By DAVID McGOVERAN

Valuing Data, Part 2

Some writers seem to think that data is free. At least, the claim is that data is worthless. Such conclusions are a bit strong and appear to be based on unclear reasoning about the relationship between data and information. These philosophers pass off the obvious counterexample of data as commodity on its own, claiming that only information — data interrelated and used — has value. We need merely observe that nothing has value unless it has utility (however abstract) to begin to understand the source of the "no value" analysts' confusion. Such arguments reduce to the claim that data has no intrinsic value, only utilitarian value. However, like any capital asset, data is a factor in the production of goods and services and so has utility. The simple conclusion is that we must analyze and account for the various forms of data utility if we are to discuss its value. As we'll see, the analysis cannot end there.

It may seem that we need consider nothing else to develop a scheme for data valuation. However, the market sets the value of those goods and services. Market price is, of course, controlled by supply and demand. Because data is usually only indirectly connected to sales, we must somehow distribute the value of goods and services across the potentially many assets used to create those goods and services. This is a complex, but solvable problem. In fact, it's just a new take on the old problem of the economic relationship between raw materials, consumables, and capital equipment on the one hand, and finished goods on the other.

The value of one or more data elements as isolated assets isn't the same as data organized as information. By organization, we mean the inter-relationships among data. This is similar to the value of a manufacturing line in producing a particular product vs. the value of a single piece of capital equipment. Relationships, once manifest, add value to the individual components. This value can be recognized in the manufacturing example if we consider replacement cost of the entire line. With data relationships, the problem is easily addressed by considering the relationships as data (as in a relational database). Moreover, we can "chunk" data into business objects (also known as abstract data types or classes), thus considering their value as a unit.

That raw materials, consumables, and capital equipment aren't free has a direct impact on our data valuation scheme. On the supply side, we try not to sell below cost. This sets a minimum value controlled by cost. Thus, we have to take into account the costs associated with raw data acquisition, derivation, presentation, and retention. Among the costs associated with retention are the cost of subsequent recovery, including disaster recovery planning, disaster recovery insurance, and business interruption insurance. The risk of data loss or corruption must be considered in determining retention costs.

Too often, these costs are treated as undifferentiated operating costs to be indiscriminately allocated. This practice prevents IT departments from developing a rational approach to managing the ever-increasing flood of data. By contrast, if costs can be allocated to specific groups of data (or even specific data elements), many IT decisions can be based on how much those costs are increased or decreased. By comparing anticipated costs with potential data value, rational business decisions can be made in light of the net value of those groups of data. For example, no reasonable business manager would insist on maintaining a piece of capital equipment for which the maintenance and operating costs exceed the maximum potential value contribution of that equipment. Should data be managed with less sophistication? Surely not!

The risk of data loss must be considered in determining retention costs.

From the many considerations we've introduced, we can establish a scheme of analysis that will result in a methodology for data valuation. First, identify the various uses of data that must be considered in establishing a utilitarian model of data value. Second, identify the various direct and indirect costs associated with data acquisition, retention, use, and even divestiture to develop a model of data costs. Third, consider the effects of degradation, depreciation, and appreciation. Then combine these components to develop a cost/benefit model of data valuation.

Once we have established this data valuation methodology, we'll look at some practical uses for the cost-benefit model. Hopefully, you'll be able to use data valuation to make important decisions that affect the cost of your enterprise integrity. eAI

David McGoveran is president of Alternative Technologies, Inc. He has more than 20 years' experience with mission-critical applications and has authored numerous technical articles on application integration. e-Mail: mcgoveran@eaijournal.com; Website: www.alternativetech.com.