typically priced based on the quality of the data (as outlined above). Overall, Data Exchanges are markets that allow for the sale of valuable goods: datasets.

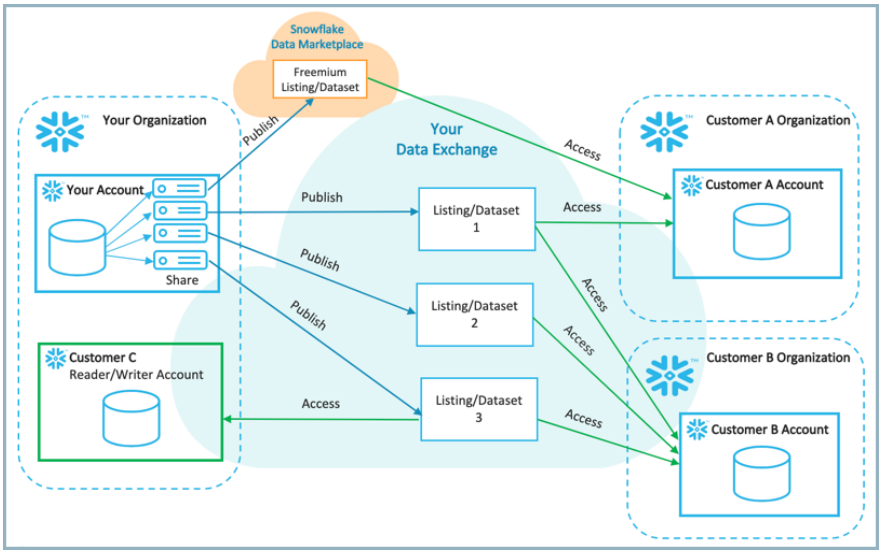

Fig 1. Securely Share Data with Customers. (Snowflake)

Alternative Data

Alternative Data is a term generally used in the finance industry by giants such as J.P. Morgan and Goldman Sachs to refer to data collected from non-traditional sources (Big Data in other industries). In the past traditional data sources such as SEC filings, stock prices etc. were the primary source of data but today the number of alternative data providers has grown exponentially as seen below.

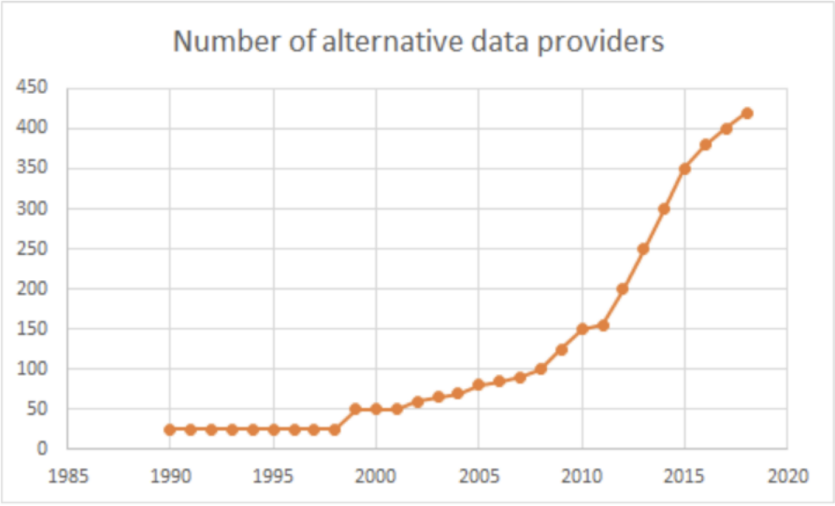

Fig 2. Number of Alternative Data Providers

As seen in the image below there are three main categories of Alternative Data:

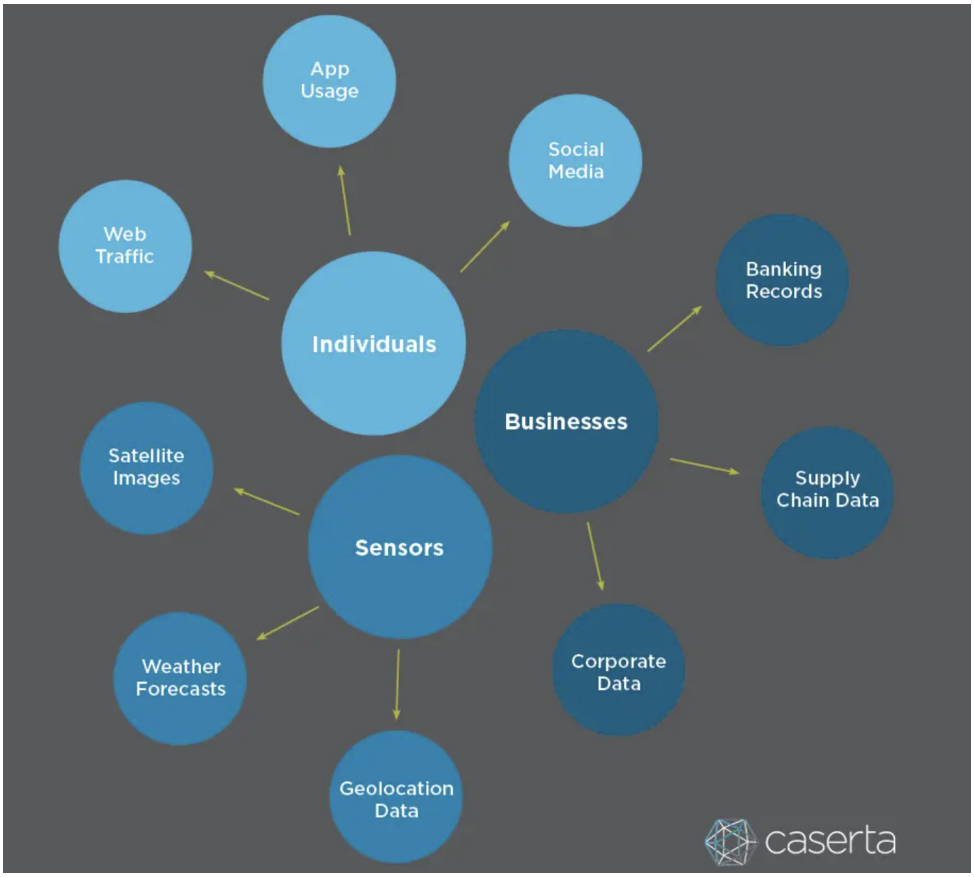

Figure 3: Types of Alternative Data

1. Individuals

A common source of Alternative Data is from individuals. This data is usually unstructured due to the various sources and can be challenging to process due to the large amount of data generated every minute. Data collected from web traffic, apps, and social media posts can be an important indicator of customer sentiment and can be used to derive market insights.

2. Business Processes

Data from business processes, also known as “exhaust data” since it is a by-product of business processes, is an important source of data and is often structured. Some common courses of business processes data are credit card transactions, sales transactions, commercial transactions, and supply chain data.

3. Sensors

Another source of alternative data is sensors. The explosion of deployed Edge and IoT sensors has unlocked a ripe new era for multiple industries. Now, sensors capturing data from previously inaccessible or cost preventative locations allow operators to make smart business

decisions. These include use cases in retail, oil & gas, and finance, to name a few. For example in retail, bluefox.io is helping customers answer questions such as “who is the most profitable customer?” In industrial applications, HMS Networks is developing solutions which answer questions such as “where do we see the most machine failures?” Finally, companies such as IntelligentVoice are able to help call center analysts detect fraud through real-time voice analysis. These examples follow a broader trend in the rise of alternative data, which are disrupting how financial markets are examining and pricing assets. As sensors continue to proliferate, we should expect to see more use cases become unlocked. Solutions, such as NVIDIA’s Jetson system, allow for powerful computing to occur at the edge in a small form-factor.

As companies turn to making more of their decisions to be data-driven, it is not enough to simply have the ability to analyze the existing data. Further efforts must be made to collect larger quantities of meaningful data, and transmit & store the data effectively. These underlying systems supporting data collection and transmission are just as important as the data itself. Ultimately, the collection of systems and data can result in businesses making smarter decisions based on analysis of cleaner, more relevant, and more timely data.