Is the **Document Dead?** Documents, **Data,** and Datamentsby **William J. "Bill" McCalpin, EDPP**

For some years, we in Xplor have struggled with the question, "Is the document dead?" Clearly, the ease with which the Internet enables the facile exchange of information has sent shudders through the document industry. But for Xplor, the word "document" has been the mainstay of the Xplor vision. Since 1992, we have been the association for electronic *document* systems.

Yet it is precisely that focus which is the source of so much *angst* in the current Xplor world: Should Xplor remain focused on an object which may be obsolete in the light-speed universe of the Internet? After all, Keith Davidson, Ph.D., EDPP, Xplor's evangelical leader for many years, has frequently told audiences of the doom of the International Word Processors Institute (IWPI), which went from thousands of members to zero seemingly over-night when technology rendered the job of many human word processors obsolete. This lesson has some long-time Xplor members looking over their shoulders at the Internet tidal wave.

It is the Internet that is causing the current FUD (fear-uncertainty-doubt) in the minds of many regarding the state of the document. With such dynamism in data on the Internet, is the document finally dead?

The Power of Data

The Internet has made the exchange of data extremely easy. Now, because the Internet simulates the networks that companies have used for many years, data-based transactions that were formerly confined within the enterprise can now just as easily be performed between enterprises.

In the past, the only common mode of communicating information between enterprises was through paper-based documents. Although the technology of producing and reading such documents has long been readily available, it was the widespread use of computers that naturally led to the creation of documents whose size made them virtually unusable.

Imagine the telephone bill for a large corporation. Some phone bills are 60,000 pages. The size of such a document can make it cheaper to simply pay the bill rather than study the calls and properly allocate them to the appropriate cost centers.

Clearly, the recipient of a 60,000-page telephone bill would much rather have the data in an electronic, data-friendly format. For example, if the data in the phone bill were presented as a spreadsheet, then the recipient of the bill could easily (as easily as sorting a 60,000-row spreadsheet can be) sort the telephone calls into caller, called number, length of call, area code, cost of call, and so on. This makes the job of properly allocating the calls to the correct cost center much easier, which can potentially reduce the telephone expense to the enterprise.

The Virtue of the Document

By the nature of the persistence of the media on which documents have traditionally been placed—paper, parchment, clay tablets, and so on—documents have had one great benefit: They are very useful for preserving information as records. In fact, the American Heritage Unabridged Dictionary of the English Language

defines a document as "information in writing placed on a medium such as paper, often used as a record."

Thus, any time two people agreed on something and needed to preserve the detail of that agreement, a document was created. Or any time someone discovered some fact that needed to be preserved for posterity, the fact was placed into a document so that people whom the author had never met could share the information. Documents are so critical to civilization that extraordinary measures were taken over the years to preserve them. There is no doubt that the invention of the printing press—the principal source of documents for the last 500 years—almost single-handedly helped enable modern civilization.

The Problem of Dynamic Data

However, there can be problems in a world where all data is dynamic. Remember the 60,000-page telephone bill? This bill contains hundreds of thousands of transactions. If the list of transactions is dynamically generated (by software fetching entries from a database) every time the bill is presented, how sure will we be that the bill detail presented today will be the same detail which will be presented upon a reprint request six months from now? Can anyone be positive that updates, changes, and maintenance to databases, extraction software, presentation software, viewers / browsers, and operating systems won't cause any changes?

In fact, one of the great benefits of a document is to isolate data collected together at a particular point in time, so that no matter what happens to the data generators in the future, the original data is preserved.

Companies that are proud of their dynamic web sites face a similar challenge. If the web pages built for a customer are highly dynamic, then at the consummation of a transaction, how can you be sure of what your customer saw on your web site? If, in some jurisdictions, advertising brochures for insurance policies can be just as binding as the contracts themselves, at what point will companies be hit with legal actions because they didn't know what their dynamic web sites were displaying? Recently, I had the pleasure of watching the web site of an airline's mileage program display the wrong data for my account every day as I checked it, only to see the monthly statement come out correct. Clearly, that airline already understands the difference between dynamic data and the needs of accuracy in statement presentment.

We Need More Than Just Data or a Document

The Internet is leading to the creation of things that even people in our industry are reluctant to call a document. We know that objects on the Internet are electronic in nature, that they may never be on a persistent medium, and—worse for traditionalists—may not use writing at all.

Over the last three years, I have urged Xplorers in sessions on XML to think of documents for both human and non-human readers. That is, XML makes it possible for two software processes to exchange information in lieu of paper documents, and a human being might never see these electronic documents at all. Many veterans in our industry find it impossible to think of documents apart from the human context. As Neil Merchant, EDPP, of American Express says, "The essence of a document is something for human consumption—without regard to the underlying medium." Neil is expressing a sentiment widely felt—that whatever new data objects the Internet is spawning, these objects are not documents if they're not meant for people.

Is it time for a new term that describes the union of the power of data with the security of a document?

Datament: The Union of Data and Document

The American Heritage Dictionary defines a document as "information in writing placed on a medium such as paper, often used as a record." If you use the same dictionary to compare the definitions of data and data bank, you can see that data is—more or less—"information (facts or figures) from which conclusions can be inferred, and which are normally stored in electronic formats for easy retrieval by computer."

To represent the definition of data in a table, we see:

Object	Data
Content	Information
Common Content Format	Data Format
Common Medium	Electronic
Presentation	Dynamic
Presentation Control	Reader
Common Audience	Machine, Human
Internal Integrity	No

To represent the definition of a document in a table, we see:

Object	Data	Document
Content	Information	Information
Common Content Format	Data Format	Writing
Common Medium	Electronic	Persistent (paper, etc.)
Presentation	Dynamic	Static
Presentation Control	Reader	Author
Common Audience	Machine, Human	Human
Internal Integrity	**No**	**Yes**

If we accept the argument that while the ability to handle and exchange dynamic data is a great benefit of the Internet, yet the information-preserving nature of documents is still a requirement in human intercourse, then we have to propose a new information object, the datament, which will combine the dynamism of data with the assurance of documents.

Now that the Internet has enabled the easy exchange of information, users of all kinds are demanding the ability to control the presentation of the information. At the same time, both businesses and regula-

tors have a legitimate vested interest in exploiting the benefits of documents as a vehicle for information exchange, particularly in the realms of financial agreements and knowledge retention and management.

A datament is an organized collection of information in time which has the following characteristics:

Object	Data	Document	Datament (Class 1 - B2C)	Datament (Class 2 - B2B)
Content	Information	Information	Information	Information
Common Content Format	Data Format	Writing	Writing & Structured Data	Structured Data
Common Medium	Electronic	Persistent (paper, etc.)	Electronic	Electronic
Presentation	Dynamic	Static	Static (Prime View) Dynamic (Data View)	Dynamic (Data View) (if any)
Presentation Control	Reader	Author	Author (Prime View), Reader (Data View)	Reader (Data View) (if any)
Common Audience	Machine & Human	Human	Machine & Human	Machine
Internal Integrity	No	Yes	Yes	Yes

Like data, a datament places the power of viewing information with the reader. Like a document, however, there is a default prime view determined by the author, and which can be used to satisfy regulatory needs. A Class 1 datament intended for humans must carry both the presentation information for the prime view and the data formatted for a satisfyingly dynamic data view. Yet a Class 2 datament that is never intended for humans might carry no formatting information.

Datamentsmust also have an internal organization. The integrity of the organization assures browsers and other machine readers that data are intact. Think of the structure of AFP that allows error recovery within a print file or of the byte count in PDF that inhibits casual data editing. Most data formats do not contain such integrity because the nature of the data is transitory, and the usual response to the erroneous presentation of data (say, via transmission error) is simply to fetch the data again.

Datamentsand All Those e-Spaces

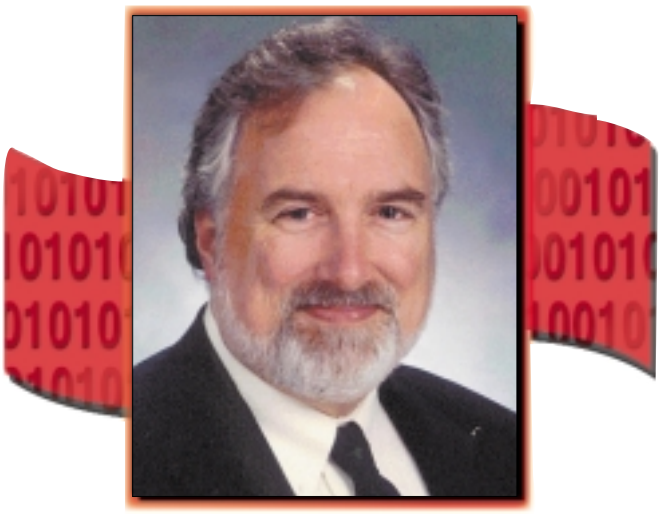
A datament is an object that fuses the power of dynamic data with the assurance of a document. Datamentscan be used within any of the business processes enabled by the Internet: EBPP, electronic statement presentment (ESP), electronic invoice presentment (EIP), and so on.

Datamentscan be expressed in multiple underlying formats as well. Clearly, XML is well-suited for describing datamentsin that the tagged data in the XML datament is easily parsable by dynamic data presentation engines, yet at the same time, the XML datament can carry an eXtensible Style Language (XSL) representation for human readers. The Document Type Definition (DTD) or XML schema provides structural integrity for the tagged format. And the XSL itself can also be used to present alternative views of the data so the author can still enable some dynamic presentation even if the reader's engine is only a browser rather than a full-featured data presentation engine. However, the standard for an XML datament is not just any XML object, but one that satisfies the requirements listed.

PDF may also be suited for datamentsIts internal integrity and presentation features are well known, but if the ability of PDF to have embedded Javascript means that readers may also have data presented dynamically, then a datament standard for PDF may be possible as well.

Killen and Associates says that by 2005 as many bills will be delivered by e-mail (push) as by web delivery (pull). If this is so, what will be the format of an effective vehicle for the electronic communication of billing information and transactions that is regulatory-friendly and data-friendly, internally consistent, and persistent?

The datament appears to be the proper format. It fuses together the best of new technology with the necessary business processes and procedures that make e-commerce and information exchange not only technically possible but eminently practical. Datament: Data that can be treated like a document. ■



William J. "Bill" McCalpin, EDPP, is one of the founders of MHE, a consulting firm for the worlds of electronic printing and imaging. Formerly with Xenos Group, Bill is on the Xplor Board of Directors and is also the editor-in-chief of Xploration Journal. Reach him at mccalpin@mhe-consulting.com.