

# How Data Creates Wealth

# Published on Published on March 6, 2018

## Dr. Ted Gaubert, PhD, MBA

In just over a decade, technology companies have displaced oil companies as some of the most valuable companies in the world. Amazon, Google, Apple, Facebook and Microsoft have grown into seemingly unstoppable titans because of the data they have and the way they use it.

The common cry of "data is the new oil" compares data to the natural resource that has created international tensions and wars over the past century. While the data/oil analogy isn't perfect, it is easy to relate to. To analyze the wealth creation process in the data economy, it is necessary to look at both the up-stream and down-stream processes of the data value creation chain.

### Up-stream Processes - Oil/Data Extraction

With the discovery of a new oilfield, a small economically disadvantaged town can quickly be transformed into a wealthy oil boomtown as long as the oil continues to flow from the well. However, history has shown that when the wells dry up, workers lose their jobs and the previously prosperous oil town turns into a ghost town.

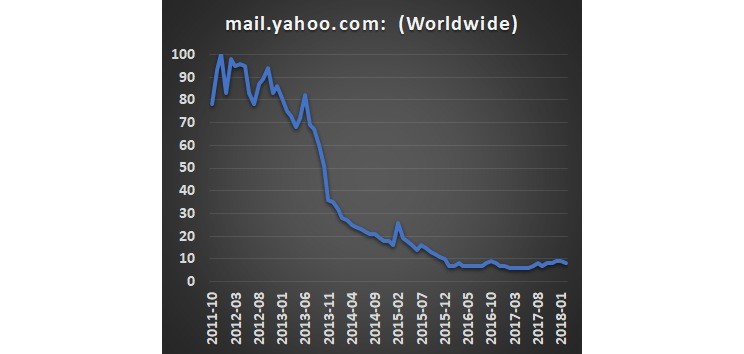
The wealth of a data company comes from the continuous extraction of data. Consumer trust in the company, product or services is a big enabler for companies to continually acquire new data. Facebook is trusted with data about our social lives. Google is trusted with our internet search and browsing history. Amazon is trusted with our purchase history, viewing patterns and shopping habits. Our trust has become a corporate asset for extracting new data.

 “Our trust has become a corporate asset for extracting new data”

Millions of internet users trust internet-based companies with private and confidential data. Until they don’t.

Trust takes years to build but only seconds to break. The sudden loss of user trust has been responsible for many internet 'ghost towns'. This is when users abruptly abandon the service or product.

In 2013, Yahoo experienced a data breach that exposed all of its 3 billion user accounts. The high-profile breach was a significant blow to user trust of their mail service. Subsequently, Yahoo's mail service experienced sharp decline in usage and never recovered.



The rapid decline of Yahoo's mail service highlights two important aspects of the data value creation chain:

1. **Criticality of continuous data extraction to create down-stream value**

With the sudden decline of Yahoo's Mail service use, Yahoo lost the ability to read new users emails and perform user profiling. This means they gradually reduced their ability to drive targeted advertising and revenue.

2. **Data has a shelf-life whose value declines over time**

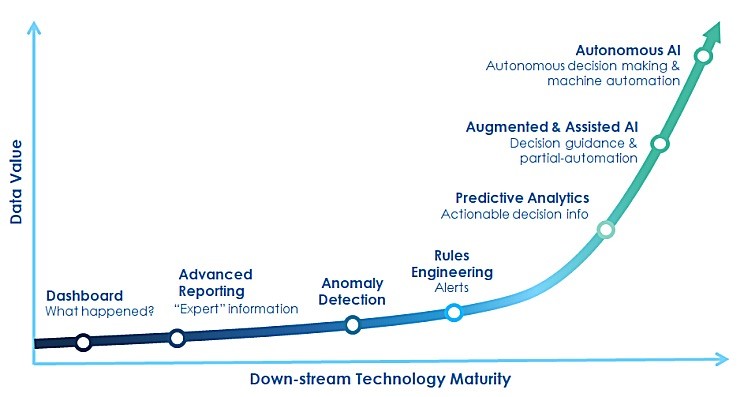
For a user with an active email account from 2003 - 2013, Yahoo had a decade worth of sent and received email. This is a significant amount of user data and could be used to build a reasonably accurate profile of the person's interests during this time period. In 2018 this dataset is far less accurate than what it was in 2013 in respect to representing the user's interest. In most cases, data has a shelf-life where the value of data set decreases over time.

### ****Down-stream Processing****

Like oil, data has “intrinsic value”. Intrinsic value is the estimation of potential economic value that would be created if either were transformed into something that creates value in the marketplace. In the oil industry, down-stream processing and refinement are the processes that transform oil into various products like gasoline, plastics, asphalt, and lubricants.

Over the past century, down-stream oil processing technologies have improved and enabled more products to be produced from a single barrel of oil at greater efficiencies. These technology advances have thus increased the intrinsic value of a single barrel of oil.

Data goes through a similar process of being refined and turned into useful products or services that increase revenue. As data related technologies have improved, the intrinsic value of data stores has also increased. Recently, advances in artificial intelligence technologies have helped to generate more value from data than could be in the past.



Problems involving large data-sets have typically had an easier time moving further "right" on the curve above. This is because machine learning and artificial intelligence tend to perform well with large datasets.

However, much of today's enterprise labeled and structured data is “small-to-medium data” and lives in an ERP system or data warehouse based on SAP, Oracle or Microsoft SQL server. Often, traditional machine learning methods can perform poorly on small or sparse datasets in comparison to much larger data sets. This has posed a challenge in pushing "small data" problems further to the "right" on the curve above. Simply obtaining additional data is often impractical or cost prohibitive.

One area I work on is developing next generation AI technologies that can extract greater value from small datasets. This is exciting because it makes it possible to solve problems that couldn’t adequately be solved previously. Ultimately, the goal of the advancements is to move "small-to-medium data" problems further "right" on the curve above.

### ****Conclusions****

There is more to the data wealth building story than amassing a huge data store. It also involves both the up-stream and down-stream components to truly unlock the most value from the data. The advancements in down-stream technologies, like AI, are allowing companies to unlock more value from their data, much like advancements in oil refining technology. We’re only at the beginning of the AI era, and as a result there remain vast opportunities for wealth creation from small, medium and big data.