enterprise integrity

By DAVID McGOVERAN

Valuing Data, Part 5

In previous columns, we've examined data valuation with the assumption that data quality is perfect. We now examine ways in which that quality might be less than perfect and some ways that result in data value degradation and depreciation. By "data value degradation," we mean the financial impact of imperfect data quality. So far, our model has assumed that data quality was as good as it can be. But data can be wrong in many ways, some of which have devastating consequences. In effect, data quality acts to modify potential value. So we model it as a multiplier of value having the range zero to one; this value represents an estimate of probability the data is correct. How you assign these probabilities depends on your particular business, but why is determined by the ways in which data quality is degraded.

One cost of data in the acquisition state is data cleansing. You must verify that your data is correct, whether you purchased it or obtained it through data entry. If data is incorrect, cleansing is required. Those involved in such operations almost always can assign an intuitive data correctness probability to the result.

Although data value may appreciate for many reasons, the good news is that appreciation is easy to take into account.

For data in an inventory state, storage and retrieval operations or media failures may cause degradation. The probability of these is due almost exclusively to information system characteristics (e.g., mean-times-between-failure) and system administrators can estimate it. Operational state data degrades due to re-entry errors and errors in data usage (using the correct data incorrectly).

Once in the forecasting state, data quality has a complex impact. For each use, how critical the data is must be evaluated. However, the same causes of degradation occur as for the operational state. The reasons for degradation are the same as for the inventory state. Finally, data in the divestiture state may degrade during transfer. The transfer mechanism determines the probability that quality will degrade. Of course, we can ignore degradation if divestiture means disposal.

Although partially controllable through proper data maintenance, data value depreciation is an automatic consequence of aging. Data is valuable only insofar as we can understand what it means and how to use it properly. Stored data has to be carefully maintained — with context of relationships to other data elements intact — if its meaning is to be preserved. Database backups and snapshots, for as long as they're recoverable given changes in technology, serve this purpose. Over time, data relationships and even the data type meanings in a business change. Surprisingly, even the most uncorrupted and fully recoverable database backup undergoes depreciation for this reason.

Think of the data as a collection of interconnected data elements, each element of which represents a specific meaning of a data type and its instances, and each connection of which represents a semantic relationship to other data types. The formal name for such a construct is a semantic net. One incorrect meaning of one data type in this semantic net causes all its relationships to be broken. Where those relationships define the meaning of other data types, the meanings of those data types are incomplete and so incorrect. Likewise, if a relationship is incorrect, data types interconnected by that relationship are less well-defined and so partially incorrect. A database backup or snapshot in effect stores such a semantic net, albeit the semantic relationships and definitions are more likely implicit than explicit. Over time, expected meanings and assumed relationships don't keep pace with the business's current meanings and relationships. The costs of recovering intended meanings and relationships from stored data, and then translating into expected meanings and relationships tracks data value depreciation. Its pace accelerates pace over time, resulting in a near exponential decay of data value as the semantic net breaks down. We must keep in mind, however, that depreciation costs undergo catastrophic increases when, for example, backup tapes become unreadable because of changes in the technology or physical decay.

Although data value may appreciate for many reasons, the good news is that appreciation is easy to take into account. An inflation factor can account for general business growth, and appreciation due to new data uses represents merely new instances of utility already described by our model. As you contemplate the impact of degradation, depreciation, and appreciation on your data value, consider that all value derives from integrity. So hold on to your (enterprise) integrity; without it, you give up everything. BIJ

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