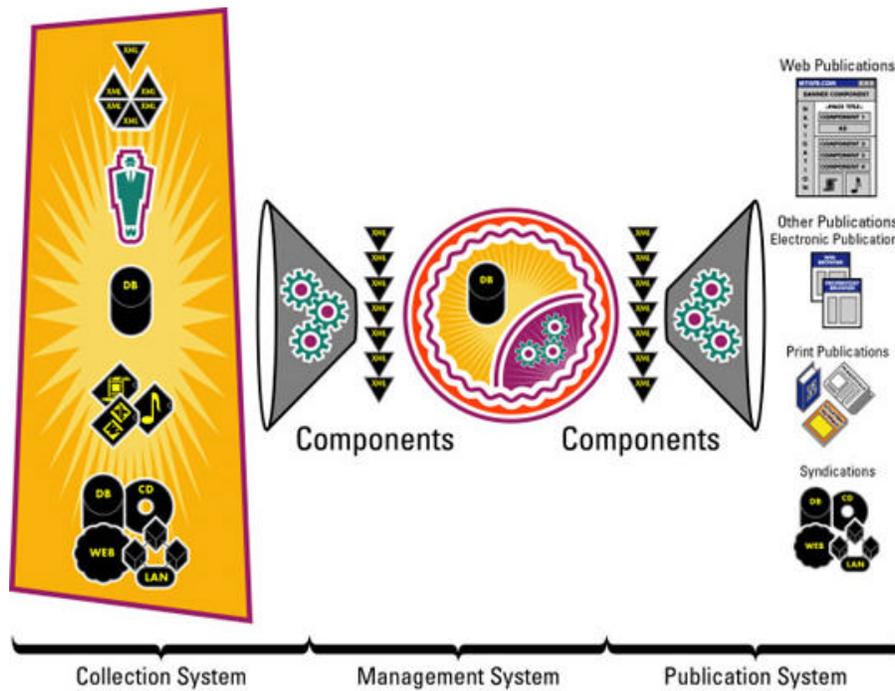


From Data to Wisdom

A CM Domain White Paper

By Bob Boiko



This white paper is produced from the Content Management Domain which features the full text of the book "Content Management Bible," by Bob Boiko. Owners of the book may access the CM Domain at www.metatorial.com.

This paper describes the increasing level of abstraction from the most concrete notion of information (data) to its most ethereal (wisdom).

Table of Contents

Table of Contents _____	2
Data is raw and discrete _____	2
Data is nondiscursive and out of context _____	4
Information is processed and continuous _____	4
Information is discursive and full of context _____	5
Knowledge and wisdom can be information _____	6

The contrast between data and information is all you really need to know to manage content. You take the methodologies of data processing and wrap them around human-created information to create information methodologies. Still, to put content in the context of the wider world of communication and meaning, I'd like to reach beyond the basics, moving from data, the most concrete communication, to wisdom, the most abstract.

Data is raw and discrete

Data is unprocessed and uninterpreted. Data is the starting place. All the interpretation and most of the meaning is yet to come. Think of the data behind a payroll system for a small company: 100 employee names, 1,000 deduction entries, 10,000 time entries and 1,000,000 calculations before 1,000 numbers can go on 100 checks. The data here is a very large number of small, uncooked ingredients. You don't need to know or understand much about any particular datum. Only after they're cooked together into a savory stew (read: processed) do they yield interest. Now, in the payroll system as a whole, you find plenty of processing and interpretation, but it's all in the software and accounting staff, not in the data.

Data comes in *snippets*. Each snippet is complete unto itself and doesn't rely on its neighbors for its meaning. Each datum stands alone as one whole integer, string, bit, date, or what have you. Because they're discrete, you can disassociate each datum from the others, use it interchangeably with any other of the same type, and create any sort of complex processing routine that works on the data.

Table 5-1 describes the more common types of data that you find in Microsoft's version 7.0 of its SQL Server database.

Table 5-1 Common Data Types in the Microsoft SQL Server Database

Data type	Description
Binary	Fixed length binary data (8,000 bytes max)
Bit	Either 0 or 1 (minimum 1 byte)
Char	Fixed length, nonunicode data (8,000 bytes max)

Datetime	Date and time data with an accuracy of 3.33 ms. (8 bytes)
Decimal	Fixed precision data with a range of -10^{38} to $10^{38} - 1$ (5-17 bytes depending on the precision)
Float	Approximate precision data with a range of $-1.79E+308$ to $1.79E+308$ (4-8 bytes)
Int	Integer data with a range of -2^{31} to $2^{31} - 1$ (4 bytes)
Real	Approximate precision data with a range of $-3.4E+38$ to $3.4E+38$ (4 bytes)
Smalldatetime	Date and time data with an accuracy of one minute (4 bytes)
Smallint	Integer data with a range of -2^{15} to $2^{15} - 1$ (2 bytes)
SmallMoney	Currency value with a fixed scale of four (4 bytes)
Text	Variable length, nonunicode data ($2^{31} - 1$ bytes max)
Timestamp	A value unique to the database and updated whenever a column changes (8 bytes)
Varbinary	Variable length binary data (8,000 bytes max)
Varchar	Variable length, nonunicode data (8,000 bytes max)

You don't need to be a programmer to see that the form for each data type is strictly defined - so strict, in fact, that if you try to disobey and call a Char a Bit or a Real an Int, your program can go down in flames. These discrete info snippets that I call data are collected into named buckets called *database fields*, or *columns*. The assumption that programmers must make to ensure that their programs work is that any operation that you can perform on one snippet you can perform on any other in the same bucket - or, in any other bucket of the same type, for that matter.

On the other hand, look at the *Text* data type. It's not nearly as restrictive as the others are. It can contain anything up to about a couple billion text characters. The data type *Binary* is even less restrictive; if you have bits and less than 8,000 bytes, you can put it in that bucket. So just about anything that you can make into bits, including stories, sound, pictures, and motion, you can put in a bucket and treat as data. Well, then, isn't everything that you communicate just data?

If the answer to this question were yes, then you'd have no need for this white paper. The methods that are currently in use would prove adequate to deal with the richness of human communication. Obviously, I believe that the answer is no - everything is *not* just data. I believe that the confusion here is that, although you can represent any communication as data (a series of ones and zeros stored in a database), if that data is information, you can't assume that any operation you perform on one info snippet, you can also perform on any other.

From another standpoint, you can say that, although you can represent all communication as data, it isn't all *just* data. It's often much, much more.

Data is nondiscursive and out of context

By design, data is too raw and fragmented to ever form the basis of a conversation. In fact, the whole point of data is to make information so raw and discrete that no conversation is necessary. What this situation means is that a datum has no life of its own. It's always an object, never a subject. The same qualities that make the datum easy to work with preclude it from ever being interesting itself.

A short list of data types can explain most of what a programmer needs to know to work successfully with data. Contrast this list with what you need to know to work with the following sentence: "Pick me up at 8 and don't be late." You have no little table that you can look in to see what this sentence means. In fact, without a certain set of life experiences (and perhaps a love of the Big Bopper - the younger of you, read on for an explanation), you'd stand little chance of making use of the sentence. The difference is that data is stripped of any connotations or context. It is content reduced to so simple a form that a dumb program can safely and expediently handle it. The sentence, on the other hand, is an unsanitary rat's nest of connotation and context. The way to use a datum is closely prescribed and easily determined, simply by knowing the type of the data. The way to "use" a sentence is nowhere prescribed, and, at best, you can only guess at it.

The long and the short of data is that it's easy to work with, but it carries very little interpretation or meaning. Data reminds me of the story about the guy who lost his wallet on the north side of the street. As I came across him, it was already night, and he was looking carefully for his wallet - but on the south side of the street.

"Why are you looking over here?" I asked. "Didn't you lose your wallet on the other side?"

"I did," he replied. "But the light is better over here."

Information is processed and continuous

Data, at least, has a natural center somewhere in the region where I've defined it. The word "information," on the other hand, has all meanings and no meaning at the same time. Anything, including data, you can rightly call information. Still, I'd like to try to nail this concept to a particular kind of communication. I take information to mean all the common forms of recorded communication that you find around you: writing, recorded sound, images, video, and animations. You can see right away that information is fundamentally more messy than data.

To find a solid center for the concept of information, I use the same distinctions that I laid out for data.

Before you ever see a piece of information, someone's done a lot of work. Someone's formed a mental image of a concept to communicate. The person uses creativity and intellect to craft words, sounds, or images to suit the concept (thus crystallizing the concept). The person then records the information in some presentable format and finally publishes it in some way (even if only to leave it lying by the printer).

Information doesn't naturally come in distinct little buckets, all displaying the same structure, and all behaving the same way. You can still put methodologies, processes, and procedures in place to "handle" information. As you see throughout the rest of this white paper, however, we are all just beginning to learn how. Information tends to be continuous - similar to a conversation - with no standard start, end, or attributes. You disrupt this continuity at your own peril. If you break up information, you always run the risk of changing or losing the original intelligence and creativity that the information is meant to express.

Information is discursive and full of context

A piece of data is the end of a conversation, but a piece of information is the beginning. Douglas Adams understood this concept perfectly in his *Hitchhiker's Guide to the Galaxy*, where the universe's biggest computer finally ended the conversation on the meaning of life, the universe, and everything, with the response "42." The computer did respond to the question, but in typical computer fashion, whether or not the answer was correct, the answer itself was irrelevant to a human.

To possess a piece of data, you simply must remember it. To possess a piece of information, you must interpret it in the light of your current beliefs and knowledge. The information starts a conversation within your head that seeks to surmise what it is, what it means, and what effect it should have on you.

The fact that information is continuous and discursive doesn't imply that it's unstructured. In fact, as the rest of this white paper reveals, information can be quite well organized.

What do you need to know to work with the sentence, "Pick me up at 8 and don't be late?" You certainly need to know what "work with" means. Are you supposed to translate it? Obey it? Put it to music? You may need to know who "me" is and to whom he or she is talking. You certainly need to know any number of other details that you nowhere find in the actual information. In fact, what you need to know to work with this information is nowhere near the actual information. If it exists, which it certainly may not, then it lies embedded deep within your psyche in some dark recess. Miraculously, it's also instantly accessible. Suppose that I stand by the printer and say to each person who arrives, "Pick me up at 8 and don't be late!" Most will probably say "Huh?" or "For what?" or even, "Oh, sorry, I must have forgotten. What did we have planned?" Now, suppose that I say the same thing but with a booming bass voice. The younger ones probably just scrunch their eyes and think, "One of those Bob-isms" (which is one of those things that yours truly is known for among friends). The older ones smile and the bolder of them may actually do the work intended by Jiles Perry Richardson, otherwise known as The Big Bopper, in his song "Chantilly Lace" and respond, "Oooo, Baby, you *know* what I like!"

So, information requires a Web of unstated relationships - a context - to use. Data has these relationships stripped out. Some say that the United States is a low-context culture - that is, in the United States, people use a lot of words and make relatively few assumptions about what the other person knows or understands before talking. Japan, on the other hand, is considered a high-context culture, where most people assume, in no uncertain terms, that you're fully steeped in a deep understanding. The smallest gesture ought to invoke from within you a world of associations and appropriate behaviors. But even in the United States, people spend a lifetime learning to hear what isn't said.

Although data is a lot easier to pin down and discuss, information is the normal mode of life. Data is the newcomer. The quintessential Western idea that you can quantify and categorize meaning

isn't normal even for the most scientifically minded, northern European, white, middle-aged male. You must learn it. Computer applications that work with data are commonplace. But as anyone who's worked much with computers can tell you, you must change your mindset to match the computer's total ignorance of what you intend and the context in which you're performing an action (not to mention its lack of basic politeness). Why? The computer's lack of context is simply leaking through to the user interface. Computers are the ultimate low-context culture. Computers force you to shed your entire context and leave nothing unsaid.

Computers, perhaps, were invented to work with data, but they're now fully pointed at information. Why? Because users have collectively gotten to the point where they can begin to see the possibility of computers moving toward human mindsets rather than people always needing to go to theirs. Users want computers capable of working with information and not just data.

The information technology (IT) departments of the world, whose names by rights should be data technology, are staffed with people who know exceedingly well how to create and run data systems. As the specter of information rises before them, in all of its dirty complexity, human interference, and esthetic considerations, however, they shrink back. They respond by either mistreating information as data (thus denying or ignoring the problem) or by throwing their hands up in disgust and declaring the project impossible. And, from their perspective, it is. These people know how to reduce information to a strictly defined set of small, discrete buckets. To succeed in an information project, however, IT professionals need to change perspective and understand how computers can facilitate the creation and handling of continuous, context-rich intensely human information. Having learned the data craft myself, I'm the last to condemn it. The reductionist methodology created all of science. In particular, it was responsible for the vast power of computers. Only now are people beginning to use that power to grasp and harness information (and not just data) by using computers.

Computers deal with facts very well because facts fit the basic way that computers work - put the world into discrete little buckets and work on these buckets by using long chains of very simple instructions. I suspect that a basic overhaul in this model is in the offing. (People seem to be culturally and technologically ready.) In fact, the very concerns that I'm discussing may drive the overhaul. Until then, you can expect that whatever people make computers to do, they're sure to do it with the same lack of subtlety that they now exhibit. In this white paper, I try to make the best fit between the constraints of information and the capabilities of computers.

In summary, information is fundamentally messier than data. It contains enough human qualities to make it hard to parse, obscure in its interpretation, and complex to handle and use. Successfully working with it requires a different mentality and skill set. But when has that sort of thing ever stopped anyone before?

Knowledge and wisdom can be information

Data are material facts; information is matter-of-fact; knowledge is a matter of dispute; and wisdom is nonmaterial. Although both data and information have a face value, you must synthesize and extract knowledge and wisdom from a wealth of communication and direct experience.

" Knowledge," according to Merriam-Webster OnLine (www.m-w.com) is as follows: "(1) : the fact or condition of knowing something with familiarity gained through experience or association..."

Knowledge, by this definition, is inside a person. It is a mental state, not a communication. So, although today's computers can conceivably store information, they can't store knowledge. But why can't computers store knowledge if books can? Books can't have a mental state anymore than computers can, but aren't books where people currently store the world's knowledge?

Both books and computers store information. If information that the book presents to a person inspires a state of knowledge, then great - the storage and communication was successful. The

communication from the book or computer screen doesn't cause the knowledge to happen; rather, the person in conversation with the communication causes knowledge to happen.

If you don't buy this argument, then I've proved my point. I'm trying my hardest to give you some knowledge that I have (or think I do). Despite my best efforts, however, I can't do it. To my mind, codified knowledge (that is, knowledge that's written or otherwise recorded) is simply information. This concept is a good thing. It means that codified knowledge is manageable by using the same techniques as you use for information. The act of knowledge, on the other hand, is an interpretive act and is, at least for the moment, reserved for humans and maybe a few animals.

Knowledge happens within a *domain*. I know about cars, information, human nature, and God. Wisdom, on the other hand, is knowledge that transcends domains.

Again, according to Merriam-Webster OnLine (www.m-w.com), wisdom is defined as follows:

"a : accumulated philosophic or scientific learning : KNOWLEDGE b : ability to discern inner qualities and relationships : INSIGHT c : good sense : JUDGMENT d : generally accepted belief <challenges what has become accepted wisdom among many historians - Robert Darnton>..."

Wisdom encompasses discerning inner qualities, exceptional insight, and good sense. Wisdom is a highly synthetic process that's beyond simple knowledge. Consider any of the following pieces of folk wisdom:

☞☞ A stitch in time saves nine.

☞☞ Necessity is the mother of invention.

☞☞ Every cloud has a silver lining.

What's interesting about these wise sayings is that, although their literal meanings may be true, their bigger truths transcend the bounds of their subject matter and provide insight into the universe at large.

Wisdom isn't just a statement of fact, as is data. It's not a matter-of-fact statement, as is information. It's not a mental state of understanding, as knowledge is. Rather, it's an ultimately synthetic act where one expresses deep understanding, simply and universally. Consider these other wise statements:

☞☞ The opposite of a correct statement is a false statement. But the opposite of a profound truth may well be another profound truth. (*Niels Bohr*)

☞☞ When we try to pick out anything by itself, we find it hitched to everything else in the universe. (*John Muir*)

☞☞ I do not know whether I was then a man dreaming I was a butterfly, or whether I am now a butterfly dreaming I am a man. (*Chang-tzu*)

Each distills and synthesizes vast realms of knowledge into simple statements. Often using contradiction to capture and coalesce worlds, they pack as much as is humanly possible into a few words.

Like codified knowledge, you can treat codified wisdom as information. After you record it, wisdom functions the same as any other information (even if it's pretty tinny if you compare it to live communication with a wise person).

The works of Lao Tzu, Socrates, or (in my personal opinion) Kurasawa (the Japanese filmmaker) are wise indeed, but you can still collect, store and distribute them as if they were "just" information.