Intellor Research Summary

XML Database
Trends And Influences

Intellor Group, Inc.
Six Montgomery Village Avenue,
Suite 400
Gaithersburg, MD 20879
Phone: 301-208-6766
Fax: 301-208-6768
www.Intellor.com

Situation Analysis

This research summary reviews the responses of executive and IT professionals from a survey conducted by Intellor Group and Wilshire Conferences at the Wilshire Meta-Data Conference and DAMA Symposium that they completed about their current and future XML initiatives. The focus is on how the need to manage unstructured data, in addition to structured data, is influencing the adoption of XML databases.

Intellor Group shall publish additional research summaries on the acceptance, understanding, and priority of XML adoption. These research summaries will be made available at no cost to members of Intellor.com.

To register to become a member of the Intellor.com eBII Web community, visit http://www.intellor.com/links/?pg=RF. Membership is FREE to IT professionals.

Background

The Focus

The majority of traditional data warehouses, transactional systems and enterprise applications rely on relational databases to maintain their data. As portals, knowledge management systems and even humble e-mail have joined the mainstream and become indispensable daily tools, a typical organization's enterprise information is no longer maintained as structured data alone — much of it is now in the form of unstructured data as well.

Structured data is typically data with regular repeated structure that can be easily stored in the data tables of a relational database. Unstructured data constitutes everything else — such as text documents, e-mails and web pages that have no regular structure.

This Intellor Research Summary examines respondent's opinions about:
- How widespread the management of unstructured data has become
- The use of XML to manage all data, structured or not
- Whether or not it makes sense to store XML in native XML databases or are traditional databases, primarily relational, up to the task.

This builds on the results of a previous Intellor Research Summary, called XML Adoption: Benefits And Challenges (http://www.intellor.com/links/?pg=RS4), where 58% of respondents rated "Common data access techniques for structured and unstructured data" a benefit of XML adoption.

Background
(continued)

The 232 respondents were attendees of the Wilshire Meta-Data Conference and DAMA Symposium on March 4-8, 2001.

The Questions
Attendees were asked to indicate their responses to the following questions:

1.) Assess the ratio of expenditure of your organization's IT resources on the management of structured data versus unstructured data:
- Today
- Within One Year
- Within Three Years

Responses were selected from:
- 90% Structured – 10% Unstructured
- 80% Structured – 20% Unstructured
- 70% Structured – 30% Unstructured
- 60% Structured – 40% Unstructured
- 50% Structured – 50% Unstructured
- 40% Structured – 60% Unstructured
- 30% Structured – 70% Unstructured
- 20% Structured – 80% Unstructured
- 10% Structured – 90% Unstructured

2.) Indicate the impact the convergence trend between mark-up languages and data management has on your organization.
- Today
- Within One Year
- Within Three Years

Responses were selected from:
- Do Not Know
- No Impact
- Minimal Impact
- Key Criteria In Any Decision

3.) Rate the implementation plans of the following in your organization:
- Native XML databases (XML is stored as XML internally in the database)
- XML-enabled databases (XML is stored in conventional DBMS formats internally in the database)

Responses were selected from:
- Do Not Know

Background
(continued)

- No Plans To Implement
- Implementation Started During Last 12 Months
- Plan To Implement In Next 12 Months

Respondent Demographics

The respondents represent over twenty-five (25) different industries ranging from Aerospace to Utilities. No single industry provided more than 20 percent of the respondents. Government is the single largest category with 16 percent of the respondents. Respondents are split 52 - 48 percent between companies with annual revenue above or below $1 billion. Titles range from CEO to IT programmer, and 14 percent of the respondents have a title of director or higher.
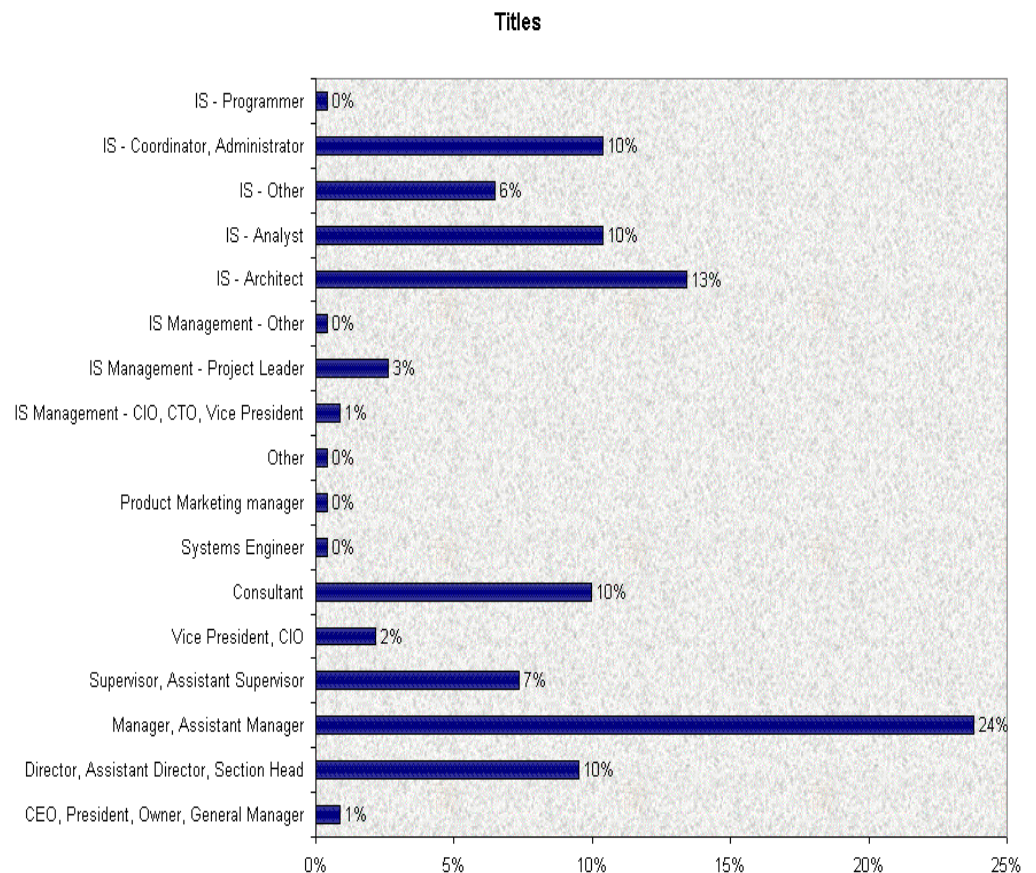
Titles

Title	Percent
IS - Programmer	0%
IS - Coordinator, Administrator	10%
IS - Other	6%
IS - Analyst	10%
IS - Architect	13%
IS Management - Other	0%
IS Management - Project Leader	3%
IS Management - CIO, CTO, Vice President	1%
Other	0%
Product Marketing manager	0%
Systems Engineer	0%
Consultant	10%
Vice President, CIO	2%
Supervisor, Assistant Supervisor	7%
Manager, Assistant Manager	24%
Director, Assistant Director, Section Head	10%
CEO, President, Owner, General Manager	1%

Figure 1 -
Respondents By Title

Background
(continued)

Industries

Education — 2%
Other — 1%
Utility — 2%
Transportation — 2%
Telecommunications — 6%
Retail — 5%
Print/Publishing — 1%
Pharmaceutical — 2%
Petrochemical — 2%
Medical — 1%
Manufacturing — 3%
Insurance — 13%
Healthcare — 6%
Government — 16%
Financial Services — 12%
Entertainment — 0%
Consumer Packaged Goods — 1%
Consulting — 6%
Computer - Software/Services — 9%
Computer - Hardware/Services — 0%
Computer - ASP — 1%
Chemical — 0%
Banking — 4%
Automotive — 1%
Agriculture — 0%
Aerospace — 3%

Figure 2 -
Respondents By Industry

Background
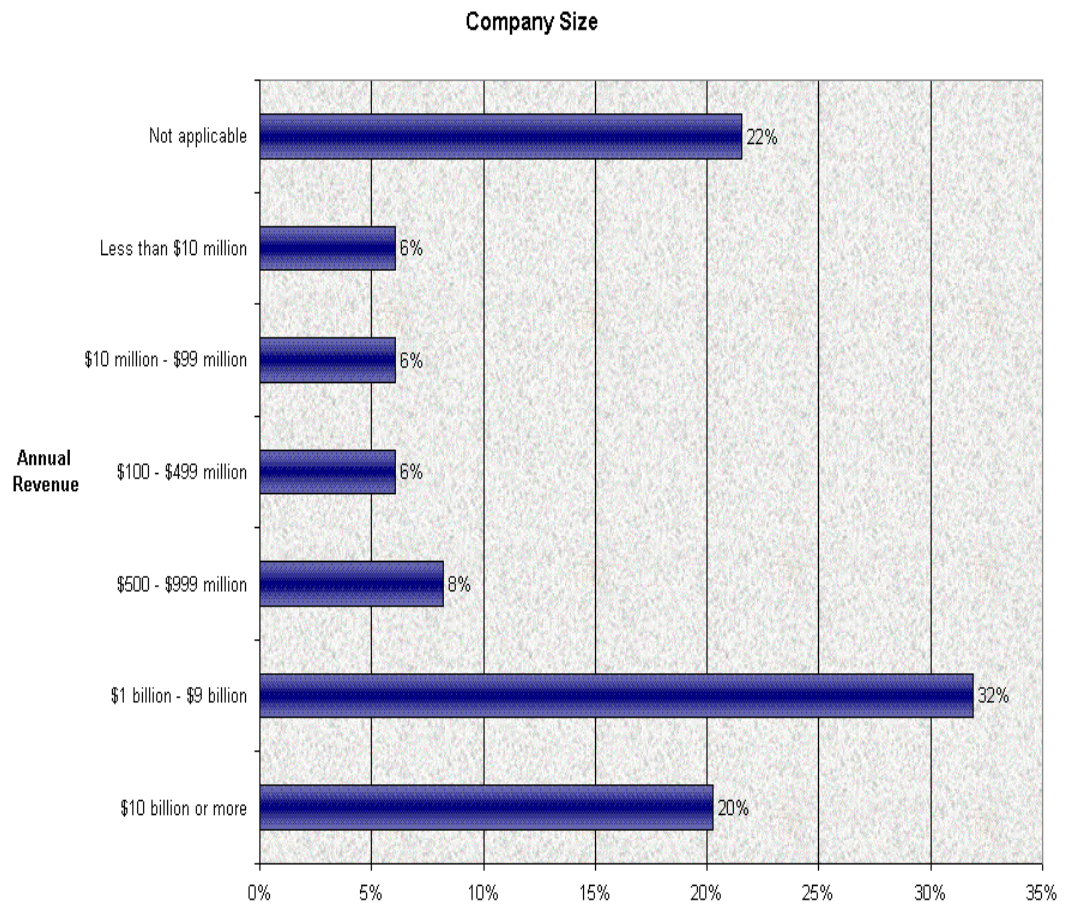**(continued)**

Company Size

Figure 3 -
Respondents By Company Size

Managing Unstructured Data

How Important Is Management Of Unstructured Data?

Respondents were asked to rank the ratio of expenditure of IT resources on the management of structured versus unstructured data. They were asked to select from nine ratio percentage bands in answering this question and apply it to their organizations today and also estimate what they expect the situation to be in one year and in three years.
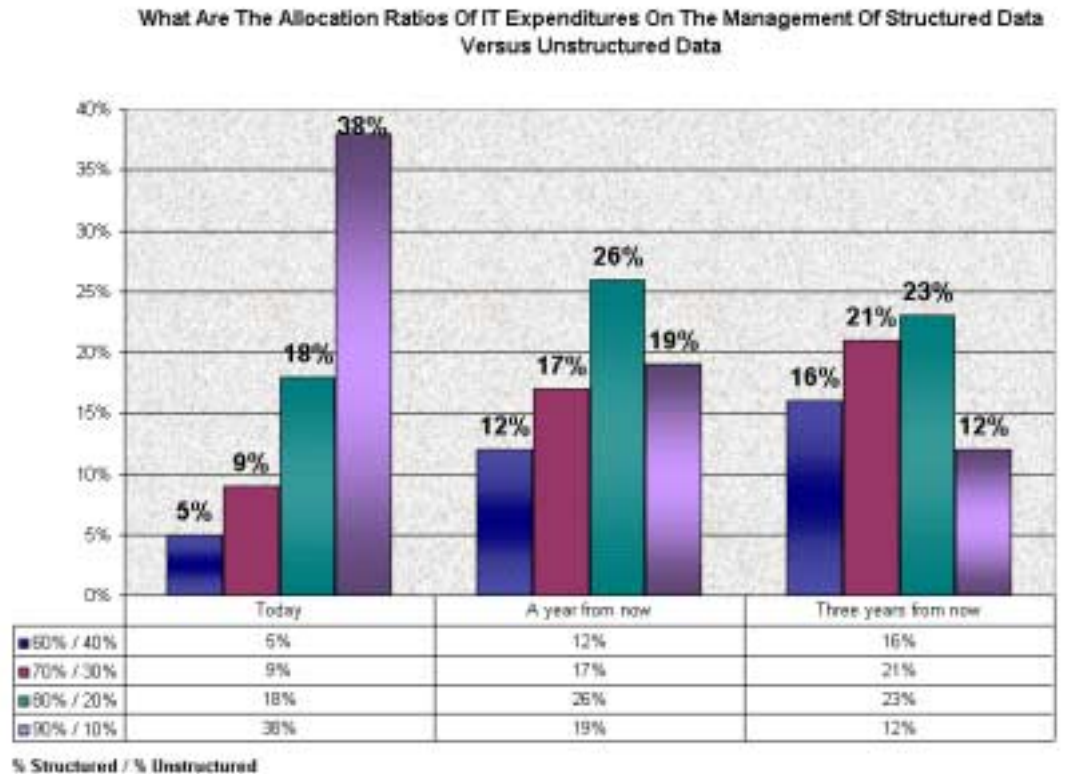
What Are The Allocation Ratios Of IT Expenditures On The Management Of Structured Data Versus Unstructured Data

	Today	A year from now	Three years from now
60% / 40%	5%	12%	16%
70% / 30%	9%	17%	21%
80% / 20%	18%	26%	23%
90% / 10%	38%	19%	12%

% Structured / % Unstructured

Figure 4 -
Allocation Ratios Of Actual And Anticipated IT Expenditures On
The Management Of Structured And Unstructured Data

The Situation Today

Thirty-eight percent of respondents claim a 90 percent/10 percent split in IT resource allocation today between structured data and unstructured data management. This contrasts with 18 percent who claim an 80 percent/20 percent split today, 9 percent with a 70 percent/30 percent split today and 5 percent with 60 percent/40 percent split today. Clearly, the bulk of respondents belong to organizations with extreme disparity in resources allocated to manage structured data versus resources allocated to manage unstructured data.

**Managing
Unstructured
Data**
(continued)

Expectations For The Future

When respondents were asked the same question based on their expected IT resource expenditure ratios a year from now and three years from now, the numbers tell a different story.

Nineteen percent of respondents expect a 90 percent/10 percent split in their IT resource allocation a year from now between these two forms of data management. This contrasts with 26 percent who expect an 80 percent/20 percent split, 17 percent who expect a 70 percent/30 percent split and 12 percent who expect a 60 percent/40 percent split. Clearly, the trend is towards a greater share of resources being allocated to *unstructured* data management in future.

The numbers for expected IT resource expenditure ratios in three years confirm the trend. Twelve percent of respondents expect a 90 percent/10 percent split, 23 percent expect an 80 percent/20 percent split, 21 percent expect a 70 percent/30 percent split and 16 percent expect a 60 percent/40 percent split.

The trend is for fewer and fewer organizations to have such an extreme 90 percent/10 percent split in resource allocation. This indicates that management of unstructured data is moving out of the niche role it occupies today in many organizations, and is going to occupy an increasingly larger share of most IT departments' time, energy and money.

Note: Approximately 23 percent of respondents voted "Do Not Know" to this question. The analysis above is based on the responses of the remainder constituting the total. The analysis does not include the ratio percentage bands below 60 percent/40 percent as they had no bearing on the outcome.

**Convergence Of
Mark-Up
Languages And
Data Management**

How influential Is The Convergence Of Mark-Up Languages And Data Management?

Respondents rated the impact that the convergence of mark-up languages and data management has on their organizations today and is expected to have in a year and in three years.

It is clear that expectations are high that XML is going to revolutionize IT structured and unstructured data management. This survey question aimed to quantify that influence in terms of the current and expected impact on IT decision-making.

Convergence Of Mark-Up Languages And Data Management
(continued)

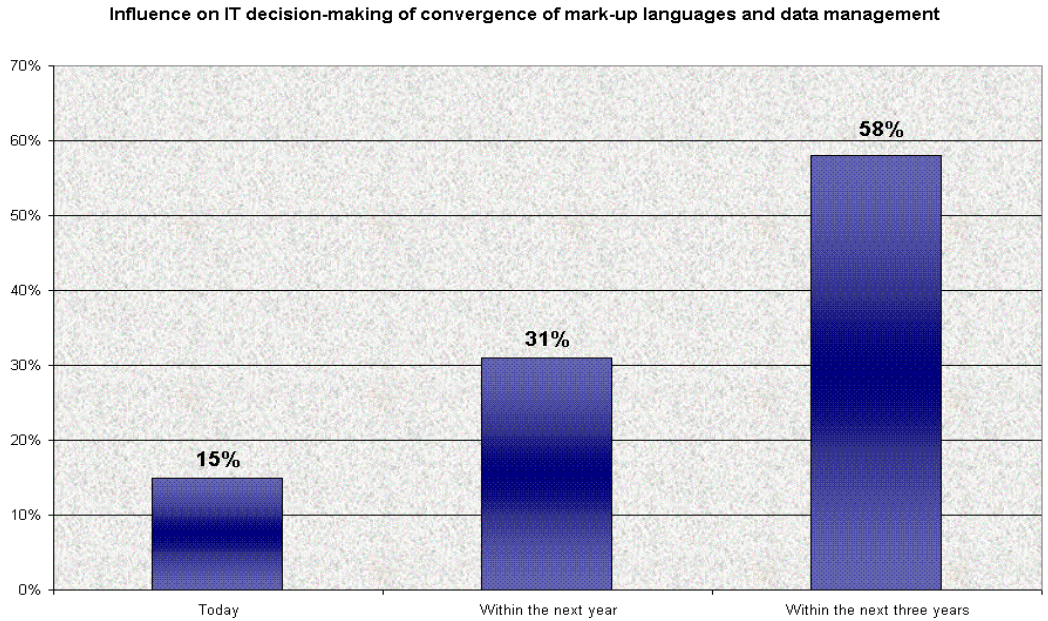Influence on IT decision-making of convergence of mark-up languages and data management

Figure 5 -
Shows the percentage of respondents who voted the convergence of mark-up languages and data management to be a "Key Criteria In Any Decision"

Respondents who consider the convergence of mark-up languages and data management to have little or no impact on decision making

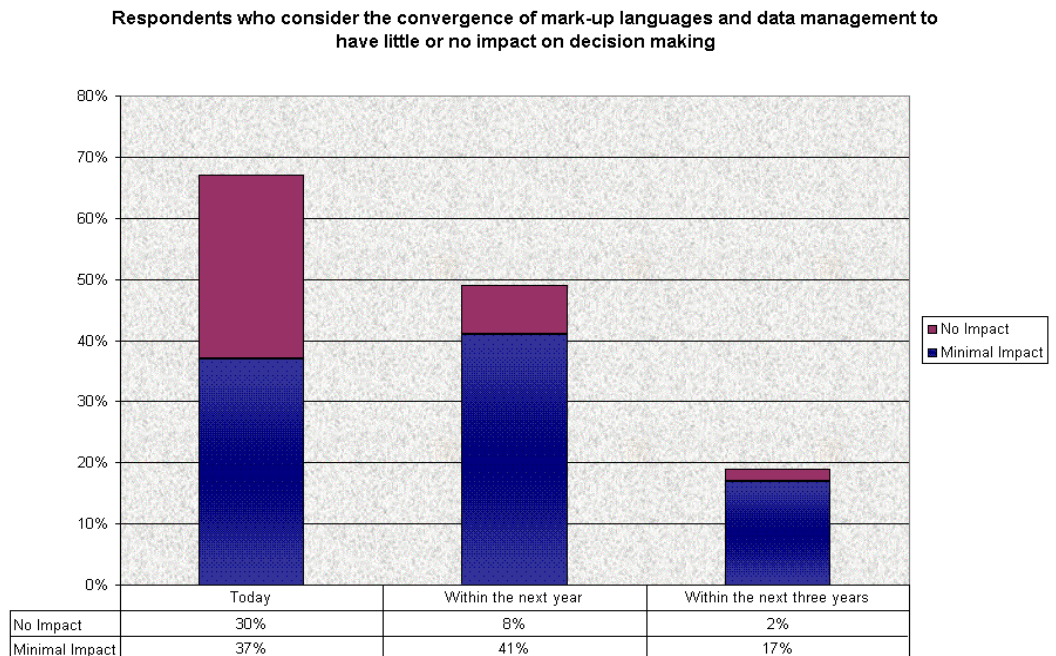|	Today	Within the next year	Within the next three years
No Impact	30%	8%	2%
Minimal Impact	37%	41%	17%

Figure 6 -
Shows the percentage of respondents who voted the convergence of mark-up languages and data management to have either "No Impact" or "Minimal Impact" on their organization.

Convergence Of Mark-Up Languages And Data Management
(continued)

Today, 15 percent of respondents believe that the convergence of mark-up languages and data management is a "Key Criteria In Any Decision", and 67 percent believe that this convergence has "Minimal Impact" or "No Impact" at all on their organization.

Thirty-one percent expect that a year from now, this convergence shall be a "Key Criteria In Any Decision", and 49 percent believe it will have "Minimal Impact" or "No Impact" at all on their organization.

Projecting forward three years into the future, 58 percent expect that this convergence shall be a "Key Criteria In Any Decision", and 19 percent believe that it will have "Minimal Impact" or "No Impact" on their organization.

These percentages speak for themselves. We are in the midst of profound changes in IT thinking about data management. Mark-up languages, as personified by XML, are reaching critical-mass acceptance in being perceived as moving from their original role to one that permeates multiple types of IT application environments.

Note: Approximately 21 percent of respondents voted "Do Not Know" to this question. The analysis above is based on the responses of the remainder constituting the total.

XML Databases

What Role Do XML Databases Play?

The two questions above establish the emergence of unstructured data and the application of common XML tools to blur the distinction and differentiation between handling structured and unstructured data. In this section, we address the most interesting topic of all and one that database vendors are clamoring to answer and influence—How do IT organizations plan to store XML data?

XML Database Background

XML database vendors fall into two camps:
- Native XML databases and
- XML-enabled databases.

The following definitions, assertions and assumptions constitute some of the prevailing, evolving, conventional, and, sometimes, conflicting wisdom about XML databases:
- Native XML databases are new custom-architected databases designed from the ground-up to manage XML

XML Databases
(continued)

and which allow XML documents to be stored as XML internally.

- XML-enabled databases are defined as conventional relational or object-oriented databases that have been fitted with some kind of front-end XML adaptor to manage the storage of data from XML documents.
- Users are familiar with and understand the behavior of relational or object databases with regard to performance. On the other hand, native XML databases are an unknown quantity with regard to scalability.
- XML documents have a hierarchical structure, which is foreign to the relational model, resulting in complex mappings with processor-intensive translation having to be performed each time an XML document is stored or retrieved from an XML-enabled database.
- XML documents can be either data-centric or document-centric. Data-centric equates to structured data and document-centric equates to unstructured data using the traditional terminology from earlier in this summary.
- Determining what type of XML documents predominate in the systems you are building is an important factor in deciding whether native XML or XML-enabled databases suit your application.
- Data-centric XML documents are more appropriate for XML-enabled databases.
- Document-centric XML documents are more appropriate for native XML databases.

This list above is shared without comment or endorsement. Except, it should be emphasized that it is still very early in the debate, and no conclusions can be drawn with any assurance that they are based on long-term, real-world experience.

XML Databases
(continued)

XML Database Implementation

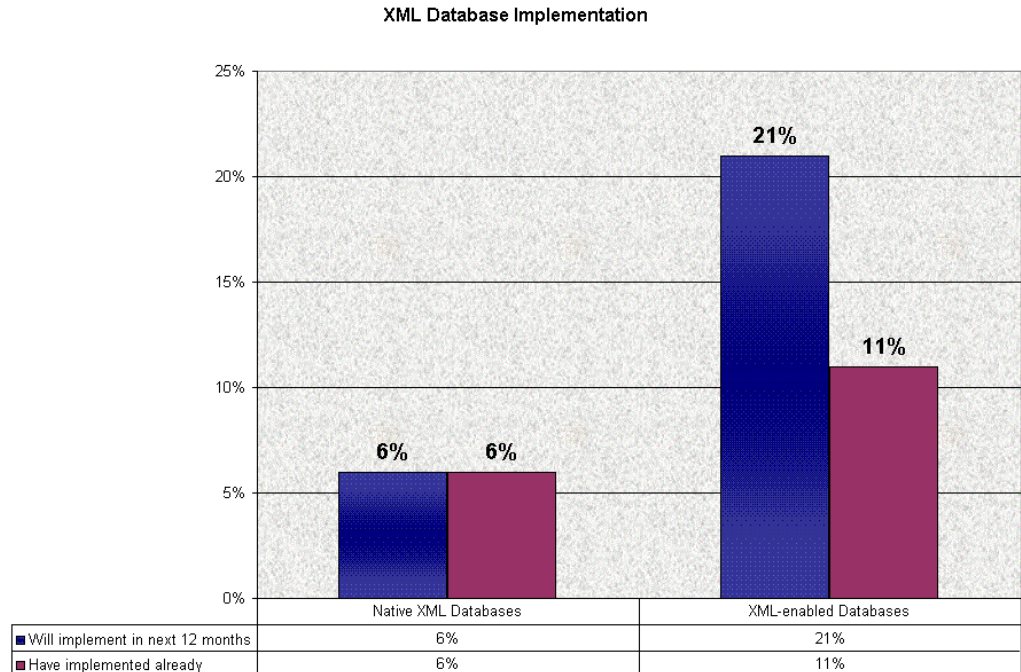|	Native XML Databases	XML-enabled Databases
■ Will implement in next 12 months	6%	21%
■ Have implemented already	6%	11%

Figure 7 -

Shows percentages of respondents who either have or intend to implement XML databases.

XML Database Adoption Trends

Twelve percent of respondents have already implemented native XML databases or plan to implement them within the next 12 months. The corresponding percentage for XML-enabled databases is 32 percent, nearly three times greater.

This finding should surprise no one given the completely new, bleeding edge aura that surrounds native XML database technology and contrasting this with the feeling of familiarity engendered by the well known, established relational database market. Even though the percentage is significantly higher for XML-enabled databases, these vendors should not become complacent.

Native XML technology is advancing in leaps and bounds, and it will not always be at the disadvantage of not having a proven track record. There is the sense that the primary claim of native XML database proponents that XML documents should be stored as XML internally in a database for best performance and management has an intuitive appeal and cannot be lightly dismissed. In all likelihood native XML databases will do well in

XML Databases
(continued)

certain applications such as content-management data stores and XML-enabled databases will predominate in more data-centric XML applications.

Note: Forty-four percent and 38 percent of respondents voted "Do Not Know" to the native XML and XML-enabled database questions respectively. The analysis above is based on the responses of all respondents.

About
Intellor Group, Inc.

About Intellor Group, Inc.

Intellor Group, Inc., located in Gaithersburg, Maryland, is an independent knowledge exchange company focused on promoting the success of organizations' e-business initiatives in the e-Business Intelligence and Integration (eBII) marketplace through the rapid and comprehensive exchange of knowledge.

Intellor Group's unique business model connects companies seeking knowledge with vendors, thought-leaders and peers who can provide that knowledge. Intellor Group provides a next-generation learning environment for Information Technology (IT) users and solution providers that surpasses the capabilities of traditional learning methods by combining the input and dissemination power of the Internet via an Internet community-based knowledge exchange – www.intellor.com – with real-world education and in-depth, community-based research and analysis.

For more information about Intellor Group's knowledge exchange offerings, visit Intellor Group at www.intellor.com, or call 301-208-6766, x116. **To register to become a member of the Intellor.com eBII Web community visit** www.intellor.com **. Membership is FREE to IT professionals.**

Why Should You Join The Intellor Group's eBII Web Community?

Ten Great Reasons To Join Intellor.com:

- Interact with your peers who have solved your challenge(s) before.
- Get the no-marketing-hype scoop on products, services and solutions.
- Eliminate time-consuming Internet surfing to research technologies and solutions.
- Access to real critiques of products and services from people like you.

About Intellor Group, Inc.
(continued)

- Aid your decision-making process, and reduce your technology investment risk.
- Expedite approvals and funding for projects through access to the information necessary to make your business case.
- Get opinions from industry experts.
- Influence product development with your critiques and feedback to industry vendors.
- Showcase your success, and gain industry recognition.
- It's **FREE** to IT professionals!

About Wilshire Conferences

About Wilshire Conferences

Wilshire Conferences provides high quality educational conferences to Information Technology (IT) professionals. In particular, Wilshire focuses on programs in the areas of data management and application development.

The company's philosophy is to provide educational environments that allow a high degree of interaction between participants. Wilshire believes that shared experience among peers provides the most cost-effective learning opportunity for IT professionals.

For more information, visit Wilshire Conferences at www.wilshireconferences.com.