

THE GILBANE REPORT

Content Technology Works!

Vol. 11, No. 8
October 2003

www.gilbane.com

Published by:
Bluebill Advisors, Inc.
763 Massachusetts Ave.
Cambridge, MA 02139
USA
(617) 497.9443
Fax (617) 497.5256
www.bluebilladvisors.com

Editor:
Frank Gilbane
frank@gilbane.com
(617) 497.9443

Editors Emeriti:
Tim Bray
tbray@textuality.com
(604) 708.9592
David Weinberger
self@evident.com
(617) 738.8323

Senior Editors:
Sebastian Holst
sebastian@gilbane.com
Bill Trippe
bill@gilbane.com
(617) 497.9443

Recent Contributors:
Kathleen Reidy
kathleenoreidy@yahoo.com
Bob Doyle
bobdoyle@skybuilders.com

Production Assistant:
Sarah G. Dionne
sarah@gilbane.com
(617) 497.9443

Subscriptions:
sales@gilbane.com
(617) 497.9443

Customer Service:
customersevice@gilbane.com

Consulting Inquiries:
consulting@gilbane.com
(617) 497.9443

Gilbane Report White Papers:
Sebastian Holst
sebastian@gilbane.com
(973) 543.8267

Speaking Engagements:
speakersbureau@gilbane.com
(617) 497.9443

MICROSOFT, ADOBE & XFORMS TO SHAKE UP ELECTRONIC FORMS MARKET

Our title this month reads like a news headline on purpose. There are a number of new, and upcoming, developments in electronic forms (eForms) technology that should be grabbing your attention. Some of these are of major importance on their own, but taken together, they signal the start of a major improvement in businesses' ability to easily collect, integrate, and process information.

"Electronic forms" have been around for years, but the term refers to a wide variety of technologies – from scanned image applications to HTML forms – that are not at all similar and far from equal in their ability to accelerate and smooth business processes. What eForm technology has shared is: a level of difficulty that kept it out of the reach of office professionals who were comfortable enough with documents and spreadsheets, but scared-off by forms, and proprietary data formats that made information integration costly and complex. This month Bill explains why all this is changing. The effects of new eForm technology will be far-reaching, and we will be looking closely at eForm developments in these pages as well as in our conferences. This is a critical technology for improving content management and information integration capabilities and ROI. Stay tuned.

CONTENTS

Microsoft, Adobe & Xforms to Shake Up Electronic Forms Market	2
Industry News	11
Back Issues.....	27
Subscription Form & Calendar	28

MICROSOFT, ADOBE & W3C TO SHAKE UP ELECTRONIC FORMS MARKET

Electronic forms (eForms) have always represented a significant piece of the Enterprise Content Management puzzle. On one end of the marketplace, eForms have been implemented to replace traditional paper processes, such as in government and paper-intensive industries such as financial services. In a number of other applications, such as Web Content Management, eForms are the *de facto* user interface for such tasks as content entry, editing and system administration.

As eForms have proliferated in both of these types of applications and others, the functional and architectural requirements for eForms have grown. Where early eForms were successful merely for capturing and perhaps storing data, it didn't take long for developers to want to manipulate and work with the captured data. On the end of the market where dedicated eForms tools were being used to automate paper processes, such development typically involved working with the proprietary data structures and programming interfaces of the eForms vendor. In applications such as Web content management, the functionality and architecture of eForms were bounded mainly by HTML and related technologies such as JavaScript. As organizations have moved toward application server-centric architectures such as J2EE and .NET, both the proprietary approaches and HTML-based forms have failed to keep up.

From the perspective of commercial software, eForms are a substantial business, but one that historically has been seen from one end of the content management spectrum—that being the scan and capture world. That is, many e-forms products and efforts have been concerned with taking traditional paper processes and beginning to make them more electronic. More recently, some e-forms technologies have begun focusing on taking these now-electronic processes and making them more robust, especially on the front end. By robust, we mean that the front ends have more logic, more customization and adaptation to the user experience, and more fidelity with the intended look and feel of the forms and documents being processed.

This evolution of the traditional eForms business is now intersecting with an evolution in the way forms are being used in Internet applications. Since the beginning of the Web, HTML forms have been the *de facto* and dominant mechanism for user interaction with Web-enabled content sources and applications. The good news about HTML forms is that they are easy to assemble, ubiquitous, and well understood by Web developers. The bad news is that HTML forms are as flawed as HTML itself. Because they are loosely structured and lack a standard programming interface, HTML forms remain an incomplete technology.

Enter Xforms, a W3C technology to bring XML, well developed programming interfaces such as the Document Object Model, and a standard data model to Web forms. In response, the traditional forms vendors have, to varying de-

grees, made their technology Xforms compatible – some by providing mapping of their proprietary data structures to Xforms, others by making their data structures and tools less proprietary. The result should be a new generation of forms technology that will work much better with the growing technical infrastructure of J2EE and .NET.

Significantly, though, both Adobe and Microsoft have stepped into the mix. Adobe has been targeting the forms market since its Accelio (Jetform) acquisition. Adobe's efforts are both an attempt to broaden the Acrobat/PDF franchise and to expand into new markets. Microsoft's introduction of InfoPath can be viewed in a similar light – an attempt to broaden the franchises of both Microsoft Office and Microsoft's developer tools, as well as an effort to expand into new markets (and notably some of Adobe's). How successful will Adobe and Microsoft be, and how successful will some of the established e-forms companies be in retaining their market share in the face of these new and larger competitors?

The fundamental change is that the new generation of eForms technologies will combine interactive forms based on structured markup with tools accessible to normal office workers. eForm applications and XML content could proliferate both throughout larger enterprises as well as smaller businesses. More tools—and more affordable tools—are emerging because the technology is maturing and because of the number of potential applications is enormous. eForms should emerge as a significant technology in the ECM mix.

FORMS & EFORMS IN THE ENTERPRISE

Forms are ubiquitous in business, government, and education. One of the first things that even a small organization does is develop forms for standard business processes—invoices, time cards, employment applications, and so on. For customers and citizens, forms are often the first (and hopefully not the last) interface to an organization. For large organizations, it is not unusual to have hundred, even thousands, of unique forms. Often, these forms have evolved and been refined over a long period of time with great care being paid to both the information being collected and the design of the form. A well-designed form is a successful blend of information design, graphic design, and attention to the needs of the business.

But the traditional paper form left this information at arms-length—or more—from the business systems that needed to process this information. As businesses automated, paper forms often were lashed to electronic processes, making for complex and inefficient workflows. Thus was the scan and capture business born, and many organizations have lived for years with various kinds of scanning, data entry, and other kinds of semi-automatic processes.

Form technology has have given organizations the ability to tie forms—and the level of information capture that comes with forms—more closely to business processes. eForms can reduce inefficiencies, and increase the accuracy of the information that is captured. Such improvements come from simply allowing the information to be captured once, instead of relying on error-

prone techniques such as scanning and re-entering. By capturing the information earlier—and with greater accuracy—eForms can help organizations avoid data quality problems downstream and provide more automation to common business processes.

eForms technology has grown by providing better functionality in at least three areas:

1. *Improvements in the rendering of the forms.* The precise rendering of some forms is very important for certain applications. Many eForms applications have been driven by the requirement to make the electronic version of the forms as much like the paper as possible.
2. *Improvements in the validation and user interface.* Early eForms provided rudimentary data capture, and little validation of the captured information. These technologies have been improved by providing more validation, and also by improving the user interface to allow for features such as having the form customize based on certain user input.
3. *Improvements in the interoperability of the forms technology with other software.* Originally, eForms solutions were highly proprietary, with proprietary repositories and programming interfaces. As organizations have emphasized open standards and greater interoperability, the eForms technologies have been enhanced to improve more standard data structures and programming interfaces.

HTML forms have followed a similar evolution. Early HTML forms provided rudimentary data capture, little or no validation, and were loosely coupled with databases and other applications. As more applications were tied to browser interfaces, developers devised a number of techniques to improve the user experience and improve the data capture and validation. Ultimately, though, the result is a hodgepodge of scripting, style sheets, HTML, and a variety of supporting technologies. The fact that HTML—with its inability to describe and enforce a standard data model—is at the core of the approach is a fatal problem in and of itself.

The shortcomings of HTML forms are widely known and understood. Indeed, the standardization effort that became Xforms grew out of the frustrations developers had with HTML forms—and the realization that a better forms technology was needed to help make the Web the ubiquitous interface for all applications and content. Similarly, the commercial eForms products needed to conform to the current architectural requirements of enterprise content and data management systems.

This brings us to the current state of eForms, and the near-simultaneous emergence of several new technologies:

- The release of the W3C's Xforms 1.0 as a *recommendation*.

-
- The release of Microsoft's InfoPath forms technology as part of the general release of the latest version of Microsoft Office.
 - The release of new versions of eForms products from established vendors such as Cardiff, PureEdge, and Texcel. The product releases are addressing both Xforms and compatibility with InfoPath.
 - The upcoming release of Adobe's new *Forms Designer* product, which is intended as a complete overhaul of the product line that Adobe acquired from Accelio

It's important to note that these specific technology innovations are not happening in a vacuum. At least one broader technical trend is driving this focus, and there are several business drivers. *The broad technical trend* is the need for an improved experience for the end user as more and more applications rely on a browser interface. *N-tier* computing is here to stay, as more and more business applications are brought to portal and browser interfaces through the loose coupling of disparate systems and data sources. For presentation, HTML and Cascading Style Sheets are giving way to XML, XHTML, and more flexible and variable style sheet technologies such as XSLT and XSL-FO. For data structure, HTML forms will give way to Xforms.

The business drivers include the recent deadlines for HIPAA (the Health Insurance Portability and Accountability Act) and GPEA (the Government Paperwork Elimination Act). Both of these government initiatives have led to significant investments by healthcare organizations, insurers, and government agencies to digitize more content and information, automate more workflows, and ensure greater controls and protection over the information. And while these initiatives are important, and have led organizations to make specific changes, there clearly is a broad effort to not only bring more content and information under electronic workflows, but to also ensure that these new workflows are effective. Even small organizations are doing more in this direction and realizing short-term goals of cost savings on the way to longer-term goals of greater efficiencies and improved customer service.

The result is a healthy and active corner of the ECM market. It's worth looking at Microsoft's InfoPath to see if it represents another evolutionary step in this market, or something else.

INFOPATH & XML EVERYWHERE

Microsoft is not new to forms, of course. Various Microsoft end-user and developer tools have rich capabilities for forms development. This begins with business tools such as Word and Excel, and continues with HTML tools such as FrontPage, and development tools such as Visual Basic and others. What these other tools lack, despite their excellent overall capabilities, is a unified data model. Microsoft's own illustration for InfoPath shows how such disparate tools for collecting information and content can lead to the *stovepipe* problem in application development. As with the client-server development model, the

immediate problem is solved once a user interface is in place to connect users with the back end systems.



Figure 1. . Source: Microsoft

With XML-based forms, the information and content can be captured in a generic way that can then be integrated with multiple applications. Where the W3C's Xforms specification is an attempt to provide a generic data structure for eForms that everyone would use, InfoPath is an attempt to provide an XML-centric platform for forms development. As Microsoft's illustration shows, they see it as a general-purpose forms development environment, and, with the other primary applications of Microsoft Office, as the ubiquitous user interface for the business user.

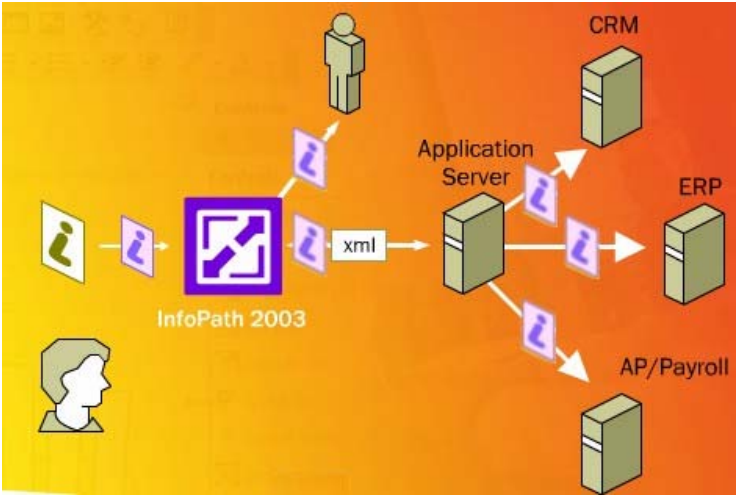


Figure 2. Source: Microsoft

InfoPath is a desktop application that uses XML as its native format, and provides features such as XML document editing, document viewing, and document input and output. InfoPath's working document is a *form template*, a file or set of files that specifies the form's data structure, appearance, and behav-

ior. Typically, the form template contains a schema and one or more XSLT files to define the structure and view(s) of the data to be captured by the forms. It can also contain image files, scripts, and something Microsoft calls a *manifest file* for containing form customizations. Once data is entered in the form, the form's contents are handled in memory as a DOM tree, making the core of InfoPath largely based on industry standard XML and closely related standards.

At the same time, InfoPath is a Microsoft client application. It requires installation of a client to each machine, and is tied closely to the Microsoft Office suite. The good news with that, of course, is that so many users are accustomed to the Office environment, and the Office installation includes a great deal of useful general purpose tools for rich text editing and formatting, table handling, and spell checking. The bad news is that the eForms market has been trending toward a thin client in the recent few years. Established eForms vendors that have been unable to deploy thin client solutions have lagged, and newer vendors that have embraced thin client approaches (such as PureEdge) have gained traction against older competitors.

It is also interesting that InfoPath is bucking another trend that many analysts had felt was a foregone conclusion—that not only would the browser be the ubiquitous user interface, but also that some kind of portal infrastructure would do the job of bringing specific application interfaces to the browser. Microsoft seems to be offering another strategy—that Windows, Office, and interface tools such as InfoPath will be the ubiquitous user interface.

Yet it appears Microsoft needs the Office environment—and the rather heavy client installation that goes with it—to accomplish one of the key features of InfoPath, which is its ease of use. Microsoft has designed InfoPath to be easy to use, and it includes a WYSIWYG design mode to let users design or modify *form templates* without having to interact with the XML schemas or XSLT scripts. InfoPath comes with a selection of canned forms, and forms can be designed from scratch using the WYSIWYG design tool. The WYSIWYG tool then generates the necessary XML schema, XSLT files, and related supporting files.

This emphasis on ease of use likely signals part of Microsoft's aim with InfoPath, which is to offer organizations an alternative to Acrobat and PDF. Despite Microsoft's overwhelming success with Microsoft Office, Adobe Acrobat has stubbornly maintained—and grown—its position as the primary means of representing and distributing office documents. While Adobe has been looking to eForms to further solidify the Acrobat franchise, Microsoft has positioned InfoPath as a more flexible platform for eForms development. The conventional wisdom in advance of InfoPath's release was that it was indeed more flexible but was perhaps more programmer friendly than user friendly. That Microsoft is emphasizing the WYSIWYG design and customization tools suggests that Microsoft is trying to answer the criticism that InfoPath is too programmer friendly. It also suggests that Microsoft is doing some positioning in advance of the release of Adobe's new forms designer

product, which is coming out in Beta this November and will be in commercial release next Spring.

One of the interesting applications of InfoPath is as a reporting mechanism. Because of its core reliance on XML schema and XSLT, InfoPath offers a powerful and flexible mechanism for generating ad hoc views of data. When coupled with the interface tools and widgets, InfoPath gives developers an impressive toolkit for forms generation. While Microsoft's FAQ stresses that InfoPath is not optimized for reporting—rather it is optimized for applications where the information and content need to be edited—developers will find a lot of opportunities to use InfoPath as a reporting tool.

ADOBE'S CONTINUING EFFORTS

Adobe, of course, has a long history in the eForms market, and they have been working with both the evolution of their core products and through acquisition to broaden that franchise. Because eForms has grown from the print incarnation of forms, Adobe has always had a presence in the eForms market. Beginning with Acrobat 3, users began using Acrobat for the distribution of forms. The IRS, for example, was an early and significant organization to begin distributing forms in Acrobat for distributed printing. End users would download and print important forms locally. Indeed, this was one of the early Web applications that drove the browser companies to aggressively develop plug-in interfaces for applications such as Acrobat.

Such an application, though, immediately raises the question of allowing people to view and fill in the form online. So while Acrobat 3 itself did not have a capability for filling out forms via a browser interface, Adobe did release a plug-in for Acrobat that did support forms, and then provided this functionality in Acrobat 4.

Acrobat 4 also introduced a dynamic format for creating forms in Acrobat—Forms Data Format, or FDF, which, like PDF, was an all-ASCII format that could be manipulated and generated using a toolkit. The toolkit supported C and Java, and enabled developers to create applications where forms could be generated and regenerated dynamically. While this proved very useful, this was also the timeframe in which XML was emerging as the *lingua franca* for data interchange on the Web. As a result, later versions of Acrobat focused on first opening up FDF to XML and later to making XML support truly generic. Thus, Acrobat 5 included XFDF, with XML-based forms encoding that supported a single DTD (XFDF), and Acrobat 6 includes support for any XML schema.

Adobe clearly views eForms as an important market. While Acrobat has perhaps the largest audience of any single software product, the majority of Adobe's end-user and developer products are marketed at professionals in various creative fields. The eForms market itself is perhaps almost as broad as the market for Acrobat. Nearly every kind of consumer interacts with forms of some kind or another; eventually the same thing can be said for eForms.

Thus Adobe's acquisition of Accelio made perfect sense from a market perspective. In Accelio they acquired an established player on both the server and client side of the eForms market. They also acquired a company with an XML focus and a great deal of XML technology, though the earlier versions of the Accelio product lines didn't necessarily feature it prominently. Acrobat's newer capabilities, for instance, of allowing arbitrary XML in and out of the document, are based on the Accelio technology.

Adobe's approach to eForms is intended to leverage all of their strengths in document creation and processing. Indeed, Adobe views the problem set of eForms and dynamic documents as being closely intertwined, so the clear long-term strategy for Adobe is to provide tools that allows a designer to create both documents and forms together. As such, it will be very interesting to see what Adobe does with the new version of its Forms Designer product, which is due out as a public Beta before the end of 2003.

Adobe's clear strength in the forms market will be to continue to leverage their unmatched capabilities in document distribution, and to stress the necessity of page fidelity in many eForms applications. While Adobe was one of the original contributors to the Xforms effort and is still involved, Adobe is counting on people caring about the presentation of the eForm. This is a pretty safe bet, in our eyes.

CONCLUSION: INFOPATH, ADOBE, AND THE COMPETITION

InfoPath is an impressive new offering, but it will not immediately dominate the eForms market. The early conventional wisdom is that it is good, solid 1.0 offering. InfoPath is also, intentionally, not a total eForms solution. Several of the existing eForms vendors have more comprehensive product offerings (*e.g.*, Cardiff's Liquid Office), some of them do a better job of providing a more open and standards-based solution (*e.g.*, PureEdge), and several of them successfully deliver *page fidelity* (what others might call "pixel perfect form") to the original paper forms. Finally, InfoPath will propagate with the latest version of Microsoft Office. It typically takes more than a year for the latest version of office to replace earlier versions on the majority of desktops.

In the meantime, Adobe will be introducing their new Forms Designer product, and continuing to emphasize the need for page fidelity and presentation in eForms applications. Adobe also can already point to the significant number of applications that already leverage Acrobat and the product lines they added in the Accelio acquisition.

The good news for the eForms market is that Microsoft and Adobe bring new strategic thinking to what has been a relatively small market. InfoPath will have the immediate effect of bringing eForms to the attention of the CIO, and will help bring a new focus to improving the client experience for the business user. As organizations deploy more applications to a distributed

workforce and partners, eForms will become a more strategic piece of the ECM mix.

Indeed, eForms have a growing role beyond ECM itself, as they are emerging as the primary interface between people, process and programs. It is no accident that the significant initiatives now—Sarbanes-Oxley, HIPAA, and the like—are forms-centric. Moreover, initiatives such as Sarbanes-Oxley are all about improving business process management while making access to both content and data more transparent and comprehensive. To this end, eForms must continue to evolve from a standalone artifact to a flexible interface intimately connected to enterprise infrastructure. The implications of this are profound. The vendors and organizations that can successfully manage this evolution will realize more success, more quickly, and will lead the next wave in integrated content and information technology.

Some Representative eForms Vendors	
Company	Comments
Cardiff, http://www.cardiff.com	Established provider of eForms solutions; maker of a comprehensive eForms solution, LiquidOffice.
Adobe, http://www.adobe.com	A growing presence in the eForms space through Acrobat and acquisition of Accelio.
Microsoft, http://office.microsoft.com/infopath	A newly focused presence in the eForms space with the introduction of InfoPath
PureEdge Solutions, http://www.pureedge.com/	Established provider of eForms solutions. Along with Cardiff, has taken active role in development of Xforms specification.
Texcel Systems, http://www.texcel.com	Smaller but growing company that provides tools for converting various legacy document formats into eForms.
Mosquito orgabit GmbH, http://mozquito.markuplanguage.net	mozquito is the maker of the DENG XBrowser. DENG is a modular XML browser, that supports XForms, SVG, XHTML, arbitrary XML with CSS, XFrames and any other custom XML namespace.
ScanSoft, http://www.scansoft.com/	ScanSoft is a supplier of productivity software with concentrations in imaging, speech, and language solutions. Their Omni-Page and PDF Converter tools are used for converting documents to various eForm formats.

Bill Trippe, bill@gilbane.com