

XML Database Technology and Tax Administration

An Oracle White Paper
November 2008

XML Database Technology and Tax Administration

As the use of XML in Tax Administration continues to expand, tax agencies need to form a strategy to take advantage of these information assets.

EXECUTIVE OVERVIEW

The use of extensible markup language (XML) based data as a medium of integration in tax administration continues to grow and expand in importance. This whitepaper explores the relationship between XML and XML database technology in key tax administration business processes.

The paper makes a set of recommendations for tax administration organizations that are interested in gaining maximum advantage of this technology trend for their business process improvement initiatives.

INTRODUCTION

Extensible markup language (XML) is a de facto international standard for defining and exchanging data. The use of XML as a medium for data exchange and storage in the government tax administration domain continues to grow and expand.

XML usage in tax administration business process ranges from tax return filing, to inter-governmental data exchanges, support for internal systems integration, and the use of XML and other types of data for analytical, reporting and publishing purposes.

Some examples of current trends in the use of XML based data in tax administration include:

- In the United States, the Internal Revenue Service continues to expand the use of XML as the medium of exchange for tax return filing. Between 2008 and 2012, XML-based filing of tax return information will quickly become the medium for the majority of electronically filed returns in the United States.¹
- At the level of the individual States in the United States, a number of XML-based standards are being implemented to support interstate filing of Sales and Use, Fuel Tax, and other taxes that have interstate data sharing implications. These initiatives lead to lower taxpayer filing and payment

¹ Reference: <http://www.irs.gov/efile/content/0,id=171945,00.html>

burdens and improve the transparency of tax flows across jurisdictional boundaries.²

- In the Netherlands, the Ministry of Finance and Ministry of Justice have encouraged the use of extensible business reporting language (XBRL – a language based on XML) as a means for taxpayers to file their tax returns and supporting documentation. This initiative has resulted in a major positive reduction in taxpayer filing burden, estimated at approximately 1 billion Euros in savings to taxpayers.³
- In Belgium, the national social security organization uses XML as the medium of storage for company filings and citizen benefits. Social contribution systems such as social security have close affinities with tax administration processes. The Belgian system stores and processes hundreds of millions of XML based documents.⁴
- In Australia, the government has established a standard business reporting and tax filing portal site, based on XBRL. This service is estimated to save business around 675 million Australian Dollars in reduced filing burdens.⁵

Clearly, the advantages of using XML as a medium of exchange are becoming firmly established in the tax administration domain. The advantages include:

- XML is an international standard – makes it easier to build integrations between trading partners at different levels of government and even internationally. These integrations are based on a shared set of data models which themselves are defined using XML technology.
- XML is flexible and adaptable – the underlying data model of the information being exchanged is easy to modify – this allows tax administration organizations the flexibility they require to rapidly adapt to changing legislative requirements
- XML is the foundation for web services, allowing administrations the ability to incorporate transactional interfaces – not just data exchange channels – into their business process improvement initiatives.

THE RELATIONSHIP BETWEEN XML AND XML DATABASES

The increasing volume of XML based transactions in tax administration raises a number of benefits and challenges.

² Reference: <http://www.taxadmin.org/fta/edi/standard.html>

³ Reference: <http://www.xbrl-nederland.nl/>

⁴ Reference: <http://www.oracle.com/technology/oramag/oracle/07-jan/o17xml.html>

⁵ Reference: XBRL Australia, Business Plan 2007 - 2010

The potential benefits include reduced taxpayer burden, reduced processing and paper handling costs, increased volume and quality of supplied tax information, and increased potential for tax analytics on the basis of received XML data.

One of the key challenges of the increasing volumes of XML is how to store and use the data. XML data can be stored in an unstructured, structured or semi-structured storage model. The choice of which model to use will depend on the range of requirements for each XML data object being used.

The emerging best practice is to store these XML documents in a database that has special features to store, publish, manipulate and query the XML-based data. In such an XML enabled database, performing analytical queries such as a ratio analysis on data from tax return objects would be as easy as performing other SQL-based query exercises.

In general, the evolutionary inclusion of XML as a first class data type into the traditional SQL relational technologies has two major technical implications:

- First, it allows organizations to incorporate XML-based data directly into their existing relational database management storage systems. This allows organizations to benefit from the same infrastructure for data management, data security and application development that they have already been using for many years.
- Second, incorporation of XML-based data into these traditional database management systems means that the standards around storage and query manipulation will evolve to provide the level of service and support that is required in the industry.

However, standards around XML storage and query perfection are still evolving. In such a situation, multiple database vendors will compete to establish their technologies as the best model for standardization. For example, standards around the best “native XML” storage model for XML documents are still being developed.

This situation also makes it difficult for application developers to create generic XML-database improvements, since taking advantage of specific vendor techniques around XML could lead to solutions, which are locked, into a specific vendor’s approach.

In this situation it is often advisable to be clear about which features in the new technology are most important for the tax organization, and of those, which are closest to being defined and constrained through an industry standard. For example, tax organizations may benefit from the ability to perform hybrid SQL-XML queries on tax return information stored as XML in their databases. Following industry standards in the development of the query syntax would then allow the organization to use this feature without being locked into the specific database implementation.

THE RELATIONSHIP BETWEEN XML DATABASES AND TAX ADMINISTRATION

One way to understand the significance of XML and XML database technology in tax administration is to review some essential use cases for tax administration and describe the role of the XML database within each business oriented scenario.

Document and Content Publication

The business of tax administration is, in many respects, a supply chain management business. Tax authorities supply the tax administration guidelines and filing and payment rules to the taxpayer communities within their jurisdiction.

In Information Technology (IT) jargon, these guidelines, forms, rules, regulations and instructions are all instances of management *content*. This includes the tax returns that are published and used by taxpayers to report and assess their tax liabilities.

And this content is increasingly being composed and managed as XML documents. XML-based content management addresses the heart of tax administration. The technology needs to be flexible enough to handle the frequent changes that tax authorities are required to respond to in their legislative environments. Tax return forms and related content typically change on an annual basis to reflect the changing tax policies and rules of evolving legislation. XML-based definitions of this kind of content provide the right balance of flexibility and meta-data integrity (through the use of governed schema).

At the same time, the technology also has to be robust enough to enforce strict version control management for all the changes that are applied to compose the correct content over the lifecycle of each piece of content.

In summary, tax authorities require the ability to securely and efficiently store and manage very large volumes of XML-based content for publication purposes.

Tax Return and Informational Filing

The taxpayer community responds to the supply of rules, forms and regulations by supplying reporting information on their tax situation, and making payments to the government as appropriate.

Tax authorities typically receive and processes very large volumes of tax return, tax payment, and tax information data feeds.

And the clear trend in the industry is that these inbound data feeds are increasingly taking the form of XML documents--sometimes very, very large XML documents.

The advantages of using XML-based data feeds are clearly mapped to the content management topic described above. Each feed is associated with the meta-data definition of its form that is produced on an on-going basis in response to legislative and other change drivers. This provides the tax authorities the flexibility

and adaptability that they require to operate according to the legal and political constraints of their jurisdictions.

Tax agencies are increasingly moving to this form of tax reporting, filing and payment. Supporting electronic filing of XML documents can save significant amounts of money by avoiding paper handling and processing costs, while at the same time, ensuring higher quality data, and greatly lowering the filing and reporting burdens for taxpayers.

Receiving this data electronically means that most tax agencies are able to capture, retain and use more of the information supplied by taxpayers on their returns than had previously been the case under paper-based or other limited filing and reporting options.

But where are these returns and informational documents stored? How can they be effectively indexed and managed in databases so as to maximize their value to the wide variety of tax administration business processes that rely on them as a key input. This is where XML database technology becomes a key enabler.

Government to Government (G2G) Information Interchange

Tax authorities exist within very specific governmental ecosystems. Each system is unique. But each of them have certain features in common, and one of these features is that the tax authorities are typically required to supply and exchange information with other governmental bodies on a frequent basis.

On the supply side, tax agencies are typically required to provide reports on their collections and costs to the government authorities that use this information to manage budgetary and other financial functions in government.

On the receiving side, tax agencies are typically dependent on other government agencies to validate taxpayer-supplied information. For example, in some jurisdictions the current identity and demographic information related to a taxpayer is maintained by an agency outside of the tax authority and the tax authority is therefore dependent on an efficient communications channel to be able to query and use that information as required.

These government-to-government exchanges are increasingly taking the form of XML-based document exchanges, very similar to the return and informational data feeds received from the taxpayer population.

In some cases this integration is further enhanced through Web Services transactions that send the XML documents in real-time. The use of Web Services provides advantages in reducing business process cycle times while ensuring transactional fidelity.

Web Services and Integration

In the context of this whitepaper, Web Services are viewed in their role in consuming and supplying XML documents that are managed, stored and used within an XML-enabled database.

So, in this perspective, XML-based content and documents form the basis of the data components that are the inputs and outputs of key tax administration business processes. The XML data components are stored and managed as first class data types in an XML-enabled database.

Alternatively, XML-enabled databases can provide functions to query data held in relational data types, and transform these data elements into the XML-based documents that are then used in Web Services transactions. And some databases expose data oriented Web Services directly from the database tier.

In a nutshell, tax administration business processes benefit from the business process integration potential made available through Web Services. These Web Services are dependent on XML and the data elements that make up the Web Service requests and responses are expressed as XML.

XML-enabled databases make the link between Web Services and database content easier to navigate and use. A good example is the ability to query stored tax return filings and integration of the query results into various business processes, including case management and taxpayer service.

Data Analytics for Enforcement (Audit and Collections)

We have already seen how increasing volumes of taxpayer information are being supplied and managed as XML-based content and how this content is received and provided through real-time, Web Services-based functions.

In addition, a critical consideration for tax administration is how to get maximum value from available information resources, regardless of whether the resources are persisted as structured, relational database tables, or unstructured or semi-structured data such as XML documents of various types, including web pages and standard word processing documents.

The business requirement is exploiting multiple data sources to uncover taxpayer behaviors and patterns that identify tax revenue at risk. So it is important to maintain a strong capability level for analyzing data captured and stored in multiple formats, including XML.

XML databases must be effectively integrated as a source for enforcement related data analytics and data mining.

Tax form filings provide a good example – stored as structured or semi structured.

WHAT SHOULD TAX ADMINISTRATION ORGANIZATIONS DO NOW?

There is no one size fits all answer for any technology initiative in taxation administration and this is true, too, in the area of XML databases.

However, there are clear trends in the industry that point to an increasing volume and importance of XML as a medium in tax administration. With that trend in mind, the following bullet points are important to keep in mind:

- Adapt an Evolutionary Approach – There will be instances where XML storage makes sense – for example in storing XML (e.g., XBRL-based) taxpayer reporting and filing documents. Tax administrations should understand the range of uses of the stored data – including document retrieval and update, and data analytics. This approach will reveal opportunities for implementing different XML storage and retrieval models beyond the current practice of using database CLOB types.
- Avoid XML-only Solutions – Pure play XML-only databases probably do not make sense for organizations such as large tax agencies that already have a heavy infrastructure investment in relational database technologies. In light of that, it makes sense to build on top of a relational database infrastructure versus building a parallel infrastructure just to manage XML data.
- Training – Training initiatives in XML database technology should cover infrastructure and operations training, as well as developer and end user training for applications that take advantage of XML Database capabilities.
- Ensure core application processing systems are XML-enabled – The trend in the IT industry is clear: business applications are required to process and store XML based data. Sometimes, this data is used with standard office automation applications such as word processors and spreadsheets. In other applications, taxpayer return and entity data is exchanged and stored as XML documents. Therefore, it is important for the core application systems in tax administration to be XML-enabled and compatible with the data schemas in use by the tax department.
- Implement a governance model – Effective use of XML goes beyond just being able to store and retrieve XML documents in a native and indexed manner. As with other forms of data, the value of XML is dependent on its shared semantic meaning. Shared meaning and business process integration and improvements go hand in hand – and this is true whether the business process is internal to the tax agency, or exchanged between the tax agency and external actors such as taxpayers and other government agencies. Therefore, it is never too early to begin implementing a strong governance model over the schema and definitions of the data elements that make up this shared data. The governance of XML extends naturally into the processes associated with Service Oriented Architecture (SOA) and web services governance in the tax department.
- Collaborate – Governance and shared meaning is the key to effective collaboration and improvements in business process collaboration are often the key to achieving improvement in process performance. In tax

administration, the use of XML to support collaborative business process improvements is becoming a mature practice. As the trend continues to grow, the strategic use of XML database technology will also grow.

Industry Trends and Standards

The domain of XML Database technology is driven in large part by trends and capabilities built out of the XML standards community. Therefore, it is important for tax administration organizations to remain aware of trends and changes to some of these standards.

The entire standards community for XML is too complex to summarize here, but Figure 1 and Table 1 provide a high level overview of XML Database technology and the current status of the standards that are driving the technology.

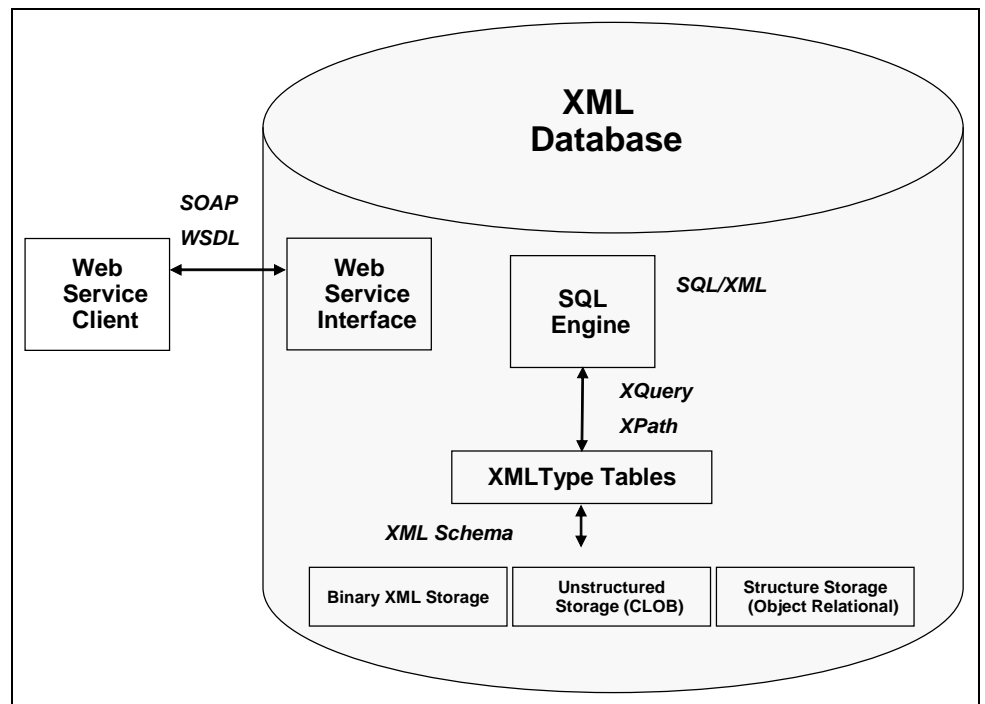


Figure 1 - XML Database and Related Standards

Table 1 – Some Key Industry Standards for XML Database

<i>Standard/Process</i>	<i>Description</i>	<i>Current Status</i>
XML	<p>The base standard itself. Extensible Markup Language (XML) is itself derived from the Standard Generalized Markup Language (ISO 8879:1986) standard.</p> <p>Standard tax documents, such as tax returns, payments, notices and so on can be represented as XML documents and stored in an XML Database.</p>	<p>1.1 W3C Recommendation (16 August 2006)</p>
XQuery	<p>XQuery is a query language for performing query and update functions on XML documents. It is similar to Structured Query Language (SQL) in nature.</p> <p>Tax documents stored in an XML Database can be retrieved, stored and updated using the XQuery language.</p>	<p>1.0 W3C Recommendation (23 January 2007)</p>
XPath	<p>XPath is an expression language for manipulating data elements in an XML document. It is complementary with XQuery.</p>	<p>2.0 W3C Recommendation (23 January 2007)</p>
XML Schema	<p>XML Schema are XML documents that are used to define the structure, content and semantics of other XML documents.</p> <p>An XML Schema can be</p>	<p>1.0 W3C Recommendation (28 October 2004).</p> <p>Version 1.1 is currently under development.</p>

<i>Standard/Process</i>	<i>Description</i>	<i>Current Status</i>
	thought of as a data dictionary that guides the construction and interpretation of XML documents	
SQL/XML	SQL/XML is an ANSI and ISO standard that defines the base capabilities of XML Database technology. It includes specification of features for storing, manipulating and querying XML data.	2003 with addendum ANSI/ISO/IEC 9075-14:2003

Summary

There are clear trends in the tax administration domain that point to increasing use of XML as a medium for data representation, format and exchange.

The XML database market is evolving quickly to support customers who want to expand their capabilities around storing, querying, publishing, and managing resources that are based on XML.

Tax administration organizations should continue to monitor the XML database market, and invest in on-going efforts to remain current with the standards that drive the XML database market, and the capabilities provided by the major database vendors.

Pilot projects that attain business value by exploiting XML database capabilities are an intelligent approach towards understanding the applicability and growth potential of this technology to tax administration environments.



XML Databases Technology and Tax Administration
November 2008

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com

Copyright © 2008, Oracle and/or its affiliates. All rights reserved.
This document is provided for information purposes only and the contents hereof are subject to change without notice.
This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission. Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.