

ISO/IEC JTC 1/SC 32 N 0912

Date: 2003-01-02

REPLACES: --

<p style="text-align: center;">ISO/IEC JTC 1/SC 32</p> <p style="text-align: center;">Data Management and Interchange</p> <p style="text-align: center;">Secretariat: United States of America (ANSI)</p> <p style="text-align: center;">Administered by Pacific Northwest National Laboratory on behalf of ANSI</p>
--

DOCUMENT TYPE	Other document (Open)
TITLE	W3C Liaison Report
SOURCE	Jim Melton
PROJECT NUMBER	
STATUS	This is placed on the SC 32 Agenda at Santa Fe
REFERENCES	
ACTION ID.	FYI
REQUESTED ACTION	
DUE DATE	
Number of Pages	8
LANGUAGE USED	English
DISTRIBUTION	P & L Members SC Chair WG Conveners and Secretaries

Douglas Mann, Secretariat, ISO/IEC JTC 1/SC 32

Pacific Northwest National Laboratory *, 13600 Angelica Court, Chantilly, VA, 20151-3360,
United States of America

Telephone: +1 202-566-2126; Facsimile: +1 202-566-1639; E-mail: MannD@battelle.org

available from the JTC 1/SC 32 WebSite <http://www.jtc1sc32.org/>

*Pacific Northwest National Laboratory (PNL) administers the ISO/IEC JTC 1/SC 32 Secretariat on behalf of ANSI



ISO
International Organization for
Standardization

ANSI
American National Standards Institute



ANSI TC X3H2
Database
ISO/IEC JTC 1/SC 32
Data Management and Interchange
WG 3
Database Languages

- Project:** (None)
- Title:** W3C Liaison Report
- Status:** Liaison Report
- Author:** Jim Melton
- Abstract:** The author is liaison from ISO/IEC JTC1/SC32 to the W3C. This document is intended to inform ISO/IEC JTC1/SC32 of activities of the W3C that are likely to be of particular interest to ISO/IEC JTC1/SC32. This Liaison Report updates information reported almost a year ago in [Previous].
- References:**
- [Previous]** SC32 N0786 ≈ WG3:VIE-030R1 = ANSI H2-2002-086, *W3C Liaison Report*, April 11, 2002
- [XML-WD]** WG3:ZSH-020 = ANSI H2-2002-574, *(Working Draft) XML-Related Specifications (SQL/XML)*, December, 2002
- [XML]** *Extensible Markup Language (XML) 1.0 Second Edition*, W3C Recommendation, 6 October, 2000 (<http://www.w3.org/TR/REC-xml>)
- [XML1.1]** *Extensible Markup Language (XML) 1.1*, 15 October 2002 W3C Candidate Recommendation (<http://www.w3.org/TR/xml11/>)

- [Namespaces]** *Namespaces in XML*, W3C Recommendation, 14 January, 1999 (<http://www.w3.org/TR/REC-xml-names>)
- [Namespaces1.1]** *Namespaces in XML*, W3C Recommendation, 5 September, 2002 (<http://www.w3.org/TR/xml-names11/>)
- [Infoset]** *XML Information Set*, W3C Recommendation, 24 October 2001 (<http://www.w3.org/TR/xml-infoset/>)
- [XPath1.0]** *XML Path Language (XPath) Version 1.0*, W3C Recommendation, 16 November 1999 (<http://www.w3.org/TR/xpath/>)
- [DataModel]** *XQuery 1.0 and XPath 2.0 Data Model*, W3C Working Draft 15 November 2002 (<http://www.w3.org/TR/query-datamodel/>)
- [XQuery]** *XQuery 1.0: An XML Query Language*, W3C Working Draft 15 November 2002 (<http://www.w3.org/TR/xquery/>)
- [Semantics]** *XQuery 1.0 and XPath 2.0 Formal Semantics*, W3C Working Draft 15 November 2002 (<http://www.w3.org/TR/query-semantics/>)
- [F&O]** *XQuery 1.0 and XPath 2.0 Functions and Operators Version 1.0*, W3C Working Draft 15 November 2002 (<http://www.w3.org/TR/xquery-operators/>)
- [XPath2.0]** *XML Path Language (XPath) 2.0*, W3C Working Draft 15 November 2002 (<http://www.w3.org/TR/xpath20/>)
- [Schema]** *XML Schema Part 1: Structures*, W3C Recommendation 2 May 2001 (<http://www.w3.org/TR/xmlschema-1/>), *XML Schema Part 2: Datatypes*, W3C Recommendation 2 May 2001 (<http://www.w3.org/TR/xmlschema-2/>), and *XML Schema 1.0 Specification Errata*, W3C Working Draft, 2 December, 2002 (<http://www.w3.org/2001/05/xmlschema-errata>)
- [TechArch]** *W3C Technical Architecture*, 15 November, 2002 (<http://www.w3.org/TR/webarch>)

1. W3C Projects Related to WG3

The World Wide Web Consortium (W3C) is a consortium that develops and published specifications (“Recommendations”) related to the World Wide Web. The Recommendations published by the W3C are widely accepted as *de facto* standards.

Among the specifications developed by the W3C are several that are related to the Extended Markup Language known as XML, including several specifications that concern XML-related metadata and querying of XML documents.

Before examining the individual specifications, readers should be aware that every well-formed XML document has an Information Set (Infoset, specified in [Infoset]) that describes the content of every node in the tree corresponding to the document. (In fact, an Infoset is a tree structure that is isomorphic to the XML document with which it is associated, or that it describes.)

1.1 Specifications

Most of the several documents cited above in the References for this paper are W3C documents that the author of this paper believes to be of significant interest to WG3 participants. The W3C’s practices include a requirement that every Working Group demonstrate their activity through frequent publication of their working drafts (commonly known as the Group’s “heartbeat”). At least once every three months, active WGs are expected to publish one or more documents that represent the current status of their development efforts; however, not every document being developed by a WG has to be published in every heartbeat cycle.

In addition to the specifications discussed below in more detail, it will be of interest to SC32 participants to know that the specification for the XML language itself [XML] is in the process of being revised [XML1.1]. The primary purpose of XML version 1.1 is to provide support for much more (perhaps “all”) of the Unicode character set. In addition, the namespaces specification [Namespaces] is also being revised [Namespaces1.1].

This report explicitly omits reference to additional documents under preparation by the XML Query Working Group (XML Query Requirements, XML Query Use Cases, and XQueryX). These documents were mentioned in earlier liaison reports and will no doubt be discussed in future reports. They do not seem to this author to be especially urgent for the present report.

The text in this section is largely unchanged since the previous version of my W3C Liaison Report, and is repeated here for the readers’ convenience.

1.1.1 XML Schema ([Schema])

The XML Schema Recommendation was published by the W3C in three parts. Part 0 is non-normative and is a *primer* that explains many aspects of the other two parts. Part 1, named “Structures”, is arguably the most complex of the three parts. It “defines facilities for describing the structure and constraining the contents of XML 1.0 documents, including those which exploit the XML Namespace facility”. Part 2, named “Datatypes”, specifies a taxonomy for XML Schema’s “built-in types”, using facilities defined in Part 1.

The facilities defined in Part 1 are intended to allow the definition of metadata for XML documents of significant complexity. This metadata differs in several important ways from the metadata associated with SQL databases. The most critical difference lies in the fact that XML documents can exist independent of any metadata, that such metadata as may be available can be provided in any of several forms (*e.g.*, DTD, or

Document Type Definition, XML Schema, and other schema-like formats), and that the metadata to which an XML document conforms is only loosely coupled with the document itself.

Another significant complexity introduced by Part 1 is the notion of complex types, both named and anonymous. The “types” covered by Part 1 are *element types*, which do not correspond terribly well with anything in the SQL language, not even user-defined types. XML Schema types can be subtyped, both by restriction and by extension, while SQL user-defined structured types are subtyped only through extension.

Part 2’s datatypes correspond more reasonably with SQL’s predefined data types. There are rather more XML Schema datatypes than SQL has predefined data types, and XML Schema partitions its datatypes into *primitive* types and *derived* types. For example, there are three primitive numeric types, called “decimal”, “real”, and “double”, while there are 13 derived numeric types, such as “integer”, “int”, and “nonPositiveInteger”. XML Schema’s derived types are derived from its primitive types by restriction through the use of *facets* that identify (arguably orthogonal) characteristics of primitive types such as precision, scale, and such.

When an XML document has been validated against one or more XML Schemata, the document’s Infoset is augmented with additional information provided during the course of that Schema validation. The augmented Infoset is known as a Post-Schema Validation Infoset, or PSVI. Among other information, a PSVI provides details of the types of each typed node in the XML document’s Infoset. (In fairness, I should note that the PSVI is technically specified as a set of annotations that are used in conjunction with a corresponding Infoset; however, it is appropriate for most uses to consider the PSVI to include both the original Infoset plus the annotations.)

1.1.2 XPath [XPath2.0]

XPath is a language for “addressing parts of an XML document”. In particular, it is a language whose syntax bears some resemblance to the syntax of file paths in UNIX filesystems. It is used to “navigate” the content of XML documents. XPath generally makes no use of the information (especially the type information) provided by an XML schema, but relies on an XPath-specific data model that can be derived from the document’s Infoset.

[XPath1.0] is among the most widely used of the W3C’s Recommendations, so development of [XPath 2.0] has proven to be quite challenging, particularly in terms of balancing the need for 1.0 compatibility against the demand for new features and the desire for consistency with XQuery.

The most prominent features of XPath are its use of *paths* and *steps*, its dependency on *node sets*, its use of predicates (placed in square brackets in path steps), and—perhaps most interesting—its use of *axes* along which document navigation occurs. XPath defines 13 axes (such as “parent”, “child”, “ancestor-or-self”, and “preceding-sibling”) and provides syntactic sugar for several of the axes (allowing a more compact notation).

1.1.3 XQuery ([DataModel], [Semantics], [XQuery], and [F&O])

In 1999, the W3C established a new Working Group to pursue specifications for XML Query. The group has faced enormous challenges, including (but hardly limited to) development of yet another data model for XML, development of a typing philosophy, and establishment of formal semantics for the query language’s syntax.

The four XQuery-related documents referenced in this paper represent the status of the Working Group’s efforts at the end of 2002.

The foundation of the XQuery efforts is, naturally, the data model described in [DataModel]. The XQuery data model is shared by XPath 2.0 and, consequently, by XSLT 2.0 (not referenced in this document). The XQuery

data model is frequently described as a superset, or extension, of the PSVI (and its associated Infoset). However, in reality, the XQuery data model explicitly omits some PSVI information on the grounds that it is not needed. This has sometimes led to confusion, when other Query-related specifications have assumed that all PSVI information is present—the resolution frequently being the enhancement of the XQuery data model. Most (perhaps all) of the contents of the XQuery data model are made available to the XQuery semantics through the provision of *accessors* that return the value of each attribute. By no means are these accessors required to be available to XQuery application code.

The XQuery syntax is specified in [XQuery] and continues to evolve, although at a relatively slow rate compared with the other specifications (which says volumes about its basic soundness). XQuery syntax is strongly based on a language named Quilt, defined in part by the “father of the SQL language”, Don Chamberlin of IBM.

The [XQuery] document provides the syntax of the language, but only minimal semantics. The formal semantics of XQuery are specified in [Semantics] and are presented in several aspects. First, and perhaps most controversial, are the *static semantics* of the language, which primarily concerns itself with the determination of the types of expressions. While this is in many ways analogous to SQL’s syntax rules that determine the data types of expressions, the presentation in [Semantics] is (in my humble opinion) needlessly baroque and mathematical in style. On the other hand, it is arguably very precise. (Since my previous liaison report [Previous], the static semantics specified in [Semantics] have been revised and are now based on named typing, instead of structural typing.) Somewhat less controversial are the *dynamic semantics*, which is concerned with evaluating expressions at runtime to determine the results of those expressions. Readers should note that “expressions” in XQuery usually have types that are XML Schema types, but they may have static types that cannot be expressed in XML Schema (however, the runtime values of these expressions always have types that can be expressed by XML Schema). The dynamic semantics are expressed in the same notation as the static semantics. It is probably because of the complexity of the typing system (and the fact that it is poorly understood by most XML Query WG participants) that [Semantics] has not been thoroughly reviewed by many people.

Finally, [F&O] is the specification that provides definitions of a great many functions that are used in the definition of XQuery and XPath and that are useful to XQuery and XPath application code. Some functions in [F&O] are specified only to assist in the definition of many XQuery and XPath operators—for example, the addition operator “+” is defined in terms of a function named “numeric-add”. Other functions in [F&O] are accessible to XQuery and XPath code. Some are accessible for both purposes. In general, the functions that are intended primarily for operator definition are defined in a namespace that is identified by the namespace prefix “op:”, while the remaining functions are defined in a namespace identified by the prefix “fn:”.

The most recent publications of [XQuery] and [XPath2.0] are considerably more stabilized than their predecessors, providing an indication that the end of the tunnel is in sight. [DataModel] was rewritten since its previous publication; the major new feature of this rewrite is its handling of namespaces. The alignment of [Semantics] with [XQuery], [XPath2.0], and [DataModel] is almost complete, providing another indication that success is coming closer. The recent publication of [F&O] includes definition of two new subtypes of the XML Schema type “duration”; these two new types are fully ordered and may be adopted in some future edition of [Schema].

1.2 Status

XML Schema (all three parts) became a W3C Recommendation in May, 2001. The XML Schema Working Group remains active and is known to have been reviewing various errors that have been reported in the

Recommendation documents. The XML Schema WG is correcting these errors by progressing an errata document with corrigenda.

Through the public mailing lists, the Schema WG has been gathering requirements for future versions of XML Schema. Whether those requirements result in publication of a “minor version” (e.g., Schema 1.1) or a “major version” (e.g., Schema 2.0) has not been publicly stated, but there are strong indications that a minor version is being considered. It is common for the elapsed time between gathering requirements and publication of a Last Call Working Draft to range between one and three years.

XPath 1.0 achieved Recommendation status in November, 1999, and is currently under revision as XPath 2.0. No public statement has been made regarding the planned publication schedule for XPath 2.0, but the state of the document would suggest that it could be released for Last Call some time in the first half of 2003. Note that the previous version of this Liaison Report [Previous] suggested “the second half of 2002”. That did not take place, but it is clear from reading the public Working Drafts that significant progress has been made towards stabilization of the specification.

The XQuery documents ([DataModel], [XQuery], [Semantics], and [F&O]) continue to be in a state of flux that reflects the complex nature of the technological problems being addressed by the WG. Some public statements have been made that would suggest publication of Last Call for most of the XQuery-related specifications some time in the first half of 2003. The previous version of this Liaison Report [Previous] suggested “the second half of 2002”. That did not take place, but it is clear from reading the public Working Drafts that significant progress has been made towards stabilization of the specifications.

It is reasonable to expect a “heartbeat” publication of the XPath and XQuery working drafts during the first quarter of 2003, since that would comprise a 3-month period following the preceding publication cycle.

2. WG3 Feedback

While there is no formal liaison relationship between the W3C and SC32 or SC32/WG3 that *explicitly* identifies SQL or XML Query, the XML Query WG in particular has made it quite clear that feedback from the standardizers of SQL is useful and interesting. (There is a Class C liaison between the W3C and SC32 that appears to concern XML as a whole, as well as the Dublin Core specifications, which are probably of interest primarily to SC32/WG2.) It is unlikely in the extreme that any formal written liaison from WG3 to the W3C as a whole would find any audience—indeed, this author would not know how to submit such a liaison other than to the W3C Director (Tim Berners-Lee), who would no doubt be astonished and nonplussed to receive it. Therefore, any liaison that WG3 might wish to send to the XML Query WG can be transmitted informally or in writing by your humble servant, but expectations of a formal response should be set accordingly low.

The nature of the XML Query Working Group’s responses to comments on the documents being developed by that WG will change when those documents reach Last Call Working Draft status, as the process requires more formality after that point.

Individuals and organizations can submit comments even now on the Working Drafts by sending email to one of several “comments list”s. The comments lists associated with documents being developed in whole or in part by the XML Query WG are:

- www-ql@w3.org — The public mailing list on query languages. Subscription to this list is open to everybody. The list is publicly archived at <http://lists.w3.org/Archives/Public/www-ql/>.
- www-xml-query-comments@w3.org — The public mailing list for submitting comments on the publications of the XML Query WG. This is not a discussion list, but a way for people to provide their

comments to the XML Query WG, and for the WG to reply. Therefore, it is not a list to which you can subscribe, but one to which you can submit your comments. The list is publicly archived at <http://lists.w3.org/Archives/Public/www-xml-query-comments/>.

3. Technical Architecture Group (TAG)

The W3C has established a Technical Architecture Group (TAG) whose responsibility it is to consider the architecture of the World Wide Web, as well as the architecture of the W3C's specifications related to the Web. The current draft of the Technical Architecture document is [TechArch].

End of paper