Establishing a Business Value for Data

Problem Statement

Information located in different parts of organizations is presumed to be of immense value, but there is currently no generally accepted approach to objectively measure as a financial asset. Knowing the actual value of information assets within an organization can lead to a better understanding of the more valuable and less valuable information. This in turn can lead to better decision making in terms of IT investment and strategy.

Proposed Methodology

This proposal offers an approach to valuing information which is both practical to apply and consistent with accepted accounting principles. The method is theoretical and based on a modified version of the research work done by Moody and Walsh [MO]. Concepts from the research paper ñInformation Assets and their Valueò by Wilco Englesman [WE] was also leveraged for this methodology.

The fundamental concept is to use a decision tree to build a relative index of business value based on economic attributes and maturity measures to guide the respondent to which valuation paradigm would be the best fit for the information asset.

The valuation model ultimately chosen will be one of the three major paradigms in accounting theory: Cost (Historical Cost), Market (Current Cash Equivalent) or Utility (Present Value).

A core premise to this model that differs from the work of Moody and Walsh [MO] is that the maturity of information - in terms of readiness to apply to a specific business problem, greatly affects its value. At its most immature, information is just data. Without a specific use case defined, attempting a utility valuation becomes highly subjective and as a result would be a better fit for a modified historical cost valuation. This is inline with the Moody and Walsh [MO] concept that information that is not used (or ready to be used) will have a cost but no value.

On the other hand, information that has been data mined and modeled to solve a specific business problem or objective can be more readily quantified in terms of benefits that can be derived from it, in terms of future cash flows. The value of this mature information is better measured using the utility (present value) method, as its intended use (and impact to the business) is better understood and quantified.

This concept is also in agreement with Cronin and Gudim position (as outlined by Moody [MO]) that information has no intrinsic value and that value depends on its context and its use by particular users on particular occasions.

This proposal suggests leveraging the Cross Industry Standard Process for Data Mining (CRISP-DM) to assist in determining the maturity of the information to be valued. As part of the decision tree that examines information attributes, the six phases of data mining are examined to determine the readiness of information specific to business use cases.

Wilco Englesman [WE] asserts that Jeff Wilkins, Bert Van Wegen, and Robert de Hoog, in their ñUnderstanding and Valuing Knowledge Assets: Overview and Methodò that the added value of a knowledge asset can be determined by systematically propagating the requirements of the product that are relevant to the value added, to knowledge asset requirements. ñWith the relevance of each product requirements, and the contribution of each knowledge asset, the portion of the added value of that product that is achieved by the knowledge asset can be estimatedò. [WE]

The difference between ñknowledgeò asset and ñinformationò are minimal and both can be discussed interchangeably. [WE]. The assertion of the author of this paper is that maturity of information directly correlates to the ability to accurately quantify the degree in which an information asset contributes to the value add of a product.

The valuation ñdecision treeò as described in this paper leverages the concept as described by Englesman [WE] (citing research done by Gunnarson and Steinarrossen) that in order to value information in a successful manner, it ñrequires a complex and extensive operational focus on how the information is used, how frequently and how users rank (value) the information, within the organization as well as how information affects different organizational functions.

The ñdecision treeò provides the respondent (the information user) a structured way to rank the value of the information at hand (based on economic attributes) so that strategic IT investment decisions can be made based on the estimated extent that the given information will increase revenue or reduce cost in the future.

Step 1: Information Assets – Attribute Baseline

The first step in the valuation process is to create a baseline for the importance of particular attributes for a given information type. This valuation will be specific to a particular organization based on their business activities. The definitions of the information categories below are detailed in the paper ñThe attributes of information as an asset, its measurement and role in enhancing organizational effectivenessò by Charles Oppenheim and Joan Stenson. [OS].

The attributes indicated in the left column are recommended in the Englesman paper [WE] and are defined in the next section.

Weights from 1-5 should be assigned by an internal individual who has responsibility for strategic and short term decision making about information (i.e. Finance Director). A ranking of ñ1ò should be considered a minor attribute while a ranking of ñ5ò would be considered critical for the specific information type.

The assigned weights will be used as a decision factor as part of the ñdecision treeò used to evaluate the value of specific information (dataset).

	Customer Information	Competitor Information	Product Information	Business Process	Management Information	People Management	Supplier information	Legal and Regulatory	Organizational Information
Usage									
Timeliness									
Accuracy									
Shareable									

Note: The information categories (column headers) in the table above were based on research work by Chales Oppenheim and Joan Stenson [OS] as an attempt to standardize differing information types. This baseline exercise can be easily adjusted to support the business data types as outlined in the Innocentive Challenge requirements (below).

	In-Service Product Data	Customer Data	Supply Chain Data	Maintenance/Repair Overhaul Data	Financial Data
Usage					
Timeliness					
Accuracy					
Shareable					

The weights assigned represent the relative importance of each attribute type when describing a particular type of information in terms of value to the business as a whole.

Once the rankings are completed the next step is to calculate a weighted average indicating the relative importance as a percentage for each attribute. That is accomplished by summing the attribute rankings and then dividing the sum of the rankings by the individual ranking for each category.

As an example, a Finance Director of a firm ranked ñIn-Service Product Dataò to indicate the relative importance of each attribute as detailed below:

	In-Service Product Data
Usage	5
Timeliness	3
Accuracy	5
Sharable	2

In this example the ability to use the data is ranked indicating critical importance. The fact that the ñsharableò attribute is ranked ñ2ò could indicate that usage is extremely important for key business units, but not necessarily for the majority.

The sum of the attribute rankings is 15. The weighted average for each is the following:

	In-Service Product Data	Ranking/Sum of Ranking	Weighted Average
Usage	5	5/15	.33
Timeliness	3	3/15	.20
Accuracy	5	5/15	.33
Sharable	2	2/15	.13

The weighted average will be used in a later step when evaluating the value of a specific dataset.

Information Attribute Definitions

These definitions below outline why the information attributes have economic value.

Usage: According to the ñSecond Lawò as outlined in the Moody paper [MO], ñThe Value of Information Increases with Useò and implies that information exhibits ñincreasing returns to useò where the major cost of information is in its capture, storage and maintenance while the marginal costs of using it are almost negligible.

When evaluating usage itôs important to weigh the costs of not using the data. Moody [ME] argues that unused information is really a liability. The cost of capture, maintenance and storage of the data type should be reviewed here. Also Moody [MO] mentions four prerequisites for using data:

- knowing where it exists
- knowing where it is located
- having access to it
- knowing how to use it

Timeliness: According to the ñThird Lawò as outlined in the Moody paper [MO], ñInformation is perishableò and has a shelf life. The speed in which information loses value depends on the type of information.

Accuracy: According to the ñFourth Lawò as outlined in the Moody paper [MO], ñThe value of information increases with accuracyò. The accuracy required is highly dependent on the type of information. Inaccurate information can be costly to organizations in terms of incorrect decision making and operational errors.

Shareable: According to the ñFirst Lawò as outlined in the Moody paper [MO], ñInformation is (Infinitely) Shareableò and implies that information value increases with usage across the organization. From the firmôs perspective, value is cumulative as it is used by different users or groups across the organization.

Step 2: Information Asset Valuation Flowchart

Once a baseline has been established that provides a weighted average for specific attribute types for different information categories, the next step is to evaluate how the specific information asset to be valued will be used. This is done by consulting the ñInformation Asset Valuation Flowchartò.

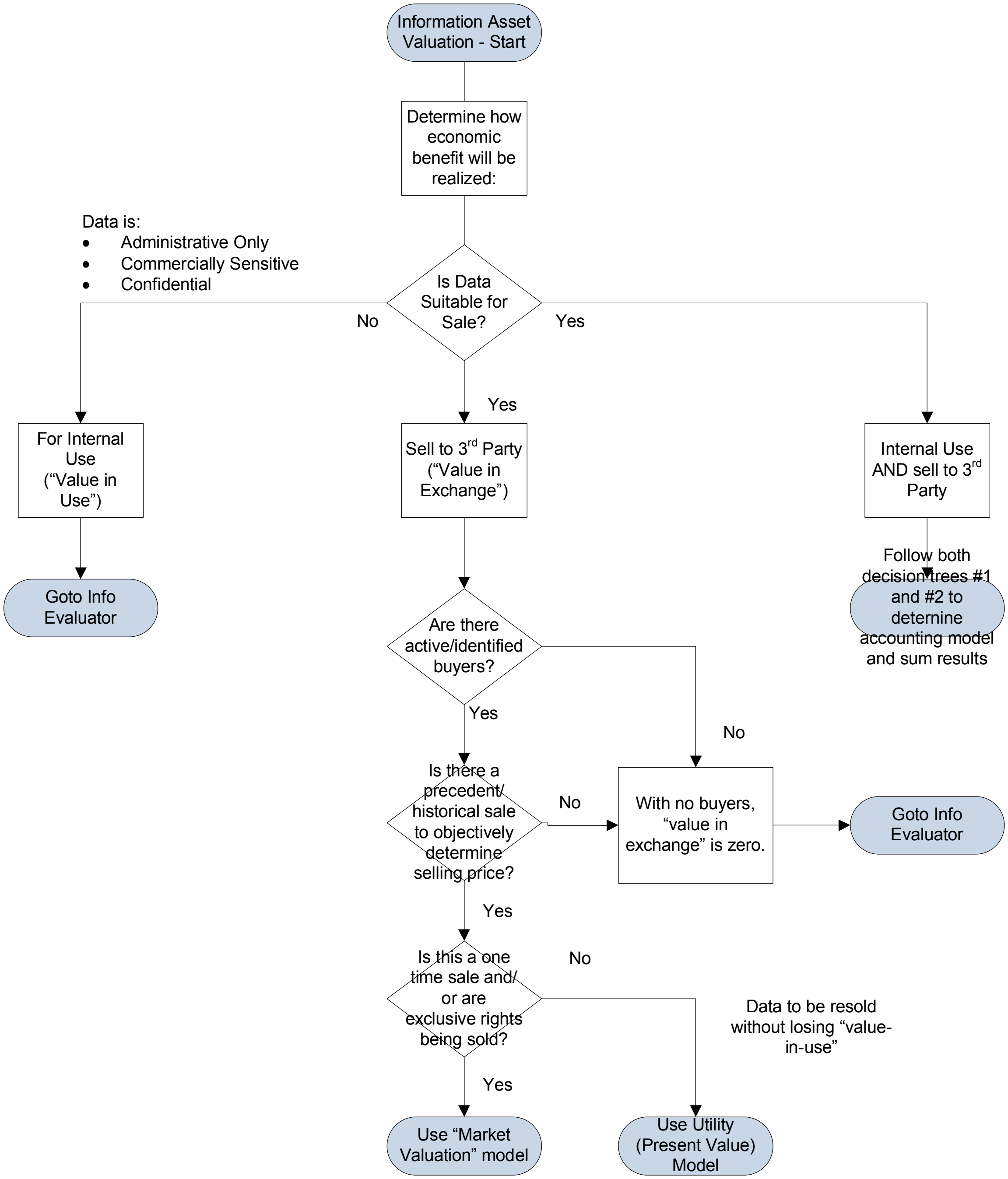
The first step is determining whether the data in question will be used internally (value in use), externally (value in exchange) or a combination of both. In the case where the information will be used both internally and externally the value calculation should be summed. The argument made in this paper is that a market valuation model or a utility (present value) model can be used in cases where the information is to be sold to external parties.

If information is targeted to be sold, but there are no identified buyers, the ñvalue in exchangeò is zero. In this case, the information needs to be further analyzed via the ñinformation evaluatorò (or decision tree) to determine the best fit valuation model.

If thereôs an identified external buyer for the information and itôs a one time sale, the information can most likely be valued by the estimated selling price, therefore using a market (current cash equivalent) model. If the data can be resold without losing ñvalue in-useò, itôs possible to estimate future cash flows if thereôs a precedent selling price. In this case a utility (present value) model can be used.

If the information is to be used internally, the information needs to be further analyzed via the ñinformation evaluatorò (or decision tree) to help determine the correct validation paradigm.

Information Asset Valuation – Flowchart



Step 3 – Information Evaluator "Decision Tree"

If a valuation model was not determined by the flowchart, the next step in the valuation process is to have the data owner walk through a decision tree to determine the best fit model based on economic attributes and maturity. (Maturity is the dimension that measures how fit the information is for a specific business problem or objective).

The end result of the decision tree described below is a determination of whether to use the modified historical cost valuation model as outlined by Moody [MO], or a utility (present value) model. Moody [MO] asserts that conceptually the utility model is the best approximation to the true economic value of an asset, but unfortunately in most cases its highly difficult to estimate future economic benefits. This proposal offers a differing view in that information that is mature and specific to a business use can be quantifiably valued in terms of impact to revenue or cost savings (estimated future economic benefit). The premise being that data sitting on a shelf without a clear purpose does not have as much value compared to data that can be easily leveraged to make a quantifiable business impact.

The Cross Industry Standard Process for Data Mining [CRISP-DM] is a data mining process model that describes commonly used approaches that data miners use to tackle business problems. The process model describes the journey that information takes from raw data to business knowledge thatõs specific and actionable. The process can be cyclical as information can generate additional information of value, but it provides a useful framework to determine value at a given point in time.

After evaluating information maturity, an additional exercise is performed to calculate the ñvalue indexò of the information asset. This is done by comparing the attribute baseline for an information category against a quality ranking of a specific information asset. This provides a relative value index that can be used when evaluating investment objectives.

Decision Tree Steps:

Step 1:
The first step is to perform a quality ranking on the specific dataset to be valued following a similar process as the baseline. The baseline describes the relative importance for a given economic attribute. For the specific dataset to be valued, rankings are based on the actual characteristics of the data itself.

For the same economic attributes as described in the baseline process, the data owner is to rank the quality of the given information from a scale of 1 to 5. A ranking of 5 would mean that the data perfectly aligned with its given attribute. Lower rankings indicate deficiencies.

For example, a ranking of ñ1ò for the ñUsageò attribute would likely indicate that thereõs very little ability to actually access the data in question.

Once the dataset is ranked based on quality, the next step is to evaluate each ranking based on the weighted average that was determined for the generic information class during the baseline exercise. This is done by multiplying the weighted average for an attribute by the ranking that was given to the specific information set for the same attribute.

The result is a relative index that can be used to objectively compare the value of information across datasets.

Information of high value will approach a perfect score of ñ5ò indicating high quality.

Example 1:
As an example using the same ñIn-Service Product Dataò weighted average from the baseline exercise, the next step would be to compare against In-Service data for a specific product.

Below is the weighted average for the In-Service Product Data category as calculated during the baseline exercise:

	In-Service Product Data	Ranking/Sum of Ranking	Weighted Average
Usage	5	5/15	.33
Timeliness	3	3/15	.20
Accuracy	5	5/15	.33
Sharable	2	2/15	.13

In this example we have a line manager who is responsible for a specific product (a water pump) and owns a dataset. He ranks the ñqualityò of his dataset as follows:

	In-Service Product Data (Water Pump)
Usage	1
Timeliness	5
Accuracy	5
Sharable	1

Given the baseline weighted average for each attribute, the relative value of this dataset can be calculated as follows:

	In-Service Product Data (Water Pump)	Weighted average of each attribute as determined during baseline process	Value Scoring
Usage	1	.33	.33
Timeliness	5	.20	1.0
Accuracy	5	.33	1.65
Sharable	1	.13	.13

Total value = 3.11 (out of 5 maximum).

In a perfect scenario the information value score would be a 5; information with deficiencies (lower accuracy, difficult to access) will result in a lower score.

The result of this process is to determine the degree of variance between the desired index value (5) and the calculated value based on data quality.

Example 2:
During the baseline exercise the financial director of a firm weighted the generic ñMaintenaince/Overhaul/Repairò data category as follows:

	Maintenance/Overhaul/Repair	Ranking/Sum of Ranking	Weighted Average
Usage	2	2/15	.13
Timeliness	5	5/15	.33
Accuracy	5	5/15	.33
Shareable	3	3/15	.20

His rankings imply that timeliness and accuracy are critical to this category of data.

A line manager owns a dataset that shows repair frequency for components that belong to a specific product. The data is accurate, but thereôs typically a lengthy delay before itôs accessible.
His weighting for the data set might look like the below:

	Product x - Maintenance/Overhaul/Repair
Usage	2
Timeliness	2
Accuracy	5
Shareable	3

Given the baseline weighted average for each attribute, the relative value of this dataset can be calculated as follows:

	Product x ï Maintenance/Overhaul/Repair	Weighted average of each attribute as determined during baseline process	Value Scoring
Usage	2	.13	.26
Timeliness	2	.33	.66
Accuracy	5	.33	1.65
Sharable	3	.20	.60

Total value = 3.17 (out of 5 maximum).

Step 2:
After determining the value index for the information in quesiton, the next step is to determine the maturity of the information asset based on the CRISP-DM data mining phase. This is done to provide a way to measure how fit the data is for actual use. Proceeding through these phases results in a better understanding of the data set, and as a result, the attribute weights first assigned might need to be changed along the way. Over time information which was initially graded with a low value might become more valuable and warrant further IT investment. The phases are outlined below:

1) Business Understanding: This is the least mature level ï there are business requirements applicable to the information asset, a problem definition regarding how the data can be exploited and a preliminary plan. *In this case the modified historical cost model will need to be used as the economic benefit can not be accurately estimated as attributes such as data quality (accuracy) and usage might not be known.*
2) Data Understanding: During this phase activities are being done in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypothesis for hidden information. *In this case most likely the modified historical cost model will still need to be used as the economic benefit can not be accurately estimated as attributes such as data quality (accuracy) and usage are still not entirely known.*
3) Data Preparation/Modeling: During these phases activities are being done to construct the final data set and model for a specific business problem. Tasks might include translating and cleaning the data along with the application of modeling techniques to extract value. This phase represents data that is not quite ready for evaluation/deployment, but thereôs an understanding of the specific problem that it can solve. *At this point the information can be valued using either the modified historical cost or utility (present value) model if expected future benefit can be quantified.*
4) Evaluation: During this phase data analysis models have been built that achieve business objectives. A key objective during this phase is to determine if some important business problem has not been considered. At the end of this phase decision to use the data should have been reached. *It's the assertion of this author that a utility (present value) model can be used to value the data at this point. If the data is mature enough to have been modeled to answer a specific business problem, then there should be an ability to estimate the impact in terms of future revenue increases or cost savings.*
5) Deployment: At this phase it is understood as to what actions will be needed to be carried out in order to make use of the information. *As in the evaluation phase, a utility (present value) valuation should be possible at this point as the expected future economic benefits can be estimated.*

The table below summarizes the CRISP-DM phases and the corresponding valuation models.

CRISP-DM Phase (Information Maturity)	**Accounting Valuation Model**
Business Understanding	Modified Historical Cost
Data Understanding	Modified Historical Cost
Data Preparation/Modeling	Modified Historical Cost or Utility (Present Value)
Evaluation	Utility (Present Value)
Deployment	Utility (Present Value)

Step 3:
The last decision step is to merge the proposed valuation model with the index value that was determined during the attribute ranking exercise. The end result is information value represented in two dimensions ï quality (based on economic attributes) and maturity (in terms of readiness to be used for a specific business objective).

In summary, the dataset value calculation (based on attribute rankings) can be viewed as a relative index of ñvalueò. Coupled with a accounting valuation model based on data maturity and/or the asset evaluation flowchart provide a method to objectively value information assets when making investment decisions.

Valuation Models

Modified Historical Cost
Moody [MO] suggests using a modified version of the historical cost valuation method as the standard model does not incorporate any concept of use ï that information that is used would be valued the same as information not used when the collection costs are the same. Aspects of the modified historical cost model are detailed in his research paper, but the primary benefit is that the laws of information are considered to provide additional weight to information that is used by more people and in multiple ways.

Utility (Present Value)
Moody [MO] argues that ñthe utility valuation method is theoretically the best valuation approach in terms of validity but is impractical to apply in practice, because of the cost of applying it and the subjective nature of the results (reliability).ò The argument made in this proposal is that the utility valuation method can be used reliably if the information is mature in that it has been modeled or fit for a specific business problem. In that case, costs to maintain the data can be weighed against the potential economic benefit in terms of cost reductions or revenue gains to provide a valuation.

Use Case Examples
In theory, this proposed methodology can be applied across different business models as the evaluation is not tied to specific information types.

For example a company that sells DVD subscriptions might be interested in evaluating prior customer purchases to develop a recommendation engine with the goal of increasing sales. The ñcustomer purchaseò data can be represented as ñcustomer informationò in the attribute baseline/ranking exercise. The ñAsset Valuationò flowchart can be consulted to determine the correct valuation modeling technique to use. The customer data can be compared to the CRISP-DM phases to determine the maturity of the information in terms of readiness for deployment.

In another example an airline manufacturer might use the methodology to justify the investment in increased inspections of planes during production/overhaul cycles.

In this case, the financial director would need to perform a baseline ranking indicating importance of the economic attributes (usage, timeliness, accuracy and shareable) for the ñMaintenance Repair and Overhaul Dataò information category. A weighted average is then calculated based on the importance rankings.

Following the ñInformation Asset Valuation Flowchartò itôs determined that the maintenance data represents an internal use case (value in use) and will require analysis of the specific information asset.

In this case, a line manager owns repair history for all mechanical components for a specific type plane for the last 5 years. The data is very accurate, and timely, but is not in a format that is readily usable. When he performs a ñqualityò ranking and calculates a ñvalue indexò using the weighted averages from the baseline, itôs determined that the data has an index value of ñ4ò because accuracy and timeliness is deemed critical for this particular information asset.

Consulting the CRISP-DM phases itôs determined that the data is not mature in terms of being ready for use. The data is in the ñdata modelingò phase as a ñvalueò is at least theorized, the accuracy is known, but needs to be prepared and modeled prior to use. Given this insight, the Moody [MO] modified historical cost model is chosen as a valuation method. Given the relatively high index value of ñ4ò and an accounting method to determine the asset value, a decision can be objectively made to justify further investment.

References

[CRISP-DM] Cross Industry Standard Process for Data Mining. *http://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining*

[MO] Daniel Moody and Peter Walsh. Measuring The Value of Information: An Asset Valuation Approach. *http://wwwinfo.deis.unical.it/zumpano/2004-2005/PSI/lezione2/ValueOfInformation.pdf*

[OS] Charles Oppenheim and Joan Stenson. The attributes of information as an asset, its measurement and role in enhancing organizational effectiveness. *http://www.libqual.org/documents/admin/stenson.pdf*

[WE] Wilco Engelsman. Information Assets and their Value. *http://referaat.cs.utwente.nl/conference/6/paper/6807/information-assets-and-their-value.pdf*