



# Data's Value: Digital Transformation's Dirty Secret

**Steve Todd**

Dell Technologies Fellow

Dell EMC

[steve.todd@dell.com](mailto:steve.todd@dell.com)



## Table of Contents

I'll Keep You (Data Value) My Dirty Little Secret .....	4
We've Only Just Begun .....	6
Hang On, Help Is On Its Way .....	9
We Could Be Heroes .....	15

Disclaimer: The views, processes or methodologies published in this article are those of the authors. They do not necessarily reflect Dell EMC's views, processes or methodologies.

This paper will increase your knowledge on the topic of data valuation. It will also give you a clear understanding of how to build an IT infrastructure that supports data valuation (and why you should do so).

This will be accomplished in four sections:

### **I'll Keep You (Data Value) My Dirty Little Secret**

- You will laugh as you realize that you truly have no idea what the business value of your data is. You will cry as you realize that the purpose of digital transformation is to extract value from data.

### **We've Only Just Begun**

- This section will emphasize that not only do YOU have no idea how to calculate the business value of your data; nobody else in the industry does either. By some estimates, the majority of corporate executives believe that data is reflected on the corporate balance sheet (but it is not).

### **Hang On, Help Is On Its Way**

- You will be rescued from your helpless state via a 5-point plan for implementing a data valuation framework within your IT infrastructure. This plan begins by taking the initiative to learn the state of data valuation research before advocating the research to your executive team.

### **We Could Be Heroes**

- Your implementation of valuation within the IT infrastructure results in transformational new business processes that leverage data as a capital asset. Your company achieves a competitive advantage as competitors play catch-up trying to imitate your data valuation capability.

By the time you finish this paper, you will have a straightforward strategy for turning 0s and 1s into dollar signs.



## **I'll Keep You (Data Value) My Dirty Little Secret**

One of the main goals of digital transformation is to enable the rapid ingest (velocity) of massive amounts of data (volume) of all kinds (variety). Once your infrastructure has undergone the proper “digital transformation” to support the 3 Vs of Big Data (velocity, volume, and variety), rest assured that you can now begin extracting massive amounts of value from that data. Cost savings will be realized! New revenue sources will emerge!

Infrastructure vendors are eager to share with you how to build this Big Data system. In spite of everybody's agreement that “data is the new oil,” there's a hidden problem that vendors will not mention.

*The value of your data before the transformation (X) and after the transformation (Y) will be identical.*

How can this be? The answer is simple. For any data set, the pre-transformation value of that data set is unknown. After the transformation, the value is equally unknown.

This disappointing outcome is largely due to the fact that most CIOs will not ask the emperor why he has no clothes on. According to Gartner (for two years running), “How do I calculate data's business value” is the number one question CIOs should be trying to answer. In March of 2015, the Gartner [Big Data Insights](#) report included the graphic shown in Figure 1.

# The Top Big Data Challenge Remains the Same

What are your organization's top 3 hurdles or challenges with big data?



© 2015 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

Figure 1 – The Top Big Data Challenge Remains the Same. Source: The Gartner Group

Figure 1 clearly shows that the value of data is not only the top problem on the minds of Big Data executives, but it makes the top three of more than half of the respondents!

Why does the rise of the Chief Data Officer not result in an increased focus on the value of data? Gartner's VP of research Doug Laney believes that executive culture may be partly to blame. According to Laney, over 80% of the executives he works with mistakenly believe that the value of data is reflected somewhere on the corporate balance sheet <sup>i</sup>.

The problem has been evident for years. Consider a quote from a joint Big Data Report resulting from a survey conducted by Capgemini and EMC in 2014:

*"Among our respondents, 63% consider that the monetization of data could eventually become as valuable to their organizations as their existing products and services".*

Three years ago this statement was a cry for help. The industry was believing that data was going to become a capital asset with value (a value equivalent to the products and services they were expecting to sell)! Yet no digital transformation guidance was given for treating data as a capital asset.

In 2014, in order to determine whether or not this cry for help was real, the Dell EMC Office of the CTO turned to academia to determine just how very real this problem was (and began to search for solutions).

## We've Only Just Begun

Dr. Jim Short of the University of California San Diego works at the San Diego Supercomputer Center (SDSC). The SDSC had just completed a “How Much Data” research project which noted that the amount of internet data flowing into a home was far out-pacing the ability of the home’s residents to consume it all. This scenario set Dr. Short down the path of considering which data was “most valuable” to the consumer (e.g. which data did they choose to consume and why).

With the data monetization cry for help ringing in the ears of EMC researchers, a collaborative data valuation research project was established. The “Architecting for Value” research conducted by SDSC would focus on two areas:

1. Broad industry studies and targeted interviews that would paint a picture of the current state of data valuation business processes.
2. IT architectures that would advance the state of the art for data valuation.

Dr. Short explained the research approach<sup>ii</sup>:

*“Our methodology for answering our core question – what’s data really worth? – included interviews and secondary data collection on information-related activities in 36 companies and nonprofit organizations in North America and Europe. Most had annual revenues of more than \$1 billion. They represented a wide range of industry sectors, including retail, healthcare, entertainment, manufacturing, transportation, energy, education, and government. Despite our focus on data value, most organizations were focused on the challenges of storing, protecting, accessing and analyzing massive amounts of data, efforts for which the IT organization is primarily responsible. While the IT organizations in our study were highly effective in storing and protecting data, they alone cannot make the key decisions that transform data into business value. Our study lens, therefore, quickly expanded to include chief financial and marketing officers, and in the case of regulatory compliance, legal officers. Importantly, as the majority of companies did not have formal valuation practices, we adjusted our methodology to focus on significant business events triggering the need the valuation, such as mergers and acquisitions, bankruptcy filings, or the acquisition and sale of data assets”.*

Dr. Short’s research uncovered striking cases of extremely high data valuation amounts that were unaccompanied by formal valuation business processes. He confirmed that not only was the industry struggling with the problem of data valuation, but staggering amounts of value (e.g. hundreds of millions of dollars per use case) were at stake.

Some of the more high-profile examples are depicted in Figure 2.

# DATA VALUATION USE CASES



Figure 2 - Data Valuation Use Cases. Source: Dr. Jim Short, San Diego Supercomputer Center

Dr. Short uncovered a variety of data valuation use cases:

- Creditor valuation: The highest priced asset in Caesar's Palace's bankruptcy proceedings was its Total Rewards Customer Loyalty database. The database was valued at one billion dollars<sup>iii</sup>.
- Mergers & Acquisitions: LinkedIn acquired Lynda.com's video library for over one billion dollars<sup>iv</sup>.
- Data Insurance: Third party insurance companies like Liberty Mutual place a value on data during the underwriting process. A lack of accurate valuation can result in huge losses for the insurer<sup>v</sup>.
- Data Monetization: DNA company 23andMe monetized DNA data by providing access to Genzyme. This resulted in millions of dollars of revenue that augmented their more traditional product sales<sup>vi</sup>.
- Data Sale: UK grocer Tesco sold their Dunnhumby database of UK shoppers' purchasing habits for over 900 million dollars.<sup>vii</sup>

What's remarkable about these high profile valuation use cases is the conclusion that the lack of formal business processes for handling data valuation can result in hundreds of millions of dollars of corporate loss. Dr. Short drilled down into several of these cases to highlight how poor data valuation processes can damage the business.

For example, how will the Caesar's Palace creditors split up the 1 billion dollar asset represented by the Total Rewards Customer Loyalty database?

Similarly, consider the Tesco use case. Dr. Short comments<sup>viii</sup>:

*“In early 2015, Tesco decided to sell its customer data science and loyalty program, Dunhumby, which it had internally valued between £1 and £2 billion. Potential buyers were reported to be WPP and Google Ventures, as well as several private equity firms, including Apax Partners, CVC Capital Partners, and TPG Capital. But UK and US analysts found Dunhumby difficult to define: was it an advertising agency or a data platform containing information on the buying habits of its database of 770 million shoppers? Ostensibly on the market for many months, talks with potential buyers and equity partners stalled, and the price of the Dunhumby unit reportedly fell to £700 million..... Data contracts were seen to be at the heart of the issue that saw the reduction in the value of Dunhumby over the course of sale negotiations..... Tesco's machinations over the sale, transfer and partial retention of Dunhumby assets echoes the Total Rewards "divisibility" issues before the Federal Court in the Caesars Palace bankruptcy case. Tesco and the examples above all share a common characteristic: data value is difficult to define, but is judged in these cases to be worth hundreds of millions of dollars of value.”*

In summary, the first year of research by Dr. Short painted a gloomy picture of an industry struggling to answer some of the most simple, basic questions related to data value:

- What is a good price to pay when purchasing a data set?
- What is the ultimate return on this data set post-purchase?
- Which data sets within the enterprise are the most monetizable? For sale? For rent?
- What is the process to convert a data set to a data product or service?
- What kind of ROI (return on investment) is being generated by data scientists? For which data sets?
- What are proper premiums and limits for data insurance?
- How does data's value dictate how it is stored and managed by IT?

Dr. Short entered year two of the research ready to answer those questions.

## Hang On, Help Is On Its Way

Year two of SDSC's data value research would focus on "getting started" recommendations for advancing the cause of data valuation. The industry would once again be surveyed to determine who (if anyone) was already working on solutions to the problem. During the course of the second year, an interesting set of valuation initiatives were discovered. An exploration of these initiatives resulted in a set of five research recommendations that were first presented at the October 2016 Evanta Global CIO (Chief Information Officer) Executive Summit. These recommendations identified the CIO as a key player in the introduction of data valuation capabilities to the business. The CIO plays a pivotal role in digital transformation: their infrastructure stores the data itself. By adding data valuation as a transformation activity they can position their company for competitive advantage and significantly reduce the possibility of losing hundreds of millions of dollars due to data valuation immaturity (as outlined above).

The five recommendations for the CIO are:

1. Choose a framework (a variety of options have been uncovered during the research).
2. Select the IT touch points for implementing data valuation.
3. Combine business and technical metadata to better enable data valuation.
4. Provide a data catalog with annotation capabilities.
5. Implement new valuation business processes on top of this architecture.

Each of these recommendations is described below.

### **1. Choose a framework for introducing data value to your business**

The CIO needs to insert him- or herself into business discussions regarding data value. This can include:

- Approaching the party responsible for purchasing data sets and discussing whether or not the company is receiving a good deal.
- Determining whether or not business processes are in place for measuring the ROI for purchased data sets.
- Meeting with data science teams to discuss ROI on data analytic experiments.
- Any other relevant valuation-related use cases.

The research discovered a set of emergent valuation frameworks that can be used as a basis for these discussions. CIOs should familiarize themselves with these approaches and work with executives to choose valuation approaches that work best for the business. In this section, we will highlight two frameworks.

One framework that has been introduced is the brainchild of Dell EMC's Dean of Big Data, Field CTO Bill Schmarzo. Bill believes that CIOs should shift their mindset from data being "a cost to be minimized" to "an asset to be monetized". Bill's approach (currently the focus of a research project at the University of

San Francisco) begins with the listing of the top business decisions being made in the near future. His approach includes a step-by-step process to trickle down the potential value of these decisions as a way of assigning “potential value” to data sets<sup>ix</sup> *as well as* the analytic models<sup>x</sup> that extract insight. Bill calls his approach the CDO (Chief Data Officer) Toolkit. The CIO is an excellent person to bring this toolkit to data-savvy executives.

The second framework is known as “Infonomics” and it is the brainchild of Gartner VP of Research Doug Laney. Doug has created a set of foundational and financial models that can be used as a starting point for calculating data’s value (e.g. priority rankings and/or dollar amounts). One such equation, the Business Value of Information (BVI), calculates the relative ranking of various data sets within a corporation. While it does not provide the absolute value of data, it is still a solid starting point.

$$BVI = \sum_{p=1}^n (\text{Relevance}_p) * \text{Validity} * \text{Completeness} * \text{Timeliness}$$

Figure 3 - The Business Value of Information. Source: Doug Laney, Gartner Group

In Figure 3, the “Business Value of Information” (BVI) is created by identifying each line of business that consumes the data (“n”) and calculating the relevance of the data to that business unit (e.g. on a scale of “0” to “1” with “1” being highly relevant). A summation of each of these relevance scores is then multiplied by the following data attributes:

- Validity: what is the overall quality of this data?
- Completeness: is 100% of the data currently in possession by the company?
- Timeliness: how quickly can the data be retrieved for usage?

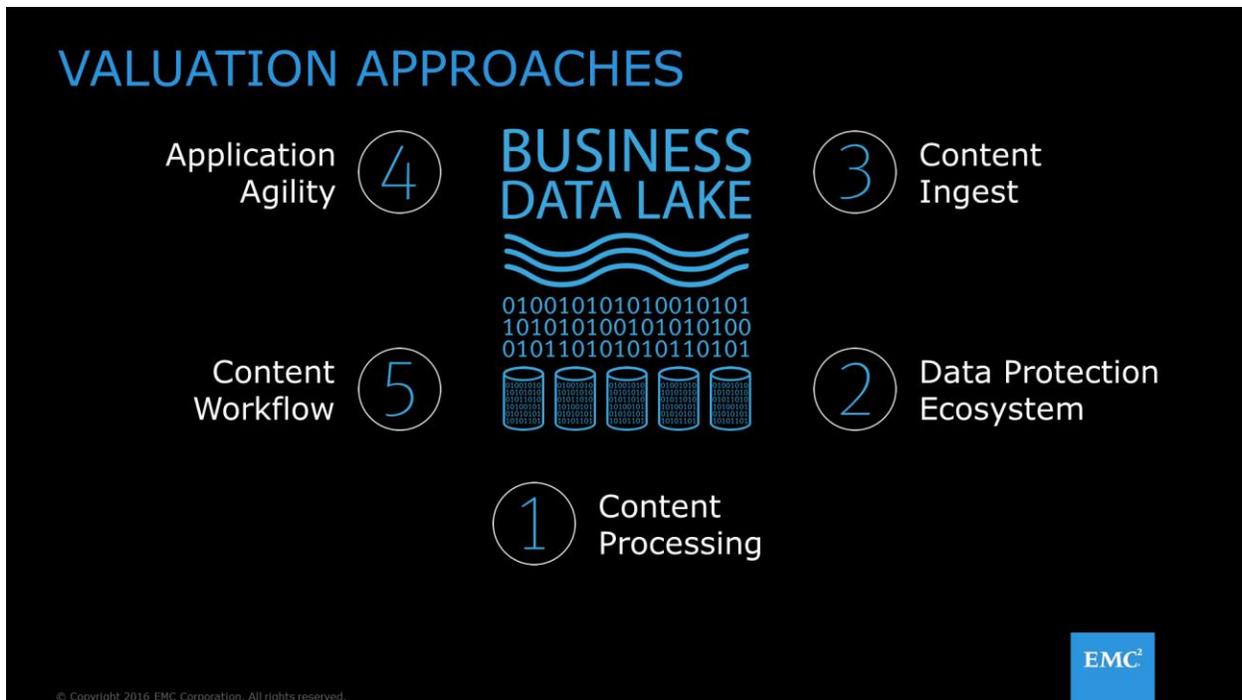
Calculating the BVI of a data set allows the business to begin to understand, for the first time, the overall business importance of a given data set. It allows for comparison between data sets and can be the first step towards a new culture of “data valuation fluency” being introduced.

In addition to Schmarzo and Laney, the Wikibon analyst team is conducting open research into data valuation. They have proposed a Digital Business Platform<sup>xi</sup> that formally treats data as a capital asset. Their approaches are likewise worthy to be considered.

Using any of these approaches would establish the CIO as a “data thought leader” within the business. Based on which valuation algorithms are chosen, the next step considers where to run those algorithms.

## 2. Selecting IT “touch points” for data valuation

The next step for the CIO is to explore all of the possible options for calculating data’s value within the infrastructure. During the course of the research, a set of innovative ideas were proposed for consideration by the CIO. These ideas are depicted in Figure 4.



**Figure 4 - Valuation Approaches in an IT Architecture**

The CIO could choose to run valuation algorithms in a variety of locations, including:

1. Performing content processing on the production data. This approach would perform in-place analysis (e.g. semantic parsing of the content for use in relevance calculations).
2. Performing content processing on a backup copy of the data. This not only has the benefit of eliminating the load on the production system, but it has the additional benefit of providing richer metadata (e.g. application, user, and backup schedules) for consideration by valuation algorithms.
3. Executing valuation algorithms while data is streaming into a company has the advantage of assigning an immediate valuation score. It also prevents the erroneous assignment of value that could result from potentially stale backup data.
4. Digital transformation implies increasingly frequent deployment of fresh applications that collect and score a wider and wider variety of data. Measuring the velocity (update frequency) of application updates and correlating that frequency with value is another potential way to assess data's value.
5. Generating valuation metadata during content workflow (e.g. as a data scientist publishes analytic models and results into a data lake) is an excellent way to tie business context to technical data.

A CIO needs to consider which approach (or approaches) represents the optimal location to calculate the value. Within Dell EMC, the Chief Data Governance Officer (CDGO) has implemented options 3 (content ingest) and 5 (content workflow). The success of the CDGO in implementing these two touch points was made possible by combining business and technical metadata.

### 3. Combine Business and Technical Metadata

During the course of the research, it was discovered that Dell EMC’s CDGO was working closely with IT on the implementation of a Data Lake that contained any and all data related to EMC’s customers. The CDGO decided to create a metadata repository alongside the data lake. This metadata repository combined business metadata (which line of business, which customer, etc.) with technical metadata (which file, which database rows, what schema, etc.). Figure 5 highlights this architecture.

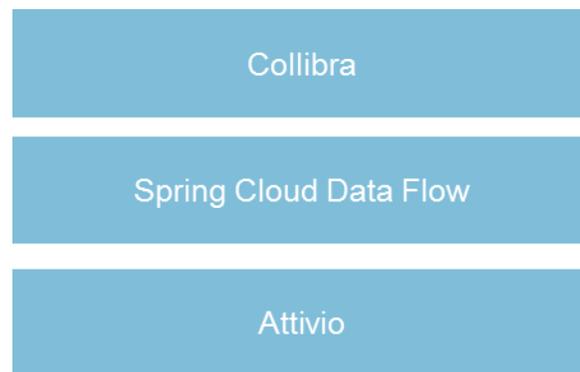


Figure 5 - Data Governance Architecture. Source: Office of the CDGO, Dell EMC

Collibra is used by the business for any and all data-related business activities (e.g. ingest and data science workflow). This layer generates business-related metadata that is critical for understanding data’s value. As this business metadata is generated it needs to be joined with technical metadata coming up from the bottom-most level.

The business and technical metadata are joined together via the use of Spring Cloud Data Flow. This results in metadata that is ripe for valuation algorithms. The CDGO team uses the Spring Cloud Data Flow layer to generate new forms of valuation metadata (e.g. flags that identify critical data elements). This real-world implementation had not been discussed in previous methods (e.g. the CDO Toolkit or Gartner BVI).

At the bottom layer, Attivio can be thought of as a software package that “discovers” data assets and generates corresponding technical metadata. It inspects each and every data asset in a data lake and generates useful metadata, such as schema descriptions.

Once business and technical metadata have been combined, and valuation algorithms are running in areas such as ingest and workflow, a set of valuation scores will be generated. These valuation scores need to be recorded properly.

### 4. Annotate a Data Catalog with Valuation Metadata

As a last implementation step, the CIO should consider that literally all traditional balance sheet assets (e.g. computer equipment, employees, real estate) are located in a catalog and their usage (e.g. upgrades, salary, maintenance) is tracked as a way of measuring value. In order to begin treating data as a capital asset with value, CIOs need to do the same. Dell EMC has introduced a data catalog with

metadata hooks that can be used for this purpose. The catalog and metadata capabilities are part of a recently introduced product called the Analytic Insights Manager (AIM), depicted in Figure 6.

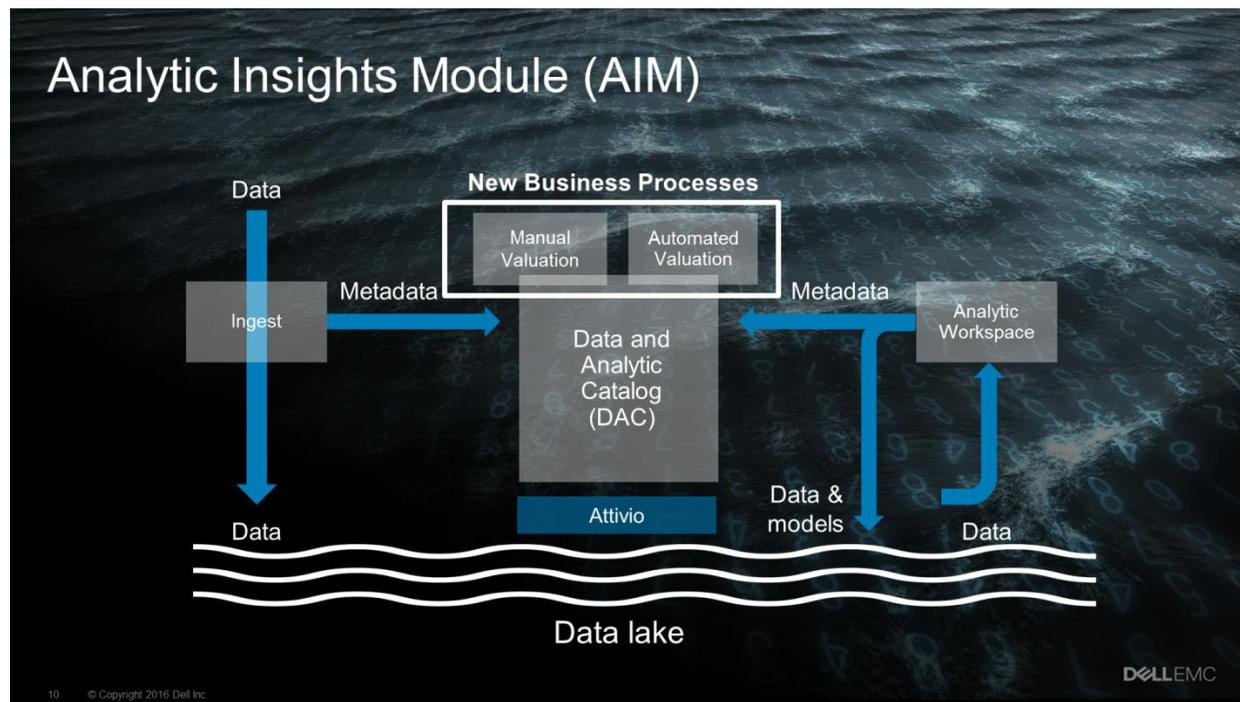


Figure 6 - Analytics Insight Module Architecture

At the heart of AIM is a data and analytic catalog (DAC). It is used to create an entry for every data set and every analytic model that is stored within a data lake. Note that when a new data set enters into the system (e.g. on Ingest), a DAC entry is created and business metadata is stored. Similarly, when new models and data outputs are published into the data lake, new DAC entries are likewise created and associated with a business context. Attvio is also used to discover existing data sets and technical metadata is published.

One of the features of the DAC is that it has an API and a GUI that allows an administrator (or an automated script) to associate key/value pairs with DAC entries in the catalog.

For example, using a Schmarzo “CDO Toolkit” business process may result in an administrator launching a GUI and associating an economic value with a data lake asset that is used to drive a high-value business decision.

Similarly, an automated script could perform a BVI calculation and call an API to associate the BVI score with a data element stored in the data lake.

After driving the dialogue to establish corporate valuation policies, and after creating an IT architecture that fully supports valuation, cataloging, and annotation, it is time to start the annotation process by choosing some initial use cases.

## 5. Implement the business process on top of the valuation architecture

The diagram below steps through the Schmarzo approach of assigning business value to data sets and analytic models. Figure 7 depicts a simplified version of the CDO Toolkit approach.

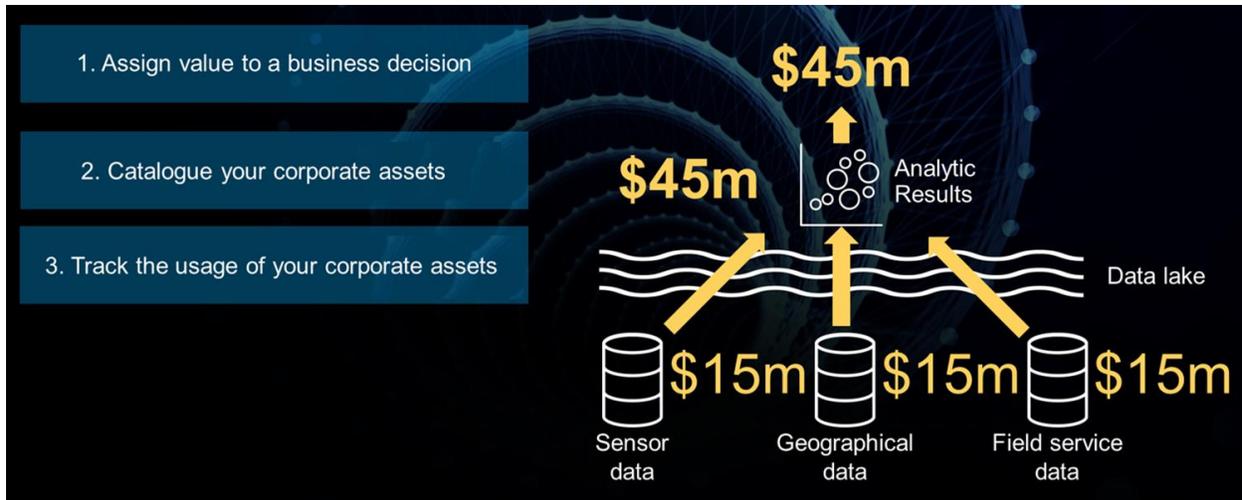


Figure 7 - Data Valuation Use Case: Energy Company

Consider the use case of a CIO at an energy company that believes it can save a significant amount of money by reducing energy theft. The Schmarzo approach, at the highest level, would work as follows:

- Step 1: assign value to a business decision. In this case, the energy company believes it can save \$45 million by using data sets to determine where energy theft is occurring on the grid.
- Step 2: locate the data assets that will assist the company in identifying the areas of theft (e.g. sensor data, geographical customer data, field service logs), centralize those assets into a data lake, and create entries into a data and analytic catalog
- Step 3: Begin tracking usage of the cataloged assets. This involves creation of an analytic model that produces a recommended list of addresses where theft is likely to occur. The analytic model is entered into the lake, and its relationship to the three input data sets is noted. The recommended list is likewise cataloged. The potential value of \$45 million is then annotated within the data lake. This annotation should be recorded against all input data sets, the model, and the output. In the energy company use case depicted above a very simple algorithm is used. The algorithm divides the potential value (\$45 million) by the number of input data sets (3) and distributes the result (\$15 million) to the input data sets. For the analytic model and the output, the algorithm performs a simple annotation of \$45 million for each.
- Step 4 (not pictured): Over time, evaluate whether or not the hypothesis of \$45 million was achieved. The annotation of “potential value” can then be followed with an annotation of “actual value”. This type of feedback loop will result in a more mature model that will enable more accurate calculations of data’s value in the real world.

## We Could Be Heroes

As the SDSC enters its third year of researching the broad topic of data valuation, a community is forming. In SDSC's first-ever Data West Conference<sup>xii</sup> in December of 2016, data valuation was featured as a prominent topic on the agenda. Communities of practice will inevitably form around the topic. The straightforward research results described above, if implemented correctly, could provide a significant competitive advantage in the new digital world order.

As the CIO continues to annotate more and more data assets within the enterprise, a full catalog of data assets can be brought to bear against the financial performance of the company. For example:

- When purchase prices for data sets are negotiated, similar data sets can be identified within the DAC and their values retrieved. Knowledge of the business value for those identified data sets can significantly reduce the risk of overpaying for new data sets. Knowledge of previous purchases and their resulting ROIs could inform the decision on whether purchasing is likely to be a very bad idea.
- High-value data sets can easily be identified and routinely be discussed as potential sources of revenue for the corporation. This revenue could be received by selling or renting. Strategies for creating data products and services around high-value data sets could be implemented (the CDGO at Dell EMC participated in the creation of Dell EMC's first data service: MyService360<sup>xiii</sup>).
- The ROI for data science teams can more easily be calculated. Specific data scientists that tend to write high-value data models can easily be identified.
- CIOs can manage data sets based on value, e.g. the higher the value, the higher the level of data protection.
- Data insurance conversations for limits and claims on data loss or theft can more easily be evaluated.

Each of the bulleted items listed above can be tied to reduced risk, cost savings, or revenue generation.

Now you know the dirty little secret. Digital transformation should not only be about choosing the right infrastructure provider. It should also be about a new framework for tracking data as a capital asset.

You also know that we've only just begun. The industry as a whole is suffering from non-existent valuation business processes.

Help, however, is definitely on the way. There is a clear and concise 5-step plan to implementing valuation. Hang on: start interacting with the consultants and communities mentioned in this paper.

Will corporate data appear on the balance sheet as an asset? FASB is working on it<sup>xiv</sup>! Will you be prepared?

There is no doubt that if you stick to the 5-step plan and transform your company into a data-savvy juggernaut, you will be a hero.

Even if it's just for one day.

---

<sup>i</sup> <http://reflectionsblog.emc.com/much-data-worth/>

<sup>ii</sup> Taken from “Valuing Billion Dollar Data”, to be published by Sloan Management Review in mid-2017.

<sup>iii</sup> <http://www.wsj.com/articles/in-caesars-fight-data-on-players-is-real-prize-1426800166>

<sup>iv</sup> <https://press.linkedin.com/site-resources/news-releases/2015/linkedin-to-acquire-lyndacom>

<sup>v</sup> <http://www.bostonglobe.com/business/2014/02/17/more-companies-buying-insurance-against-hackers-and-privacy-breaches/9qYrvlhskcoPEs5b4ch3PP/story.html>

<sup>vi</sup> <https://www.forbes.com/sites/matthewherper/2015/01/06/surprise-with-60-million-genentech-deal-23andme-has-a-business-plan/>

<sup>vii</sup> <http://adexchanger.com/ecommerce-2/tesco-eyes-sale-of-dunnhumby-its-nearly-1-billion-shopper-data-business/>

<sup>viii</sup> Valuing Billion Dollar Data.

<sup>ix</sup> <http://www.cio.com/article/3130334/analytics/chief-data-officer-toolkit-leading-the-digital-business-transformation-part-i.html>

<sup>x</sup> <http://www.cio.com/article/3131035/analytics/chief-data-officer-toolkit-leading-the-digital-business-transformation-part-ii.html>

<sup>xi</sup> <http://wikibon.com/transforming-to-digital-business/>

<sup>xii</sup> <http://www.datawest.org/>

<sup>xiii</sup> <https://www.emc.com/collateral/solution-overview/h15047-svo-emc-myservice360.pdf>

<sup>xiv</sup>

[http://www.fasb.org/cs/ContentServer?c=Document\\_C&pagename=FASB%2FDocument\\_C%2FDocumentPage&cid=1176168366904](http://www.fasb.org/cs/ContentServer?c=Document_C&pagename=FASB%2FDocument_C%2FDocumentPage&cid=1176168366904)

Dell EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” DELL EMC MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying and distribution of any Dell EMC software described in this publication requires an applicable software license.

Dell, EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries.