

(or)

*How XML-based portals can allow you to develop information delivery systems in weeks instead of years and often without requiring application programmers*

As a DAMA International Advisor and an associate professor of information systems at a Carnegie Research-I University, I am often asked about emerging trends in information management. One of the most exciting I have worked with recently is the promise of XML-based portals. During 1998, I was fortunate enough to co-author the book *Building Corporate Portals Using XML* (1999 McGraw-Hill) with Clive Finkelstein. In the years since we wrote the text, a number of significant development trends have occurred in the portal field that reinforce my initial excitement.

The somewhat casual and tongue-in-cheek title to this short paper is designed to draw your attention to two specific points.

1. Portals can be effective and efficient information delivery vehicles for organizations to employ. However because *XML*-based portals are based on some of the most promising integration technologies, they are of particular interest to information managers. Of most importance, they represent a single technology that can be deployed to handle the majority of current and future information delivery needs in a manner that can easily be demonstrated to be cost effective.
2. Combined with an XML-enabled environment, information managers are able to deliver personalized information to knowledge workers with a 10X reduction in time, money, and effort – when compared with more traditional approaches. In many instances this business values is delivered virtually without significant programming requirements.

Organizations are claiming hundreds of millions of US dollars annually in savings from their XML-based portals as they provide effective information integration and delivery for hundreds of uses. I believe the current state of affairs presents a somewhat unique opportunity for information management professionals to rapidly demonstrate business value – something has not always been easy.

Everyone agrees on the utility of portal technologies as a means of delivering lots of information content via a single interface. A prominent example is yahoo.com – seeking to seamlessly deliver text, music, and video – content in any form – indeed its attempt is to be a single point of access for all an individual's web content access. When you add XML to a portal – creating an XML-based portal – you are add the ability to incorporate virtually any XML-wrapped data to the scope of the portal information management capabilities.

XML-based portals differ from portals that are not based on XML in significant ways including the ability to directly integrate information from previously un-integrated information sources. That is – XML-based

The Institute for Data Research (IDR) is a research and development center located in Richmond, Virginia and affiliated with Virginia Commonwealth University. The Institute is an interdisciplinary research organization founded to further the study of data and metadata as valuable organizational assets. The key to IDR's successful practice of creating business value is the application of innovative data engineering and management techniques for recovering, understanding, improving and re-using organizational metadata assets.



# XML-based Portals Save Money

(Page 2 of 3)

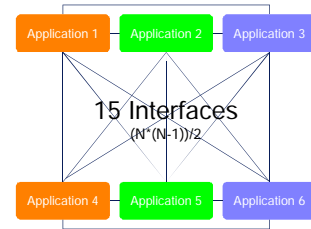


INSTITUTE FOR DATA RESEARCH

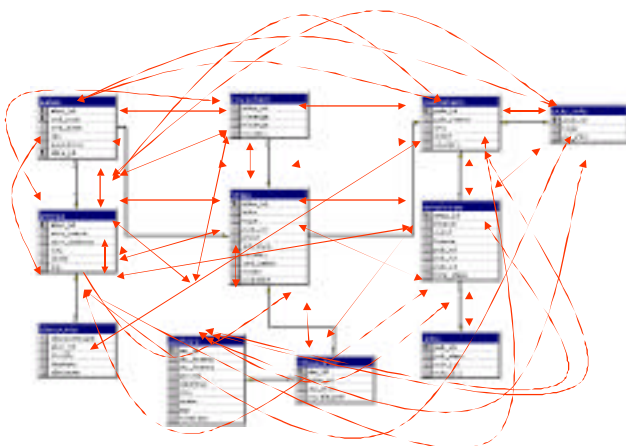
portals can "understand" the potential integration that can occur among any similar, XML-wrapped data. This integration can occur in response to user drag-and-drop requests without the portal first having to know that the user might want to combine the two formerly disparate pieces of data.

Consider an example: a requirement calls for integration of data from two data sets that have not been previously integrated. Prior to XML there were four options: 1) do nothing 2) integrate the first data set into the second 3) integrate the second data set into the first 4) combine both data sets into a third data set. The major weakness with these approaches has been that all require development of situation-specific integration solutions. That is whatever solution is developed; it will be specific to that data integration problem.

XML technologies allow another option – the two data sets can be wrapped in metadata. Data integration is reduced from an  $(N - (N - 1))/2$  problem to a hub and spoke model. Each new data set that is wrapped in XML becomes a new spoke that can exchange data with all other spokes via the hub. XML wrapping allows two disparate data items to be integrated using an XML-based portal – which can also act as the hub of subsequent data integration.



In traditional information delivery, possible queries against a data set must be anticipated. After a prioritization screening, some subset of queries are made available to the knowledge workers via SQL, 4<sup>th</sup> GL languages, warehousing technologies, etc. Others queries remain unmet.



(Adapted from SAP's TopTier Portal Presentation)

Making a data set's metadata known to the portal is accomplished when the data set becomes wrapped in XML that correctly describes the characteristics of the data. The XML tags form the basis for automated integration by the portal and other technologies. Key to effectively managing this process is that the tags are mapped to transformation rules and both the set of tags and the transformation rules are managed by modern data management technologies. In this manner organizational data management evolves from interface-based to integration-based. Unlike ERP deployment, organizational costs begin to noticeably decrease as soon portal technologies are deployed.

The industry is still at the early stages of product innovation – implementation costs will continue to decrease! XML-based portals essentially make any valid set of data combinations that are understood by the portal to be feasibly accessed using browser-based interface technologies (i.e., reports can be bookmarked). Using XML as a means of precisely identifying the content and structure of an existing data set, data integration can evolve from a labor-intensive, one-time solution to a structured means of evolving toward programmatic management of organizational metadata. XML-tagged repositories will soon succeed the current state of ill-suited organizational repository technologies (as will a new model for organizational computer aided system engineering (CASE) technologies)!

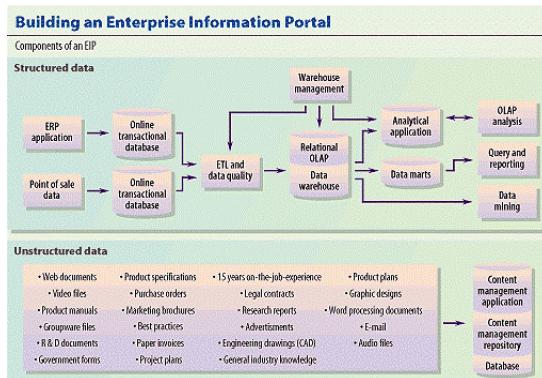
Three other aspects of the current state of the practice combine to enhance the attractiveness of XML-based portals.

# XML-based Portals Save Money

(Page 3 of 3)



INSTITUTE FOR DATA RESEARCH



First, an expanded definition of data management. Experts estimate that only 20% of organizational data is managed by modern data management technologies such as RDBMS. The remaining 80% of organizational data is stored in items such as purchase orders, PowerPoint presentations, legal contracts documents, and of course let us know forget the ubiquitous spreadsheet. The implementation of XML-based integration as part of the Office 2000 and Office 2001 suites sets a standard by which other package will have to begin including XML as the basis for their document management technologies.

Saving an Office 2000+ document to a BizTalk (by Microsoft) server now makes the wealth of automatically created document content and metadata accessible to organizational searchers. (See the various "File -> Properties ..." tags.) This development alone greatly increases the range of information over which information management professional can extend the reach of modern data management practices. Now we can include previously unstructured organizational data available via our intranets using XML-based portals.

Second, is an expanded role for information managers in this newly defined XML-based portal environment. Now you are responsible and known to be capable of delivering information by searching through the larger, formerly unstructured portion of organizational data. Consider your domain increasing by a factor of four. In addition, as the ROI on strategic XML-based portal investments becomes more widely known, organizational will increasingly demand that their existing environment be migrated into XML-portal environments. Beyond the metadata engineering is the application engineering – possible in the XML-based portal environment because so much can be accomplished via direct metadata tag integration. For application developers this will be an unfamiliar paradigm and they will require leadership and training or they might not succeed.

Third, as organizations demand increased means of accessing more and more organizational data, organizational data quality issues will become increasingly important. Particularly for e-transactions – where data errors move at the speed of the Internet – data preparation will become an increasingly programmatically managed activity and the price of errors will multiply. The price of not participating in the various online communities and exchanges may rapidly become too high for some. XML-based portals give organizations an opportunity to begin branding organizational data as meeting certain quality standards. Using the proper semi-automated data engineering analysis technologies organizations must be prepared to successfully support this evolution.

All of these developments are coming to make this one of the most opportune times to work in information management. At last we have our tools (XML-based portals), our techniques (data engineering), and our technologies (TCP/IP-based information deliver) ready to tangibly and noticeably contribute to our organizations bottom line. Let's get started!