
**Information technology — Generic
applications of ASN.1: Fast Infoset**

*Technologies de l'information — Applications génériques de ASN.1:
Infoset rapide*

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2007

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

CONTENTS

	<i>Page</i>
1 Scope	1
2 Normative references	1
2.1 Identical Recommendations International Standards	2
2.2 Additional references	2
3 Definitions	3
3.1 ASN.1 terms	3
3.2 ECN terms	3
3.3 ISO/IEC 10646 terms	3
3.4 Additional definitions	3
4 Abbreviations	4
5 Notation	4
6 Principles of vocabulary table construction and use	5
7 ASN.1 type definitions	6
7.1 General	6
7.2 The Document type	6
7.3 The Element type	11
7.4 The Attribute type	12
7.5 The ProcessingInstruction type	12
7.6 The UnexpandedEntityReference type	13
7.7 The CharacterChunk type	13
7.8 The Comment type	14
7.9 The DocumentTypeDeclaration type	14
7.10 The UnparsedEntity type	15
7.11 The Notation type	15
7.12 The NamespaceAttribute type	16
7.13 The IdentifyingStringOrIndex type	16
7.14 The NonIdentifyingStringOrIndex type	17
7.15 The NameSurrogate type	18
7.16 The QualifiedNameOrIndex type	19
7.17 The EncodedCharacterString type	20
8 Construction and processing of a fast infoset document	21
8.1 Conceptual ordering of components of an abstract value of the Document type	22
8.2 The restricted alphabet table	22
8.3 The encoding algorithm table	22
8.4 The dynamic string tables	23
8.5 The dynamic name tables and name surrogates	23
9 Built-in restricted alphabets	24
9.1 The "numeric" restricted alphabet	24
9.2 The "date and time" restricted alphabet	24
10 Built-in encoding algorithms	24
10.1 General	24
10.2 The "hexadecimal" encoding algorithm	25
10.3 The "base64" encoding algorithm	25
10.4 The "short" encoding algorithm	25
10.5 The "int" encoding algorithm	26
10.6 The "long" encoding algorithm	26
10.7 The "boolean" encoding algorithm	26
10.8 The "float" encoding algorithm	27
10.9 The "double" encoding algorithm	27
10.10 The "uuid" encoding algorithm	27

	<i>Page</i>
10.11 The "cdata" encoding algorithm	28
11 Restrictions on the supported XML infosets and other simplifications.....	28
12 Bit-level encoding of the Document type.....	29
Annex A – ASN.1 module and ECN modules for fast infoset documents	31
A.1 ASN.1 module definition	31
A.2 ECN module definitions	33
Annex B – The MIME media type for fast infoset documents	53
Annex C – Description of the encoding of a fast infoset document.....	55
C.1 Fast infoset document	55
C.2 Encoding of the Document type	55
C.3 Encoding of the Element type	57
C.4 Encoding of the Attribute type	58
C.5 Encoding of the ProcessingInstruction type.....	58
C.6 Encoding of the UnexpandedEntityReference type.....	59
C.7 Encoding of the CharacterChunk type	59
C.8 Encoding of the Comment type	59
C.9 Encoding of the DocumentTypeDeclaration type.....	59
C.10 Encoding of the UnparsedEntity type	60
C.11 Encoding of the Notation type	60
C.12 Encoding of the NamespaceAttribute type.....	61
C.13 Encoding of the IdentifyingStringOrIndex type.....	61
C.14 Encoding of the NonIdentifyingStringOrIndex type starting on the first bit of an octet	61
C.15 Encoding of the NonIdentifyingStringOrIndex type starting on the third bit of an octet	62
C.16 Encoding of the NameSurrogate type	62
C.17 Encoding of the QualifiedNameOrIndex type starting on the second bit of an octet	62
C.18 Encoding of the QualifiedNameOrIndex type starting on the third bit of an octet	63
C.19 Encoding of the EncodedCharacterString type starting on the third bit of an octet	63
C.20 Encoding of the EncodedCharacterString type starting on the fifth bit of an octet.....	64
C.21 Encoding of the length of a sequence-of type.....	64
C.22 Encoding of the NonEmptyOctetString type starting on the second bit of an octet	64
C.23 Encoding of the NonEmptyOctetString starting on the fifth bit of an octet	65
C.24 Encoding of the NonEmptyOctetString type starting on the seventh bit of an octet	65
C.25 Encoding of integers in the range 1 to 2 ²⁰ starting on the second bit of an octet.....	65
C.26 Encoding of integers in the range 0 to 2 ²⁰ starting on the second bit of an octet.....	66
C.27 Encoding of integers in the range 1 to 2 ²⁰ starting on the third bit of an octet	66
C.28 Encoding of integers in the range 1 to 2 ²⁰ starting on the fourth bit of an octet.....	66
C.29 Encoding of integers in the range 1 to 256.....	67
Annex D – Examples of encoding XML infosets as fast infoset documents	68
D.1 Introduction of examples	68
D.2 Size of example documents (including redundancy-based compression).....	68
D.3 UBL order example	69
D.4 UBL Order fast infoset document with an external vocabulary	71
D.5 UBL order fast infoset document without an initial vocabulary	79
Annex E – Assignment of object identifier values.....	90
BIBLIOGRAPHY	91

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 24824-1 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 6, *Telecommunications and information exchange between systems*, in collaboration with ITU-T. The identical text is published as ITU-T Rec. X.891.

ISO/IEC 24824 consists of the following parts, under the general title *Information technology — Generic applications of ASN.1*:

— *Part 1: Fast infoset*

— *Part 2: Fast Web Services*

The following part is under preparation:

— *Part 3: Fast infoset security*

Introduction

This Recommendation | International Standard specifies a representation of an instance of the W3C XML Information Set using binary encodings (specified using the ASN.1 notation and the ASN.1 Encoding Control Notation). The encoding specified in this edition of this Recommendation | International Standard is identified by the version number 1 (see 12.9).

The technology specified in this Recommendation | International Standard is named Fast Infoset. It provides an alternative to W3C XML syntax as a means of representing instances of the W3C XML Information Set. This representation generally provides smaller encoding sizes and faster processing than a W3C XML representation.

The representation of an instance of the W3C XML Information Set specified in this Recommendation | International Standard is called a fast infoset document. Each fast infoset document is an encoding of an abstract value of an ASN.1 data type (the **Document** type – see 7.2) representing an instance of the W3C XML Information Set.

This Recommendation | International Standard specifies the use of several techniques that minimize the size of a fast infoset document and that maximize the speed of creating and processing such documents.

These techniques are based on the use of vocabulary tables, which allow typically-small integer values (vocabulary table indexes) to be used instead of character strings that form (for example) the names of elements or attributes in an XML 1.0 serialization of an instance of the W3C XML Information Set.

There are a number of vocabulary tables (see clause 8), of which the most basic (the eight character string tables) map typically-small integers to strings of characters. There are, however, also vocabulary tables (the element name table and the attribute name table) that provide a further level of indirection, with a vocabulary table index mapping to a set of three vocabulary table indexes, identifying a prefix, a namespace name, and a local name.

Another important technique is the use of a restricted alphabet vocabulary table. This contains entries that list a subset of ISO/IEC 10646 characters. If a character string needs to be encoded for which there is an entry in this table, then it can be encoded by identifying that this vocabulary table is being used, giving the vocabulary table index, and then encoding each character in the minimum number of bits needed for that particular subset of ISO/IEC 10646 characters. There are a number of built-in restricted alphabets that always form the first few entries of this table, covering such commonly occurring strings as dates and times, and numeric values.

A further important optimization uses the encoding algorithm vocabulary table. This table identifies specialized encodings that can be employed for commonly occurring strings, again with a number of built-in algorithms. For example, if there is a string which looks like the decimal representation of an integer in the range –32768 to 32767, then that string can be encoded by identifying that this vocabulary table is being used, giving the vocabulary table index, and then encoding the integer as a two-octet signed integer. Floating-point numbers and arrays of such numbers are supported in the same way.

In order to ensure fast processing without sacrificing compactness, many components of a fast infoset document (such as character strings and components representing information items of the XML infoset) are octet-aligned, while other components (such as lengths and vocabulary table indexes) are not necessarily octet-aligned but always end on the last bit of an octet. To provide a formal specification of these optimized encodings, the ASN.1 Encoding Control Notation (defined in ITU-T Rec. X.692 | ISO/IEC 8825-3) is used (see A.2), but use of ECN tools for implementation is not necessary and a complete description of the encoding is provided (see Annex C).

The vocabulary tables for a particular fast infoset document can be initialized by information at the head of the document, and are normally added to dynamically, providing flexibility for an encoder. The initial vocabulary tables can be provided by a reference to the set of final vocabulary tables of some other identified fast infoset document (or by other means). This vocabulary reference can then be supplemented by further table additions to provide the initial vocabulary tables for this document. Further dynamic additions are normally made to the tables during the creation or the processing of the document.

Finally, a mechanism is provided for the generator of a fast infoset document to include data (called additional processing data) related to optional additional processing of the fast infoset document, together with a URI that identifies a complete specification of the form and semantics of that additional processing data. The optional additional processing data is ignored by any subsequent processor of the fast infoset document if the URI is not known, or the processing that it specifies is not supported or not required.

NOTE – An example of such additional processing data would be data that provides indexes that enable immediate access to parts of the fast infoset document, so that the whole document need not be processed if the only interest is in those parts of the fast infoset document that correspond to a specific XML tag.

Annex A forms an integral part of this Recommendation | International Standard, and contains an ASN.1 module (see ITU-T Rec. X.680 | ISO/IEC 8824-1) and two ECN modules (EDM and ELM – see ITU-T Rec. X.692 | ISO/IEC 8825-3) which together specify the abstract content and the bit-level encoding of a value of the **Document** type, which conveys the value of an instance of the W3C XML Information Set.

Annex B forms an integral part of this Recommendation | International Standard, and contains the specification of a MIME media type identifying a fast infoset document.

Annex C does not form an integral part of this Recommendation | International Standard, and provides a complete description of the encodings formally specified in clause 12 and A.2.

Annex D does not form an integral part of this Recommendation | International Standard, and provides examples of fast infoset documents generated from some XML documents. Annex D also gives the size of the XML representation and the Fast Infoset representation of these examples.

**INTERNATIONAL STANDARD
ITU-T RECOMMENDATION****Information technology – Generic applications of ASN.1: Fast infoset****1 Scope**

This Recommendation | International Standard specifies an ASN.1 type (see ITU-T Rec. X.680 | ISO/IEC 8824-1) whose abstract values represent instances of the W3C XML Information Set. It also specifies binary encodings for those values, using ASN.1 Encoding Control Notation (see ITU-T Rec. X.692 | ISO/IEC 8825-3).

NOTE – These encodings are called fast infoset documents.

This Recommendation | International Standard also specifies techniques that:

- minimize the size of fast infoset documents;
- maximize the speed of creating and processing fast infoset documents;
- allow the specification (by the generator of a fast infoset document) of additional processing data.

The first two techniques involve the use of conceptual vocabulary tables. The set of vocabulary tables and the nature of their entries is fully defined in this Recommendation | International Standard, but their representation in computer memory is outside the scope of this Recommendation | International Standard. Provision for transfer or storage of, or a formal notation for displaying or specifying, vocabulary tables to be used as an external vocabulary is also outside the scope of this Recommendation | International Standard.

The third technique involves the provision of additional processing data and a URI that identifies the form and semantics of that data. The specification of specific forms of additional processing data and their use is outside the scope of this Recommendation | International Standard.

URIs can be used to identify final vocabularies that can be used as either part or all of some new initial vocabulary, but the assignment of specific URIs to specific final vocabularies is outside the scope of this Recommendation | International Standard.

This Recommendation | International Standard specifies built-in restricted alphabets, the addition to vocabulary tables of further restricted alphabets by enumeration, and the use of these vocabulary tables for efficient encoding of character strings.

This Recommendation | International Standard further specifies built-in encoding algorithms for the optimum encoding of certain character strings, and the addition to vocabulary tables of further encoding algorithms identified by URIs, but the definition of these further encoding algorithms and their associated URIs is outside the scope of this Recommendation | International Standard.

In addition, this Recommendation | International Standard specifies a Multipurpose Internet Mail Extensions (MIME) media type that identifies a fast infoset document.

2 Normative references

The following Recommendations, International Standards and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation | International Standard. At the time of publication, the editions indicated were valid. All Recommendations, International Standards and other references are subject to revision, and parties to agreements based on this Recommendation | International Standard are encouraged to investigate the possibility of applying the most recent edition of the Recommendations, International Standards and other references listed below. The Telecommunication Standardization Bureau of the ITU maintains a list of currently valid ITU-T Recommendations. Members of IEC and ISO maintain registers of currently valid International Standards. The IETF maintains a list of RFCs, together with those that have been obsoleted by later RFCs. The W3C maintains a list of currently valid W3C Recommendations. The reference to a document within this Recommendation | International Standard does not give it, as a stand-alone document, the status of a Recommendation or International Standard.

2.1 Identical Recommendations | International Standards

- ITU-T Recommendation X.667 (2004) | ISO/IEC 9834-8:2005, *Information technology – Open Systems Interconnection – Procedures for the operation of OSI Registration Authorities: Generation and registration of Universally Unique Identifiers (UUIDs) and their use as ASN.1 Object Identifier components*.
- ITU-T Recommendation X.680 (2002) | ISO/IEC 8824-1:2002, *Information technology – Abstract Syntax Notation One (ASN.1): Specification of basic notation*.
- ITU-T Recommendation X.681 (2002) | ISO/IEC 8824-2:2002, *Information technology – Abstract Syntax Notation One (ASN.1): Information object specification*. †
- ITU-T Recommendation X.682 (2002) | ISO/IEC 8824-3:2002, *Information technology – Abstract Syntax Notation One (ASN.1): Constraint specification*. †
- ITU-T Recommendation X.683 (2002) | ISO/IEC 8824-4:2002, *Information technology – Abstract Syntax Notation One (ASN.1): Parameterization of ASN.1 specifications*. †
- ITU-T Recommendation X.690 (2002) | ISO/IEC 8825-1:2002, *Information technology – ASN.1 encoding rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER)*. †
- ITU-T Recommendation X.691 (2002) | ISO/IEC 8825-2:2002, *Information technology – ASN.1 encoding rules: Specification of Packed Encoding Rules (PER)*. †
- ITU-T Recommendation X.692 (2002) | ISO/IEC 8825-3:2002, *Information technology – ASN.1 encoding rules: Specification of Encoding Control Notation (ECN)*.
- ITU-T Recommendation X.693 (2001) | ISO/IEC 8825-4:2002, *Information technology – ASN.1 encoding rules: XML Encoding Rules (XER)*. †

NOTE – The complete set of ASN.1 Recommendations | International Standards are listed above, as they can all be applicable in particular uses of this Recommendation | International Standard. Where these are not directly referenced in the body of this Recommendation | International Standard, a † symbol is added to the reference.

2.2 Additional references

- ISO 8601:2004, *Data elements and interchange formats – Information interchange – Representation of dates and times*.
- ISO/IEC 10646:2003, *Information technology – Universal Multiple-Octet Coded Character Set (UCS)*.
- *The Unicode Standard, Version 4.0*, The Unicode Consortium (Reading, MA, Addison-Wesley).
NOTE 1 – The graphics characters (and their encodings) defined by Unicode are identical to those defined by ISO/IEC 10646-1, but Unicode is included as a reference because it also specifies the names of control characters and defines the abbreviation UTF-16BE.
- W3C XML 1.0:2004, *Extensible Markup Language (XML) 1.0 (Third Edition)*, W3C Recommendation, Copyright © [4 February 2004] World Wide Web Consortium (Massachusetts Institute of Technology, Institut National de Recherche en Informatique et en Automatique, Keio University), <http://www.w3.org/TR/2000/REC-xml-20040204/>.
- W3C XML 1.1:2004, *Extensible Markup Language (XML) 1.1*, W3C Recommendation, Copyright © [4 February 2004] World Wide Web Consortium (Massachusetts Institute of Technology, Institut National de Recherche en Informatique et en Automatique, Keio University), <http://www.w3.org/TR/2000/REC-xml11-20040204/>.
NOTE 2 – References to both W3C XML 1.0 and W3C XML 1.1 are included as neither is a subset of the other. These references are used solely in 3.4.10.
- W3C XML Information Set:2004, *XML Information Set (Second Edition)*, W3C Recommendation, Copyright © [04 February 2004] World Wide Web Consortium (Massachusetts Institute of Technology, Institut National de Recherche en Informatique et en Automatique, Keio University), <http://www.w3.org/TR/2004/REC-xml-info-20040204/>.
- W3C XML Namespaces 1.0:1999, *Namespaces in XML*, W3C Recommendation, Copyright © [14 January 1999] World Wide Web Consortium (Massachusetts Institute of Technology, Institut National de Recherche en Informatique et en Automatique, Keio University), <http://www.w3.org/TR/1999/REC-xml-names-19990114/>.
- W3C XML Namespaces 1.1:2004, *Namespaces in XML 1.1*, W3C Recommendation, Copyright © [4 February 2004] World Wide Web Consortium (Massachusetts Institute of Technology, Institut National de Recherche en Informatique et en Automatique, Keio University), <http://www.w3.org/TR/2004/REC-xml-names11-20040204/>.

NOTE 3 – References to both W3C XML Namespaces 1.0 and W3C XML Namespaces 1.1 are included as neither is a subset of the other. These references are used solely in 3.4.10.

- IETF RFC 2045 (1996), *Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies*.
- IETF RFC 2396 (1998), *Uniform Resource Identifiers (URI): Generic Syntax*.
- IEEE 754-1985, *IEEE Standard for Binary Floating-Point Arithmetic*.