

# Why are data hard to value? Data's unique attributes

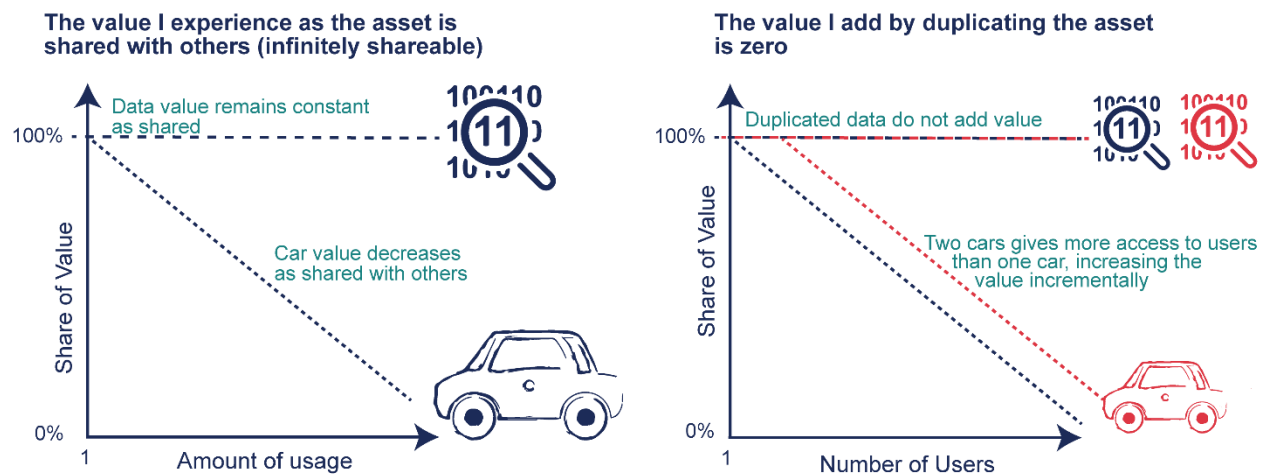
Last Updated November 26, 2018

*What is the value of my data? What is the value of sharing my data with others? These seem like simple questions, but data are not simple assets. The unique attributes of data making them more difficult to value than traditional assets.*

Most organizations need to demonstrate a return on investment (ROI) when undertaking new initiatives. There are many methods available to estimate the ROI for traditional assets; however, data do not behave like traditional assets. For data, costs tend to be easy to estimate while the benefits are complicated. To understand why these benefits are difficult to estimate, let's compare data with a common asset, such as a car.

## (1) Finite versus infinite users

Cars are finite. If I lend you my car 25% of the time, then I will only be able to use my car 75% of the time. In contrast, **data are infinitely shareable** (a non-rival good). You and I can easily use the same data at the same time. This means that the value of data is cumulative between individuals, rather than proportioned amongst individuals (Figure 1). Stu Hamilton (2014) proposed that the value of data is a function of its quality and usage:  $\text{Value} = (\text{Data} * \text{Quality})^{\text{Sharing}}$ . This suggests the value of data increases exponentially with usage. Whereas with cars, the value is proportioned among its users. This means that if I had two identical cars, you and I could each use a car 100% of the time. The same is not true for data. **The value of redundant, duplicated data is zero** (Figure 1). Indeed, data that are replicated many times complicate ownership and add cost to gather the data, store it, and reconcile multiple versions of 'truth'.



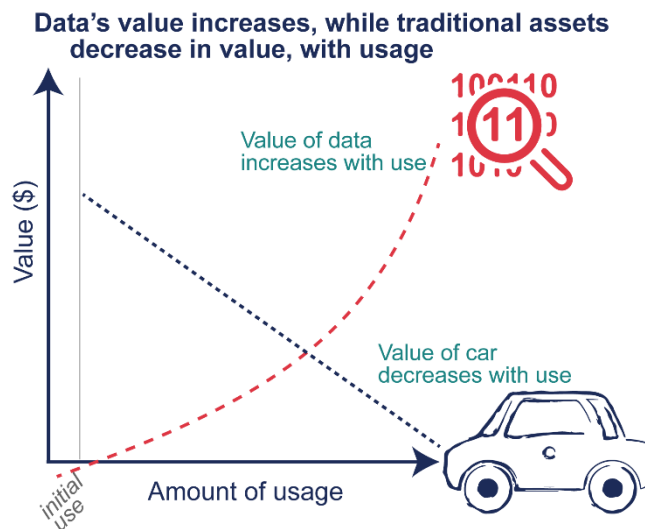
**Figure 1:** (Left) The full value of data can be realized by each user while only a portion of the value of a car is available to each user. (Right) Increasing the number of cars can increase usage, while duplicated data don't create additional value.

## (2) Amount of use matters

Most assets are not valuable until they are used. Both unused cars and unused data are liabilities. The upfront capital investments have already been made and the only way to regain those costs is to use the resource. In most cases, once a resource is put to use it, its value begins to decrease with wear and tear. For a car, the number of miles driven will impact how frequently maintenance is required. But, **the more data are used, the greater their value** (Figure 2). Data have the highest potential value when all who want to use the data have access and know how to put it to use. Data may not be used if people:

- ◆ do not know data exist,
- ◆ cannot discover data easily,
- ◆ do not have access, or
- ◆ do not have the information (metadata) or knowledge to put the data to appropriate use.

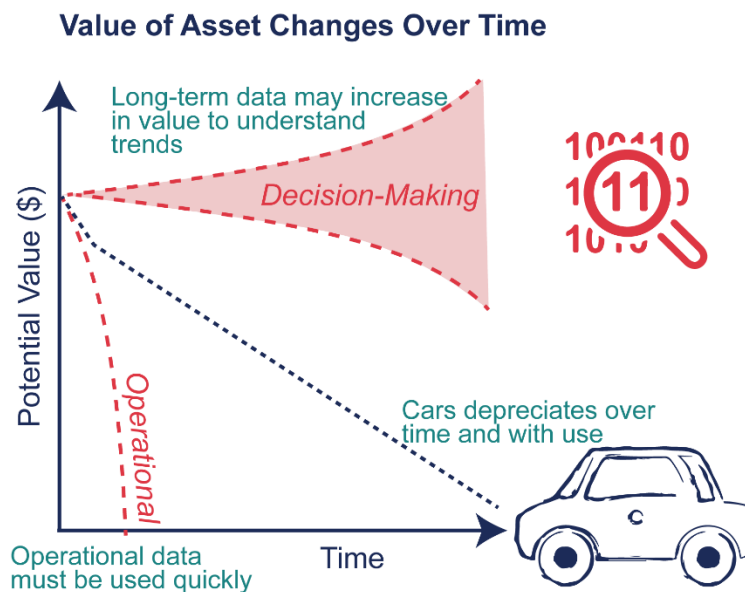
Value creation requires the data producer to make the data and metadata available and the data user to have good data literacy.



**Figure 2:** The value of a car decreases with usage while the value of data increases with usage.

### (3) Usable lifespan varies

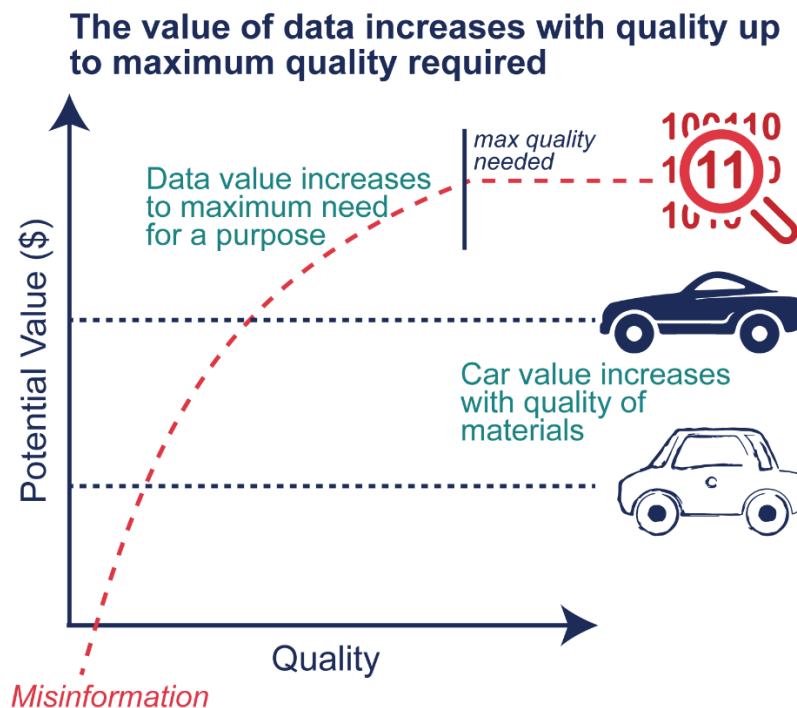
Most assets depreciate over time. The value of a car decreases the moment it is driven off the lot. Assuming the value of the car depreciates by 10% each year, the car will be worth 41% of its initial value by year five. In terms of data, *the speed at which the value of the data decreases depends on the purpose of the data* (Figure 3). Data used to support day-to-day operations must be put to use quickly and their value decreases instantly (many organizations discard operational data once their capacity to inform operations has passed). In contrast, long-term records used to understand broad trends and support decision-making may increase in value over time ([knowing how streamflow has fluctuated over the past 30 years provides more insight than 10 years of data](#)). Typically, it is easier to assess the value of operational data since their use are directly tied to an outcome. Assessing the value of data for decision-making, regulatory purposes, or research is often challenging and ambiguous.



**Figure 3:** The value of a car depreciates with time. Data's value over time depends on their purpose.

#### (4) Higher quality, higher value

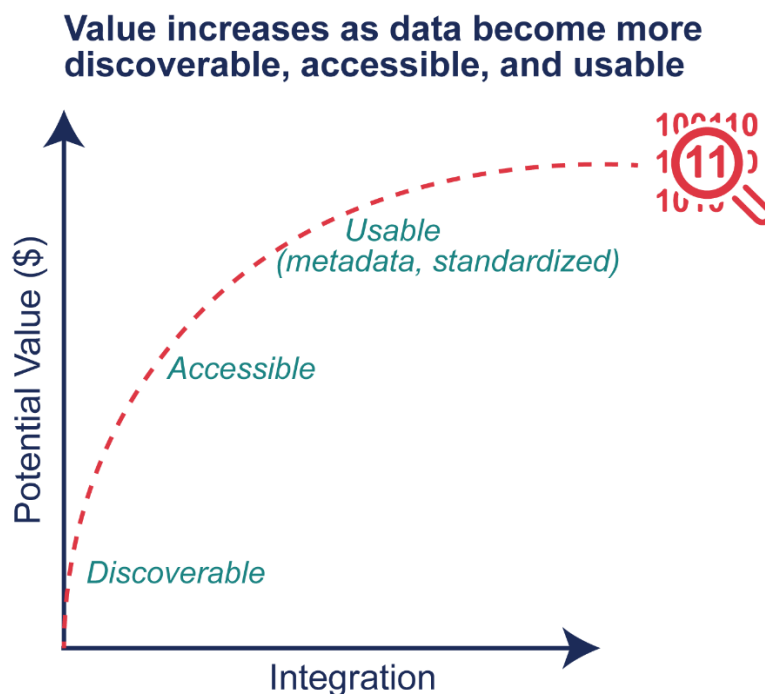
Higher quality assets are worth more. The price for a basic Honda Civic is \$19,000 while a Honda Civic with all-wheel drive, navigation, and an anti-lock braking system costs \$28,000. Similarly, **higher quality data have the potential to create better information**. For instance, fifteen minute streamflow data provide better flood forecasts than daily data. And just as a poorly constructed car can be a dangerous liability, inaccurate data can be very costly if they lead to bad decisions. It is imperative for users to have a sense of the data's quality to know whether the data are fit for a particular purpose. Hamilton's equation  $\text{Value} = (\text{Data} * \text{Quality})^{\text{Sharing}}$  places quality as a multiplicative factor. The caveat is that data exceeding the necessary quality (or temporal or spatial resolution) of a specific purpose do not add value. For instance, 15 second water use data is irrelevant to a homeowner, who wants to see hourly or daily data to understand their water use (Figure 4).



**Figure 4:** Higher quality goods, whether cars or data, have greater value.

## (5) Combining assets creates greater value

Combining assets can lead to greater value. A car is useful, but a car with gas is even better. The same is true for data. **Data can become more valuable when combined with other data to produce new information and insights.** For example, knowing a stream’s water quality is useful, but knowing the water quality in combination with intake locations and health standards provides context and greater insights. In order to combine data, the data must be discoverable, accessible, and usable (Figure 5). Today, organizations spend much of their time finding, cleaning, and reconciling water data, with little time to spend on creating value. Another unique attribute of data is that combining data can create new data (derived products). The volume of data, and associated management challenges, can grow quickly.

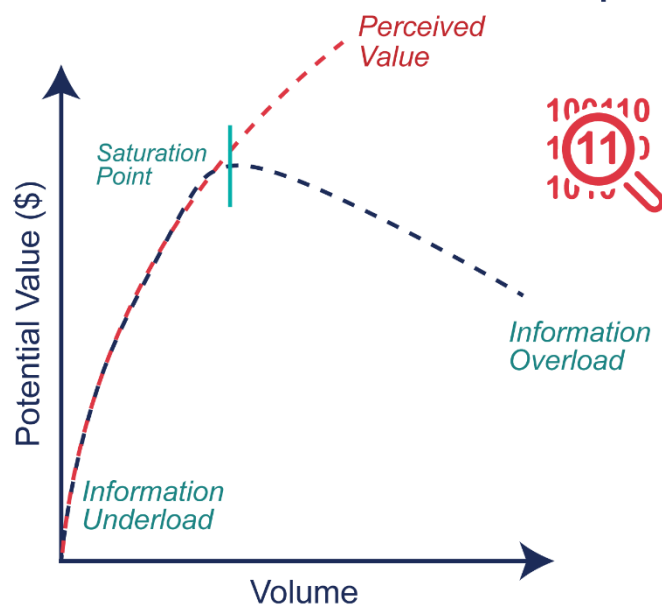


**Figure 5:** The value of data increases as the data are more discoverable, accessible, and usable.

## (6) More is not necessarily better

Most individuals have the perception that more is better: more money, more time, more cars. And in some cases, this holds true. A family of four driving adults would likely benefit by having more than one car to increase flexibility and capacity. But 20 cars would probably create more problems than help. Similarly, as the amount of data available increases, it can become overwhelming and result in analysis paralysis. Here, **decision-making performance decreases once the amount of data exceeds the capacity for individuals to digest** (Figure 6). All data are not created equal and it is important to prioritize the data to avoid overload.

**Value of data increases to a saturation point**



**Figure 6:** Data and information overload can lead to design and expertise gaps ('analysis paralysis').

**For more information:**

Blue Ribbon Task Force. 2010. [Sustainable Economics for a Digital Planet: Ensuring Long-term Access to Digital Information](#).

Hamilton, Stu. 2014. [Shared Data, Complete with Discoverable Provenance, Have Unbounded Value](#).

Moody and Walsh. 1999. [Measuring the Value of Information: An Asset Valuation Approach](#). European Conference on Information Systems.