Using open document data to build innovative new solutions
White paper
October 2006

Opportunities for innovation with OpenDocument Format XML.

Contents

Introduction

On May 1, 2006, by a unanimous vote of participating ISO members, the OASIS OpenDocument Format for Office Applications V1.0 specification became ISO/IEC 26300. As a newly minted ISO standard, OpenDocument Format (ODF) is now viewed by senior public policymakers and information technology officers in governments around the world as a legitimate international standard. Public policies articulating required support for ODF will be increasingly published, creating a new wave of emerging marketplace demand for solutions that integrate support for the format. Spurred by this increased demand, new and innovative solutions will be developed that take advantage of ODF XML.

For now, the ODF XML technology standard for document data is most closely associated with today's office application suites, e.g., Sun Microsystems StarOffice and its open source sibling, OpenOffice. The IBM productivity editors, which are a standard component of IBM Lotus® Notes® software, also are now supporting ODF XML as a default read/write file format. But, more significantly, ODF as an XML technology opens the gates for new programmatic access to the document data held within. In a closed, proprietary document file format, document data is out of bounds—meaning that the value of document data has been unrealized due to the closed nature of today's dominant office application suite. But all of this is about to change.

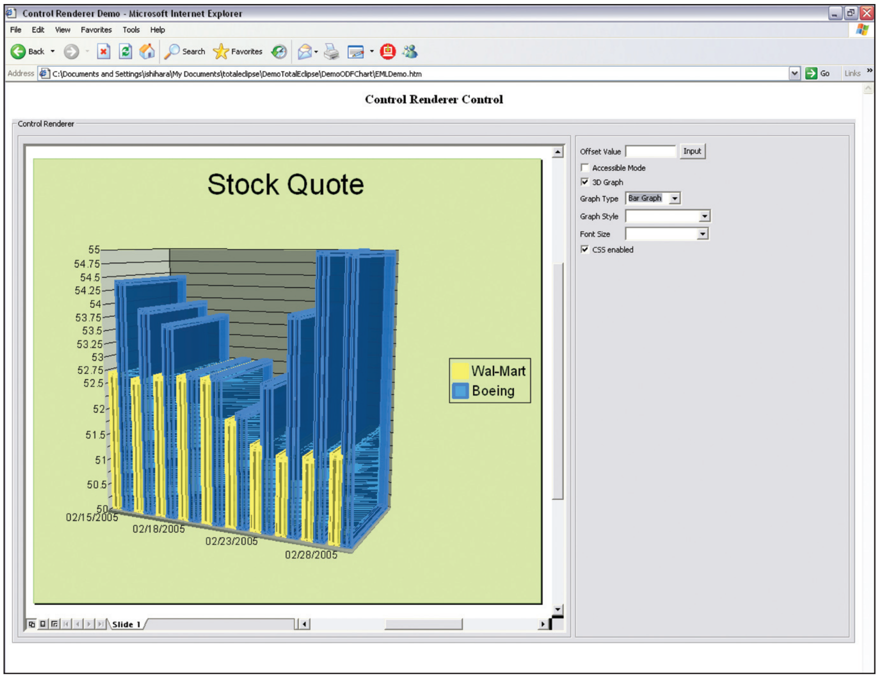
IBM sees an opportunity to leverage ODF XML data in new and innovative ways. This short paper will discuss and demonstrate how ODF technology can benefit new types of solutions that redefine how documents, and the data contained within them, may now begin to participate in the context of business processes, workflows and transactions. Many of these new solutions may be deployed directly on the Internet via industry standard, open source Web browsers, or they may be deployed within any ODF-supporting application—such as OpenOffice, KOffice or the IBM productivity editor software. In each instance, end users will be able to collaborate and interact with document data in ways never before possible.

Integration through programmability

Today, programming with document data is too complex and platform dependent. Microsoft® Office software requires developers to use Microsoft Visual Basic for Applications, a Microsoft Windows® software-only and Microsoft Office software-only programming language and interface. Alternatively, developers working with StarOffice or OpenOffice must rely on the Universal Network Objects (UNO) application programming interface. None of these technologies interoperates well with independently developed third-party technologies.

The Document Object Model (DOM) used by all modern Web browser applications is a powerful way to functionally (not just visually) integrate various kinds of documents. It is also widely used in the context of Web applications on the server side in languages such as Java™. Thus, it is one of the few interfaces known and understood by browser script-based programmers as well as by traditional programmers who use procedural languages such as Java.

IBM is developing a new, simplified DOM-based programming model for ODF. It leverages the ODF XML format but, more important, it uses a DOM as the document's run-time model. This means that it's now possible to dynamically control an ODF document using a variety of scripting and other languages. It is also possible to programmatically integrate run-time behavior of an ODF document with other DOM-based open-standard documents such as XForms and Scalable Vector Graphics (SVG). And all browser-based technologies such as Cascading Style Sheets (CSS) can be leveraged for personalization or accessibility. Moreover, with a truly open format that has open access to document elements at all levels, accessibility itself becomes open and programmable—no longer constrained by a static realization of predetermined policies. This will truly enable ODF documents to participate in and contribute to a wide ecosystem of documents and deliver an enriched user experience obtained via easy, open composition of standard elements.



Offset Value □ Input
□ Accessible Mode
☑ 3D Graph
Graph Type: Bar Graph
Graph Style:
Font Size:
☑ CSS enabled

Figure 1. ODF run-time integration in the browser

Figure 1 shows a screen shot of one instance of such a run-time composition in a browser. The user has navigated to a Web page with some ODF content and an HTML form. Note the HTML form and the ODF chart showing stock quotes are visually integrated in the same way as HTML gets extended by other markups such as XForms and SVG. Further, the user can enter different symbols in the form and the attached JavaScript handler. This in turn makes an Asynchronous JavaScripting and XML (AJAX) call to the server side to fetch the desired stock quote data and insert it into the ODF run-time data model via DOM-based access. The ODF chart service then dynamically updates the presentation according to the changes to the data model—without a page refresh.

As another example, one can think of a scenario where companies with domain expertise in accessibility can build both accessibility compliance tools as well as run times that check document compliance. In addition, these tools may be used to correct the documents in place within the browser. Imagine the screen shot in Figure 1 with an accessibility tool user interface to the right of the chart, showing where the current composition is noncompliant.

IBM views the composability of components to be an orthogonal feature of the end-user platform, and so one should be able to experience these composite applications from platforms including standard Web browsers (e.g., Mozilla Firefox, Microsoft Internet Explorer), an Eclipse run time, an OpenOffice application or IBM client software supporting ODF.

By providing this visual and functional integration, the new composition result is much greater than the sum of the individual parts. Further, the power of such composability (programmability and integration) will spark new innovation and new marketplaces where companies can independently build and offer widgets that rely on open-standards-based access to ODF-supporting run times.

Document-centric collaboration

There is an emerging trend toward online rich document-based collaboration. Google Docs and Spreadsheets, Zoho Writer, ajaxWrite and social networking startups Zimbra, Socialtext, Jotspot and Alfresco are moving in this direction. In the past, commercial document processing systems (e.g., Microsoft Office, IBM Lotus SmartSuite® software) supported some forms of collaboration. Today, wikis and blogs are beginning to represent new approaches to collaboration on the so-called Web 2.0 platform. However, wikis and blogs do not have a structured information model below them. Without this foundation, it is difficult to support content-based access control, history, versions, views and live collaboration.

Coupled with the integration technology above, ODF's XML-based document model can unlock new paradigms in document-based collaboration on the Web. This easily allows multiple authors to interact in real time with the document and its information, allowing role-based access control, views, versions and history. Combine this concept with specific business templates for documents, spreadsheets and presentations, and the document lifecycle model evolves to one where interaction and collaboration over content or information (data) in the context of a business document is radically different.

For example, authoring teams can easily come together and edit their documents in real time over the network using their preferred ODF editor(s) in any combination. Or teams can simply edit within the Web browser. To facilitate this editing, the ODF document is treated as a shared data model and is "rendered" into different forms: one used by the native ODF editor and the other into HTML for the rich text editor.

Modifications to the content flow both ways and users can productively collaborate on the content freed from the document format. This is possible because open standards are developed and specified with the help and contributions of multiple stakeholders in an open community. The open-standards process plays an important role. As standards are defined and evolve, developers increasingly recognize the opportunity for new marketplaces for these tools and run times. With this revolution in the open-document standards domain, industry leaders will pave the way for content-based collaboration across different types of users, editors and devices. This is the same phenomenon that accelerated the development of the Internet and its subsequent adoption in commerceand daily life.

Implications for enterprise document and content management systems

Today's enterprise document and content management solutions manage large repositories of all types of information documents, images and multimedia. Banking and insurance companies depend on these systems for mission-critical business processes such as claims processing and credit approval. Business information does not always appear in machine-readable structured forms; often, it exists as semistructured templates, such as claim documents and loan documents. Document information must be indexed to be efficiently searched. Indexing technology typically depends on the degree of metadata associated with the document, since search engines are challenged to crawl and retrieve meaningful information when the internals of a document or image are stored in opaque, binary data form.

With the advent and anticipated wide-scale adoption of ODF, and the view of the future where document data will be stored in XML format, these systems will be much more effective in terms of their ability to programmatically index, query, search, retrieve and assemble through transformation operations into new compound documents. These new techniques and methods open up new horizons for developing business solutions distinctly set apart from yesterday's office application suite model. More significantly, they create the opportunity for the development of new software that will programmatically converge many different data sources into a new document, further automating business processes and creating new efficiencies.

An open-standards format is critical because it enables the creation of relational or XQuery-type operators on a document; it also guarantees the document's semantics. For example, in an insurance scenario you could select all claim documents where the claim was about US$20,000—or, join a set of auto claims and home claims documents to create a document with claim amount, claim type and customer name; then assemble the new compound document on the fly.

In fact, document management systems can provide mini-search/relationship mining engines and can suggest new links between projects or assets within the organization, and contribute to the overall efficiency of the enterprise.

It's important to note that ODF is currently in its first-version form. As a de jure open standard, ongoing stewardship and development of the ODF specification continues at OASIS with many vendors and individuals from diverse organizations participating and providing leadership. Significant new work in three subcommittees will conclude before the end of 2006. Accessibility, formula and metadata extensions will update the ODF specification and continue to support ongoing creative innovation. Therefore, we are just seeing the first entry of the specification with much more to look forward to in the near future. It will only improve.

In the meantime, the industry is rapidly moving forward to enhance developer productivity using ODF XML. The projects and opportunities for innovation described above are indicative of this progress, and are expected to continue to accelerate, possibly reaching marketplace entry status in the not too distant future.

For more information

To learn more about ODF; familiarize yourself with ODF resources; or become part of the ODF community, visit:

opendocument.xml.org

Lotus Software
IBM Software Group
One Rogers Street
Cambridge, MA 02142
U.S.A.

Produced in the United States of America
10-06
All Rights Reserved

The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this documentation, it is provided AS IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.