Database Design -
Why Doctypes Are A Problem In So Many Databases

I've received quite a few contacts these past two months asking for information on 1) how to write coding instructions (a coding manual) to keep data entry consistent and 2) how to build a database. As a result, I hope you will attend DAPA's Tech Section meeting on Feb 17, 2005, at the BELO for a great session on what needs to go into a coding manual. In the meantime, let me tackle just one aspect of the coding manual - document types.

Database design is one of the most critical elements in successfully tracking your documents. Databases for document organization and retrieval give you the ability to manage huge amounts of paper that seem to surface in even simple litigation cases. It is important to remember that the database is a living document inventory, which will be added to and used many times over during discovery, depositions, in preparation for trial, during trial and after.

In this article, we are going to cover one field - document types or doctypes, as they are affectionately known. The only subjective field in bibliographic coding - the document types (known by the combination of words as doctypes). This is the only subjective part of bibliographic coding, but as such can be misconstrued to include subjects and issues. Since the purpose of the database is to locate documents quickly and efficiently for review, the document type list becomes one of the critical items (fields). Receiving too little attention, it often gets expanded beyond reason to a point where you can't find consistent categories for review.

In a project planning meeting last year, I noted an important point regarding doctypes to the paralegals and attorneys in attendance, all who admitted that it makes sense and which had often been violated in the past. The point has to do with keeping the list short as possible while encompassing the entire collection. The purpose of a good doctype list is to classify the collection by categories for research and review. You want to drive as many similar document types as make sense into a single category.

Many workgroups inadvertently lengthen the doctype list by default, by giving each document in the collection a specific type as they go through them, rather than planning or meeting ahead of time, talking over what make sense for this matter and making a list based on those needs. If more than one person is working on the collection and no instruction manual (coding manual) exists, it leads to inconsistencies. Different persons looking at the same document (via multiple copies) will often use a similar, but different document type to categorize the same document if no list exists. This makes it more difficult to review.

Even when only one person works on the database, you can still find inconsistencies, especially as time elapses in between coding or review sessions. For example, sometimes a letter-formatted document gets called a letter; other times correspondence, and at other times an agreement. An agreement you say, how so? When the agreement was in the form of a letter agreement. One person called the format an agreement, while most others

called it a letter. This can make categorized documents more difficult to find, as one must try to remember all the types that may have been used during the review or coding.

Another good example is the doctype Financial. The title field will allow you to find 10K, Balance Sheet, P&L, and other types of specific financial data without having to create a separate doctype for each one.

On occasion, there will be specific items that need to be part of the doctype list. For example, you have a case where the tax returns are extremely important. Create specific doctypes for your needs, for example: add the doctypes Taxforms and TaxRegs, but leave financial for all other financially related items. This will help to maintain the data consistency that is so necessary for a good, reliable database, while allowing flexibility for the specific nuances of a case.

One firm had a document type list that had grown to 63 types. They were obviously having problems finding things, but not understanding quite why. As we reviewed the list, we determined that some doctypes were found to be ambiguous, others overlapped in meaning, while still others were truly issues and topics.

You must be careful that your doctype list truly represents categories of documents, not issues, topics or subjects. Save issues, topics, and subjects for the issues field, summary field or comments field, which can easily be updated during your or attorney review. Remember that you may have other fields, such as title or OCR that will help you locate more specific documents.

When using a database, it is not necessary to repeat items in other fields; it defeats the purpose of the database. Don't use the database to try your case - use it as a research tool to locate the relevant and important documents for the case.

Open Door Solutions, LLP is a Dallas-based company providing litigation support services and document management solutions to law firms and corporations nationwide. Author Bob Sweat received his education in Business Administration and Economics at the University of Wisconsin and advanced work at Purdue University. He holds a Paralegal Certificate in Civil Litigation with Computer Emphasis from The Center for Legal Technology, Milwaukee, WI, and has years of experience working with local and national vendors on large, complex litigations. Bob is currently a partner at Open Door Solutions.