

Design Tip #71 The Naming Game

By Warren Thornthwaite

The issue of field naming rears its ugly head while you're creating the dimensional data model. Naming is complex because different people have different meanings for the same name, like revenue, and different names with the same meaning, like sales. The difficulty comes from human nature: most of us don't want to give up what we know and learn a new way. The unenviable task of determining names typically falls on the data steward. If you are responsible for dealing with this political beast, you will find the following three-step approach helpful. Steps 1 and 2 generally happen before the model is presented to the business users. Step 3 usually happens after business users have seen and understand the model.

Step 1—Preparation

Begin by developing skills at thinking up succinct, descriptive, unique names for data elements. Learn your organization's (and team's) naming conventions. Study the table and column names in the various systems. If you don't have established naming conventions, now's a good time to do so. A common approach is to use a column name standard with three parts:

PrimeWord_ZeroOrMoreQualifiers_ClassWord

The prime word is a categorization word that often corresponds to the entity the column is from, and in some cases, qualifiers may not be necessary. So the field in the Sales Fact table that represents the amount sold might be Sales_Dollar_Amount. You can research different naming conventions on the Internet. Here are some links to get you started:

<https://dwr.ais.columbia.edu/info/Data%20Naming%20Standards.html>

<http://ss64.com/ora/syntax-naming.html>

Step 2—Creating a Solid Starting Point Name Set

During the modeling process, work with the modeling team (including a representative or two from the business) to draft an initial set of names and the rationale. Once the model is near completion, hold a review session with the modeling team to make sure the names makes sense -- this is also good practice for the next step.

In addition to the review session, it helps to have one-on-one meetings with the key stakeholders. This typically includes the core business users and any senior managers whom you have a sense might have an opinion. If their preferred name for any given column is different from your suggested name, try to figure out why. Help them be clear on their definition of the data element by asking them to explain what the term means to them. Look for missing qualifiers and class words to clarify the meaning. For example, a sales analyst would be interested in Sales numbers, but it turns out that this Sales number is really Sales_Commissionable_Amount, which is different from Sales_Gross_Amount and Sales_Net_Amount.

The resulting name set should be used by the data modeling team to update the current version of the data model. Keep track of the alternative names for each field and the reasons people offered for their preferred choices. This will be helpful in explaining the derivations of the final name set.

Step 3—Building Consensus

Once you have a solid, tested name set, and the core users have seen the data model presentation, gather all the stakeholders in a conference room for at least half a day (count on more if you have a

lot of columns or a contentious culture) and work through it. Start from the high level model and progress through all the columns, table by table. Generally, there have been enough iterations of model reviews and naming discussions so that many of the issues have already been resolved and the remaining issues are reasonably well understood.

The goal of this session is to reach consensus on the final name set. Often this means someone has to accept the will of the majority and let go of their favorite name for a given column. It is surprising how emotional this can be. These names represent how we view the business, and people feel pretty strongly about getting them "right." Don't let people get out of the room without reaching agreement if it is at all possible. If you have to reconvene on the same issues, it will take extra time to re-hash the various arguments.

Once you have reached agreement on the final name set, document it carefully and take it back to the data modelers so they can work it into the final data model.