# Valuing Data as a New Asset Type

**Professor Laura VELDKAMP (Columbia University)**

**Keynote at the ABFER Welcome Dinner**
**20 May 2024**

The most valuable firms today are valued primarily for their data. Data, therefore, has emerged as a whole new asset type with central importance for firm valuation.

However, despite its centrality, we lack tools to value data: the classic asset pricing tools are inappropriate for this new asset class.

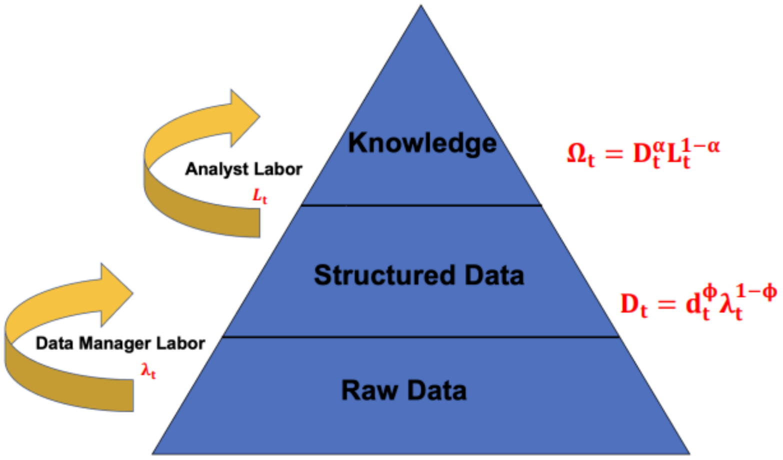
Against this backdrop, Laura Veldkamp, Professor of Finance at Columbia University, delivered a keynote address titled "Valuing Data as a New Asset Type" at the 11th Annual Conference of ABFER. The address described a set of approaches to value data.

The Professor started by defining data. Data or big data is digitized information used for forecasting. This data is usually generated as a by-product of economic activity and acts as a key input to modern prediction technologies like AI and machine learning to forecast outcomes.

Data as a by-product of economic activity leads to the data feedback loop: a firm with more transactions and customers generates more data, which in turn helps improve product quality, efficiencies, or customer satisfaction. This improved performance spurs growth, leading to more transactions and customers, and consequently more data.

This feedback loop can help firms improve profitability through superior decision-making; gaining market power; and mitigating uncertainty by enhancing predictive accuracy.

The Professor then explained how data's value can be gauged by observing the firm's data-related hiring patterns. Data managers convert raw data into structured data, while analysts transform it into actionable insights. Measuring this labour and wages can provide insights into the firm's data value.




The Professor also touched upon data depreciation, focusing on how fast specific data loses value over time. This depreciation rate is informed by Bayes' law and influenced by factors like volatility and persistence of the object forecasted. Once the persistence, volatility, and data stock are known, the depreciation rate can be measured using Bayes' law. Notably, the depreciation rate is faster for highly volatile processes and larger datasets.

The Professor then turned to explaining the six data valuation approaches.

The **Cost Accounting** approach to data valuation equates an asset's value with its production cost when transaction prices are unavailable. However, while purchased data can be valued at acquisition cost, data from firm transactions lacks clear production costs, resulting in a zero book value and complicating data valuation. A potential solution is data barter. Firms offer discounts for customer data. By considering these discounts and identifying instances where firms charge more for data non-disclosure, it is possible to determine data production cost.

The Professor then discussed the **Complementary Inputs** approach. By estimating the number of data-related laborers and their wages, we can infer the value of the data necessary to justify these hires. Likewise, investments into IT infrastructure can help value data – significant investments suggest high data value necessary to justify the expenses.

The **Value Function** approach links the amount of data to future revenues. Specifically, the value of the data is the gross revenue a firm produces with that data, minus its costs, plus a discounted value of the data the firm will have in the next period.

$$V(\text{data}_t) = max_{K,L} A(\text{data}_t) K_t^{\alpha} L_t^{1-\alpha} - wL_t - rK_t + \beta V(\text{data}_{t+1})$$

The state law of motion is the depreciation equation:

$$data_{t+1} = \underbrace{(\rho^2 data_t^{-1} + \sigma_\epsilon^2)^{-1}}_{depreciated\ \ data} + \underbrace{n_s \sigma_s^{-2}}_{new\ \ data\ \ inflows}$$

Likewise, the **Revenue Approach** equates data value to the present discounted value of the revenue it generates, adjusted for risk. However, isolating data-driven revenue from other sources poses a challenge, requiring a clear understanding of how data contributes to revenues. Hence, a model is essential to delineate data's impact on revenue through the computation of counterfactuals – what the firm would earn without the data.

Using the revenue approach, the Professor examined data used in purchasing a risky asset portfolio, leveraging a clear model of data-driven revenue generation in this sector. She employed an equilibrium model, incorporating heterogeneous investors with correlated information and noisy prices, and identified three key statistics to estimate data value: expected return, the variance of return, and the variance of return conditional on data. She found that the value of the same data varied significantly, from $10 to $1.2 million, depending on factors like investor wealth, investment style, price impact, or trading frequency, highlighting the data's private-value nature.

The **Choice Covariance** approach to data valuation is to measure covariances between a firm's choices and payoffs. Data enables better prediction, informing covariance between choices and outcomes. This covariance is vital for better decision-making. In certain instances, such covariances are measurable and can be used as data's value.

Better choices means actions $q_t$ that covary with payoffs $r_t$.

$$E[q_t r_t] = E[q_t]E[r_t] + cov(q_t, r_t)$$

Lastly, the **Market Prices** approach involves analyzing market prices of data in online marketplaces like Snowflake. However, these prices are a function of supply and demand, offering insights different from revenue-based valuations. Similarly, the difference between the firm's market value and book value can be argued as data value. However, while high market value may indicate the value of data assets, they could also reflect other intangible assets like patents and branding. Thus, market prices alone may not entirely capture the data's value.

Concluding her address, the Professor suggested that a thorough exploration across industries – involving aspects like data supply, markets, ownership, and demand – is crucial for gaining a comprehensive understanding of data valuation.