

PROJET DE NORME INTERNATIONALE

ISO/DIS 24613-4

ISO/TC 37/SC 4

Secrétariat: KATS

Début de vote:
2020-02-17

Vote clos le:
2020-05-11

Gestion des ressources linguistiques — Cadre de balisage lexical (LMF) —

Partie 4: Sérialisation TEI

*Language resource management — Lexical markup framework (LMF) —
Part 4: TEI serialization*

ICS: 01.020

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
<https://standards.iteh.ai/catalog/standards/sist/700be78f-2bd5-42c9-80cb-b2b71c10a725/iso-dis-24613-4>

CE DOCUMENT EST UN PROJET DIFFUSÉ POUR OBSERVATIONS ET APPROBATION. IL EST DONC SUSCEPTIBLE DE MODIFICATION ET NE PEUT ÊTRE CITÉ COMME NORME INTERNATIONALE AVANT SA PUBLICATION EN TANT QUE TELLE.

OUTRE LE FAIT D'ÊTRE EXAMINÉS POUR ÉTABLIR S'ILS SONT ACCEPTABLES À DES FINS INDUSTRIELLES, TECHNOLOGIQUES ET COMMERCIALES, AINSI QUE DU POINT DE VUE DES UTILISATEURS, LES PROJETS DE NORMES INTERNATIONALES DOIVENT PARFOIS ÊTRE CONSIDÉRÉS DU POINT DE VUE DE LEUR POSSIBILITÉ DE DEVENIR DES NORMES POUVANT SERVIR DE RÉFÉRENCE DANS LA RÉGLEMENTATION NATIONALE.

LES DESTINATAIRES DU PRÉSENT PROJET SONT INVITÉS À PRÉSENTER, AVEC LEURS OBSERVATIONS, NOTIFICATION DES DROITS DE PROPRIÉTÉ DONT ILS AURAIENT ÉVENTUELLEMENT CONNAISSANCE ET À FOURNIR UNE DOCUMENTATION EXPLICATIVE.

Le présent document est distribué tel qu'il est parvenu du secrétariat du comité.



Numéro de référence
ISO/DIS 24613-4:2020(F)

© ISO 2020

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
<https://standards.iteh.ai/catalog/standards/sist/700be78f-2bd5-42c9-80cb-b2b71c10a725/iso-dis-24613-4>



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO 2020

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en oeuvre, 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, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Tél.: +41 22 749 01 11
Fax: +41 22 749 09 47
E-mail: copyright@iso.org
Website: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos.....	v
1 Domaine d'application.....	1
2 Références normatives.....	1
3 Termes et définitions.....	1
4 Principes généraux.....	2
5 Sérialisation du modèle de base LMF (ISO 24613-1).....	2
5.1 Implémentation de la classe LexicalResource.....	2
5.2 Implémentation de la classe GlobalInformation.....	3
5.3 Implémentation de la classe Lexicon.....	3
5.4 Implémentation de la classe LexiconInformation.....	4
5.5 Implémentation de la classe LexicalEntry.....	4
5.6 Implémentation des classes Form et Lemma.....	5
5.6.1 Classe Form.....	5
5.6.2 Classe Lemma.....	6
5.7 Implémentation de la classe GrammaticalInformation.....	6
5.8 Implémentation de la classe Sense.....	8
5.9 Implémentation de la classe Definition.....	8
5.10 Implémentation de la classe OrthographicRepresentation.....	8
5.11 Implémentation de l'allocation CrossREF.....	8
5.12 Sélection des catégories de données.....	10
6 Sérialisation de l'extension MRD (ISO 24613-2).....	10
6.1 Représentations de forme pour la classe Form.....	10
6.2 Classes dérivées de la classe Form.....	11
6.2.1 Principes généraux.....	11
6.2.2 Implémentation de la classe WordForm.....	12
6.2.3 Implémentation de la classe Stem.....	12
6.2.4 Implémentation de la classe WordPart.....	12
6.2.5 Implémentation de la classe RelatedForm.....	12
6.3 Implémentation de la classe Bibliography.....	13
6.4 Implémentation du mécanisme CrossREF de renvoi à des fichiers multimédia externes.....	13
6.5 Implémentation de la classe Example.....	13
6.6 Implémentation de la classe Translation.....	14
6.7 Implémentation de la classe TextRepresentation.....	14
6.8 Implémentation de la classe SubjectField.....	15
7 Implémentation des classes à partir de l'extension étymologique (ISO 24613-3).....	15
7.1 Implémentation de la classe Etymology.....	15
7.2 Implémentation de la classe Etymon.....	15
7.2.1 Référencement des formes dans un étymon.....	16
7.2.2 Représentation de la signification d'un étymon.....	16
7.2.3 Représentation de la langue d'un étymon.....	16
7.2.4 Datation d'un étymon.....	16
7.2.5 Citation de sources associées à un étymon.....	17
7.2.6 Élément de prose supplémentaire dans une description étymologique.....	17
7.3 Implémentation de la classe EtymLink (liens étymologiques).....	17
7.4 Implémentation de la classe CognateSet.....	17
7.5 Implémentation de la classe Cognate.....	17

8	Mécanismes supplémentaires	18
8.1	Vue d'ensemble	18
8.2	Représentation de signes de ponctuation avec <pc>	18
8.3	Représentation de diverses étiquettes avec <lbl>	18
8.4	Codage de caractères de séparation simples avec <metamark>	18
8.5	Transmission d'informations de rendu avec l'attribut @rend	19
8.6	Exemple d'usage des traits supplémentaires présentés à l'Article 8	19
Annexe A (informative) Spécification ODD de la TEI concernant la présente partie de l'ISO 24613		20
Annexe B (informative) Exemples de codage		21
Bibliographie		22

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
<https://standards.iteh.ai/catalog/standards/sist/700be78f-2bd5-42c9-80cb-b2b71c10a725/iso-dis-24613-4>

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (IEC) en ce qui concerne la normalisation électrotechnique.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier de prendre note des différents critères d'approbation requis pour les différents types de documents ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives).

L'attention est attirée 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 ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la nature volontaire des normes, la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir le lien suivant : www.iso.org/iso/fr/avant-propos.

Le présent document a été élaboré par le comité technique ISO/TC 37, *Langage et terminologie*, sous-comité SC 4, *Gestion des ressources linguistiques*.

Cette première édition de l'ISO 24613-4, conjointement avec l'ISO 24613-1, l'ISO 24613-2, l'ISO 24613-3 et l'ISO 24613-5, annule et remplace l'ISO 24613:2008, qui a été scindée en plusieurs parties et a fait l'objet d'une révision technique.

Les principales modifications par rapport à l'édition précédente sont les suivantes :

- révision complète du contenu et de sa subdivision.

Une liste de toutes les parties de la série ISO 24613 se trouve sur le site web de l'ISO.

Il convient que l'utilisateur adresse tout retour d'information ou toute question concernant le présent document à l'organisme national de normalisation de son pays. Une liste exhaustive desdits organismes se trouve à l'adresse www.iso.org/fr/members.html.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Full standard:
<https://standards.iteh.ai/catalog/standards/sist/700be78f-2bd5-42c9-80cb-b2b71c10a725/iso-dis-24613-4>

Gestion des ressources linguistiques — Cadre de balisage lexical (LMF) — Partie 4 : Sérialisation TEI

1 Domaine d'application

Le présent document décrit la sérialisation du modèle LMF défini en tant que modèle XML conforme aux lignes directrices de la TEI (Text Encoding Initiative). Cette sérialisation couvre les classes du modèle de base du LMF (ISO 24613-1) ainsi que les classes fournies par les parties supplémentaires suivantes de l'ISO 24613 : dictionnaires lisibles par ordinateur (ISO 24613-2) et étymologie (ISO 24613-3).

2 Références normatives

Les documents suivants sont cités dans le texte de sorte qu'ils constituent, pour tout ou partie de leur contenu, des exigences du présent document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document de référence s'applique (y compris les éventuels amendements).

ISO 24613-1, *Gestion des ressources linguistiques — Cadre de balisage lexical (LMF) — Partie 1 : Modèle de base*.

ISO 24613-2, *Gestion des ressources linguistiques — Cadre de balisage lexical (LMF) — Partie 2 : Modèle de dictionnaire lisible par ordinateur (MRD)*.

ISO 24613-3, *Gestion des ressources linguistiques — Cadre de balisage lexical (LMF) — Partie 3 : Extension étymologique*.

BCP 47 Tags for Identifying Languages, A. Phillips; M. Davis. IETF. Septembre 2009. IETF Best Current Practice. URL : <https://tools.ietf.org/html/bcp47>.

LANGUAGE E.M. (XML) 1.0 (Cinquième édition), W3C Recommendation, 26 novembre 2008 : <http://www.w3.org/TR/xml/>.

CONSORTIUM T.E.I., ed. TEI P5: Guidelines for Electronic Text Encoding and Interchange. [Version numéro : 3.6.0]. [Date de la dernière modification : 2019-07-16]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> ([2019-12-02]).

3 Termes et définitions

Pour les besoins du présent document, les termes et définitions de l'ISO 24613-1 et de l'ISO/DIS 24613-3 s'appliquent.

L'ISO et l'IEC tiennent à jour des bases de données terminologiques destinées à être utilisées en normalisation, consultables aux adresses suivantes :

- ISO Online browsing platform : disponible à l'adresse <https://www.iso.org/obp> ;
- IEC Electropedia : disponible à l'adresse <http://www.electropedia.org/>.

4 Principes généraux

Le présent document est destiné à fournir des constructions essentiellement univoques pour chaque classe LMF issue du modèle de base (ISO 24613-1), de l'extension MRD (ISO 24613-2) et de l'extension étymologique (ISO/DIS 24613-3), en rendant ainsi les recommandations d'origine de la TEI du chapitre sur les dictionnaires plus laconiques (cf. [8],[10],[13],[7] et [3]), tout en spécifiant les contraintes spécifiques nécessaires pour s'aligner sur l'ISO 24613 en général. Par exemple, des listes de valeurs précises sont données pour certains attributs tels que @type. De plus, le présent document respecte les cardinalités exprimées dans l'ISO 24613-1, l'ISO 24613-2 et l'ISO/DIS 24613-3.

Sauf indication explicite, toutes les constructions résultantes doivent être des représentations valides de la TEI, ce qui implique que la spécification décrite dans le présent document correspond à un sous-ensemble strict des lignes directrices de la TEI.

De la même manière, des représentations précises sont fournies pour toutes les catégories de données citées dans l'ISO 24613-1, l'ISO 24613-2 et l'ISO/DIS 24613-3.

Néanmoins, le présent document n'aborde pas les aspects des métadonnées de LMF car l'en-tête TEI, c'est-à-dire le composant de métadonnées associé à tout document TEI, est par essence nettement plus riche car il représente tous les aspects liés à la création, au contenu, au versionnage et à la publication d'un document textuel dans son ensemble.

Dans tous les exemples XML du présent document et afin de simplifier les représentations réelles, l'hypothèse sera faite, sauf indication contraire, que les éléments XML appartiennent à l'espace de nommage TEI, ce qui suppose que l'exemple est dans le domaine d'application de la déclaration d'espace de nommage XML suivante :

```
xmlns="http://www.tei-c.org/ns/1.0"
```

5 Sérialisation du modèle de base LMF (ISO 24613-1)

5.1 Implémentation de la classe LexicalResource

La classe LexicalResource doit être implémentée au moyen de l'élément <teiCorpus>¹⁾ qui regroupe un ou plusieurs documents TEI pour chaque lexique spécifique. L'élément <teiCorpus> doit inclure un attribut @type avec la valeur « lexicalResource ». Si la ressource lexique ne contient qu'un seul lexique, ce niveau peut être omis, de sorte que le document débute directement par un élément racine <TEI> (voir 5.3).

Tableau 1 — Classe LexicalResource

Classe LMF	Construction TEI
/LexicalResource/	<teiCorpus type="lexicalResource">... </teiCorpus>

¹⁾ <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-teiCorpus.html>

5.2 Implémentation de la classe GlobalInformation

La classe GlobalInformation doit être implémentée au moyen de l'élément TEI <teiHeader>²⁾, en tant qu'enfant direct de l'élément <teiCorpus>. Pour l'utilisation effective de l'élément <teiHeader>, les contraintes supplémentaires indiquées dans les lignes directrices de la TEI doivent être respectées lors du codage de tout type d'information se rapportant à des informations administratives, documentaires et bibliographiques rattachées à la ressource lexicale correspondante.

Tableau 2 — Classe GlobalInformation

Classe LMF	Construction TEI
/GlobalInformation/	<teiHeader>...</teiHeader>

Les exigences supplémentaires concernant la classe GlobalInformation sont implémentées de la façon suivante :

- comme la sérialisation TEI repose sur la recommandation W3C pour XML, elle implémente l'attribut @xml:lang pour indiquer l'information sur la langue de travail correspondant au contenu de tout élément spécifique. D'après la recommandation W3C, le contenu de @xml:lang doit être conforme à BCP 47. Il n'est donc pas nécessaire d'effectuer une implémentation spécifique de la catégorie de données /language coding/ afin de garantir la conformité du présent document à l'ISO 24613-1 ;
- BCP 47 repose également sur l'ISO 15924^[1] pour le codage des noms d'écritures. Il n'est donc pas nécessaire de disposer d'un mécanisme spécifique dans le présent document pour implémenter la catégorie de données /script coding/ ;
- la catégorie de données /character encoding/ est implémentée dans la déclaration XML d'un document conforme à la TEI, en utilisant l'attribut @encoding. Par exemple, un document XML-TEI codé en UTF-8 conformément à la norme Unicode doit débiter par la déclaration suivante :

```
<?xml version="1.0" encoding="UTF-8" ?>
```

5.3 Implémentation de la classe Lexicon

La classe Lexicon doit être implémentée au moyen de l'élément TEI <TEI>³⁾, qui regroupe toutes les entrées lexicales d'un lexique spécifique. L'élément <TEI> doit avoir un attribut @type avec la valeur « lexicon ».

Tableau 3 — Classe Lexicon

Classe LMF	Construction TEI
/Lexicon/	<TEI type="lexicon">...</TEI>

2) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-teiHeader.html>

3) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-TEI.html>

D'après les lignes directrices de la TEI, le contenu principal d'un document TEI est incorporé dans un élément enfant <text>⁴⁾ qui est ensuite décomposé en sous-éléments <front>⁵⁾, <body>⁶⁾ et <back>⁷⁾. Toutes les entrées lexicales d'un lexique donné doivent être des éléments descendants du sous-élément <body> de <text>, et peuvent être réorganisées en sous-ensembles au moyen de l'élément <div>⁸⁾. <front> et <back> peuvent être utilisés pour incorporer des ressources complémentaires le cas échéant, par exemple des références bibliographiques.

L'exemple suivant illustre une organisation documentaire simple pour un lexique (sans les éléments facultatifs <front> et <back>, ainsi que les subdivisions) :

EXEMPLE

```
<TEI type="lexicon">
  <teiHeader> ... </teiHeader>
  <text>
    <body>
      <entry> ... </entry>
      <entry> ... </entry>
      ...
    </body>
  </text>
</TEI>
```

5.4 Implémentation de la classe LexiconInformation

La classe LexiconInformation doit également être implémentée au moyen de l'élément TEI <teiHeader>, en tant qu'enfant direct de l'élément <TEI>. Pour l'utilisation effective de l'élément <teiHeader>, les contraintes supplémentaires indiquées dans les lignes directrices de la TEI doivent être respectées pour coder tout type d'information se rapportant à des informations administratives, documentaires et bibliographiques rattachées au lexique correspondant.

Tableau 4 — Classe LexiconInformation

Classe LMF	Construction TEI
/LexiconInformation/	<teiHeader>...</teiHeader>

5.5 Implémentation de la classe LexicalEntry

La classe LexicalEntry doit être implémentée au moyen de l'élément <entry>⁹⁾ des lignes directrices de la TEI. Il convient de coder les informations lexicales à l'intérieur de l'élément <entry> en utilisant les éléments enfants suivants :

- <form>¹⁰⁾ pour les informations morphologiques et grammaticales liées à toute l'entrée ;
- <etym>¹¹⁾ pour les aspects étymologiques ;

4) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-text.html>

5) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-front.html>

6) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-body.html>

7) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-back.html>

8) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-div.html>

9) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-entry.html>

10) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-form.html>

11) <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-etym.html>