INTERNATIONAL STANDARD

ISO 11238

Second edition
2018-07

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

Informatique de santé — Identification des produits médicaux — Eléments de données et structures pour l'identification unique et l'échange d'informations réglementées sur les substances

Reference number
ISO 11238:2018(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

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

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

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.

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 19844:2018, *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

— IEC Electropedia: available at 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.

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)[1)].

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[2)]
unique numerical identifier of a substance in the CAS Registry system

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

1) https://www.whocc.no/atc/structure_and_principles/

2) https://www.cas.org/content/chemical-substances/faqs

3.10
chemical bond
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

EXAMPLE 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

EXAMPLE 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

3.37
isotope
variants of a chemical element that differ by atomic mass, having the same number of protons and differing in the number of neutrons in the nucleus

Note 1 to entry: Radionuclides or nuclides with a non-natural isotopic ratio are shown in the structural representation with the nuclide number displayed. Natural abundance isotopes are represented by an elemental symbol without a nuclide number.

EXAMPLE ^{13}C refers to a carbon atom that has an atomic mass of 13.

3.38
manufactured item
qualitative and quantitative composition of a product as contained in the packaging of the Medicinal Product

Note 1 to entry: A Medicinal Product may contain one or more manufactured items. In many instances the manufactured item is equal to the pharmaceutical product. However, there are instances where the manufactured item(s) undergo a transformation before being administered to the patient (as the pharmaceutical product) and the two are not equal.

3.39
manufacturer
organization that holds the authorization for the manufacturing process

Note 1 to entry: In this document the definition refers to a company responsible for the manufacturing of the substance

3.40
manufacturing
process of production for a substance or medicinal product from the acquisition of all materials through all processing stages

Note 1 to entry: The critical process, critical process steps, starting and processing materials and critical production parameters are included.

3.41
material
entity that has mass, occupies space and consists of one or more substances

3.42
medicinal product
pharmaceutical product or combination of pharmaceutical products that can be administered to human beings (or animals) for treating or preventing disease, with the aim/purpose of making a medical diagnosis or to restore, correct or modify physiological functions

Note 1 to entry: A Medicinal Product may contain in the packaging one or more manufactured items and one or more pharmaceutical products. In certain regions, a Medicinal Product may also be defined as any substance or combination of substances which may be used to make a medical diagnosis.

3.43
microheterogeneity
substances isolated together that contain minor differences in structure between essentially identical substances that are isolated/source material (e.g. sequence heterogeneity) and/or post-translational modification such as glycosylation

Note 1 to entry: Microheterogeneity is not a defining characteristic of substances but can be a defining one at the specified substance group 1 information level, e.g. differences in glycans.

Note 2 to entry: Microheterogeneity consists of variability in the type of glycosylation (biantennary, triantennary), extent of glycosylation at a given site (site occupancy), sequence heterogeneity due to polymorphism in source material, translation errors or variable proteolytic processing or other.

3.44
mixture
type of polydisperse substance that is a combination of single substances isolated together or produced in the same synthetic process

Note 1 to entry: Single substances of diverse origin that are brought together and do not undergo a chemical transformation as a result of that combination are defined as multi-substance materials (Specified Substance Group 1) and not as mixture.

EXAMPLE 1 Gentamicin is defined as a mixture substance of Gentamicin C1, Gentamicin C1A, Gentamicin C2, Gentamicin C2A and Gentamicin C2B. Glyceryl monoesters are defined as mixture substances of two single substances which differ in the position of esterification. Simethicone, which consists of dimethicone and silicon dioxide, is not defined as a mixture substance since these are diverse materials brought together to form a multi-substance material.

EXAMPLE 2 Glyceryl monoesters could be defined as a mixture of two single substances which differ in the position of esterification.

Note 2 to entry: Mixture could be used for a homologous group of structurally diverse single substances used as starting materials in order to prepare an allergen extract. The extract is further described by using the class 'Source Material', element group 'Fraction Description' (allergen preparation) obtained from the structurally diverse single substances (starting materials) as parent substances. This substance (allergen extract) is the result of the same (synthetic) process and hence the extract is considered as a mixture substance.

3.45
moiety
entity within a substance that has a complete and continuous molecular structure

EXAMPLE The strength of a medicinal product is often based on what is referred to as the active moiety of the molecule, responsible for the physiological or pharmacological action of the drug substance. To avoid ambiguity, the free acid and/or free base should be used as the moiety upon which strength is based.

Note 1 to entry: The active moiety of a stoichiometric or non-stoichiometrical substance molecule is considered that part of the molecule that is the base, free acid or ion molecular part of a salt, solvate, chelate, clathrate, molecular complex or ester.

3.46
molecular formula
chemical formula that shows the total number and kind of atoms in a molecule indicating atomic proportional ratios (the numerical proportions of atoms of one type to those of other types) which is a way of expressing information about the proportions of atoms that constitute a particular chemical compound, using a single line of chemical element symbols, numbers, and sometimes other symbols

Note 1 to entry: The molecular formula could contain other symbols, such as parentheses, dashes, brackets, and plus (+) and minus (–) signs. These are limited to a single typographic line of symbols, which may include subscripts and superscripts. A chemical formula contains no words (e.g. Glucose is represented as $C_6H_{12}O_6$).

Note 2 to entry: Molecular formulas are written in accordance with the Hill system/ Hill notation such that the number of carbon atoms in a molecule is indicated first, the number of hydrogen atoms next, and then the number of all other chemical elements subsequently, in alphabetic order. When the formula contains no carbon, all the elements, including hydrogen, are listed alphabetically.

Note 3 to entry: Inorganic acids and metal salts are shown without charges or bonds: $HClO_4$ and $KMnO_4$ respectively. If metal salts of inorganic acids include several metals, the symbols for the metals are shown in alphabetic order, e.g. K_2NaPO_4.

3.47
molecular formula by moiety
way of describing of the molecular formula of a stoichiometric or non-stoichiometric substance existing of two or more moieties, the molecular formula of each moiety shall be described separated by a dot

Note 1 to entry: The molecular formula of the chemical salt Amlodipine besilate is described as $C_{20}H_{25}ClN_2O_5$. $C_6H_6O_3S$.

Note 2 to entry: In non-cyclic linear structures like sodium nitroprusside: $Na_2[Fe(CN)_5(NO)].2H_2O$, a non-cyclic structure is constructed in the following order:

a) symbol of the central atom placed on the left;

b) ionic ligands with cations first then anions;

c) neutral ligands.

3.48
molecular fragment
portion of a molecule that has one or more sites of attachment to other fragments or moieties

Note 1 to entry: Molecular fragments are used in the description of polymers to represent substituents and in structural modifications to a substance.

3.49
molecular structure
unambiguous representation of the arrangement of atoms

Note 1 to entry: For the purposes of defining substances, the three-dimensional conformations are not captured. Individual conformations or conformers of substances would only be captured in either a general sense for proteins (i.e. denatured) or when a given rotation about a single bond is restricted in such a way that the two different conformers are isolatable from each other and do not interconvert at room temperature (e.g. substituted biphenyls).

Note 2 to entry: This representation should be generally translatable into a graphical representation.

3.50
molecular weight
mass of one molecule of a homogenous substance or the average mass of molecules that comprise a heterogeneous substance, which is derived from the molecular structure or the molecular formula

Note 1 to entry: It is calculated as the sum of the mass of each constituent atom multiplied by the number of atoms of that element in the molecular formula. The unified atomic mass unit is the unit of molecular weight and includes the type of molecular weight (g/mol).

Note 2 to entry: For stoichiometric chemicals, the molecular weight is calculated from the molecular formula using standard masses for each of the elements. The molecular mass refers to the complete structure or a moiety or a fragment.

Note 3 to entry: The unified atomic mass unit or Dalton is the unit of molecular weight. The type of molecular weight should always be captured.

Note 4 to entry: For polymers, there are several different types of molecular weight (weight average, number average, etc.).

Note 5 to entry: For a substance not described in a Pharmacopoeia a mass spectrum may be provided to substantiate the calculated molecular weight.

3.51
monodisperse substance
single substance that is homogeneous in molecular weight, that is, it does not have a distribution of different molecular weight chains within the total mass

3.52
multi-substance material
single substances or specified substances of diverse origin that are brought together and do not undergo a chemical transformation

EXAMPLE Materials such as human insulin isophane, simethicone, aluminium lakes, nicotine polacrilex, and phosphate buffered saline are all multi-substance ingredients.

Note 1 to entry: Each substance part of the multi-substance material is a parent substance of the multi-substance material and should be registered first. The multi-substance material is captured at the Specified Substance Group 1 information level.

3.53
multi-substance starting material
mixture of multi-substances of which another substance will be prepared

EXAMPLE Aqueous extract of a mixture of homologous group of allergenic source material.

3.54
nucleic acid substance
type of substance that can be defined by a linear sequence of nucleosides typically linked through phosphate or phosphate-like diester bonds

Note 1 to entry: The type of nucleic acid substance, e.g. ribonucleic acid (RNA) and deoxyribonucleic acid (DNA), is also identified. Oligonucleotides and gene elements, e.g. promoters, enhancers, coding sequences and silencers, are defined as nucleic acid substances.

3.55
official name
name given by an official registration authority/organization

3.56
organism
individual living entity anatomical origin that can react to stimuli reproduce, grow, and maintain homeostasis

3.57
part
entity of anatomical origin and location of source material within an organism

Note 1 to entry: Entity is a thing with distinct and independent existence.

3.58
pharmaceutical product
qualitative and quantitative composition of a Medicinal Product in the dose form approved for administration in line with the regulated product information

Note 1 to entry: In many instances, the pharmaceutical product is equal to the manufactured item. However, there are instances where the manufactured item shall undergo a transformation before being administered to the patient (as the pharmaceutical product) and the two are not equal.

3.59
physical form
physical state, either gas, liquid or solid, and the type of organization for solid matter

Note 1 to entry: Solids can be either crystalline or amorphous and can show polymorphism. Amorphous material is characterized by the absence of distinct reflections in the X-ray powder diffraction (XRPD) pattern. Polymorphism which is ability of crystalline materials to exist in more than one form, can also be captured.

3.60
polydisperse substance
single substance containing multiple related entities

Note 1 to entry: Polydisperse substances include polymers, mixture and structurally diverse material isolated from a single source. Chemical substances, proteins and nucleic acids with defined sequences are not described as polydisperse substances.

3.61
polydispersity
measure of the range of molecular masses in a polymer substance

Note 1 to entry: The dispersity of polymers is typically calculated by the ratio of weight average molecular weight to number average molecular weight

3.62
polymer connectivity
copolymer sequence type (Polymer connectivity) can either be random, statistical, alternating, periodic, block, graft or mixed

3.63
polymer substance
type of polydisperse substance that contains structural repeat units linked by covalent bonds

Note 1 to entry: Monodisperse proteins and nucleic acids with defined sequences shall not be defined using the polymer substance elements.

3.64
post-translational modification
modification of a protein that typically occurs *in vivo* during or after translation

Note 1 to entry: Post-translational modification is described within the structural representation and not as a modification of a protein.

3.65
potentization
process by which dilutions and triturations are obtained from homeopathic stocks in accordance with a homeopathic manufacturing procedure: this means successive dilutions and successions, or successive appropriate triturations, or a combination of the two processes

Note 1 to entry: The number of potentization steps defines the degree of dilution; for example, "D3", "3DH" or "3X" means three decimal potentization steps (diluted 1:1 000), and "C3" or "3CH" means three centesimal potentization steps (diluted 1:1 000 000).

3.66
processing material
type of material essential to the manufacturing process that is not incorporated into the resultant material but can be present in the resultant material as constituent

3.67
protein sequence
order and identity of amino acids within a protein or peptide

Note 1 to entry: Protein sequences will be represented by single letter Dayhoff codes and listed from the N-terminal to the C-terminal.

3.68
protein substance
type of substance with a defined sequence of alpha-amino acids connected through peptide bonds

Note 1 to entry: A protein consists of one or more chains with a length of more than 40 amino acids. A peptide is defined as a linear sequence consisting of 2 to 40 amino acid residues.

Note 2 to entry: Synthetic peptides and proteins with defined sequences, recombinant proteins and highly purified proteins extracted from biological matrices are described as protein substances. Sites of glycosylation, disulphide linkages and glycosylation type (e.g. fungal, plant, anthropoid, avian mammalian, human) are defining elements of protein substances, when known. A graphical molecular structure is also included in the definition of all peptides of 40 amino acid residues or less.

Note 3 to entry: The absolute configuration at the a-carbon atom of the a-amino acids is designated by the prefixed small capital letter D or L to indicate a formal relationship to D- or L-serine and thus to D- or L-glyceraldehyde. The prefix ξ (Greek xi) indicates unknown configuration[17].

3.69
protein sub-unit
linear sequence of amino acid residues connected through peptide bonds

Note 1 to entry: Protein sub-units usually adopt specific structural motives (e.g. β-sheets, α-helical structures) and have specific functions/roles within the protein (e.g. catalytic centre, substrate binding). Sub-units are linked to each other to make up the fully functional protein.

Note 2 to entry: Repeated sub-units in proteins are captured.

EXAMPLE Monoclonal antibodies typically consist of four sub-units.

3.70
resultant material
material that is the result of a manufacturing process

Note 1 to entry: Resultant material may be the starting material of the next process step or the final material or actual specified substance.

3.71
salt
ionic substances formed from the neutralization reaction of an acid and a base

Note 1 to entry: Salts are ionic compounds composed of cations (positive ions) and anions (negative ions).

3.72
signature substance
substance used to be able to follow the quality of the production or extraction process

Note 1 to entry: Signature substance is not necessarily the intended active substance or resultant material. It is a substance used for analytical purpose which is representative for the quality of the manufacturing or extraction process.

3.73
silencer
DNA sequence that suppresses transcription

3.74
single substance
substance that can be described by a single representation or set of descriptive elements

Note 1 to entry: A single substance can be described using one or more of five types of elements: chemical, protein, nucleic acid, polymer or structurally diverse substances.

Note 2 to entry: Racemates and substances with unknown, epimeric or mixed chirality can be defined as single substances because a single structural representation may be generated and the stereochemistry indicated as descriptive text.

3.75
solvate
substance formed through association of a solvent molecule (e.g. water, alcohol) with another moiety

Note 1 to entry: Solvates can be either stoichiometric or non-stoichiometric and are predominately present in the solid form of substances.

Note 2 to entry: Solvates formed with water as solvent are referred to as hydrates.

3.76
source material
material from which a substance is derived, which is defined based on taxonomic and anatomical origins

Note 1 to entry: Source material is used to define structurally diverse, chemical, mixture, polymer and protein substances isolated from biological matrices.

3.77
specification
list of tests, references to analytical procedures, and appropriate acceptance criteria, which are numerical limits, ranges, or other criteria for the tests described. It establishes the set of criteria to which a drug substance should conform to be considered acceptable for its intended use

Note 1 to entry: "Conformance to specifications" means that the drug substance, when tested according to the listed analytical procedures, will meet the listed acceptance criteria. Specifications are critical quality standards that are proposed and justified by the manufacturer and approved by regulatory authorities as conditions of approval.

3.78
specified substance
substance defined by groups of elements that describes multi-substance materials or specifies further information on substances relevant to the description of Medicinal Products

Note 1 to entry: This could include grade, units of measure, physical form, constituents, manufacturer, critical manufacturing processes (e.g. extraction, synthetic or recombinant processes), specification and the analytical methods used to determine whether a substance is in compliance with a specification. There are four different groups of elements that can be used to define a given specified substance and specific relationships between each group of elements.

3.79
starting material
material from which the manufacturing process of the substance starts. This material is a building block which will be (partly) incorporated in the structure of the resultant product

3.80
stereochemistry
relative spatial arrangement of atoms within molecules

3.81
stoichiometric substance
substance that contains moieties in simple integral ratios

Note 1 to entry: Defined composition stoichiometry shall be represented in the structural representation of a given substance. Moieties shall be represented using the lowest common factors such that a fractional representation is avoided. Substances will either be defined as stoichiometric or non-stoichiometric.

Note 2 to entry: Chemicals have defined composition stoichiometry when the ratio of all moieties (ion, counter ion and solvate) can be represented as simple integral ratios.

3.82
structural repeat unit
fundamental descriptor of a polymer typically derived from a monomer that is used to synthesize the polymer

3.83
structurally diverse substance
type of polydisperse substance isolated from a single source that is a complex combination which cannot be described as a mixture of a limited number of single substances

Note 1 to entry: Structurally diverse substances are defined based on immutable properties of a given material. Modifications that irreversibly alter the structure of the material, distinctive physical properties or components subsumed into the material, e.g. a gene in gene therapy substances, are defining elements for structurally diverse substances. Fractions derived from source material (oils/juices and extracts) are also captured in the definition. Protein mixtures containing a large number of diverse sequences such as polyclonal immunoglobulins are defined as structurally diverse substances.

3.84
substance
matter of defined composition that has discrete existence, whose origin may be biological, mineral or chemical

Note 1 to entry: A substance can be a moiety. A moiety is an entity within a substance that has a complete and continuous molecular structure. The strength of a pharmaceutical product is often based on what is referred to as the active moiety of the molecule, responsible for the physiological or pharmacological action of the drug substance. Chemically, the active moiety of a stoichiometric or non-stoichoimetrical substance molecule is considered that part of the molecule that is the base, free acid or ion molecular part of a salt, solvate, chelate, clathrate, molecular complex or ester.

Note 2 to entry: In this document substances are further described as single substances, mixture substances or one of a group of specified substances. Single substances are defined using a minimally sufficient set of data elements divided into five types: chemical, protein, nucleic acid, polymer and structurally diverse. Substances may be salts, solvates, free acids, free bases or mixtures of related compounds that are either isolated or synthesized together. Pharmacopoeial terminology and defining characteristics will be used when available and appropriate. Defining elements are dependent on the type of substance.

Note 3 to entry: Discrete existence refers to the ability of a substance to exist independently of any other substance. Substances can either be well-defined entities containing definite chemical structures, synthetic (e.g. isomeric mixtures) or naturally-occurring (e.g. conjugated oestrogens) mixtures of chemicals containing definite molecular structures, or materials derived from plants, animals, microorganisms or inorganic matrices for which the chemical structure may be unknown or difficult to define.

3.85
substituent
molecular fragment attached to a structural repeat unit of a polymer that typically replaces a hydrogen atom

Note 1 to entry: This information is captured as part of the structural repeat unit when the position of substitution is fully occupied. When occupancy of a site is incomplete, the amount of a substituent is specified as either a fragment or moiety structural modification.

3.86
succession
process of agitating a liquid preparation in the manufacturing of homeopathic medicinal products.

Note 1 to entry: As intermediate step in the process of serial dilution, it is part of the potentization process.

3.87
tautomer
molecular structure capable of interconversion with an isomeric molecular structure that typically involves facile migration of a hydrogen atom between two adjacent atoms

Note 1 to entry: It is anticipated that a single tautomeric form will be associated with each substance and detailed rules will be developed within the implementation guide to indicate the tautomeric form associated with each chemical substance. If individual isomers may be isolated under normal conditions and are known to have distinct molecular properties, they are defined as separate substances [see ISO/TS 19844 (liquid dextrose)].

3.88
taxonomy
scientific organism classification system needed to describe the origin of source material in substances isolated from biological matrices

Note 1 to entry: Taxonomic information is captured to the species level for all polydisperse substances isolated from biological matrices, if such information is available and the source material is consistently derived from the species. Taxonomic family, genus and species along with the taxon author are necessary to identify the source organism. Kingdom, phylum, class and order are also captured when available. Intraspecific information (e.g. subspecies, strain or variety) is captured when the forms exhibit consistent differences in either material content or function.

3.89
unitage
specifications of the amount constituting a unit

3.90
vehicle
excipient used for the preparation of certain homeopathic stocks or for the potentization process

Note 1 to entry: They may include, for example, purified water, ethanol of a suitable concentration, glycerol 85 per cent and lactose monohydrate.

Note 2 to entry: Vehicle description is in reference to homeopathy formulations and not in relation to other vehicles which are carriers or inert media used as solvents (diluents) in medicinally active agent formulations and/or their administration.

4 Symbols and abbreviated terms

NOTE Only general abbreviations are listed. They are used either within ISO 11238 or ISO/TS 19844 since the two documents are regarded as inseparable.

4.1

ACS

American Chemical Society[3)]

4.2

ASK Number

ID of a substance in German “Arzneistoffkatalog” (Pharmaceutical Substance Dictionary)

4.3

BAN

British Approved Name[4)]

4.4

DCF

Dénominations Communes Françaises (French approved drug name)[5)]

4.5

EVcode

EudraVigilance Code (Unique Identifier) used for a substance in the Extended EudraVigilance Medicinal Product Dictionary (XEVMPD)[6)]

4.6

HAB

3) https://www.acs.org/content/acs/en.html

4) https://www.tsoshop.co.uk/Medicine/Pharmacopoeia/&Action=Book&ProductID=9780113230334

5) http://ansm.sante.fr/Mediatheque/Publications/Pharmacopee-francaise-Denominations-communes-et-scientifiques-de-medicaments

6) http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000596.jsp

Homöopathisches Arzneibuch, Amtliche Ausgabe

4.7

HTS

high-throughput sequencing

4.8

INCI

International Nomenclature of Cosmetic Ingredients[7)]

4.9

INN

International Nonproprietary Name [also consider as rINN (recommended International Nonproprietary Name) or pINN (proposed International Nonproprietary Name)][8)]

4.10

JAN

Japanese Approved Name[9)]

4.11

JP

Japanese Pharmacopoeia[10)]

4.12

NDF-RT

National Drug File — Reference Terminology, produced by the U.S. Department of Veterans Affairs, Veterans Health Administration (VHA)[11)]

4.13

NLT

not less than

4.14

NMT

not more than

4.15

OMG

7) http://www.cirs-reach.com/Cosmetic_Inventory/International_Nomenclature_of_Cosmetic_Ingredients_INCI.html

8) http://www.who.int/medicines/services/inn/en/

9) http://jpdb.nihs.go.jp/jan/index.aspx

10) https://www.pmda.go.jp/english/rs-sb-std/standards-development/jp/0019.html

11) https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/

Object Management Group[12)]

4.16

Ph.Eur.

European Pharmacopoeia (Pharmacopée Européenne)[13)]

4.17

UCUM

Unified Code for Units of Measure[14)]

4.18

UML

Unified Modeling Language[15)]

4.19

UNII

Unique Ingredient Identifier. Identifier of a substance in the FDA Global Substance Registration System (G-SRS)[16)]

4.20

USAN

United States Adopted Name[17)]

4.21

USP

United States Pharmacopeia[18)]

4.22

WHO-ATC

World Health Organization – Anatomical Therapeutic Chemical Classification System[19)]

12) http://www.omg.org/

13) https://www.edqm.eu/en/european-pharmacopoeia-9th-edition

14) http://unitsofmeasure.org/

15) http://www.uml.org/

16) https://www.fda.gov/forindustry/datastandards/substanceregistrationsystem-uniqueingredientidentifierunii/

17) https://www.ama-assn.org/about/united-states-adopted-names-council

18) http://www.usp.org/

19) https://www.whocc.no/atc/structure_and_principles/

5 Description of the information modelling principles and practices

5.1 General considerations

The information modelling in this document follows the general principles described in other ISO IDMP Standards and Technical Specifications (e.g. ISO 11615, ISO 11616, ISO/TS 20443) and uses the Unified Modeling Language (UML) which is maintained by the Object Management Group (OMG).

UML has different styles and patterns that may be followed. The use of UML in this document has been kept very simple, using classes, attributes, datatypes, and basic association relationships mostly. Some constructs, such as stereotypes and complex relationships have been avoided for this reason. In addition, colour coding has been used in the diagrams to help visualize groups of associated entities (for example, see Figure 1).

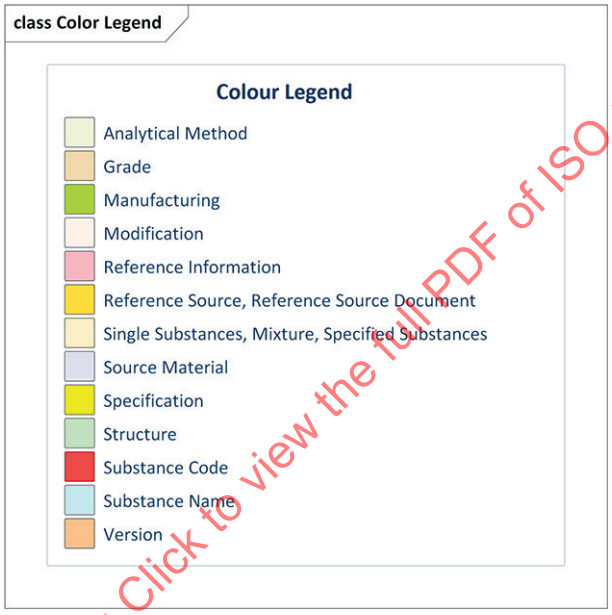

Figure 1 — Legend for colour coding of model classes

The following aims to explain the style that has been followed in this document.

5.2 Conceptual overview diagrams

The conceptual overview diagram provides a framework with which to view the more detailed descriptions of information.

The conceptual overview diagram (see Figure 2) provides a framework with which to view the more detailed descriptions of information. The Substance and Specified Substance high level information models (see Figure 10 and Figure 11) show a single representative class from each information section, related to the core concept (Substance).

Basic cardinalities between the Substance or the Specified Substance and these core classes are shown, but not the detailed entities, relationships or attributes.

NOTE 1 The expressions 'Element Group' and 'Class' have the same meaning in the context of ISO 11238 and ISO/TS 19844.

NOTE 2 The terms cardinality and multiplicity (used by UML) are interchangeable in the context of ISO 11238 and ISO/TS 19844.

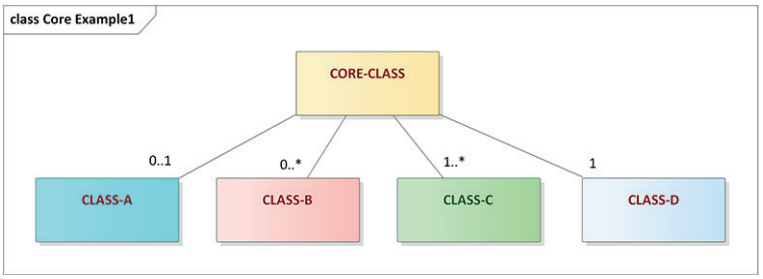

Figure 2 — Example conceptual overview diagram

5.3 Section high-level diagrams

The high-level diagrams (see Figure 3) show all the classes required to describe the information for that section and the conceptual relationships between those classes, with the starting point always as the section's core class.

No attributes or detailed cardinalities are shown in these conceptual diagrams, as their primary purpose is to provide a framework for viewing the detailed diagrams which follow.

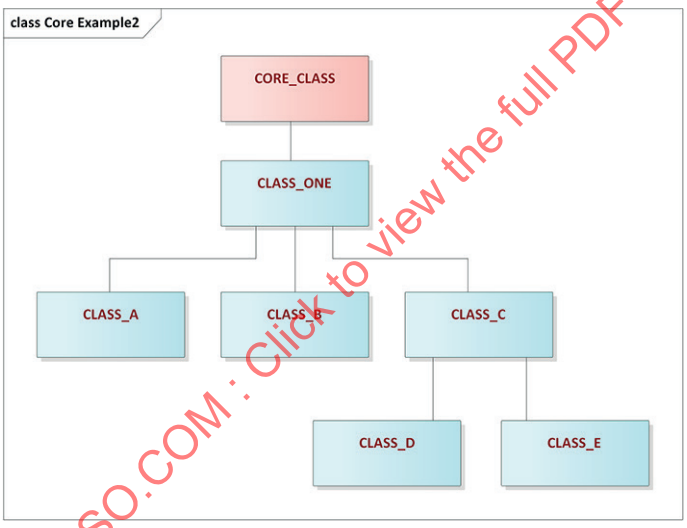

Figure 3 — Example high-level diagram

5.4 Detailed diagrams

The detailed description diagrams (see Figure 4) for each section show all the classes and all the attributes required to describe the information for that section and the detail of the conceptual relationships between those classes.

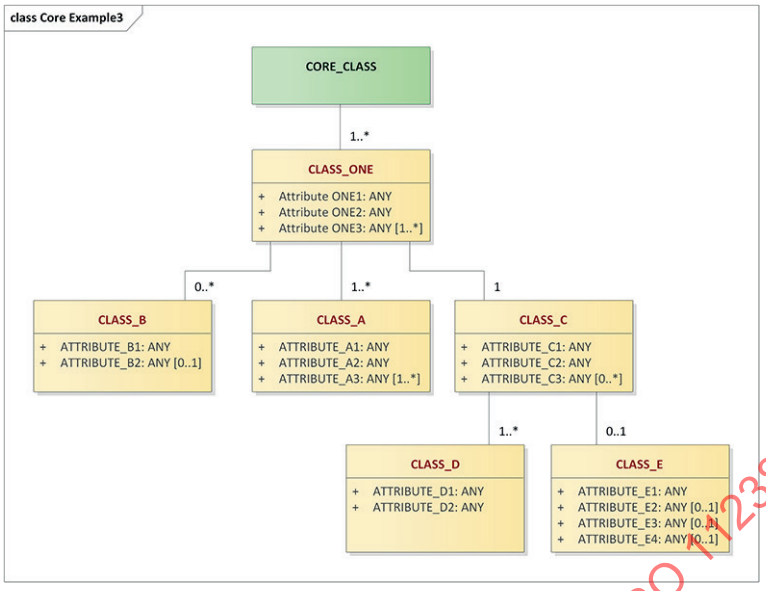

Figure 4 — Example detailed description diagram

5.5 Relationships between classes

Classes related to each other in specific ways. This document uses only a few of the connection features that are possible in UML: association, directed association, reflexive association, multiplicity, aggregation, composition, inheritance/generalization and realization. In the context of this document, the association (Figure 5), multiplicity (Figure 6) and inheritance/generalization (Figure 7).

An association is any logical connection or relationship between classes. For example, MOIETY and AMOUNT may be linked as shown in Figure 5.

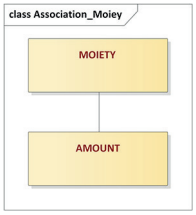

Figure 5 — Association

UML multiplicity is when the number of repetitions of a class in relation to another is shown. For example, one structure may contain zero to many structural representations. The notation 0..* in the diagram means 'zero to many' (see Figure 6).

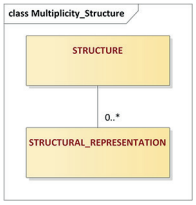

Figure 6 — Multiplicity

Inheritance/generalization (see Figure 7) refers to a type of relationship where one class is a child of another, and assumes the same functionalities of the parent class. For example, the child class Chemical

is a specific type of the parent class Single Substance (which itself is a specific type of the parent class Substance). To show inheritance in a UML diagram, a solid line from the child class to the parent class is drawn using an unfilled arrowhead.

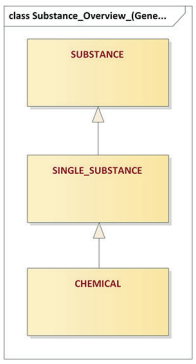

Figure 7 — Inheritance/generalization

NOTE When displaying a Class with attributes, the inherited attributes from parents can be shown (see Figure 8), including parents of parents.

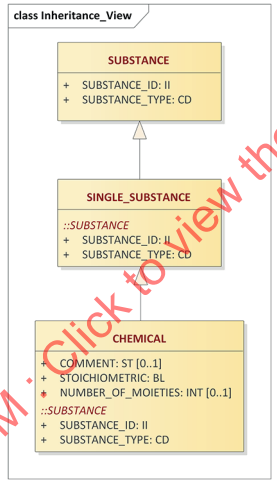

Figure 8 — Display of Inherited Attributes

To keep the model simple, relationships between classes/element groups are generally described as associations having multiplicities, with no further qualification as to the role or type of the association.

Cardinalities on relationships are given in a single direction only: the direction that is away from the Substance/Specified Substance. The rationale for this is that the scope of this document is to describe the Substance/Specified Substance and their associated information. Having these classes always as the source entity avoids describing complex many-to-many cardinalities that might occur in a reverse direction from an entity towards the Substance/Specified Substance.

A cardinality of "1" is synonymous with a cardinality of "1..1".

A cardinality of "1" between entities indicates that the information for that entity shall be specified and that only one set of the entity information shall be given.

A cardinality of "1..*" between entities indicates that the information for that entity shall be specified and that one or more sets of the entity information shall be given.

A cardinality of "0..1" between entities indicates that the information for that entity can be specified and that one set of the entity information can be given.

A cardinality of "0..*" between entities indicates that the information for that entity can be specified and that one or more sets of the entity information can be given.

Some optional entities can be elevated to mandatory if some conditions are met (see 5.9).

See ISO 21090[10] for more information on composition of attributes. An ISO 21090 datatype for the data in each attribute is shown directly in the model and, the text description for each attribute, indicates the form in which data should be specified.

5.6 Notes

Notes are comments in the diagrams and may stand on their own or be linked by a dashed line to the elements they are referring to (see Figure 9).

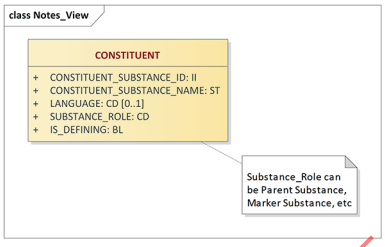

Figure 9 — Note used as a comment on a diagram

5.7 Attributes

Attributes of a class are described using an attribute name in the model. The definition, description and example values for the attribute are given in the text following the model diagram.

An attribute showing no explicit cardinality means that the attribute shall have one value (this is the equivalent to [1...1]).

An attribute showing a cardinality of [1...*] means that the attribute shall have one or more values.

An attribute showing a cardinality of [0...1] means that the attribute may have one value.

An attribute showing a cardinality of [0...*] means that the attribute may have one or more values.

All optional attributes can be conditionally elevated to mandatory if certain conditions are met (see 5.9).

See ISO 21090 for more information on composition of attributes.

5.8 Message exchange format

HL7 messaging standards are widely implemented globally. ISO recognizes HL7 as an accredited partner organization for mutually issuing standards. The first mutually published standard was ISO/HL7 21731[11][12].

HL7 V3 was developed to address the complex requirements of health information technology. The HL7 model-driven methodology is used to develop consensus-based standards for healthcare system interoperability and information exchange. The HL7 Reference Information Model (RIM) is the cornerstone of V3 and the essential model from which all HL7 V3 messages are derived. The RIM defines data content needed in a specific context and provides an explicit representation of the semantic and lexical connections that exist between the information items carried. HL7 V3 messages are based on an XML encoding syntax.

HL7/FHIR (Fast Healthcare Interoperability Resources)[15] is a new specification based on emerging industry approaches, but informed by years of lessons learned and experience gained of requirements, successes and challenges gained through implementing HL7 V2, HL7 V3 and CDA. FHIR was built

using the modern technology of the Internet, particularly RESTful interfaces. It is organized around the concept of "resources" and defines many types of resources that describe the healthcare space. FHIR is rapidly becoming the worldwide standard for health and life sciences data exchange. It has an open licence, a focus on implementation, and a formal maturity process linked to implementation outcomes[18].

The ISO IDMP standards were designed to specify the necessary data elements and associated standards to be used to support unique identifiers. These were developed as an integral part of the IDMP consensus requirements and they are consistent with the HL7 Common Product Model (CPM) V4[13] and the HL7 Structured Product Labeling V8 (SPL)[14]. The IDMP data elements represent a subset of those in the CPM. In addition, the FHIR SubstanceSpecification and MedicinalProduct resources will ensure that FHIR is one of the messaging options for IDMP in the short and long term.

It is anticipated that implementation of ISO 11238 will use HL7 messaging standards. The normative use of HL7 standards will facilitate the integration of IDMP into the broader healthcare community.

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

— *Mandatory*: Defining elements necessary for the unique identification of Substances and Specified Substances per the ISO IDMP standards/technical specifications.

— *Conditional*: Conditional applies to the 'within category' data elements, as applicable, when there are alternative data sources for a given data element(s) to identify a Substance/Specified Substance. Regional implementation of the ISO 11238 and ISO/TS 19844 may elevate the conditional conformance categories to ***'mandatory'*** per regional requirements.

— *Optional*: When listed at the category level (e.g. Specified Substance), optional corresponds to ISO categories or data elements that are not absolutely necessary for the unique identification of Substances/Specified Substances as per ISO 11238. Regional implementation of ISO 11238 and ISO/TS 19844 may elevate the optional conformance categories to 'mandatory' or 'conditional' per regional requirements.

6 Requirements

6.1 General

Substances and specified substances shall be defined in a manner consistent with the elements and relationships present in the figures within this clause and the ISO/TS 19844 implementation guide, which defines these elements and relationships further.

6.2 Concepts required for the unique identification and description of substances

Substances shall be single substances, mixture substances or specified substances.

NOTE The term 'substance' as used below generally refers to a single substance or mixture substance. A specified substance is generally a further specification of a substance that captures information on manufacture, specifications, physical form or multi-substance materials that are components of a medicinal product formulation.

This document defines the concepts required for the unique identification of substances at an international level, whenever such recognition is required. Such identification shall be based on the following principles:

1) a substance shall generally be defined based on what the material is rather than on how it is made or used;

2) a substance shall be defined based on immutable properties independent of physical form, grade or level of purity;

3) substances can be single molecular entities or mixtures of single molecular entities either synthesized or isolated together;

4) to avoid ambiguity and facilitate implementation, a mixture shall be defined as a combination of single substances either synthesized or isolated together;

5) substances shall not be diverse materials brought together to form a medicinal product or multi-substance material.

EXAMPLE 1 Simethicone would not be defined as a substance because it consists of two substances, dimethicone and silicon dioxide, which are of diverse origin and typically not isolated together. Simethicone is defined as a Specified Substance Group 1.

Complex materials from biological matrices and mixtures that cannot be defined or represented by a limited number of chemical structures are defined based on source taxonomy, part and fraction. Materials containing interactions of an indefinite nature and indefinite composition stoichiometry shall not be defined as substances.

NOTE Because of the difficulties in determining the extent, strength and composition stoichiometry of non-covalent interactions, these types of interactions are not taken into account when defining a substance. The only exceptions would be ionic (salt) and solvate (hydrate) interactions of simple chemicals, peptides and well-defined polymers. Materials that contain moieties that interact with polymers, complex matrices or cyclodextrins will typically not be defined as substances, but can be described as components. Simple polymeric salts such as sodium polystyrene sulfonate would be defined as a single substance.

EXAMPLE 2 Nicotine polacrilex is defined as two distinct substances: nicotine and polacrilex. Human insulin isophane would also be defined as two distinct substances: protamine and human insulin. Nicotine polacrilex and human insulin isophane, however, could be defined as single specified substances and are classified as Specified Substance Group 1. Liposomal doxorubicin would be defined as a Specified Substance Group 1 that contains doxorubicin and the components that make up the liposome.

Substances shall be defined using one or more of the following groups of elements:

— chemical;

— protein;

— nucleic acid;

— polymer;

— structurally diverse;

— mixture.

All types of substances shall have the ability to capture official names, synonyms, isotopic and other reference information.

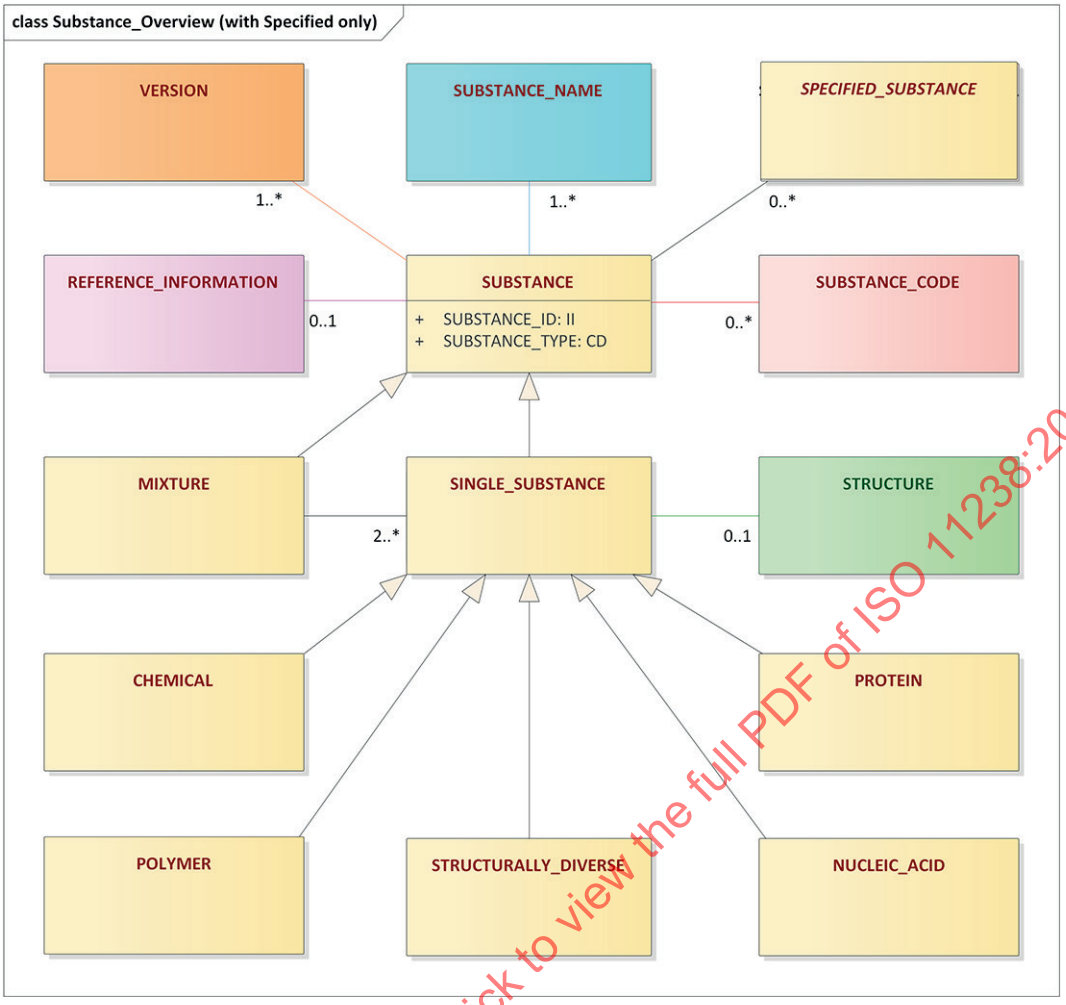

Figure 10 — High-level information model of substances

6.3 Concepts required for the description of specified substances

Specified substances shall include further information for substances and multi-substance materials. A specified substance shall capture more detailed characteristics of single substances or the composition of material that contains multiple substances or different physical forms.

The elements necessary to define specified substances shall be divided into four groups to facilitate implementation.

These groups shall be delineated as follows.

— Group 1: element group 'Constituent' (Constituent Substance: including components for material containing multiple substances, parent substance, marker substance and extraction solvents for herbals, allergenic extracts), element group 'Characteristic Attribute' for e.g. herbals, vaccines, homeopathic substances and plasma-derived substances, element group 'Physical Form' and any physical property that is essential for defining the specified substance (e.g. size of liposomes) and 'Fraction Description'. Together with the element group 'Amount' and in some cases element group 'Attribute Parameters' the 'Constituent Substance' or 'Characteristic Attribute' can be quantified. See Figure 25.

NOTE The element group 'Characteristic Attribute' is meant to capture all kinds of properties and additional information of a substance. These properties and additional information are not attributes in their own right but are instances of characteristic attributes. Examples of characteristic attributes are e.g. for Herbals: 'Degree of Comminution', 'Process State', 'Wild/Cultivated', 'Growth State', 'Feeding composition' (e.g. zoological source material), 'Harvesting Time', 'Decontamination Process', 'Country of Origin' (also applicable to Plasma-derived substances), 'Geographical Location', 'Storage Time' and 'Storage Condition'. In addition, physical properties like boiling point, triple point, density and solubility are also described by this element group. Detailed information is provided by the annexes of the ISO/TS 19844 Implementation Guide.

— Group 2: limited manufacturing information of the Substance or Specified Substance Group 1 information, overall Production Method Type (e.g. synthetic, extractive, recombinant), high-level Production Method Description, Production System Type (e.g. cell line, plant or animal tissue), Production System (specific cell line) as well as Critical Process Version Number, Version and Version Date; see Figure 26. The Manufacturer is captured by the element group 'Organization'. The attribute Issuer of (Manufacturer) ID is the maintenance organization who keeps track of the Manufacturing information e.g. EMA-OMS[20] service or DUN & Bradstreet to provide a D-U-N-S number[21], which is a worldwide identification number for corporations.

  This high-level manufacturing information may be extended with limited information:

  — recursive manufacturing or fractionation steps such as in case of herbal and homeopathic preparations;

  — modification step such as for allergen extracts;

  — fractionation steps such as in case of plasma-derived substances;

  — starting and processing materials such as in case of chemical and protein substances, see Figure 27.

— Group 3: Substance or Specified Substance Group 1 coupled to the grade and reference source of grade (pharmacopeia, technical). The element group 'Grade' consists of two attributes: 'Grade Type' and 'Grade Name', which is the name of the Substance or Specified substance in accordance with the Pharmacopoeial name of the monograph, see Figure 28.

  NOTE Group 2 is also connected to Group 3. This is important when a manufacturer makes use of the specifications of more than one Pharmacopoeial Grade. Each grade might differ in specifications and therefore an In-house specification should be laid down in order to cover the specifications of all grades.

— Group 4: detailed manufacturing information, constituents (impurities, degradants which are not captured in Group 1), and specifications, see Figure 32.

The relationship between substance and specified substance is shown in Figure 11.

20) https://spor.ema.europa.eu/omswi/#/

21) https://www.dnb.co.uk/duns-number.html

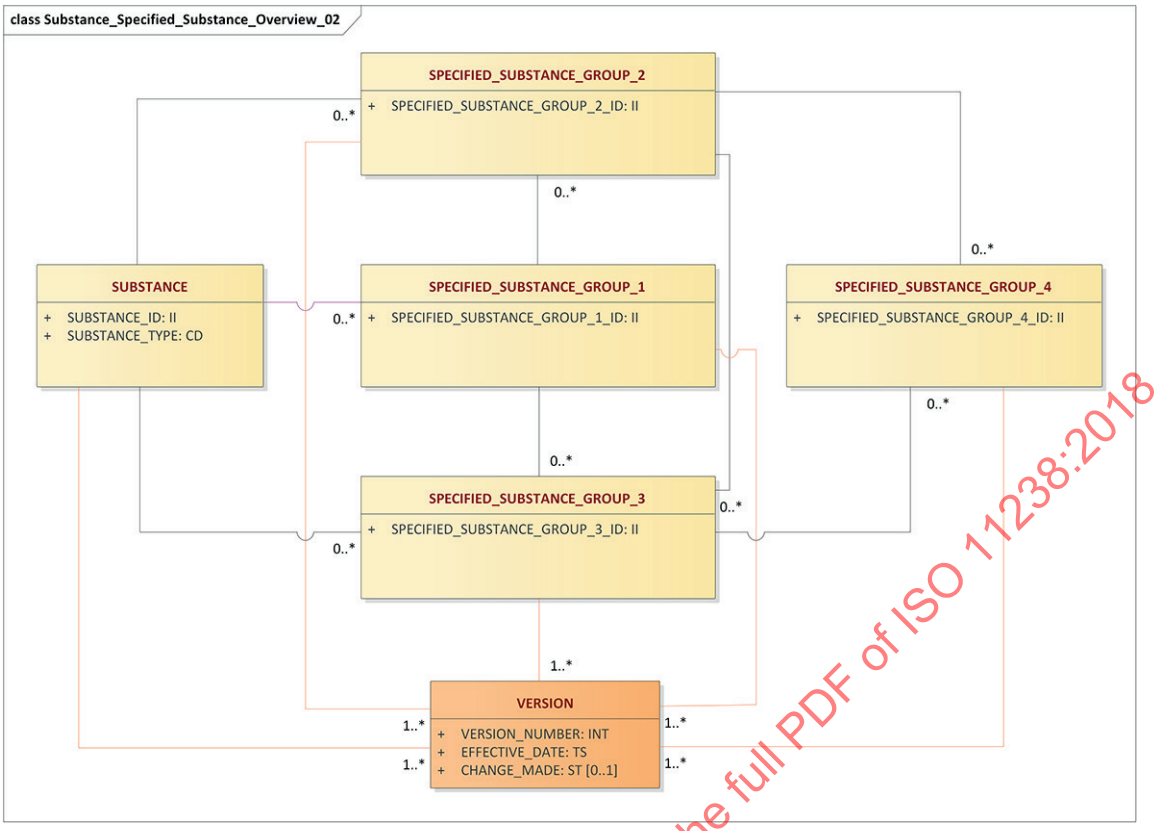

Figure 11 — High-level Substance - Specified substance information model

6.3.1 Relationship between Substances and Specified Substance Groups

Every specified substance is related to its antecedent (parent) substance or specified substance.

EXAMPLE 1 The substance Triamcinolone acetonide is the parent substance of Triamcinolone acetonide, micronized, captured at the Specified Substance Group 1 information level. The particle size can be a defining element and shall be captured at the Specified Substance Group 1 information level, element group 'Characteristic Attribute' together with the amount.

Triamcinolone acetonide, micronized — Company AA is the Specified Substance Group 2 Name of the parent substance Triamcinolone, micronized with reference to a specific manufacturer, Company AA.

Triamcinolone acetonide — Ph.Eur. is the Specified Substance Group 3 Name of the parent substance Triamcinolone acetonide with reference to a specific grade, Ph.Eur. Triamcinolone acetonide — Ph.Eur. shall have a different Specified Substance Group 3 ID from Triamcinolone acetonide — USP.

Triamcinolone acetonide, micronized — Ph.Eur. is the Specified Substance Group 3 Name of the Substance Triamcinolone acetonide with reference to the Ph.Eur. which has an additional specification for its particle size. The 'In house' specification covers both the specifications as laid down in the Ph.Eur. monograph as well as the particle size specification as laid down by the manufacturer.

The parent substance of Triamcinolone acetonide, micronized–Ph.Eur. is Triamcinolone acetonide, micronized captured at the Specified Substance Group 1 information level, which in turn has Triamcinolone acetonide as parent substance, see Figure 12.

For the Substance captured at the Specified Substance Group 1 information level, the Parent (substance) and ID can be described by the element group 'Constituents' with the attributes: 'Constituent Substance Name' and ''Constituent Substance ID', and the 'Substance Role' with the value 'Parent'.

The corresponding Substance Name and ID are captured at the Substance information level and referred to as 'Parent' of the particular specified substance captured at the Specified Substance Group 1 information level.

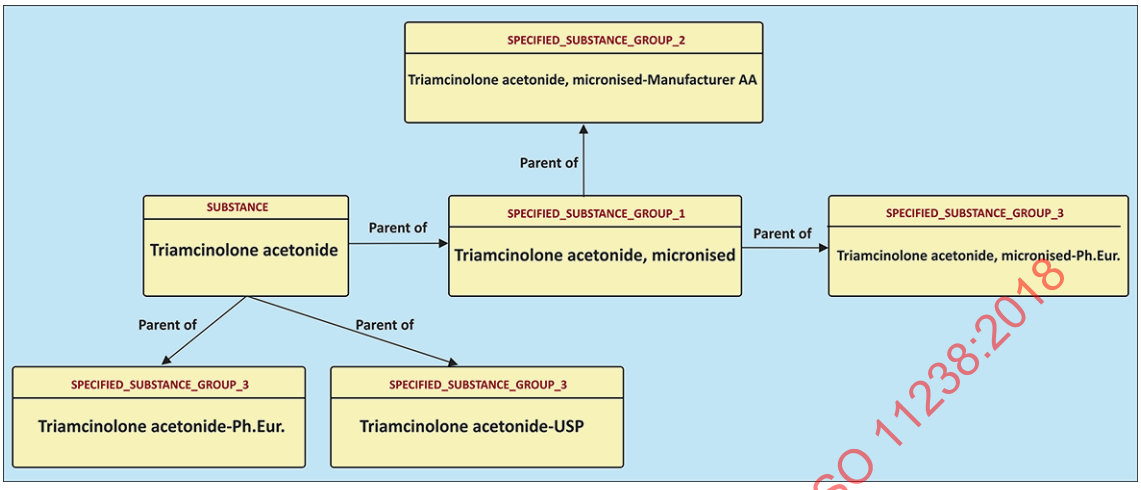

Figure 12 — Parent Substance and Specified Substances Groups relationships of Triamcinolone acetonide

EXAMPLE 2 Simethicone is a multiple substance material whose components are the substances dimethicone and silicon dioxide. Simethicone may encompass a number of Specified Substances Group 1 depending on the type of dimethicone and the particle size and/or surface area of silicon dioxide. The USP or Ph.Eur. monographs cover a broad range of simethicone Specified Substances Group 1 materials.

EXAMPLE 3 For structurally diverse substances the 'Parent Substance ID' and 'Parent Substance Name' are described by the element group 'Source Material', e.g. the (Herbal) Substance (fresh) '*Ginkgo biloba* L., Leaf' is the Parent (substance) of the Herbal Drug '*Ginkgo biloba*, Leaf' (with reference to a Pharmacopoeial monograph), which is in turn the Parent substance of the Herbal preparation '*Ginkgo biloba*, Leaf, Dry Extract all captured at the Substance information level.

6.4 Naming of substances

At least one substance name or company code shall be associated with each substance. The element group 'Substance Name' consists of the following attributes: 'Substance Name', 'Substance Name Type', 'Language', 'Substance Name Domain' and 'Jurisdiction'.

If the name is an official name, the naming authority used shall be identified by the 'Official Name Type' as well as the 'Official Name Status'.

This document shall be neutral with respect to any given systematic or official nomenclature.

NOTE It is anticipated that every substance will have a name in English. Synonyms can be associated with a substance. Translations of English names to other languages can also be accommodated. Language and Jurisdiction will be described using ISO standards.

The information model for the class name is shown in Figure 13.

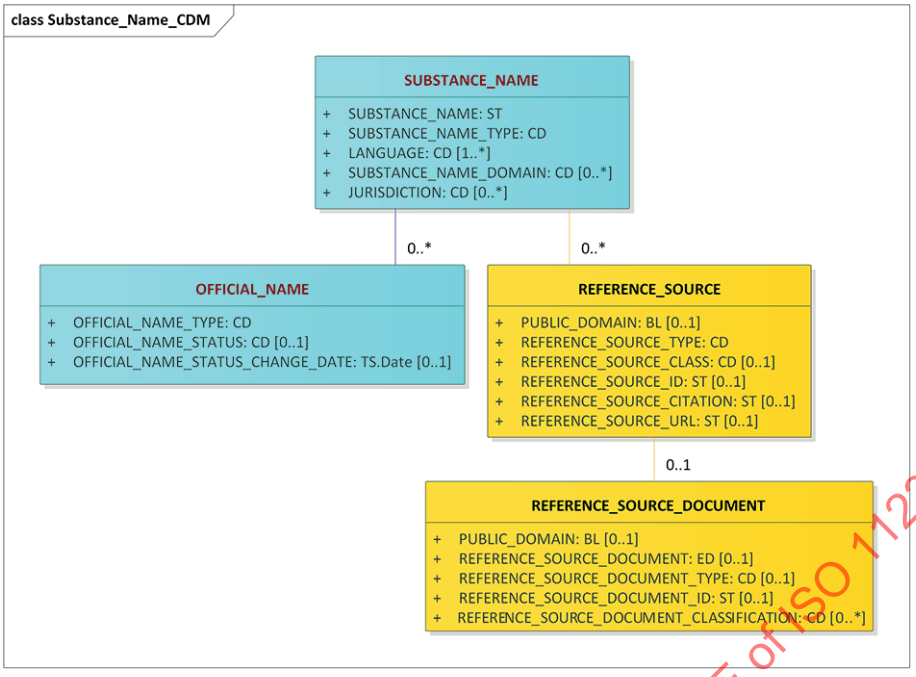

Figure 13 — Information model for substance names

6.5 Requirements for unique identifiers

Each substance and specified substance shall have only one permanently associated unique identifier (called respectively Substance-ID, resp. Specified Substance Group 1, 2, 3 or 4-ID) that shall not indicate the order of submission to the registration process. The Substance ID and Specified Substance Group 1, 2, 3 or 4-ID are per definition unique and therefore global.

The unique identifier shall be non-semantic, random and of fixed length with an internal integrity check.

The unique identifiers can be publicly available when the defining information along with the name or company code is publicly available in a single reference. The use of the identifier shall be royalty free.

A unique identifier shall be assigned to approved and investigational substances, excipients and impurities, solvents, ions, fragments and moieties, each of which shall be defined as a Substance.

NOTE 1 A variety of chemical and biological nomenclature systems have been developed that describe the pharmacological actions of drugs. Functional naming systems such as INN or USAN are valuable in either describing molecular structure or the biological actions of a substance. However, a unique identifier based on such classification systems may result in greater maintenance requirements because classification schemes often require broad ranges of expertise as well as a controlled terminology. Translation is also always a problem with any semantic system.

Once a substance has been defined and assigned a unique identifier it is essential that this identifier be permanently associated with the substance. A substance shall only have one unique identifier. This will necessitate the generation of detailed rules to define substances that will be presented in the ISO/TS 19844 implementation guide.

NOTE 2 A major purpose of the unique identifier is its use in electronic data systems. An identifier of fixed length with an internal integrity check would facilitate the use of the identifier and help identify errors that may occur in data systems that use the identifier.

6.6 Existing identifiers and molecular structure representation

Existing identifiers and molecular structure representations are discussed in Annex A of this document.

7 Types of substances

7.1 General

If it is possible to represent a substance as either a single substance or as a mixture substance, the substance shall be represented as a single substance. All single substances shall be defined as one of the five types: Chemical, Protein, Nucleic acid, Polymer or Structurally Diverse.

NOTE 1 Racemic substances will be represented as single substances because they can be described with a single structural representation and distinguished from chiral substances.

NOTE 2 Some substances have characteristics belonging to more than one of the single substance types e.g. a PEGylated protein would be defined as a protein with the polymer captured as a structural modification. In addition, natural occurring plasma-derived proteins are described by their sequence and the source material. Another example is Heparin, which is defined as a polymer and described by the source material.

7.2 Element sets common to multiple types of substances

7.2.1 Structure

The structure shall contain a sufficient amount of graphical and textual information to define the underlying atoms and the connectivity between atoms as well as the composition ratio of moieties.

Structural representations shall include the complete molecular structure with all known stereochemistry indicated. Molecular fragments and moieties shall also contain structural representations. The structure is a defining element for chemicals, polymers and structural modifications. It should be defined in a consistent and unambiguous manner. The Structural Representation can cover complex substances like proteins and nucleic acid by graphical representation.

7.2.2 Isotope

Radionuclides and other non-naturally abundant nuclides present in a substance shall be defined as isotopes and associated with characteristics using a controlled terminology derived from an internationally recognized reference source.

The presence of isotopes shall also be indicated in structural representations.

Radiopharmaceuticals shall be defined based on the type of the underlying substance and not a type of substance in and of itself.

NOTE Characteristics for each nuclide shall include half-life (energy of emission and type of emission, parent and daughter nuclides could be sourced from a standard reference table).

EXAMPLE Yttrium 90Y ibritumomab tiuxetan would be described as a protein substance. Thyroxine 131I would be described as a chemical substance.

The information model for structure and isotope is shown in Figure 14.

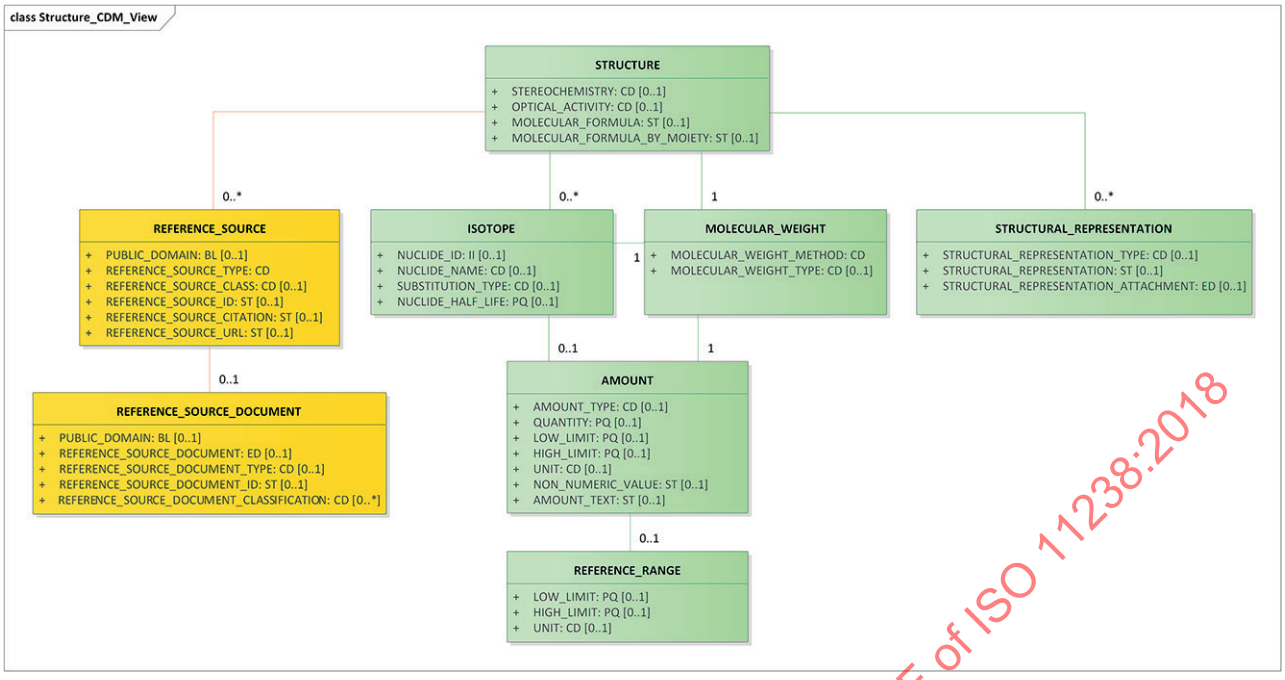

Figure 14 — Information model for structure and isotope

7.2.3 Modification

Irreversible changes in the underlying molecular structure of a substance shall be described as a modification of the antecedent material. Modification of a chemical substance will typically result in a new chemical substance.

NOTE Modifications of chemical substances are inherently captured in the structural representation.

Irreversible changes in the underlying structure of polymers, proteins, nucleic acids, structurally diverse material, mixture substances and Specified Substance Group 1 substances shall be captured using modification elements. The modifications may be physical, chemical, enzymatic etc. Modifications shall be represented as the addition of moieties, substitution of moieties to residues or molecular fragments to the underlying material when definitive structural modifications occur, but the actual position of substitution may be unknown or variable. Physical treatments that result in irreversible structural modifications shall also be captured.

EXAMPLE Process modifications such as thermal curing can be captured as physical modifications. Thermally aggregated albumin is a distinct substance from albumin and albumin aggregated using chemical crosslinking agents. A minimal description of the modification process shall be generated when a definitive structural modification cannot be determined.

The information model for the class modification group is shown in Figure 15.

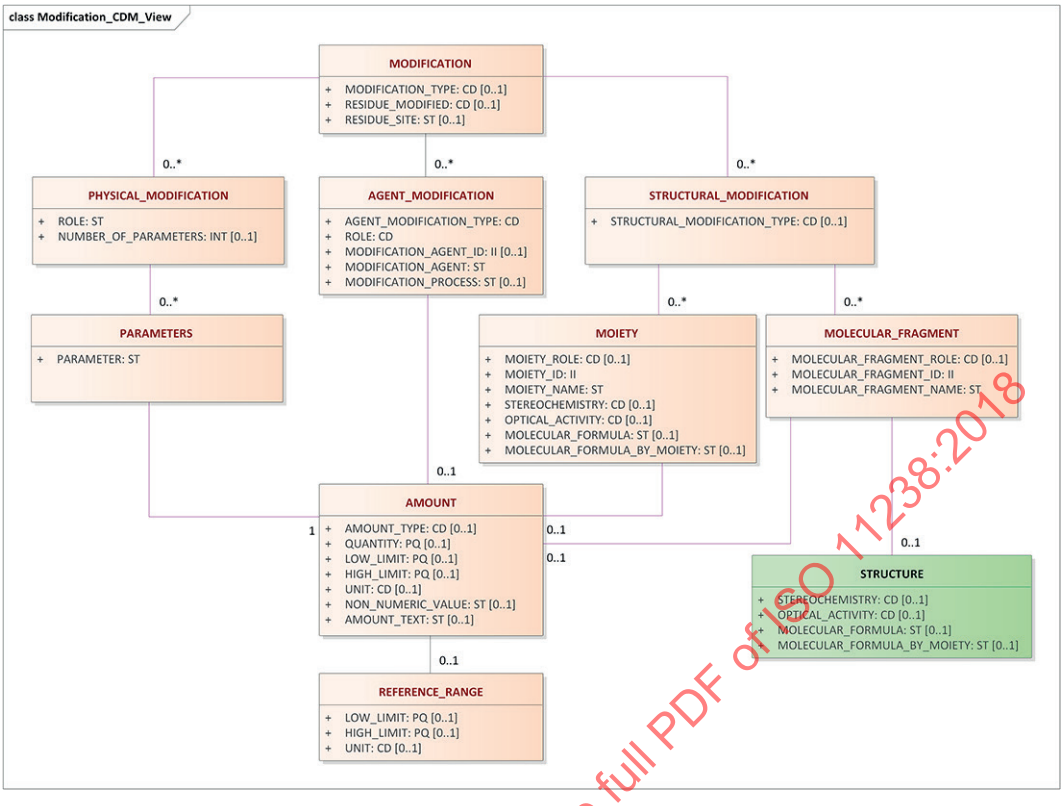

Figure 15 — Information model for modification

7.2.4 Reference information

7.2.4.1 General

Additional types of informative reference information shall be captured for each type of substance in a consistent manner. Such information may include both classification and target information for active substances.

This document does not provide any guidance on the classification of pharmacological effects or the determination of the putative targets for any substance or specified substance. This document does allow for the capture of such information if available and provided. This information shall not affect the generation of a new unique identifier, i.e. it is not defining.

Reference information shall be captured for all types of substances and Specified Substance Group 1.

NOTE Genes from which proteins are derived, target information and codes from code systems also constitute reference information for which this document provides a consistent structure to capture and link to a substance. Classification systems such as the WHO ATC and the United States Veterans Administration NDF-RT, which code classification information for substances, are particularly important. Target information is important for monoclonal and polyclonal antibodies and small molecules directed against specific molecular targets.

The relationships involving reference information are shown in Figure 16.

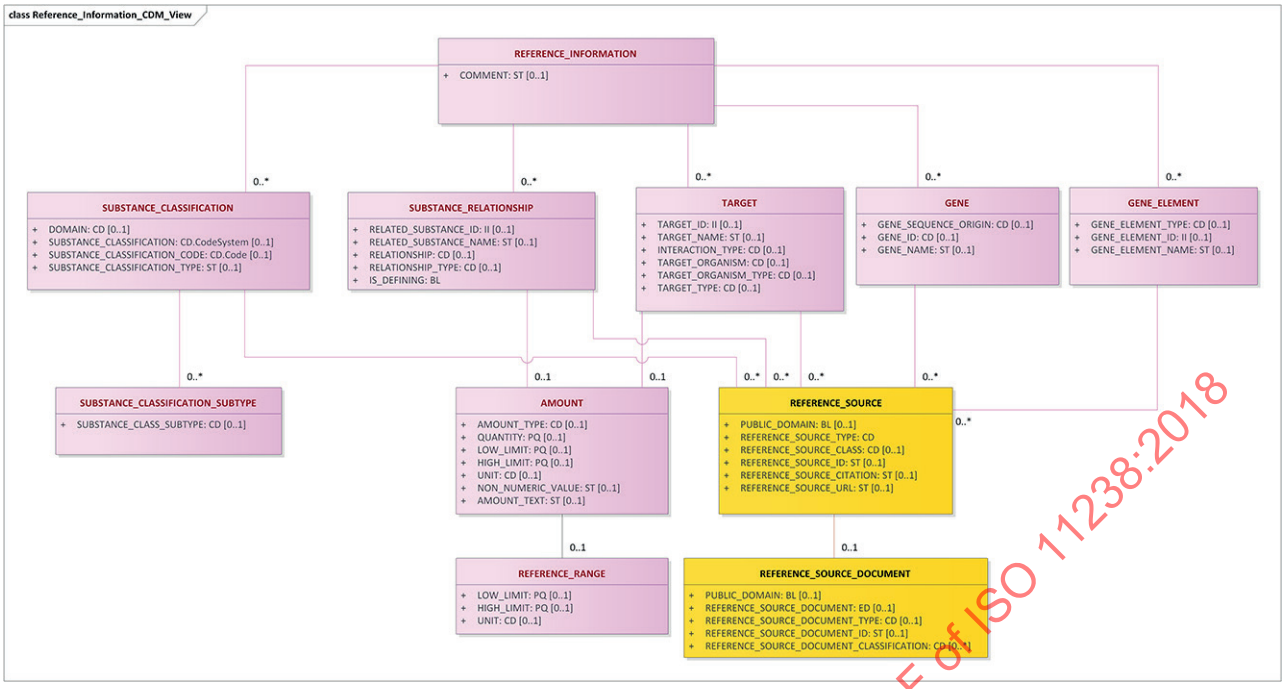

Figure 16 — Information model for reference information

7.2.4.2 Substance classification

Substance classification, although not defining elements for a given substance, can be captured according to the information model presented below. Multiple classifications and variable levels of classification can be captured for a substance.

Classification systems are typically based on molecular structure, chemical properties, pharmacological effects, mechanism of action, therapeutic targets or indication.

Although most classification will be associated with an external classification system, ad-hoc classification of substances may be developed within this terminology as needed.

7.2.5 Source material

Source Material captures the taxonomic and anatomical origins as well as the fraction of a material that can result in or can be modified to form a substance. The Source Material shall be used to define the structurally diverse and polymer substances isolated from biological matrices.

Taxonomic and anatomical origins shall be described using controlled vocabularies as required.

The information model for Source Material is shown in Figure 17.

Fresh plant material, Herbal Drugs and Herbal preparations, including extracts with limited information, will all be captured as Substances, defined using the elements of the structurally diverse substance type. The elements necessary for defining the source material of structurally diverse substances, i.e. organism, part and fraction, are illustrated in Figure 17.

NOTE See ISO/TS 19844:2018, Figure E.3 (Overview of Naming for Herbal Substance, Herbal Drug and Herbal preparation in relation to the information levels).

Human plasma-derived substances, allergens, vaccines and some homeopathic substances, when the homeopathic stock is derived from botanical or zoological material, will also be captured as structurally diverse substances and the source material will be a predominant element group.

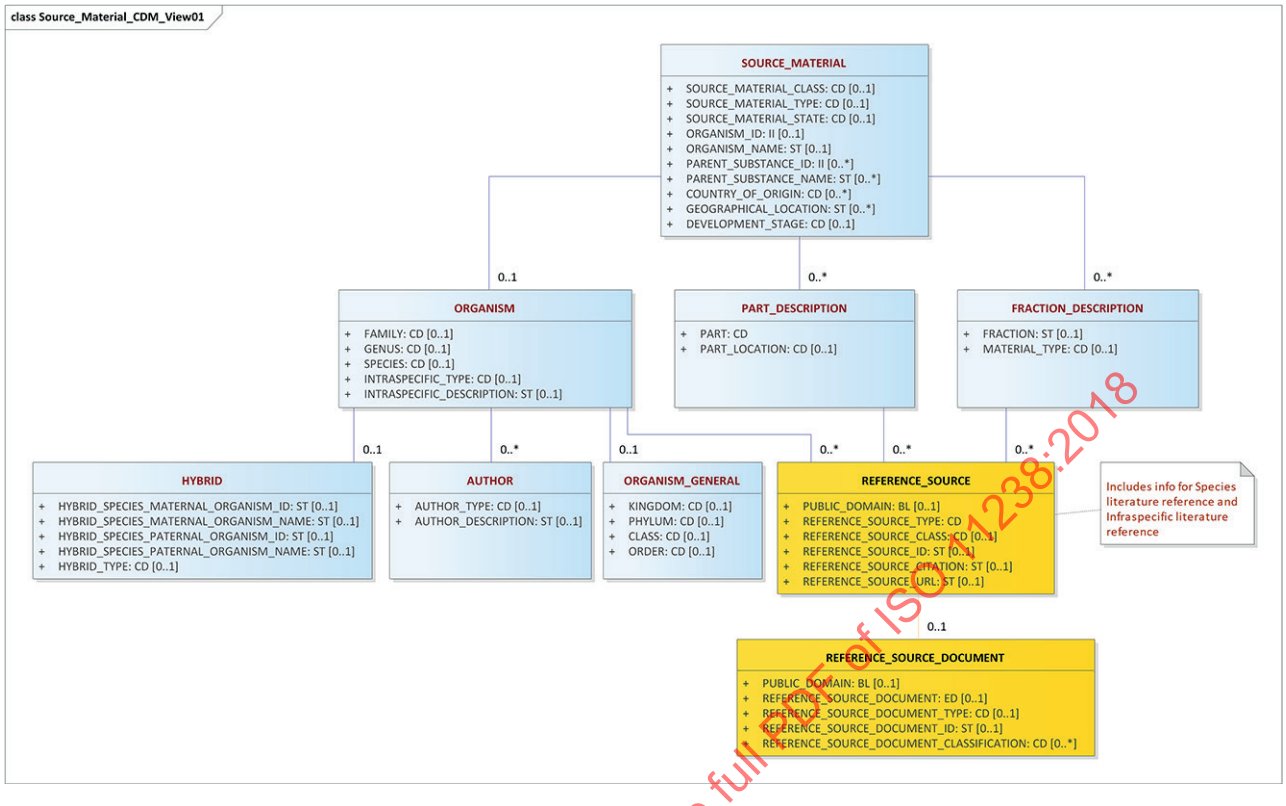

Figure 17 — Information model for source material information

7.2.6 Taxonomy

Taxonomic information shall be captured for polymer, protein, structurally diverse substances, and mixture derived from biological matrices. This document does not provide any guidance on the generation or qualification of taxonomic information. Consistent taxonomic information shall be derived from a limited number of authoritative sources.

Taxonomic information, particularly the scientific name of a medicinal plant, is essential for defining herbal substances.

All scientific names of the Substance (fresh), Herbal Drug and Herbal preparations will be in compliance with an authoritative source which also maintains a list of plant parts and fractions. A controlled vocabulary for medicinal plant taxonomy has been developed and will be maintained by this authoritative source, e.g. Kew Gardens Medicinal Plant Names Services.

7.2.7 Authentication of Herbal Drugs

DNA methods are increasingly being developed and used for the authentication of Herbal Drugs[19], and have already been added to a number of monographs, e.g. in the Chinese Pharmacopoeia. DNA authentication should be added to the Structurally Diverse Substance type to capture the relevant attributes. These can include the main DNA barcodes for particular organisms, captured as defining properties associated with source material for a whole organism.

Looking to the future, recent development in high-throughput sequencing (HTS) techniques are also likely to be adopted as an authentication method and can also be captured or linked to a defining property for source material.

7.2.8 Substance codes

Substance codes and related substance code systems, although not defining elements for a given substance should be captured according to the information model presented below.

Codes typically facilitate mapping and linking of substances to a variety of information sources. All the codes that are captured should be associated with a publicly recognized code system and map directly to a given substance. It should be noted that company codes are not captured in this section but are considered a type of name for a given substance.

EXAMPLE These codes include Chemical Abstract Service (CAS) Registry Numbers, European Inventory of Existing Commercial Chemical Substances (EINECS), European Drug Codes (XEVMPD) and Japanese Drug Codes.

The information model for the substance code is shown in Figure 18.

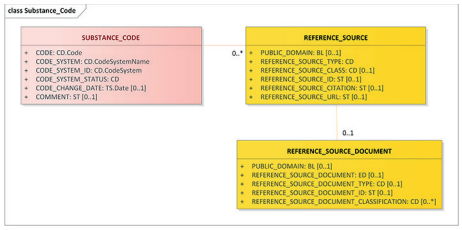

Figure 18 — Information model for substance code

7.3 Chemical substances

Chemical substances shall be defined by a representation of the complete covalent molecular structure including the presence of a salt (counter-ion) and/or solvates and, when necessary, stereochemical and related physical characteristics. The molecular structure, the molecular formula, the molecular weight and optical activity, together with the representation of the stereochemistry are mandatory elements to be provided. These elements are inherited from the element groups 'Structure', 'Structural Representation' and 'Molecular Weight'.

Each chemical substance shall be associated with a single structural representation.

Stereochemistry shall be completely defined when known. If not known, positions where stereochemistry is unknown shall be clearly identified.

Underlying the graphical representation of the structure shall be a textual format that indicates the atoms and the connectivity between atoms that represent a molecular structure.

Fixed and variable stoichiometric ratios of moieties within a substance shall be captured. For substances that have moieties with variable composition stoichiometry, the range of composition shall be captured.

Unknown composition stoichiometry of a given moiety or moieties shall also be clearly identified. Composition stoichiometry shall be defined as fully as possible; unknown and variable composition stoichiometry shall also be allowed.

Physical properties shall only be used to define single substances that have variable or unknown composition stoichiometry. Physical properties shall only be captured when they are necessary to distinguish two substances from one another.

Isotopes shall be described in the structural representation; the specific position or positions of substitution shall be provided, if known. Substances shall be defined independently of the extent of isotopic enrichment of a given radioisotope.

The information model for the Chemical Substance is shown in Figure 19.

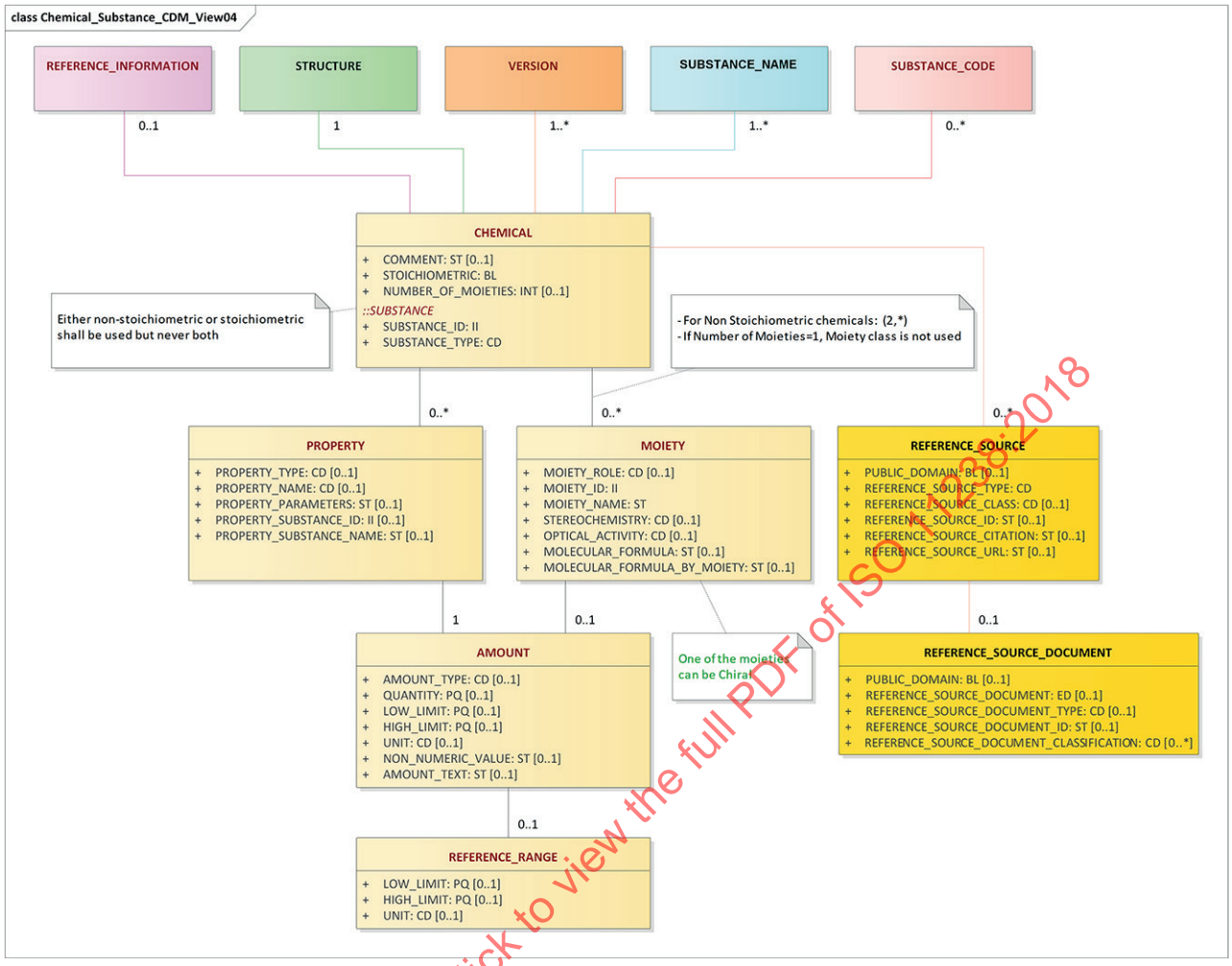

Figure 19 — Information model for the chemical substance

7.4 Protein substances

A protein consists of a defined sequence of alpha-amino acids connected through peptide bonds folded into 3-dimensional structures and consists of one or more chains where each chain is a separate sub-unit.

NOTE 1 Mixtures of proteins, such as immunoglobulins, that have a large number of individual proteins with diverse sequences will be described as structurally diverse substances.

Proteins that differ in protein sequence, type of glycosylation, disulfide linkages or glycosylation site shall be defined as separate substances. Detailed information on Glycosylation including the types of glycans and the extent of site of occupancy can be captured at the Specified Substance Group 1 information level.

EXAMPLE Interferon alfa-2a and interferon alfa-2b, whose sequences differ at a single residue, would be defined as different substances.

The structural representation, the molecular formula, and the molecular weight are mandatory elements to be provided for non-glycosylated peptides and small proteins. These elements are inherited from the element groups ‘Structure’ and ‘Structural Representation’. The element group ‘Molecular Weight’ is part of the Protein class information model. See Figure 20. For all proteins, the molecular weight or molecular weight range should be provided, if known. Multiple molecular weights that either depend on the method or type should also be provided.

All non-glycosylated proteins shall be defined without regard to the method of synthesis, the cell line or organism biological matrix from which the protein was produced or isolated.

Proteins shall be described without regard to microheterogeneity.

Like chemical substances, protein substances and nucleic acid substances (described in 7.5) shall be described as single defined molecular entities. Microheterogeneity shall not be described because of inherent variability. Cyclic peptides and those derived largely from non-proteogenic amino acids as well as extensively-modified oligonucleotides shall be defined as chemical substances.

The type of glycosylation shall reflect significant differences in overall glycosylation and is determined from the species of the cell or tissue from which the protein was isolated. A limited set of controlled terminologies shall be used to describe the type of glycosylation.

Proteins shall be defined by the final expressed sequence; pre-pro-proteins and pro-proteins shall not be described.

Proteins that are irreversibly modified by either chemical or physical processes shall be defined as different proteins.

The description of modified proteins shall capture structural changes that result from the modification when a definitive structure is known.

Structural modifications shall be described using either moieties or molecular fragments that are added to the protein structure or by a description of the modification process if a definitive structural modification does not occur.

The molecular fragment or moiety may have a functional role and that role shall be captured using controlled terminology.

For specific modifications, the site and residue modified shall be described. When the site or sites are not definite the amino acid residue or residues modified will be captured along with the overall extent of modification.

Post-translational modifications shall only be captured if they are essential for activity or present on the predominant forms of the proteins.

In some instances, the modification will not result in a definitive structure. In these instances, the modification process shall be described in a minimal manner, capturing the modifying agent or physical conditions that result in an irreversible change.

Purified blood, or tissue materials whose putative functionality is attributed to a protein or a limited number of proteins with distinct and known amino acid sequences, shall be described as a protein.

Non-covalent interactions between proteins or peptide chains shall not be captured, with the exception of protein chains that are tightly associated with well-defined composition stoichiometry.

Non-defining elements as described below can also be captured at the substance level or/and Specified Substance Group 1 information level using the reference information model, see Figure 16:

— ligand, substrate or target;

— type of interaction of the protein;

— gene from which the protein was derived.

Reference information shall be captured using controlled vocabularies where available.

NOTE 1 Monoclonal immunoglobulins are described as proteins.

NOTE 2 Somatropin, a non-glycosylated protein that can be produced in *E.coli*, yeast or mammalian cells, is defined as the same single substance regardless of the cell line it was produced in.

NOTE 3 Examples of glycosylation types include fungal, plant, anthropoid, avian, mammalian and human.

NOTE 4 Differences in even a single amino acid would result in two distinct substances. For example, interferon alfa-2a and interferon alfa-2b will be defined as separate substances because the sequences differ by a single amino acid. Aggregated human serum albumin, which is formed by irreversible partial physical denaturation, would be defined as a separate substance from human serum albumin.

The information model for the Protein Substance is shown in Figure 20.

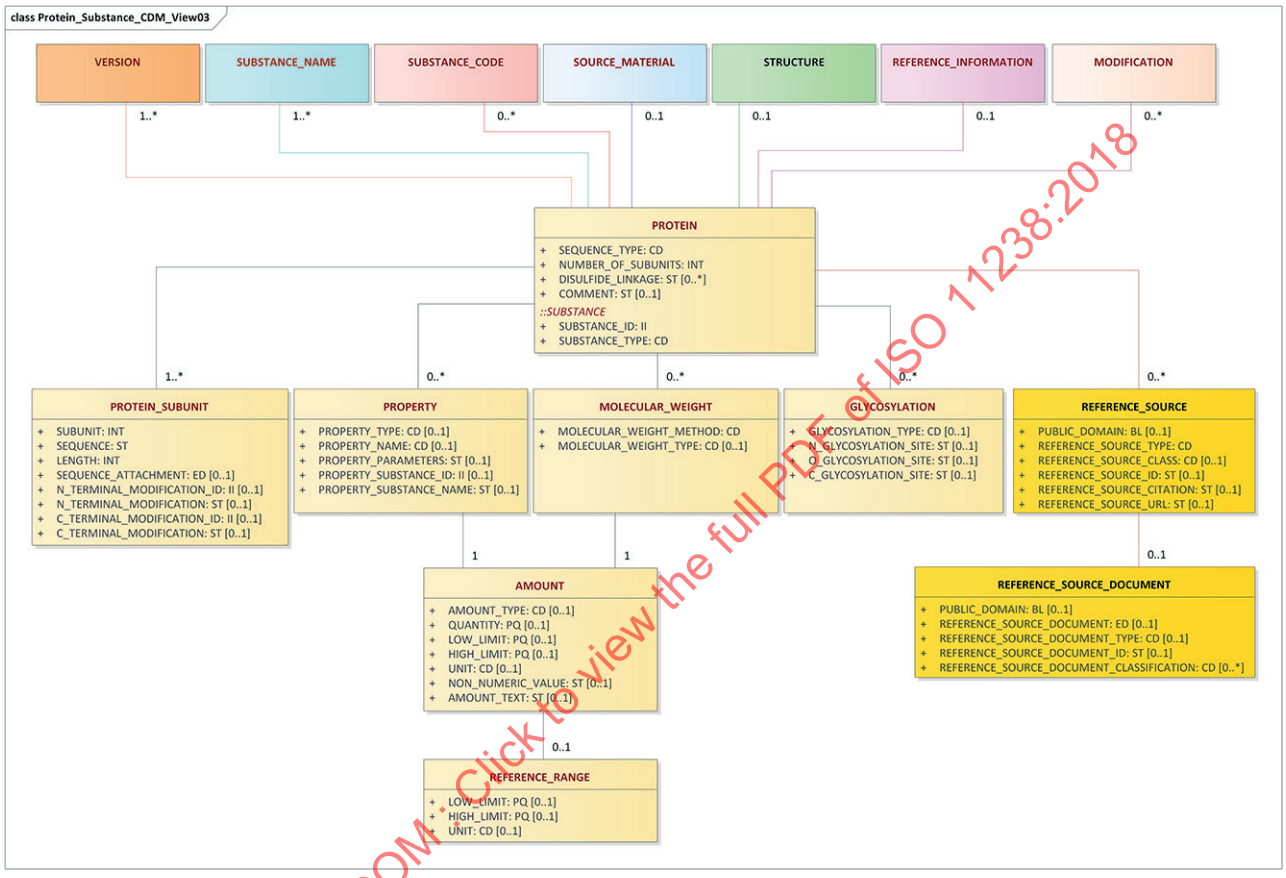

Figure 20 — Information model for the protein substance

7.5 Nucleic acid substances

The sequence of the nucleic acid, the type (RNA, DNA, plasmid, single or double stranded), sugar or sugar-like entities, linkage (typically phosphate), together with any modifications that affect the molecular structure, shall be the defining elements for nucleic acid substances.

Genes, plasmids and the nucleic acid portion of viral vectors used in gene therapy shall also be described as nucleic acid substances.

Individual gene elements shall be described and defined as nucleic acid substances.

Modifications, either physical or chemical, that irreversibly modify the underlying molecular structure shall be described using modification elements.

For gene therapy, the entire sequence of the transforming/transducing vector shall be used as the defining element. Each gene element shall also be captured and defined as a substance.

NOTE 1 A gene is composed of coding and noncoding sequences as well as regulatory elements. In describing the gene all the regulatory gene elements will be described and captured in the description of substance. Regulatory elements include transcriptional elements: enhancers, promoters, silencers, insulators, locus control regions, activators, repressors, coactivators and chromosome remodelling factors.

The information model for the Nucleic Acid substance is shown in Figure 21.

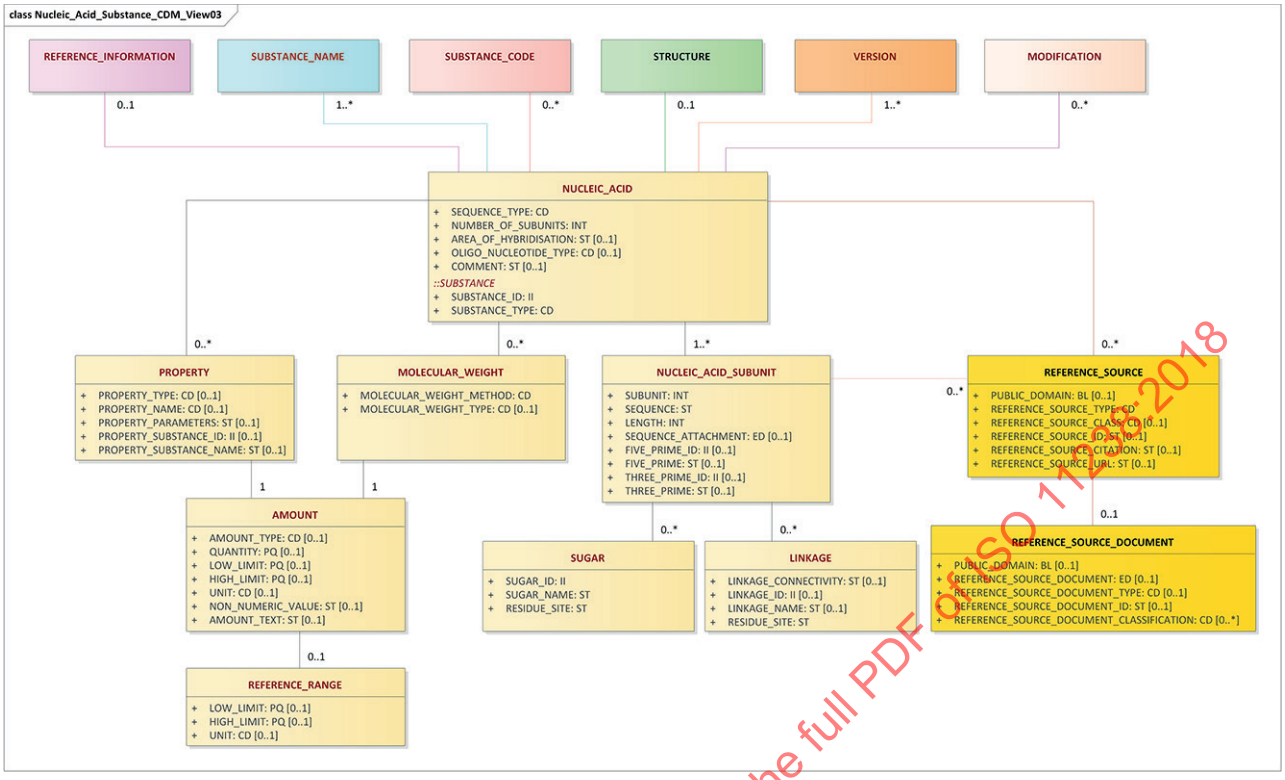

Figure 21 — Information model for the nucleic acid substance, high level view

7.6 Polymer substances

Polymers shall refer to material that is polydisperse and contains structural repeat units.

Polymers shall be defined using a combination of controlled vocabularies and representations of the molecular structure of the structural repeat units, substituents that are attached to the structural repeat unit, which are described as either fragment or moiety modifications, molecular weight or the polydispersity of the material. The degree of polymerization, monomer description, polymer starting material used to synthesize synthetic polymers or copolymers, the source material for naturally derived polymers, polymeric end groups, and physical or biological properties shall also be captured when known and needed to distinguish material. Polymers shall be defined to the level of specificity needed to distinguish materials, and broad polymeric definitions shall be discouraged.

EXAMPLE Polymers containing polyethylene glycol structural repeat units are defined based on either degree of polymerization or molecular weight. A generic polyethylene glycol substance is not defined as a substance because of the wide variation in the functionality of these types of materials and safety concerns related to the degree of polymerization.

The polymer class shall be defined by the number of structural repeat units and the connectivity between them. A controlled vocabulary shall be developed as required to describe the polymer class, polymer geometry and polymer connectivity (copolymer sequence type).

Physical and biological properties shall only be a defining element if they are necessary to distinguish polymeric substances from one another and are related to the underlying molecular structures of the polymeric ensemble.

NOTE Values for polymer class would include homopolymer, copolymer; values for polymer geometry would include linear, branched, cross-linked and network or dendritic; values for copolymer sequence type would include random, statistical, alternating, periodic, block, mixed, graft or cross. Dispersity is usually determined from the ratio of the weight average molecular weight to the number average molecular weight. Properties such as viscosity, light scattering or sedimentation velocity, which are indicative of molecular weight, and biological properties such as enzymatic inhibition can also be distinguishing properties.

The structural repeat unit shall have a distinct stoichiometric composition. However, if the substituents within SRU are variable the structural information of the substituents can be partially described e.g. SRU of polysaccharides sourced from biological matrixes (vaccines).

The information model for the polymer substance is shown in Figure 22.

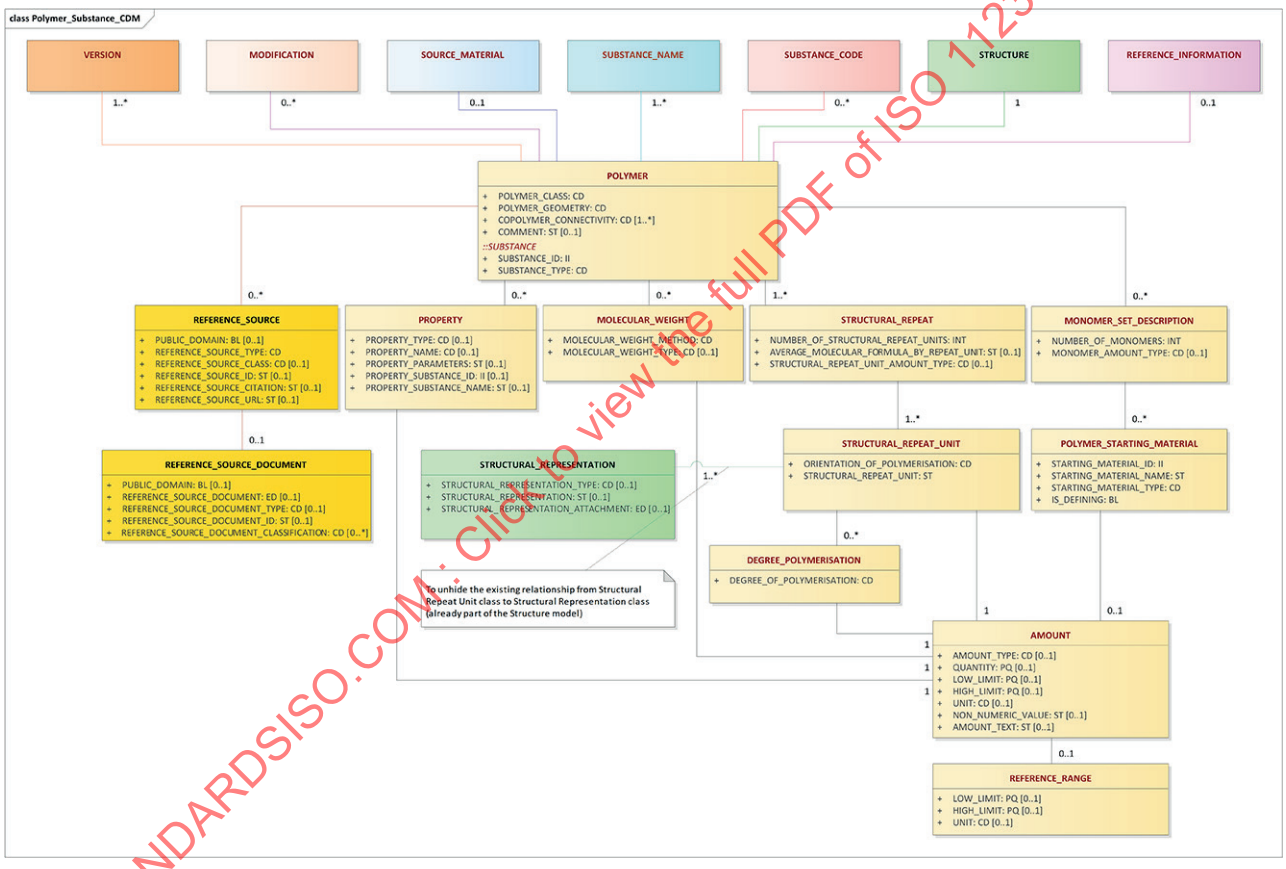

Figure 22 — Information model for the polymer substance

7.7 Structurally diverse substances

There is a wide variety of substances described as structurally diverse substances. A Structurally Diverse substance is a type of polydisperse substance isolated from a single source that is a complex mixture which cannot be described as a mixture of a limited number of single substances.

NOTE 1 Structurally diverse substances are defined based on immutable properties of a given material. Modifications that irreversibly alter the structure of the material, distinctive physical properties or components subsumed into the material, e.g. a gene in gene therapy substances, are defining elements for structurally diverse substances. Fractions derived from source material (oils/juices and extracts) are also captured in the definition. Protein mixtures containing a large number of diverse sequences such as polyclonal immunoglobulins are defined as structurally diverse substances. Minerals are also defined as structurally diverse substances.

This category would be used to describe the following class of substances:

— Herbals: The Substance (fresh), the Herbal Drug and the Herbal preparation and substances used in the preparation of plant-based allergenic extracts.

  The Substance (fresh) name is referred to the scientific genus/binomial/trinomial with author and part, e.g. *Ginkgo biloba* L., Leaf; the Substance name equivalent to the Herbal Drug name refers to the scientific genus/ binomial/ trinomial without the author plus the part, e.g. *Ginkgo biloba*, Leaf; the Substance name referring to the Herbal preparation consists of the scientific genus/binomial/ trinomial without author, plus the part and fraction, e.g. *Olea europaea*, Fruit, Oil. For the common name: Olive Oil, Virgin. The source material for such a Herbal preparation is a separate substance (fresh) defined as *Olea europaea* L., Fruit., which is the parent substance of *Olea europaea*, Fruit, Oil.

— Plasma-derived substances and polyclonal antibodies will be defined as structurally diverse substances. Purified blood substances, distinct clotting factors and human serum albumin can also be described as proteins. Polyclonal immunoglobulins are described as structurally diverse materials and require identification of the immunoglobulin type and targeted antigen if applicable.

— An Allergen substance is a structurally diverse substance derived from biological matrices. Many of the specific allergenic proteins responsible for the allergenic response in the majority of patients have been isolated and characterized, e.g. Fel d 1 protein in cat saliva or Der p 1 (*Dermatophagoides pteronyssinus* group 1) protein or its counterpart Der f 1 (*Dermatophagoides farinae* group 1) protein. These can be substances in their own right described as proteins and related to allergenic extract as a constituent.

— Substances used in Advanced Therapies and Vaccines constitute a wide range of structurally diverse substances: modified viruses, bacteria, cells or tissues, autologous (from the patient's own body) or allogeneic cells that express new proteins or the silencing of an expressed protein; modified lineage specific stem cells used to treat inherited metabolic disorders; antigen primed dendritic cells directed to provide an immune response against cancer cells; retroviruses designed to deliver specific genes to specific cell populations; viruses modified to express an antigenic protein of a human pathogen or acellular matrices to assist wound healing, bone or neural regeneration.

— Homeopathic substances prepared from materials of botanical, mineral, or zoological origin will be described as structurally diverse substances. They are prepared in accordance with a homeopathic manufacturing procedure described in official pharmacopoeias. The Homeopathic Substance called Homeopathic Stock used as starting material for the production of homeopathic preparations is described for the raw material of botanical, zoological or human origin by the source material from which the stock is derived.

Structurally diverse substances shall be defined by the source material from which the substance is derived, modifications that result in irreversible changes in the underlying material and/or physical or biological properties related to underlying molecular composition of the material will also be captured.

Physical or biological properties shall only be used when they are essential to defining and distinguishing the material.

NOTE 2 The majority of structurally diverse substances are derived from biological organisms. They might also be complex natural materials such as coal tar or mineral oil.

EXAMPLE 1 Light mineral oil is distinguished from mineral oil on the basis of the viscosity and specific gravity.

For organism-based structurally diverse substances, the parent organism from which the source material was derived is essential to the definition of the substance. Parent organisms shall be defined from the family to at least the species level. Varieties, cultivars, strains or sub-strains of biological material shall be defining information if intraspecific differences are distinct and reflect consistent differences in functionality or composition. Kingdom, phylum, class and order can also be captured

when available but these levels of taxonomy will generally not be defining, see class Source material/ Organism.

NOTE 3 Herbals are typically described by parent organism family, genus, species and part or parts. If specific parts of a plant are used, identification requires lists of individual parts such as the flower, leaf and stem. An indication of the plant life cycle segment may also be necessary, e.g. whole flowering. Because of variability in constituents due to extraction processes (solvent, temperature, time) and growing conditions (season and place of harvest, type of soil, use of fertiliser, amount of daylight and water), biological extracts shall be identified by their source unless they represent a particular fraction or class of chemicals, e.g. sennosides (*Senna alexandrina* anthraquinone glycosides). Substance (fresh), Herbal Drug and Herbal preparation (oils/ juices and extracts) are considered as different substances.

A cultivar or variety of a plant shall be defined as a different substance if differences exist in constituents or functionality. Other organisms, typically bacteria and viruses, shall require the identification of subspecies, variety, strain or type, in order to be accurately described and distinguished from related substances.

EXAMPLE 2 Broccoli and cauliflower, which are different cultivar groups or varieties of *Brassica oleracea*, are defined as different substances even though they share the same genus and species because there are considerable differences in appearance and constituents. Influenza viruses would be defined at a level that allows the distinction of various vaccine strains.

Commodity oils, juices and exudates of plants shall be separate substances. Oils and juices are Herbal preparations and shall be described as fractions of the material from which they are isolated. The materials and processes (i.e. time, temperature, solvent) used to prepare extracts vary and are captured at the Specified Substance Group 1 information level.

EXAMPLE 3 Olive oil is *Olea europaea*, Fruit, Oil. Orange juice is *Citrus aurantium*, Fruit, Juice. Dry green tea shall be defined as the Herbal Drug and green tea extracts shall be defined as the Herbal preparation (liquid extract) of leaves of *Camellia sinensis*.

The information model for the Structurally Diverse substance is shown in Figure 23.

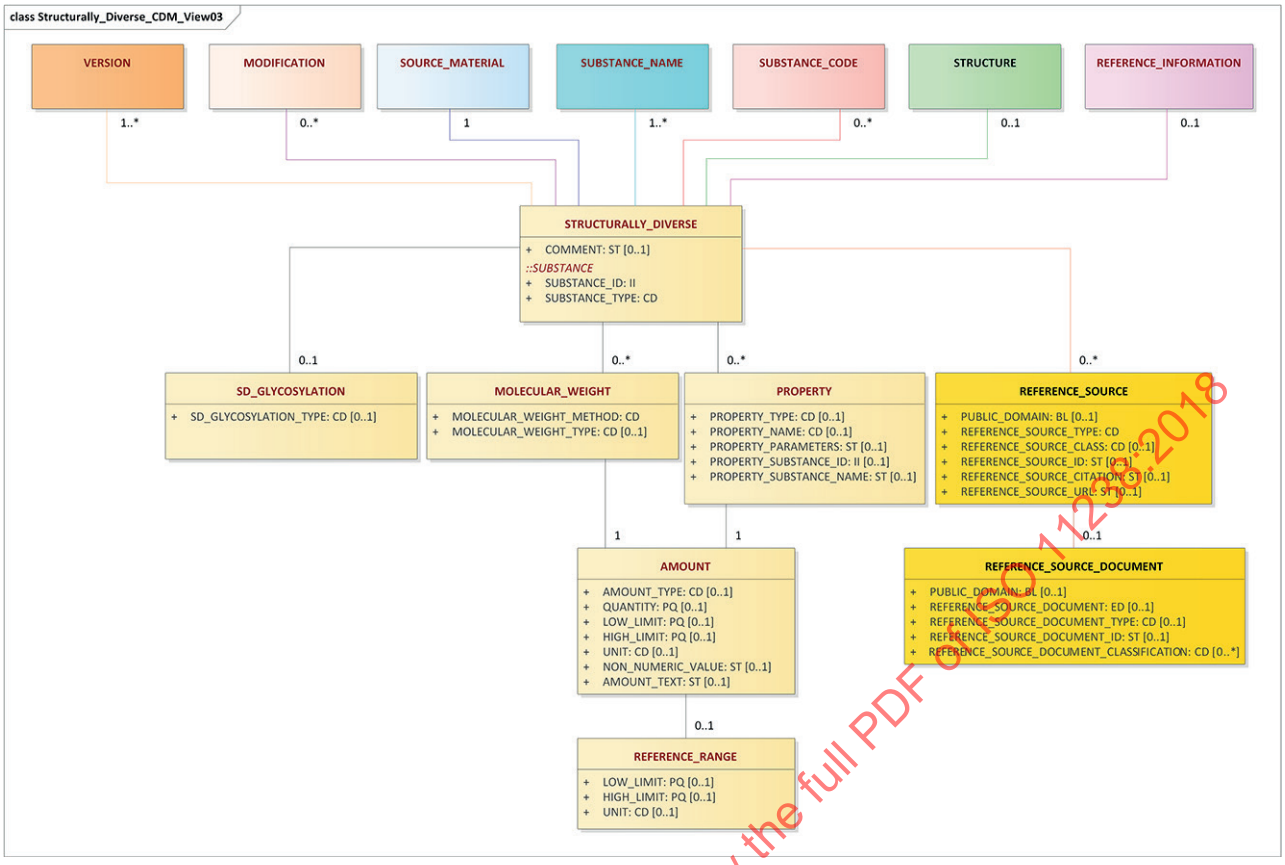

Figure 23 — Information model for the structurally diverse substance

7.8 Mixture

Mixture describes a type of polydisperse substance that is a combination of single substances isolated together or produced in the same synthetic process.

For mixtures derived from natural sources, the source material from which the mixture was derived shall be identified.

Mixture substances shall not be combinations of diverse material brought together to form a product.

EXAMPLE Simethicone, which consists of dimethicone and silicon dioxide, would not be defined as a mixture substance because the substances are not typically isolated or synthesized together; it would be defined at the Specified Substance Group 1 information level.

The extract of a multi-substance material (homologous group of allergen source material) can be described using the class 'Constituent Component' and the class 'Source Material', element group 'Fraction Description', because the extract is obtained from structurally diverse single substances (starting materials) as parent substances. This substance (the allergen extract) is the result of the same (synthetic) process and hence the extract is considered as a mixture substance.

There shall be three types of mixture substance:

— "All Of" in which all of the single substances are required to be present;

— "Any Of" in which one or more of the single substances are required to be present;

— "One Of" in which only one of the single substances is present.

"Any Of" mixtures shall indicate whether a given single substance is always present. The relative amount of each single substance shall not be captured.

Relative amounts of substances in a mixture substance shall be captured at the substance or specified substance level consistent with either a pharmacopoeial or manufacturer specification.

All mixture substances shall consist of mixtures of single substances.

Mixtures of mixture substances shall not be allowed.

Mixtures of mixture substances shall be represented as a single mixture of all the underlying substances.

All related substances in a mixture present in an amount greater than one percent shall be constituent components of the mixture substance.

Impurities and degradants shall generally not be considered constituent components of a mixture substance.

Mixtures that cannot be described by a limited number of related single substances shall be described as structurally diverse substances.

The information model for the Mixture is shown in Figure 24.

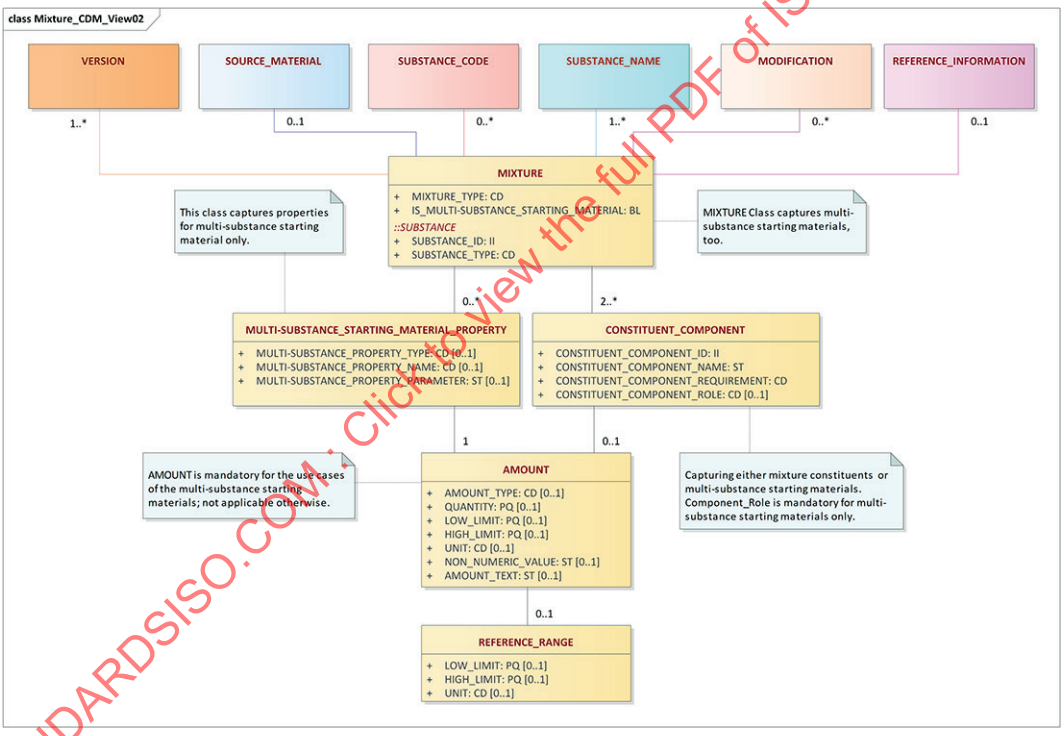

Figure 24 — Information model for the mixture

8 Defining specified substances

8.1 General

Although the substance model captures information essential to the description of materials in medicinal products, there is often a strong regulatory need for additional information that is not captured at the substance level. Specified substances provide a general information model that shall be used to further define materials present in medicinal products.

The specified substance shall be organized to capture diverse information in a consistent manner. This information shall include:

— purity or grade;

— manufacturer data including information on the manufacturer and processes in manufacturing;

— analytical data in view of the tests and specifications;

— analytical methods used for potency determination;

— constituent substances, including amounts and role when known and relevant;

— specifications for identity, impurities, degradants, related substance limits would be captured using constituent substances and potency;

— unitage;

— reference material.

To meet the needs of medicinal product identification, the elements of the specified substance shall be divided into four groups and a specified substance identifier shall be associated with each group of elements.

NOTE The grouping of elements simplifies the data model and allows for both regional and incremental implementations.

8.2 Specified Substance Group 1

Material containing multi-substance material of diverse origin, physical and polymorphic forms of material as well as standardized herbal and allergenic extracts are defined by Specified Substance Group 1 elements. In addition, extended information regarding plasma-derived substances sourced from a specific Cryopoor plasma or Cryoprecipitate flow will be defined as part of Specified Substance Group 1 based on the tests and acceptance criteria of the plasma sourced from a selection of the countries. The following paragraphs discuss additional elements that may be defining.

Solvents used in the preparation of herbal or allergenic extracts, specific marker or signature substances present in materials derived from biological matrices, the physical form of a substance when relevant, fraction information and any properties captured as characteristic attributes essential to the description of the material. Elements of micro-heterogeneity for proteins such as details of glycosylation and other post-translational modifications can also be captured at the Specified Substance Group 1 level as modifications. The information model is shown in Figure 25.

The 'Constituent' element group shall consist of substances that are components of a multi-substance material, marker or signature substances present in botanical-, animal- or human-derived material and constituent substance(s) with the substance role of parent. In contrast to mixture substances, the amount of the above substances shall be captured. In all cases, it will be described if the information provided shall be defining or not defining.

Impurities or degradants shall not be captured as constituent substances at the Specified Substance Group 1 information level but will de described at the Specified Substance Group 4 information level.

NOTE Grouping of constituent substances is allowed for the definitions of many materials in commerce that are used in the formulation of medicinal products although the individual ingredients of a multi-substance material should be provided as much as possible.

There are two other element groups which are used to capture additional information about substances. The element group ‘Characteristic Attribute’ is used to capture various information depending on the Substance Type and the element group ‘Attribute Parameter’ is used to capture information of specific

conditions on which the instance of the Characteristic Attribute is measured in combination with the element group 'Amount'.

EXAMPLE 1 Description of density of the chemical substance 'Nitrous oxide, gas' using the element groups 'Characteristic Attribute' and 'Attribute Parameter' in relation to the class 'Amount'. The values of the attributes are provided below:

Element Group: Characteristic Attribute
Attribute Type: Physical
Attribute Name: Density
Attribute Substance ID: PHJSGT785G (Artificial ID)
Attribute Substance Name: Nitrous oxide, gas
Is Defining: Yes.

Element Group: Amount
Amount Type: Exact
Quantity: 85,76
Unit: kg/m^3

Element Group: Attribute Parameter
Attribute Parameter Name: Density condition
Attribute Parameter Value: Physical state: Gas, at 0 °C, 31,29 atm at equilibrium.

EXAMPLE 2 Description of the selection (subset) of the Country of Origin described in a plasma master file used for the manufacturing of a plasma-derived substance by a specific defined Cryopoor plasma flow related to the testing strategy:

Element Group: Fraction Description
Fraction: Process Flow
Material Type: Protein

Element Group: Modification
Modification Type: Physical

Element Group: Physical Modification
Role: Isolation of a specific Cryopoor plasma Process flow by precipitation and adsorption steps by means of the modified Cohn fractionation process.

Element Group: Characteristic Attribute
Attribute Type: Cryopoor plasma Process Flow
Attribute Name: Country of Origin
Is Defining: Yes

Element Group: Amount
Amount Text: NL, DE, FR.

Element Group: Characteristic Attribute (Repeat)
Attribute Type: Cryopoor plasma Process Flow
Attribute Name: Pathogen test, test strategy
Is Defining: Yes.

Element Group: Amount
Amount Type: NLT
Low limit: 8,4
Unit: ^{10}logarithmic virus reduction factors

By using the element groups 'Fraction Description', 'Constituent' and 'Characteristic Attribute' in combination with the element group 'Amount' any additional extended information can be described for the Structurally Diverse Substances: Herbals, Herbal preparations, Homeopathics, Plasma-derived proteins, Allergens and Allergen extracts, Vaccines and Advanced Therapies at the Specified Substance Group 1 Information level.

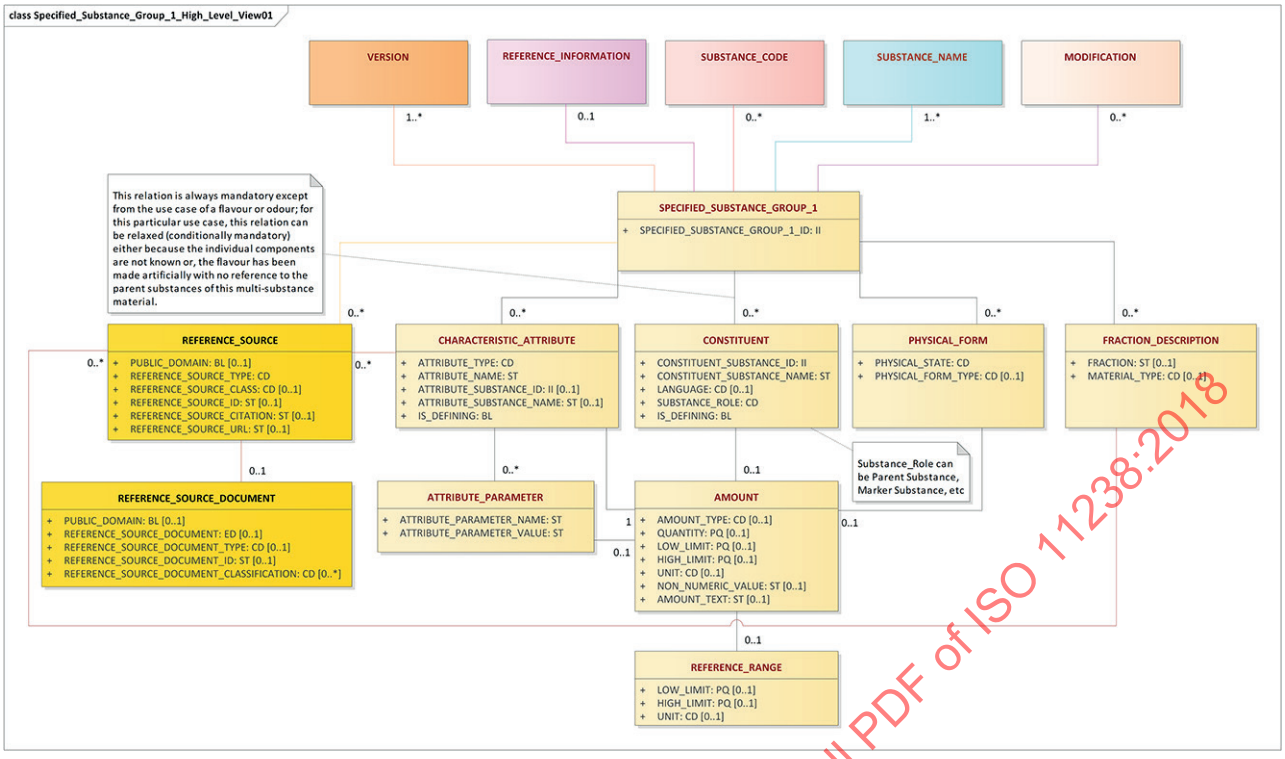

Figure 25 — High level information model for the Specified Substance Group 1

For the (Herbal) substance (fresh) or the Herbal Drug or allergens sourced from botanical or zoological material instances of the characteristic attributes are: Degree of comminution, process state, wild/cultivated, growth state, harvesting time, decontamination process, country of origin, geographical region, drug extract ratio, extract ratio for allergens, feeding composition of zoological material (e.g. Mites), storage time and storage condition. All these elements are in most cases not defining but can be used to describe additional information which has effect on the content of the markers or major allergens.

In addition, the Specified Substance Group 1 information model includes the element group 'Fraction Description' in which further details are described for the Herbal preparation 'extract' with limited information e.g. liquid or dry extract. Extended fraction information of a herbal preparation such as (quantitative) 'Extraction Solvent Composition' information is captured in combination with the element group 'Constituent' for the solvent composition and by the element group 'Characteristic Attribute' for the 'Drug Extract Ratio'.

The extraction solvent composition and the drug extract ratio are defining elements for herbal preparations described at the Specified Substance Group 1 information level.

For homeopathic substances, additional information of the 'Vehicle' composition is captured by the 'Constituent' class in combination with the class 'Amount'. The element group 'Characteristic Attribute' is used to describe the 'Substance Dilution Grade' and, the element group 'Amount' is used to describe the dilution/potentization value. The attribute 'Unit' is used to describe the unit of dilution grade e.g. DH (Decimal).

For plasma-derived substances additional information will be captured by the element group 'Fraction Description' which captures the elements 'Fraction' (Cryopoor plasma or Cryoprecipitate). A specific case of Cryopoor plasma/Cryoprecipitate Process Flow can be used to describe the manufacture of an intended blood coagulation factor or a plasma-derived substance (e.g. Serum albumin) from a subset of the countries as described in the plasma master file in combination of a specific testing strategy. The selected countries of origin and the testing strategy can be described by using the element group 'Characteristic Attribute', see example 2 above.

For Allergenic substances of zoological source material important characteristic attributes are 'Wild/ Cultivated', 'Growth Stage', 'Feeding Composition', 'Harvesting/Killing process', 'Storage Time', and 'Storage condition' when appropriate.

For Vaccines Specified Substance Group 1 is used to describe the constituent substance e.g. haemagglutinin and neuraminidase as active marker, sodium deoxycholate for virus disruption of an influenza virus inactivated split vaccine. Physical form, e.g. Virus Like Particle size and characteristic attribute, e.g. History of the Strain, Passage information of the Master seed and Working seed are further used to describe additional information of vaccines.

8.3 Specified Substance Group 2

Elements shall be used to capture the manufacturer of either a Substance or a Specified Substance Group 1, along with minimal manufacturing information, see Figure 26.

The minimal manufacturing information shall include the overall production method type (e.g. synthetic, extractive, recombinant), limited production method description, production system type (e.g. cell line, plant or animal tissue) and production system (specific cell line). Critical Process Version Number shall be used to distinguish Specified substances that have undergone a major change in the Critical Process used in the manufacturing of the (specified) substance, e.g. a change which needs regulatory approval. The initial Critical Process Version Number shall be one and each subsequent number shall be increased sequentially.

NOTE 1 The Specified Substance Group 2 elements would allow the tracking of the substance to the manufacturer in a 1 to 1 relationship. This is important for biosimilar and other generic products. It also allows the distinguishing of synthetic peptides from recombinant peptides and the capture of the production cell lines.

NOTE 2 For substance (active ingredients intended to be used in the medicinal product) the element 'Manufacturing Type', 'Production Method Type', 'Production System Type' and 'Critical Process Version Number' is always mandatory.

For excipients (non-active ingredients intended to be used in the medicinal product) the information is optional. But for certain excipients that cause an intolerance or allergenic reaction (e.g. sesame oil) this information shall always be provided.

The manufacturer information is laid down in the element group 'Organization' with the attributes 'Manufacturer ID', 'Manufacturer Name', 'Issuer of ID' and the 'Manufacturer Role'. The Manufacturer ID could be obtained from a maintenance organization keeping track of all organization information needed by regulators. The Manufacturer Role can be e.g. the actual manufacturer of the bulk substance, a manufacturer specialized in grinding the bulk chemical substance into a particular particle size range or a harvester of botanical material.

The information model for Specified Substance Group 2 is shown in Figure 26; Figure 27 provides the extended manufacturing information needed to substantiate the change of the critical process version number.

The extended manufacturing information is used to provide information about the extraction solvent composition used for Herbal preparations. In many situations, there is a stepwise approach in making extract of botanical material in which the composition of the extraction solvent changes from step to step. A change in the subsequent extraction solvent composition would change the Specified Substance Group 2 ID even when the parent substance is the same substance (fresh) or Herbal Drug. The extended information model is used to further describe multiple extraction steps or an extraction step followed by a modification e.g. modified allergen extracts.

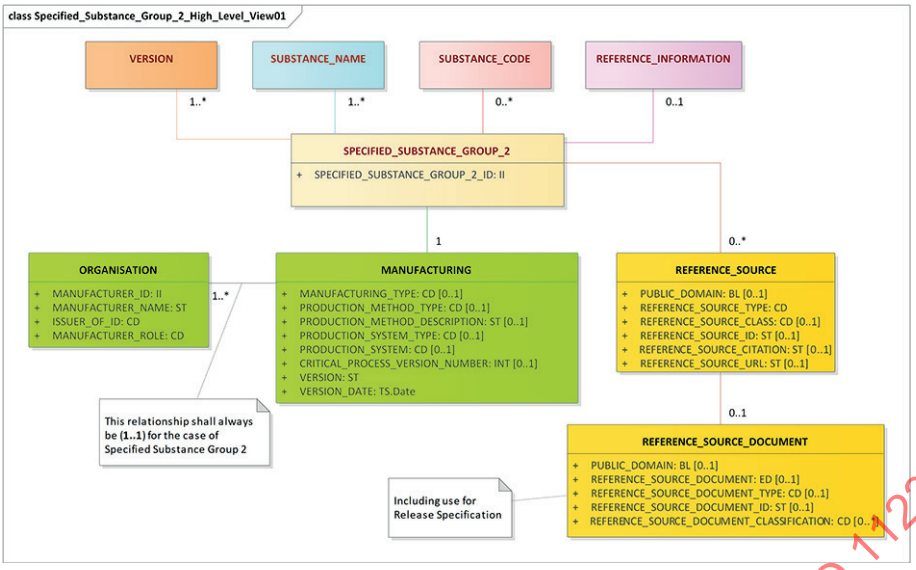

Figure 26 — High level information model for the Specified Substance Group 2

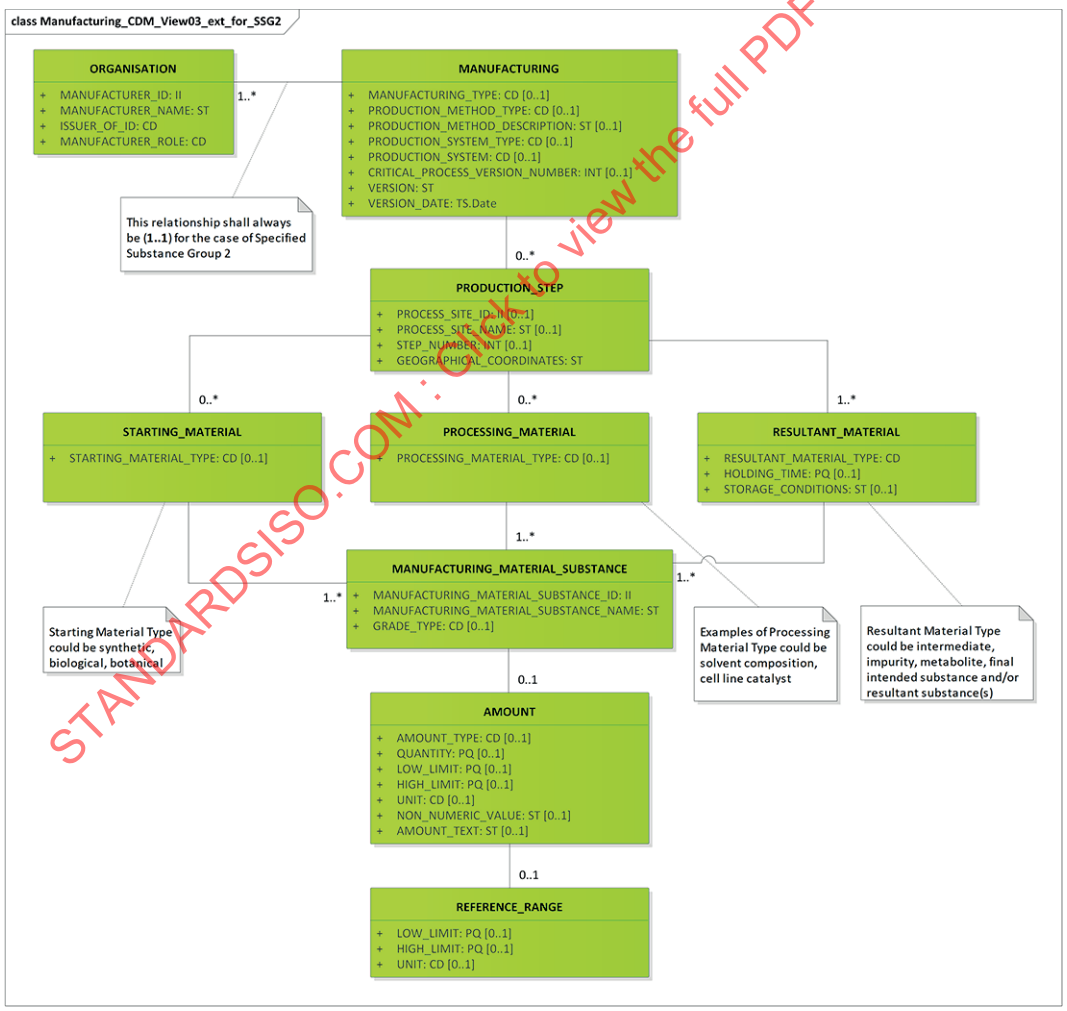

Figure 27 — Extended manufacturing information model for the Specified Substance Group 2

The starting material is usually the botanical material of which the extract is made, the processing material is usually the extraction solvent composition and the resultant material is the obtained crude