
The Role of XML in Information Portals

Sangayya S Sirurmah

Ashis Kumar Pani

Abstract

XML based portals are specifically delineated subsets of enterprise portal technologies that provide inherent XML based data exchange and interface capabilities to access information. Portals act as information distribution hubs, providing access to the XML wrapped metadata. This paper analyzes the impact of XML on Portal Application Integration (PAI) architecture and portal systems as a whole. It also described the XML in databases and content management. Finally highlighted the solutions, search functionality and future of portals for accessing information.

Keywords : Information Portals, Data Exchange, Standards, XML

0. Introduction

XML is an open standard for expressing structured content and for structuring unstructured content that has been promulgated by the World Wide Web consortium. It is a standard method for defining "tags" that may be used to describe documents. Moreover, it is not a fixed collection of tags, but a standard for constructing meta-languages. XML is also an information exchange format, because it's tagged instructions for applications and its document type definitions provides context to either the encoded data or encoded content in XML documents or messages.

Information is frequently data extracted, filtered, or formatted in some way. An XML message or document is a perfect example of formatted raw content or data that is of information. XML is a powerful trend because it is increasingly accepted as a standard of exchange among differing nonstandard formats. It is thought by many to provide the ultimate answer to the "Island of information" problem.

The software industry is going through the process of creating interfaces between every other format and XML. It has structuring unstructured content by modeling it using XML tagging, document type definition specifications, and in the end object modeling.

Presently portal product manufactures have joined the XML bandwagon. Information portal products are also very much immersed in XML technology and has the capability to convert content into XML format. This paper analyzes the impact of XML on Portal Application Integration (PAI) architecture and portal systems as a whole.

1. Background of XML

The term 'markup' dates back to the days before electronic documents, when publishing professionals used to take basic text from authors on paper and write markup instructions to tell the typesetter how to make the document look good on the final printed page e.g. make this world bold, double space etc.

The digital world is having different kinds of markup languages. Firstly, we have the proprietary markup languages used by word processor, desktop publishing packages, etc. Then we have open nonproprietary markup languages such as text, troff, etc., - the most famous nonproprietary markup language is HTML (Hyper Text Markup Language).

XML (eXtensible Markup Language) is also a markup language. It is not a software programme and thus does not do anything unaided. XML provides a standard approach for describing, capturing,

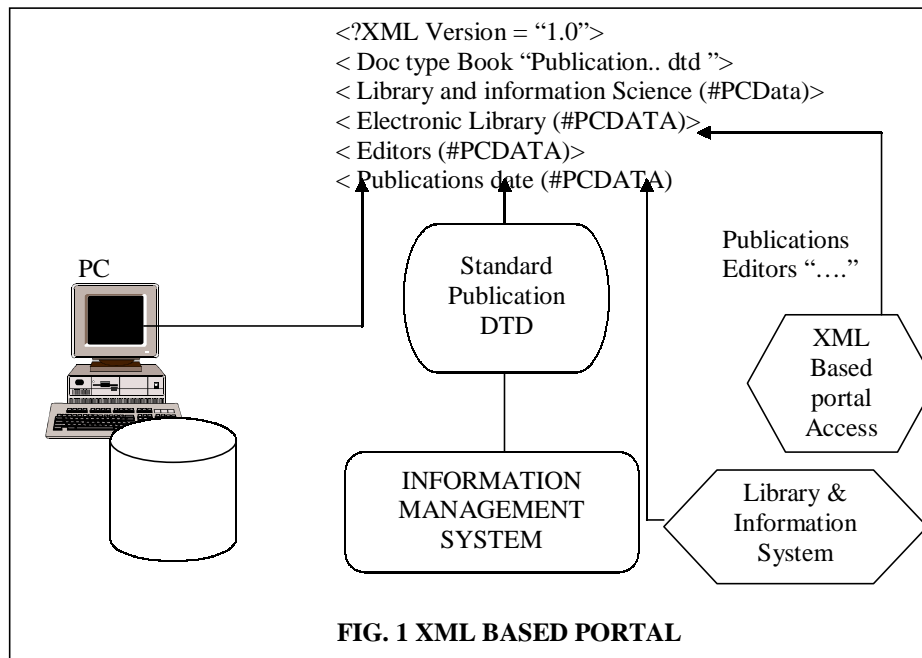
processing and publishing information. It is a language that has significant benefits over HTML. HTML is a fixed markup language. That is to say, it provides a certain feature set in their markup, and that set is fixed in the design of the language. HTML, for example, has a fixed set of tags with which we craft our documents — <H1> <P>, <Table>, etc. XML on the other hand, does not define any particular set of tags. Rather, it provides a standardized framework with which to define our own, or to use those defined by other that befit their needs.

The core philosophy of XML has come about as result of a long and thoughtful analysis of what is really mean by the term “Document” in the digital world. By and large, documents consist of three distinct components, viz., data content, structure and presentation. The central idea of XML is that significant benefits accrue to the document owner if those three aspects of a document are kept separate and made explicit in a computer system.

2. XML based Portal Technologies

XML based portals are specifically delineated subsets of enterprise portal technologies that provide inherent XML based data exchange and interface capabilities to all knowledge workers who access the technology. Disparate, unforeseen data integration queries and exchanges can be accomplished by having the portal interpret the query, translate it in to XML and access the data required to supply the requested information. The result is based on an understanding of XML and the associated Meta data wrappers or tags. Portals act as information distribution hubs, providing access to the XML wrapped metadata that describes the legacy environment. No longer is integration handled on a case-by-case, legacy system-to-legacy system basis, but instead on a portal integration basis. The single fixed target mapping on to the portal environment simplifies all aspects of this previously difficult task.

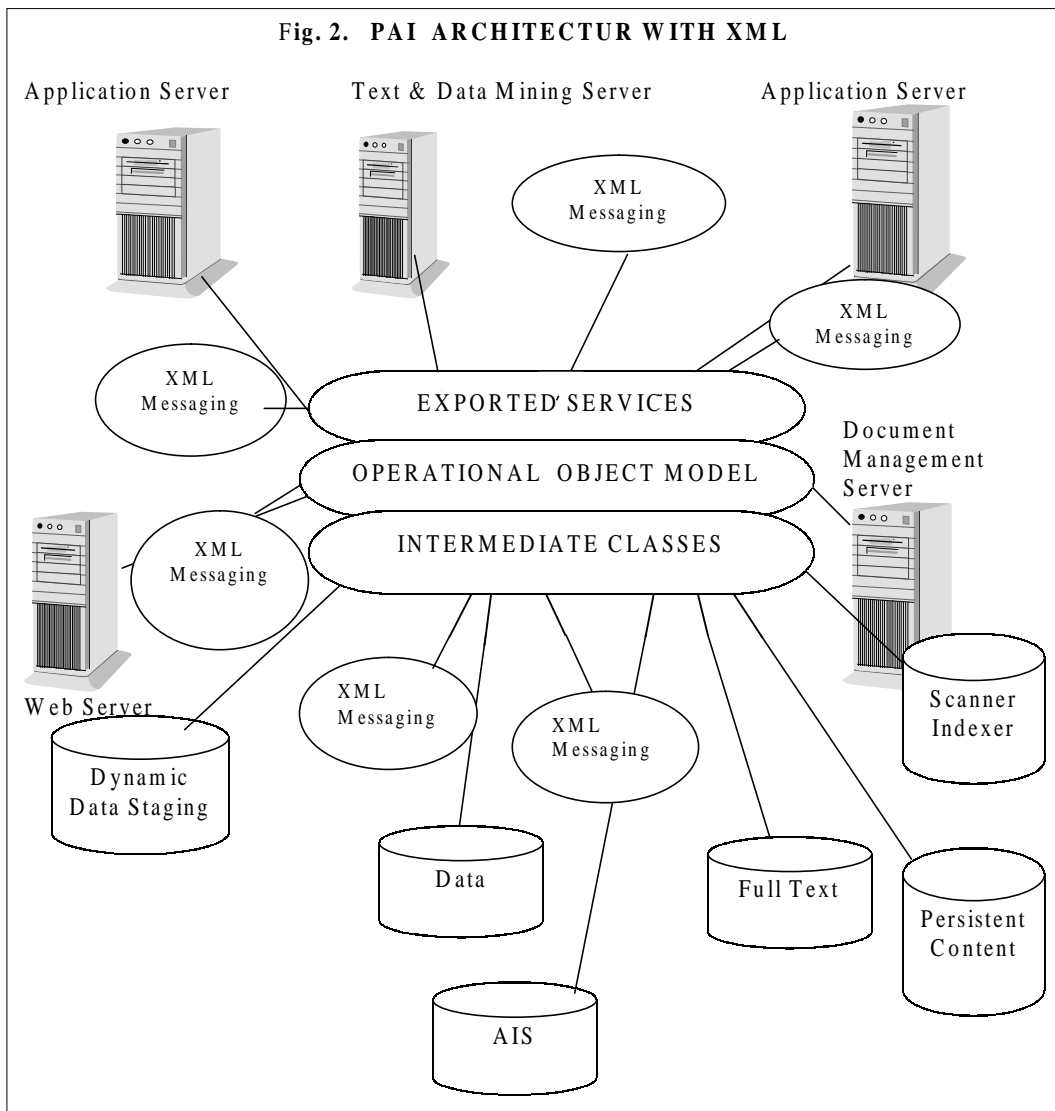
As Fig. 1 illustrates, portals provides opportunities to manage organizational metadata programmatically in an XML based environment, mapping a data object known to the portal is roughly equivalent to publishing it from an object-oriented perspective. Once data objects are known to the portal, they are accessible within security partitions.



The portal is able to “know” for more about its data, structures, and quality than just XML wrapped data. The portal can be used to represent complex data structures that have been virtually integrated among legacy systems and infer meanings from meta data component. The information that becomes associated with the portal information is now a brand. This “brand” represents whatever the organization chooses to use as an imprint on the data produced by the portal.

2.1 XML and PAI for Information

Fig. 2 Shows PAI (Portal Application Integration) architecture with XML messaging. In this architecture, XML provides universal connectivity : from the front end of clients and client based agents, to web servers, page builders and IAS. XML provides a universal exchange format for all communications in the system.



synchronization and integration by removing much, but not all of the work of interface modeling from the mix.

2.2 XML for Messaging and Connectivity in Portal Systems.

Following are the categories of XML connectivity in portal systems Viz.,

- Front end to the Web server and back: Library users makes request to the portal system using browser that supports XML. The XML request, using HTTP for communication, goes to the portal Web server and then passed on to the AIM . When the XML replaying to a request is delivered to the web server by the AIM, the web server sends back a raw XML stream to the client's browser.
- Web server to the AIM and back: Once the processing by web server, requests are passed to the AIM in XML, using either HTTP or TCP/IP. Using an XML connector / proxy facility the AIM maps the incoming XML data stream to its own object model. When the AIM distributes the request, it has received to other application servers or to database servers and results are returned to it, if these are in XML, it passes them back to the web server. IF they are not in XML, the AIM with its object model, it's mapping to XML intermediate classes, and its export services is used to cover them.
- AIM to the middle tier and back: The middle tier is composed of a variety of application servers, both stateful and stateless. In XML based information portals with PAI, architecture such application servers 'Speak' XML. The AIM maps their object models to its own, sends XML messages to, receives them from the application servers, and passes them on either to the portal job server, other application servers or to the portal web server.
- AIM to data and content stores and back: The IAM issues XML queries or content oriented searches of data and contact stores in response to user queries and searches coming from the portal web server. The content oriented searches are routed through the portal job server and rely on its content management capability and XML transformation capability. In the XML based information portals with PAI architecture, all data stores produce XML data in response to queries, either because the data stores are " Warped " so that legacy data is converted to XML "on the fly" or because the data stores being queried are already in XML.

2.3 XML in Clients

The XML is transmitted to clients from the portal web server, it is formatted using XSL to provide HTML for the browser. In addition, since the XML data is available to client-side applications through the XML document object model (DOM) and client based XML parsers, it may also processed by client side applications of any complexity.

This opens the way for knowledge to use the processing power of "clients" for a variety of analytical purposes involving processing of XML data. By distributing, analytical takes across the information centers, some of the processing load is taken from the application servers, improving the load balancing in the information system. With client processing based on XML, all LAN/WAN resident client data may be exchanged among different client applications. Workflow and information process automation applications can integrate client side applications, while subject matter, integration in the portal interface can be provided by a cognitive style interface.

2.4 XML in Databases and Content Management

PAI XML based architecture may also use XML in databases and content stores. Major relational database vendors already made provision for parsing and mapping XML data to database tables and storing in

there. XML may be stored in file systems, flat files, relational databases, object-oriented database system and native XML format. Performance in PAI architecture is improved if XML is stored either in object or in native XML form. Flat files and file systems of various kinds cannot scale to the performance in PAI architecture is improved if XML is stored either in object, or in native XML form. Flat files and file systems of various kinds cannot scale to the performance required in information systems.

The relational solution has the difficulty of requiring a mapping of the hierarchical structure of XML data to a non-hierarchical format of flat relational tables. That format is not particularly responsive to change in the system of XML tags that occur frequently in XML systems when new types of documents are added to the system.

3. Information Portal Solution (IPS)

The Internet solution can be implemented to deliver the information service that library and information centers wants to have available for their users. It has to establish an appropriate information sharing strategy. For example, portals are to provide information, techniques, tools and learning opportunities for users to successfully perform objectives. Library users, however, are having trouble finding the information they are looking for and it has gotten so difficult that at times users do not know what information and other tools even exist for them to use. For the solution information portals, wants to implement as a central location that logically catalog and organize the information.

A portal is an internet “window” that presents information to users and an internet “ door” that allows users to pass through to reach selected destinations. The information portals creates the central location where navigation services are available for users, to find information, launch applications, internet with data, identify collaborations, share knowledge and make decisions. Therefore information portal is “a browser- based application that allows information access to, collaborate with, make decisions and take action on a wile variety of information regardless of the location of the information or the format in which the information is stored”.

Information portals can provide a successful Internet strategy for the organization. IT has several features that are used to implement information management initiatives in organization viz.,

- A constant view of your library and information centers.
- Information organization and search capabilities
- Direct link and access different racecourses
- Direct link to relative data and experts
- Individual identity and personalized.

3.1 Benefits of IPS

LICs can gain several advantages by implementing information portal solution such as;

- Better decision making capabilities
- Improved semantics
- Improved information organization and search capabilities
- Direct access to knowledge and resources
- Direct links to reports, analysis and quarries
- Direct link to related or relative data points
- Personalized access to content.

These benefits are major elements that are built into the portal desktop used by users. This software function work together in different combinations to achieve the portal benefits identified as important by organizations, users interact with portal desktop, which allows organization to address its objects.

3.2 Information portals search functionality

Library and information centers will want to include several common search features in the search software function of its information portal solution. The standard search features that are available from Internet sites viz.,

- Exact phrase search: The query returns results that contain at least one occurrence of the string of words, entered in the search box.
- Key word search: The query returns results that contain at least one occurrence of at least one of the words entered in the search box.,
- Category search: The query returns results that are associated with type of information the user is interested in.

The search engines provide services to execute predefined queries that return the data points or information from multiple or identified sources that match the criteria in the query. These matches are stored in a index or catalog by the search application. The index or catalog is used to provide reports or information in content phase windows presented to the user. The data points and information returned to the information portal from the search catalog can be presented in several formats. Some of these formats are;

- A link to an internet website
- A listing of available sources within organization
- A collection of hyperlinks that might provide relevant information related to the content page.

There is a broad set of requirements that LICs must implement to provide the searching features needed to quickly find information within LICs and on the Internet. When users create and submit their own search requests the features fall into three basic categories: standard search features, concept based search and metadata search. There fore portal software vendors support search features and several third party search engines.

4. Future of Portal

The future is bright for the information portals or to be more precise, for integrated computing platforms, which provide portal access to universe of knowledge . Such platforms are becoming the standard in organizational computing and they are growing and will continue to grow increasingly sophisticated technologically. Soon, everyone will have access to portal workplaces to do both their internally and externally facing work, and no one will think of providing access to computing resources through a means other than a portal interface (whether desktop, mobile or wireless). Thus, the information portal revolution will be complete in the sense that everyone will use a portal for access their required information.

5. Conclusion

The PAI architecture is to use the distributed AIM to integrate all data, content and applications, in online information systems. The appearance of XML makes the task of developing connectors and proxies and

communicating among application servers to support integration much easier as the XML standard spreads. The XML is not the answer to all communication. The combination of RDF (Resource Description Framework), XTM (XML Topic Maps) and MDI (Meaning Definition Language) promises to provide on increased capability to support the continuous mapping of XML data streams to object models that is at the heart of the AIM integration of the portal system in PAI architecture through its network of servers and intelligent agents.

6. References

1. Ceponkus, A., and Hoodbhoy, F. "Applied XML", New York, John Wiley, 1999.
2. Collins, H "Corporate portals: revolutionizing information access to increase productivity and drive the bottom line. New York, AMACOM. 2001
3. Finkelstein, C and Aiken, P. "Building corporate portals with XML" New York, McGraw Hill, 1999.
4. Firestone, J.M. "Approaching enterprise information portals: Executive information system industry report". Wilmington, EIS, 1999.
5. Firestone, J.M. "Enterprise information portals and knowledge management". Amsterdam, Butterworth Heinemann, 2003.
6. McGrath, S. "XML by example building e-commerce applications", New Jersey, Prentice Hall, 2000.
7. Tannenbaum, A "Metadata solutions: Using Meta models, repositories, XML, and enterprise portals to generate information on demand." Boston, Addison Wesley, 2002. Pp 227-228.
8. <http://www.w3.org>
9. <http://www.ipedo.com>
10. <http://www.exceloncorp.com>

About Authors



Dr. Sangayya S Sirurmath is Librarian & Head – NICMAN at Indian Institute of Management, Ahmedabad-380 015, India. He holds M.L.I.Sc, M.A, M.Com, PGDLAN and Ph.D degrees with him. He was with XLRI, Jamshedpur prior to joining IIMA. He has coordinated management development programmes, organized and attended national conferences, published a book and more than 18 papers. He is a member of ILA., IASLIC., MANLIBNET.

E-mail : sssmath@iimahd.ernet.in



Dr. Ashis Kumar Pani is Professor in MIS at XLRI – School of Management, JAMSHEDPUER 831 001, India. He holds M.Tech. (IITM) and Ph.D. (IIT Kharagpur). He has published and presented several papers. Current areas of interest are e-Business, e-SCM, e-CRM, IT Security, IT Outsourcing, web mining and application of AI techniques. Presently member of IARCS) and CSI.

E-mail : akpani@xlri.ac.in