

# NORME INTERNATIONALE **ISO 24617-11**

Première édition  
2021-08

---

---

## **Gestion des ressources linguistiques — Cadre d'annotation sémantique (SemAF) —**

### **Partie 11: Informations quantitatives mesurables (MQI)**

*Language resource management — Semantic annotation framework (SemAF) —*

*Part 11: Measurable quantitative information (MQI)*

[ISO 24617-11:2021](https://standards.iteh.ai/standards/iso/4a03ace4-8e0e-463b-b60f-8269aa3b8ad5/iso-24617-11-2021)

<https://standards.iteh.ai/catalog/standards/iso/4a03ace4-8e0e-463b-b60f-8269aa3b8ad5/iso-24617-11-2021>



Numéro de référence  
ISO 24617-11:2021(F)

© ISO 2021

iTeh Standards  
(<https://standards.iteh.ai>)  
Document Preview

[ISO 24617-11:2021](https://standards.iteh.ai/catalog/standards/iso/4a03ace4-8e0e-463b-b60f-8269aa3b8ad5/iso-24617-11-2021)

<https://standards.iteh.ai/catalog/standards/iso/4a03ace4-8e0e-463b-b60f-8269aa3b8ad5/iso-24617-11-2021>



**DOCUMENT PROTÉGÉ PAR COPYRIGHT**

© ISO 2021

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, 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, Genève

Tél.: +41 22 749 01 11

E-mail: [copyright@iso.org](mailto:copyright@iso.org)

Web: [www.iso.org](http://www.iso.org)

Publié en Suisse

## Sommaire

Page

<b>Avant-propos</b> .....	<b>iv</b>
<b>Introduction</b> .....	<b>v</b>
<b>1 Domaine d'application</b> .....	<b>1</b>
<b>2 Références normatives</b> .....	<b>1</b>
<b>3 Termes et définitions</b> .....	<b>1</b>
<b>4 Spécification abstraite de QML</b> .....	<b>3</b>
4.1 Vue d'ensemble.....	3
4.2 Caractéristiques de QML.....	4
4.3 Métamodèle.....	4
4.4 Syntaxe abstraite de QML (QML_as).....	5
4.5 Syntaxes concrètes de QML (QML_cs) et de ses sous-ensembles.....	6
<b>5 Syntaxe concrète de QML basée sur XML (QML_csx)</b> .....	<b>6</b>
5.1 Généralités.....	6
5.2 Noms de balises avec préfixes d'ID.....	6
5.3 Spécification des attributs de la racine <MQI>.....	7
5.4 Spécification des attributs des types d'éléments de base.....	7
5.5 Spécification des attributs des types de liens.....	8
5.6 Illustrations de QML_csx.....	8
5.6.1 Généralités.....	8
5.6.2 Échantillons de données.....	8
5.6.3 Procédure d'annotation.....	9
<b>6 Syntaxe concrète de QML basée sur la TEI (QML_cst)</b> .....	<b>11</b>
6.1 Syntaxes concrètes de QML (QML_cst).....	11
6.1.1 Généralités.....	11
6.1.2 Noms de balises avec préfixes d'ID.....	11
6.1.3 Spécification des attributs des types d'éléments de base.....	11
6.1.4 Spécification des attributs des deux types de liens.....	12
6.2 Illustrations de QML_cst.....	12
6.2.1 Généralités.....	12
6.2.2 Échantillons de données.....	13
6.2.3 Illustrations de la syntaxe concrète basée sur la TEI.....	13
<b>Annexe A (informative) Illustrations de QML_csx avec davantage d'échantillons</b> .....	<b>16</b>
<b>Annexe B (informative) Énoncés informels de MQI</b> .....	<b>19</b>
<b>Annexe C (informative) Représentation des unités</b> .....	<b>20</b>
<b>Bibliographie</b> .....	<b>21</b>

## 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](http://www.iso.org/directives)).

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 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](http://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](http://www.iso.org/iso/fr/avant-propos).

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

Une liste de toutes les parties de la série ISO 24617 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](http://www.iso.org/fr/members.html).

## Introduction

Les informations quantitatives mesurables (MQI, Measurable Quantitative Information) telles que «165 cm» ou «60 kg» de «John» qui s'appliquent à la taille ou au poids de la personne sont très courantes dans le langage ordinaire. Les MQI décrivent l'une des propriétés de base qui est associée à l'aspect quantitatif d'une grandeur. Les principales caractéristiques de la norme MQI sont que les informations quantitatives sont présentées sous forme de mesures exprimées en termes de paire  $\langle n, u \rangle$ , consistant en une grandeur exprimée numériquement  $n$  et une unité  $u$ , qui est une unité de base ou une unité dérivée, ou encore une unité normalisée ou utilisée par convention. Ces informations sont beaucoup plus abondantes dans les publications scientifiques ou les rapports techniques au point qu'elles constituent une part essentielle des segments communicatifs du langage en général. Le traitement de ces informations est donc nécessaire pour une gestion réussie des ressources linguistiques.

À l'époque du «big data», les demandes de l'industrie et des milieux universitaires pour une acquisition précise des informations quantitatives mesurables ont augmenté. Par exemple, les sociétés d'investissement dans les entreprises ont fréquemment besoin d'agréger différents types d'informations couvrant les ventes nettes, la marge brute, les frais d'exploitation, le bénéfice d'exploitation, les frais d'intérêt, le bénéfice net avant impôts, le revenu net, etc. des sociétés cibles à partir de leurs rapports annuels. La recherche en informatique médicale, en plein essor, a également besoin de traiter une grande quantité de textes médicaux pour analyser la dose de médicament, les critères d'éligibilité des essais cliniques, les caractères phénotypiques des patients, les essais en laboratoire dans les dossiers cliniques, etc.<sup>[8]</sup>. Toutes ces demandes, qu'elles soient liées à l'industrie ou à la recherche médicale, exigent la représentation précise et cohérente des informations quantitatives mesurables afin de permettre un traitement, un calcul et un échange automatisés.

Cependant, en IR et en PNL, il n'existe actuellement aucun moyen normalisé de représenter les informations quantitatives mesurables. Chaque système d'application développé dans les secteurs industriels utilise jusqu'à présent son propre format pour annoter les informations quantitatives mesurables. Un format de représentation des informations quantitatives mesurables qui soit flexible, interopérable et normalisé est nécessaire pour permettre aux tâches d'IR et de PNL de fonctionner avec de nombreux systèmes d'application différents.

Le présent document vise à formuler un schéma d'annotation général en suivant les principes d'annotation sémantique définis dans l'ISO 24617-6 en général et les exigences de base de l'ISO 24611, qui facilite le traitement des MQI dans le langage scientifique et technique et afin de le rendre interopérable avec d'autres schémas d'annotation sémantique, tels que l'ISO 24617. Le schéma d'annotation est conçu pour être interopérable avec les autres parties de l'ISO 24617. Il s'appuie également sur diverses normes ISO relatives aux ressources lexicales et aux cadres d'annotation morpho-syntaxique. Il vise à être compatible avec les autres normes pertinentes existantes.

**NOTE** L'ISO 24617-1 et l'ISO 24617-7, par exemple, ont proposé un moyen d'annoter les mesures de temps (durées ou quantités de temps) et d'espace (distances), respectivement. L'ISO 24612 fournit un formulaire pivot (cadre d'annotation graphique) qui permet de réaliser toutes les annotations de mesures de temps et d'espace dans ces deux schémas d'annotation.

Le QML est normalisé à un niveau abstrait qui permet divers formats de sérialisation représentant les informations quantitatives mesurables annotées, tels qu'une représentation basée sur XML. La normalisation de l'annotation QI (information quantitative) est indiquée au niveau abstrait de l'annotation, et le format d'annotation déportée est adopté au niveau concret de la sérialisation.

Axé sur les mesures en langage scientifico-technologique, le présent document est censé contribuer aux applications d'extraction d'information (IR)<sup>[9]</sup>, de réponse aux questions (QA), de résumé de texte (TS) et autres applications de traitement du langage naturel (NLP)<sup>[10]</sup>.



# Gestion des ressources linguistiques — Cadre d'annotation sémantique (SemAF) —

## Partie 11: Informations quantitatives mesurables (MQI)

### 1 Domaine d'application

Le présent document porte sur l'aspect mesurable ou quantitatif de la grandeur, de sorte qu'il est possible de se concentrer sur l'utilisation technique ou pratique des mesures dans les applications IR (recherche d'informations), QA (réponse aux questions), TS (résumé de texte) et autres applications NLP (traitement du langage naturel). Il s'applique aux domaines technologiques qui présentent plus d'intérêt sur le plan de l'application que certains problèmes théoriques rencontrés dans l'utilisation ordinaire du langage.

NOTE L'ISO 24617-12 traite des questions plus générales et théoriques de la quantification et de l'information quantitative.

Le présent document traite également des durées temporelles qui sont abordées dans l'ISO 24617-1 et des mesures spatiales telles que les distances qui sont traitées dans l'ISO 24617-7, tout en les rendant interopérables avec d'autres types de mesures. Il intègre également le traitement des mesures ou des montants qui sont introduits dans l'ISO 24617-6:2016, 8.3.

### 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 24612, *Gestion des ressources linguistiques — Cadre d'annotation linguistique (LAF)*

### 3 Termes et définitions

Pour les besoins du présent document, les termes et définitions suivants 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 <https://www.electropedia.org/>.

#### 3.1

##### **grandeur**

propriété d'un objet mesurable se référant à son ampleur ou à sa multiplicité

[SOURCE: ISO/IEC Guide 99:2007, 1.1, modifiée — La définition a été considérablement remaniée et les notes ont été supprimées.]

### 3.2

#### **grandeur de base**

*grandeur* (3.1) d'un sous-ensemble choisi par convention dans un système de grandeurs donné de façon qu'aucune grandeur du sous-ensemble ne puisse être exprimée en fonction des autres grandeurs de ce sous-ensemble

Note 1 à l'article: La nature des grandeurs comprend sept grandeurs de base définies par le Système international de grandeurs (ISQ).

[SOURCE: ISO/IEC Guide 99:2007, 1.4, modifiée — L'expression «des autres» a été remplacée par «des autres grandeurs de ce sous-ensemble», et les notes ainsi que l'exemple ont été supprimés.]

### 3.3

#### **grandeur dérivée**

*grandeur définie* (3.1), dans un système de grandeurs, en fonction des *grandeurs de base* (3.2) de ce système

EXEMPLE La vitesse est une grandeur dérivée définie par la longueur (distance) par rapport au temps ( $LT^{-1}$ ), où la longueur (L) et le temps (T) sont des grandeurs de base.

[SOURCE: ISO/IEC Guide 99:2009, 1.5, modifiée — L'exemple a été remplacé.]

### 3.4

#### **information quantitative**

##### **QI**

mesure associée à la *grandeur* (3.1) d'un objet mesurable

### 3.5

#### **information quantitative mesurable**

##### **MQI**

*information quantitative* (3.4) qui peut être exprimée en termes numériques unifiés

### 3.6

#### **langage de balisage des informations quantitatives mesurables**

#### **langage de balisage quantitatif**

##### **QML**

langage de spécification pour l'annotation des *informations quantitatives mesurables* (3.5) extractibles de textes ou d'autres types de support de langage

### 3.7

#### **unité de mesure**

##### **unité**

base scalaire, définie et adoptée par convention, de la mesure des objets par multiplication de leurs valeurs quantitatives exprimées en nombres réels

Note 1 à l'article: Les expressions utilisées en mesurage telles que «mètre», «litre» et « $\mu\text{mol/kg}$ » sont des unités selon la définition donnée ci-dessus. Les expressions de multiplicité telles que «bouteilles», «boîtes» ou «deux» comme dans «deux bouteilles de lait», «une boîte de pommes» et «deux cafés» ne sont parfois pas considérées comme des unités, mais elles peuvent l'être si elles sont acceptées comme unités par convention ou accord dans certaines communautés. L'ISO 24617 SemAF Partie 12: Quantification traite ces expressions de multiplicité comme de véritables unités.

[SOURCE: ISO/IEC Guide 99:2007, 1.9, modifiée — La définition a été considérablement remaniée, les notes d'origine ont été supprimées et une nouvelle Note 1 à l'article a été ajoutée.]

### 3.8

#### **unité de base**

*unité de mesure* (3.7) adoptée par convention pour une *grandeur de base* (3.2)

Note 1 à l'article: Il existe sept unités de base choisies par le Système international d'unités (SI) associées à sept grandeurs de base ISQ pour mesurer les grandeurs, comme indiqué dans le [Tableau 1](#).

Tableau 1 — Unités de base

Unité SI de base (symbole de l'unité)	Grandeur de base de l'ISQ associée (symbole de la grandeur de base)
mètre (m)	longueur (L)
kilogramme (kg)	masse (M)
seconde (s)	temps (T)
ampère (A)	courant électrique (I)
kelvin (K)	température thermodynamique (È)
mole (mol)	quantité de matière (N)
candela (cd)	intensité lumineuse (J)

[SOURCE: ISO/IEC Guide 99:2007, 1.10, modifiée — Les notes et les exemples ont été supprimés, et une nouvelle Note 1 à l'article ainsi que le [Tableau 1](#) ont été ajoutés.]

### 3.9

#### unité dérivée

unité de mesure (3.7) d'une grandeur dérivée (3.3)

EXEMPLE L'unité «newton» (N) est une unité dérivée pour une grandeur dérivée «force» (F), qui est définie comme la «masse multipliée par l'accélération» ( $MLT^{-2}$ ), où la grandeur «accélération» est une grandeur dérivée définie par la «vitesse divisée par le temps» ( $VT^{-1}$ ) et la «vitesse» définie par la «longueur (distance) divisée par le temps» ( $LT^{-1}$ ).

Note 1 à l'article: Le [Tableau 2](#) illustre certaines des unités dérivées.

[SOURCE: ISO/IEC Guide 99:2007, 1.11, modifiée — Les exemples ont été supprimés et un nouvel exemple ainsi que la Note 1 à l'article ont été ajoutés.]

Tableau 2 — Unités dérivées

Unité dérivée (symbole de l'unité)	Grandeur dérivée associée
kilomètre par minute (km/min)	vitesse = longueur(L)/temps(T)
gramme par mètre cube ( $g/m^3$ )	masse volumique = masse(M)/volume( $L^3$ )
kilogramme mètre par seconde carrée ( $kg \times m/s^2$ )	force = masse (M) x longueur(L)/temps( $T^2$ )
lumen par mètre carré ( $lm/m^2$ )	éclairage lumineux = intensité lumineuse (J)/aire( $M^2$ )

## 4 Spécification abstraite de QML

### 4.1 Vue d'ensemble

Le langage de balisage quantitatif (QML) (3.6) est spécifié à deux niveaux, abstrait et concret. Certaines caractéristiques de QML sont énumérées en 4.2. La structure globale de QML est représentée par un métamodèle, tel que présenté en 4.3. La syntaxe abstraite de QML comme QML\_as doit être une spécification ensembliste de QML en termes conceptuels qui sont indépendants des manières de représenter l'annotation (contenu) des informations quantitatives mesurables. La syntaxe concrète de QML comme QML\_cs doit être une spécification d'un ensemble de formats de représentation, basé sur QML\_as, pour l'annotation des informations quantitatives mesurables d'une manière traçable informatiquement. QML\_as est présenté en 4.4, tandis que QML\_cs est présenté en 4.5. Les syntaxes concrètes équivalentes, dont une syntaxe concrète QML\_csx basée sur XML et une syntaxe concrète QML\_cst basée sur la TEI, sont décrites à l'Article 5 et à l'Article 6 respectivement.

NOTE Il peut y avoir de nombreuses syntaxes concrètes équivalentes définies sur une seule syntaxe abstraite.

## 4.2 Caractéristiques de QML

Le QML doit présenter les caractéristiques suivantes:

- le QML doit être axé sur l'annotation des attributs mesurables des entités. Par exemple, «IMC entre 10-20 kg/m<sup>2</sup>»;
- le QML doit permettre d'annoter les relations des mesures. Par exemple, «âge 40 ou plus» et «fpg >= 100 mg/dl ou a1c pas moins de 5,8 %»;
- le QML doit couvrir les utilisations complexes de grandeurs numériques unifiées. Par exemple, «14,0 × 10<sup>9</sup>», «hémoglobine glyquée (hba1c) < 1,15 fois la limite supérieure de la normale»;
- le QML doit faciliter l'identification d'unités et numériques normalisés en tant qu'attribut mesurable d'une entité associée.

NOTE Le QML ne spécifie pas les moyens d'annoter la normalisation (par exemple, «millimoles par litre» est normalisé par «mmol/l») ou la spécification complète (par exemple, «kg/m» s'écrit «kg/m<sup>2</sup>» pour l'IMC) des unités, ce qui sera abordé dans une autre partie de l'ISO 24617 traitant de la mise en œuvre automatisée des MQI.

## 4.3 Métamodèle

La structure globale des informations quantitatives mesurables est représentée par le métamodèle de la [Figure 1](#).

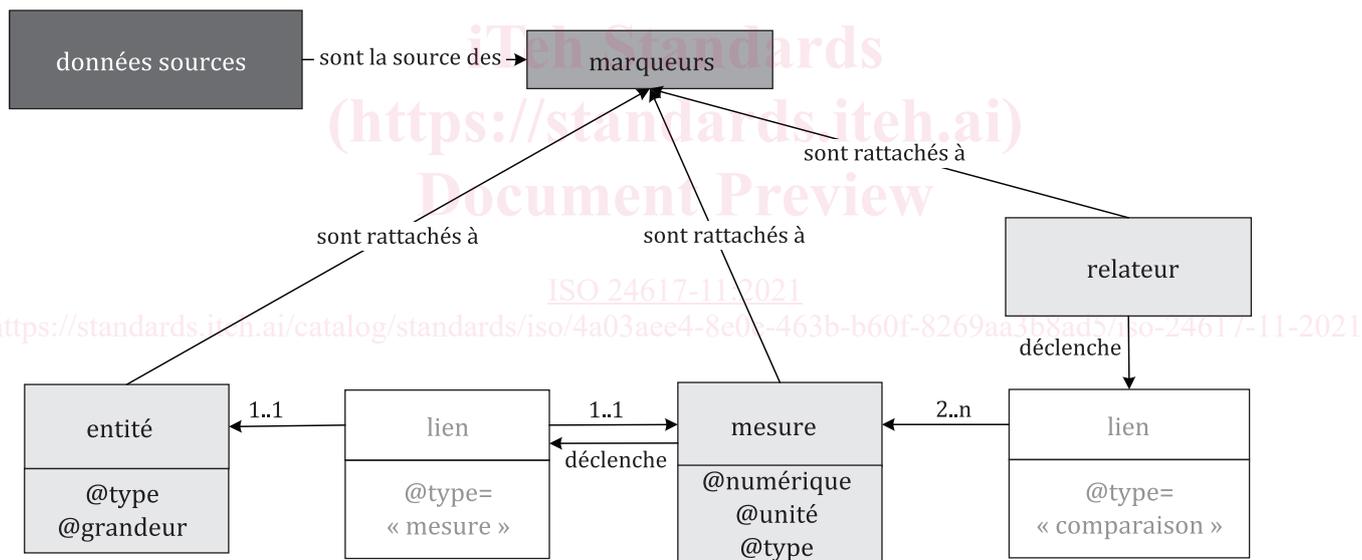


Figure 1 — Métamodèle des informations quantitatives mesurables

Ce métamodèle doit se composer de sept composantes de classe, représentées par des cases carrées à la [Figure 1](#):

- données sources** en entrée pour l'annotation des MQI;
- marqueurs** extraits des sources de données;
- trois types d'éléments de base: **entité**, **mesure**, et **relateur**;
- deux types de liens: **lien de mesure** et **lien de comparaison**.

L'élément «entité» doit être tout objet qui a la propriété d'une grandeur mesurable, représentée par «@ grandeur», comme l'une de ses propriétés. L'« entité », telle qu'elle est utilisée dans le présent document, doit être un terme très général qui fait référence à tout objet, non seulement à des entités individuelles,