Examples of encoding XML infosets as fast infoset documents

WARNING: This document is based on Annex D of ITU-T Rec. X.891 | ISO/IEC 24824-1 as it was on Jan. 24, 2005. At that time, the standard was under ballot. Annex D, and thus this document, will be updated due to ballot comments that require slight modifications to the encoding and to the examples.

D.1 Introduction of examples

D.1.1 This Annex uses the following typographical conventions for numbers:

a) for a number represented in base ten **bold Courier** is used for the digits of the number, followed by the subscript "10" (for example, **11**$_{10}$); and

b) for a number represented in base sixteen (a hexadecimal number) **bold Courier** is used for the digits of the number, followed by the subscript "16" (for example, **0b1f**$_{16}$); and

c) if the base of a number is explicitly stated, then the subscript is omitted.

D.1.2 This Annex presents two examples of possible encodings of a Universal Business Language (UBL 1.0) order into a fast infoset document. UBL is designed to provide a universally understood and recognized commercial syntax for legally binding business documents.

D.1.3 The XML infoset for the example UBL order is presented in D.3.

D.1.4 The first fast infoset document has an initial vocabulary that references an external vocabulary. Subclause D.4 describes the content of the external vocabulary, the octets of the fast infoset document, and explanations of some octet sequences.

D.1.5 The second fast infoset has no initial vocabulary. Subclause D.5 describes the octets of that fast infoset document, and explanations of some octet sequences.

NOTE – The final vocabulary of this fast infoset document is the same as the final vocabulary of the fast infoset document described in D.4.

D.1.6 The octets of D.4 and D.5 are presented in a series of tables each with two columns. The first column lists the starting position in hexadecimal of 32 consecutive octets of the fast infoset document, and the second column lists the octets in hexadecimal notation. Those hexadecimal characters containing bits that correspond to the identification and termination of information items are underlined.

NOTE – The encoding examples of this document were produced with a Fast Infoset SAX serializer.

D.1.7 The explanations of some octet sequences of the fast infoset documents (in D.4 and D.5) are presented in tables with the following columns:

a) Column 1 presents the position, in hexadecimal, of the octet(s) listed in column 2.

b) Column 2 presents the octet(s) of the fast infoset document associated with a relevant information item and the item's properties. An octet is represented in base two followed by the same octet represented in base sixteen (hexadecimal) in brackets, for example, **11110000 (f0).**

c) Column 3 presents, in detail, a description of the octets in column 2.

d) Column 4 presents a portion of the XML infoset or a portion of the XML 1.0 document (if applicable) corresponding to the octet(s) in column 2.

D.1.8 In these examples all chunks of **character** information items containing less than 6 characters are added to the CONTENT CHARACTER CHUNK table, and the **[normalized value]** property of all **attribute** information items containing less than 6 characters are added to the ATTRIBUTE VALUE table.

D.1.9 The sizes of the XML 1.0 document and of the fast infoset documents, and the compressed sizes (using GZIP) of those documents are listed in D.2

D.2 Size of example documents (including redundancy-based compression)

D.2.1 Table 2 presents the sizes of all documents. Column 1 lists the UBL documents, column 2 lists the document sizes, and column 3 lists the GZIP (with default options, see IETF RFC 1952:1996, *GZIP file format specification version 4.3*) compressed sizes of documents.

NOTE 1 – The UBL Order XML 1.0 document contains no white spaces (see D.3.1.2).

NOTE 2 – For each document all characters are encoded using the UTF-8 character encoding.

NOTE 3 – No XML declaration is serialized for the fast infoset documents.

Table 2 – Initial sizes and GZIP compressed sizes of documents

UBL document	Size	GZIP compressed size
XML 1.0 document	3311	909
Fast infoset document with an external vocabulary	684	527
Fast infoset document with no initial vocabulary	1322	861

D.2.2 The size of the fast infoset document with a reference to an external vocabulary is the smallest in size, and also the smallest in GZIP compressed size. The ratio of GZIP compressed size over the size of the fast infoset document implies that this fast infoset document has little redundant information.

D.2.3 In all cases the GZIP compressed sizes of the fast infoset documents are smaller than the GZIP compressed size of the XML 1.0 document. Furthermore the size of the fast infoset document with a reference to an external vocabulary is smaller than the GZIP compressed size of the XML 1.0 document.

D.3 UBL order example

D.3.1 Joinery Order example

D.3.1.1 The UBL order example is taken. Specifically, the Joinery Order example has been chosen (see xml/joinery/UBL-Order-1.0-Joinery-Example.xml) for the following reasons:

a) it is a real world example developed independently of this Recommendation | International Standard with no particular bias towards Fast Infoset;

b) it is freely available; and

b) it makes extensive use of XML namespaces and thus is a good example to present how Fast Infoset supports XML namespaces.

D.3.1.2 The Joinery Order example has been modified with the following:

a) the last three **OrderLine** elements have been removed; and

NOTE – This reduces the XML 1.0 document to reasonable size for presentation in this Recommendation | International Standard.

b) all white spaces have been removed.

NOTE – This represents a more realistic use case for XML infosets that may be serialized, transmitted over a network, and parsed.

D.3.2 Joinery Order XML 1.0 document

The Joinery Order XML 1.0 document with the modifications as stated in D.3.1.2 a), but with white spaces retained for readability, is presented as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<Order xmlns:res="urn:oasis:names:tc:ubl:codelist:AcknowledgementResponseCode:1:0"
xmlns:cbc="urn:oasis:names:tc:ubl:CommonBasicComponents:1:0"
xmlns:cac="urn:oasis:names:tc:ubl:CommonAggregateComponents:1:0"
xmlns:cur="urn:oasis:names:tc:ubl:codelist:CurrencyCode:1:0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="urn:oasis:names:tc:ubl:Order:1:0"
xsi:schemaLocation="urn:oasis:names:tc:ubl:Order:1:0 ../../xsd/maindoc/UBL-Order-1.0.xsd">
        <BuyersID>S03-034257</BuyersID>
        <cbc:IssueDate>2003-02-03</cbc:IssueDate>
        <cac:BuyerParty>
                <cac:Party>
                        <cac:PartyName>
                                <cbc:Name>Jerry Builder plc</cbc:Name>
                        </cac:PartyName>
                        <cac:Address>
                                <cbc:StreetName>Marsh Lane</cbc:StreetName>
                                <cbc:CityName>Nowhere</cbc:CityName>
                                <cbc:PostalZone>NR18 4XX</cbc:PostalZone>
                                <cbc:CountrySubentity>Norfolk</cbc:CountrySubentity>
                        </cac:Address>
                        <cac:Contact>
```

```
                        <cbc:Name>Eva Brick</cbc:Name>
                    </cac:Contact>
                </cac:Party>
            </cac:BuyerParty>
            <cac:SellerParty>
                <cac:Party>
                    <cac:PartyName>
                        <cbc:Name>Specialist Windows plc</cbc:Name>
                    </cac:PartyName>
                    <cac:Address>
                        <cbc:BuildingName>Snowhill Works</cbc:BuildingName>
                        <cbc:CityName>Little Snoring</cbc:CityName>
                        <cbc:PostalZone>SM2 3NW</cbc:PostalZone>
                        <cbc:CountrySubentity>Whereshire</cbc:CountrySubentity>
                    </cac:Address>
                </cac:Party>
            </cac:SellerParty>
            <cac:Delivery>
                <cbc:RequestedDeliveryDateTime>2003-02-
24T00:00:00</cbc:RequestedDeliveryDateTime>
                <cac:DeliveryAddress>
                    <cbc:StreetName>Riverside Rd.</cbc:StreetName>
                    <cbc:BuildingName>Plot 17, Whitewater Estate</cbc:BuildingName>
                    <cbc:CityName>Whetstone</cbc:CityName>
                    <cbc:CountrySubentity>Middlesex</cbc:CountrySubentity>
                </cac:DeliveryAddress>
            </cac:Delivery>
            <cac:OrderLine>
                <cac:LineItem>
                    <cac:BuyersID>A</cac:BuyersID>
                    <cbc:Quantity quantityUnitCode="unit">2</cbc:Quantity>
                    <cac:Item>
                        <cac:SellersItemIdentification>
                            <cac:ID>236WV</cac:ID>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>wood</cac:AttributeID>
                                <cbc:Description>soft</cbc:Description>
                            </cac:PhysicalAttribute>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>finish</cac:AttributeID>
                                <cbc:Description>primed</cbc:Description>
                            </cac:PhysicalAttribute>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>fittings</cac:AttributeID>
                                <cbc:Description>satin</cbc:Description>
                            </cac:PhysicalAttribute>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>glazing</cac:AttributeID>
                                <cbc:Description>single</cbc:Description>
                            </cac:PhysicalAttribute>
                        </cac:SellersItemIdentification>
                    </cac:Item>
                </cac:LineItem>
            </cac:OrderLine>
            <cac:OrderLine>
                <cac:LineItem>
                    <cac:BuyersID>B</cac:BuyersID>
                    <cbc:Quantity quantityUnitCode="unit">3</cbc:Quantity>
                    <cac:Item>
                        <cac:SellersItemIdentification>
                            <cac:ID>340TW</cac:ID>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>hand</cac:AttributeID>
                                <cbc:Description>RH</cbc:Description>
                            </cac:PhysicalAttribute>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>wood</cac:AttributeID>
                                <cbc:Description>hard</cbc:Description>
                            </cac:PhysicalAttribute>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>finish</cac:AttributeID>
                                <cbc:Description>stain</cbc:Description>
                            </cac:PhysicalAttribute>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>fittings</cac:AttributeID>
                                <cbc:Description>brass</cbc:Description>
                            </cac:PhysicalAttribute>
                            <cac:PhysicalAttribute>
                                <cac:AttributeID>glazing</cac:AttributeID>
                                <cbc:Description>double</cbc:Description>
                            </cac:PhysicalAttribute>
                        </cac:SellersItemIdentification>
                    </cac:Item>
                </cac:LineItem>
            </cac:OrderLine>
```

</Order>

D.4 UBL Order fast infoset document with an external vocabulary

The external vocabulary of the fast infoset document is presented in D.4.1. The octets (as hexadecimal characters) of the fast infoset document are presented in D.4.2. Detailed explanations of some octet sequences in D.4.2 are presented in D.4.3. The fast infoset document cannot be considered self-describing because external information is required (the external vocabulary) to produce complete XML infoset.

NOTE – The fast infoset document can still be processed by a fast infoset parser that cannot obtain the vocabulary tables given the URI but vocabulary table indexes cannot be de-referenced to obtain the necessary information to generate properties of information items.

D.4.1 The UBL Order external vocabulary

D.4.1.1 The external vocabulary of the fast infoset document is specified to be the final vocabulary obtained from the example UBL order XML infoset (see D.3.1.2) that is further modified to contain:

a) no **character** information items; and

b) empty **[normalized value]** properties of the **attribute** information items.

NOTE 1 – This represents a realistic scenario where it is not known in advance what the application-defined content (**character** information items and or **[normalized value]** properties of the **attribute** information items) of an XML infoset will be.

NOTE 2 – In practice it is not expected that the document to be serialized will be used to generate the external vocabulary. It is anticipated that tools will make use of schema, and potentially XML infoset instances of the schema for frequency analysis of strings and qualified names such that smaller index values will be assigned to more frequently occurring information (for example, the frequency of **[local name]** properties in XML infosets may obey a power law series).

D.4.1.2 The URI of the external vocabulary is **urn:oasis:names:tc:ubl:Order:1.0:joinery:example.**

D.4.1.3 Table 3 presents the vocabulary of the UBL Order XML infoset (the vocabulary tables). Column 1 lists the vocabulary table indexes of the vocabulary tables (index), column 2 lists the vocabulary table entries of the PREFIX table (prefix entry), column 3 lists the vocabulary table entries of the NAMESPACE NAME table (namespace name entry), column 4 lists the vocabulary table entries of the LOCAL NAME table (local name entry), column 5 lists the vocabulary table entries of the ELEMENT NAME table (element name entry), column 6 lists the vocabulary table entries of the ATTRIBUTE NAME table (attribute name entry). The index values for the name surrogate entries, of the ELEMENT NAME and ATTRIBUTE NAME tables, are presented in the order as specified for the components of the **`NameSurrogate`** type (**`prefix-name-string-index`**, **`namespace-name-string-index`** and **`local-name-string-index`**). A character of "_" specifies that the value is absent (which only occurs for values of the **`prefix-name-string-index`** and **`namespace-name-string-index`** components).

NOTE 1 – The namespaces name entries (URIs) have been truncated.

NOTE 2 – For the first element name entry (index `1`) there is no reference to a prefix (since the value is absent, represented by "_"), there is a reference to the sixth namespace name entry (index `6`) for the **[namespace name]** property ("urn:oasis:names:tc:ubl:Order:1:0"), and there is a reference to the first local name entry (index `1`) for the **[local name]** property ("Order").

Table 3 – Vocabulary of the UBL Order XML infoset

Index	Prefix entry	Namespace name entry	Local name entry	Element name entry	Attribute name entry
1	res	….AcknowledgementResponseCode:1:0	Order	_ 6 1	5 5 2
2	cbc	….CommonBasicComponents:1:0	schemaLocation	_ 6 3	_ _ 23
3	cac	….CommonAggregateComponents:1:0	BuyersID	2 2 4	
4	cur	….CurrencyCode:1:0	IssueDate	3 3 5	
5	xsi	….XMLSchema-instance	BuyerParty	3 3 6	
6		….Order:1:0	Party	3 3 7	
7			PartyName	2 2 8	
8			Name	3 3 9	

Index	Prefix entry	Namespace name entry	Local name entry	Element name entry	Attribute name entry
9			Address	2 2 10	
10			StreetName	2 2 11	
11			CityName	2 2 12	
12			PostalZone	2 2 13	
13			CountrySubentity	3 3 14	
14			Contact	3 3 15	
15			SellerParty	2 2 16	
16			BuildingName	3 3 17	
17			Delivery	2 2 18	
18			RequestedDeliveryDateTime	3 3 19	
19			DeliveryAddress	3 3 20	
20			OrderLine	3 3 21	
21			LineItem	3 3 3	
22			Quantity	2 2 22	
23			quantityUnitCode	3 3 24	
24			Item	3 3 25	
25			SellersItemIdentification	3 3 26	
26			ID	3 3 27	
27			PhysicalAttribute	3 3 28	
28			AttributeID	2 2 29	
29			Description		

D.4.2 Octets (as hexadecimal characters) of the fast infoset document

Table 4 presents the octets of the fast infoset document for the UBL order example presented in D.3.

NOTE – Hexadecimal characters containing bits that correspond to the identification and termination of information items are underlined.

Table 4 – Octets (as hexadecimal characters) of fast infoset document

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
000000	e00100001010002f75726e3a6f617369733a6e616d65733a74633a75626c3a4f
000020	726465723a313a303a6a6f696e6572793a6578616d706c6578cf8080cf8181cf
000040	8282cf8383cf8484cd85f00000083b75726e3a6f617369733a6e616d65733a74
000060	633a75626c3a4f726465723a313a30202e2e2f2e2e2f7873642f6d61696e646f
000080	632f55424c2d4f726465722d312e302e787364f00182075330332d3033343235
0000a0	37f002820732303033 2d30322d3033f003040506820e4a65727279204275696c
0000c0	64657220706c63ff070882074d61727368204c616e65f00982044e6f77686572
0000e0	65f00a82054e52313820345858f00b82044e6f72666f6c6bff0c068206457661
000100	20427269636bffff0d040506821353706563696 16c6973742057696e646f7773
000120	20706c63ff070e820b536e6f7768696c6c20576f726b73f009820b4c6974746c

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
000140	6520536e6f72696e67f00a8204534d3220334e57f00b8207576865726573 6869
000160	7265ffff0f108210323030332d30322d32345430303a30303a3030f01108820a
000180	52697665727369646520526 42ef00e8217506c6f742031372c20576869746577
0001a0	617465722045737461746 5f00982065768657473746f6e65f00b82064d696464
0001c0	6c65736578fff01213149041f0550143756e6974f09032f01617189202323336
0001e0	5756f0191a9201776f6f64f01b9201736f6674ff191a820366696e697368f01b
000200	82037072696d6564ff191a820566697474696e6773f01b9202736174696eff19
000220	1a8204676c617a696e67f01b820373696e676c65ffffff1213149042f0550180
000240	f09033f016171892023334305457f0191a920168616e64f01b915248ff191aa3
000260	f01b920168617264ff191a820366696e697368f01b9202737461696eff191a82
000280	0566697474696e6773f01b92026272617373ff191a8204676c617a696e67f01b
0002a0	8203646f75626c65ffffffff
0002ac	

D.4.3 Explanation of encoding

D.4.3.1 Encoding of the document information item and the Order element information item

The following explanation details the initial encoding of the fast infoset document (including the URI of the external vocabulary) and the root element information item. In particular, the encoding of a **document** information item, a sequence of **namespace** information items, an **element** information item and an **attribute** information item are explained. Table 5 presents the fragment of the fast infoset document for encoding of the **document** information item and the **Order element** information item of D.3.2. Table 6 details this encoding. The fragment in XML 1.0 is presented as follows:

```
<Order xmlns:res="urn:oasis:names:tc:ubl:codelist:AcknowledgementResponseCode:1:0"
xmlns:cbc="urn:oasis:names:tc:ubl:CommonBasicComponents:1:0"
xmlns:cac="urn:oasis:names:tc:ubl:CommonAggregateComponents:1:0"
xmlns:cur="urn:oasis:names:tc:ubl:codelist:CurrencyCode:1:0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="urn:oasis:names:tc:ubl:Order:1:0"
xsi:schemaLocation="urn:oasis:names:tc:ubl:Order:1:0 ../../xsd/maindoc/UBL-Order-1.0.xsd">
```

Table 5 – Octets (as hexadecimal characters) of fragment

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
000000	e00100001010002f75726e3a6f617369733a6e616d65733a74633a75626c3a4f
000020	726465723a313a303a6a6f696e6572793a6578616d706c6578cf8080cf8181cf
000040	8282cf8383cf8484cd85f00000083b75726e3a6f617369733a6e616d65733a74
000060	633a75626c3a4f726465723a313a30202e2e2f2e2e2f7873642f6d61696e646f
000080	632f55424c2d4f726465722d312e302e787364f0

Table 6 – Encoding details

	Octet(s)	Description	XML infoset or XML
00 01	11100000 (e0) 00000000 (00)	The octets are present at the beginning of every fast infoset document.	**document** information item
02 03	00000000 (00) 00000001 (01)	The octets are the encoding of the version number.	

	Octet(s)	Description	XML infoset or XML
05 06 07	00010000 (10) 00010000 (10) 00000000 (00)	The octets are the encoding of the presence of an initial vocabulary and a reference to an external vocabulary of the initial vocabulary. The octet at position 05_{16}, value 10_{16}, has three '0' (padding) for the first three bits (first to the third bit). The fourth bit is '1' denoting that the **initial-vocabulary** component is present, and the other four optional components are absent. The octet at position 06_{16}, value 10_{16}, has three '0' (padding) for the first three bits. The fourth bit is '1' denoting that the **external-vocabulary** of the **initial-vocabulary** is present. The last four bits are '0' (fifth to eighth bits) denoting that the four of the twelve other optional components are absent. The octet at position 07_{16}, value 00_{16}, has '0' for all bits denoting that the last eight of the twelve optional components are absent.	
08 09 37	00101111 (2f) 01110101 (75) 01100101 (65)	The octets are the encoding of the URI of the external vocabulary. The octet at position 08_{16}, value $2f_{16}$, has a '0' (padding) for the first bit. The URI is encoded as UTF-8 characters. The second bit is '0' denoting that the length of the URI is greater than or equal 1_{10} octet and less than or equal to 64_{10} octets, and that the length, minus the lower bound, is encoded in bits three to eight as an unsigned integer. The unsigned integer is 47_{10} and the length is 48_{10} (the lower bound is 1). The 48_{10} octets of the encoded UTF-8 characters (of the URI) are encoded from the octet at position 09_{16} to the octet at position 37_{16}.	
38	01111000 (78)	The octet is the initial encoding of a child of the **document** information item. The octet at position 38_{16}, value 78_{16}, has a '0' (identification) for the first bit denoting that there is a child of the **document** information item, and the child is an **element** information item. The second bit is '1' denoting that the **element** information item has attributes. The third to sixth bits are '1110' followed by '00' (padding) on the seventh and eighth bits, denoting that namespace **attribute** information items are present.	**element** information item with **[namespace attribute]** property
39 3a 3b	11001111 (cf) 10000000 (80) 10000000 (80)	The octets are the encoding of namespace **attribute** information item with indexed **[prefix]** and **[normalized value]** properties. The octet at position 39_{16}, value cf_{16}, has '110011' (identification) for the first to sixth bits (the first to the fifth bit) denoting that a namespace **attribute** information item is present. The seventh bit is '1' denoting that the **[prefix]** property is present. The eighth bit is '1' denoting that the **[normalized value]** property is present. The octet at position $3a_{16}$, value 80_{16}, has '1' for the first bit denoting that an index is encoded, and the index into the PREFIX table will identify the **[prefix]** property. The second bit is '0' denoting that the index is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the index is encoded in bits three to eight as an unsigned integer. The unsigned integer is 0_{10} and the index is 1_{10} (the lower bound is 1_{10}), which results in the **[prefix]** property "res" when de-referenced from the PREFIX table. The octet at position $3b_{16}$, value 80_{16}, has '1' for the first bit denoting that an index is encoded, and the index into the NAMESPACE NAME table will identify the **[normalized value]** property. The second bit is '0' denoting that the index is greater	**xmlns:res= "….ResponseCode:1:0"**

	Octet(s)	Description	XML infoset or XML
		than or equal to 1_{10} and less than or equal to 64_{10}, and the index is encoded in bits three to eight as an unsigned integer. The unsigned integer is 0_{10} and the index is 1_{10} (the lower bound is 1_{10}), which results in the **[normalized value]** property "….ResponseCode:1.0" when de-referenced from the NAMESPACE NAME table.	
3c 3d 3e	11001111 (cf) 10000001 (81) 10000001 (81)	The octets are the encoding of namespace **attribute** information item with indexed **[prefix]** and **[normalized value]** properties. The index of **[prefix]** property is 2_{10}, which results in a value of "cbc" when de-referenced from the PREFIX table. The index of **[normalized value]** property is 2_{10}, which results in a value of "….sicComponents:1.0" when de-referenced from the NAMESPACE NAME table.	**xmlns:cbc= "….sicComponents:1:0"**
3f 40 41	11001111 (cf) 10000010 (82) 10000010 (82)	The octets are the encoding of namespace **attribute** information item with indexed **[prefix]** and **[normalized value]** properties.	**xmlns:cac= "….ateComponents:1:0"**
42 43 44	11001111 (cf) 10000011 (83) 10000011 (83)	The octets are the encoding of namespace **attribute** information item with indexed **[prefix]** and **[normalized value]** properties.	**xmlns:cur= "….CurrencyCode:1:0"**
45 46 47	11001111 (cf) 10000100 (84) 10000100 (84)	The octets are the encoding of namespace **attribute** information item with indexed **[prefix]** and **[normalized value]** properties.	**xmlns:xsi= "….Schema-instance"**
48 49	11001101 (cd) 10000101 (85)	The octets are the encoding of namespace **attribute** information item with an indexed **[normalized value]** property. The octet at position 48_{16}, value cd_{16}, has a seventh bit of '0' denoting that the **[prefix]** property is absent, and an eighth bit of '1' denoting that the **[normalized value]** property is present.	**xmlns="….Order:1:0"**
4a	11110000 (f0)	The octet is the encoding of the terminator for the sequence of namespace **attribute** information items. The octet at position $4a_{16}$, value $f0_{16}$, has '1111' (terminator) for the first four bits (the first to the fourth bit) and is the terminator for the sequence. Four out of the six '0' (padding) are present on the fifth to eighth bits.	
4b	00000000 (00)	The octet is the encoding of an indexed qualified name of the **element** information item. The octet at position $4b_{16}$, value, 00_{16}, has the last two out of the six '0' (padding) on the first and second bits. The third bit is '0' denoting that the qualified name is not a literal qualified name and is indexed. The index is greater than or equal to 1_{10} and less than or equal to 32_{10}, and the index is encoded in bits four to eight as an unsigned integer. The unsigned integer is 0_{10} and the index is 1_{10} (the lower bound is 1_{10}), which results in qualified name with a **[namespace name]** property of "….Order:1.0" and a **[local name]** property of "Order" (there is no **[prefix]** property for this qualified name) when de-referenced from the ELEMENT NAME table.	**<Order ….**
4c 4d 4e 4f 92	00000000 (00) 00001000 (08) 01111011 (3b) 01110101 (75) 01101000 (64)	The octets are the encoding of an **attribute** information item with an indexed qualified name and a **[normalized value]** property. The presence of **attribute** information items was denoted in the octet at position 38_{16} (second bit is '1'). The octet at position $4c_{16}$, value 00_{16}, has a first bit of '0' (identification) denoting that an **attribute** information item is present. The second bit is '0' denoting that the qualified name is not a literal qualified name and is indexed. The index is greater	**xsi:schemaLocation="…."**

	Octet(s)	Description	XML infoset or XML
		than or equal to 1_{10} and less than or equal to 64_{10}, and the index is encoded in bits three to eight as an unsigned integer. The unsigned integer is 0_{10} and the index is 1_{10} (the lower bound is 1_{10}), which results in qualified name with a **[prefix]** property of "xsi", a **[namespace name]** property of "….Schema-instance" and a **[local name]** property of "schemaLocation" when de-referenced from the ATTRIBUTE NAME table. The octet at position $4d_{16}$, value 08_{16}, is the initial encoding of a non identifying string or index for the **[normalized value]** property. The first bit is '0' denoting that a literal character string is present. The second bit is '0' denoting that the literal character string should not be added to ATTRIBUTE VALUE table. The third and fourth bits, both '0', denote that the encoding format of the string is UTF-8. The fifth and sixth bits are '1' and '0' respectively denoting that length of the octets of the encoded UTF-8 characters (the **[normalized value]** property) is greater than or equal to 9_{10} octets and less than or equal to 264_{10} octets, and that the length, minus the lower bound, is encoded in eight bits on the next octet as an unsigned integer. The seventh to eighth bits are '0' (padding). The octet at position $4e_{16}$, value $3b_{16}$, is the encoding of the unsigned integer. The length of octets of the encoded UTF-8 characters is 68_{10} (the lower bound is 9_{10}). The 68_{10} octets of the encoded UTF-8 characters (of the **[normalized value]** property) are encoded from the octet at position $4f_{16}$ to the octet at position 92_{16}.	
93	11110000 (f0)	The octet is the encoding of the terminator for the sequence of **attribute** information items. The octet at position 93_{16}, value $f0_{16}$, has '1111' for the first four bits (the first to the fourth bit) and is the terminator for the sequence. Four '0' (padding) are present (the fifth to the eighth bit) since the **Order element** information item has children.	

D.4.3.2 Encoding of the Address element information item of the BuyerParty element information item

The following explanation details the encoding of the **Address element** information item of the **BuyerParty element** information item of the fast infoset document. In particular, the encoding of **element** information items and **character** information items are explained. Table 7 presents the fragment of the fast infoset document for encoding of the **Address element** information item of the **BuyerParty element** information item of D.3.2. Table 8 details this encoding. The fragment in XML 1.0 is presented as follows:

```
<cac:Address>
        <cbc:StreetName>Marsh Lane</cbc:StreetName>
        <cbc:CityName>Nowhere</cbc:CityName>
        <cbc:PostalZone>NR18 4XX</cbc:PostalZone>
        <cbc:CountrySubentity>Norfolk</cbc:CountrySubentity>
</cac:Address>
```

Table 7 – Octets (as hexadecimal characters) of fragment

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
0000c0	070882074d6172736204c616e65f00982044e6f77686572
0000e0	65f00a82054e52313820345858f00b82044e6f72666f6c6bff

Table 8 – Encoding details

	Octet(s)	Description	XML infoset or XML
c8	00000111 (07)	The octet is the encoding of the **Address element** information item. The octet at position $c8_{16}$, value 07_{16}, has a '0' (identification) for the first bit denoting that there is a child of an **element** information item (child of the **Party element** information item), and the child is an **element** information item. The second bit is '0' denoting that the **element** information item does not have attributes. The third bit is '0' denoting that the qualified name is not a literal qualified name and is indexed. The index is greater than or equal to 1_{10} and less than or equal to 32_{10}, and the index is encoded in bits four to eight as an unsigned integer. The unsigned integer is 7_{10} and the index is 8_{10} (the lower bound is 1_{10}), which results in qualified name with a **[prefix]** of "cac", a **[namespace name]** property of "….gateComponents:1.0" and a **[local name]** property of "Address" when de-referenced from the ELEMENT NAME table.	**<cac:Address>**
c9	00001000 (08)	The octet is the encoding of the **StreetName element** information item. The **element** information item has an index of 9_{10}, which results in qualified name with a **[prefix]** of "cbc", a **[namespace name]** property of "….BasicComponents:1:0" and a **[local name]** property of "StreetName" when de-referenced from the ELEMENT NAME table.	**<cbc:StreetName>**
ca	10000010 (82)	The octets are the encoding of the **character** information items of the **StreetName element** information item. The octet at position ca_{16}, value 82_{16}, has '10' (identification) for the first two bits (the first to second bit) denoting that there is a child of **element** information item (child of the **StreetName element** information item), and the child is a chunk of **character** information items. The third bit is '0' denoting that a literal character string is present. The fourth bit is '0' denoting that the literal character string should not be added to CONTENT CHARACTER CHUNK table. The fifth and sixth bits, both '0', denote that the encoding format of the chunk is UTF-8. The seventh and eighth bits are '1' and '0' respectively denoting that length of the octets of the encoded UTF-8 characters (the chunk of **character** information items) is greater than or equal to 3_{10} octets and less than or equal to 258_{10} octets, and that the length, minus the lower bound, is encoded in eight bits on the next octet as an unsigned integer. The octet at position cb_{16}, value 07_{16}, is the unsigned integer. The length of octets of the encoded UTF-8 characters is 10_{10} (the lower bound is 3_{10}). The 10_{10} octets of the encoded UTF-8 characters are encoded from the octet at position cc_{10} to the octet at position $d5_{10}$.	**character** information items "Marsh Lane"
cb	00000111 (07)		
cc	01001101 (4d)		
....			
d5	01100101 (65)		
d6	11110000 (f0)	The octet is the terminator for the **StreetName element** information item. The octet at position $d6_{16}$, value $f0_{16}$, has '1111' (terminator) for the first four bits (the first to the fourth bit) and is the terminator for the **StreetName element** information item. The fifth to eighth bits are '0' (padding) since a further child (peer) occurs (**CityName element** information item).	**</cbc:StreetName>**
d7	00001001 (09)	The octet is the encoding of the **CityName element** information item. The **element** information item has an index of 10_{10}, which results in qualified name with a **[prefix]** of "cbc", a **[namespace name]** property of "….BasicComponents:1:0" and a **[local name]**	**<cbc:CityName>**

	Octet(s)	Description	XML infoset or XML
		property of "CityName" when de-referenced from the ELEMENT NAME table.	
d8	10000010 (82)	The octets are the encoding of the **character** information items of the **CityName element** information item.	**character** information items "Nowhere"
d9	00000100 (04)		
da	01001110 (4e)	The 7_{10} octets of the encoded UTF-8 characters are encoded from the octet at position da_{16} to the octet at position $e0_{16}$.	
....			
e0	01100101 (65)		
e1	11110000 (f0)	The octet is the terminator for the **CityName element** information item.	**</cbc:CityName>**
e2	00001010 (0a)	The octet is the encoding of the **PostalZone element** information item. The **element** information item has an index of 11_{10}, which results in qualified name with a **[prefix]** of "cbc", a **[namespace name]** property of "….BasicComponents:1:0" and a **[local name]** property of "PostalZone" when de-referenced from the ELEMENT NAME table.	**<cbc:PostalZone>**
e3	10000010 (82)	The octets are the encoding of the **character** information items of the **PostalZone element** information item.	**character** information items "NR18 4XX"
e4	00000101 (05)		
e5	01001110 (4e)	The 8_{10} octets of the encoded UTF-8 characters are encoded from the octet at position $e5_{16}$ to the octet at position ec_{16}.	
....			
ec	01011000 (58)		
ed	11110000 (f0)	The octet is the terminator for the **PostalZone element** information item.	**</cbc:PostalZone>**
ee	00001011 (0b)	The octet is the encoding of the **CountrySubentity element** information item. The **element** information item has an index of 12_{10}, which results in qualified name with a **[prefix]** of "cbc", a **[namespace name]** property of "….BasicComponents:1:0" and a **[local name]** property of "CountrySubentity" when de-referenced from the ELEMENT NAME table.	**<cbc:CountrySubentity>**
ef	10000010 (82)	The octets are the encoding of the **character** information items of the **CountrySubentity element** information item.	**character** information items "Norfolk"
f0	00000100 (04)		
f1	01001110 (4e)	The 7_{10} octets of the encoded UTF-8 characters are encoded from the octet at position $f1_{16}$ to the octet at position $f7_{16}$.	
....			
f7	01101011 (6b)		
f8	11111111 (ff)	The octet is the terminator for the **CountrySubentity element** information item and the **Address element** information item. The octet at position $f8_{16}$, value ff_{16}, has '1111' (terminator) for the first four bits (the first to the fourth bit) and is the terminator for the **CountrySubentity element** information item. The last four bits (the fifth to eighth bit) are '1111' and is the terminator for the **Address element** information item.	**</cbc:CountrySubentity>** **</cac:Address>**

D.5 UBL order fast infoset document without an initial vocabulary

The octets (as hexadecimal characters) of the fast infoset document are presented in D.5.1. Detailed explanations of some octet sequences in D.5.1 are presented in D.5.2. The final vocabulary of this fast infoset document and the former fast infoset document will be the same since the vocabulary table indexes of the tables in the external vocabulary are generated in the same order. Since the strings are embedded in the fast infoset document this will result in a larger size. The cost of including the strings is 635_{10} bytes (the size of this fast infoset document minus the size of the former fast infoset document), which makes up for approximately half the document size (for larger documents this difference should be less as the

vocabulary will tend to be a fixed cost). Unlike the former fast infoset document this document can be considered self-describing because the XML infoset can be produced without any external information (an external vocabulary).

D.5.1 Octets (as hexadecimal characters) of the fast infoset document

Table 9 presents the octets of the fast infoset document for the UBL order example presented in D.3.

NOTE – Hexadecimal characters containing bits that correspond to the identification and termination of information items are underlined.

Table 9 – Octets (as hexadecimal characters) of fast infoset document

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
000000	e00100000078cf027265733e75726e3a6f617369733a6e616d65733a74633a75
000020	626c3a636f64656c6973743a41636b6e6f776c656467656d656e74526573706f
000040	6e7365436f64653a313a30cf026362632f75726e3a6f617369733a6e616d6573
000060	3a74633a75626c3a436f6d6d6f6e4261736963436f6d706f6e656e74733a313a
000080	30cf026361633375726e3a6f617369733a6e616d65733a74633a75626c3a436f
0000a0	6d6d6f6e41676772656761746543 6f6d706f6e656e74733a313a30cf02637572
0000c0	2f75726e3a6f617369733a6e616d65733a74633a75626c3a636f64656c697374
0000e0	3a43757272656e6379436f64653a313a30cf027873692868747 4703a2f2f7777
000100	772e77332e6f72672f323030312f584d4c536368656d612d696e7374616e6365
000120	cd1f75726e3a6f617369733a6e616d65733a74633a75626c3a4f726465723a31
000140	3a30f03d85044f726465727b84840d736368656d614c6f636174696f6e083b75
000160	726e3a6f617369733a6e616d65733a74633a75626c3a4f726465723a313a3020
000180	2e2e2f2e2e2f7873642f6d61696e646f632f55424c2d4f726465722d312e302e
0001a0	787364f03d850742757965727349448207533033 2d303334323537f03f818108
0001c0	4973737565446174658207323030332d30322d3033f03f828209427579657250
0001e0	617274793f828204506174 74793f828208506174 7274794e616d653f8181034e61
000200	6d65820e4a6572727920427569 6c64657220706c63ff3f828206416464726573
000220	733f8181095374726565744e616d6582074d61727368204c616e65f03f818107
000240	436974794e616d6582044e6f7768657265f03f818109506f7374616c5a6f6e65
000260	82054e52313820345858f03f81810f436f756e747279537562656e7469747982
000280	044e6f72666f6c6bff3f828206436f6e7461637406820645766120427269636b
0002a0	ffff3f82820a53656c6c6572506172747904050682135370656369616c697374
0002c0	2057696e646f777320706c63ff073f81810b4275696c64696e674e616d65820b
0002e0	536e6f7768696c6c20576f726b73f009820b4c6974746c6520536e6f72696e67
000300	f00a8204534d3220334e57f00b8207576865726573686972 65ffff3f82820744
000320	656c69766572793f8181185265717565737465644656c697665727944617465
000340	54696d658210323030332d30322d32345430303a30303a3030f03f82820e4465
000360	6c697665727941646472657373 08820a52697665727369646520526 42ef00e82
000380	17506c6f742031372c20576869746577617465722045737461746 5f009820657
0003a0	68657473746f6e65f00b82064d6964646c65736578fff03f8282084f72646572
0003c0	4c696e653f8282074c696e654974656d3f8282829041f07f8181075175616e74
0003e0	697479780f7175616e74697479556e6974436f646543756e6974f09032f03f82
000400	82034974656d3f82821853656c6c6572734974656d4964656e74696669636174

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
000420	696f6e3f828201494492023233365756f03f828210506879736963616c417474
000440	7269627574653f82820a41747472696275746549449201776f6f64f03f81810a
000460	446573637269707469 6f6e9201736f6674ff191a820366696e697368f01b8203
000480	7072696d6564ff191a820566697474696e6773f01b9202736174696eff191a82
0004a0	04676c617a696e67f01b820373696e676c65ffffff1213149042f0550180f090
0004c0	33f0161718920233343054 57f0191a920168616e64f01b915248ff191aa3f01b
0004e0	920168617264ff191a820366696e697368f01b9202737461696eff191a820566
000500	697474696e6773f01b92026272617373ff191a8204676c617a696e67f01b8203
000520	646f75626c65ffffffff
00052a	

D.5.2 Explanation of encoding

D.5.2.1 Encoding of the document information item and the Order element information item

The following explanation details the initial encoding of the fast infoset document and the root element information item. In particular, the encoding of a **document** information item, a sequence of **namespace** information items, an **element** information item and an **attribute** information item are explained. Table 10 presents the fragment of the fast infoset document for encoding of the **document** information item and the **Order element** information item of D.3.2. Table 11 details this encoding. The fragment in XML 1.0 is presented as follows:

```
<Order xmlns:res="urn:oasis:names:tc:ubl:codelist:AcknowledgementResponseCode:1:0"
xmlns:cbc="urn:oasis:names:tc:ubl:CommonBasicComponents:1:0"
xmlns:cac="urn:oasis:names:tc:ubl:CommonAggregateComponents:1:0"
xmlns:cur="urn:oasis:names:tc:ubl:codelist:CurrencyCode:1:0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="urn:oasis:names:tc:ubl:Order:1:0"
xsi:schemaLocation="urn:oasis:names:tc:ubl:Order:1:0 ../../xsd/maindoc/UBL-Order-1.0.xsd">
```

Table 10 – Octets (as hexadecimal characters) of fragment

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
000000	e00100000078cf027265733e75726e3a6f617369733a6e616d65733a74633a75
000020	626c3a636f64656c6973743a41636b6e6f776c656467656d656e74526573706f
000040	6e7365436f64653a313a30cf026362632f75726e3a6f617369733a6e616d6573
000060	3a74633a75626c3a436f6d6d6f6e426173696436f6d706f6e656e74733a313a
000080	30cf026361633375726e3a6f617369733a6e616d65733a74633a75626c3a436f
0000a0	6d6d6f6e4167677265676174 65436f6d706f6e656e74733a313a30cf02637572
0000c0	2f75726e3a6f617369733a6e616d65733a74633a75626c3a636f64656c697374
0000e0	3a43757272656e6379436f64653a313a30cf027873692868747 4703a2f2f7777
000100	772e77332e6f72672f323030312f584d4c536368656d612d696e7374616e6365
000120	cd1f75726e3a6f617369733a6e616d65733a74633a75626c3a4f726465723a31
000140	3a30f03d85044f726465727b84840d736368656d614c6f636174696f6e083b75
000160	726e3a6f617369733a6e616d65733a74633a75626c3a4f726465723a313a3020
000180	2e2e2f2e2e2f7873642f6d61696e646f632f55424c2d4f726465722d312e302e
0001a0	787364f0

Table 11 – Encoding details

	Octet(s)	Description	XML infoset or XML
00	11100000 (e0)	The octets are present at the beginning of every fast infoset document.	**document** information item
01	00000000 (00)		
02	00000000 (00)	The octets are the encoding of the version number.	
03	00000001 (01)		
04	00000000 (00)	The octets are encoding of the presence of an initial vocabulary and other components of the **Document** type. The octet at position 05_{16}, value 10_{16}, has three '0' (padding) for the first three bits (first to the third bit). The fourth to eighth bits are '00000' denoting that all optional components of the **Document** type are absent (including the **initial-vocabulary** component whose absence is denoted on the fourth bit).	
05	01111000 (78)	The octet is the initial encoding of a child of the **document** information item. The octet at position 05_{16}, value 78_{16}, has a '0' (identification) for the first bit denoting that there is a child of the **document** information item, and the child is an **element** information item. The second bit is '1' denoting that the **element** information item has attributes. The third to sixth bits are '1110' followed by '00' (padding) on the seventh and eighth bits, denoting that namespace **attribute** information items are present.	**element** information item with **[namespace attribute]** property.
06	11001111 (cf)	The octets are the encoding of namespace **attribute** information item with literal **[prefix]** and **[normalized value]** properties. The octet at position 06_{16}, value cf_{16}, has '110011' (identification) for the first to sixth bits denoting that a namespace **attribute** information item is present. The seventh bit is '1' denoting that the **[prefix]** property is present. The eighth bit is '1' denoting that the **[normalized value]** property is present. The octet at position 07_{16}, value 02_{16}, has '0' for the first bit denoting that a literal character string is encoded for the for the **[prefix]** property. The second bit is '0' denoting that the length of encoded UTF-8 characters is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the length is encoded in bits three to eight as an unsigned integer. The unsigned integer is 2_{10} and the length is 3_{10} (the lower bound is 1_{10}). The 3_{10} octets of the encoded UTF-8 characters (of the **[prefix]** property) are encoded from the octet at position 08_{16} to the octet at position $0a_{10}$. The string "res" will be added to the PREFIX table (with an index of 1_{10}). The octet at position $0b_{16}$, value $3e_{16}$, has '0' for the first bit denoting that a literal character string is encoded for the for the **[normalized value]** property. The second bit is '0' denoting that the length of encoded UTF-8 characters is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the length is encoded in bits three to eight as an unsigned integer. The unsigned integer is 62_{10} and the length is 63_{10} (the lower bound is 1_{10}). The 63_{10} octets of the encoded UTF-8 characters (of the **[normalized value]** property) are encoded from the octet at position $0c_{16}$ to the octet at position $4a_{16}$. The string "….ResponseCode:1:0" will be added to the NAMESPACE NAME table (with an index of 1_{10}).	**xmlns:res=** **"….ResponseCode:1:0"**
07	00000010 (02)		
08	01110010 (72)		
....	....		
0a	01110010 (73)		
0b	00111110 (3e)		
0c	01110101 (75)		
....	....		
4a	01110000 (30)		
4b	11001111 (cf)	The octets are the encoding of namespace **attribute** information item with literal **[prefix]** and **[normalized value]** properties.	**xmlns:cbc=**

Octet(s)		Description	XML infoset or XML
4c	00000010 (02)	The 3_{10} octets of the encoded UTF-8 characters (of the **[prefix]** property) are encoded from the octet at position $4c_{16}$ to the octet at position $4f_{16}$. The string "cbc" will be added to the PREFIX table (with an index of 2_{10}).	"….sicComponents:1:0"
4d	01100011 (63)		
....	....		
4f	01100011 (63)	The 48_{10} octets of the encoded UTF-8 characters (of the **[normalized value]** property) are encoded from the octet at position 51_{16} to the octet at position 80_{16}. The string "….sicComponents:1:0" will be added to the NAMESPACE NAME table (with an index of 2_{10}).	
50	00101111 (2f)		
51	01110101 (75)		
....	....		
80	01110000 (30)		
81	11001111 (cf)	The octets are the encoding of namespace **attribute** information item with literal **[prefix]** and **[normalized value]** properties.	xmlns:cac=
82	00000010 (02)		"….ateComponents:1:0"
83	01100011 (63)	The 3_{10} octets of the encoded UTF-8 characters (of the **[prefix]** property) are encoded from the octet at position 83_{16} to the octet at position 85_{16}. The string "cac" will be added to the PREFIX table (with an index of 3_{10}).	
....	....		
85	01100011 (63)		
86	00110011 (33)	The 52_{10} octets of the encoded UTF-8 characters (of the **[normalized value]** property) are encoded from the octet at position 87_{16} to the octet at position ba_{16}. The string "….ateComponents:1:0" will be added to the NAMESPACE NAME table (with an index of 3_{10}).	
87	01110101 (75)		
....	....		
ba	01110000 (30)		
bb	11001111 (cf)	The octets are the encoding of namespace **attribute** information item with literal **[prefix]** and **[normalized value]** properties.	xmlns:cur=
bc	00000010 (02)		"….CurrencyCode:1:0"
bd	01100011 (63)	The 3_{10} octets of the encoded UTF-8 characters (of the **[prefix]** property) are encoded from the octet at position bd_{16} to the octet at position bf_{16}. The string "cur" will be added to the PREFIX table (with an index of 4_{10}).	
....	....		
bf	01100011 (63)		
c0	00101111 (2f)	The 48_{10} octets of the encoded UTF-8 characters (of the **[normalized value]** property) are encoded from the octet at position $c1_{16}$ to the octet at position fo_{16}. The string "….CurrencyCode:1:0" will be added to the NAMESPACE NAME table (with an index of 4_{10}).	
c1	01110101 (75)		
....	....		
f0	01110000 (30)		
f1	11001111 (cf)	The octets are the encoding of namespace **attribute** information item with literal **[prefix]** and **[normalized value]** properties.	xmlns:xsi=
f2	00000010 (02)		"….Schema-instance"
f3	01111000 (78)	The 3_{10} octets of the encoded UTF-8 characters (of the **[prefix]** property) are encoded from the octet at position $f3_{16}$ to the octet at position $f5_{16}$. The string "xsi" will be added to the PREFIX table (with an index of 5_{10}).	
....	....		
f5	01101001 (69)		
f6	00101000 (28)	The 41_{10} octets of the encoded UTF-8 characters (of the **[normalized value]** property) are encoded from the octet at position $f7_{16}$ to the octet at position $1ff_{16}$. The string "….Schema-instance" will be added to the NAMESPACE NAME table (with an index of 5_{10}).	
f7	01101000 (68)		
....	....		
11f	01110111 (77)		
120	11001101 (cd)	The octets are the encoding of namespace **attribute** information item with an indexed **[normalized value]** property.	xmlns="….Order:1:0"
121	00011111 (1f)		
122	01110101 (75)	The 32_{10} octets of the encoded UTF-8 characters (of the **[normalized value]** property) are encoded from the octet at position 122_{16} to the octet at position 141_{16}. The string "….Order:1:0" will be added to the NAMESPACE NAME table (with an index of 6_{10}).	
....	....		
141	00110000 (30)		
142	11110000 (f0)	The octet is the encoding of the terminator for the sequence of namespace **attribute** information items.	
		The octet at position 142_{16}, value $f0_{16}$, has '1111' (terminator) for the first four bits (the first to the fourth bit) and is the terminator for the sequence. Four out of the six '0' (padding) are	

Octet(s)		Description	XML infoset or XML
		present on the fifth to eighth bits.	
143	00111101 (3d)	The octets are the encoding of a literal qualified name of the **element** information item.	**<Order ....**
144	10000101 (85)		
145	00000100 (04)	The octet at position 143_{16}, value $3d_{16}$, has the last two out of the six '0' (padding) on the first and second bits. The third to fifth bits are '1111' denoting that the qualified name is a literal qualified name. The seventh bit is '0' denoting that the qualified name does not have a **[prefix]** property. The eighth bit is '1' denoting that the qualified name has a **[namespace name]** property.	
146	01001111 (4f)		
....	....		
14a	01110010 (72)		
		The octet at position 144_{16}, value 85_{16}, has '1' for the first bit denoting that the **[namespace name]** property is not a literal string and is indexed. The second bit is '0' denoting that the index is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the index is encoded in bits three to eight as an unsigned integer. The unsigned integer is 5_{10} and the index is 6_{10} (the lower bound is 1_{10}), which results in a **[namespace name]** property of "….Order:1:0" when de-referenced from the NAMESPACE NAME table.	
		The octet at position 145_{16}, value 04_{16}, has '0' for the first bit denoting that that a literal character string is encoded for the for the **[local name]** property. The second bit is '0' denoting that the length of encoded UTF-8 characters is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the length is encoded in bits three to eight as an unsigned integer. The unsigned integer is 4_{10} and the length is 5_{10} (the lower bound is 1_{10}).	
		The 5_{10} octets of the encoded UTF-8 characters (of the **[local name]** property) are encoded from the octet at position 146_{16} to the octet at position $14a_{16}$. The string "Order" will be added to the LOCAL NAME table (with an index of 1_{10}).	
		The qualified name with no **[prefix]** property, a **[namespace name]** property of "….Order:1:0" (index 6_{10}), and a **[local name]** property of "Order" (index 1_{10}) will be added to the ELEMENT NAME table (with an index of 1_{10}).	
14b	01111011 (7b)	The octets are the encoding of an **attribute** information item with a literal qualified name and a **[normalized value]** property. The presence of **attribute** information items was denoted in the octet at position 05_{16} (second bit is '1').	**xsi:schemaLocation="…."**
14c	10000100 (84)		
14d	10000100 (84)		
14e	00001101 (0d)	The octet at position $14b_{16}$, value $7b_{16}$, has a first bit of '0' (identification) denoting that an **attribute** information item is present. The second to fifth bits are '1111' denoting that the qualified name is a literal qualified name. The sixth bit is '0' (padding). The seventh bit is '1' denoting that the qualified name does has a **[prefix]** property. The eighth bit is '1' denoting that the qualified name has a **[namespace name]** property.	
14f	01110011 (73)		
....	....		
15c	01101110 (6e)		
15d	00001000 (08)		
15e	00111011 (3b)		
15f	01110101 (75)	The octet at position $14c_{16}$, value 84_{16}, has '1' for the first bit denoting that the **[prefix]** property is not a literal string and is indexed. The second bit is '0' denoting that the index is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the index is encoded in bits three to eight as an unsigned integer. The unsigned integer is 4_{10} and the index is 5_{10} (the lower bound is 1_{10}), which results in a **[prefix]** property of "xsi" when de-referenced from the NAMESPACE NAME table.	
....	....		
1a2	01100100 (64)		
		The octet at position $14d_{16}$, value 84_{16}, has '1' for the first bit denoting that the **[namespace name]** property is not a literal string and is indexed. The second bit is '0' denoting that the index is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the index is encoded in bits three to eight as an unsigned integer. The unsigned integer is 4_{10} and the index is	

Octet(s)		Description	XML infoset or XML
		5_{10} (the lower bound is 1_{10}), which results in a **[namespace name]** property of "….Schema-instance" when de-referenced from the NAMESPACE NAME table. The octet at position $14e_{16}$, value $0d_{16}$, has '0' for the first bit denoting that that a literal character string is encoded for the **[local name]** property. The second bit is '0' denoting that the length of encoded UTF-8 characters is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the length is encoded in bits three to eight as an unsigned integer. The unsigned integer is 13_{10} and the length is 14_{10} (the lower bound is 1_{10}). The 14_{10} octets of the encoded UTF-8 characters (of the **[local name]** property) are encoded from the octet at position $14f_{10}$ to the octet at position $15c_{10}$. The string "SchemaLocation" will be added to the LOCAL NAME table (with an index of 2_{10}). The qualified name with a **[prefix]** property of "xsi" (with index 5_{10}), a **[namespace name]** property of "….Schema-instance" (index 5_{10}), and a **[local name]** property of "schemaLocation" (index 2_{10}) will be added to the ATTRIBUTE NAME table (with an index of 1_{10}). The octet at position $15d_{16}$, value 08_{16}, is the initial encoding of a non identifying string or index for the **[normalized value]** property. The first bit is '0' denoting that a literal character string is present. The second bit is '0' denoting that the literal character string should not be added to ATTRIBUTE VALUE table. The third and fourth bits, both '0', denote that the encoding format of the string is UTF-8. The fifth and sixth bits are '1' and '0' respectively denoting that length of the octets of the UTF-8 characters (the **[normalized value]** property) is greater than or equal to 9_{10} octets and less than or equal to 264_{10} octets, and that the length, minus the lower bound, is encoded in eight bits on the next octet as an unsigned integer. The seventh to eighth bits are '0' (padding). The octet at position $15e_{16}$, value $3b_{16}$, is the encoding of the unsigned integer. The length of octets of the encoded UTF-8 characters is 68_{10} (the lower bound is 9_{10}). The 68_{10} octets of the encoded UTF-8 characters (of the **[normalized value]** property) are encoded from the octet at position $15f_{16}$ to the octet at position $1a2_{16}$.	
1a3	1111<u>0000</u> (f0)	The octet is the encoding of the terminator for the sequence of **attribute** information items. The octet at position $1a3_{16}$, value $f0_{16}$, has '1111' for the first four bits (the first to the fourth bit) and is the terminator for the sequence. Four '0' (padding) are present (the fifth to the eighth bit) since the **Order element** information item has children (see D.3.2).	

D.5.2.2 Encoding of the Address element information item of the BuyerParty element information item

The following explanation details the encoding of the **Address element** information item of the **BuyerParty element** information item of the fast infoset document. In particular, the encoding of **element** information items and **character** information items are explained. Table 12 presents the fragment of the fast infoset document for encoding of the **Address element** information item of the **BuyerParty element** information item of D.3.2. Table 13 details this encoding. The fragment in XML 1.0 is presented as follows:

```
<cac:Address>
        <cbc:StreetName>Marsh Lane</cbc:StreetName>
        <cbc:CityName>Nowhere</cbc:CityName>
        <cbc:PostalZone>NR18 4XX</cbc:PostalZone>
        <cbc:CountrySubentity>Norfolk</cbc:CountrySubentity>
</cac:Address>
```

Table 12 – Octets (as hexadecimal characters) of fragment

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
000200	3f828206416464726573
000220	733f818109537472656574e616d6582074d61727368204c616e65f03f818107
000240	436974794e616d6582044e6f7768657265f03f818109506f7374616c5a6f6e65
000260	82054e52313820345858f03f81810f436f756e747279537562656e7469747982
000280	044e6f72666f6c6bff

Table 13 – Encoding details

Octet(s)	Description	XML infoset or XML
216 00111111 (3f) 217 10000010 (82) 218 10000010 (82) 219 00000110 (06) 21a 01000001 (41) 220 01110011 (73)	The octets are the encoding of the **Address element** information item. The octet at position 216_{16}, value $3f_{16}$, has a '0' (identification) for the first bit denoting that there is a child of an **element** information item (child of the **Party element** information item), and the child is an **element** information item. The second bit is '0' denoting that the **element** information item does not have attributes. The third to fifth bits are '1111' denoting that the qualified name is a literal qualified name. The seventh bit is '1' denoting that the qualified name has a **[prefix]** property. The eighth bit is '1' denoting that the qualified name has a **[namespace name]** property. The octet at position 217_{16}, value 82_{16}, has '1' for the first bit denoting that the **[prefix]** property is not a literal string and is indexed. The second bit is '0' denoting that the index is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the index is encoded in bits three to eight as an unsigned integer. The unsigned integer is 2_{10} and the index is 3_{10} (the lower bound is 1_{10}), which results in a **[prefix]** property of "cac" when de-referenced from the PREFIX table. The octet at position 218_{16}, value 82_{16}, has '1' for the first bit denoting that the **[namespace name]** property is not a literal string and is indexed. The second bit is '0' denoting that the index is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the index is encoded in bits three to eight as an unsigned integer. The unsigned integer is 2_{10} and the index is 3_{10} (the lower bound is 1_{10}), which results in a **[namespace name]** property of "….ateComponents:1:0" when de-referenced from the NAMESPACE NAME table. The octet at position 219_{16}, value 06_{16}, has '0' for the first bit denoting that that a literal character string is encoded for the for the **[local name]** property. The second bit is '0' denoting that the length of encoded UTF-8 characters is greater than or equal to 1_{10} and less than or equal to 64_{10}, and the length is encoded in bits three to eight as an unsigned integer. The unsigned integer is 6_{10} and the length is 7_{10} (the lower bound is 1_{10}). The 7_{10} octets of the encoded UTF-8 characters (of the **[local name]** property) are encoded from the octet at position $21a_{16}$ to the octet at position 220_{16}. The string "Address" will be added to the LOCAL NAME table (with an index of 9_{10}).	**<cac:Address>**

Octet(s)		Description	XML infoset or XML
		The qualified name with a **[prefix]** property of "cac" (index 3_{10}), a **[namespace name]** property of "….ateComponents:1:0" (index 3_{10}), and a **[local name]** property of "Order" (index 1_{10}) will be added to the ELEMENT NAME table (with an index of 1_{10}).	
221	00111111 (3f)	The octets are the encoding of the **StreetName element** information item.	**<cbc:StreetName>**
222	10000001 (81)		
223	10000001 (81)	The **[local name]** property "StreetName" will be added to the LOCAL NAME table (with an index of 10_{10}).	
224	00001001 (09)		
225	01000001 (53)	The qualified name with a **[prefix]** property of "cbc" (index 2_{10}), a **[namespace name]** property of "….BasicComponents:1:0" (index 2_{10}), and a **[local name]** property of "StreetName" (index 10_{10}) will be added to the ELEMENT NAME table (with an index of 9_{10}).	
....	....		
22e	01100101 (65)		
22f	10000010 (82)	The octets are the encoding of the **character** information items of the **StreetName element** information item.	**character** information items "Marsh Lane"
230	00000111 (07)		
231	01001101 (4d)	The octet at position $22f_{16}$, value 82_{16}, has '10' (identification) for the first two bits (the first to second bit) denoting that there is a child of an **element** information item (child of the **StreetName element** information item), and the child is a chunk of **character** information items. The third bit is '0' denoting that a literal character string is present. The fourth bit is '0' denoting that the literal character string should not be added to CONTENT CHARACTER CHUNK table. The fifth and sixth bits, both '0', denote that the encoding format of the chunk is UTF-8. The seventh and eighth bits are '1' and '0' respectively denoting that length of the octets of the encoded UTF-8 characters (the chunk of **character** information items) is greater than or equal to 3_{10} octets and less than or equal to 258_{10} octets, and that the length, minus the lower bound, is encoded in eight bits on the next octet as an unsigned integer. The octet at position 230_{16}, value 07_{16}, is the unsigned integer. The length of octets of the encoded UTF-8 characters is 10_{10} (the lower bound is 3_{10}). The 10_{10} octets of the encoded UTF-8 characters are encoded from the octet at position 231_{16} to the octet at position $23a_{16}$.	
....			
23a	01100101 (65)		
23b	11110000 (f0)	The octet is the terminator for the **StreetName element** information item.	**</cbc:StreetName>**
23c	00111111 (3f)	The octets are the encoding of the **CityName element** information item.	**<cbc:CityName>**
23d	10000001 (81)		
23e	10000001 (81)	The **[local name]** property "CityName" will be added to the LOCAL NAME table (with an index of 10_{10}).	
23f	00000111 (07)		
240	01000011 (43)	The qualified name with a **[prefix]** property of "cbc" (index 2_{10}), a **[namespace name]** property of "….BasicComponents:1:0" (index 2_{10}), and a **[local name]** property of "CityName" (index 11_{10}) will be added to the ELEMENT NAME table (with an index of 10_{10}).	
....	....		
247	01100101 (65)		
248	10000010 (82)	The octets are the encoding of the **character** information items of the **CityName element** information item.	**character** information items "Nowhere"
249	00000100 (04)		
24a	01001110 (4e)	The 7_{10} octets of the encoded UTF-8 characters are encoded from the octet at position $24a_{16}$ to the octet at position 250_{16}.	
....			
250	01100101 (65)		
251	11110000 (f0)	The octet is the terminator for the **CityName element**	**</cbc:CityName>**

Octet(s)		Description	XML infoset or XML
		information item.	
252	00111111 (3f)	The octets are the encoding of the **PostalZone element** information item.	**<cbc:PostalZone>**
253	10000001 (81)		
254	10000001 (81)	The **[local name]** property "PostalZone" will be added to the LOCAL NAME table (with an index of 12_{10}).	
255	00001001 (09)		
256	01000011 (50)	The qualified name with a **[prefix]** property of "cbc" (index 2_{10}), a **[namespace name]** property of "….BasicComponents:1:0" (index 2_{10}), and a **[local name]** property of "PostalZone" (index 12_{10}) will be added to the ELEMENT NAME table (with an index of 11_{10}).	
....	....		
25f	01100101 (65)		
260	10000010 (82)	The octets are the encoding of the **character** information items of the **PostalZone element** information item.	**character** information items "NR18 4XX"
261	00000101 (05)		
262	01001110 (4e)	The 8_{10} octets of the encoded UTF-8 characters are encoded from the octet at position 262_{16} to the octet at position 269_{16}.	
....			
269	01011000 (58)		
26a	1111<u>0000</u> (f0)	The octet is the terminator for the **PostalZone element** information item.	**</cbc:PostalZone>**
26b	00111111 (3f)	The octets are the encoding of the **CountrySubentity element** information item.	**<cbc:CountrySubentity>**
26c	10000001 (81)		
26d	10000001 (81)	The **[local name]** property "CountrySubentity" will be added to the LOCAL NAME table (with an index of 13_{10}).	
26e	00001111 (0f)		
26f	01000011 (43)	The qualified name with a **[prefix]** property of "cbc" (index 2_{10}), a **[namespace name]** property of "….BasicComponents:1:0" (index 2_{10}), and a **[local name]** property of "CountrySubentity" (index 13_{10}) will be added to the ELEMENT NAME table (with an index of 12_{10}).	
....	....		
27e	01111001 (79)		
27f	10000010 (82)	The octets are the encoding of the **character** information items of the **CountrySubentity element** information item.	**character** information items "Norfolk"
280	00000100 (04)		
281	01001110 (4e)	The 7_{10} octets of the encoded UTF-8 characters are encoded from the octet at position 281_{16} to the octet at position 287_{16}.	
....			
287	01101011 (6b)		
288	11111111 (ff)	The octet is the terminator for the **CountrySubentity element** information item and the **Address element** information item.	**</cbc:CountrySubentity>** **</cac:Address>**
		The octet at position 288_{16}, value ff_{16}, has '1111' (terminator) for the first four bits (the first to the fourth bit) and is the terminator for the **CountrySubentity element** information item. The last four bits (the fifth to eighth bit) are '1111' and is the terminator for the **Address element** information item.	

D.5.2.3 Encoding of the BuyersID element information item of the first LineItem element information item

The following explanation details the encoding of the **BuyersID element** information item of the first **LineItem element** information item of the fast infoset document. In particular, the encoding of an **element** information item whose **[local name]** property has been indexed prior to this information item is explained. Table 14 presents the fragment of the fast infoset document for encoding of the **BuyersID element** information item of the first **LineItem element** information item of D.3.2. Table 15 details this encoding. The fragment in XML 1.0 is presented as follows:

```
<cac:LineItem>
    <cac:BuyersID>A</cac:BuyersID>
```

Table 14 – Octets (as hexadecimal characters) of fragment

	000102030405060708090a0b0c0d0e0f101112131415161718191a1b1c1d1e1f
0003c0	3f8282074c696e654974656d3f8282829041f0

Table 15 – Encoding details

	Octet(s)	Description	XML infoset or XML
3c4 3c5 3c6 3c7 3c8 3cf	00111111 (3f) 10000010 (82) 10000010 (82) 00001111 (07) 01001010 (4c) 01101101 (6d)	The octet is the encoding of the **LineItem element** information item. The **[local name]** property "LineItem" will be added to the LOCAL NAME table (with an index of 21_{10}). The qualified name with a **[prefix]** property of "cac" (index 3_{10}), a **[namespace name]** property of "….ateComponents:1:0" (index 3_{10}), and a **[local name]** property of "LineItem" (index 21_{10}) will be added to the ELEMENT NAME table (with an index of 20_{10}).	**<cac:LineItem>**
3d0 3d1 3d2 3d3	00111111 (3f) 10000010 (82) 10000010 (82) 10000010 (82)	The octet is the encoding of the **BuyersID element** information item. The **[prefix]** property, **[namespace name]** property and **[local name]** property have all been indexed as the associated strings have all occurred before this information item. The **[local name]** property was indexed by processing the first child of the **Order element** information item, namely the **BuyersID element** information item with the **[namespace name]** property "….Order:1:0". The qualified name with a **[prefix]** property of "cac" (index 3_{10}), a **[namespace name]** property of "….ateComponents:1:0" (index 3_{10}), and a **[local name]** property of "BuyersID" (index 3_{10}) will be added to the ELEMENT NAME table (with an index of 21_{10}).	**<cac:BuyersID>**
3d4 3d5	10010000 (90) 01000001 (41)	The octets are the encoding of the **character** information items of the **BuyersID element** information item. The octet at position $3d4_{16}$, value 90_{16}, has '10' (identification) for the first two bits (the first to second bit) denoting that there is a child of **element** information item (child of the **BuyersID element** information item), and the child is a chunk of **character** information items. The third bit is '0' denoting that a literal character string is present. The fourth bit is '1' denoting that the literal character string should be added to CONTENT CHARACTER CHUNK table (in this example strings less than 6_{10} characters are added to the CONTENT CHARACTER CHUNK table or the ATTRIBUTE VALUE table). The fifth and sixth bits, both '0', denote that the encoding format of the chunk is UTF-8. The seventh bit is '0' denoting that length of the octets of the encoded UTF-8 characters (the chunk of **character** information items) is greater than or equal to 1_{10} octet and less than or equal to 2_{10} octets, and that the length, minus the lower bound, is encoded in the eighth bit as an unsigned integer. The unsigned integer is 0_{10} and the length is 1_{10}. The octet of the encoded UTF-8 character is encoded by the octet at position 41_{16}.	**character** information item "A"
3d6	11110000 (f0)	The octet is the terminator for the **BuyersID element** information item.	**</cac:BuyersID>**