

Unified Information Management Using Xpiori XMS

by Chris Brandin

Release 1.1

Xpiori, LLC
2864 S. Circle Dr.
Ste. 401
Colorado Springs, CO 80906
(719) 425-9840
www.xpiori.com

© 2007 by Xpiori, LLC. All rights reserved.

Version 1.1

Copyright © Xpiori, LLC All Rights Reserved

Xpiori technology is protected by the following patents:

US Patent #5,742,611 (21 Apr 98)

US Patent #5,942,002 (8 Aug 99)

US Patent #6,157,617 (5 Dec 00)

US Patent #6,167,400 (26 Dec 00)

US Patent #6,324,636 (27 Nov 01)

US Patent #6,493,813 (10 Dec 02)

US Patent #6,792,428 (14 Sept 04)

Other U.S. and international patents pending.

The information in this white paper has been provided by Xpiori, LLC. To the best knowledge of Xpiori, it contains information concerning the current state of information processing technology. Xpiori, LLC disclaims any and all liabilities for and makes no warranties, expressed or implied, with respect to products described in this paper, including, without limitation, the implied warranties of merchantability and fitness for a particular purpose. No specific reliance should be made on the material provided herein without thorough investigation of the technology and its proposed application to specific circumstances. Product and technology information is subject to change without notice.

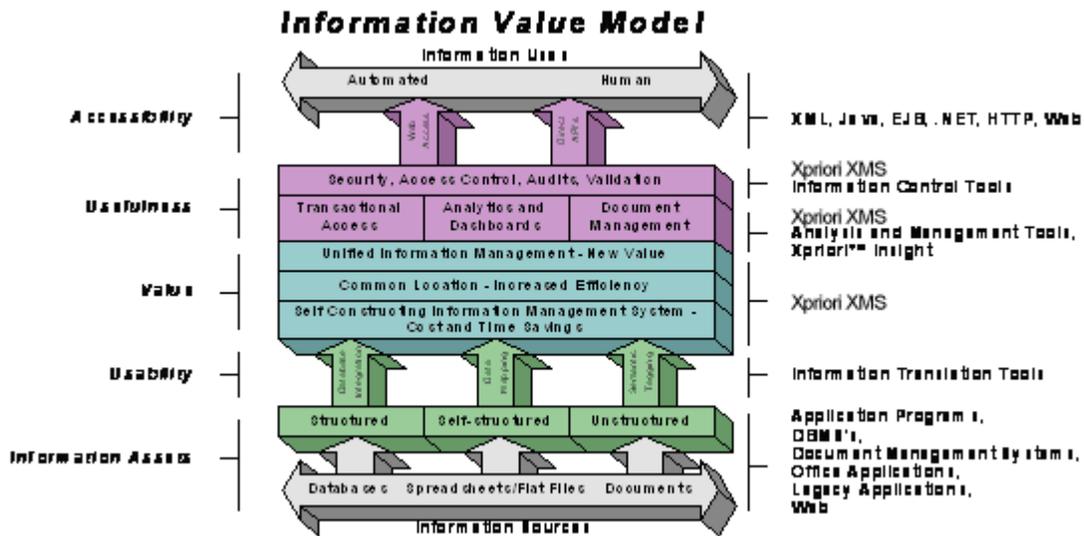
Introduction

Although today's enterprise possesses vast amounts of information, much of the value of that information remains unrealized. Xpiori's products are targeted at enabling organizations to reap the full benefit of their information assets quickly and efficiently in a unified environment.

There are several reasons why information assets remain largely untapped today:

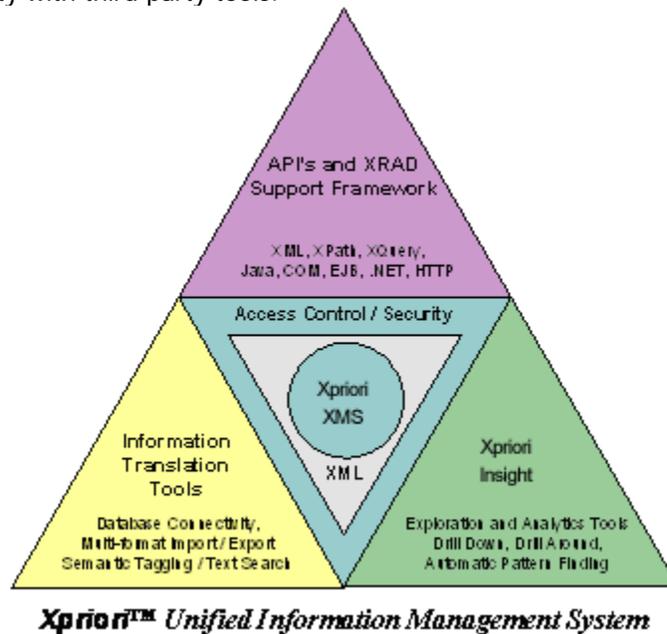
- Different forms of information require different programs to manage them. Approximately 80% of information assets in a typical enterprise are not managed by a database at all. This includes items like spreadsheets, flat files, semi-structured information, and unstructured forms of information such as email and word processor documents. There has been no unified way to explore or discover such information so it has remained largely invisible and unusable at the enterprise level.
- Having access to different types of information about a common subject available at the same time from a common source is just now becoming possible.
- Even information managed by similar database management systems has proven to be burdensome to aggregate when a common view of disparate data sources is needed.
- There are systems that can "contain" information from disparate sources – but they store information in a generally unusable form. The problem remains that most of the information has to be exported and "unwrapped" before it can be used.
- When information has to be used in a new way, often it must be re-cast for the new use case. The process of re-working data for new, or modified applications is expensive and time-consuming.
- New information sources come into existence faster than effective information management systems can be built to accommodate them.

Information systems have traditionally done a fine job of managing specific information for specific purposes. Xpiori has crafted its Unified Information Management System according to an information value model that accounts for new, broader capabilities: The ability to assimilate information from disparate sources and in different formats. The ability to make information available in different formats for a variety of uses. The ability to make previously disconnected information assets viewable, usable and manageable in a unified environment. The ability to easily re-task information for new uses. The ability to simultaneously accommodate data and document centric application environments. The ability to accommodate new information management requirements quickly and easily.



Unified Information Management System Overview

The Xpiori Unified Information Management System consists of four basic categories of components. At the core of Xpiori's offering is Xpiori XMS – a self-constructing and auto-indexing XML information management system capable of accommodating any kind of information expressed in XML in a controlled and secure environment. Java, COM and .NET development environment integration tools are provided to serve eXtreme Rapid Application Development (XRAD) requirements. The Information translation tools are used to convert all formats of information into XML, and to deliver information in various forms. Xpiori Insight provides a continuous view and analytical capability for all information managed by XMS. All components of the Xpiori Unified Information Management System are open standards based, insuring compatibility with third party tools.



Combining these components in an open standards based platform, Xpiori offers a means to leverage information assets in new and more productive ways without disrupting existing systems:

- Using a technology neutral format (XML), Xpiori XMS offers common, unified way to manage data, documents, and everything in-between. Information translation tools are available for converting virtually any data format into XML, and visa-versa. XMS can also be accessed directly through standard API's, operating as a fully transactional database.
- Information managed by XMS is in fully query-able and manageable form, rather than simply being "contained". Xpiori™ Insight comprises a set of tools for exploring and analyzing information in a unified way while it is in XMS without the need to create separate data repositories like OLAP cubes. XMS always indexes everything expressed in XML, so information can be used in new ways and by new application programs without having to be reorganized.
- By eliminating database design and indexing instructions, XMS reduces the time it takes to develop and deploy new applications by up to 70%. Xpiori XMS, like XML, is fully extensible - allowing for on-the-fly changes or additions to applications and their underlying information models. Java, COM and .NET environments are supported seamlessly, and can be used together.
- XMS has an unusually flexible access control mechanism. Access to data and functions can be controlled based not just on user profiles, but on data field contents as well.

Xpiori XMS – following XML's lead

Much like a Web browser can determine how to display HTML information based on presentation metadata embedded in the HTML, Xpiori XMS determines how to manage and index XML information based on metadata embedded in the XML. XML differs from HTML in that where HTML was targeted to a single function – the presentation of information, XML fulfills a more universal purpose – the complete characterization of information. XML consists primarily of tags, attributes and data elements. Tags provide context; describing what data elements in their scope are.

Attributes provide information about, or indicate how to interpret data elements in their scope. Data elements represent data in the traditional sense. The structure of XML also provides information – about hierarchy, groupings, relationships, etc. To understand how Xpiori XMS identifies and uses these natural patterns consider the following XML fragment:

```
<Telephone_Book_Listing>
  <Name>
    <Last> Smith </Last>
    <First> Tracy </First>
  </Name>
  <Telephone>
    <Area_Code> 719 </Area_Code>
    <Number> 555-1234 </Number>
  </Telephone>
</Telephone_Book_Listing>
```

XML Fragment

The explicit patterns in the example above are used to determine what to index and make available as information access terms:

- This is a telephone phone book listing.
- The listing's last name is "Smith".
- The listing's first name is "Tracy".
- The telephone number's area code is "719".
- The telephone number is "555-1234".

The implicit patterns in the example above are kept intact and are also used to automatically determine where information access terms converge based on groupings and hierarchy:

- "Smith" and "Tracy" belong to the instance group "Name" and are the last and first name respectively.
- "719" and "555-1234" belong to the instance group "Telephone" and are the area code and number respectively.
- Everything belongs to the instance group "Telephone_Book_Listing".

Xpiori XMS organizes itself automatically; there is no separate database design process and there are no indexing instructions required. The behavior of XMS is determined entirely by the structure of the XML information managed by it. Although XML schemas can be enforced to control information, they are not needed by XMS. This is an important feature because while it is important to enforce schemas in traditional database-like applications, it is not practical to create schemas for every instance of unstructured information. Schema validation need not be enforced at the core XMS level; rather it can be applied when necessary, at the user's discretion.

Information Translation Tools

Xpiori XMS uses XML as a technology neutral format expressing all information managed by it. The Information translation tools play a critical role in Xpiori™'s Unified Information Management System because they are the means by which information assets are transformed into and back out of XML to be managed by XMS. Information assets basically take one of three forms:

- Centrally structured information. This consists primarily of information managed by database management systems. Xpiori™ offers a robust set of tools for connecting to traditional databases. According to industry estimates, approximately 20% of a typical enterprise's information is managed this way; the rest is either unmanaged or not managed in any centralized way.
- Unstructured information. This consists of word processor documents, emails, text files and the like. Xpiori offers tools for creating metadata containers for unstructured information of any type – even binary files. Also, tools are available to embed semantic tags into text files of any type, making document contents available as query terms based on ontological rules.
- Self-structured information. This includes office documents (spreadsheets, etc.), ad hoc database files, flat files, Web pages, reports – anything that follows some sort of predefined format. Xpiori™ offers an ever-growing set of connectors for dealing with standard formats as well as tools for defining and creating custom connectors and "information scrapers". These connectors can be used independently for content gathering and delivery; for example, information gathered from a database could be delivered as a spreadsheet or a flat file.

Supported Formats (partial list)

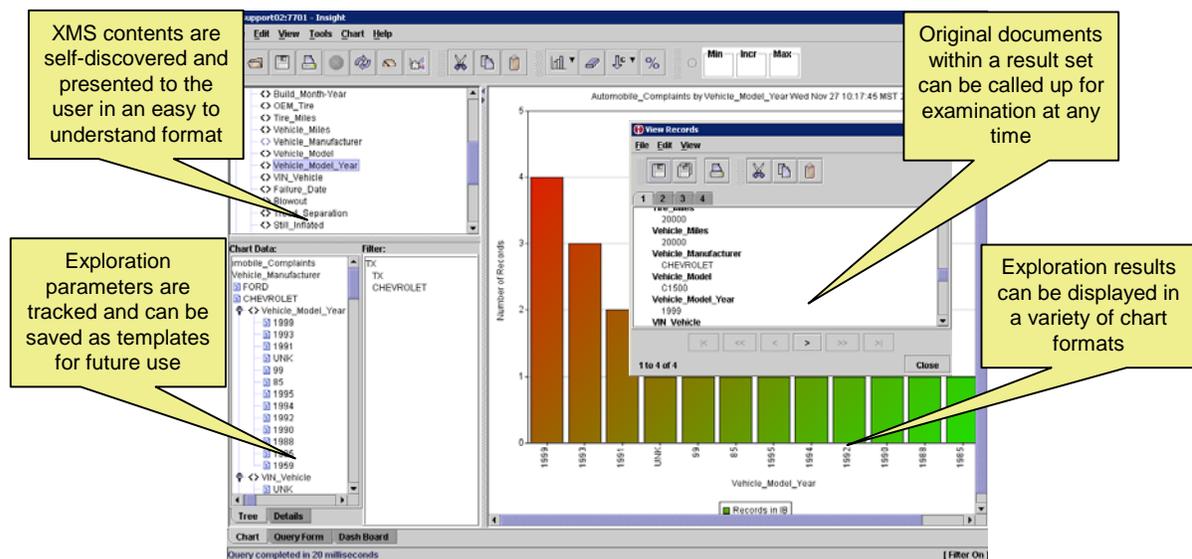
Access 2.x / 95	DIALOG Market Research	LDAP	Report Reader (source only)
Access 97	DIALOG Patents	Lotus 1-2-3 r1A	Salesforce.com
Access 2000	DIALOG Sci / Tech & BioMed	Lotus 1-2-3 r2	SAP (IDoc)
Access XP	DIALOG Trade Names & Products	Lotus 1-2-3 r3	SAS
AccountMate	DIALOG Trademarks	Lotus 1-2-3 r4	SBT Accounting (FoxPro)
ACT! for Windows	DIF	Lotus Notes	Scalable SQL
Acucobol - 85	Dodge Bidders	Lotus Notes 5	Sequential Binary
ADABAS (NatQuery)	Dodge Firms	Lotus Notes Structured Text	Solomon Accounting (Btrieve)
Alpha Four	Dodge Firms & Bidders	Macola Accounting (Btrieve)	SPLUS
Apache Common Log File Format	Dodge Projects	Magic PC	SPSS
ASCII (Delimited)	EDI (EDIFACT)	Mailer's+4 (dBASE)	SQL Server
ASCII (Fixed)	EDI (X12)	MicroFocus COBOL	SQL Server 2000
Content Extractor / Cambio	Enable	Microsoft COBOL	SQL Server 7 (OLE DB)
Binary	Excel XP	Microsoft IIS Extended Log File Format	SQL Server BCP
Binary Line Sequential	Excel 95	MQ Series (Extractor)	SQL Base
BizTalk XML	Excel 97	MQ Series (Fixed ASCII)	SQL Base 6.x
Btrieve v5	Excel v2	MQ Series (XML)	Statistica
Btrieve v6	Excel v3	MSMQ (Extractor)	Sybase Adaptive Server 11.x
Btrieve v7	Excel v4	MSMQ (Fixed ASCII)	Sybase Adaptive Server 12.x
C-ISAM	Excel v5	MSMQ (XML)	Sybase BCP
C-tree 4.3	eXcelon	MySQL	Sybase SQL Anywhere
C-tree Plus	Extractor	Navision Financials (ODBC 3.x)	Sybase SQL Anywhere 6
Cambio	Foxbase+	Null Connection	Sybase SQL Server System 11
Clarion	FoxPro	ODBC 2.x	SYSTAT
Clipper	Fujitsu Cobol	ODBC 3.x	Tape Drive Sequential
Cloudscape (ODBC 3.x)	GoldMine	ODBC 2.x	Text (Delimited – EDI)
COBOL Flat File	GoldMine Import File (dbf)	ODBC 3.x	Text (Delimited – EDIFACT)
Common Log File Format Webserver	Great Plains DOS (Btrieve)	Oracle 7.x	Text (Delimited – HL7)
Content Extractor	Great Plains UNIX / Mac (C-tree)	Oracle 8.x	Unicode (Delimited)
Data Junction Log File	HIPAA	Oracle 9i	Unicode (Fixed)
DataEase	Hitachi HiRDB (ODBC)	Oracle Direct Path	USMARC
DataFlex	IBM DB2 Loader	Oracle SQL Loader	Variable Sequential (MVS)

Supported Formats Continued (partial list)

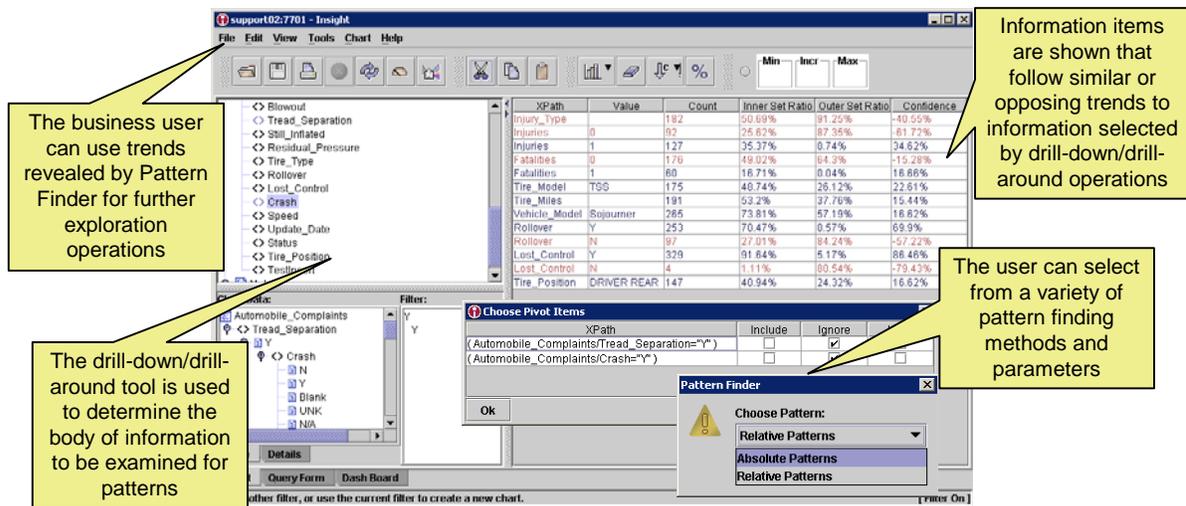
DBASE II	IBM DB2 Universal Database	Paradox v5 (IDAPI)	Variable Sequential (Record – V UniKix)
DBASE III	IDAPI	Pervasive.SQL	Variable Sequential (SyncSort)
DBASE III+	Informix (ODBC 3.x)	Platinum Accounting (Btrieve)	Variable Sequential (User Defined)
DBASE IV	Informix DB Loader	PostgreSQL (ODBC 3.x)	Velocis (ODBC 3.x)
DBASE V (IDAPI)	Informix - Online DS	Progress (ODBC 3.x)	Visual dBASE 5.5
DIALOG Biographical	Informix / SE	Quattro Pro Windows v5	Visual FoxPro
DIALOG Business & News	Ingres (ODBC 3.x)	Rbase (ODBC 3.x)	Watcom SQL v5
DIALOG Commerce Business Daily	JMS (Extractor)	RealWorld Accounting (MF COBOL)	WordPerfect 6.0 (Mail Merge)
DIALOG Company Directory	JMS (Fixed Ascii)	Red Brick	XDB (ODBC 3.x)
DIALOG Federal Register	JMS (XML)	Remedy ARS	XML

Xpiori™ – unified information visibility

It is important to not only contain information assets, but to have a continuous view into them and be able to use them together without having to export blocks of information first. An archiving system, for example, is much more useful if one does not have to know where information is contained before it can be seen. Fine-grained content based query and analytical tools significantly raise the usability and value of an information management system. Xpiori™ Insight comprises a rich set of exploration and lightweight analytical tools which offer the business user means to see and explore all XML content, all at once, and in a unified manner. This includes drill-down, drill around, executive dashboard, and automatic pattern finder capabilities in an interactive environment. *Xpiori™ exploration tools can self-discover content that is expressed in XML and automatically present query terms for the user.* These tools integrate with the access control system, so information visibility can be controlled.



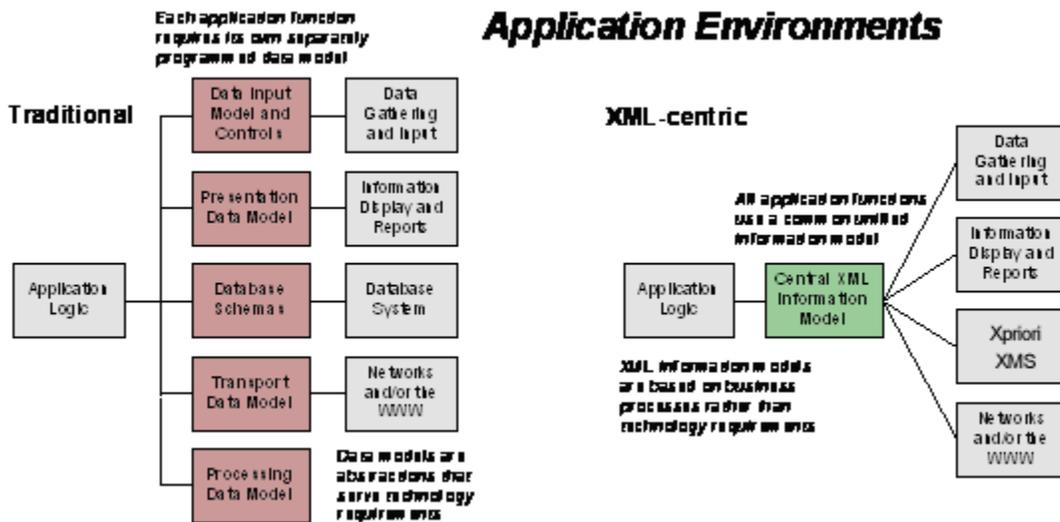
Drill-down, Drill-around Tool



Pattern Finder Tool

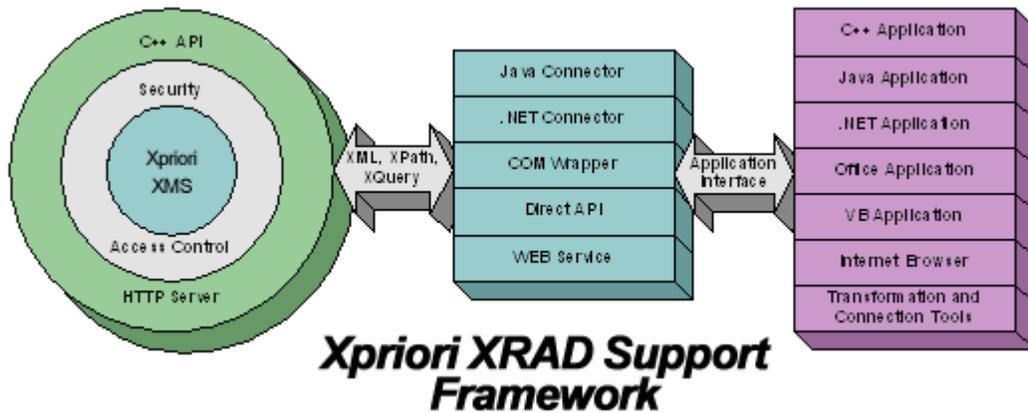
Xpiori™ XRAD Support Framework

In this mode of use Xpiori XMS operates as a transactional XML database, offering fine-grained access to all information managed by XMS. Most application programs encompass the same basic functions - input, presentation, communication, processing, and information management. Although the same information underlies each of these functions, different data models have traditionally been employed to accommodate application components accomplishing these tasks. Remodeling the same information in different ways for each component of an application is inefficient and yields programs that are unreliable and difficult to maintain and change. The Xpiori XRAD support framework adheres to an XML-centric model for defining applications, offering an open standards based, fully transactional information management framework.



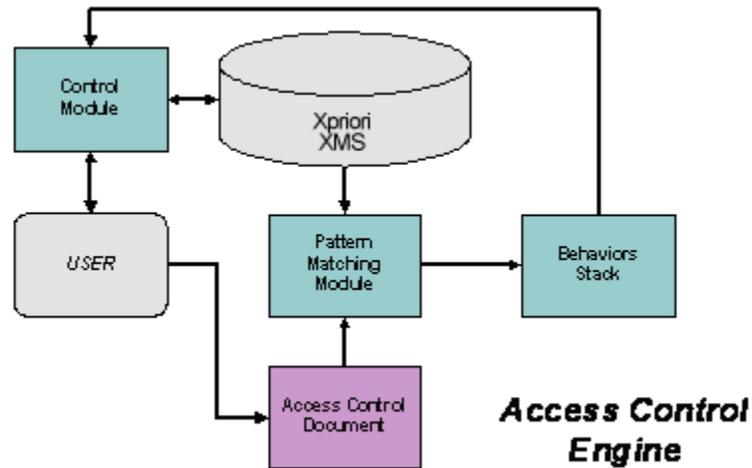
Because Xpiori XMS is entirely self-constructing and auto-indexing, database design and indexing instructions are eliminated; resulting in up to a 70% reduction in the time it takes to develop application programs and even more time savings when they need to be changed. Application programs can be built incrementally, and expanded as requirements change – without the need for offline database reorganization. In many cases the database creation process can be entirely automatic. For example, using the .NET development platform, a developer simply creates the application front-end. The .NET connector tools are used to create and access XML information

sources transparently, or through simple property assignments. Connector tools also provide a number of convenience functions, like “query by example” XPath statement generation.

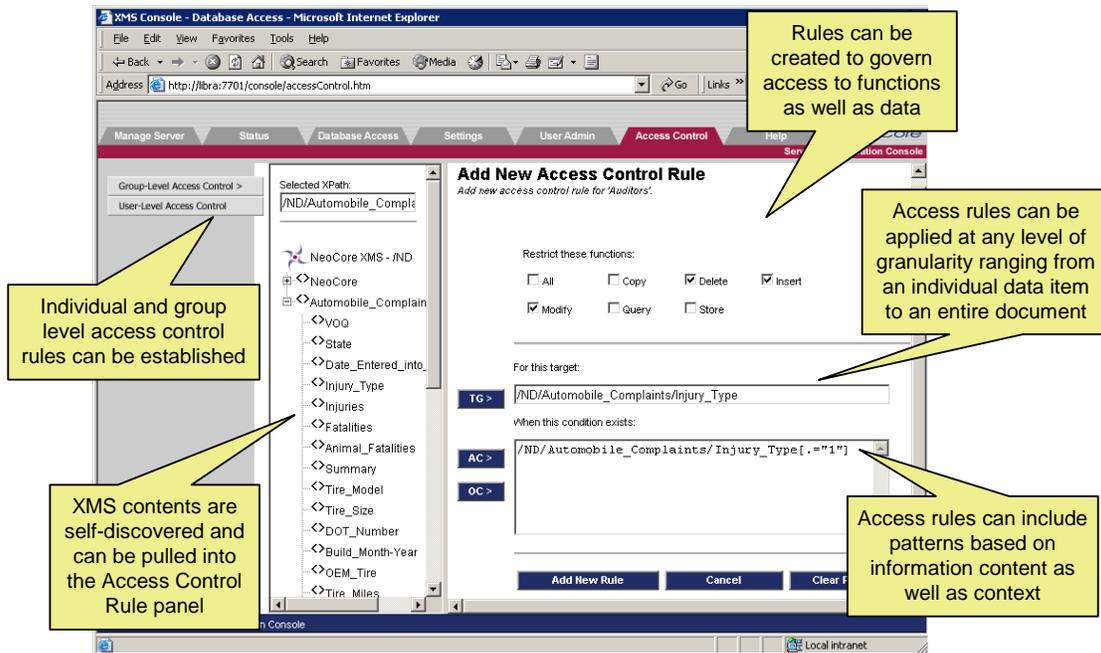


Access Control

Xpiori XMS uses a novel mechanism to implement access control. Pattern matching is used to provide a greater degree of control than is normally available. Rules governing access to information and XMS functions can be created that take user context *and* XML information contents into account. This makes it possible to not only control access to items of information within XML documents, but to also create virtual collections of information ranging from the entire contents of XMS - to a single XML document type - to a collection of items fulfilling a common arbitrary pattern.



Each user (whether a person or a process) can be associated with individual and group access control documents. When XML information is accessed, the pattern-matching module examines the XML and identifies patterns as described in the control document. If a match occurs – whether it is context and/or content based - the appropriate access control behavior is invoked.



Access Control Manager

Epilog

There are many XML products on the market. The Xpiori™ Unified Information Management System was not designed so much to address XML information management challenges as it was to address a new breed of business requirements using XML as part of the solution. The components of the Xpiori™ Unified Information Management System operate together to allow the enterprise to realize the full value of their information assets in new and better ways. Although each component plays an important role by itself, it is the interaction between them that is responsible for the unique value the platform brings. While architecting the platform and the products it encompasses, Xpiori™'s engineers evaluated the challenge from the ground up coming to a series of technical assumptions, sometimes departing significantly from traditional thought:

- XML is the only suitable standard for a technology neutral format. All information would have to be transformed to (or contained in) XML.
- The core of the platform – the XML information management system (Xpiori XMS) – would have to be completely self-organizing, requiring no database design or indexing instructions. An XML database requiring separate database design or indexing instructions offers little advantage over traditional database solutions and would not suffice. Aside from the obvious economic value of eliminating database design, there was another reason for this requirement – one that was not optional. Unstructured information is by definition heterogeneous, and having to produce schemas for each item of such information is simply not practical. Furthermore, if the content of unstructured information is not explorable and usable while it is in Xpiori XMS, the value of the information is diminished. In order to maintain performance, semantic tagging techniques would have to be employed, and Xpiori XMS would have to automatically support the embedded tags for direct queries.

- Data and metadata would have to both be managed dynamically. This departs significantly from traditional thought, where metadata is typically predefined and represented statically, such as with tables and columns.
- The original structure of XML information would have to be preserved. Often, XML information is decomposed into tables and columns. When this is done, all the information implicit to the structure of XML information is either lost or must be stated explicitly – this includes information about relationships and groupings. In order to preserve the original structure of XML information, XMS would have to avoid relying on a structure of its own, using pattern recognition technology instead of the traditional structural approaches to managing information.
- Data, documents, and everything in-between would have to be managed in a unified way, in a common repository, and simultaneously accessible in transactional data-centric and document-centric ways.
- Intrinsic characteristics of XML – heterogeneity, extensibility, and flexibility – would have to be preserved during the process of storing and managing information.
- As a unified information management system aggregating multiple information sources, Xpiori XMS would have to be scalable, maintaining performance irrespective of size.
- Xpiori XMS would have to be schema independent at the core in order to be able to manage all types of information – structured and unstructured. Schema validation would be applied when required either as an application function or as a server-side extension.

It was decided that only when *all* of these requirements were met, could a unified information management system be built. Using a technology invented at Xpiori™ – Digital Pattern Processing – a new type of information management system was developed fulfilling all the requirements listed above. Xpiori XMS was combined with information translation, analytical and exploration, and XRAD development tools to offer a complete unified information management solution. Using the Xpiori Unified Management System, users can now realize increased value from their information resources, going beyond what they could do before.

#