Constrain Your Way to Better Data Definitions

Richard Larson, PMP, CBAP
Principal, Watermark Learning, Inc.

Industry Article

Constrain Your Way to Better Data Definitions

By Richard Larson, PMP, CBAP
Principal, Watermark Learning, Inc.

Thanks to Tim Ward, CBAP for suggesting this article. If you have suggestions for tips or would like to write one yourself, please give us your thoughts at www.watermarklearning.com/feedback.

Most IT projects create some sort of new data. You might be working on a small effort to refine a report or add data to an online display. Or, you could be creating a whole new system. In the former case, you would probably add a few columns to an existing table. For a new system you need to create a completely new database. (I always like those kinds of projects because I can really sink my teeth into the data model). Whatever the scope or size of a project, you usually have to deal with the data.

Let's say you are adding flexibility to an order entry system. Most processing in the current system is at the order level, and the company now wants to apply product discounts, shipping location, and shipping costs at the line item level. The project will allow processing complex orders the way customers want them. None of this data is captured or available for reporting currently. Your job is to define the data requirements to make sure the right data is collected and reported.

Data Type. Let's start with discounts as an example, using the term "attribute" to refer to a detailed data requirement. If you defined an attribute for discount as a number, that's a good start. But, is the discount a percentage or a decimal? Could it be either? For each type, how many decimals are required? These are all good questions to determine the "data type" of the attribute. The data type is the highest level of data definition and should be noted for every attribute.

Constraints. Another important part to defining data is to note any constraints that its values can take. Let's say a discount is a percentage or number with 2 decimal places. Here are some things I'd want to know about the data:

- Are there upper limits to a discount? One would assume that 100% is the maximum, but it would be good to find that out. What if a discount is a flat amount? Are there upper limits to the flat amount?
- Can a discount ever be negative? I know, that sounds weird, but maybe there is a good business reason for a discount to be negative. Or, maybe the organization *never* wants a discount to be negative, so the constraint would be it could only be zero or a positive amount up to 100%.

Watermark Learning
ENHANCED PERFORMANCE. ENDURING RESULTS.

The point about constraints is to never assume and never default. If the discount was a default number, it could potentially be negative, or potentially be 900%. Hmmm, is that one reason why bad data gets into a database?

Integrity Rules. The last consideration for defining data is to find out any integrity rules for an attribute. An integrity rule is a constraint that one attribute's value places on another attributes value. Let's say that that a standard discount percentage is set depending on a customer's status, another attribute. If a customer's *status* is "gold," they receive a 5% discount. If their status is "platinum," they get a 10% discount. Regular customers don't automatically receive discounts, and their default discount is zero.

Another example is the *sales tax amount* to apply to a line item. Depending on the *state* attribute, a sales tax percentage is applied. Again, one attribute's value constrains the value of another attribute. A final example is the *discount code*, which is a percentage in most cases, but could be a dollar amount. What determines the percent vs. amount? Why, an integrity rule, which dictates that certain products get discounted by amounts and not percentages.

The three types of data definitions discussed here aren't easily portrayed in an Entity-Relationship Diagram (ERD) or a class diagram. You typically have to document them using text statements. That can be tricky to pass on to developers or database staff for creating databases or code. Data modeling tools like Erwin Data Modeler® can help, but you might have to resort to Microsoft Word® if you don't have a formal tool.

To summarize, the three facets of data definition we are suggesting include:
- **Data type** – the classification of data, including character, date, integer, decimal, percentage, etc.
- **Constraints** – any limits imposed on the values that data can take.
- **Integrity rules** – constraints that one attribute's value places on another attributes value.

Without good data definition, developers are forced to ask questions about the data that should be raised during analysis. Worse yet, developers and DBAs may make assumptions or choose defaults that are incorrect, and the customer is not happy. Without constraining data values, absurd results like a 900% discount can also occur, which lessens the value of the data collected, or worse.

If you do define data using data types, constraints, and integrity rules, though, developers and DBAs can make good use of them. All three types of definitions can be readily incorporated into column definitions in a relational database table. The resulting databases and reporting will be more valuable and customers will be happier with the results. So, get busy and start constraining!

Watermark Learning

ENHANCED PERFORMANCE. ENDURING RESULTS.

About Watermark Learning

Watermark Learning helps improve project success with outstanding project management and business analysis training and mentoring. We foster results through our unique blend of industry best practices, a practical approach, and an engaging delivery. We convey retainable real-world skills, to motivate and enhance staff performance, adding up to enduring results.

With our academic partner, Auburn University, Watermark Learning provides Masters Certificate Programs to help organizations be more productive, and assist individuals in their professional growth. Watermark is a PMI Global Registered Education Provider, and an IIBA Endorsed Education Provider.

Contact us at **800-646-9362** or at **www.watermarklearning.com**.

Watermark Learning
ENHANCED PERFORMANCE. ENDURING RESULTS.