

Checkpoint Contents

Accounting, Audit & Corporate Finance Library

Editorial Materials

Cost Management

Cost Management [formerly Journal of Cost Management]

Archive

March/April 2015

CAN YOU QUANTIFY THE VALUE OF YOUR DATA?

CAN YOU QUANTIFY THE VALUE OF YOUR DATA?

JEFF LAWTON

JEFF LAWTON is a director of IT in the Global Public Sector division of Grant Thornton LLP. He received his MBA from the University of Georgia and B.A. from the University of Virginia. He holds certifications as a Project Management Professional, Certified Business Intelligence Professional, IT Infrastructure Library, and Agile Scrum Master. For 20 years with Grant Thornton and other Big 4 consultancies, he has led data quality assessment and remediation projects, data warehouse and business intelligence projects, and systems integration projects that solve problems for public-sector agencies. He leads the Intelligent Data Quality Management special interest group at CAM-I.

The IDQM working group is addressing the problem of poor data management by creating an equation that places a dollar value on any record of data.

The use of data has evolved. At first data were used to record an event or a state of being for future reference. Now data can serve multiple purposes, such as improving business performance, complying with legal mandates, detecting improper payments, and making wiser choices with regard to business partners and customers. We are getting better at making data sweat, so the intrinsic value of data as an asset is increasing. Yet it is the most poorly managed asset in every organization. Why is this? Can we do better? The Consortium for Advanced Management International (CAM-I) Intelligent Data Quality Management (IDQM) working group is tackling this very problem by creating an equation that can place a dollar value on any record of data. And by knowing the value of a record, intelligent decisions regarding its management can be made.

Data are a corporate asset that reduces uncertainty about decisions, affects behavior, and can even have its own market value. When data are compromised by quality problems, not only does the value of data as an asset decrease, but there can also be more extensive consequences (e.g., cost efficiency, risk management, regulatory compliance, agility, revenue growth, readiness).

According to a study performed by Gartner in 2011, poor data quality was a primary reason for why 40 percent of all business initiatives failed to achieve their targeted benefits, and data quality can also affect overall labor productivity by as much as 20 percent. Most importantly, Gartner predicted that as more business processes become automated, data quality will become the rate-limiting factor for overall process quality. ¹ This means that no matter how well we improve the performance of a process, eventually we will be constrained by the data that govern that process.

This certainly warrants further investigation. A problem with data quality indicates that there is a problem with the data's ability to be fit for the intended purpose, which in turn means that an organization's asset is broken and needs to be remediated. While this sounds straightforward, data's unique combination of characteristics defies traditional asset management techniques. These traits include the following:

- easily created;
- easily duplicated;
- easily destroyed;
- easily altered;
- easily transported and transferred;
- easily misunderstood or subject to interpretation;
- easily stolen;
- capable of being aggregated or disaggregated with other data;
- difficult to preserve as storage media changes over time (e.g., reel-to-reel tape, floppy disks, CD-ROMs, thumb drives); and
- loses value with age.

The trouble with data

Perhaps the most compelling differentiator between data and traditional assets is that data do not change color, stink, smoke, rattle, catch fire, melt, or collapse upon failure. Data quality problems are quite difficult to detect; in fact, they can only be found when the data are used by someone who understands well enough to know when the data have failed to adequately capture the real-world condition.

So we need a new approach that can allow us to manage data the same way we manage any other physical asset. The IDQM working group believes that the key is the ability to place a dollar value on any data record to reflect its value to an organization. A data record, in this case, is the collection of data elements that are needed to adequately describe a condition or state of being such that the organization finds it useful. This means a data record can be a virtual concept under which records from multiple tables in a database that describe a single customer are aggregated (e.g., customer attributes, customer address, customer purchase history).

Expressing a record of data in monetary terms will allow us to make asset management decisions about data intelligently. Do we clean the data or live with the problems? Do we upgrade to a new financial system? Do we change our retention policy? All of these decisions can be made by comparing the current value of the data, the projected value of the data after the change, and the costs of implementing the change and by selecting the decision where the value is greater than the cost.

Building the value framework and equation

The mission of the IDQM working group is to create data quality and data management frameworks for well-informed business decision-making, improved investment analysis and allocation of appropriate funding/resources, reduced risk exposure, and controlled improvement. In this regard, our first task was to conduct a survey to determine if a data valuation equation or framework already exists elsewhere in the industry, and we found no indication of such artifacts. Gartner coined this approach as "infonomics" in several presentations, citing the need for such framework but never offering a solution. Strangely, many organizations were able to cite the costs of their data quality on their business but, when contacted, could not provide a reproducible approach for obtaining such figures.

Framework component: Data life cycle.

The IDQM working group began working on the framework piecemeal, beginning with components of the framework likely to have the largest effect on the value of a data record. The first component dealt with the data life cycle, and our hypothesis was that the value of a record of data is a function of where that record exists at this moment in the data life cycle.

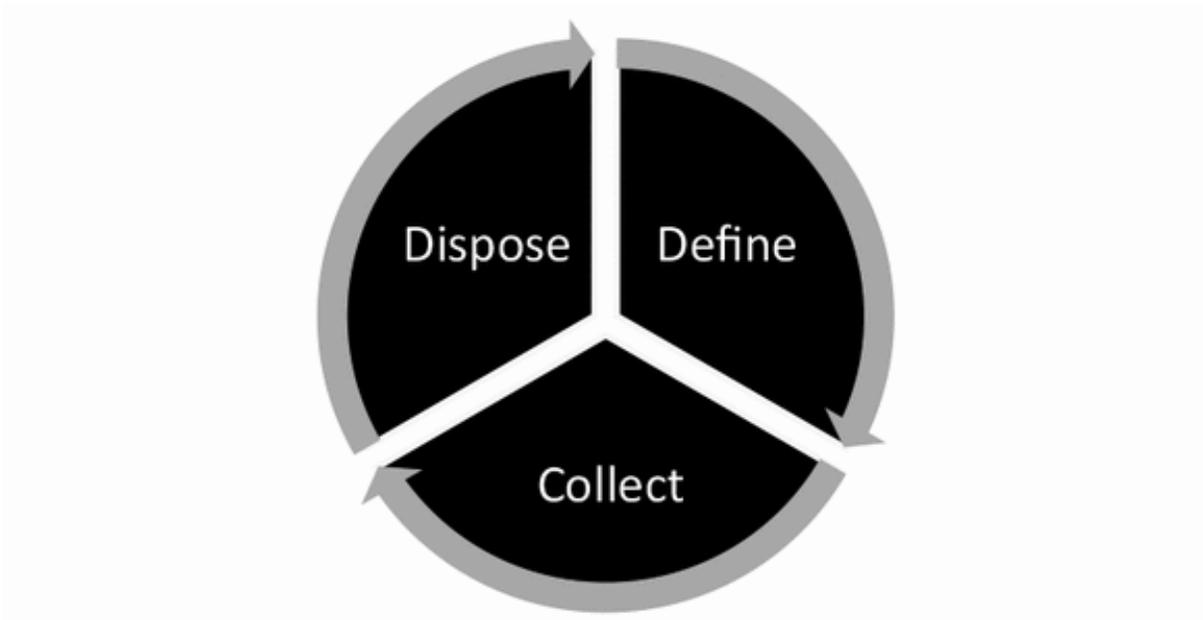
This required having a defined data life cycle. We researched nearly 60 data life cycle models in the industry and dismissed them all for one of three reasons.

- The model was circular in nature, such as the example shown in Exhibit 1. Circular life cycle models imply that the cycle is continuous in nature and that the disposal of data necessarily will lead to the defining of a new record of data. We know from everyday experiences that this is not always the case. Take as an example a customer record. When the customer dies, the record may eventually be purged, but this does not necessarily mean that the record will be recreated again for that customer. Rather, the data life cycle is linear in nature, and when a record of data is purged, that signals the completion of its life cycle. Any recreation of that record constitutes a new instantiation of a linear life cycle for a new record of data.
- The model was too focused on a single industry (e.g., life sciences) to the exclusion of all others. Our goal is to develop a framework that can accommodate any record of data from any system in any organization, so our model must reflect a more universal set of life cycle phases.
- The model was too basic. Our data model must be robust enough to accommodate data management best practices such that valuation can take into account whether or not those

practices are in use.

Exhibit 1.

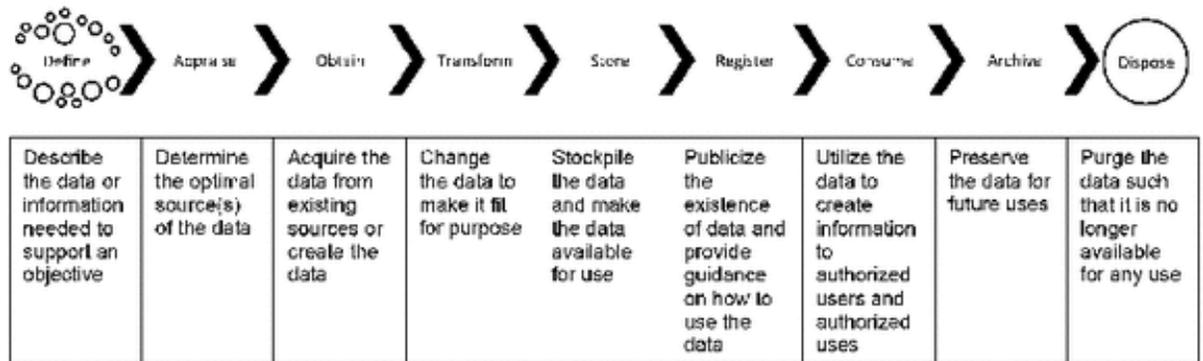
A Typical Circular Life Cycle Model Does Not Work Well for Data



Ultimately, the IDQM working group selected the data life cycle model shown in Exhibit 2.

Exhibit 2.

IDQM Life Cycle Model



We characterized data across a number of activities that are performed in each of these phases, including:

- design and architecture activities;
- governance activities;
- communication and outreach activities; and
- quality assurance and quality control activities.

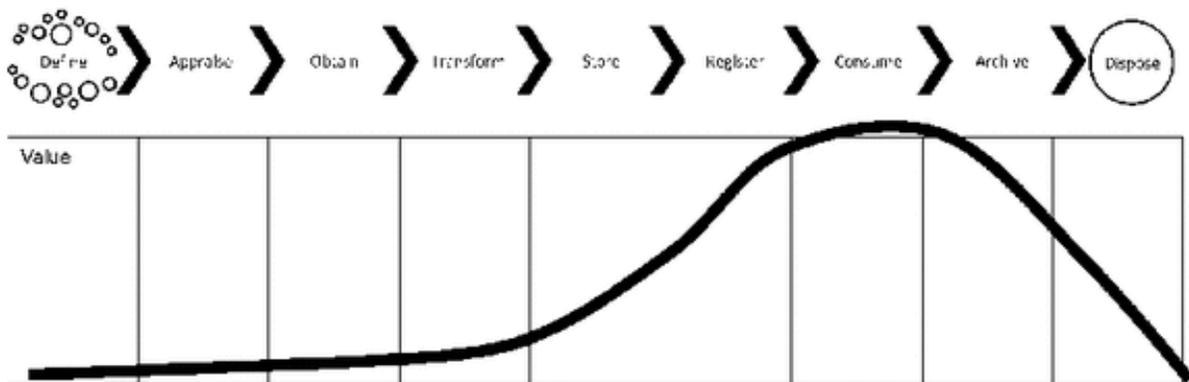
We also characterized data across a number of enablers and how they are used in each of these phases. Examples of these enablers include the following:

- tools;
- standards;
- technical metadata;
- business metadata; and
- change management.

Finally, we explored the nature of valuation concepts, costs, and risks involved in each stage of our life cycle. When examining the end product, it became obvious that the very nature of data, their management concepts, their value, and their costs change dramatically across each of the different phases of our life cycle model, thus validating our hypothesis that where a record exists in the context of its life cycle greatly affects the value of that record. A curve, such as the one depicted in Exhibit 3, can then be generated for the purposes of valuation. At present, the IDQM working group believes that the data record is serving its ultimate purpose at the consume phase and therefore is generating its maximum value (a multiplier of 1.0). Upon disposal, the record no longer exists and thus has no value (a multiplier of 0). At other phases in the life cycle, the value is fractional to that of the consume phase and receives multipliers between 0.1 and 0.75, with an increase beginning at the store phase, where the data record actually exists within the organization in a consumable format.

Exhibit 3.

The Value of Data: Changes Throughout Its Life Cycle



At this point, our data record valuation model appears as such:

$$\text{Value of a Record of Data} = f \left[\begin{array}{c} \text{Position Within} \\ \text{Data Life Cycle} \\ \text{Min 0, Max 1} \end{array} \right] \times \sum (\text{Multiple Other Factors})$$

Framework component: Data management maturity.

The next part of our equation was another component that the IDQM working group felt had a multiplicative effect on the value of a record of data - the level of data management maturity exhibited by the organization. The idea here is that organizations that have strong, mature data management practices, tools, and policies are more likely to be able to do more with their data and adapt more quickly to change than organizations with low maturities. Thus, the higher the maturity, the more intrinsic value a record of data offers an organization.

The IDQM working group first listed common techniques and enablers that are used for data management and narrowed them down to 14 distinct techniques. Each technique was defined and assigned prerequisite techniques, affinity techniques, and typical activities that comprise the technique. An example is provided in Exhibit 4.

Exhibit 4.

An Example Enabler: Metadata Management

Technique	Definition	Affinity to Other Technique Categories	Prerequisite Technique Categories	Activities
Metadata management	Information that describes a data set with regard to definition and domain (e.g., range of valid responses), instructions for usage, design and specifications of data, parameters, quality rules, creation date, associated supporting processes, etc. The purpose of this information is to inform users in the proper and effective use or understanding of the data.	<ul style="list-style-type: none"> • Enterprise data architecture • Data quality • Data governance • Master data management • Data design 	<ul style="list-style-type: none"> • None 	<ul style="list-style-type: none"> • Training on the collection and documentation of good metadata • Collect metadata from source systems • Reuse of external metadata constructs that apply • Create metadata • Create search capabilities • Create presentation mechanism for access • Expose data lineage • Expose business rules • Expose transformations

Based on the affinities and prerequisites, we assigned the techniques to one of the four maturity levels

shown in Exhibit 5. We selected this model so that IDQM would be aligned with the CAM-I performance management maturity model.

Exhibit 5.

CAM-I's Four Maturity Levels

1. Rudimentary	2. Established	3. Effective	4. Adaptive
Non-systematic, non-periodic, and reactive	Stable and repetitive	Internally efficient and continuously improving	Externally efficient and dynamic

After organizing the techniques by maturity level, our model appeared per Exhibit 6. This means that the data management maturity of an organization must be at a certain level of maturity in order to use a technique effectively.

Exhibit 6.

Data Management Enablers by Maturity Level

Enablers/Technique Categories	Controls											
	Data Governance	Data Quality	Metadata Management	Process Design	Data Warehouse and Business Intelligence	Enterprise Data Architecture	IT Portfolio Management	Data Innovation	Data Mining	Master Data Management		
Information Management	L1	L1	L1	L1	L1	L2	L2	L2	L3	L3	L3	

As such, the presence of successfully implemented techniques provides an indication of an organization's data management maturity. For example, if an organization has successful data warehouse, IT portfolio management, and data governance programs in place but no data innovation, data-mining, or master data management programs, that organization is likely at maturity level two (established). Knowing where an organization is with regard to data management maturity allows us to plot the multiplier effect along a curve that looks something like that in Exhibit 7, where the organizations with higher maturities are more likely to extract greater value from a record of data than organizations at lower maturities. The slope of the curve is likely to increase but not necessarily at a logarithmic rate, so a

Fibonacci sequence of one, two, three, five for levels one through four, respectively, seems to align well.

Exhibit 7.

Changes in Data Value as Maturity Increases

1. Rudimentary	2. Established	3. Effective	4. Adaptive
Non-systematic, non-periodic, and reactive	Stable and repetitive	Internally efficient and continuously improving	Externally efficient and dynamic
Multiplier: 1	Multiplier: 2	Multiplier: 3	Multiplier: 5

Similar to the data life cycle, the maturity level is likely to exhibit a multiplicative effect. Our data record valuation model now appears as such:

$$\text{Value of a Record of Data} = f \left[\begin{array}{c} \text{Position Within} \\ \text{Data Life Cycle} \\ \text{Value} = \text{Min } 0, \text{ Max } 1 \end{array} \right] \times f \left[\begin{array}{c} \text{Assessed Data} \\ \text{Management Maturity} \\ \text{Value} = \left\{ \begin{array}{l} \text{Level 1} = 1 \\ \text{Level 2} = 2 \\ \text{Level 3} = 3 \\ \text{Level 4} = 5 \end{array} \right\} \end{array} \right] \times \sum (\text{Multiple Other Factors})$$

Framework component: Other factors and future directions for IDQM.

Over the past two years, we have noted more than 125 other possible components of this framework that will need further exploration. It is likely that some of these components will have a significant effect on value, and some will be insignificant. Examples include the following:

- politics;
- costs of noncompliance;
- culture;
- opportunity as a function of total customer value;
- acceptable tolerance of invalid information;
- purpose/use of data;

- number of copies of data;
- value as a function of potential uses/innovation potential;
- degree of usage;
- degree to which reputation is threatened; and
- tie to core business function.

Summary

The era of big data and an information-driven economy are exciting times for organizations, and the successful ones will be those that are able to make the best decisions in the shortest time that drive the greatest value at the least cost. To do this, we must manage our data as carefully as we do any other asset. The possibilities for being able to do so are compelling. Consider these decision-making events in a future with an equation that can derive the monetary value of data.

- A data quality problem is detected that causes a mistake in the shipping of a product to the wrong customer address. The reaction from the top is to get the data cleaned to prevent further mistaken deliveries. The cost of cleaning the address data records is estimated in manpower hours to be the equivalent of \$2 million. The value of those address data records is calculated to be \$1 million, and the cost of the bad records is calculated to be \$100,000. The decision to live with the bad data saves \$900,000.
- An organization contemplates migrating to a different financial management system. It calculates the value of all the records in the current system at \$100 million, while the value of those same records in the new system would be \$200 million because of the additional features the new system offers. The cost of moving to the new system is \$50 million. The decision is made to migrate because it offers a net \$50 million in value.

In summary, data are an asset with unbelievable untapped potential, but because data can be tricky to manage, a special approach is required. Data quality is a problem that won't go away and will only increase as data volume and variety increase. **2**

1 Friedman, T. and Smith, M., Measuring the business value of data quality (Gartner, 2011).

2 The IDQM has operated as an active CAM-I interest group since September 2012. Jeff Lawton has served as the chair of the interest group. Key contributors to the interest group include Mr. Don Carlson of Bank of America, Ms. Nancy Lashbrook of Boeing, Mr. Mike Lanouette of Definitive Logic, and Ms. Pamela Henderson of the U.S. Army.

