

# BUSINESS VOCABULARIES

*By: Dave McComb*



**semantic arts**

Most companies of any size have created an internal Tower of Babel that frustrates all attempts to integrate their disparate systems.

The Tower of Babel is the difference in language and data names that spring up from the development of dozens or hundreds of separate but highly interdependent systems. A business vocabularies project is one of the most cost-effective means to resolve the problems that have been created by the Tower of Babel. In this white paper we will discuss some of the key considerations that go into watching a business vocabulary project.

## **Packages**

Two of the most prevalent trends in the information system business over the last decade have further exacerbated the problem of the inconsistency of terminology naming, definition, and meaning. These two trends are the implementation of packaged software and the outsourcing of key IS functions.

The reason that these lead to further fractioning of the namespace of the business is that a given business will have a number of terms by which they shape their world. These terms have often evolved over a long period of time and have an internal meaning to the enterprise. However, packaged software vendors have built their packages, their meanings, their names, and their labels from a different environment. In some cases, they were built in different industries and were morphed over into the industry in question.

In many cases, the names are intentionally abstract which allows the package vendor to offer a "flexible" system. The flexibility, of course, comes in the interpretation of what the abstract fields really mean. Outsourcing has a similar net effect; both because many outsourcers use packaged systems and very often process data for many organizations and therefore are less likely to adopt a terminology set of any one organization. If all the companies in a given industry shared the same vocabulary, and if the packaged vendors implemented that vocabulary, then this would not be an issue. However, experience and observation has shown that this is not what has occurred, at least not to the present.

## **Agreement vs. Reconciliation**

The fact that different applications over time have evolved slightly different definitions, different categorizations, and the like, is not up for debate. Anyone who has done any systems integration, data warehouse projects, data cleanup, or data profile projects has seen this firsthand.

The issue is not that there are different versions of entities and attributes in different systems and the issue is not that they have different names or that they are structured and stored differently. This is all true. The issue is that one can either understand and document the differences upfront before another project is initiated, or deal with this in the execution of the project. Generally, what this means is that we end up reconciling the differences between the systems at some point in the project and in many cases with much greater effort than we had planned for.

For instance, in a systems integration project, if we discover late in the project subtle differences between what a customer is in one system versus another system, they will have to be resolved in the integration phase; they just may not have been accounted for in the estimating. And it's the same with the Extraction Transform and Load process for a data warehouse or the development of the interface specification for a Web services project. The intent of a business vocabulary project is to bring these issues into the open where they can be estimated and resolved before subsequent projects begin.

## **Scope**

One of the most interesting concerns for a business vocabulary project is the scope of the agreement. In other words, how many units, how many departments, how many applications, how many other organizations will have to be involved and ultimately come to an agreement about the meaning of the terms for the information that they will exchange.

This is a tricky issue because the smaller the scope, the easier the projects will be to complete. At the same time, the smaller the scope, the less valuable the project will be. For as the scope of the agreement extends, so does the potential value, and the ease of integration extends to a broader and broader network.

Our general counsel is to strike a balance on the scope decision such that it encompasses a broad enough scope to cover the main integration areas in the predictable future and to use some of the techniques that we describe later in this paper to help with the eventual further integration with a wider range of enterprises.

## **Consortia**

One area that may be helpful in getting a kick start to this process would be to review available material from industry consortia. These consortia very often have come together for the express purpose of allowing members of the industry to trade with one another. As such, one of the first things they often have to do is resolve differences in vocabulary.

One caution is that many industry consortia are actually consortia of software companies and their primary motivation is to ensure that the vocabulary, as captured in the schema of their software package, is faithfully reproduced in the consortia. In many cases, this is not particularly helpful for the business vocabulary project, although, in almost every case, they do contain clues to the important terms to be defined and what their definitions might be.

In some cases, consortia have come together led primarily by the business and user community. In these cases the vocabularies are often much more useful. One such example is [Justice XML](#), which is a very user- and business-centric description of the key terms in use in the criminal justice system.

## **Common Usage and Precision**

One of the first goals of the business vocabulary process is to come up with a set of terms that is agreeable and is in common use by the participants. This of course is easier said than done; had they had a common vocabulary, there would not be a need for the project. What you'll find in the process of conducting the project is that there are many terms that are in common agreement. In many cases, they may not be in agreement with common usage outside the organization or industry and that should be noted, but wherever possible we should be using the terms that the people in the company recognize and use on a daily basis. However, even where all the participants agree on the meaning of a particular term, it is necessary to get a very precise and context-independent definition.

There are two reasons for this. One is to ensure that the participants are in fact agreeing on the same thing. The other is because the project will eventually expand its scope beyond the participants at this level. At that point, it will be very necessary to have a context-free, precise definition of the term such that it can be compared with other industries and organizations. Also, in the course of finding and documenting the common usage terms, you will often find many synonyms for these terms. We did a project once where we found over 20 terms that referred to an approximate amount of money (estimated, anticipated, projected, planned, allotted, accrued, allocated, etc.).

Two things need to be done in these cases. First, you need to tease apart the actual definitions of these terms. You will likely find as we did that out of 20-some terms, there are three to four validly different conceptual meanings. For instance, some of these terms referred to the result of a decision that allocated resources, whereas many of the other terms were a projection or a guess for planning purposes. Once you determine the different meanings, it's merely a matter of determining among the audience which synonyms they use for each of the different concepts. At that point you will have the equivalent of a synonym dictionary for each of the key concepts.

The opposite problem occurs very frequently when you determine that one term has two conceptual meanings. For instance, the term "employee" may sometimes refer to employees of the company that owns the system and may in other contexts refer to employees of the company that owns the system it does business with. For instance, in a worker's compensation insurance company there is a major distinction between the employees of the employers, who potentially were hurt on-the-job, and the employees of the insurance company, who are employees in the human resources sense.

## **Archaeology**

One of the first things to do in a project of this type is what we often referred to as archaeology: raking through the ruins, if you will, and finding meta data descriptions, documentation, and any form of information where the business terms in this business have been used. We catalog, organize, and spot check just to get some information and examples through which to have a conversation. These projects cannot be done entirely in systems and through paperwork because, at their heart, the projects are about discovering and delivering agreement on

terms, and the system and the documentation are merely clues in that regard.

## **Anthropology**

The second and most important part of the process we refer to as anthropology, "living with the natives." The process is typically a series of elicitation meetings where we bring the broadest possible cross-section of interested parties into a room at one point in time to discuss a topical area. One of the keys to this is to keep the topic small enough to be discussed by the entire team but large enough to be interesting. It is also key to make sure that we have the broadest number of points of view on any given topic without making the room overpopulated to the point that conversation is nonproductive. We have found that much beyond 12 or 13 people in the room significantly drops the level of interaction, brainstorming, and contribution, as does having fewer than five or six.

## **Top down**

Another trap that we have heard of from other endeavors in this area is that they will get a number of people together and spend days and days arguing over some very detailed distinction. Our experience has been that the arguments are because the broader contexts and agreements have not been reached. Our council is to start with some of the broadest and largest concepts that must be dealt with in the business, and get an agreement at that level before proceeding to the minutia. So, for instance, in the healthcare setting, you may want to start with, "What is a patient?" While that sounds like a ludicrously simple question, you'd be surprised. In the aircraft industry, a discussion on "What is a customer?" uncovered that a customer is a member of its frequent-flier program. What we might think of as a customer (someone who pays to fly for instance) they call a passenger.

## **Time frames**

A key question is, "How long will this take?" Our advice is that these projects will work better, will be more economic, and will result in more buy-in if they are done in a very part-time fashion over a long period of time. Obviously sometimes these projects have mandates, sometimes run a critical path, and sometimes this luxury just cannot be afforded. However, in the cases where it is possible to do it over a longer period of time and more gradually, you'll find much higher quality and quantity of input and more time for participants to agree.

As the time frame becomes more compressed, key business users will delegate initially to subordinate users and eventually to IT or other staff functions, very much to the detriment of both the quality of the output and especially the buy-in on the part of the businesses to the output. A typical topic session lasts about two hours. While it's possible to do more than one in a day, and we have done as many as four in a day, it is both exhausting and very difficult to assimilate what has been learned in any given session.

Additionally, it's possible to schedule many of these topics in a given week but because very often the same key people need to be present, this often creates enough calendar pressure that you will not get their participation. With those facts in mind, it typically takes one to two dozen sessions to get a good working set of vocabulary. It can take more for a very large and complex domain and if there is a broader and more uncoordinated group of users to bring together. How much calendar time this takes is then just a function of scheduling. We have known of some efforts, indeed some of the best efforts, such as Justice XML, which were on the order of two years in the making.

## **Other Benefits**

The main benefit for doing a business vocabulary project is to reduce the risk of other projects such as systems integration, data warehousing, system conversion, Web services implementation, and the like. However, there are many other benefits that accrue from this approach. We have found that this guided exploration uncovers better and more interesting designs for any systems that are being contemplated. The dialogue that goes along with the vocabulary discovery often also highlights what is relatively important, what is meant to be changed, what is meant to be left constant, and other considerations that are key to the development of other systems. We have also found that the process brings a much higher degree of involvement, agreement, and cooperation between the business units and the information technology groups, who in the past were often standoffish partly due to their different vocabularies.

## **Business Vocabulary and Ontology**

The final question that you may be asking yourself, "Is a business vocabulary project the same thing as an ontology project?" Yes, in most regards they are very similar. The disciplines of the ontology community can be very readily and productively brought to bear on a business vocabulary project.

Certainly the rigor of breaking terms into their concepts, defining these concepts much more precisely, and establishing the rule-in and rule-out criteria, as well as concepts of scoping by namespace and logical inference, can all be used to great advantage in a business vocabulary project. However, in the spirit of business vocabulary, we have found that very few business people refer to their vocabulary as an ontology. As such, to the extent that we are trying to use terms that are shared by the participants, we have found the term "vocabulary" to be far more widely shared and at least approximately understood. So, our council is to learn from the discipline of ontology, perhaps even use an ontology editing tool, but use terms of the business community starting with its own business vocabulary.

11 Old Town Square  
Suite 200  
Fort Collins, CO 80524

970-490-2224  
305-425-2224  
info@semanticarts.com

© Semantic Arts, Inc.