ISO 11238:2018

ISO TC 215/WG 6

Secretariat: ANSI

Health informatics – Identification of medicinal products – Data elements and structures for the unique identification and exchange of regulated information on substances

Contents

~~**Foreword** 7~~

~~**Introduction** 7~~

~~**1 Scope** 10~~

~~**2 Normative references** 10~~

~~**3 Terms and definitions** 10~~

~~**4 Symbols and abbreviated terms** 25~~

~~**5 Description of the information modelling principles and practices** 28~~

~~5.1 General considerations 28~~
~~5.2 Conceptual overview diagrams 28~~
~~5.3 Section high-level diagrams 29~~
~~5.4 Detailed diagrams 29~~
~~5.5 Relationships between classes 30~~
~~5.6 Notes 32~~
~~5.7 Attributes 32~~
~~5.8 Message exchange format 33~~
~~5.9 Conformance terminology and context as it relates to the ISO 11238 and ISO/TS 19844 33~~

~~**6 Requirement** 34~~

~~6.1 General 34~~
~~6.2 Concepts required for the unique identification and description of substances 34~~
~~6.3 Concepts required for the description of specified substances 35~~

6.3.1 Relationship between Substances and Specified Substance Groups ... 37
6.4 Naming of substances ... 38
6.5 Requirements for unique identifiers ... 39
6.6 Existing identifiers and molecular structure representation ... 40
7 Type of Substances ... 40
7.1 General ... 40
7.2 Element sets common to multiple types of substances ... 40
7.2.1 Structure ... 40
7.2.2 Isotope ... 40
7.2.3 Modification ... 41
7.2.4 Reference information ... 42
7.2.5 Source material ... 43
7.2.6 Taxonomy ... 44
7.2.7 Authentication of Herbal Drugs ... 44
7.2.8 Substance codes ... 45
7.3 Chemical substances ... 45
7.4 Protein substances ... 46
7.5 Nucleic acid substances ... 49
7.6 Polymer substances ... 50
7.7 Structurally diverse substances ... 51
7.8 Mixture ... 54
8 Defining specified substances ... 56
8.1 General ... 56
8.2 Specified Substance Group 1 ... 56
8.3 Specified Substance Group 2 ... 59
8.4 Specified Substance Group 3 ... 61
8.5 Specified Substance Group 4 ... 62
8.5.1 General ... 62
8.5.2 Specified Substance Group 4 Name ... 63
8.5.3 Grade ... 63
8.5.4 Use of Analytical data ... 64
8.5.5 Manufacturing ... 64
8.5.6 Version and Specification ... 65
Annex A (informative) Existing identifiers and molecular structure representations ... 69
A.1 Identifiers ... 69
A.1.1 General ... 69
A.1.2 CAS Registry numbers ... 69

~~A.1.3 InChI and InChIKey 69~~

~~A.1.4 EC Number 69~~

~~A.1.5 UNII 70~~

~~A.1.6 ASK Number 70~~

~~A.1.7 EV Code 70~~

~~A.2 Molecular structure representations 71~~

~~A.2.1 General 71~~

~~A.2.2 Molfile 71~~

~~A.2.3 SMILES 72~~

~~A.2.4 InChI 72~~

~~A.2.5 CDX and CDXML file format 72~~

~~Bibliography 73~~

Foreword 8

Introduction 9

1 Scope 11

2 Normative references 11

3 Terms and definitions 11

4 Symbols and abbreviated terms 24

5 Description of the information modelling principles and practices 28

5.1 General considerations 28

Figure 1 — Legend for colour coding of model classes 28

5.2 Conceptual overview diagrams 28

Figure 2 — Example conceptual overview diagram 29

5.3 Section high-level diagrams 29

Figure 3 — Example high-level diagram 29

5.4 Detailed diagrams 29

Figure 4 — Example detailed description diagram 30

5.5 Relationships between classes 30

Figure 5 — Association 30

Figure 6 — Multiplicity 30

Figure 7 — Inheritance/generalization 31

Figure 8 — Display of Inherited Attributes 31

5.6 Notes 32

Figure 9 — Note used as a comment on a diagram 32

5.7 Attributes 32

5.8 Message exchange format 32

5.9 Conformance terminology and context as it relates to ISO 11238 and ISO/TS 19844 33

6 Requirements 33

6.1 General 33

6.2 Concepts required for the unique identification and description of substances 33

Figure 10 — High-level information model of substances 35

6.3 Concepts required for the description of specified substances 35

Figure 11 — High-level Substance - Specified substance information model 37

6.3.1 Relationship between Substances and Specified Substance Groups 37

Figure 12 — Parent Substance and Specified Substances Groups relationships of Triamcinolone acetonide 38

6.4 Naming of substances 38

Figure 13 — Information model for substance names 39

6.5 Requirements for unique identifiers 39

6.6 Existing identifiers and molecular structure representation 39

7 Types of substances 40

7.1 General 40

7.2 Element sets common to multiple types of substances 40

7.2.1 Structure 40

7.2.2 Isotope 40

Figure 14 — Information model for structure and isotope 41

7.2.3 Modification 41

Figure 15 — Information model for modification 42

7.2.4 Reference information 42

Figure 16 — Information model for reference information 43

7.2.5 Source material 43

Figure 17 — Information model for source material information 44

7.2.6 Taxonomy 44

7.2.7 Authentication of Herbal Drugs 44

7.2.8 Substance codes 44

Figure 18 — Information model for substance code **45**

7.3 Chemical substances 45

Figure 19 — Information model for the chemical substance **46**

7.4 Protein substances 46

Figure 20 — Information model for the protein substance **48**

7.5 Nucleic acid substances 48

Figure 21 — Information model for the nucleic acid substance, high level view **49**

7.6 Polymer substances 49

Figure 22 — Information model for the polymer substance **50**

7.7 Structurally diverse substances 50

Figure 23 — Information model for the structurally diverse substance **52**

7.8 Mixture 52

Figure 24 — Information model for the mixture **54**

8 Defining specified substances **54**

8.1 General 54

8.2 Specified Substance Group 1 55

Figure 25 — High level information model for the Specified Substance Group 1 ..... **56**

8.3 Specified Substance Group 2 57

Figure 26 — High level information model for the Specified Substance Group 2 ..... **58**

Figure 27 — Extended manufacturing information model for the Specified Substance Group 2 ..... **59**

8.4 Specified Substance Group 3 59

Figure 28 — Information model for the Specified Substance Group 3 ..... **60**

8.5 Specified Substance Group 4 60

8.5.1 General 60

Figure 29 — High level information model for the Specified Substance Group 4 ..... **61**

8.5.2 Specified Substance Group 4 Name 61

8.5.3 Grade 61

8.5.4 Use of analytical data 62

8.5.5 Manufacturing 62

Figure 30 — Information model for manufacturing **63**

8.5.6 Version and specification 63

Figure 31 — Specified Substance Group 4 information model/view of specification **64**

Figure 32 — Detailed Information model for the Specified Substance Group 4 ..... **66**

Annex A (informative) Existing identifiers and molecular structure representations 67
A.1 Identifiers 67
A.1.1 General 67
A.1.2 CAS Registry numbers 67
A.1.3 InChI and InChIKey 67
A.1.4 EC Number 67
A.1.5 UNII 67
A.1.6 ASK Number 68
A.1.7 EV Code 68
A.2 Molecular structure representations 68
A.2.1 General 68
A.2.2 Molfile 68
A.2.3 SMILES 69
A.2.4 InChI 69
A.2.5 CDX and CDXML file format 70
Bibliography 71

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. ~~www.iso.org/directives~~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. ~~www.iso.org/patents~~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 on 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 WTO principles in the Technical Barriers to Trade (TBT) see the following URL: ~~Foreword - Supplementary information~~Foreword - Supplementary information.

This document was prepared by ISO/TC 215, *Health informatics*.

This second edition cancels and replaces the first edition ISO 11238:2012[2], which has been technically revised.

Introduction

This document was developed in response to a worldwide demand for internationally harmonized specifications for medicinal products. It is one of a group of five standards and four technical specifications which together provide the basis for the unique identification of medicinal products. The group of standards and technical specifications comprises:

ISO 11615[3], *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated medicinal product information*

ISO 11616[4], *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated pharmaceutical product information*

ISO 11238, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on substances*

ISO 11239[5], *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on pharmaceutical dose forms, units of presentation, routes of administration and packaging*

ISO 11240[6], *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of units of measurement*

ISO/TS 19844, *Health informatics — Identification of medicinal products — Implementation guidelines for data elements and structures for the unique identification and exchange of regulated information on substances*

ISO/TS 20440[7], *Health informatics — Identification of Medicinal Products — Implementation guide for ISO 11239 data elements and structures for the unique identification and exchange of regulated information on pharmaceutical dose forms, units of presentation, routes of administration and packaging*

ISO/TS 20443[8], *Health informatics — Identification of Medicinal Products — Implementation guide for ISO 11615 data elements and structures for the unique identification and exchange of regulated Medicinal Product information*

ISO/TS 20451[9], *Health informatics — Identification of Medicinal Products — Implementation guide for ISO 11616 data elements and structures for the unique identification and exchange of regulated pharmaceutical product information*

These standards for the identification of medicinal products (IDMP) support the activities of medicines regulatory agencies worldwide by jurisdiction. These include a variety of regulatory activities related to development, registration and life cycle management of medicinal products, as well as pharmacovigilance and risk management.

To meet the primary objectives of the regulation of medicines and pharmacovigilance, it is necessary to reliably exchange medicinal product information in a robust and reliable manner. The IDMP standards therefore support the following interactions:

— between one medicine regulatory agency and another, e.g. European Medicines Agency to the US Food and Drug Administration (FDA), or vice versa; and between the European Medicines Agency and the National Competent Authorities in the EU, vice versa;

— between pharmaceutical companies and medicine regulatory agencies, e.g. "Pharma Company A" to Health Canada;

— between the sponsor of a clinical trial to a medicine regulatory agency, e.g. "University X" to the Austrian Agency for Health and Food Safety (AGES);

— between a medicine regulatory agency and other stakeholders, e.g. UK Medicines and Health Care Products Regulatory Agency (MHRA) to the National Health Service (NHS);

— between medicine regulatory agencies and worldwide-maintained data sources, e.g. the Pharmaceutical and Medical Device Agency (PMDA) and the organization responsible for assigning substance identifiers.

Unique identifiers produced in conformance with the IDMP standards will support applications for which it is necessary to reliably identify and trace the use of medicinal products and the ingredients within medicinal products.

This document provides a structure that enables the assignment and maintenance of unique identifiers for all substances in medicinal products. This document sets out the general rules for defining and distinguishing substances, and provides a high-level model for substances and specified substances to support the organization and capturing of data.

It is anticipated that implementation will use the ISO/TS 19844 and HL7 messaging (see 5.8) to deliver a strong, non-semantic unique identifier for every substance present in a medicinal product. It is anticipated that a single maintenance organization will be responsible for the generation of global identifiers for every substance and that such an organization would retain the defining elements upon which the substance identifier was based. At the specified substance level, a more regional approach may be necessary because of the proprietary nature of much of the information.

The use of the identifier is essential for the description of substances in medicinal products on a global scale. This document does not involve developing nomenclature for substances or specified substances, but common and official substance names in current use can be mapped to each identifier.

Ingredients used in medicinal products range from simple chemicals to gene-modified cells to animal tissues. To unambiguously define these substances is particularly challenging. This document defines substances based on their scientific identity (i.e. what they are) rather than on their use or method of production. Molecular structure or other immutable properties, such as taxonomic, anatomical and/or fractionation information, are used to define substances. This document contains five single substance types and a mixture substance class that are sufficient to define all substances. Although it is certainly possible to define or classify substances in other ways, this document uses a minimalistic structured scientific concept approach focusing on the critical elements necessary to distinguish two substances from one another. There are frequently interactions between substances when they are mixed together, but this document has intentionally not included these supramolecular interactions at the substance level because of the variable nature and strength of such interactions. This document also allows for the capture of multiple terms which refer to a given substance and a variety of reference information that could be used to classify substances or relate one substance to another.

In addition to the substance level, this document also provides elements for the capture of further information on substances that make up the defining characteristics of specified substances, such as grade, manufacturer, manufacturing information and specifications, and also to capture information on substances that are frequently combined together in commerce but are not strictly a medicinal product. At the specified substance level, four groups of elements provide information essential to the tracking and description of substances in medicinal products.

The basic concepts in the regulatory and pharmaceutical standards development domain use a wide variety of terms in various contexts. The information models presented in this document depict elements and the relationship between elements that are necessary to define substances. The terms and definitions described in this document are to be applied for the concepts that are required to uniquely identify, characterize and exchange information on substances in regulated medicinal products.

The terms and definitions adopted in this document are intended to facilitate the interpretation and application of legal and regulatory requirements, but they are without prejudice to any legally binding document. In case of doubt or potential conflict, the terms and definitions contained in legally binding documents prevail.

In this document, “% (V/V)” is used in place of “% volume fraction”.

Health informatics – Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on substances

CAUTION — This document uses colour. This should be taken into consideration when printing.

1 Scope

This document provides an information model to define and identify substances within medicinal products or substances used for medicinal purposes, including dietary supplements, foods and cosmetics. The information model can be used in the human and veterinary domain since the principles are transferrable. Other standards and external terminological resources are referenced that are applicable to this document.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/TS 19844~~:—[1]~~:2018, *Health informatics — Identification of medicinal products (IDMP) — Implementation guidelines for ISO 11238 for data elements and structures for the unique identification and exchange of regulated information on substances*

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 ~~http://www.iso.org/obp~~http://www.iso.org/obp

— IEC Electropedia: available at ~~http://www.electropedia.org/~~http://www.electropedia.org/

3.1
adjuvant
component that potentiates the immune response to an antigen and/or modulates it towards the desired immune response

3.2
active marker
constituent or groups of constituents of a (herbal) Substance (fresh), Herbal Drug, Herbal preparation or herbal medicinal product which are of interest for control purposes and are generally accepted to contribute to therapeutic activity

Note 1 to entry: Active markers are not equivalent to analytical or signature markers that serve solely for identification or control purposes.

3.3
allergen
material of concern used as ingredient or in a device capable of stimulating a type-I hypersensitivity or allergic reaction in atopic individuals

Note 1 to entry: In this document the definition is specified to a molecule (substance) capable of inducing an immunoglobulin E (IgE) response and/or a Type I allergenic reaction.

~~[1] To be published. Stage at time of publication ISO/PRF TS 19844:2018~~

3.4
allergoids
allergen extracts chemically modified (e.g. by formaldehyde or glutaraldehyde) in order to reduce allergenicity while maintaining immunogenicity

Note 1 to entry: Formaldehyde and glutaraldehyde react with primary amino groups in the polypeptide chain of the allergen leading to intramolecular and intermolecular cross-linked high-molecular-weight allergen polymers; in this way conformational IgE epitopes should be destroyed while the linear T-cell epitopes remain unaffected.

3.5
analytical data
set of elements to describe and capture methods and reference material used to determine purity, potency or identity in a specified substance

3.6
analytical marker
constituent or groups of constituents that serve for analytical purposes

Note 1 to entry: Active markers are not equivalent to analytical or signature markers that serve solely for identification or control purposes.

3.7
ATC Code
Anatomical Therapeutic Chemical Classification code
substance classification code
code used for the classification of drugs

Note 1 to entry: It is controlled by the WHO Collaborating Centre for Drug Statistics Methodology (WHOCC)[2].

Note 2 to entry: This pharmaceutical coding system divides drugs into different groups according to the organ or system on which they act and/or their therapeutic, pharmacological and chemical properties. Each bottom-level ATC code stands for a pharmaceutically used substance or a combination of substances in a single indication (or use). This means that one drug can have more than one code: Acetylsalicylic acid, for example, has A01AD05 as a drug for local oral treatment, B01AC06 as a platelet inhibitor, and N02BA01 as an analgesic and antipyretic. On the other hand, several different brands share the same code if they have the same active substance and indications.

3.8
CAS Index name
Chemical Abstracts Service Index name
CAS Registry name
identifier that usually identifies a single substance

Note 1 to entry: For further explanations see subclause A.1.2.

3.9
CAS Registry Number
CAS number[3]
unique numerical identifier of a substance in the CAS Registry system

Note 1 to entry: For further explanations see subclause A.1.2.

3.10
chemical bond

[2] https://www.whocc.no/atc/structure_and_principles/
[3] https://www.cas.org/content/chemical-substances/faqs

condition that occurs when forces acting between two atoms or groups of atoms lead to the formation of a stable discrete molecular entity

3.11
chemical substance
type of substance that can be described as a stoichiometric or non-stoichiometric single molecular entity and is not a protein, nucleic acid or polymer substance

Note 1 to entry: Chemical substances are generally considered "small" molecules which have associated salts, solvates or ions and may be described using a single definitive or representative structure.

3.12
chiral substance
substance whose molecular structure is not superimposable on its mirror image

3.13
co-crystals
homogenous (single phase) crystalline structures made up of two or more components in a definite stoichiometric ratio where the arrangement in the crystal lattice is not based on ionic bonds

3.14
component
substance which is part of a mixture and that defines a multi-substance material at the Specified Substance Group 1 level

EXAMPLES Dimethicone and silicon dioxide are components of simethicone. Human insulin and protamine are the components in human insulin isophane.

Note 1 to entry: Components are used to describe a multi-substance material.

3.15
composition stoichiometry
quantitative relationships between the chemical elements or moieties that make up a substance

EXAMPLES Disodium hydrogen phosphate heptahydrate and disodium hydrogen phosphate dihydrate are defined as different substances because they differ in composition stoichiometry.

3.16
configuration
method for indicating the three-dimensional arrangement of atoms at a stereogenic carbon, phosphorous, sulfur centre or stereocenter

3.17
constituent
substance present within a Specified Substance or a parent substance

Note 1 to entry: Constituents can be impurities, degradants, extraction solvents, vehicles, active markers or signature substances, parent substances or single substances mixed together to form a multi-substance material.

Note 2 to entry: Constituents shall have an associated role and amount at the Specified Substance Group 1 information model. Constituent specifications shall be used to describe components as well as limits on impurities or related substances for a given material.

EXAMPLE The substance, triamcinolone acetonide is the parent (constituent) substance of the Specified Substance Group 1 substance, triamcinolone acetonide, micronized.

Note 3 to entry: Constituent component is part of a mixture belonging to a homologous group of individual components, described as parent substances for the manufacture of an allergenic extract.

3.18
controlled vocabulary
finite set of values that represent the only allowed values for a data item

Note 1 to entry: The allowed values can be codes, text or numeric.

[SOURCE: CDISC Clinical Research Glossary V10.0, 2016, modified][16]

3.19
copolymer
polymer with more than one type of structural repeat unit linked through covalent bonds

Note 1 to entry: Copolymers are obtained by copolymerization or sequential polymerization of two or more different monomers. Copolymers can be random, statistical, alternating, periodic, block, cross, graft or mixed.

3.20
critical process parameter
process parameter whose variability has an impact on a critical quality attribute and therefore should be monitored or controlled to ensure the process produces the desired quality

Note 1 to entry: A manufacturing parameter is considered "critical" and necessary for production of Substance or Specified Substance e.g. inclusion of chromatographic step for removal or reduction of impurities, viruses.

Note 2 to entry: The critical process is tied to the Production Method type.

3.21
cytokine
small protein released by cells that has a specific effect on the interactions between cells, on communications between cells or on the behaviour of cells

3.22
degree of polymerization
average number of monomers or repeat units in a polymeric block or chain

Note 1 to entry: Applies to both homopolymers and block copolymers where it refers to the degree of polymerization within a block.

3.23
diverse origin
substances that are not isolated together or the result of the same process

3.24
drug extract ratio
ratio of the quantity of the (herbal) substance (fresh), or herbal drug to the quantity of the resulting herbal preparation

3.25
enhancer
cis-acting sequence of DNA that increases the utilization of some eukaryotic promoters and which can function in either orientation and in any location (upstream or downstream) relative to the promoter

3.26
extract ratio for allergens
extraction ratio indicating the relative proportions (m/V) of allergenic source materials and solvents

Note 1 to entry: This ratio is a minimal requirement for allergens for which there are not enough patients to determine the total allergenic activity *in vivo* or *in vitro*.

3.27
extraction solvents
solvents which are used for the extraction process

3.28
fraction
distinct portion of material derived from a complex matrix, the composition of which differs from antecedent material

Note 1 to entry: This concept is used to describe source material and is recursive in that a subsequent fraction can be derived from an antecedent fraction.

EXAMPLE Serum immunoglobulins to polyclonal IgG is an example of recursive fractionation.

3.29
gene
basic unit of hereditary information composed of chains of nucleotide base pairs in specific sequences that encodes a protein or protein subunit

3.30
gene element
individual element within a gene such as a promoter, enhancer, silencer or coding sequence

3.31
glycosylation
enzymatic process that links saccharides or oligosaccharides to substances

3.32
glycosylation type
significant differences in glycosylation between different types of organisms

Note 1 to entry: This distinguishes the pattern of glycosylation across organism types, e.g. human, mammalian and avian. The glycosylation type is a defining element when a glycosylated protein exists as a substance.

3.33
grade
set of specifications indicating the quality of a substance or specified substance

3.34
harvesting
process of collecting a (herbal) substance (fresh) or parts of botanical material from the field or process of collecting viral or bacterial material from its production/manufacturing site

3.35
homeopathic stocks
substances, products of preparations used as starting materials for the production of homeopathic preparations.

Note 1 to entry: A stock is usually one of the following: a mother tincture or a glycerol macerate, for raw materials of botanical, zoological or human origin, or the substance itself, for raw materials of chemical or mineral origin.

3.36
homopolymer
polymer containing a single structural repeat unit

