
Informatique génomique — Langage de balisage Omics (OML)

Genomics informatics — Omics Markup Language (OML)

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

ISO 21393:2021

<https://standards.iteh.ai/catalog/standards/iso/2cb358e3-3ff1-4d41-b290-a285028cc807/iso-21393-2021>



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

ISO 21393:2021

<https://standards.iteh.ai/catalog/standards/iso/2cb358e3-3ff1-4d41-b290-a285028cc807/iso-21393-2021>



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

Web: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos	iv
Introduction	v
1 Domaine d'application	1
2 Références normatives	1
3 Termes et définitions	1
4 Spécifications de l'OML	6
4.1 Exigences de spécification et positionnement de l'OML.....	6
4.2 Structure de l'OML.....	7
4.3 DTD de l'OML et schéma XML.....	7
5 Processus de développement de l'OML	8
6 Figures	8
Annexe A (informative) Travaux de référence	29
Bibliographie	47

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

[ISO 21393:2021](https://standards.iteh.ai/catalog/standards/iso/2cb358e3-3ff1-4d41-b290-a285028cc807/iso-21393-2021)

<https://standards.iteh.ai/catalog/standards/iso/2cb358e3-3ff1-4d41-b290-a285028cc807/iso-21393-2021>

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 www.iso.org/avant-propos.

Le présent document a été élaboré par le comité technique ISO/TC 215, *Informatique de santé*, sous-comité SC 1, *Informatique génomique*, en collaboration avec le comité technique CEN/TC 251, *Informatique de santé*, du Comité européen de normalisation (CEN) conformément à l'Accord de coopération technique entre l'ISO et le CEN (Accord de Vienne).

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.

Introduction

Dans cette ère post-génomique, la gestion des données de santé devient de plus en plus importante tant pour la recherche omique que pour la médecine basée sur les sciences omiques^[1]. Les approches informationnelles de la gestion des données cliniques, d'images et d'omique commencent à avoir autant de valeur que les recherches ordinaires en laboratoire. Le monde électronique actuel est marqué par de nombreux types de données différents en matière de soins de santé, comme indiqué à la [Figure 1](#). En outre, il existe aujourd'hui de nombreux types de données omiques de par le monde qui attendent une utilisation efficace dans le domaine de la santé humaine. Le développement d'un format de données et de normes de message pour prendre en charge l'échange de données omiques cliniques est nécessaire. Les données omiques comprennent la séquence omique, la variation de séquence et d'autres données d'expression, les données protéomiques, le réseau moléculaire, etc. Comme point d'entrée, le présent document se concentre sur l'échange de données.

Dans les circonstances actuelles, on s'attend à ce que l'omique soit une clé pour comprendre la réponse humaine aux stimuli externes tels que n'importe quels types d'invasions étrangères, de thérapies, et d'interactions environnementales^[2]. L'infection bactérienne est un exemple d'invasion étrangère et les réponses aux infections diffèrent d'un individu à l'autre. Selon la thérapie utilisée, les effets secondaires d'un médicament diffèrent d'un patient à l'autre. Ces réponses diffèrent également d'un environnement à l'autre. Le nombre de ces recherches omiques ayant explosé récemment, les données expérimentales s'accumulent en grande quantité dans de nombreuses bases de données sous différents types de formats de données. Ces données attendent d'être utilisées dans la découverte de médicaments, le diagnostic clinique et les recherches cliniques.

Le langage de balisage est un ensemble de symboles et de règles permettant de les utiliser dans le balisage d'un document^[3]. Le premier langage de balisage normalisé a été le langage normalisé de balisage généralisé (SGML)^[4] de l'ISO 8879, qui présente de fortes similitudes avec les langages de présentation de texte troff et nroff qui accompagnent les systèmes Unix. Le langage HTML (langage de balisage hypertexte) est basé sur SGML^[5]. XML (langage de balisage extensible) est une version réduite du SGML, conçue particulièrement pour les documents Web^[6]. XML sert de base au XHTML (HTML extensible)^[7] et au WML (langage de balisage sans fil)^[8] ainsi qu'à des définitions normalisées d'interaction système telles que SOAP (Simple Object Access Protocol)^[9]. Par contre, la présentation de texte ou la sémantique est souvent définie sous une forme purement interprétable par machine, comme dans la plupart des formats de fichiers de traitement de texte^[10].

Le langage de balisage dans le domaine biomédical basé sur XML est en cours de développement depuis plusieurs décennies dans le but d'améliorer l'échange de données entre chercheurs. Le BSML (langage de balisage de séquence bioinformatique)^[11], le SBML (langage de balisage en biologie des systèmes)^[12], le Cell ML (langage de balisage de cellules)^[13] et le Neuro-ML (Langage de balisage neuronal)^[14] sont des exemples de langages de balisage. Le Polymorphism Mining and Annotation Programs (PolyMAPr)^[15] est centré sur le SNP et tente de réaliser l'exploration, l'annotation et l'analyse fonctionnelle des bases de données publiques telles que dbSNP^[16] CGAP^[17], et JSNP^[18] par le biais de la programmation. Le langage de balisage de variation de la séquence génomique (GSVML) de l'ISO 25720 est le premier langage de balisage normalisé pour l'échange de données relatives à la variation de la séquence génomique dans un contexte clinique.

Le langage de balisage Omics (OML) vise à fournir le format normalisé d'échange de données pour les sciences omiques dans le domaine de la santé humaine.

L'essor récent de la recherche omique a généré d'importantes quantités de données conservées dans de nombreuses bases de données sous différents formats. La gestion, l'analyse et l'utilisation de ces données exigent une normalisation de l'échange de données. Compte tenu de l'importance des sciences omiques pour la médecine moléculaire et la pharmacogénomique, en particulier la transcriptomique, la protéomique, la signalomique et la métabolomique, le format d'échange de données est essentiel pour améliorer la recherche clinique et la médecine basées sur des approches omiques.

Les approches informationnelles ont récemment gagné en importance tant pour la recherche omique que pour la médecine basée sur les sciences omiques. Dans cette nouvelle ère, la gestion des données omiques est devenue aussi essentielle que celle des données de recherche fondamentale. Il existe de

nombreux types de données omiques dans le monde et le temps est venu d'utiliser efficacement ces données pour la santé humaine. Pour utiliser ces données de manière efficace et efficiente, il convient d'élaborer des normes pour permettre l'échange interopérable des données omiques dans le monde. Il convient que ces normes définissent le format de données ainsi que les messages qui seront utilisés pour échanger et partager ces données à l'échelle internationale.

OML est un cadre de base pour tous les types de données omiques cliniques. Chaque catégorie du domaine omique sera présentée sous la forme d'une composante complémentaire spécifique. Par exemple, le langage de balisage du séquençage de génome complet formera une composante complémentaire spécifique pour des données de séquençage de génome complet, et le langage de balisage de la variation de la séquence génomique formera une composante complémentaire spécifique des données de variation de la séquence génomique.

Pour utiliser les données omiques cumulées à l'échelle internationale, il convient de définir des normes autour de l'échange de données omiques. Il convient que ces normes définissent un format de données et de messages d'échange. Le langage de balisage est un choix raisonnable pour répondre à ce besoin. Quant à la gestion des messages de données omiques, le groupe de travail de génomique clinique^[19] au sein du Health Level Seven®¹⁾ a récapitulé les cas d'utilisation clinique pour les données omiques générales. Le projet OML a contribué à ces efforts. En outre, ces travaux ont incorporé des cas d'utilisation basés sur le «Millennium Project» japonais^[20]. Sur la base de ces contextes et investigations, le présent document élucide les besoins et les exigences pour l'OML et propose ensuite la spécification de l'OML en vue de la normalisation internationale.

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

ISO 21393:2021

<https://standards.iteh.ai/catalog/standards/iso/2cb358e3-3ff1-4d41-b290-a285028cc807/iso-21393-2021>

1) Health Level Seven (HL7) est une marque déposée de Health Level Seven International. Cette information est donnée à l'intention des utilisateurs du présent document et ne signifie nullement que l'IEC approuve l'emploi du produit ainsi désigné.

Informatique génomique — Langage de balisage Omics (OML)

1 Domaine d'application

Le présent document est applicable au format d'échange de données qui est conçu pour faciliter l'échange de données omiques dans le monde entier sans imposer le moindre changement de schéma de base de données.

Le présent document spécifie les caractéristiques de l'OML selon les perspectives suivantes.

D'un point de vue informatique, OML définit le format d'échange de données basé sur XML. Le présent document établit des lignes directrices pour la spécification du format d'échange de données, mais il exclut le schéma de base de données proprement dit.

Du point de vue moléculaire, le présent document est applicable à toutes les sortes de données omiques bien qu'il exclue les détails relatifs aux molécules (par exemple, les détails des variations de la séquence génomique ou la séquence génomique complète). Le présent document est également applicable aux annotations moléculaires, y compris les questions cliniques et les relations avec les autres questions omiques.

Du point de vue de l'application, le présent document est applicable à la santé humaine, y compris les pratiques cliniques, la médecine préventive, la recherche translationnelle et la recherche clinique, notamment la découverte de médicaments. Le présent document ne s'applique pas à la recherche fondamentale et aux autres domaines scientifiques.

Du point de vue des espèces biologiques, le présent document est applicable aux espèces associées à la santé humaine telles que l'homme, les animaux en préclinique et les lignées cellulaires. Le présent document ne s'applique pas aux autres espèces biologiques.

<https://standards.iteh.ai/catalog/standards/iso/2cb358e3-3ff1-4d41-b290-a285028cc807/iso-21393-2021>

2 Références normatives

Le présent document ne contient aucune référence normative.

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 <http://www.electropedia.org/>

3.1

acteur agent

entité qui fournit un stimulus au système

Note 1 à l'article: Les acteurs englobent tant les humains que d'autres entités quasi autonomes, telles que machines, tâches informatiques et systèmes.

[SOURCE: ISO 25720:2009, 4.1]