



# Quantitative Value of Data and Data Management

## Co-Authors:

Paul Haines – Noah Consulting

Mark Wiseman – Hess

## Abstract:

Data is at the heart of all exploration and production workflows and is always a pivotal factor in any business outcome (be it good or bad!). The E&P business customer consistently acknowledges that data availability, integrity, and readiness are crucial to the success of their business. Yet formally accepting data as a corporate asset, both in terms of "book value" and workforce behavior, remains elusive and undermines any strategic approach and long-term investment in data management programs. In IM conferences, meetings, and discussions we, as IM professionals, continually hear about the difficulty or inability to create a robust business case for IM projects and/or value for data as a corporate asset in dollars and cents. We often only think about trying to calculate a value for data as a sub-task (activity) of "due diligence" when acquiring new E&P assets, divesting E&P assets, etc. Having valid and repeatable methods to calculate the value of data would be a great benefit for many reasons in an E&P company. Additionally, even if you have a good and valid number for the value of data, it is more difficult to get a valid and repeatable method for calculating the value of data management on top of that.

This paper will discuss these challenges and provide a repeatable method to calculate the value of data to an E&P company – a true value in terms that business sponsors will stand behind and defend when funding approval, business support, and business participation reaches critical decision points. The method will analyze data from a cost standpoint, a value standpoint, and a risk standpoint. Once established, we will use this number as the "basis" for analyzing the value of data management; bad data management practices will detract from the base number and good data management will add to the base value of the data asset.

Finally, we will use historical data to demonstrate the validity of our approach.

## Introduction & Problem Statement:

We all know that data has value. We all know that data is an asset. Oil and gas data costs a lot to acquire, process, interpret, and store. If there was no value, then E&P companies would not spend the

money described. E&P companies exist to discover, produce, and sell hydrocarbons to maximize shareholder value. They do that by using data to a better advantage than their competitors.

One of our challenges as data managers is helping the business understand and appreciate the value of data by turning data into an asset with measurable value to the shareholder.

- Metrics
- Tangible asset
- Measurable value

This is not meant to be a definitive treatise on the subject as much as it is meant to be thought provoking. Although there may not be general agreement in the actual factors or numbers involved, the anticipation is that this can be used to solidly demonstrate data can and should be thought of as an asset.

What is the path to get there and establish a value versus cost?

How can we measure value in a quantitative manner?

Can we measure cost of data in a quantitative manner?

Can we divide value into components that are more measurable than the esoteric “value” concept?

Are there other factors that go into determining the value of data as an asset?

How does the management of data – good or bad – factor into our equation? Can managing data poorly actually detract from the ability to use data to make meaningful decisions? Can managing data well increase the value to such a point that the value of the data is greater than the total cost of managing it?

This paper will answer some of these questions and issues and provide a repeatable method to calculate the value of data to an E&P company. A goal is for our business advocates to have the means to justify costs when funding approval, business support, and business participation reaches critical decision points. The method will analyze data from a cost standpoint, a value standpoint, and a risk standpoint. We use this number as the “base case” for analyzing the net worth of data; bad data management practices will detract from the base number and good data management will add to the base value of the data asset.

## **Develop Solution Thesis:**

Our underlying thesis:

1. Data can be and should be treated as an asset.
2. Data is worth more than it costs – it has a net value – and we can approximate it as a base case.
  - There are certain components that go into the total costs of data.
  - There are certain components that go into the value proposition of data and data usage.

- There are certain components of business risk that data has an effect on. Realizing those risks can increase costs and mitigating risks can reduce costs.
3. Data management has an effect on the value of the data asset.
    - Good data management has a positive effect on the base case.
    - Poor data management has a negative effect on the base case.
  4. We can develop and implement a methodology to quantify the factors to come up with a “number” as a data asset net worth.
  5. The methodology and resultant value can be directly related to data acquisition cost.
  6. Data and data management has an ROI that should be directly related to data acquisition cost.

And further to the underlying thesis, if data is treated as an asset and not strictly a cost, then:

1. As an asset, it should be added to a company value in a similar manner as reserves.
2. It will increase shareholder value.
3. Good data management can be shown to also increase shareholder value even more.

## Qualitative Details:

Data as an asset:

An asset has implicit and explicit attributes that contribute to any value assessment. One of the fundamental concepts is that a value has to be based on “something.” What is an attribute that can be used to determine value and how is it reliably, consistently, and convincingly be determined? It is a fact that all data assets have an acquisition cost that can be somewhat reliably and consistently be determined. In developing our ideas, that was our thought – we can use acquisition cost to formulate a method to determine value.

We need to look at data is an asset and treat it as an asset. What does that mean? Your car and your house are important assets. Your assets have a cost associated with them, and you get more benefit from them than they cost you. Taking care of and managing your assets makes them more valuable, while not taking care of your assets can deteriorate them until a point where they literally fall apart. You insure your assets because there are risks associated with assets that you want to minimize, alleviate, or transfer the risk.

Let us analyze data as an asset from the perspective of cost, value, and risk. Once these are determined, these will be called the “base case”. From that base case, then we will look at (a) how good data management affects the values related to the base case and (b) how poor data management affects the values related to the base case.

### The Base Case -

#### Base Case Costs of Data –

These base case costs are intended to be costs associated with any data that is brought into and used in an organization to make decisions. The base case was fashioned in an assumed

cost of acquisition of \$1,000,000. So everything in all of the quantitative analyses and graphs spawns from an initial acquisition cost of \$1,000,000. From that initial \$1,000,000 it was modeled that the base case for cost of data is \$10,000,000, the value of data is \$19,000,000, and the risk attribute of data is \$10,000,000.

#### Cost to Acquire –

Acquisition can come in many forms but two forms readily come to mind:

- Acquiring it directly from the field such as well logs, cores, seismic surveys, etc.
- Purchasing it from a data vendor such as wells, well logs, production, studies, etc.

So in our thesis, everything is based from the cost to acquire some “unit of data.”

There are other attributes that go into the “total cost” assessment.

Let us look at the following related to costs of assets:

- Cost to use and leverage
- Cost to replace
- Cost to maintain
- Cost of decisions based on the data usage

#### Cost to Use and Leverage –

When data is brought into an organization, it can go through many changes before it is used to help make a decision. For example, it can be put through processes, transformed, and manipulated. The basic management of data, validation, and the quality control of data, etc. is also an aspect of its use. It can be loaded into interpretations software and integrated with other data to create yet more new data. It can be used by many people in many disciplines throughout the E&P value chain during which time it is used over and over again in some form to help in the decision making process.

#### Cost to Replace –

During the useful life of a data asset, it is sometimes necessary to replace or repurchase the data that you previously acquired. This is almost never a desired outcome. Reasons for the need to replace can include losing the data asset, corruption of media, and loss of the right to use.

#### Cost to Maintain –

Maintaining data as an asset take time, effort, money, and resources, just as any other asset of value. Maintenance costs can include storing, protecting, loading, formatting, indexing, refreshing, and supporting. Data becomes stale and deteriorates over time, and maintenance of data helps to keep data in a usable form.

#### Cost of Decisions Based on the Data Usage –

Data is used in making decisions, which cause actions to occur. In the E&P space, actions can include (at a very high level) exploration to add more reserves to the portfolio, and production of hydrocarbons at annual BOE targets to generate revenue for the company. So, in an attempt to characterize all costs associated with data as an asset, this is meant to represent a portion of a company’s total spend attributed to decisions based on data that

would not have otherwise been spent.

**Base Case Value of Data-**

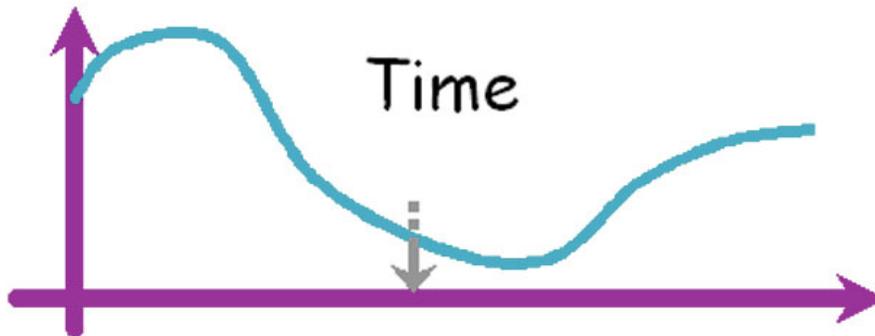
There are many possible components to putting a value on data. The following are components that we believe are independent and additive to our data value proposition:

- Time value of data
- Performance value of data that comes from increasing people productivity
- Integration of data to further its relevance and applicability
- Value of decisions based on the data usage

**Time Value of Data –**

Data is normally used almost immediately after its acquisition. At that time, it has a high value. Normally the value of data goes down over time, but at some time in the future, the value of data normally increases again. This can be due to a variety of reasons such as:

- Renewed interest in certain oil or gas fields
- Lack of ability to acquire new data over the area due to development
- Renewed interest in the data due to improvements in technology



**Performance Value of Data that comes from People Productivity –**

Data is used by many people in an E&P company to aid in finding new hydrocarbons, efficiently exploiting areas where hydrocarbons have been found, or optimizing operations to produce hydrocarbons at a profitable lifting cost. This use of data, even in its most basic form, allows people to be more effective and efficient in their daily work. The right data at the right time allows people to be productive with their time.

**Integration of data to further its relevance and applicability –**

The value of data increases when it is integrated with other data either of like kind or of distinct data types. Paleo and well core data integrated together with log data possesses a much greater value than any of the three data types alone. Seismic data integrated with well data allows for much more relevant interpretations. Culture and surface data greatly enables more accurate decisions concerning profitability of fields. These are all examples showing the importance of data integration to further its relevance and applicability in enhancing the value of the data.

#### Value of Decisions Based on the Data Usage –

More than any other component, the real value of data comes to the forefront when looking at the value of decisions made based on data. In our industry, that value is the reserves in the ground, efficiencies gained in operations, and optimized revenues in portfolio management. This is not to suggest that all the value in a company can be or should be attributed to the data, but there is a percentage that can and should be. Depending on how one looks at it, that small percentage will equate to a large sum of money. In past efforts, there have been attempts to tag 0.1% to 3% of the value to the data. It is sometime easier to attribute a value of the data to a difference in reserves or improvements. In our analysis, we are taking a different approach and stating that in the base case, the value of those decisions based on data usage is a percentage of the cost of the data acquisition. Given the other components, it can be seen that better decisions can be made when data is more leveraged and there is more of it ... at least, up to a point.

#### Base Case Risk Attribute of Data:

Using risk as a component of determining the value of data is not a straightforward as the cost and value analysis. None-the-less, risk or the mitigation of risk using data can be of high value or high cost. It is a primary factor that needs to be taken into consideration. There may be many components of risk that might be considered. The three described below can be understood by using simple examples:

- Regulatory risk of not utilizing the right data
- HSE (Health, Safety, Environment) risk of not utilizing the right data
- Risk of making bad decisions based on not utilizing the right data

#### Regulatory Risk of not Utilizing the Right Data –

Each state, province, or governmental agency has certain requirements that E&P companies have to follow in exploring, drilling, and producing hydrocarbons. These regulatory requirements seem to be increasing over time, as we can see in the news. Data is a key component of most regulatory requirements. Regulatory bodies can impose stiff fines and penalties if regulations are not followed in a timely manner, when all the required information is not provided, and/or when data is suspect and an audit is required. This component tries to capture that risk – potential fines and penalties for not following the regulations or the cost of mitigating the risk through other means.

#### HSE Risk of Not Utilizing the Right Data –

Safety is a key concern on every oilman's mind, and HSE risks are very real. The stakes are high and, although the chances of catastrophic events are very, very small, the potential impact, cost, and liability are huge. Data plays a very critical and central role in mitigating HSE risks. Data related to formation pressure, rock composition, culture data, sea floor mapping, equipment maintenance, gas composition, etc. can be easily shown as examples of the need to use the right data to analyze, quantify, and mitigate risk related to HSE. Not having or not

using the right data can potentially have very high costs that affect an E&P company's bottom line.

#### Risk of Making Bad Decisions Based on Not Utilizing the Right Data.

In a similar, but contrary, example as to how good data can enhance decisions and create more value, there are many examples of poor data management causing poor decisions to be made. That example will be more fully vetted in another section, but this risk relates to the costs or potential costs associated with bad decisions being made using data. An E&P company makes decisions and takes risks. Each decision made every day carries a certain risk associated with it. There is even a cost and risk associated with not making any decision. Part of this cost deals with the extra efforts involved in trying to lower risk associated with major decisions – e.g. measure twice and cut once. Another part of this aspect deals with making decisions that are poor by not having or not utilizing the right data. Examples can be not running a cement bond log when returns are not what were expected, drilling a dry hole due to lack of the right offset well, or perforating off depth because of an improperly interpreted gamma ray log.

### **The Good Data Management Case -**

What is good data management? It is not the intention of this paper to go into details as to what good data management is. There have been many papers and presentations relating good data management practices in our upstream E&P space. We will state that good data management is having the right data at the right time in the right format. It is managed and maintained through proper data governance and stewardship; it is of known quality, source, and lineage, and referential integrity is applied. It is indexed, mastered, and cross-referenced across systems as required. It is easily browsed and queried and both structured and unstructured data can be consumed in a variety of ways.

In a general qualitative sense, you would expect good data management to do good things to the overall asset value of a data asset. As we have seen, you would especially expect the “value” of the data utilizing good data management to be much greater than the base case.

#### Cost of Data – Good Data Management Case –

Good data management will take the base case and actually lower the cost of the data asset. In our example and our analysis, the overall cost of the data went down 20% due to good data management. Although the cost to maintain the data went up significantly (33%), the other costs decreased. Since the data is managed (see definition above), the cost to use and leverage the data goes down. Also, since the data is used so much more effectively, the cost of decisions also goes down.

#### Value of Data – Good Data Management Case –

More than any other factor, good data management greatly enhances the value of data. In our example and analysis, the overall value of the data was increased by over 300% above the value of the base case. The value of decisions was shown the greatest gain – better data from good data



management will provide for better decision analysis over time. Since E&P decisions can be so critical to the overall success of the company and these decisions provide for the direction of future efforts, even if a small percentage is attributed to data, then the resultant numbers are large.

BOE / DAY	BOE / YEAR	COST PER BARREL	Annual Gross Production (\$)
100,000	36,500,000	\$100	\$3,650,000,000
% Improvement	BOE / YEAR	COST PER BARREL	Annual Gross Production (\$)
0.1%	36,500	\$100	\$3,650,000
0.5%	182,500	\$100	\$18,250,000
1.0%	365,000	\$100	\$36,500,000
3.0%	1,095,000	\$100	\$109,500,000
5.0%	1,825,000	\$100	\$182,500,000

The integration value is enhanced because the data has been prepared properly for integration and collaboration. Understanding the quality, lineage, etc. of the data will improve the relevance and applicability. The time value of the data is increased because it is managed properly; in the future (perhaps years), it will be able to be found, be used, and be integrated with newer data. The professionals will have their productivity greatly enhanced because they will have the right data at the right time. The following example shows a quantitative analysis of the performance value strictly from productivity gains.

Time Spent Managing Data (Based on 100 PetroTechnical Professionals)				
Pure Cost Efficiencies correlating users time spent on data management tasks - Agnostic of Data Quality benefits, Data Duplication Costs, Impact of Missing data etc)				
Average Salary Per Geoscientist (Industry Average - Uplifted)	Time Spent Doing Data Management	Total User Cost on DM	Total Headcount Cost	Reducing Time Spent on Data by 50%
\$250,000	30%	\$75,000	\$7,500,000	\$3,750,000
\$250,000	50%	\$125,000	\$12,500,000	\$6,250,000
\$250,000	70%	\$175,000	\$17,500,000	\$8,750,000

Risk Attribute of Data – Good Data Management Case –

As stated before, this is a more difficult component to deal with because it is not as easy to understand. With good data management, the cost of risk mitigation and the potential cost is greatly reduced. With good data management, there is much less risk of regulatory deadlines being missed, inaccurate data being reported, etc. If all of the regulatory fines for a company were added up, it would most likely be significant. Reducing that cost by 50% would be very good.

The HSE aspects of risk reduction due to good data management spawn mainly from the fact that having the right data provides a better platform for people to do the right thing. Often, when dealing with HSE matters, speed is important and good data management will facilitate getting the data to the right people quickly.

Finally, better data management will clearly lower risk associated with decisions. It has been seen many times that having the right data at the right time allows intelligent people to make the right decision. Therefore, when looked at in a holistic sense, this decrease in risk and decrease in cost of risk mitigation will be significant.

**The Poor Data Management Case -**

What is poor data management? The simplest definition would be the opposite of good data management. It is not the intention of this paper to go into details as to what good data management is. There have been many papers and presentations contrasting good data management practices with poor ones in our upstream E&P space. For our purposes, we will state that poor data management is primarily a lack of thought as to the integrity, safety, and leveragability of data. Data of unknown

quality, lineage, and history is siloed. It is not managed properly, not stored properly, and not readily available, not searchable, not able to be browsed, and not shared. There is no data governance or stewardship.

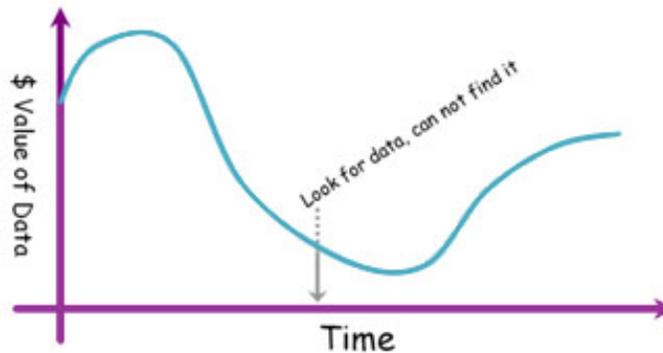
In general, in a qualitative sense, you would expect poor data management to do bad things to the overall asset value of a data asset. As we have all seen when there is a lack of data management associated with your data, the data is harder to use and you would especially expect the “value” of the data usage with poor data management to be less than the base case. The case can be made that sometimes it is better to have no data at all than to have data of poor quality or unknown quality or data that hand cuffs you because you don’t know if you can trust it.

Cost of Data – Poor Data Management Case –

Poor data management will take the base case and increase the cost of the data asset. In our example and our analysis, the overall cost of the data went up 50% due to poor data management. Although the cost to maintain the data went down significantly (33%) the other costs went up. Since the data is not well managed (see definition above) the cost to use and leverage the data goes up. Also, as suggested above, since the data is used so poorly, is difficult to find, and cannot be trusted, the cost of decisions also goes up.

Value of Data – Poor Data Management Case –

More than any other factor, poor data management greatly decreases the value of data. In fact, it can be said that in certain circumstances the data should not even be used. In our example and analysis, the overall value of the data was decreased by about 60% below the value of the base case. The value of decisions was shown the greatest loss – poorer data means that there will be more wrong decisions made and the overall value of those decisions is therefore reduced. The time-value of data is also reduced as there is a high probability that when the data is wanted in the future, it will not be able to be found or will not be otherwise usable.



The performance and integration value of the data is reduced for similar reasons. Poorly managed data is harder to integrate and can actually decrease productivity.

Risk Attribute of Data – Poor Data Management Case –

More than any of our other components, poor data management affects risk in a negative way. Poorly managed data greatly increases the risk that “something will go wrong.” Regulatory risk goes up due to the increased probability that data will be lost and, therefore, not ready to satisfy regulatory requirements. HSE factors go way up. This is a high risk scenario when data is poorly managed. Again, many examples come to mind, but a simple one would be where records for equipment maintenance are lost and equipment fails, causing

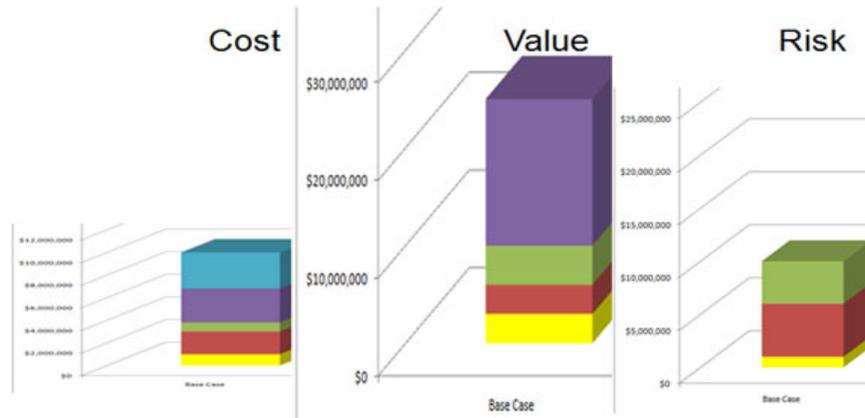
environmental or safety issues. Finally, risk on decisions in the presence of poor data management is higher.

### Quantitative Details:

#### The Base Case -

#### Base Case

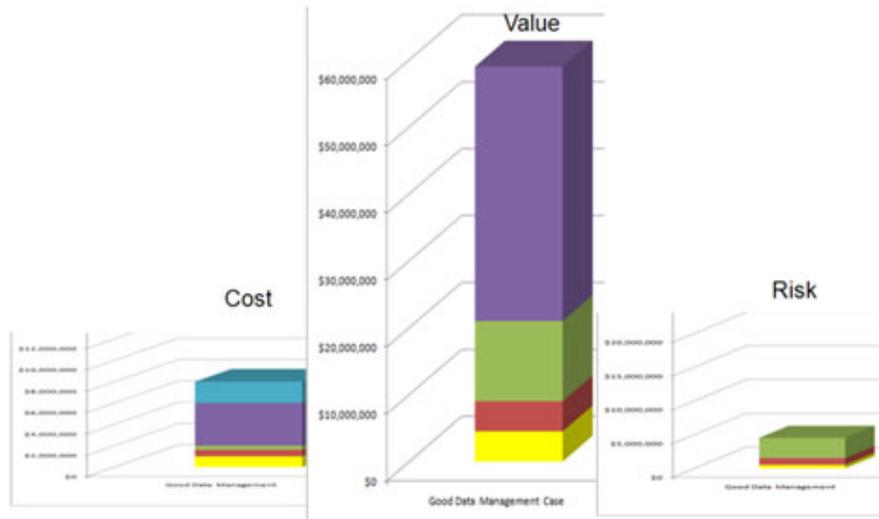
Cost	Cost to acquire	Cost to use & leverage	Cost to replace	Cost to maintain	Cost of decisions
	\$1,000,000	\$2,000,000	\$800,000	\$3,000,000	\$3,200,000
Value	Time value	Performance Value - improved productivity	Integration Value - relevance & applicability		Decision Value - data usage
	\$2,000,000	\$2,000,000	\$4,000,000		\$11,000,000
Risk	Regulatory risk - data usage		EH&S risk - data usage		Decisions risk - data usage
	\$1,000,000		\$5,000,000		\$4,000,000



#### The Good Data Management Case -

### Good Data Management Case

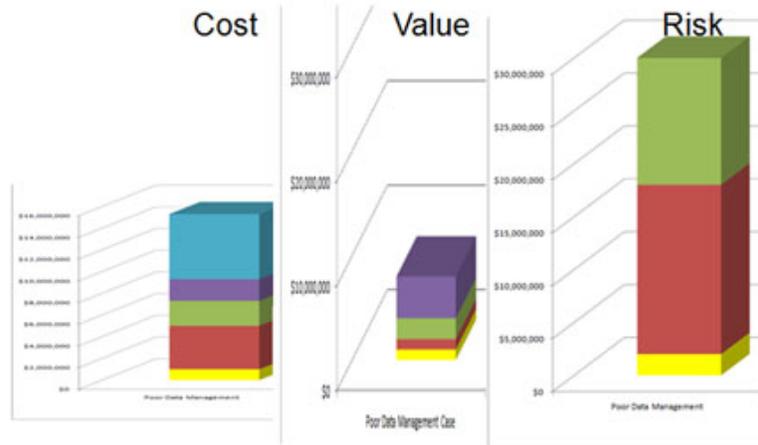
Cost	<b>Cost to acquire</b>	<b>Cost to use &amp; leverage</b>	<b>Cost to replace</b>	<b>Cost to maintain</b>	<b>Cost of decisions</b>
	\$1,000,000	\$600,000	\$400,000	\$4,000,000	\$2,000,000
Value	<b>Time value</b>	<b>Performance Value - improved productivity</b>	<b>Integration Value - relevance &amp; applicability</b>		<b>Decision Value - data usage</b>
	\$4,500,000	\$4,500,000	\$12,000,000		\$38,000,000
Risk	<b>Regulatory risk - data usage</b>		<b>EH&amp;S risk - data usage</b>		<b>Decisions risk - data usage</b>
	\$500,000		\$1,000,000		\$3,000,000



### The Poor Data Management Case -

### Poor Data Management Case

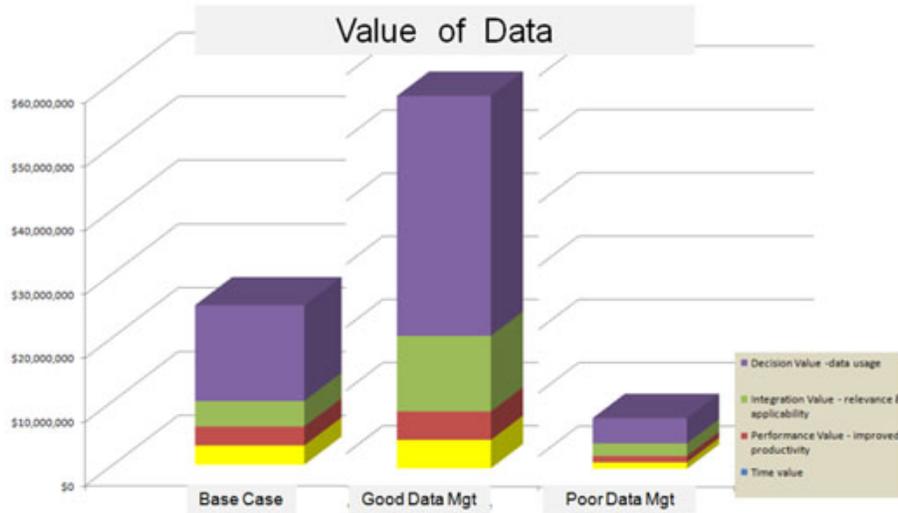
Cost	<b>Cost to acquire</b>	<b>Cost to use &amp; leverage</b>	<b>Cost to replace</b>	<b>Cost to maintain</b>	<b>Cost of decisions</b>
	\$1,000,000	\$4,000,000	\$2,300,000	\$2,000,000	\$6,000,000
Value	<b>Time value</b>	<b>Performance Value - improved productivity</b>	<b>Integration Value - relevance &amp; applicability</b>		<b>Decision Value - data usage</b>
	\$1,000,000	\$1,000,000	\$2,000,000		\$4,000,000
Risk	<b>Regulatory risk - data usage</b>		<b>EH&amp;S risk - data usage</b>		<b>Decisions risk - data usage</b>
	\$2,000,000		\$16,000,000		\$12,000,000



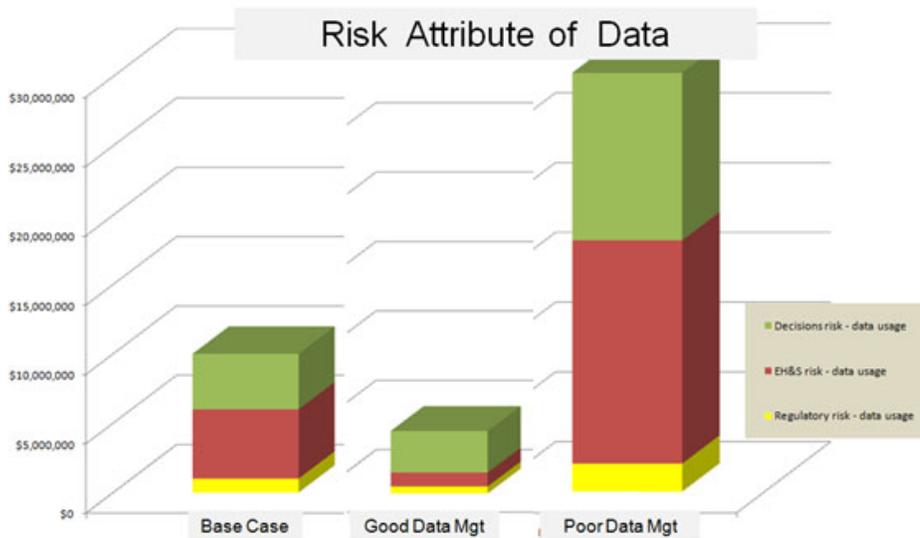
Comparing the Cost of Data Between the 3 Cases:



Comparing the Value of Data Between the 3 Cases:



Comparing the Risk Attribute of Data Between the 3 Cases:



### Test Thesis and Methodology with Actual Data:

Hess has not given permission to release actual data into the published paper. We will try to get approval to make it part of the presentation before the conference.

### Summary and Conclusions:



Using some simple assumptions, data value can be quantified from known parameters. Based on these parameters and these assumptions, it is clear that the value of data far outweighs its cost.

Even more apparent in our analysis is how good data management is worth much more than it costs and the resultant "net asset value" makes the implementation of good data management practices a "must." Due to the potential risks involved in E&P and the criticality of data to mitigate risk, poor data management "can" actually create a negative value when all factors are considered.

This is not meant to be a definitive treatise on the subject as much as it is meant to be thought provoking. Although there may not be general agreement in the actual factors or numbers involved, the anticipation is that this can be used to solidly demonstrate data can and should be thought of as an asset. We hope this causes you to leverage this same methodology to calculate a net asset value for the data at your company.

Why do we jump through so many hoops to justify good data management practices and implementations? ***It is a no brainer!***