# How Quark XML Author Enhances
## Microsoft Word

## Abstract

This white paper describes Microsoft® Word's built-in XML support and shows how Quark® XML Author enhances Word to make it an easy-to-use editor for XML documents.

## Overview

Microsoft introduced XML capabilities in Office 2003 and enhanced those capabilities in Office 2007, which has helped enormously to introduce people to the value of XML. Microsoft has focused its use of XML on enabling third-party software developers to integrate data into documents, integrate documents into downstream processes, and deliver other productivity improvements for individual users.

But Microsoft has deliberately chosen not to support XML editing in Word out of the box. For those customers who want to use Word to author XML documents, Microsoft guides their customers to partners' products such as Quark XML Author.

While developers could take advantage of the built-in XML capabilities of Word to deliver XML editing functionality for simple cases, the requirements of a dynamic publishing application require the full-fledged XML document editing functionality that only Quark XML Author for Microsoft Word provides. Quark XML Author builds on Word's XML capabilities to allow authors to create and edit XML documents while preserving the simple, intuitive user experience of Microsoft Word.

# Microsoft's Support for XML

This section provides details about Microsoft's XML-related functionality, to help you both understand its value for many types of applications and appreciate its limitations as an XML editor. The XML support in Microsoft Word falls into two areas: file format and data integration. This section describes each of those areas in detail.

## File Format

Let's say that your organization has generated thousands of Word documents that contain your company logo and that now, because your logo has changed, you need to replace the logo in all of those documents. (This is a scenario we borrowed from a Microsoft article at http://msdn2.microsoft.com/en-us/library/aa338205.aspx, which goes into Microsoft's file format in more detail than we do here.)

One way to replace all of the logos is for someone to open each document, find each instance of the logo, replace it with the new logo, and save it. That might be okay for a few documents, but for hundreds or thousand of files, it would be a labor-intensive and time-consuming task that is also vulnerable to error.

An alternative would be to write software that automates the entire process of opening each document, finding and replacing the old logo, and saving the updated document. But until the release of Office 2007, Microsoft Word files were in the ".doc" format, a proprietary binary file format. To automate the logo replacement, you would first have to learn the file format (by itself, a daunting task) and then write and test software to scan the files, find the target data, and replace it.

To simplify this scenario, as well as to enable many other new features (described in the article at the above link), Microsoft decided to open up the .doc format in Office 2007 so that others could more easily access it. And while Microsoft could have stopped with just publishing the specifications for the existing .doc format, they went further and changed the default format to an XML version of .doc that Microsoft calls "OpenXML," which uses the .docx extension.

The primary advantage of basing this new file format on XML is that abundant software is readily available to process XML. This means that developing software to process OpenXML is easier and less expensive than writing software to process .doc files.

(Microsoft first introduced an XML file format in Office 2003 that they called "Word ML." In Office 2007, Microsoft made several improvements to the format, published documentation about it, and made it the default file format for Office. Although Word 2007 can open .doc files, it uses the OpenXML format by default.)
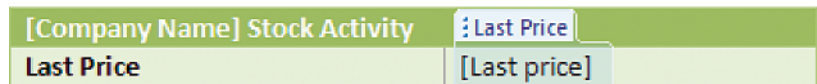
Quark XML Author builds on Word's XML capabilities to allow authors to create and edit XML documents while preserving the simple, intuitive user experience of Microsoft Word.

# Data Integration

Many Word documents contain data that comes from other business systems. For example, Word documents could be catalogs that contain part numbers, descriptions, and prices from a database, or they could be research reports that contain stock prices and other financial data for publicly traded companies. In most cases, getting that information into Word documents involves someone typing the data into the Word document.

But what do you do when the data changes? Until Microsoft released Office 2007, someone would have had to perform a procedure similar to the earlier logo example: open the old documents, find the obsolete data and update it with the latest data, and then save the documents. How many weekly or monthly reports are built using exactly that approach? And how much time and effort does that waste?

To eliminate that error-prone and labor-intensive work, Microsoft introduced XML-based data integration capabilities in Office 2003. With this feature, developers can easily set up fields within a document that connect to databases and other data feeds so that updates happen automatically or with the push of a single button. These days, XML-based data feeds are available from all major databases and enterprise systems.

To implement this capability, Microsoft introduced "content controls" in Office 2007 that a programmer uses to embed XML data in a Word document. From the Microsoft article at http://msdn2.microsoft.com/en-us/library/bb266218.aspx, the following screen shot shows the use of a content control to embed stock quotes in a document:

| [Company Name] Stock Activity | Last Price |
|---|---|
| **Last Price** | [Last price] |

A programmer can set up the content control so that when the user opens the document, Word automatically retrieves the the latest stock price and updates this table. In addition, Office 2007 allows the programmer to lock content controls so that users cannot alter their function or remove them. Locking down parts of the document enables downstream processes to rely on the documents to contain expected information every time, which simplifies the process of writing and testing those processes.

# How Quark XML Author Enhances XML Support in Microsoft

Although Word provides some excellent capabilities that make use of XML, Microsoft does not claim that Word is an "XML editor." In fact, Microsoft tells customers who want an XML editor to look at partners' products like Quark XML Author.

Before we examine how Quark XML Author enhances Word, you should first understand the key difference between an XML editor and Word: an XML editor lets you use your own tags throughout the entire document. For dynamic publishing applications — where you want to publish the same information with different types of formatting for different types of documents and media — this is a critical requirement.

In contrast, Microsoft chose the tags for its Word 2007 XML file format to control the appearance of the content in a Word file. This makes perfect sense for Word, but it is exactly the opposite of the separation of content from its formatting that dynamic publishing requires. In dynamic publishing, we want the tag to indicate the meaning of the content (for example, title, part number, or list) and not its appearance; a separate automated publishing process determines formatting based on meaning, which allows text with the same meaning to be formatted differently for different purposes.

To support the data integration capabilities of Word, the content controls in Office 2007 allow you to assign your own XML tags to specific parts of the document. You may be tempted to conclude that if you were to use content controls for every part of your document, then you could build your own XML editor. And you would be correct.

However, to replace all of the normal text editing functions with content controls, you would have to decide how much effort to invest to preserve the normal behavior of Word. If you are willing to greatly alter Word's normal behavior by creating what amounts to a very smart and very functional form, then development might take just a few weeks. But if you want to minimize disruption to users by preserving Word's normal behavior to the greatest possible degree, development would likely take years.

That's where Quark XML Author comes in: it creates a true XML editing environment within Word. In order to achieve this, developers had to inspect every function in Word, whether accessible through the keyboard, menus, toolbars or Windows operating system, and analyze its:

- Potential to corrupt the document

- User experience for normal Word operation

- Appropriate implementation for XML authoring
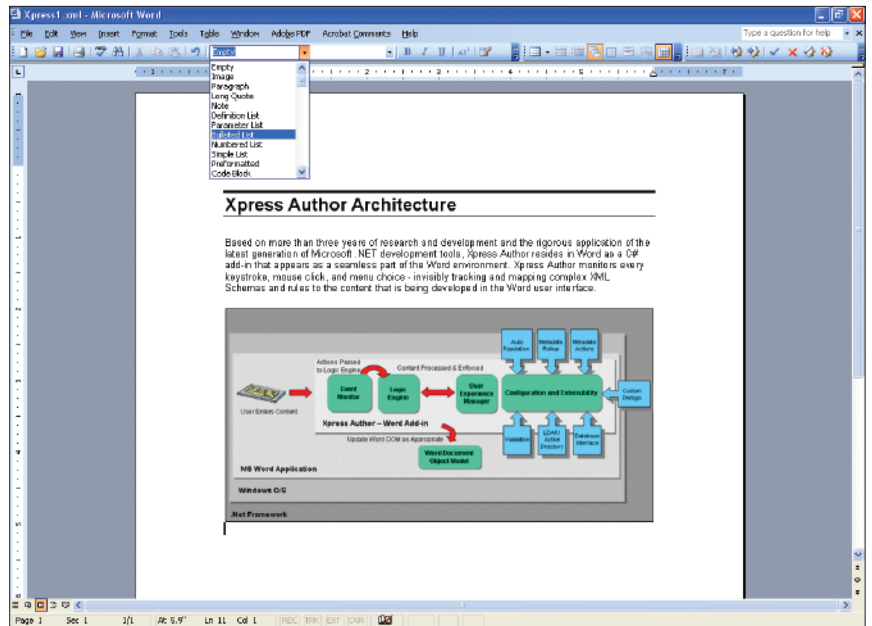
*Quark XML Author for Word works in the background and minimally modifies the normal operation of Word to let users create XML documents. In this illustration, Quark XML Author uses the Word Style dropdown list to show all of the valid types of content that the author can insert at the current cursor location.*

One example of the functions that Quark XML Author modifies is Word Styles. Quark XML Author sets up a relationship between Styles and XML elements, which it calls Smart Styles because the only Styles that appear in the drop-down are those that are valid for insertion at the current cursor location. Quark XML Author intercepts hundreds of functions such as the Styles function and must deal with thousands of permutations. To build your own XML editor within Word, you would have to duplicate all of that functionality, an extremely expensive proposition.

Quark XML Author aims to preserve the normal experience of using Word wherever possible. It uses dialog boxes and other types of controls only if they improve the user experience over the native Word interface.

To enforce compliance with XML, Quark XML Author takes the following measures:

1. Removes all irrelevant functions from the menus and toolbars

2. Intercepts invalid functions, such as an attempt to drag and drop an image into a part of the document where it is not allowed, and displays a user-friendly error message instead

## Conclusion

> The combination of Microsoft Word and Quark XML Author delivers the benefits of XML with the comfortable, convenient user experience of Word.

While recent releases of Microsoft Word provide significant improvements in its use of XML, it does not offer XML editing capabilities out of the box. But because Word has a rich development environment, it is possible — through enormous development effort — to create a full-featured, easy-to-use XML editor on top of Word. Quark XML Author for Microsoft Word delivers this experience thanks to eight years and more than 100,000 hours of development work. The combination of Microsoft Word and Quark XML Author delivers the benefits of XML with the comfortable, convenient user experience of Word.

## About Quark

### Revolutionizing Publishing. **Again™.**

Two decades ago, Quark drove the first revolution in publishing with QuarkXPress, desktop publishing software that rapidly became the industry standard. Today, not only does QuarkXPress continue to innovate in the desktop publishing market — Quark is revolutionizing publishing again. With Quark Dynamic Publishing Solution, we're helping customers meet changing requirements and develop new revenue streams by extending the benefits of advanced technologies across the publishing process. Our dynamic publishing solution is setting a new standard in automated multi-channel publishing by combining the power of flexible layout and design with automated workflows and easy XML authoring for personalized communications across print, the Web, and electronic media.

| For more information about Quark XML Author, visit www.quark.com. | North America | Europe | Asia | Japan |
|---|---|---|---|---|
| | Quark Inc. | Quark Media House Sàrl | Quark India Pvt Ltd. | Quark Japan K.K. |
| | 1800 Grant Street | Rue Du Monruz 2 | A 45 Industrial Area | West 248 bldg. |
| | Denver, CO 80203 | 2000 Neuchâtel | Phase VIII B Mohali | 2-4-8 Ebisu-nishi |
| | solutions@quark.com | Switzerland | 160059 | Shibuya-ku, Tokyo |
| | +1.800.676.4575 | solutionseu@quark.com | Punjab, India | 150-0021 Japan |
| | +1.303.894.8888 | 00800 1787 8275 | solutionsasia@quark.com | solutionsjapan@quark.com |
| | | +41 (32) 720 1414 | +91.172.2299.131 | +81.3.3476.9440 |