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

Information technology — Generic applications of ASN.1: Fast infoset

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO/IEC 24824-1:2007](https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9cede3/iso-iec-24824-1-2007)

<https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9cede3/iso-iec-24824-1-2007>

PDF – Exonération de responsabilité

Le présent fichier PDF peut contenir des polices de caractères intégrées. Conformément aux conditions de licence d'Adobe, ce fichier peut être imprimé ou visualisé, mais ne doit pas être modifié à moins que l'ordinateur employé à cet effet ne bénéficie d'une licence autorisant l'utilisation de ces polices et que celles-ci y soient installées. Lors du téléchargement de ce fichier, les parties concernées acceptent de fait la responsabilité de ne pas enfreindre les conditions de licence d'Adobe. Le Secrétariat central de l'ISO décline toute responsabilité en la matière.

Adobe est une marque déposée d'Adobe Systems Incorporated.

Les détails relatifs aux produits logiciels utilisés pour la création du présent fichier PDF sont disponibles dans la rubrique General Info du fichier; les paramètres de création PDF ont été optimisés pour l'impression. Toutes les mesures ont été prises pour garantir l'exploitation de ce fichier par les comités membres de l'ISO. Dans le cas peu probable où surviendrait un problème d'utilisation, veuillez en informer le Secrétariat central à l'adresse donnée ci-dessous.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO/IEC 24824-1:2007](https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9ccede3/iso-iec-24824-1-2007)

<https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9ccede3/iso-iec-24824-1-2007>



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO/CEI 2007

Droits de reproduction réservés. Sauf prescription différente, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'ISO à l'adresse ci-après ou du comité membre de l'ISO dans le pays du demandeur.

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

Publié en Suisse

TABLE DES MATIÈRES

	<i>Page</i>	
1	Domaine d'application	1
2	Références normatives	1
	2.1 Recommandations Normes internationales identiques	2
	2.2 Références supplémentaires	2
3	Définitions	3
	3.1 Termes ASN.1	3
	3.2 Termes ECN	3
	3.3 Termes de l'ISO/CEI 10646	3
	3.4 Définitions supplémentaires	3
4	Abréviations	4
5	Notation	5
6	Principes de construction et d'utilisation des tableaux de vocabulaire	5
7	Définitions de type ASN.1	6
	7.1 Généralités	6
	7.2 Le type Document	6
	7.3 Le type Element	11
	7.4 Le type Attribute	12
	7.5 Le type ProcessingInstruction	13
	7.6 Le type UnexpandedEntityReference	13
	7.7 Le type CharacterChunk	14
	7.8 Le type Comment	14
	7.9 Le type DocumentTypeDeclaration	15
	7.10 Le type UnparsedEntity	15
	7.11 Le type Notation	16
	7.12 Le type NamespaceAttribute	16
	7.13 Le type IdentifyingStringOrIndex	17
	7.14 Le type NonIdentifyingStringOrIndex	18
	7.15 Le type NameSurrogate	19
	7.16 Le type QualifiedNameOrIndex	19
	7.17 Le type EncodedCharacterString	21
8	Construction et traitement d'un document Fast Infoset	22
	8.1 Rangement conceptuel des composants d'une valeur abstraite du type Document	22
	8.2 Tableau d'alphabet restreint	23
	8.3 Tableau d'algorithme de codage	23
	8.4 Tableaux de chaîne dynamiques	24
	8.5 Les tableaux de nom dynamiques et les substituts de nom	24
9	Alphabets restreints prédéfinis	25
	9.1 L'alphabet restreint "numérique"	25
	9.2 L'alphabet restreint "date et heure"	25
10	Algorithmes de codage prédéfinis	26
	10.1 Généralités	26
	10.2 L'algorithme de codage "hexadécimal"	26
	10.3 L'algorithme de codage "base64"	26
	10.4 L'algorithme de codage "court"	26
	10.5 L'algorithme de codage "int"	27
	10.6 L'algorithme de codage "long"	27
	10.7 L'algorithme de codage "booléen"	28
	10.8 L'algorithme de codage "float"	28
	10.9 L'algorithme de codage "double"	28
	10.10 L'algorithme de codage "uuid"	29

	<i>Page</i>
10.11 L'algorithme de codage "cdata"	29
11 Restrictions sur les infosets XML pris en charge et autres simplifications	29
12 Codage de niveau binaire du type Document	30
Annexe A – Module ASN.1 et modules ECN pour documents Fast Infoset.....	32
A.1 Définition du module ASN.1	32
A.2 Définitions de modules ECN	34
Annexe B – Type de support MIME pour documents Fast Infoset.....	54
Annexe C – Description du codage d'un document Fast Infoset	56
C.1 Document Fast Infoset	56
C.2 Codage du type Document	56
C.3 Codage du type Element	58
C.4 Codage du type Attribute	59
C.5 Codage du type ProcessingInstruction	59
C.6 Codage du type UnexpandedEntityReference	60
C.7 Codage du type CharacterChunk	60
C.8 Codage du type Comment	60
C.9 Codage du type DocumentTypeDeclaration	60
C.10 Codage du type UnparsedEntity	61
C.11 Codage du type Notation	61
C.12 Codage du type NamespaceAttribute	62
C.13 Codage du type IdentifyingStringOrIndex	62
C.14 Codage du type NonIdentifyingStringOrIndex commençant sur le premier bit d'un octet	62
C.15 Codage du type NonIdentifyingStringOrIndex commençant au troisième bit d'un octet.....	63
C.16 Codage du type NameSurrogate	63
C.17 Codage du type QualifiedNameOrIndex commençant sur le second bit d'un octet	63
C.18 Codage du type QualifiedNameOrIndex commençant sur le troisième bit d'un octet	64
C.19 Codage du type EncodedCharacterString commençant sur le troisième bit d'un octet	64
C.20 Codage du type EncodedCharacterString commençant sur le cinquième bit d'un octet.....	65
C.21 Codage de la longueur d'un type sequence-of.....	65
C.22 Codage du type NonEmptyOctetString commençant sur le second bit d'un octet	65
C.23 Codage de la chaîne NonEmptyOctetString commençant sur le cinquième bit d'un octet.....	66
C.24 Codage du type NonEmptyOctetString commençant sur le septième bit d'un octet.....	66
C.25 Codage des entiers dans la gamme 1 à 2 ²⁰ commençant sur le second bit d'un octet	66
C.26 Codage des entiers dans la gamme 0 à 2 ²⁰ commençant sur le second bit d'un octet	67
C.27 Codage des entiers dans la gamme 1 à 2 ²⁰ commençant sur le troisième bit d'un octet	67
C.28 Codage des entiers dans la gamme 1 à 2 ²⁰ commençant sur le quatrième bit d'un octet	67
C.29 Codage des entiers dans la gamme 1 à 256	68
Annexe D – Exemples de codage d'infosets XML comme documents Fast Infoset	69
D.1 Introduction des exemples.....	69
D.2 Taille des documents exemples (y compris la compression fondée sur la redondance).....	69
D.3 Exemple d'ordre UBL	70
D.4 Document Fast Infoset en ordre UBL avec vocabulaire externe	72
D.5 Document Fast Infoset en ordre UBL sans vocabulaire initial	79
Annexe E – Allocation des valeurs d'identifiant d'objet.....	90
BIBLIOGRAPHIE	91

Avant-propos

L'ISO (Organisation internationale de normalisation) et la CEI (Commission électrotechnique internationale) forment le système spécialisé de la normalisation mondiale. Les organismes nationaux membres de l'ISO ou de la CEI participent au développement de Normes internationales par l'intermédiaire des comités techniques créés par l'organisation concernée afin de s'occuper des domaines particuliers de l'activité technique. Les comités techniques de l'ISO et de la CEI collaborent dans des domaines d'intérêt commun. D'autres organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO et la CEI participent également aux travaux. Dans le domaine des technologies de l'information, l'ISO et la CEI ont créé un comité technique mixte, l'ISO/CEI JTC 1.

Les Normes internationales sont rédigées conformément aux règles données dans les Directives ISO/CEI, Partie 2.

La tâche principale du comité technique mixte est d'élaborer les Normes internationales. Les projets de Normes internationales adoptés par le comité technique mixte sont soumis aux organismes nationaux pour vote. Leur publication comme Normes internationales requiert l'approbation de 75 % au moins des organismes nationaux votants.

L'attention est appelée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO et la CEI ne sauraient être tenues pour responsables de ne pas avoir identifié de tels droits de propriété et averti de leur existence.

L'ISO/CEI 24824-1 a été élaboré par le comité technique mixte ISO/CEI JTC 1, *Technologies de l'information*, sous-comité SC 6, *Téléinformatique*, en collaboration avec l'UIT-T. Le texte identique est publié en tant que Rec. UIT-T X.891.

<https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b59ceda3/iso-iec-24824-1-2007>

L'ISO/CEI 24824-1 comprend les parties suivantes, présentées sous le titre général *Technologies de l'information — Applications génériques de l'ASN.1*:

— *Partie 1: Infoset rapide*

— *Partie 2: Services web rapides*

La partie suivante est en cours d'élaboration:

— *Partie 3: Sécurité d'Infoset rapide*

Introduction

La présente Recommandation | Norme internationale spécifie une représentation d'une instance de l'ensemble d'informations XML W3C qui utilise des codages binaires (spécifiés à l'aide de la notation ASN.1 et de la notation de contrôle de codage ASN.1). Le codage spécifié dans l'édition de la présente Recommandation | Norme internationale est identifié par le numéro de version 1 (voir le § 12.9).

La technologie spécifiée dans la présente Recommandation | Norme internationale se nomme Fast Infoset. Elle fournit une alternative à la syntaxe XML W3C comme moyen de représenter les instances d'ensemble d'informations XML du W3C. Cette représentation donne habituellement des tailles de codage plus petites et un traitement plus rapide que la représentation XML du W3C.

La représentation d'une instance de l'ensemble d'informations XML W3C spécifiée dans la présente Recommandation | Norme internationale s'appelle un document Fast Infoset. Chaque document Fast Infoset est un codage d'une valeur abstraite d'un type de données ASN.1 (le type `Document` – voir le § 7.2) représentant une instance de l'ensemble d'informations XML W3C.

La présente Recommandation | Norme internationale spécifie l'utilisation de plusieurs techniques qui minimisent la taille d'un document Fast Infoset et qui maximisent la vitesse de création et de traitement de tels documents.

Ces techniques sont fondées sur l'emploi de tableaux de vocabulaire, qui permettent d'utiliser des valeurs entières normalement faibles (les indices de tableaux de vocabulaire) à la place des chaînes de caractères qui forment (par exemple) les noms des éléments ou attributs dans une mise en série XML 1.0 d'une instance de l'ensemble d'informations XML W3C.

Il y a de nombreux tableaux de vocabulaire (voir le paragraphe 8), dont les plus fondamentaux (les tableaux de chaînes de huit caractères) transposent des entiers normalement petits en chaînes de caractères. Il y a cependant aussi des tableaux de vocabulaire (le tableau des noms d'éléments et le tableau des noms d'attributs) qui donnent un niveau d'indication supplémentaire, avec un indice de tableau de vocabulaire transposant un ensemble de trois indices de tableaux de vocabulaire, identifiant un préfixe, un nom d'espace de nom, et un nom local.

Une autre technique importante est l'utilisation d'un tableau de vocabulaire à alphabet restreint. Il contient des entrées qui font la liste d'un sous-ensemble de caractères ISO/CEI 10646. Si une chaîne de caractères qui doit être codée possède une entrée dans ce tableau, elle peut alors être codée en notant qu'on utilise ce tableau de vocabulaire, en donnant l'indice du tableau de vocabulaire, et en codant alors chaque caractère avec le nombre minimal de bits nécessaires pour ce sous-ensemble particulier de caractères ISO/CEI 10646. Il y a un certain nombre d'alphabets restreints préconstruits qui forment toujours les premières entrées de ce tableau, couvrant les chaînes qu'on rencontre de façon usuelle comme les dates, les heures et les valeurs numériques.

Une importante technique d'optimisation utilise le tableau de vocabulaire d'algorithme de codage. Ce tableau identifie les codages spécialisés qui peuvent être employés pour les chaînes d'utilisation courante, toujours avec un certain nombre d'algorithmes incorporés. Par exemple, s'il y a une chaîne qui ressemble à la représentation décimale d'un entier dans la gamme de -32 768 à 32 767, cette chaîne peut alors être codée en précisant qu'on utilise ce tableau de vocabulaire, en donnant l'indice du tableau de vocabulaire, puis en codant l'entier comme entier algébrique de deux octets. Les nombres et matrices à virgule flottante, de tels nombres sont pris en charge de la même façon.

Pour assurer un traitement rapide sans sacrifier la concision, de nombreux composants d'un document Fast Infoset (comme les chaînes de caractères et les composants représentant des éléments d'information de l'infoset XML) sont calés à l'octet, alors que d'autres composants (comme les longueurs et les indices de tableaux de vocabulaire) ne sont pas nécessairement calés à l'octet mais se terminent toujours sur le dernier bit d'un octet. La notation de contrôle de codage ASN.1 (définie dans la Rec. UIT-T X.692 | ISO/CEI 8825-3) est utilisée pour fournir une spécification formelle de ces codages optimisés (voir le § A.2), mais l'utilisation d'outils ECN pour les implémentations n'est pas nécessaire et une description complète du codage est donnée à l'Annexe C.

Les tableaux de vocabulaire pour un document Fast Infoset particulier peuvent être initialisés par des informations en tête du document, et y sont habituellement ajoutés de façon dynamique, pour donner de la souplesse au codage. Les tableaux de vocabulaire initiaux peuvent être fournis par une référence à l'ensemble de tableaux de vocabulaire finaux d'autres documents Fast Infoset identifiés (ou par d'autres moyens). Cette référence de vocabulaire peut alors être complétée par d'autres ajouts de tableaux pour fournir les tableaux de vocabulaire initiaux pour ce document. D'autres ajouts dynamiques ultérieurs sont normalement faits aux tableaux durant la création ou le traitement du document.

Finalement, il est fourni un mécanisme de génération d'un document Fast Infoset qui inclut des données (appelées données de traitement supplémentaire) qui se rapportent au traitement supplémentaire facultatif du document Fast Infoset, lié à un URI qui identifie une spécification complète de la forme et de la sémantique des données de traitement supplémentaire. Les données de traitement supplémentaire facultatif sont ignorées par les traitements ultérieurs du document Fast Infoset si l'URI n'est pas connu, ou si le traitement qu'elles spécifient n'est pas accepté ou pas utile.

NOTE – Un exemple de telles données de traitement supplémentaire serait celui de données fournissant des indices permettant un accès immédiat à des parties du document Fast Infoset, de sorte que le document tout entier n'ait pas besoin d'être traité si son seul intérêt réside dans les parties du document Fast Infoset qui correspondent à une étiquette XML spécifique.

L'Annexe A fait partie intégrante de la présente Recommandation | Norme internationale, et contient un module ASN.1 (voir la Rec. UIT-T X.680 | ISO/CEI 8824-1) et deux modules ECN (EDM et ELM – voir la Rec. UIT-T X.692 | ISO/CEI 8825-3) qui spécifient conjointement le contenu abstrait et le codage de niveau binaire d'une valeur du type **Document**, qui porte la valeur d'une instance de l'ensemble d'informations XML du W3C.

L'Annexe B fait partie intégrante de la présente Recommandation | Norme internationale, et contient la spécification d'un type de support MIME qui identifie un document Fast Infoset.

L'Annexe C ne fait pas partie intégrante de la présente Recommandation | Norme internationale, et donne une description complète des codages spécifiés formellement au § 12 et au § A.2.

L'Annexe D ne fait pas partie intégrante de la présente Recommandation | Norme internationale, et donne des exemples de documents Fast Infoset générés à partir de documents XML. L'annexe D donne aussi la taille de la représentation XML et de la représentation Fast Infoset de ces exemples.

iTeh STANDARD PREVIEW (standards.iteh.ai)

[ISO/IEC 24824-1:2007](https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9cede3/iso-iec-24824-1-2007)

<https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9cede3/iso-iec-24824-1-2007>

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO/IEC 24824-1:2007](https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9cede3/iso-iec-24824-1-2007)

<https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9cede3/iso-iec-24824-1-2007>

**NORME INTERNATIONALE
RECOMMANDATION UIT-T**

Technologies de l'information – Applications génériques de l'ASN.1: Fast Infoset

1 Domaine d'application

La présente Recommandation | Norme internationale spécifie un type d'ASN.1 (voir la Rec. UIT-T X.680 | ISO/CEI 8824-1) dont les valeurs abstraites représentent des instances de l'ensemble d'informations XML du W3C. Elle spécifie aussi les codages binaires pour ces valeurs, en utilisant la notation de contrôle de codage ASN.1 (voir la Rec. UIT-T X.692 | ISO/CEI 8825-3).

NOTE – Ces codages sont appelés documents Fast Infoset.

La présente Recommandation | Norme internationale spécifie aussi des techniques qui:

- minimisent la taille des documents Fast Infoset;
- maximisent la vitesse de création et de traitement des documents Fast Infoset;
- permettent la spécification (grâce au générateur d'un document Fast Infoset) de données de traitement supplémentaire.

Les deux premières techniques impliquent l'utilisation de tableaux de vocabulaire conceptuel. L'ensemble des tableaux de vocabulaire et la nature de leurs entrées est défini de façon complète dans la présente Recommandation | Norme internationale, mais leur représentation dans les mémoires d'ordinateurs est en dehors du domaine d'application de la présente Recommandation | Norme internationale. Les dispositions pour le transfert ou le stockage, ou la notation formelle de l'affichage ou la spécification des tableaux de vocabulaire à utiliser comme vocabulaire externe sont également en dehors du domaine d'application de la présente Recommandation | Norme internationale.

La troisième technique implique la fourniture de données de traitement supplémentaire et d'un URI qui identifie la forme et la sémantique de ces données. La spécification de formes particulières de données de traitement supplémentaire et leur utilisation sont en dehors du domaine d'application de la présente Recommandation | Norme internationale.

Les URI peuvent être utilisés pour identifier les vocabulaires finaux qui peuvent être utilisés comme tout ou partie d'un nouveau vocabulaire initial, mais l'affectation d'URI spécifiques à des vocabulaires finaux particuliers est en dehors du domaine d'application de la présente Recommandation | Norme internationale.

La présente Recommandation | Norme internationale spécifie des alphabets restreints prédéfinis, l'ajout par énumération à des tableaux de vocabulaire d'autres alphabets restreints, et l'utilisation de ces tableaux de vocabulaire pour un codage efficace des chaînes de caractères.

La présente Recommandation | Norme internationale spécifie aussi des algorithmes de codage prédéfinis pour le codage optimal de certaines chaînes de caractères et l'ajout à des tableaux de vocabulaire d'autres algorithmes de codage identifiés par des URI, mais la définition de ces autres algorithmes de codage et de leurs URI associés est en dehors du domaine d'application de la présente Recommandation | Norme internationale.

De plus, la présente Recommandation | Norme internationale spécifie un type de support d'extensions de messagerie Internet multiobjets (MIME, *multipurpose Internet mail extensions*) qui identifie un document Fast Infoset.

2 Références normatives

Les Recommandations | Normes internationales et autres références suivantes contiennent des dispositions qui, par suite de la référence qui y est faite, constituent des dispositions valables pour la présente Recommandation | Norme internationale. Au moment de la publication, les éditions indiquées étaient en vigueur. Toutes Recommandations, Normes internationales et autres références sont sujettes à révision et les parties prenantes aux accords fondés sur la présente Recommandation | Norme internationale sont invitées à rechercher la possibilité d'appliquer les éditions les plus récentes des Recommandations | Normes internationales et autres références indiquées ci-après. Le Bureau de la normalisation des télécommunications de l'UIT tient à jour une liste des Recommandations de l'UIT-T en vigueur. Les membres de la CEI et de l'ISO possèdent le registre des Normes internationales en vigueur. L'IETF tient à jour une liste des RFC, ainsi que de celles qui ont été rendues obsolètes par des RFC plus récents. Le W3C tient à jour la liste des

spécifications du W3C en vigueur. La référence à un document figurant dans la présente Recommandation | Norme internationale ne donne pas à ce document, en tant que tel, le statut d'une Recommandation ou Norme internationale.

2.1 Recommandations | Normes internationales identiques

- Recommandation UIT-T X.667 (2004) | ISO/CEI 9834-8:2005, *Technologies de l'information – Interconnexion des systèmes ouverts – Procédures opérationnelles des organismes d'enregistrement de l'OSI: génération et enregistrement des identificateurs uniques universels (UUID) et utilisation de ces identificateurs comme composants d'identificateurs d'objet ASN.1.*
- Recommandation UIT-T X.680 (2002) | ISO/CEI 8824-1:2002, *Technologies de l'information – Notation de syntaxe abstraite numéro un: spécification de la notation de base.*
- Recommandation UIT-T X.681 (2002) | ISO/CEI 8824-2:2002, *Technologies de l'information – Notation de syntaxe abstraite numéro un: spécification des objets informationnels.* †
- Recommandation UIT-T X.682 (2002) | ISO/CEI 8824-3:2002, *Technologies de l'information – Notation de syntaxe abstraite numéro un: spécification des contraintes.* †
- Recommandation UIT-T X.683 (2002) | ISO/CEI 8824-4:2002, *Technologies de l'information – Notation de syntaxe abstraite numéro un: paramétrage des spécifications de la notation de syntaxe abstraite numéro un.* †
- Recommandation UIT-T X.690 (2002) | ISO/CEI 8825-1:2002, *Technologies de l'information – Règles de codage ASN.1: spécification des règles de codage de base, des règles de codage canoniques et des règles de codage distinctives.* †
- Recommandation UIT-T X.691 (2002) | ISO/CEI 8825-2:2002, *Technologies de l'information – Règles de codage ASN.1: spécification des règles de codage compact (PER).* †
- Recommandation UIT-T X.692 (2002) | ISO/CEI 8825-3:2002, *Technologies de l'information – Règles de codage ASN.1: notation de contrôle de codage (ECN).*
- Recommandation UIT-T X.693 (2001) | ISO/CEI 8825-4:2002, *Technologies de l'information – Règles de codage ASN.1: règles de codage XML (XER).* †

NOTE – L'ensemble complet des Recommandations | Normes internationales sur l'ASN.1 est énuméré ci-dessus, car toutes peuvent être applicables dans des utilisations particulières de la présente Recommandation | Norme internationale. Lorsqu'elles ne sont pas directement référencées dans le corps de la présente Recommandation | Norme internationale, un symbole † est ajouté à la référence.

<https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9ccde3/iso-iec-24824-1-2007>

2.2 Références supplémentaires

- ISO 8601:2004, *Eléments de données et formats d'échange – Echange d'informations – Représentation de la date et de l'heure.*
- ISO/CEI 10646:2003, *Technologies de l'information – Ensemble universel de caractères codés sur plusieurs octets.*
- Unicode, *The Unicode Standard, Version 4.0, The Unicode Consortium. (Reading, MA, Addison-Wesley).* (Norme Unicode, version 4.0)
 - NOTE 1 – Les caractères graphiques (et leurs codages) définis par Unicode sont identiques à ceux définis par la norme ISO/CEI 10646-1, mais Unicode est inclus comme référence parce qu'il spécifie aussi les noms des caractères de contrôle et définit l'abréviation 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/>.* (Langage de balisage extensible, troisième édition)
- 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/>.* (Langage de balisage extensible)
 - NOTE 2 – Les références à W3C XML 1.0 et W3C XML 1.1 sont toutes deux incluses car aucune n'est un sous-ensemble de l'autre. Ces références ne sont utilisées qu'au § 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-infoset-20040204/>.* (Ensemble d'informations XML, deuxième édition)

- 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/>. (Espaces de nom en XML)
- 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/>. (Espaces de nom en XML)
 - NOTE 3 – Les références à W3C XML Namespaces 1.0 et W3C XML Namespaces 1.1 sont toutes deux incluses car aucune n'est un sous-ensemble de l'autre. Ces références ne sont utilisées qu'au § 3.4.10.
- IETF RFC 2045 (1996), *Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies*. (*Extensions de messagerie Internet multiobjets, MIME*)
- IETF RFC 2396 (1998), *Uniform Resource Identifiers (URI): Generic Syntax*. (*Identificateurs universels de ressources (URI): Syntaxe générique*)
- IEEE 754 (1985), *IEEE Standard for Binary Floating-Point Arithmetic*. (*Norme IEEE pour arithmétique binaire à virgule flottante*)

3 Définitions

Pour les besoins de la présente Recommandation | Norme internationale, on applique les définitions suivantes.

3.1 Termes ASN.1

La présente Recommandation | Norme internationale utilise les termes suivants, définis dans la Rec UIT-T X.680 | ISO/CEI 8824-1:

- a) type choix (*choice type*);
- b) type séquence (*sequence type*);
- c) type séquence de (*sequence-of type*).

3.2 Termes ECN <https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9ced3/iso-iec-24824-1-2007>

La présente Recommandation | Norme internationale utilise les termes suivants définis dans la Rec. UIT-T X.692 | ISO/CEI 8825-3:

- a) module de définition de codage (EDM, *encoding definition module*);
- b) module de lien de codage (ELM, *encoding link module*).

3.3 Termes de l'ISO/CEI 10646

La présente Recommandation | Norme internationale utilise les termes suivants définis dans l'ISO/CEI 10646:

- a) plan multilingue de base.

3.4 Définitions supplémentaires

3.4.1 Base64: mécanisme de codage qui représente une valeur de chaîne d'octet comme une chaîne de caractères en utilisant un alphabet restreint de 65 caractères (voir le § 10.3 et la RFC 2045 de l'IETF).

3.4.2 chaîne de caractères: chaîne de caractères abstraits ISO/CEI 10646, sans aucune implication sur la façon dont ils sont codés.

3.4.3 algorithme de codage: spécification précise de la façon de coder efficacement une chaîne de caractères avec des caractéristiques spécifiées dans les octets.

NOTE – Un exemple serait le codage d'une chaîne telle que "-32176" en entier binaire complément à deux sur deux octets. Le codage sur deux octets serait accompagné d'un indice de tableau de vocabulaire identifiant cet algorithme de codage.

3.4.4 vocabulaire externe: ensemble de tableaux de vocabulaire référencé par un URI (voir au § 7.2.14).

3.4.5 document Fast Infoset: Infoset XML représenté comme spécifié dans la présente Recommandation | Norme internationale.

3.4.6 vocabulaire final: contenu des tableaux de vocabulaire à la fin de la création ou du traitement d'un document Fast Infoset.

3.4.7 élément informationnel: chacun des types d'éléments qui constitue un infoset XML.

3.4.8 vocabulaire initial: ensemble des tableaux de vocabulaire établis par information au début d'un document Fast Infoset qui fait facultativement référence à un vocabulaire externe et fournit facultativement des entrées de tableau supplémentaires.

3.4.9 suppléant de nom: ensemble de trois indices de tableau de vocabulaire (les deux premiers sont facultatifs) qui sont utilisés pour représenter un nom qualifié (voir au § 3.4.11).

3.4.10 document XML bien conformé à l'espace des noms: document XML W3C 1.0 bien conformé selon l'espace de noms XML 1.0 de la W3C ou document XML 1.1 W3C bien conformé selon l'espace de noms XML 1.1 de la W3C.

3.4.11 nom qualifié: ensemble comportant les propriétés **[prefix]**, **[namespace name]**, et **[local name]** d'un élément informationnel **element** ou d'un élément informationnel **attribute**.

3.4.12 alphabet restreint: ensemble ordonné de caractères ISO/CEI 10646 distincts, qui permet un codage compact de toute chaîne de caractères uniquement constituée de caractères pris dans cet ensemble.

3.4.13 indice de tableau de vocabulaire: valeur entière positive identifiant une entrée dans un tableau de vocabulaire.

3.4.14 tableaux de vocabulaire: ensemble de tableaux de concepts (normalement, mais pas nécessairement, construits de façon dynamique) associés à un document Fast Infoset, qui contient des chaînes de caractères ou d'autres informations, et prend en charge l'utilisation de valeurs entières positives normalement petites (indices de tableau de vocabulaire) qui identifient leurs entrées.

NOTE – Des exemples de tableaux de vocabulaire sont ceux qui contiennent des chaînes de caractères qui sont les propriétés **[local name]** des éléments d'information **attribute** ou **element**, ou des chaînes de caractère correspondant à des séquences d'éléments d'information **character** qui sont membres de la propriété **[children]** des éléments d'information **element**.

3.4.15 déclaration XML: codage UTF-8 d'une chaîne de caractère spécifiée (voir aussi le § 12.3) qui peut être incluse au début d'un document Fast Infoset pour identifier le codage comme celui d'un document Fast Infoset et pour le distinguer d'un document W3C XML 1.0 ou W3C XML 1.1.

3.4.16 infoset XML: ensemble de données abstraites qui décrit les informations dans un document XML à espace de nom bien conformé, comme spécifié dans l'ensemble d'informations XML W3C.

3.4.17 espace blanc XML: un ou plusieurs caractères TABULATION HORIZONTALE (9), SAUT DE LIGNE (10), RETOUR CHARIOT (13), ou ESPACE (32) d'Unicode.

NOTE – Ces caractères sont ceux qui correspondent à la production de "S" à la fois dans W3C XML 1.0 et W3C XML 1.1 (voir W3C XML 1.0, 2.3 et W3C XML 1.1, 2.3). Les caractères LIGNE SUIVANTE (133) et SEPARATEUR DE LIGNE (8232), qui peuvent survenir dans un document W3C XML 1.1 à espaces de nom bien conformés (voir W3C XML 1.1, 2.11), sont convertis en caractères SAUT DE LIGNE par un traitement de fin de ligne (voir W3C XML 1.1, 2.11). Si ces caractères surviennent dans un infoset XML généré à partir d'un document W3C XML 1.1 à espaces de nom bien conformés, ce ne sont pas des espaces blancs XML.

4 Abréviations

Pour les besoins de la présente Recommandation | Norme internationale, on applique les abréviations suivantes:

ASN.1	Notation de syntaxe abstraite numéro un (<i>abstract syntax notation one</i>)
BMP	Table multilingue de base (<i>basic multilingual plane</i>)
ECN	Notation de contrôle de codage (<i>encoding control notation</i>)
MIME	Extensions de messagerie Internet multiobjets (<i>multipurpose Internet mail extensions</i>)
UBL	Langage d'affaires universel (<i>universal business language</i>)
URI	Identificateur universel de ressources (<i>uniform resource identifier</i>)
UTF-8	Format de transformation UCS 8 (<i>universal transformation function 8-bit</i>) (voir l'ISO/CEI 10646, Annexe D)
UTF-16BE	Format de transformation UCS 16 bits gros-boutistes (<i>universal transformation function 16-bit Big Endian</i>) (voir Unicode, 2.6)
UUID	Identificateur unique universel (<i>universally unique identifier</i>)

XML Langage de balisage extensible (*extended markup language*)

5 Notation

5.1 La présente Recommandation | Norme internationale utilise la notation ASN.1 définie par la Rec. UIT-T X.680 | ISO/CEI 8824-1 pour la définition formelle des types de données dont les codages sont des documents Fast Infoset.

NOTE – Le paragraphe 12 spécifie l'application de la Rec. UIT-T X.692 | ISO/CEI 8825-3 aux définitions des types ASN.1, et donne le codage au niveau binaire d'un document Fast Infoset.

5.2 Dans la présente Recommandation | Norme internationale, la police de caractère **Courier gras** est utilisée pour les notations en ASN.1 et la police de caractères **Arial gras** est utilisée pour la syntaxe XML W3C et pour les noms des éléments d'information de l'ensemble d'informations XML.

5.3 Les noms des propriétés des éléments d'information sont en **Arial gras** et inclus entre crochets (par exemple, **[children]**).

5.4 Les noms des catégories de chaînes de caractères (voir le § 8.4.2) et les noms des catégories de noms qualifiés (voir le § 8.5.4) sont en MAJUSCULES.

5.5 Dans la présente Recommandation | Norme internationale, les positions des bits au sein d'un octet sont spécifiées en utilisant la terminologie "premier bit, second bit, etc., jusqu'au huitième bit", où le premier bit est le bit de plus fort poids de l'octet, et le huitième bit est le bit de plus faible poids de l'octet.

6 Principes de construction et d'utilisation des tableaux de vocabulaire

6.1 Les tableaux de vocabulaire sont des tables de concepts qui transposent un indice de tableau de vocabulaire en entrée de tableau de vocabulaire.

NOTE – La représentation des tableaux de vocabulaire dans les mémoires des ordinateurs n'est pas définie, pas plus que les moyens par lesquels une implémentation transpose un indice de tableau de vocabulaire en entrée de tableau de vocabulaire pour ce tableau.

6.2 Le créateur d'un document Fast Infoset à partir d'un infoset XML détermine le contenu des tableaux de vocabulaire.

6.3 Dans le cas le plus général, l'en-tête d'un document Fast Infoset peut faire référence à un ensemble de tableaux de vocabulaire (un vocabulaire externe), suivi par la spécification d'ajouts à ces tableaux de vocabulaire pour former le vocabulaire initial de ce document Fast Infoset. Des ajouts ultérieurs à ces tableaux de vocabulaire surviennent lors de la création et durant le traitement d'un document Fast Infoset, de sorte qu'ils croissent progressivement pour former les tableaux de vocabulaire finaux pour ce document.

6.4 Certains tableaux de vocabulaire croissent progressivement à partir d'un vocabulaire initial jusqu'à un vocabulaire final durant la création et pendant le traitement d'un document Fast Infoset, et ont donc le mot "dynamique" dans le nom du tableau de vocabulaire. Il n'existe pas de mécanisme de suppression des entrées d'un tableau.

6.5 Les indices d'un tableau de vocabulaire sont alloués de façon implicite. La première entrée de tout tableau de vocabulaire a un indice de tableau de vocabulaire de un, et chaque entrée suivante de ce tableau a la valeur de l'entier immédiatement supérieur d'indice de tableau de vocabulaire. Lorsque la présente Recommandation | Norme internationale spécifie que quelque chose doit être ajouté à un tableau de vocabulaire, cela implique qu'on doit lui allouer le prochain indice de tableau de vocabulaire disponible.

NOTE – Les indices de tableau de vocabulaire débutent à un et non à zéro parce que la valeur zéro (lorsqu'elle est permise) a la signification particulière de "chaîne de caractères vide" dans un champ qui pourrait autrement contenir un indice de tableau de vocabulaire.

6.6 Pour prendre en charge l'allocation implicite des indices de tableau de vocabulaire, l'ordre conceptuel de traitement des composants (de toute nature) d'un document Fast Infoset est entièrement défini (voir le § 8.1).

NOTE – Cet ordre est le même que l'ordre de codage des composants dans un document Fast Infoset. Cela n'implique pas nécessairement que la sémantique portée par le document soit traitée dans cet ordre. L'ordre n'est défini que pour garantir que le même indice de tableau de vocabulaire est alloué pour toute entrée de tableau de vocabulaire donnée à la fois par le créateur et par celui qui va traiter le document Fast Infoset.

6.7 Les tableaux de vocabulaire servent à de nombreux usages (voir le paragraphe 8), mais leur fonction première est de permettre l'utilisation d'un indice de tableau de vocabulaire à la place d'une entrée de tableau de vocabulaire, car de tels indices sont plus petits (et peuvent être traités plus vite) que les entrées de tableau. Un certain nombre d'entrées prédéfinies pour des tableaux de vocabulaire sont spécifiées au § 9. Ces entrées sont toujours implicitement présentes dans ces tableaux de vocabulaire, avec les indices de tableau de vocabulaire spécifiés au § 9.

6.8 Pour certaines catégories de chaînes de caractères, le créateur d'un document Fast Infoset a la faculté d'ajouter ou non une chaîne à un tableau de vocabulaire, selon le nombre attendu (ou connu) d'occurrences de cette chaîne de caractères dans l'infoset XML.

6.9 La forme et la signification précises des entrées d'un tableau de vocabulaire est spécifiée au § 8, mais ce sont dans la plupart des cas des chaînes de caractères de longueur variable, souvent courtes, mais pouvant aller jusqu'à 2^{32} octets.

6.10 Un créateur de document Fast Infoset conforme est obligé de faire tous les ajouts aux tableaux de vocabulaire spécifiés aux § 7.13.7, 7.14.6 à 7.14.7 et 7.16.7. Ceci garantit que le nombre d'entrées de tableau de vocabulaire dans chaque tableau de vocabulaire ne dépasse jamais 2^{20} .

NOTE – Une entrée de tableau de vocabulaire peut être égale à une ou plusieurs autres entrées de tableau de vocabulaire. Cela permet la création efficace de documents Fast Infoset. Toutefois, des entrées en double diminueront l'efficacité du transfert. Un processeur n'est pas affecté par des entrées en double.

6.11 Un traitement de document Fast Infoset conforme est nécessaire pour effectuer tous les ajouts aux tableaux de vocabulaire spécifiés aux § 7.13.8, 7.14.11 et 7.16.8. Ceci garantit que la restriction du § 6.10 a n'a pas été violée.

7 Définitions de type ASN.1

7.1 Généralités

7.1.1 La présente Recommandation | Norme internationale spécifie un ensemble de types ASN.1 qui prennent en charge une représentation de l'ensemble d'informations XML. Le type racine de cet ensemble de types est le type **Document**.

7.1.2 Quelques restrictions sont imposées sur le contenu des infosets XML et on a fait quelques simplifications dans la représentation (voir le § 11) afin d'améliorer la facilité d'utilisation de la spécification et l'efficacité des codages produits avec son aide.

NOTE – Un infoset XML qui ne satisfait pas à ces restrictions ne peut pas être présenté comme document Fast Infoset, et ne peut non plus être représenté normalement comme un document XML à espace de noms bien conformé.

7.1.3 Pour chaque sorte d'élément d'information spécifié dans l'ensemble d'informations XML W3C, une définition de type ASN.1 correspondante est fournie dans la présente Recommandation | Norme internationale. Cette définition de type est toujours un type de séquence, avec des composants correspondant aux propriétés de l'élément d'information.

7.1.4 Certaines propriétés des éléments d'information ne sont pas incluses dans les définitions de type ASN.1 (voir le § 11.4).

7.1.5 Dans certains cas, la valeur d'une propriété qui n'est pas incluse dans les définitions de type ASN.1 peut être déterminée à partir de la valeur d'autres propriétés du même élément d'information ou d'autres éléments d'information inclus. Dans ces cas, l'omission de cette propriété simplifie la représentation sans perte d'information. Il y a cependant, quelques cas dans lesquels la valeur d'une propriété qui n'est pas incluse ne peut être déduite des autres propriétés. Dans de tels cas, l'omission de cette propriété est une simplification qui ne limite pas l'utilité de la spécification pour la plupart des cas d'utilisation pratiques.

7.1.6 La section 12 spécifie le codage du type **Document**.

7.2 Le type **Document**

7.2.1 Le type **Document** est:

```
Document ::= SEQUENCE {
    additional-data          SEQUENCE (TAILLE(1..one-meg)) DE
        additional-datum SEQUENCE {
            id                URI,
            data              NonEmptyOctetString } FACULTATIF,
    initial-vocabulary      SEQUENCE {
        external-vocabulary  URI FACULTATIF,
        restricted-alphabets SEQUENCE (TAILLE(1..256)) DE
            NonEmptyOctetString FACULTATIF,
        encoding-algorithms SEQUENCE (TAILLE(1..256)) DE
            NonEmptyOctetString FACULTATIF,
        prefixes             SEQUENCE (TAILLE(1..one-meg)) DE
            NonEmptyOctetString FACULTATIF,
        namespace-names     SEQUENCE (TAILLE(1..one-meg)) DE
            NonEmptyOctetString FACULTATIF,
```

```

local-names          SEQUENCE (TAILLE(1..one-meg)) DE
                    NonEmptyOctetString FACULTATIF,
other-ncnames        SEQUENCE (TAILLE(1..one-meg)) DE
                    NonEmptyOctetString FACULTATIF,
other-uris           SEQUENCE (TAILLE(1..one-meg)) DE
                    NonEmptyOctetString FACULTATIF,
attribute-values     SEQUENCE (TAILLE(1..one-meg)) DE
                    EncodedCharacterString FACULTATIF,
content-character-chunks SEQUENCE (TAILLE(1..one-meg)) DE
                    EncodedCharacterString FACULTATIF,
other-strings        SEQUENCE (TAILLE(1..one-meg)) DE
                    EncodedCharacterString FACULTATIF,
element-name-surrogates SEQUENCE (TAILLE(1..one-meg)) DE
                    NameSurrogate FACULTATIF,
attribute-name-surrogates SEQUENCE (TAILLE(1..one-meg)) DE
                    NameSurrogate FACULTATIF }
(CONSTRAINED BY {
    -- Si le composant initial-vocabulary est présent, au moins
    -- un de ses composants doit être présent -- }) FACULTATIF,

notations            SEQUENCE (TAILLE(1..MAX)) DE
                    Notation FACULTATIF,
unparsed-entities    SEQUENCE (TAILLE(1..MAX)) DE
                    UnparsedEntity FACULTATIF,
character-encoding-scheme NonEmptyOctetString FACULTATIF,
standalone           BOOLEAN FACULTATIF,
version              NonIdentifyingStringOrIndex FACULTATIF
                    -- catégorie AUTRE CHAINE --,
children             SEQUENCE (TAILLE(0..MAX)) DE
                    CHOIX {
                        element          Element,
                        processing-instruction ProcessingInstruction,
                        comment           Comment,
                        document-type-declaration DocumentTypeDeclaration }}

```

où la valeur **one-meg** est:

```
one-meg ENTIER ::= 1048576 -- Deux à la puissance 20
```

Le type **NonEmptyOctetString** est: <https://standards.iteh.ai/catalog/standards/sist/939bdca4-3dd0-4521-b3ca-467b5f9cede3/iso-iec-24824-1-2007>

```
NonEmptyOctetString ::= OCTET STRING (TAILLE(1..four-gig))
```

où la valeur **four-gig** est:

```
four-gig ENTIER ::= 4294967296 -- Deux à la puissance 32
```

Le type **URI** est:

```
URI ::= NonEmptyOctetString
```

7.2.2 Les types **EncodedCharacterString**, **NameSurrogate**, **Notation**, **UnparsedEntity**, **NonIdentifyingStringOrIndex**, **Element**, **ProcessingInstruction**, **Comment** et **DocumentTypeDeclaration** sont définis respectivement aux § 7.17, 7.15, 7.11, 7.10, 7.14, 7.3, 7.5, 7.8 et 7.9.

7.2.3 Le type **URI** doit être un URI tel que spécifié dans la RFC 2396 de l'IETF.

7.2.4 Le composant **restricted-alphabets** de **initial-vocabulary** (s'il est présent) doit comporter une ou plusieurs chaînes de caractères, chacune portant les caractères d'un alphabet restreint. Chaque chaîne de caractères doit contenir au moins deux caractères, et tous les caractères de la chaîne de caractères doivent être distincts.

NOTE – L'utilisation d'alphabets restreints pour optimiser le codage des chaînes de caractères est spécifiée au § 7.17.6.

7.2.5 Le composant **encoding-algorithms** de **initial-vocabulary** (s'il est présent) doit comporter un ou plusieurs URI dont chacun identifie un algorithme de codage.

NOTE – La présente Recommandation | Norme internationale définit des algorithmes de codage prédéfinis (voir le § 10), avec des indices de tableau de vocabulaire spécifiés, mais il est en dehors du domaine d'application de la présente Recommandation | Norme internationale de définir plus avant les algorithmes de codage et les URI qui leur sont associés, ni les moyens de définir de tels algorithmes. Les informations nécessaires pour définir un algorithme de codage sont spécifiées au § 8.3.3.

7.2.6 Le type **Document** représente l'élément d'information **document** d'un infoset XML. Comme tous les autres éléments d'information d'un infoset XML sont des propriétés de cet élément d'information ou des propriétés d'un