ISO/IEC JTC 1/SC 32 N 0576

Date: 2000-10-10

REPLACES: --

ISO/IEC JTC 1/SC 32

Data Management and Interchange

Secretariat: United States of America (ANSI)
Administered by Pacific Northwest National Laboratory on behalf of ANSI

DOCUMENT TYPE	Other Document (Open)
TITLE	Live and Dud Rounds
SOURCE	Jim Melton (USA)
PROJECT NUMBER	
STATUS	Supporting material for SQL/XML Project Split Proposal
REFERENCES	
ACTION ID.	FYI
REQUESTED ACTION	
DUE DATE	
Number of Pages	6
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 *, 901 D Street, SW., Suite 900, Washington, DC, 20024-2115, United States of America

Telephone: +1 703 575 2114; Facsimile; +1 703 671 9180; 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

File: hel077.doc

Printed: Tuesday, 10 October, 2000 at 07:57

JTC 1/SC 32 N00576
H2-2000-___
WG3:HEL-077
2000-10-10 — 10 October, 2000

ISO

Authoritative Version: Adobe Acrobat Portable Document Format (PDF)

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: ANSI: 1234D — ISO: 1.32.3.4

Title: Live and Dud Rounds

Status: Supporting material for **SQL/XML Project Split Proposal**

Author: Jim Melton (USA)

Abstract: [XMLsplit] proposes a subproject split proposal to create a new part of the SQL standard tentatively called SQL/XML. WG3 has significant interests in all items contained in the Program of Work indicated that proposal, but that does not imply that the other SC32 Working Groups do not have interests in one or more of those items. The present paper suggests more specific tasks that do — and that do not — fall within the scope of WG3, indicating for some of those tasks which of SC32's other Working Groups may be most interested.

References:

1) [XMLsplit] SC32 N00536 = WG3:HEL-26R2 = H2-2000-331R2, *Subproject: "XML-Related Specs (SQL/XML)"*, 10 Octrober, 2000

Note to reviewers: In this paper, I occassionally use boxed paragraphs like this one to signify notes to reviewers. These notes are *not* to be applied to the referenced documents.

1. Discussion

[XMLsplit] proposes a subproject split of project 1.32.03.05 to establish a new part of the SQL standard dealing with XML-related facilities. During discussion of the proposal at the October, 2000, SC32 Plenary in Helsinki, there were many questions about which of the bullets under the Program of Work were within WG3's scope and which were properly within the scope of other WGs.

The author of the specific words contained in [XMLsplit] is also the author of the present paper, which is intended to give one person's view of each of the Program of Work bullets. The present paper is intended to suggest a slightly more specific breakdown of each of those bullets, indicating some that are (obviously, the author hopes!) within the scope of WG3 and others that are (also obviously) not within WG3's purview.

The author of the present paper hastens to state that the text herein is solely his responsibility and represents the viewpoints of nobody other than himself. The statements are *opinions* and are certainly not the considered positions of the United States, nor of WG3 participants, and perhaps not even of himself! They are meant to be purely illustrative and the author does not mean to imply that WG3 will or should or even can address every single item that is suggested to be within WG3's scope, nor that any other WG in SC32 will or should address every item that is suggested to be within its scope.

The author fervently hopes that the suggestions contained herein produce more light than heat and apologizes in advance for any other result.

One word of caution: The phrase "XML document" should not be interpreted to mean "books, magazines, and other textual material marked up with XML tags"; instead, it must be interpreted more liberally to mean "all forms of data encoded using XML".

1.1. *Specifications for the representation of SQL data (specifically rows and tables of rows, as well as views and query results) in XML form, and* vice versa

This first bullet includes the ability to "publish" SQL data in an XML format (as exemplified by papers such as WG3:HEL-032, WG3:HEL-033, and WG3:HEL-034, and possibly even WG3:HEL-035). There are many, many approaches to solving this problem and it is not yet clear which approaches (almost certainly more than one!) will capture the attentions of the vendors and the users. Among the variable parameters is the question of whether data is represented as the content of XML tags

```
<quantity>10</quantity>
```

or as attributes of XML tags

```
<quantity value=10/>
```

The problem of storing XML data into an SQL database is arbitrarily complex, since there are a great many ways of "shredding" an XML document into its component parts before storing its contents into SQL. These ways range from merely storing a complete XML document as a (possibly quite long) character string to completely breaking down every tag, element, and attribute into column values in rows of tables.

It is within the scope of WG3 to address the issue of generally publishing the result of queries against SQL databases in an XML form as well as the issue of accepting XML data to be stored (whether decomposed — or "shredded" — or not). However, certain special cases of the representation of SQL data in XML, such as that

required for the use of a client-server protocol like RDA, are almost certainly best defined by experts in those special cases.

1.2. Specifications associated with mapping SQL schemata to and from XML schemata. This may include performing the mapping between existing arbitrary XML and SQL schemata

In order to publish SQL data in an XML form, one must specify the relationship between the schema describing the SQL data and the schema (or, less interesting, the DTD — Document Type Descriptor) that describes the generated XML. In some situations, the XML schema has already been defined and the problem is one of describing the relationships between the components of the SQL schema and that XML schema. In other situations, the XML schema does not yet exist and it may be permissible — or required — that an XML schema be "inferred" and then generated based on the structure of the SQL data being expressed. That information may reside directly in the SQL schema, or it may be derived from an SQL query.

Some of the issues associated with this item are related to issues associated with the preceding item, particularly including decisions regarding whether specific data are to be represented as XML elements or as XML attributes.

The problems are certainly associated with metadata, which may imply that WG2 should investigate whether they have, or should initiate, work associated with this item.

1.3. Specifications for the representation of SQL Schemas in XML

This item is obviously related to the preceding item, but has special implications related to default representations of SQL schemas in an XML form.

Because of the relationship to default representations, it has implications on RDA's use of XML for transporting SQL data across the RDA client-server interface. Therefore, in addition to WG3's work on this item, it would appear that WG5 should investigate whether it should also address it.

1.4. Specifications for the representation of SQL actions (insert, update, delete)

WG3:HEL-036 describes one possible mechanism for representing SQL actions — such as insert, update, and delete, but also transaction initiation and termination, and others — in XML. Such representations have a number of uses, including (but certainly not limited to):

- Database replication, particularly in a heterogeneous environment
- Database monitoring
- Business-to-business communications
- Personalized information notification
- Capturing database changes for publication outside of the database environment

It has been noted that WG5's work on RDA requires the ability to represent, not only SQL data, but also certain SQL actions in the transmissions across the client-server boundary. Thus, it would appear that WG5 would benefit from work done in this area and may profitably contribute to that work.

1.5. Specifications for messaging for XML when used with SQL

The ability to communicate XML across an interface (network or API) often involves formatting and protocols. WG3, in general, has little expertise in protocols (particularly network protocols), but regularly deals with issues of formatting and structuring data for different purposes. The W3C has developed, and continues to develop, network protocols for transmission of XML in a web environment; however, there may be a need for additional messaging capabilities.

If such messaging capabilities involve network communication, then WG5's expertise will be required and the work will undoubtedly be used by RDA.

However, other messaging capabilities — not involving communications, *per se* — may well be required. The SOAP (Simple Object Access Protocol) being popularized by several vendors (names available upon request) is quite likely to become a *de facto* standard and to become a W3C Recommendation (a/k/a "standard"). It is not a foregone conclusion that there will be any WG3 work to be done in this area, but prudence dictates that WG3's interests be protected by inclusion of a possible work item in this area.

1.6. Specifications of the (perhaps "a") manner in which SQL language can be used with XML

This item is intended to cover several possible associations between SQL and XML, one of which is the use of SQL, suitably enhanced, to query repositories of XML documents, including (but not necessarily limited to) such documents stored in SQL database systems.

Consider an XML document such as the following, which could be associated with either a DTD or an XML schema (neither of which is shown, since it is not relevant to the present discussion):

```
<?xml version="1.0" encoding="ISO-8859-1"?>
...
  <purchaseOrder>
    <customer>
      <custID id=12345>
      <address>
        123 Some Street, Sometown, AK 99123 USA
      </address>
    </customer>
    <lineItem>
      <itemNumber SKU="ABC"/>
      <quantity number=38/>
      <inStock/>
      <shippingInstructions>
        This item must be shipped with a carrier that will guarantee
        on-time delivery, as it spoils very quickly!
      </shippingInstructions>
    </lineItem>
    <lineItem>
      <itemNumber SKU="RST"/>
      <quantity number=7/>
    </lineItem>
```

```
  </purchaseOrder>
```

In order to locate that document among possibly many in an archive of purchase orders, or to retrieve certain data from the document once it has been identified, it is necessary to use a query language. The W3C XML Query Working Group is in the process of defining an XML query language and has published both a requirements document and an algebra for querying XML documents. However, as of the writing of the present paper, it has not published a specification of an XML query language.

When it does publish a specification for an XML query language, it may be that the language is not particularly well suited for use alongside SQL for concurrently querying SQL data — or, for that matter, for querying XML documents stored inside SQL databases. If that should be the case, then it might be appropriate for WG3 to devise extensions to SQL:1999 that are suitable for implementing the W3C's query algebra.

One vital question that must be addressed regards *closure* of any such query language: does the language return a table, as SQL does today, or does it return an XML document (or document fragment) as the XML Query Working Group's requirements currently state? Indeed, it may be desirable for the user to specify what is to be returned.

By contrast, consider an XML document such as the following:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
...
  <book>
    <author>...</endauthor>
    <chapter title="An XML Adventure"/>
      <p>
        This is <emphasis>really</emphasis> exciting work.
      </p>
    </chapter>
  </book>
```

In order to locate all books with the phrase "really exciting", ordinary SQL predicates are unlikely to be convenient because a phrase like

```
     WHERE something = '<emphasis>really</emphasis> exciting'"
```

will not match

```
     really exciting
     really <emphasis>exciting</emphasis>
```

or

```
     really exciting
```

The author of the present paper believes that such queries are much more appropriately within the scope of WG4 than WG3.

End of paper