

DIN EN ISO 21393

ICS 35.240.80

**Genomik-Informatik –
Omics Auszeichnungssprache (OML) (ISO 21393:2021);
Englische Fassung EN ISO 21393:2021**

Genomics informatics –
Omics Markup Language (OML) (ISO 21393:2021);
English version EN ISO 21393:2021

Informatique génomique –
Langage de balisage Omics (OML) (ISO 21393:2021);
Version anglaise EN ISO 21393:2021

Gesamtumfang 60 Seiten

DIN-Normenausschuss Medizin (NAMed)

Nationales Vorwort

Dieses Dokument (EN ISO 21393:2021) wurde vom Technischen Komitee ISO/TC 215 „Health informatics“ in Zusammenarbeit mit dem Technischen Komitee CEN/TC 251 „Medizinische Informatik“ erarbeitet, dessen Sekretariat von NEN (Niederlande) gehalten wird.

Das zuständige deutsche Normungsgremium ist der Arbeitsausschuss NA 063-07-02 AA „Interoperabilität“ im DIN-Normenausschuss Medizin (NAMed).

Dieses Dokument enthält unter Berücksichtigung des DIN-Präsidialbeschlusses 1/2004 nur die englische Originalfassung von EN ISO 21393:2021.

Der Bedarf für eine Deutsche Fassung wird derzeit beraten.

Für die in diesem Dokument zitierten Dokumente wird im Folgenden auf die entsprechenden deutschen Dokumente hingewiesen:

ISO 13606 (all parts) siehe DIN EN ISO 13606 (alle Teile)

Aktuelle Informationen zu diesem Dokument können über die Internetseiten von DIN (www.din.de) durch eine Suche nach der Dokumentennummer aufgerufen werden.

Nationaler Anhang NA (informativ)

Begriffe

Die Benummerung der folgenden Begriffe ist identisch mit der Benummerung in der englischen Fassung.

3 Begriffe

Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

ISO und IEC stellen terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen bereit:

- ISO Online Browsing Platform: verfügbar unter <https://www.iso.org/obp>
- IEC Electropedia: verfügbar unter <http://www.electropedia.org/>

3.1

Akteur

etwas oder jemand, das/der einen Reiz an das System liefert

Anmerkung 1 zum Begriff: Akteure beinhalten sowohl Menschen als auch andere quasi-autonome Dinge wie Maschinen, Computertätigkeiten und Systeme.

[QUELLE: ISO 25720:2009, 4.1]

3.2

Allel

Gen, das in einer, zwei oder mehreren unterschiedlichen Form(en) an derselben Stelle in einem Chromosom gefunden wird

3.3

Auszeichnungssprache für bioinformatische Sequenz

BSML, en: Bioinformatic Sequence Markup Language
erweiterbare Sprachspezifikation und Container für bioinformatische Daten

[QUELLE: ISO 25720:2009, 4.2]

3.4

Cancer Gene Anatomy Project

CGAP

Datenbank zur Genexpression, gesammelt für unterschiedliche tumorerzeugende Gewebe sowohl im Menschen als auch in Mäusen

Anmerkung 1 zum Begriff: CGAP liefert außerdem Informationen zu den Verfahren und Reagenzien für die Ableitung der genomischen Daten.

[QUELLE: ISO 25720:2009, 4.4, modifiziert]

3.5

Codon

Sequenz von drei Nukleotiden, die zusammen eine Einheit des genetischen Codes in einem DNA- oder RNA-Molekül bilden

3.6

dbSNP

Datenbank des *Einzelnukeotid-Polymorphismus* (3.29), bereitgestellt durch das US National Center for Biotechnology Information (NCBI)

Anmerkung 1 zum Begriff: Verfügbar unter <https://www.ncbi.nlm.nih.gov/SNP/>.

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

3.7

digitale Bildgebung und -kommunikation in der Medizin

DICOM, en: Digital Imaging and Communications in Medicine

Standard im Feld der medizinischen Informatik zum Austausch digitaler Informationen zwischen medizinischen Bildgebungsgeräten (wie radiologische Bildgebung) und anderen Systemen zur Sicherstellung der Interoperabilität

[QUELLE: ISO 25720:2009, 4.6]

3.8

DNA-Sequenz-Variation

Unterschiede der DNA-Sequenz zwischen Individuen einer Population

Anmerkung 1 zum Begriff: DNA-Sequenz-Variation impliziert Polymorphismus (3.25).

[QUELLE: ISO 25720:2009, 4.8]

3.9

Dokumenttypdefinition

DTD, en: Document Type Definition

Dokumenttyp, der formale Definitionen aller Datenelemente eines bestimmten Typs von Dokumenten in *Hypertext-Auszeichnungssprache* (3.13), *Normierter Verallgemeinerter Auszeichnungssprache* (3.28)* oder *erweiterbarer Auszeichnungssprache* (3.36) beinhaltet

[QUELLE: ISO 25720:2009, 4.9]

3.10

Eingangsstelle

Referenzpunkt, der die Klasse(n) bezeichnet, ab der/denen die Nachrichten für die Domain beginnen

[QUELLE: ISO 25720:2009, 4.10, modifiziert]

3.11

Exon

Teil eines Gens, der einen Teil der fertigen reifen RNA kodiert, welche nach dem Entfernen der Introns (3.16) durch das RNA-Spleißen durch das Gen produziert wird

* Nationale Fußnote: CEN ISO/TS 21393 verweist an dieser Stelle fälschlicherweise auf 3.29.

3.12**Auszeichnungssprache für Genomsequenz-Variation****GSVML**, en: Genomic Sequence Variation Markup Language

Standard zum Datenaustausch von Genomsequenz-Variationsdaten

3.13**Hypertext-Auszeichnungssprache****HTML**, en: Hypertext Markup Language

Reihe von Auszeichnungssymbolen oder Codes, die in eine für die Darstellung im Browser vorgesehenen Datei eingefügt sind

[QUELLE: ISO 25720:2009, 4.12, modifiziert]

3.14**Internationale Klassifikation der Krankheiten****ICD-11**, en: International Classification of Diseases

Diagnose-Kodierungssystem für epidemiologische, gesundheitspolitische und klinische Belange

Anmerkung 1 zum Begriff: ICD-10 ist die 10. Revision und ICD-11 ist die 11. Revision.

Anmerkung 2 zum Begriff: verfügbar unter <https://icd.who.int/>**3.15****klinisches Omics-Sub-Informationsmodell für die internationale Klassifikation von Krankheiten**, en: clinical omics sub-information model for international classification of diseases**klinisches Omics-Sub-Informationsmodell für ICD**, en: Clinical Omics Sub-information model for ICD
iCOS

Unterinformationsmodell, das darauf abzielt, die Darstellungsfähigkeit des ICD-11-Inhaltsmodells durch die Einbeziehung von Omics-Informationen als zusätzlichen Teil zu verbessern

Anmerkung 1 zum Begriff: Zusätzliches Sub-Informationsmodell zum Verbessern der Darstellungsfähigkeit des ICD11-Inhaltsmodells zum Abdecken der Omics-Informationen.

3.16**Intron**

Nukleotidsequenz innerhalb eines Gens, die durch das RNA-Spleißen während der Reifung des fertigen RNA-Produkts entfernt wird

3.17**Joint Photographic Experts Group****JPEG**

Kompressionstechnik für Bilder

[QUELLE: ISO 25720:2009, 4.13]

3.18**Auszeichnungssprache****ML**, en: Markup Language

Reihe von Symbolen und Regeln zur Verwendung während der Auszeichnung eines Dokuments

[QUELLE: ISO 25720:2009, 4.15]

3.19**Auszeichnungssprache für Microarray Genexpression****MAGE-ML**, en: Microarray Gene Expression Markup Language

Datenformat zur Beschreibung von Informationen zu DNA-Array-Experimenten und Genexpressionsdaten

3.20

Neuro-Auszeichnungssprache

neuro-ML, en: Neuro Markup Language

Auszeichnungssprache (3.18) für die Beschreibung von Neuronmodellen und -netzwerken

[QUELLE: ISO 25720:2009, 4.16]

3.21

nroff

Unix-Textformatierungsprogramm, das ein Vorläufer des Unix-Dokumentenverarbeitungssystems *troff* (3.33) ist

[QUELLE: ISO 25720:2009, 4.17]

3.22

omics

Teilgebiet der Biologie, das auf -omics endet

Anmerkung 1 zum Begriff: Es behandelt u.a. Genomik, Proteomik und Metabolomik.

3.23

Pharmakogenomik

Branche der Pharmazeutik für die Entwicklung rationaler Mittel zur Verbesserung der Arzneimitteltherapie mit Bezug auf den Genotyp des Patienten

3.24

Polymorphismus-Förderung und Anmerkungsprogramme

PolyMAPr, en: Polymorphism Mining and Annotation Programs

Programme für die Förderung, Anmerkung und Funktionsanalyse von *Polymorphismus*-Datenbanken (3.25)

[QUELLE: ISO 25720:2009, 4.19]

3.25

Polymorphismus

Variation in der DNA-Sequenz zwischen Individuen

Anmerkung 1 zum Begriff: Polymorphismus impliziert Einzelnukleotid-Polymorphismus (3.29) und kurze wiederholende Polymorphismus-Einheiten (3.32).

[QUELLE: ISO 25720:2009, 4.20]

3.26

RNA-Auszeichnungssprache

RNAML

Datenformat zum Austausch von RNA-Informationen

3.27

Auszeichnungssprache für Systembiologie

SBML, en: Systems Biology Markup Language

Auszeichnungssprache (3.18) für Simulationen in der Systembiologie

[QUELLE: ISO 25720:2009, 4.21]

3.28**normierte verallgemeinerte Auszeichnungssprache****SGML**, en: Standard Generalized Markup Language*Auszeichnungssprache* (3.18) für Dokumentdarstellung, die Auszeichnung formalisiert und von System- und Prozessabhängigkeiten befreit

[QUELLE: ISO 8879:1986, 4.305, modifiziert]

3.29**Einzelnukleotid-Polymorphismus****SNP**, en: Single Nucleotide Polymorphism

Einzelnukleotid-Polymorphismus-Variation in einer Gensequenz, die in einer nennenswerten Häufigkeit in der Population auftritt

[QUELLE: ISO 25720:2009, 4.23]

3.30**systematisierte Nomenklatur der Medizin - klinische Begriffe ®******SNOMED-CT ®**, en: Systematized Nomenclature Of Medicine - Clinical Terms

dynamische, wissenschaftlich validierte Terminologie klinischer Gesundheitsversorgung und Infrastruktur

[QUELLE: ISO 25720:2009, 4.24]

3.31**einfaches Objekt-Zugangsprotokoll****SOAP**, en: Simple Object Access Protocol

einfaches Protokoll für den Austausch von Informationen in einer dezentral verteilten Umgebung

[QUELLE: ISO 25720:2009, 4.25]

3.32**kurze wiederholende Polymorphismus-Einheiten****STRP**, en: Short Tandem Repeat Polymorphism

variable DNA-Segmente, bestehend aus zwei bis fünf Basen mit zahlreichen Wiederholungen

[QUELLE: ISO 25720:2009, 4.26]

3.33**troff**

Hauptkomponente eines Dokumentenverarbeitungssystem, das von AT&T für das Betriebssystem Unix entwickelt wurde

3.34**Wireless Markup Language****WML**

erweiterbare Auszeichnungssprache zur Spezifizierung von Inhalt und Benutzeroberfläche für WAP (Wireless Application Protocol)-Geräte

[QUELLE: ISO 25720:2009, 4.29]

** SNOMED CT ist das eingetragene Warenzeichen der International Health Terminology Standards Development Organisation. Diese Informationen dienen der Bequemlichkeit der Nutzer dieses Dokuments und stellen keine Billigung des genannten Produkts durch ISO dar.

3.35

erweiterbare HTML

XHTML, en: Extensible HTML

speziell für Netzgerätedisplays entwickelter Hybrid zwischen der *Hypertext-Auszeichnungssprache* (3.13) und der *Erweiterbaren Auszeichnungssprache* (3.36)

[QUELLE: ISO 25720:2009, 4.30]

3.36

erweiterbare Auszeichnungssprache

XML, en: Extensible Markup Language

verkleinerte Version der *Normierten Verallgemeinerten Auszeichnungssprache* (3.29), speziell für Web-Dokumente entwickelt

[QUELLE: ISO 25720:2009, 4.31]

3.37

XML-Schema

Sprache zur Beschreibung der Struktur und Einschränken des Inhalts der Dokumente in erweiterbarer Auszeichnungssprache

[QUELLE: ISO 25720:2009, 4.32]

Nationaler Anhang NB (informativ)

Literaturhinweise

DIN EN ISO 13606 (alle Teile), *Medizinische Informatik — Kommunikation von Patientendaten in elektronischer Form*

English Version

Genomics informatics —
Omics Markup Language (OML)
(ISO 21393:2021)

Informatique génomique —
Langage de balisage Omics (OML)
(ISO 21393:2021)

Genomik-Informatik —
Omics Auszeichnungssprache (OML)
(ISO 21393:2021)

This European Standard was approved by CEN on 29 September 2020.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.



EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

CEN-CENELEC Management Centre: Rue de la Science 23, B-1040 Brussels

European foreword

This document (EN ISO 21393:2021) has been prepared by Technical Committee ISO/TC 215 "Health informatics" in collaboration with Technical Committee CEN/TC 251 "Health informatics" the secretariat of which is held by NEN.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by February 2022, and conflicting national standards shall be withdrawn at the latest by February 2022.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN shall not be held responsible for identifying any or all such patent rights.

Any feedback and questions on this document should be directed to the users' national standards body/national committee. A complete listing of these bodies can be found on the CEN websites.

According to the CEN-CENELEC Internal Regulations, the national standards organizations of the following countries are bound to implement this European Standard: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and the United Kingdom.

Endorsement notice

The text of ISO 21393:2021 has been approved by CEN as EN ISO 21393:2021 without any modification.

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 OML specification	6
4.1 Specification requirements and OML positioning	6
4.2 OML Structure	6
4.3 OML DTD and XML Schema	7
5 OML development process	7
6 Figures	8
Annex A (informative) Reference works	28
Bibliography	45

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 215, *Health informatics*, Subcommittee SC 1, *Genomics informatics*, in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/TC 251, *Health informatics*, in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

In this post genomic era, the management of health-related data is becoming increasingly important to both omics research and omics-based medicine.^[1] Informational approaches to the management of clinical, image and omics data are beginning to have as much worth as basic, bench top research. In the current electronic world, there are multiple different types of data for healthcare as shown in Figure 1. Besides, nowadays there are many kinds of omics data around the world awaiting effective utilization for human health. The development of data format and message standards to support the interchange of clinical omics data is necessary. Omics data includes omics sequence, sequence variation and other expression data, proteomics data, molecular network, etc. As an entry point, this document focuses on the data exchange.

In the present circumstances, omics is expected to be a key to understand human response to external stimuli such as any kinds of alien invasions, therapies, and the environmental interactions.^[2] Bacterial infection is an example of alien invasion, and the responses to the infections are different among the individuals. According to the therapy, the side effects to a drug are different among the patients. These responses are also different in various environments. As a result of recent explosive amount of these omics researches, the huge amounts of experimental data have been accumulating in many databases in various types of data formats. These data are waiting to be used in drug discovery, clinical diagnosis, and clinical researches.

The Markup Language is a set of symbols and rules for their use when doing a markup of a document.^[3] The first standardized markup language was ISO 8879 on Generalized Markup Language (SGML)^[4] which has strong similarities with troff and nroff text layout languages supplied with Unix systems. Hypertext Markup Language (HTML) is based on SGML.^[5] Extensible Markup Language (XML) is a pared-down version of SGML, designed especially for Web documents.^[6] XML acts as the basis for Extensible HTML (XHTML)^[7] and Wireless Markup Language (WML)^[8] and for standardized definitions of system interaction such as Simple Object Access Protocol (SOAP).^[9] By contrast, text layout or semantics are often defined in a purely machine-interpretable form, as in most word processor file formats^[10].

Markup Language for the biomedical field, based on XML, has been in development for several decades to enhance the exchange data among researchers. Bioinformatic Sequence Markup Language (BSML),^[11] Systems Biology Markup Language (SBML),^[12] Cell Markup Language (Cell ML),^[13] and Neuro Markup Language (Neuro-ML)^[14] are examples of markup languages. Polymorphism Mining and Annotation Programs (PolyMAPr)^[15] is centric on SNP and tries to achieve mining, annotation, and functional analysis of public database as dbSNP,^[16] CGAP,^[17] and JSNP^[18] through programming. ISO 25720 Genomic Sequence Variation Markup Language (GSVML) is the first standardized ML for clinical genomic sequence variation data exchange.

The purpose of Omics Markup Language (OML) is to provide a standardized data exchange format for omics in human health.

The recent expansion in omics research has produced large quantities of data held in many databases with different formats. Standardization of data exchange is necessary for managing, analysing and utilizing these data. Considering that omics, especially transcriptomics, proteomics, signalomics and metabolomics, has significant meaning in molecular-based medicine and pharmacogenomics, the data exchange format is key to enhancing omics-based clinical research and omics-based medicine.

Recently, informational approaches have become more important to both omics research and omics-based medicine. The management of omics data is as critical as basic research data in this new era. There are many kinds of omics data around the world, and the time has come to effectively use this omics data for human health. To use this data effectively and efficiently, standards should be developed to permit the interoperable interchange of omics data globally. These standards should define the data format as well as the messages that would be used to interchange and share this data globally.

OML is a base frame of all kinds of clinical omics data. Each omics category will be introduced as a specific add on component part. As an instance, Whole Genome sequence Markup Language will be

a specific add on component part for whole genome sequence data, and Genomic Sequence Variation Markup Language will be a specific add on component part for genomic sequence variation data.

To utilize the internationally accumulated omics data, standards for the interchange of omics data should be defined. These standards should define a data format and exchange messages. Markup Language is a reasonable choice to address this need. As for omics data message handling, Health Level Seven®¹⁾ Clinical Genomics Work Group^[19] has summarized clinical use cases for general omics data. The OML project has contributed to these efforts. Additionally, this work incorporated use cases based on the Japanese millennium project.^[20] Based on these contexts and investigations, this document elucidates the needs and the requirements for OML and after then proposes the specification of OML for the international standardization based on the elucidated needs and the requirements.

1) Health Level Seven (HL7) is the registered trademark of Health Level Seven International. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of the product named.

Genomics informatics — Omics Markup Language (OML)

1 Scope

This document is applicable to the data exchange format that is designed to facilitate exchanging omics data around the world without forcing changes of any database schema.

This document specifies the characteristics of OML from the following perspectives.

From an informatics perspective, OML defines the data exchange format based on XML. This document gives guidelines for the specifications of the data exchange format, but this document excludes the database schema itself.

From a molecular side of view, this document is applicable to all kinds of omics data, while this document excludes the details of the molecules (e.g., details of genomic sequence variations or whole genomic sequence). This document is also applicable to the molecular annotations including clinical concerns and relations with other omics concerns.

From an application side of view, this document is applicable to the clinical field including clinical practice, preventive medicine, translational research, and clinical research including drug discovery. This document does not apply to basic research and other scientific fields.

From a biological species side of view, this document is applicable to the human health-associated species as human, preclinical animals, and cell lines. This document does not apply to the other biological species.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <http://www.electropedia.org/>

3.1

actor

something or someone who supplies a stimulus to the system

Note 1 to entry: Actors include both humans and other quasi-autonomous things, such as machines, computer tasks and systems.

[SOURCE: ISO 25720:2009, 4.1]

3.2

allele

gene that is found in one of two or more different forms in the same position in a chromosome

3.3
bioinformatic sequence markup language
BSML

extensible language specification and container for bioinformatic data

[SOURCE: ISO 25720:2009, 4.2]

3.4
cancer genome anatomy project
CGAP

genomic expression data collected for various tumorigenic tissues in both humans and mice

Note 1 to entry: CGAP also provides information on methods and reagents used in deriving the genomic data

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

3.5
codon

sequence of three nucleotides which together form a unit of genetic code in a DNA or RNA molecule

3.6
dbSNP

database of *single nucleotide polymorphisms* (3.29) provided by the US National Center for Biotechnology Information (NCBI)

Note 1 to entry: Available at <https://www.ncbi.nlm.nih.gov/SNP/>.

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

3.7
digital imaging and communications in medicine
DICOM

standard in the field of medical informatics for exchanging digital information between medical imaging equipment (such as radiological imaging) and other systems, ensuring interoperability

[SOURCE: ISO 25720:2009, 4.6]

3.8
DNA sequence variation

differences of DNA sequence among individuals in a population

Note 1 to entry: DNA sequence variation implies polymorphism 3.25.

[SOURCE: ISO 25720:2009, 4.8]

3.9
document type definition
DTD

document that contains formal definitions of all of the data elements in a particular type of *hypertext markup language* 3.13, *standard generalized markup language* (3.29), or *extensible markup language* (3.36) document

[SOURCE: ISO 25720:2009, 4.9]

3.10
entry point

reference point that designate the class(es) from which the messages begin for the domain

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

3.11

exon

part of a gene that will encode a part of the final mature RNA produced by that gene after *introns* (3.16) have been removed by RNA splicing

3.12

genomic sequence variation markup language

GSVML

standard for data exchange of genomic sequence variation data

3.13

hypertext markup language

HTML

set of markup symbols or codes inserted in a file intended for display in a browser

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

3.14

international classification of diseases

ICD

diagnose coding system for epidemiology, health management and clinical purposes

Note 1 to entry: ICD-10 is the 10th revision and ICD-11th is the 11th revision.

Note 2 to entry: available at <https://icd.who.int/>.

3.15

clinical omics sub-information model for international classification of diseases

clinical omics sub-information model for ICD

iCOS

sub-information model aiming to enhance the representation ability of ICD-11 contents model with covering omics information as an add-on part.

Note 1 to entry: Add-on sub-information model to enhance the representation ability of ICD-11 contents model to cover omics information.

3.16

intron

nucleotide sequence within a gene that is removed by RNA splicing during maturation of the final RNA product

3.17

joint photographic experts group

JPEG

compression technique for images

[SOURCE: ISO 25720:2009, 4.13]

3.18

markup language

ML

set of symbols and rules for their uses when doing a markup of a document

[SOURCE: ISO 25720:2009, 4.15]

3.19

microarray gene expression markup language

MAGE-ML

data format for describing information about DNA-array based experiments and gene expression data

3.20

neuro markup language

neuro-ML

markup language (3.18) for describing models of neurons and networks of neurons.

[SOURCE: ISO 25720:2009, 4.16]

3.21

nroff

unix text-formatting program that is a predecessor of the Unix *troff* (3.33) document processing system

[SOURCE: ISO 25720:2009, 4.17]

3.22

omics

field of study in biology ending in -omics

Note 1 to entry: It includes, but is not limited to, genomics, proteomics, and metabolomics.

3.23

pharmacogenomics

branch of pharmaceutics aiming to develop rational means to optimize drug therapy, with respect to the patient's genotype

3.24

polymorphism mining and annotation programs

PolyMAPr

programs for *polymorphism* (3.25) database mining, annotation, and functional analysis

[SOURCE: ISO 25720:2009, 4.19]

3.25

polymorphism

variation in the sequence of DNA among individuals

Note 1 to entry: Polymorphism implies single nucleotide polymorphism (3.29) and short tandem repeat polymorphism (3.32) .

[SOURCE: ISO 25720:2009, 4.20]

3.26

RNA markup language

RNAML

data format for exchanging RNA information

3.27

systems biology markup language

SBML

markup language (3.18) for simulations in systems biology

[SOURCE: ISO 25720:2009, 4.21]

3.28

standard generalized markup language

SGML

markup language (3.18) for document representation that formalizes markup and frees it of system and processing dependencies

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