NORME INTERNATIONALE

ISO 21393

Première édition
2021-07

Informatique génomique — Langage de balisage Omics (OML)

Genomics informatics — Omics Markup Language (OML)

Numéro de référence
ISO 21393:2021(F)

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

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.

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.

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]

3.2
allèle
gène trouvé dans différentes formes à la même position dans un chromosome

3.3
Bioinformatic Sequence Markup Language
BSML
spécification de langage extensible et conteneur pour données bioinformatiques

[SOURCE: ISO 25720:2009, 4.2]

3.4
Cancer Genome Anatomy Project
CGAP
données d'expression génomiques recueillies pour différents tissus tumorigènes chez l'homme et chez la souris

Note 1 à l'article: Le projet CGAP fournit également des informations sur des méthodes et des réactifs utilisés pour obtenir les données génomiques.

[SOURCE: ISO 25720:2009, 4.4, modifiée]

3.5
codon
séquence de trois nucléotides qui, ensemble, forment une unité de code génétique dans une molécule d'ADN ou d'ARN

3.6
dbSNP
base de données sur les *SNP* (3.29) fournie par le National Center for Biotechnology Information (NCBI) des États-Unis d'Amérique

Note 1 à l'article: Disponible à l'adresse https://www.ncbi.nlm.nih.gov/SNP/.

[SOURCE: ISO/TS 20428:2017, 3.9]

3.7
Digital Imaging and Communications in Medicine
DICOM
norme dans le domaine de l'informatique médicale pour l'échange d'information numérique entre un équipement d'imagerie médicale (tel qu'une imagerie radiologique) et d'autres systèmes, assurant l'interopérabilité

[SOURCE: ISO 25720:2009, 4.6]

3.8
variation de la séquence d'ADN
différences de séquence d'ADN parmi des individus dans une population

Note 1 à l'article: La variation de la séquence d'ADN implique le polymorphisme (3.25).

[SOURCE: ISO 25720:2009, 4.8]

3.9
Document Type Definition
DTD
document qui contient les définitions formelles de tous les éléments de données dans un type particulier de document *HTML* (3.13), *SGML* (3.29), ou *XML* (3.36)

[SOURCE: ISO 25720:2009, 4.9]

3.10
point d'entrée
point de référence qui indique la (les) classe(s) où les messages débutent pour le domaine

[SOURCE: ISO 25720:2009, 4.10, modifiée]

3.11
exon
partie d'un gène qui encode une partie de l'ARN mature final produit par ce gène après l'élimination des *introns* (3.16) par épissage de l'ARN

3.12
Genomic Sequence Variation Markup Language
GSVML
norme pour l'échange de données de la variation de la séquence génomique

3.13
Hypertext Markup language
HTML
ensemble de symboles ou codes de balisage insérés dans un fichier destiné à l'affichage dans un navigateur

[SOURCE: ISO 25720:2009, 4.12, modifiée]

3.14
classification internationale des maladies
CIM
système de codage de diagnostic normalisé pour l'épidémiologie, la gestion de la santé et les applications cliniques

Note 1 à l'article: La CIM-10 est la dixième révision et la CIM-11 est la onzième révision.

Note 2 à l'article: disponible à l'adresse https://icd.who.int/

3.15
sous-modèle d'informations omiques cliniques pour la classification internationale des maladies
sous-modèle d'informations omiques cliniques pour la CIM
iCOS
sous-modèle d'informations destiné à renforcer la capacité de représentation du modèle de contenu CIM-11 afin de couvrir les informations omiques en tant que partie complémentaire

Note 1 à l'article: Modèle d'informations complémentaires pour renforcer la capacité de représentation du modèle de contenu CIM-11 afin de couvrir les informations relatives aux domaines omiques.

3.16
intron
séquence nucléotidique dans un gène qui est éliminée par épissage de l'ARN pendant la maturation du produit ARN final

3.17
Joint Photographic Experts Group
JPEG
technique de compression d'images

[SOURCE: ISO 25720:2009, 4.13]

3.18
langage de balisage
ML (markup language)
ensemble de symboles et règles pour leurs utilisations dans le balisage d'un document

[SOURCE: ISO 25720:2009, 4.15]

3.19
Microarray Gene Expression Markup Language
MAGE-ML
format de données pour décrire des informations sur des expérimentations basées sur le réseau d'ADN et les données d'expression des gènes

3.20
Neuro Markup Language
Neuro-ML
langage de balisage (3.18) pour décrire des modèles de neurones et les réseaux de neurones

[SOURCE: ISO 25720:2009, 4.16]

3.21
nroff
programme de formatage de texte sous Unix qui est le prédécesseur du système de document *troff* (3.33) sous Unix

[SOURCE: ISO 25720:2009, 4.17]

3.22
omique
«omics»
domaine d'étude biologique ayant le suffixe «omique»

Note 1 à l'article: Inclut, sans toutefois s'y limiter, la génomique, la protéomique et la métabolomique.

3.23
pharmacogénomique
branche de la pharmacie visant à développer un moyen rationnel permettant d'optimiser la chimiothérapie, en fonction du génotype du patient

3.24
Polymorphism Mining and Annotation Programs
PolyMAPr
programmes pour l'exploration, l'annotation et l'analyse fonctionnelle de bases données du *polymorphisme* (3.25)

[SOURCE: ISO 25720:2009, 4.19]

3.25
polymorphisme
variation de la séquence de l'ADN parmi les individus

Note 1 à l'article: Le polymorphisme implique le *SNP* (3.29) et le *STRP* (3.32).

[SOURCE: ISO 25720:2009, 4.20]

3.26
RNA Markup Language
RNAML
format de données pour l'échange d'informations ARN

3.27
Systems Biology Markup Language
SBML
langage de balisage (3.18) pour les simulations en biologie des systèmes

[SOURCE: ISO 25720:2009, 4.21]

3.28
Standard Generalized Markup Language
SGML
langage de balisage (3.18) pour la représentation de documents qui formalise le balisage et le rend indépendant des systèmes et des traitements

[SOURCE: ISO 8879:1986, 4.305, modifiée]

3.29
Single Nucleotide Polymorphism
SNP
variation d'un seul nucléotide dans une séquence génétique qui apparaît à une fréquence appréciable dans la population

[SOURCE: ISO 25720:2009, 4.23]

3.30
Systematized Nomenclature of Medicine - Clinical Terms®[2)]
SNOMED-CT®
ensemble dynamique et validé scientifiquement d'infrastructure et de terminologie de soins de santé cliniques

[SOURCE: ISO 25720:2009, 4.24]

3.31
Simple Object Access Protocol
SOAP
protocole léger pour l'échange d'informations dans un environnement réparti décentralisé

[SOURCE: ISO 25720:2009, 4.25]

3.32
Short Tandem Repeat Polymorphism
STRP
segments variables de l'ADN qui ont une longueur de deux bases à cinq bases avec de nombreuses séquences répétées

[SOURCE: ISO 25720:2009, 4.26]

3.33
troff
composant principal d'un système de traitement de documents développé par AT&T pour le système d'exploitation Unix

3.34
Wireless Markup Language
WML
langage de balisage extensible utilisé pour spécifier le contenu et l'interface utilisateur pour des dispositifs WAP (protocole d'application sans fil)

[SOURCE: ISO 25720:2009, 4.29]

2) SNOMED CT est une marque déposée de l'International Health Terminology Standards Development Organisation. 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.

3.35
eXtensible HTML
XHTML
hybride entre *HTML* (3.13) et *XML* (3.36) spécialement conçu pour les écrans d'affichage de dispositifs Net

[SOURCE: ISO 25720:2009, 4.30]

3.36
eXtensible Markup Language
XML
version réduite du *SGML* (3.29), conçue pour les documents Web

[SOURCE: ISO 25720:2009, 4.31]

3.37
schéma XML
langage servant à décrire la structure et à contraindre le contenu de documents XML

[SOURCE: ISO 25720:2009, 4.32]

4 Spécifications de l'OML

4.1 Exigences de spécification et positionnement de l'OML

Dans le contexte actuel, les informations d'annotation relatives au domaine omique vont en augmentant et ces informations tendent à incorporer les trous d'information. Les données omiques en tant que telles augmentent également, mais sont stockées dans différentes bases de données. Le piège dans le traitement des données omiques se situe dans l'absence de normalisation des formats de données pour l'omique organisée. Historiquement, les langages de balisage ont été utilisés et des programmes sont développés pour gérer l'information omique. Toutefois, il n'existait pas jusqu'ici de langages de balisage centrés sur l'omique. L'OML est le premier langage de balisage centré sur l'omique et axé sur la santé humaine. Compte tenu de l'impact considérable de l'omique notamment pour la santé et la réponse humaines, on peut affirmer que l'OML a le plus grand potentiel pour devenir le langage de balisage désigné pour les soins de santé humaine. D'autre part, la mise en place des applications à la santé humaine dans la pratique signifie qu'il doit gérer les annotations directes ou indirectes. Ici, l'annotation directe doit indiquer les informations d'annotation générales telles que l'omique associée à d'autres informations omiques et à des préparations expérimentales. L'annotation indirecte doit indiquer la totalité des données omiques et des données cliniques résultant des données omiques. Pour comprendre la situation clinique omique de chaque patient, ces types d'informations supplémentaires sont nécessaires. Eu égard à la nécessité d'ajouter de nombreux types d'informations supplémentaires, le développement et la normalisation de l'OML ne peuvent pas être isolés et doivent donc faire l'objet d'une harmonisation avec les différents documents des autres organismes internationaux de normalisation.

L'OML est destiné à être utilisé dans les messages d'échange de données liés à la santé humaine. Pour le développement et la normalisation de l'OML dans ce domaine d'application, il est impératif de toujours garder un œil sur la sécurité du patient, l'efficience clinique et les coûts médicaux. Pour la sécurité du patient du point de vue informationnel, la conservation et la protection des informations relatives au patient doivent être jugées importantes. Pour le renforcement de l'efficience clinique, la simplicité et l'intelligibilité sans peine doivent être jugées importantes. Pour la réduction des coûts médicaux, la capacité d'adaptation et la facilité d'installation doivent être jugées importantes.

L'OML tente de satisfaire à ces exigences fondamentales en fournissant le format partageable d'échange de données basé sur XML. L'OML peut être utilisé pour l'échange de données omiques d'un point de vue clinique parmi divers types de formats de données. Dans le cadre plus large de la normalisation des données cliniques, l'OML doit jouer un rôle dans la description des données omiques et de leurs informations requises.

4.2 Structure de l'OML

Une expression OML valide doit être structurée conformément aux éléments suivants, voir également Figure 2:

— le contour de la structure de l'OML est illustré à la Figure 2.

L'OML doit être constitué de trois critères de données:

— données «omiques»;

— annotation directe;

— annotation indirecte.

Le critère de données omiques doit décrire, pour chaque domaine omique,

les données omiques simples telles que:

— le type;

— la position;

— la longueur;

— la région;

— etc.

Le critère d'annotation directe doit décrire, pour chaque domaine omique,

les données jointes des données omiques telles que:

— l'analyse d'expérimentations;

— l'épidémiologie;

— l'omique associée;

— etc.

Le critère d'annotation indirecte doit décrire

les informations explicatives/de niveau supérieur des données omiques, telles que:

— les informations cliniques;

— les données environnementales.

Ces critères de données doivent intérieurement avoir des relations les uns avec les autres.

— La structure détaillée de l'OML doit être fournie telle qu'illustrée aux Figures 3 à 23.

4.3 DTD de l'OML et schéma XML

La définition de type de document (DTD) de l'OML est disponible pour information à l'adresse https://standards.iso.org/iso/21393/ed-1/en.

Le schéma XML de l'OML utilisé est disponible à l'adresse https://standards.iso.org/iso/21393/ed-1/en.

5 Processus de développement de l'OML

Étape 1: établir les éléments et les besoins selon les cas d'utilisation étudiés à l'aide de l'iCOS CIM-11 de l'OMS.

Étape 2: construire la structure de base et la DTD.

Étape 3: étudier le langage de balisage biologique existant, en particulier le GSVML (ISO 25720), et son applicabilité aux besoins. (Comparaison avec les langages MAGE-ML, BSML, SBML, RNAML[21], ProML, CellML, PolyMAPr)

Étape 4: affiner la structure de base et la DTD, construire le schéma XML (XSD).

Étape 5: étudier le format existant (comparaison de leurs formats de données).

Étape 6: vérifier la capacité d'interface au modèle de génotype du Health Level Seven®.

Étape 7: redéfinir les besoins par rapport à l'OML et ses éléments exigés.

Étape 8: affiner la structure de base, la DTD et le schéma XML.

La Figure 24 montre le contour du processus de développement. Le projet a été élaboré en harmonie avec le GT génomie clinique de HL7®, le groupe CDISC BRIDGE, le groupe FIC ITC de l'OMS aussi bien pour l'ISO 25720 (GSVML) que pour le présent document. Il y a eu des «allers et retours» entre l'élaboration du projet et le processus de normalisation.

En outre, l'interface entre l'OML, l'ISO 13606 (toutes les parties), et la SNOMED-CT® a été analysée.

L'Annexe A fournit des informations supplémentaires ayant contribué à l'élaboration du présent document.

6 Figures

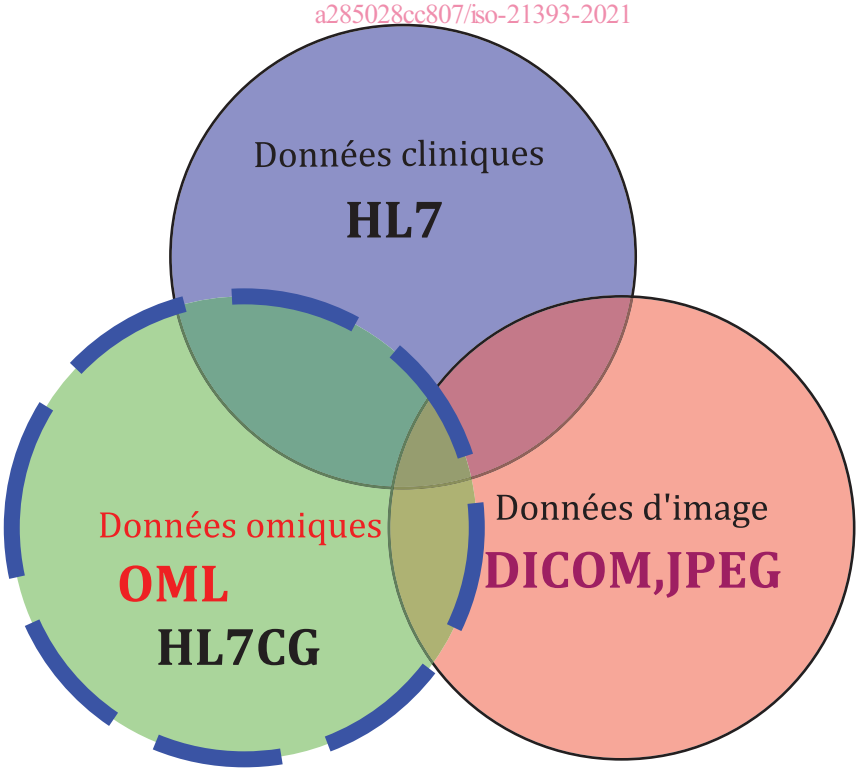

Figure 1 — Types de donnée de soins de santé les plus importants

Le monde de réseau électronique actuel est marqué par de nombreux types de données différents en matière de soins de santé, comme indiqués à la Figure 1. Au-delà des données cliniques et des données d'image, à l'heure où nous entrons dans cette nouvelle ère post-génomique, d'immenses quantités de données omiques sont créées à l'échelle internationale. Des normes applicables à ces données sont en cours d'élaboration chez les organismes de normalisation; le Health Level Seven® (HL7®) élabore des normes pour les données cliniques, DICOM et JPEG élaborent des normes pour les données d'image; et l'Omics Markup Language (OML) définit une norme pour les données omiques, en particulier des données omiques en lien avec l'homme. L'OML cible essentiellement le format d'échange de données.

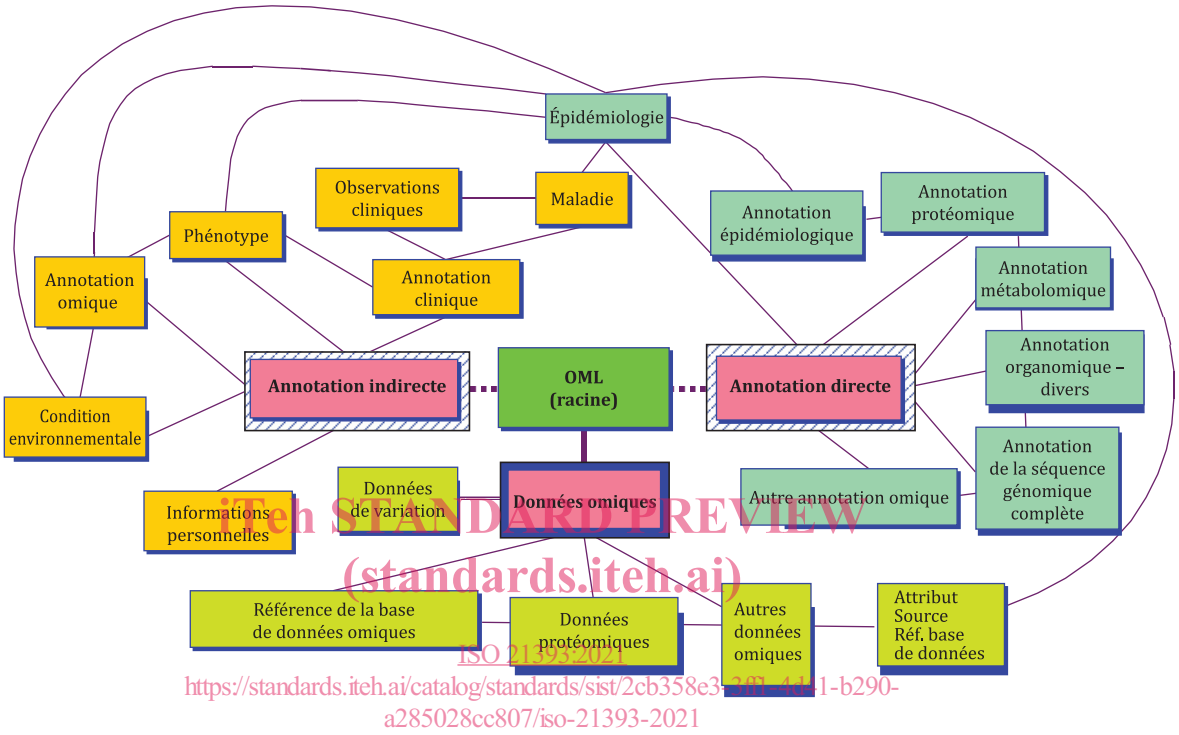

Figure 2 — Contour de la structure de l'OML

La structure globale de l'OML est centrée sur l'OML (racine) et sur le contenu des données omiques, à savoir les données de variation, les données protéomiques ou d'autres données omiques. Les informations relatives aux processus omiques ou qui ne sont autrement pas incluses dans les données omiques sont contenues dans des annotations directes. Les annotations indirectes permettent d'inclure dans le document OML des informations cliniques, phénotypiques, environnementales et autres données similaires.

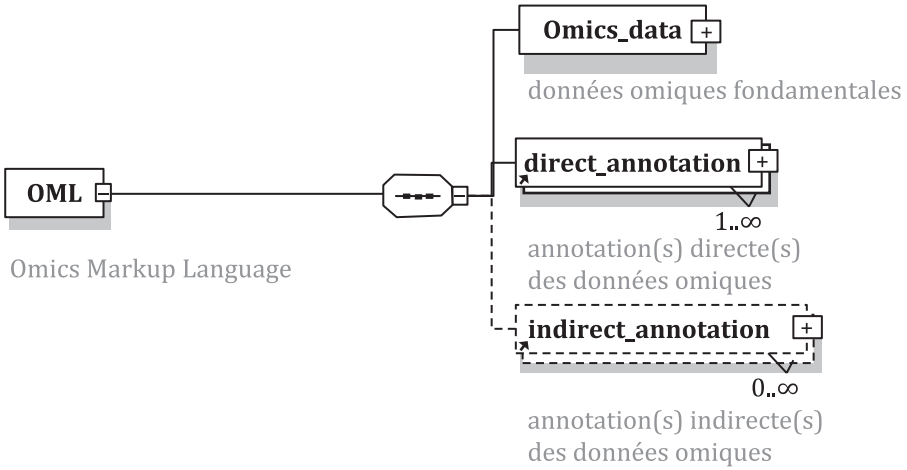

Figure 3 — Structure détaillée de l'OML: racine OML (OML)