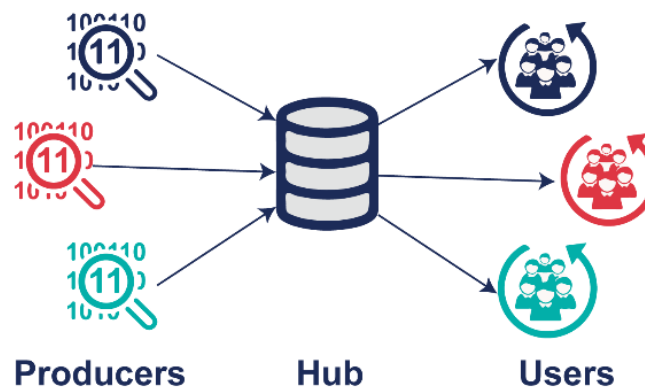


# Consumption-Based method

Last Updated November 26, 2018

*Data hubs are about sharing data and may not know how those data are being put to use. Instead hubs track the number of downloads and unique users for each dataset. A consumption-based approach to the Modified Historic Cost method can be used to assess the value of data hubs, with the underlying assumption that more data downloads is equated with more data usage and greater value.*

The Consumption-Based method is the [Modified Historic Cost method](#) adjusted to estimate the value of data hubs. This method assumes hubs receive data from different data producers and in turn share the data to multiple users (Figure 1). The levels of effort and cost to a hub to take in data varies across producers. Similarly, the hubs data will be accessed with different frequency by different users.



**Figure 1:** Producers push data to the hub while users pull data from the hub.

## Consumption-Based method

The Consumption-Based method, sets the initial value of the hub equal to its costs and then refines this number by adjusting for data usage statistics and data attributes.

### (1) Identify the cost of the hub

Data hubs may use their annual expenditures as the cost or they may track costs by dataset or by individual data producers.

## (2) Modify costs using data usage statistics

The starting value of the hub is then adjusted by data usage statistics. The value for individual datasets can be weighted at zero (no usage) to one (maximum usage) based on either the number of downloads or the percent of each dataset accessed by data users. This means the value of the data are relative to their use within the hub. Since the benefits are realized by data users, the [benefit-to-cost ratio](#) is used to compare costs and benefits (similar to a return on investment but does not assume the same organization accrues benefits and costs).

## (3) Modify costs using data attributes

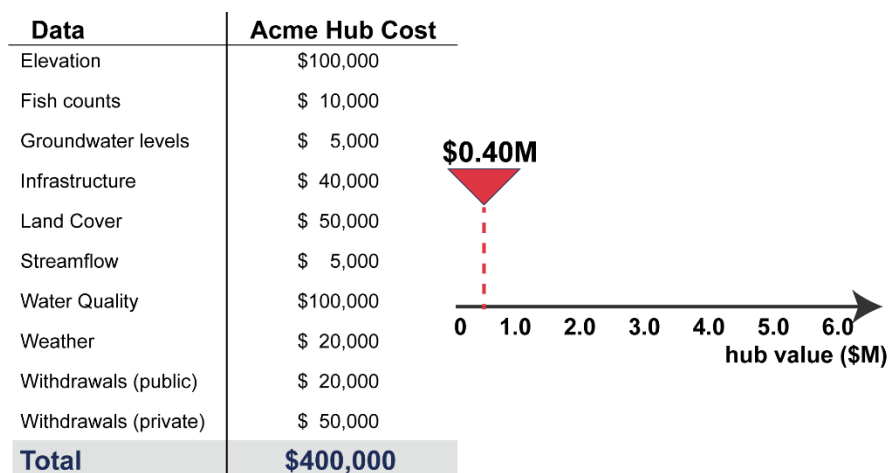
The value of the hub is then further adjusted by dataset attributes. For instance, redundant or duplicated data have no value. Data may be weighted by their quality from 0 (very poor) to 1 (excellent). For data of unknown quality, expert opinion may be used. Valuing datasets within the hub allows the hub to prioritize which datasets are of high value to their users and how to maximize the value of the hub.

## Example application

Acme Hub’s mission is to provide data for decisions related to water resources in the IoW watershed and aquifer system. Acme Hub has been online for a year and provides access to eight datasets collected from multiple data producers.

### (1) Identify the cost of the hub

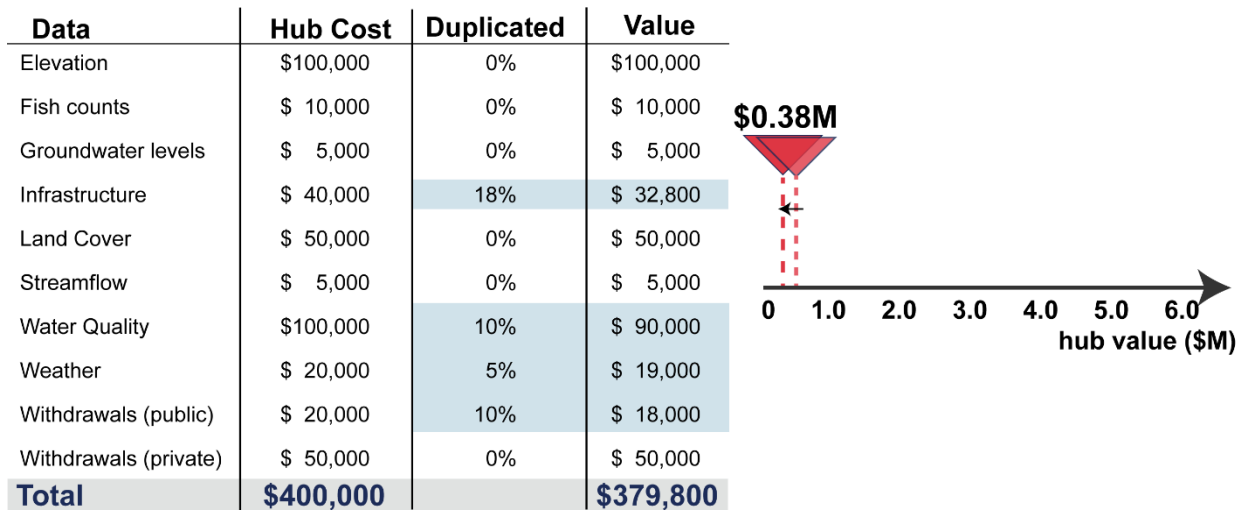
Acme Hub wants to track the costs by dataset (Figure 2). The total annual cost is \$400,000. Knowing the relative costs of each dataset enables the hub to prioritize investments based on data costs and usage.



**Figure 2:** Acme hub costs by dataset

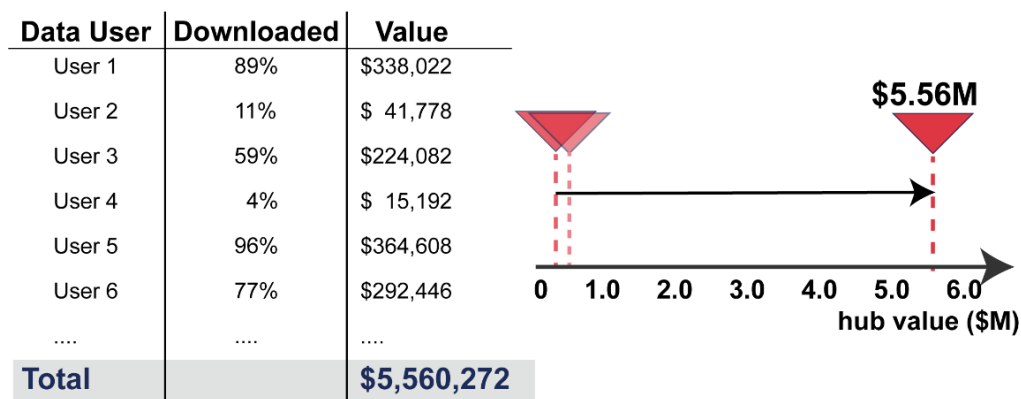
## (2) Modify costs using data usage statistics

Acme Hub tracks data on two fronts. For producers, they assess which data are duplicated within and between producers using unique identifiers. They found 5% of weather, 10% of public withdrawals, 10% of water quality, and 18% of infrastructure data were duplicated. The duplicated data have zero value to the hub, reducing the hub's value to \$387,000 (Figure 3).



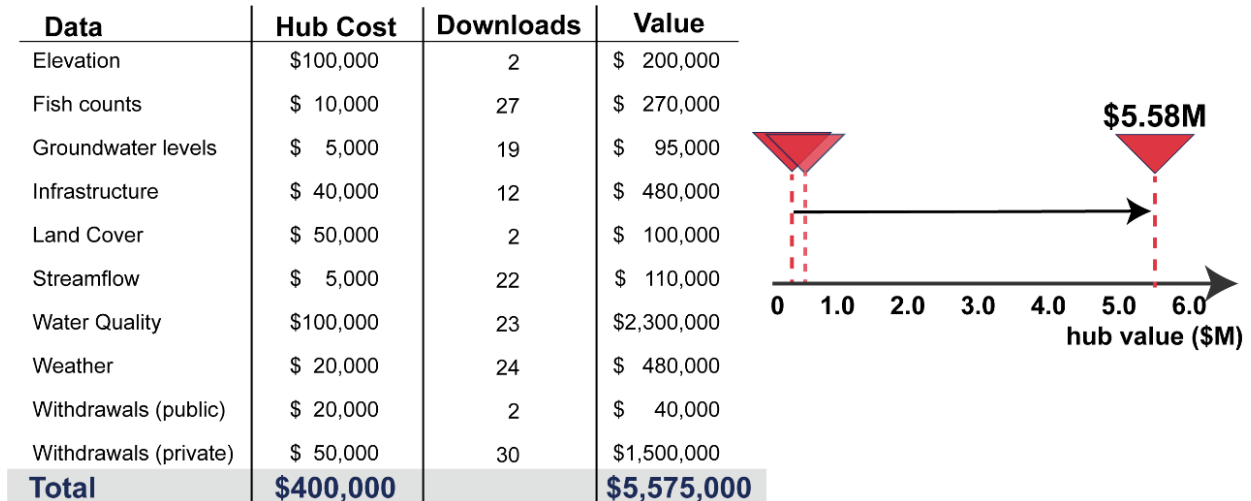
**Figure 3:** Acme Hub costs are adjusted for duplicated data.

For users, Acme Hub tracks the number of unique users (n=30) accessing data. Acme Hub explored two approaches to adjusting hub value with data usage statistics: by percent downloaded by user or number of downloads. When Acme Hub looked at the percentage of data downloaded by each of its 30 unique users, it found users downloaded between 0 to 96% of the hub's data (Figure 4). Multiplying the adjusted hub value (\$379,800) by the percent utilized for each user produces a hub value of \$5.56M with an estimated benefit-to-cost ratio of \$13.9 benefit for every \$1 spent.



**Figure 4:** Hub value based on percent of data downloaded by each user.

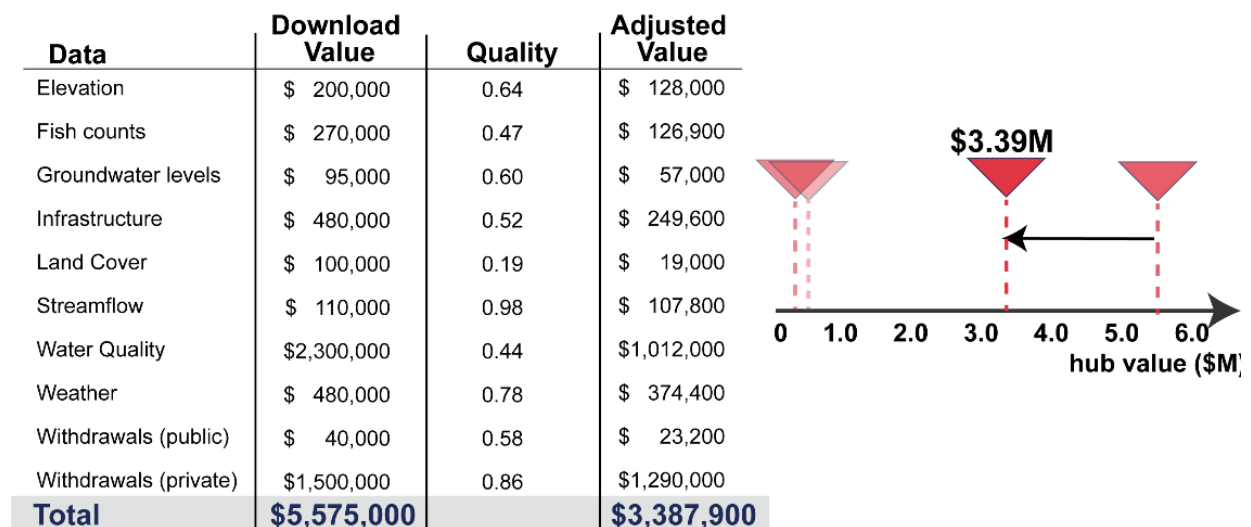
Acme Hub also looked at hub value based on the number of downloads for each dataset, irrespective of unique users (Figure 5). The estimated hub value was also \$5.58M, with the same benefit-to-cost ratio of \$13.9 for each \$1 spent.



**Figure 5:** Hub value based on number of downloads for each dataset.

### (3) Modify costs using data attributes

Next, Acme Hub adjusts the value of their hub based on data quality, which ranged between 19% and 86% accuracy. The final adjusted Acme Hub value is \$3.4M, resulting in \$8.50 in benefits for every \$1 spent (Figure 6).



**Figure 6:** Hub value adjusted for data quality.

## Final thoughts

By equating the value of each download to the full cost of the data, the Consumption-Based method will almost always result in benefits outweighing costs as long as data are being downloaded. As such, the value of data hubs can grow tremendously fast. This may be an unrealistic valuation, particularly since it is not known whether downloaded data are ever put to use. This method may be refined by including surveys to understand what percentage of downloads are put to use and how the data are used. For instance, the hub may learn that the data are put to use once for every 10 downloads, and weigh their valuation accordingly. The method would benefit from these types of additional adjustments to create a more realistic estimate of the value of data hubs.

### **For more information:**

Glue Reply. [The Valuation of Data as an Asset](#).

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