INTERNATIONAL STANDARD

ISO 11238

First edition
2012-11-01

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 médicaments — Éléments de données et structures pour l'identification unique et l'échange d'informations réglementées concernant les substances

Reference number
ISO 11238:2012(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web 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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 11238 was prepared by Technical Committee ISO/TC 215, *Health informatics*.

Introduction

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

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

ISO 11616, *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, *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, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of units of measurement*.

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;

— 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 Medicines Agency;

— 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.

The necessary messaging specifications are included as an integral part of the IDMP standards to secure the interactions listed above.

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 materials within medicinal products.

This International Standard provides a structure that enables the assignment and maintenance of unique identifiers for all substances in medicinal products or in packaging materials in which medicinal products are contained. This International Standard sets out the general rules for defining and distinguishing substances, and provides a high-level model that structures substances and specified substances for the organization and capturing of data.

This International Standard has been developed using HL7’s Common Product Model, and detailed modelling of substances and specified substances has been undertaken in that domain. It is anticipated that implementation will use the HL7 substances implementation guide and messaging to deliver a strong, non-semantic unique identifier for every substance present in a medicinal product. It is anticipated that a single organization will be

responsible for the generation of 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 International Standard 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.

Materials used in medicinal products range from simple chemicals to gene-modified cells to animal tissues. To unambiguously define these substances is particularly challenging. This International Standard 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 International Standard contains five groups of elements that are sufficient to define all substances. Although it is certainly possible to define or classify substances in other ways, this International Standard uses a minimalist 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 International Standard has intentionally not included these supramolecular interactions at the substance level because of the variable nature and strength of such interactions. This International Standard 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 International Standard also provides elements for the capture of further information on substances, such as grade, manufacturer, manufacturing 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 International Standard depict elements and the relationship between elements that are necessary to define substances. The terms and definitions described in this International Standard 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 International Standard 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.

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

1 Scope

This International Standard provides an information model to define and identify substances within medicinal products or substances used for medicinal purposes, including dietary supplements, foods and cosmetics. Other standards and external terminological resources are referenced that are applicable to this International Standard.

2 Terms, definitions, symbols and abbreviated terms

2.1 Terms and definitions

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

2.1.1
active marker
constituent, or groups of constituents, of an herbal substance, herbal preparation or herbal medicinal product which are of interest for control purposes and are generally accepted to contribute to therapeutic activity

NOTE Active markers are not equivalent to analytical or signature markers that serve solely for identification or control purposes.

2.1.2
analytical data
set of elements to describe and capture methods and reference material used to determine purity, potency or identity in a specified substance

2.1.3
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

2.1.4
chemical substance
type of substance that can be described as a stoichiometric or non-stoichiometric single molecular entity and is not a protein or nucleic acid substance

NOTE 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.

2.1.5
chiral substance
substance whose molecular structure is not superimposable on its mirror image

2.1.6
component
intended constituent of a specified substance

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

NOTE Components are used to describe the substances and specified substances that form a multi-substance material.

2.1.7
composition stoichiometry
quantitative relationships between the chemical elements or moieties that make up a substance

EXAMPLE Sodium phosphate dibasic heptahydrate and sodium phosphate dibasic dihydrate are defined as different substances.

2.1.8
constituent
substance present within a specified substance

NOTE Constituents can be impurities, degradants, active markers or signature substances, or single substances mixed together to form a product. Constituents shall have an associated role and amount. Constituent specifications shall be used to describe components as well as limits on impurities or related substances for a given material.

2.1.9
controlled vocabulary
finite set of values that represent the only allowed values for a data item

NOTE 1 The allowed values can be codes, text or numeric.

NOTE 2 Adapted from CDISC Clinical Research Glossary V8.0, 2009.

2.1.10
copolymer
polymer with more than one type of structural repeating unit linked through covalent bonds

NOTE Copolymers are obtained by copolymerization or sequential polymerization of two or more monomers. Copolymers can be random, alternating, block or graft.

2.1.11
critical process step
manufacturing step necessary for production of a specified substance

2.1.12
degree of polymerization
number of structural repeating units in a polymeric block or chain

NOTE Applies to both homopolymers and block copolymers where it refers to the degree of polymerization within a block.

2.1.13
diverse origin
substances that are not isolated together or the result of the same chemical synthetic process

NOTE Material containing multiple substances is defined either as a mixture substance or a multi-substance (group 1) specified substance based on origin. Two substances brought together that do not undergo a chemical reaction resulting in the formation or breakage of specific chemical bonds would be defined as separate substances, even if there are non-bonding interactions between the substances.

2.1.14
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

2.1.15
fraction
distinct portion of material derived from a complex matrix, the composition of which differs from antecedent material

NOTE This concept is used to describe source material and is recursive in that a subsequent fraction can be derived from an antecedent fraction, which is implied in the order of listing.

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

2.1.16
gene
basic unit of hereditary material that encodes and controls the expression of a protein or protein subunit

2.1.17
gene element
individual element within a gene such as a promoter, enhancer, silencer or coding sequence

2.1.18
glycosylation
enzymatic process that links saccharides or oligosaccharides to proteins, lipids or other organic molecules

2.1.19
glycosylation type
significant differences in glycosylation between different types of organisms

NOTE This distinguishes the pattern of glycosylation across organism types, e.g. human, mammalian and avian. The glycosylation type is a defining element when polydisperse organism-based glycosylation exists in a substance.

2.1.20
grade
set of specifications indicating the quality of a specified substance

2.1.21
homopolymer
polymer containing a single structural repeating unit

2.1.22
isotope
variants of a chemical element that differ by molecular mass

NOTE 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.

2.1.23
manufacturing
process of production for a substance or medicinal product from the acquisition of all materials through all processing stages

NOTE The critical process, starting and processing materials, and critical production parameters are included.

2.1.24
material
any entity that has mass, occupies space and consists of one or more substances

2.1.25
medicinal product
any substance or combination of substances which may be administered to human beings or animals for treating or preventing disease, with the view to making a medical diagnosis or to restore, correct or modify physiological functions

NOTE 1 A medicinal product may contain one or more manufactured items and one or more pharmaceutical products.

NOTE 2 In certain jurisdictions, a medicinal product may also be defined as any substance or combination of substances which may be used to make a medical diagnosis.

2.1.26
microheterogeneity
substances isolated together that contain minor differences in glycosylation, or post-translational modification such as glycosylation or sequence heterogeneity

NOTE 1 Microheterogeneity is not a defining characteristic of either substances or specified substances.

NOTE 2 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, variable proteolytic processing or other.

2.1.27
mixture substance
type of polydisperse substance that is a combination of single substances isolated together or produced in the same synthetic process

NOTE 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 (Group 1 specified substances) and not as mixture substances.

EXAMPLE Gentamicin is defined as a mixture substance of Gentamicin C1A, Gentamicin C1 and Gentamicin C2. 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 product.

2.1.28
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 and should be defined in a consistent manner across all products. To avoid ambiguity, the free acid and/or free base should be used as the moiety upon which strength is based.

NOTE In this International Standard, moiety shall be used in the context of non-stoichiometric chemical substances and in modification of nucleic acid, proteins, polymers and structurally diverse substances. Moieties shall be single substances, ions or solvate molecules.

2.1.29
molecular fragment
portion of a molecule that has one or more sites of attachment to other fragments or moieties

NOTE Molecular fragments are used in the description of polymers to represent substituents and in structural modifications to a substance.

2.1.30
molecular structure
unambiguous representation of the arrangement of atoms

NOTE 1 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 This representation should be generally translatable into a graphical representation.

2.1.31
molecular weight
mass of one molecule of a homogenous substance or the average mass of molecules that comprise a heterogeneous substance

NOTE 1 The unified atomic mass unit is the unit of molecular weight. The type of molecular weight should always be captured.

NOTE 2 For polymers, there are several different types of molecular weight (weight average, number average, etc.).

2.1.32
multi-substance material
multiple substances and/or specified substances of diverse origin used as a component in the formulation of a medicinal product

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

NOTE Multi-substance materials are Group 1 specified substances. Any medicinal product used to formulate another medicinal product could also be considered a multi-substance material.

2.1.33
nucleic acid substance
type of substance that can be defined by a linear sequence of nucleosides typically linked through phosphate esters

NOTE The type of nucleic acid substance (RNA, DNA) is also identified. Oligonucleotides and gene elements (i.e. promoters, enhancers, coding sequences and silencers) are defined as nucleic acid substances.

2.1.34
official name
name given by an official registration authority

2.1.35
parent organism
organism from which biological source material is derived

2.1.36
part
anatomical origin and location of source material within an organism

2.1.37
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 A medicinal product may contain one or more pharmaceutical products.

NOTE 2 In many instances, the pharmaceutical product is equal to the manufactured item. However, there are instances where the manufactured item must undergo a transformation before being administered to the patient (as the pharmaceutical product) and the two are not equal.

2.1.38
physical form
physical state, either gas, liquid or solid, and the type of organization for solid matter

NOTE Solids can be either crystalline or amorphous. Polymorphism can also be captured.

2.1.39
polydisperse substance
substance containing multiple related molecular components

NOTE Polydisperse substances include polymers and structurally diverse material isolated from a single source. Chemical substances, proteins and nucleic acids with defined sequences are not described as polydisperse substances.

2.1.40
polydispersity
measure of the range of molecular masses in a polymer substance

NOTE The polydispersity index of polymers is typically calculated by the ratio of weight average molecular weight to number average molecular weight.

2.1.41
polymer substance
type of polydisperse substance that contains structural repeating units linked by covalent bonds

NOTE Monodisperse proteins and nucleic acids with defined sequences shall not be defined using the polymer substance elements.

2.1.42
post-translational modification
modification of a protein that typically occurs in vivo during or after translation

NOTE Post-translational modification is described within the structural representation and not as a modification of a protein.

2.1.43
processing material
type of material essential to the manufacturing process that is not incorporated into the resultant material

2.1.44
protein substance
type of substance with a defined sequence of alpha-amino-acids connected through peptide bonds

NOTE 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, disulfide linkages and glycosylation type (e.g. fungal, plant, arthropod, 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 15 amino acid residues or less.

2.1.45
protein sequence
order and identity of amino acids within a protein sub-unit

NOTE Protein sequences will be represented by single letter Dayhoff codes and listed from the *N*-terminus to the *C*-terminus.

2.1.46
protein sub-unit
linear sequence of amino acid residues connected through peptide bonds

NOTE Repeated sub-units in proteins are captured.

EXAMPLE Monoclonal antibodies typically consist of four sub-units.

2.1.47
resultant material
material that is the result of a critical process

NOTE Resultant material may be the starting material of the next process or the final material or actual specified substance.

2.1.48
salt
ionic substances formed from the neutralization reaction of an acid and base

NOTE Salts are ionic compounds composed of cations (positive ions) and anions (negative ions).

2.1.49
silencer
DNA sequence that suppresses transcription

2.1.50
single substance
substance that can be described by a single representation or set of descriptive elements

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

NOTE 2 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.

2.1.51
solvate
substance formed through association of a solvent molecule (e.g. water, alcohol) with another moiety.

NOTE Solvates can be either stoichiometric or non-stoichiometric and are predominately present when substances exist in a solid form.

2.1.52
source material
material from which a substance is derived, which is defined based on taxonomic and anatomical origins

NOTE This class is used to define structurally diverse and polymer substances isolated from biological matrices.

2.1.53
specified substance
group of elements which describe multi-substance materials and specify further information on substances and multi-substance materials relevant to the description of medicinal products

EXAMPLE This could include grade, units of measure, physical form, constituents, manufacturer, critical manufacturing processes (i.e. extraction, synthetic, recombinant processes), specification and the analytical methods used to determine that a substance is in compliance with a specification.

NOTE 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.

2.1.54
starting material
type of material that is modified through a manufacturing process

2.1.55
stoichiometric
substances that contain moieties in simple integral ratios

NOTE 1 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 Chemicals have defined composition stoichiometry when the ratio of all moieties (ion, counter ion and solvate) can be represented as simple integral ratios.

2.1.56
stereochemistry
relative spatial arrangement of atoms within molecules

2.1.57
structurally diverse substance
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 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 and juices) 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.

2.1.58
substance
any matter of defined composition that has discrete existence, whose origin may be biological, mineral or chemical

NOTE 1 Substances can be 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 2 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 (i.e. isomeric mixtures) or naturally-occurring (i.e. 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. Substances may be salts, solvates, free acids, free bases and mixtures of related compounds that are either isolated or synthesized together.

2.1.59
substituent
molecular fragment attached to a structural repeat unit of a polymer that typically replaces a hydrogen atom

NOTE 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.

2.1.60
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 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.

2.1.61
taxonomy
scientific organism classification system needed to describe the origin of source material in substances isolated from biological matrices

NOTE 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.

2.2 Symbols and abbreviated terms

2.2.1
ACS
American Chemical Society

2.2.2
ASK Number
ID of a substance in German “Arzneistoffkatalog” (Pharmaceutical Substance Dictionary)

2.2.3
BAN
British Approved Name

2.2.4
DCF
Dénominations Communes Françaises (French approved drug name)

2.2.5
EP
European Pharmacopoeia

2.2.6
INCI
International Nomenclature of Cosmetic Ingredients

2.2.7
INN
International Non-proprietary Name [also known as rINN (recommended International Non-proprietary Name) or pINN (proposed International Non-proprietary Name)].

2.2.8
JAN
Japanese Approved Name

2.2.9
JP
Japanese Pharmacopoeia

2.2.10
UCUM
Unified Code for Units of Measure

2.2.11
USAN
United States Adopted Name

2.2.12
USP
United States Pharmacopeia

3 Requirements

3.1 General

Substances and specified substances shall be defined in a manner consistent with the elements and relationships present in the figures within this clause. An implementation guide shall further define these elements and relationships.

3.2 Concepts required for the unique identification and description of substances

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

NOTE 1 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 and multi-substance materials that are components of a medicinal product formulation.

This International Standard 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:

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

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

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

d) to avoid ambiguity and facilitate implementation, a mixture shall be defined as a combination of single substances;

e) 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.

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 2 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 and well-defined polymers. Materials that contain moieties that interact with polymers, complex matrices or cyclodextrins will typically not be defined as substances. 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. Liposomal doxorubicin would be defined as a specified substance 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.

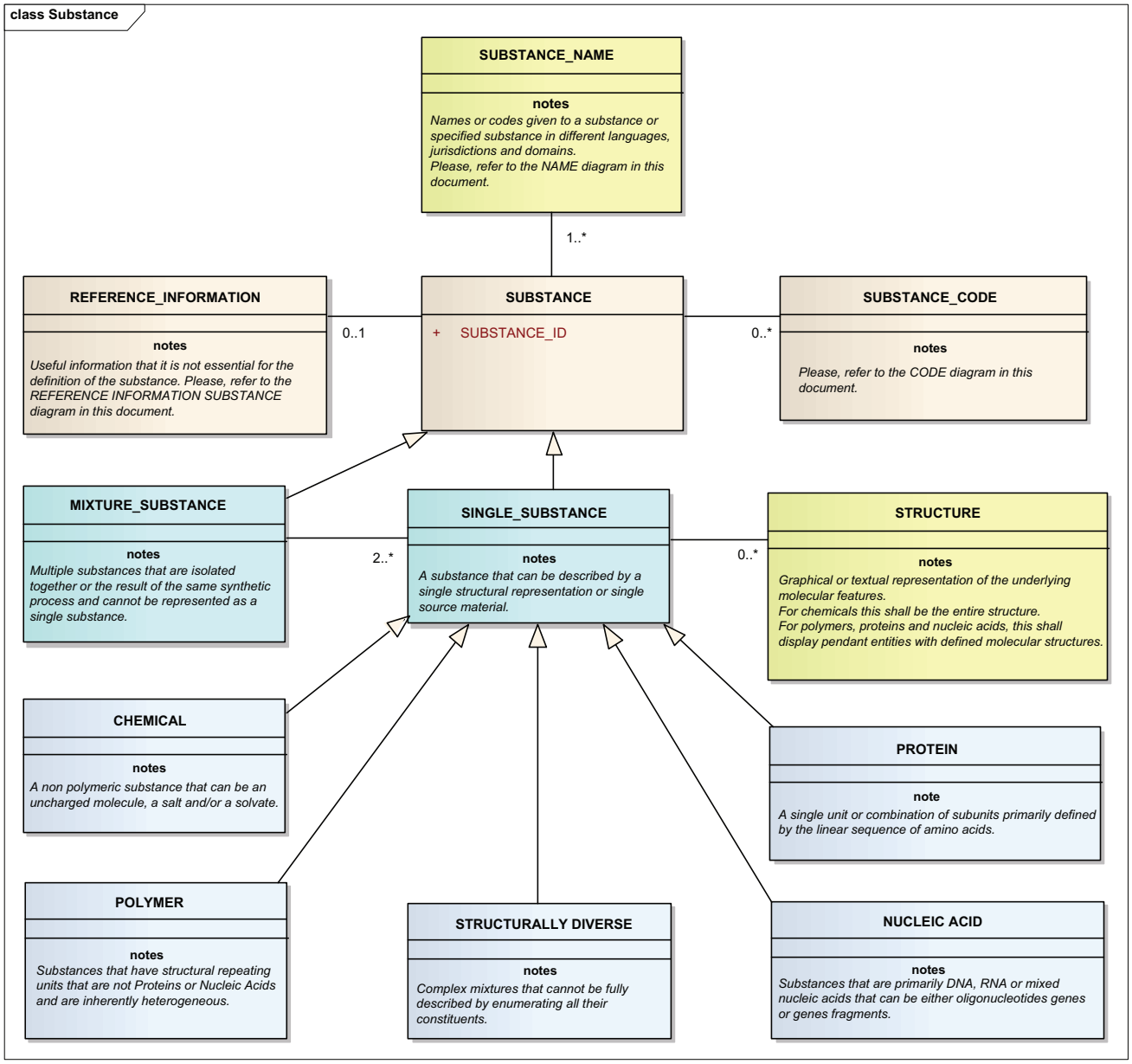

Figure 1 — High-level information model of substances

3.3 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 multiple 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: constituents (including components for material containing multiple substances, marker substance and extraction solvents for herbals and allergenic extracts), physical form and any physical property that is essential for defining the specified substance (e.g. size of liposomal preparation).

— Group 2: limited manufacturing information, parent substance or Group 1 specified substance, manufacturer, high-level production method, overall production method type (e.g. synthetic, extractive, recombinant) production system type (i.e. cell line, plant or animal tissue), production system (specific cell line).

— Group 3: parent substance or Group 1 specified substance, grade and source of grade (pharmacopeia, technical);

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

The relationship between specified substances is shown in Figure 2.

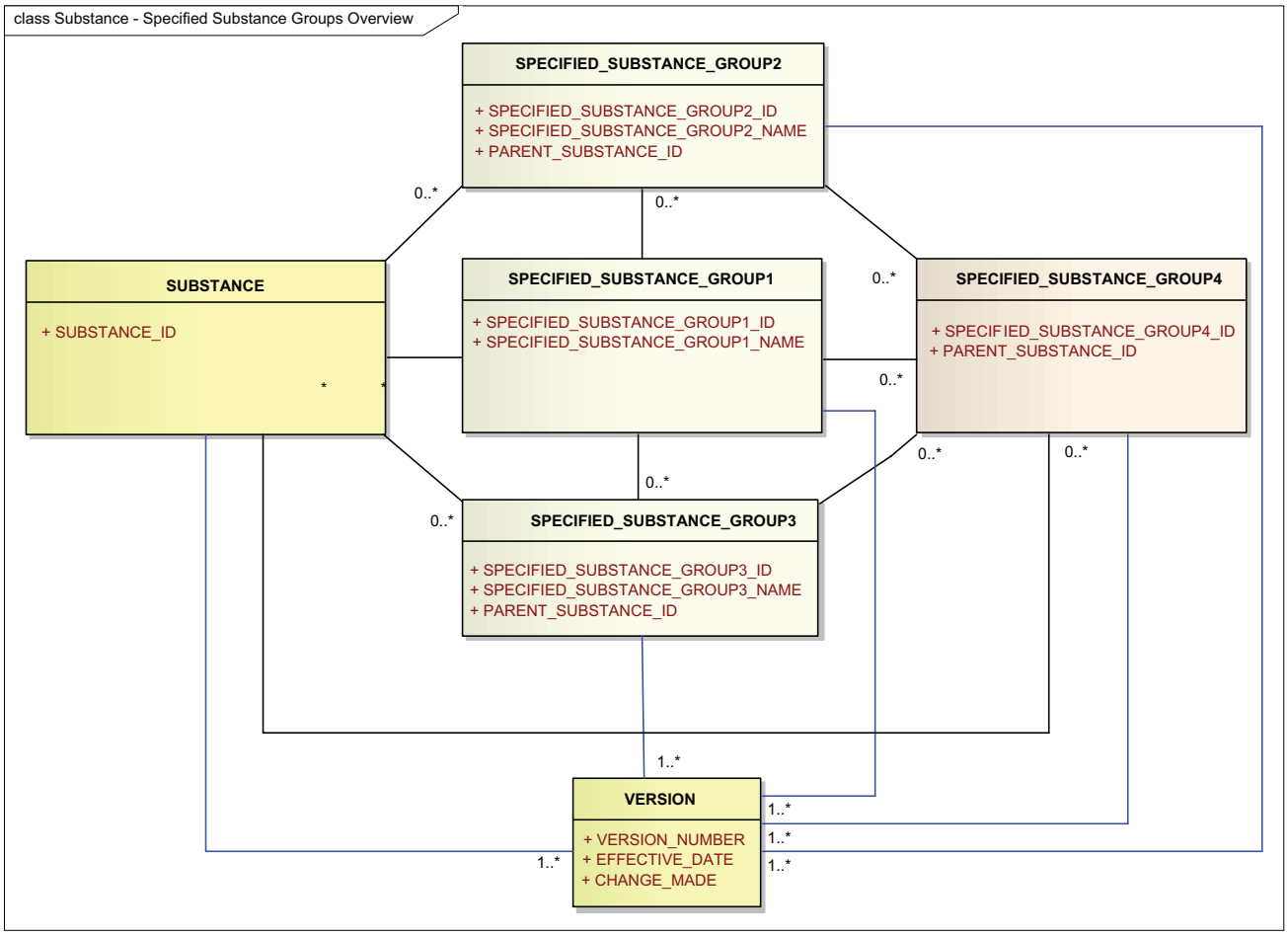

Figure 2 — High-level specified substance information model

3.4 Naming of substances

At least one substance name or company code shall be associated with each substance.

If the name is an official name, the naming authority, language and jurisdiction in which the name is used shall be identified.

This International Standard 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 in this International Standard. Language and jurisdiction will be described using ISO standards.

The information model for the class name is shown in Figure 3.

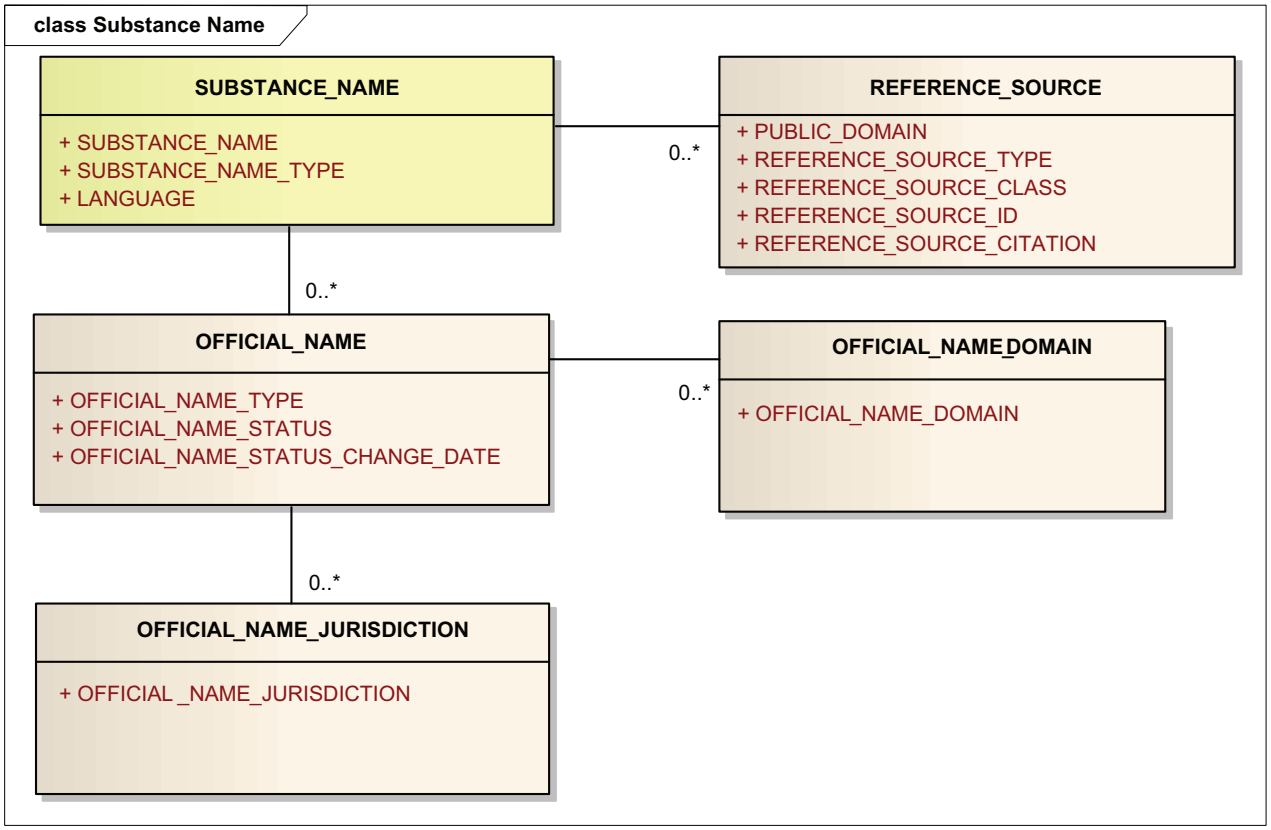

Figure 3 — Information model for substance names

3.5 Requirements for unique identifiers

Each substance and specified substance shall have only one permanently associated unique identifier that shall not indicate the order of submission to the system.

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

The unique identifiers shall be publicly available and their use royalty free.

A unique identifier shall be assigned to approved and investigational substances, excipients and impurities, solvents, ions, fragments and moieties.

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 an 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.

The information model for the relationship between names, unique identifier, substance and specified substance is shown in Figure 4.

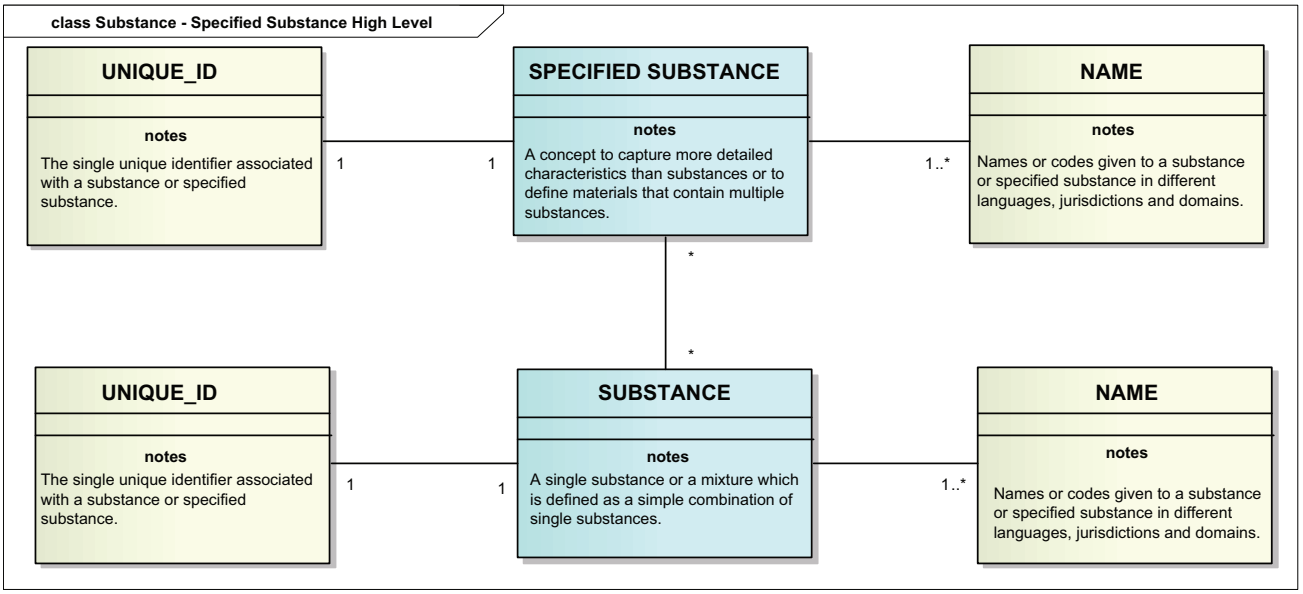

Figure 4 — Information model for substance, specified substances, names and identifiers

3.6 Types of substances

3.6.1 General

If it is possible to represent a substance as a single substance or a mixture substance, the substance shall be represented as a single substance.

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

3.6.2 Element sets common to multiple types of substances

3.6.2.1 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.

3.6.2.2 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 could include half-life, type and energy of emission, parent and daughter nuclides.

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

The information model for structure and isotope is shown in Figure 5.

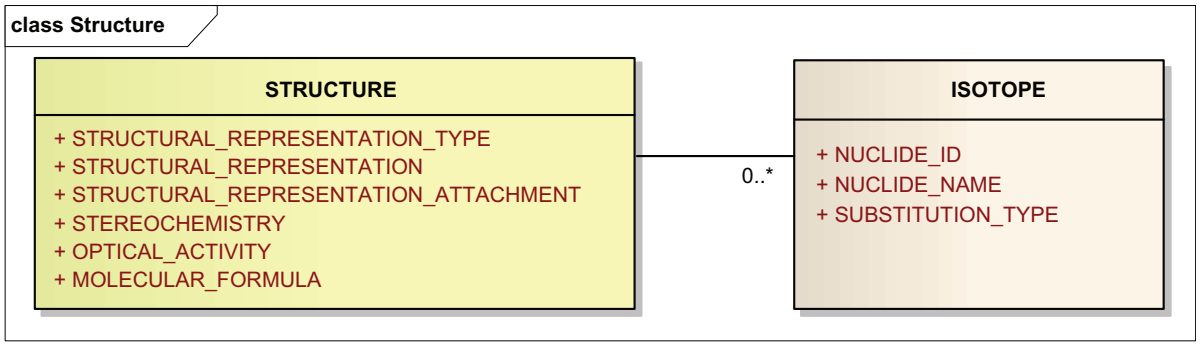

Figure 5 — Information model for structure and isotope

3.6.2.3 Modification

Irreversible changes in the underlying molecular structure of a substance shall be described as a modification of the antecedent material. Modification of a 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 and structurally diverse material shall be captured using modification elements. Modifications shall be represented as the addition of moieties or molecular fragments to the underlying material when definitive structural modifications occur, but the actual position of substitution is 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 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 6.

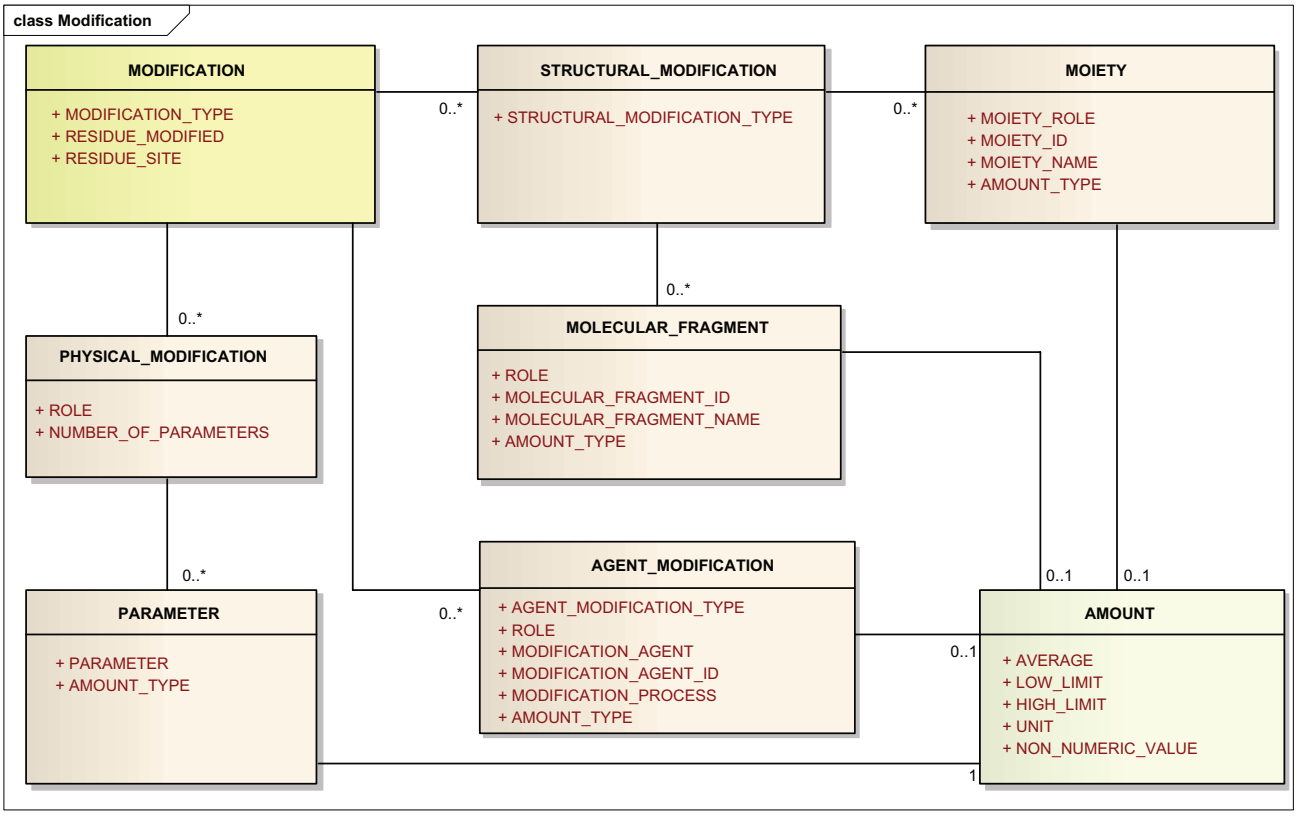

Figure 6 — Information model for modifications

3.6.2.4 Reference information

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 International Standard 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 International Standard does allow for the capture of such information if provided or available. This information shall not affect the generation of a new unique identifier.

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

NOTE Genes from which proteins are derived, target information and codes from code systems also constitute reference information for which this International Standard 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 to 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 7.

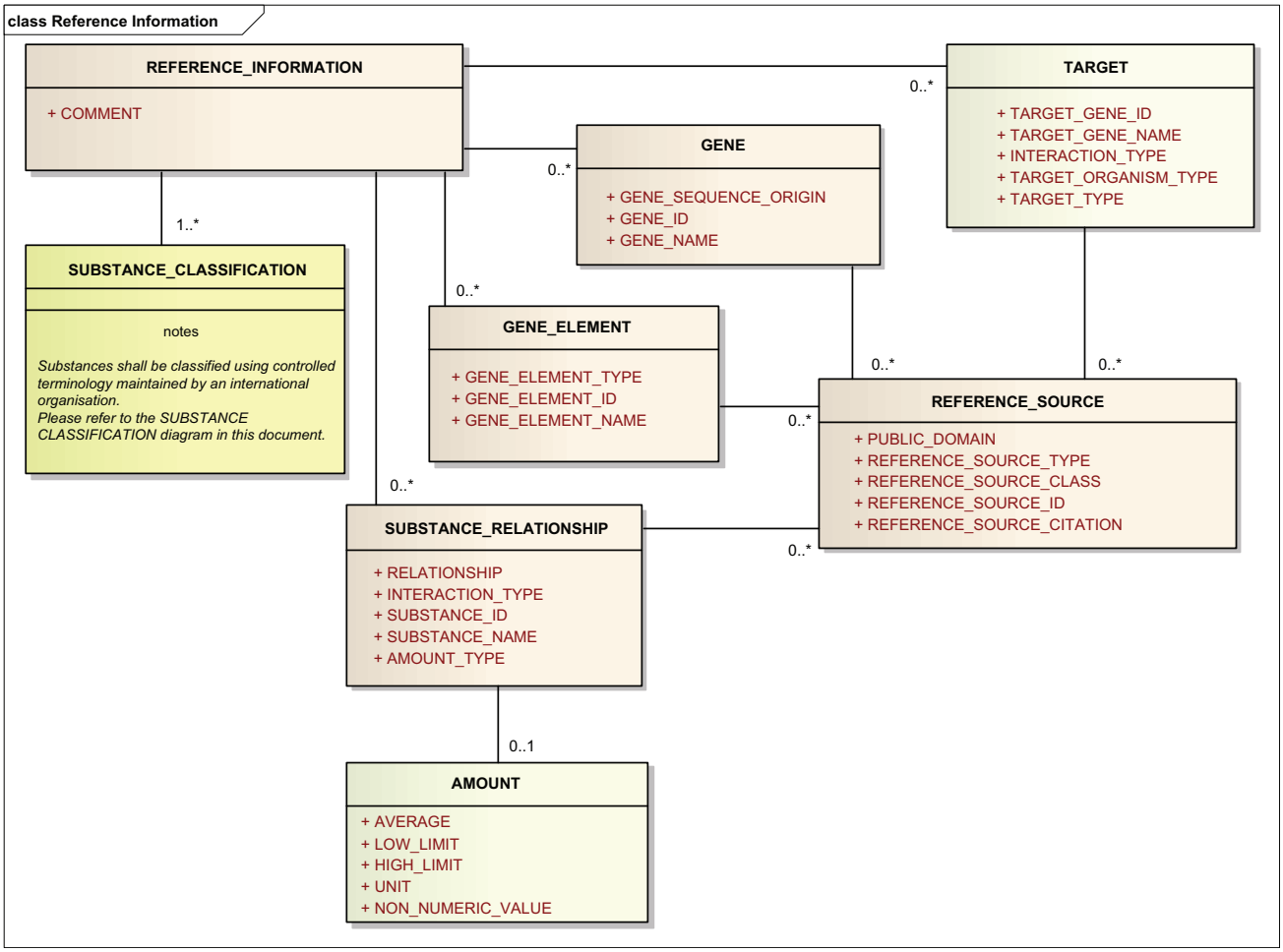

Figure 7 — Information model for reference information

3.6.2.5 Source material

Source material is an information class that shall capture 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. This class shall be used to define structurally diverse and polymer substances isolated from biological matrices.

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

The information model for the class source material is shown in Figure 8.

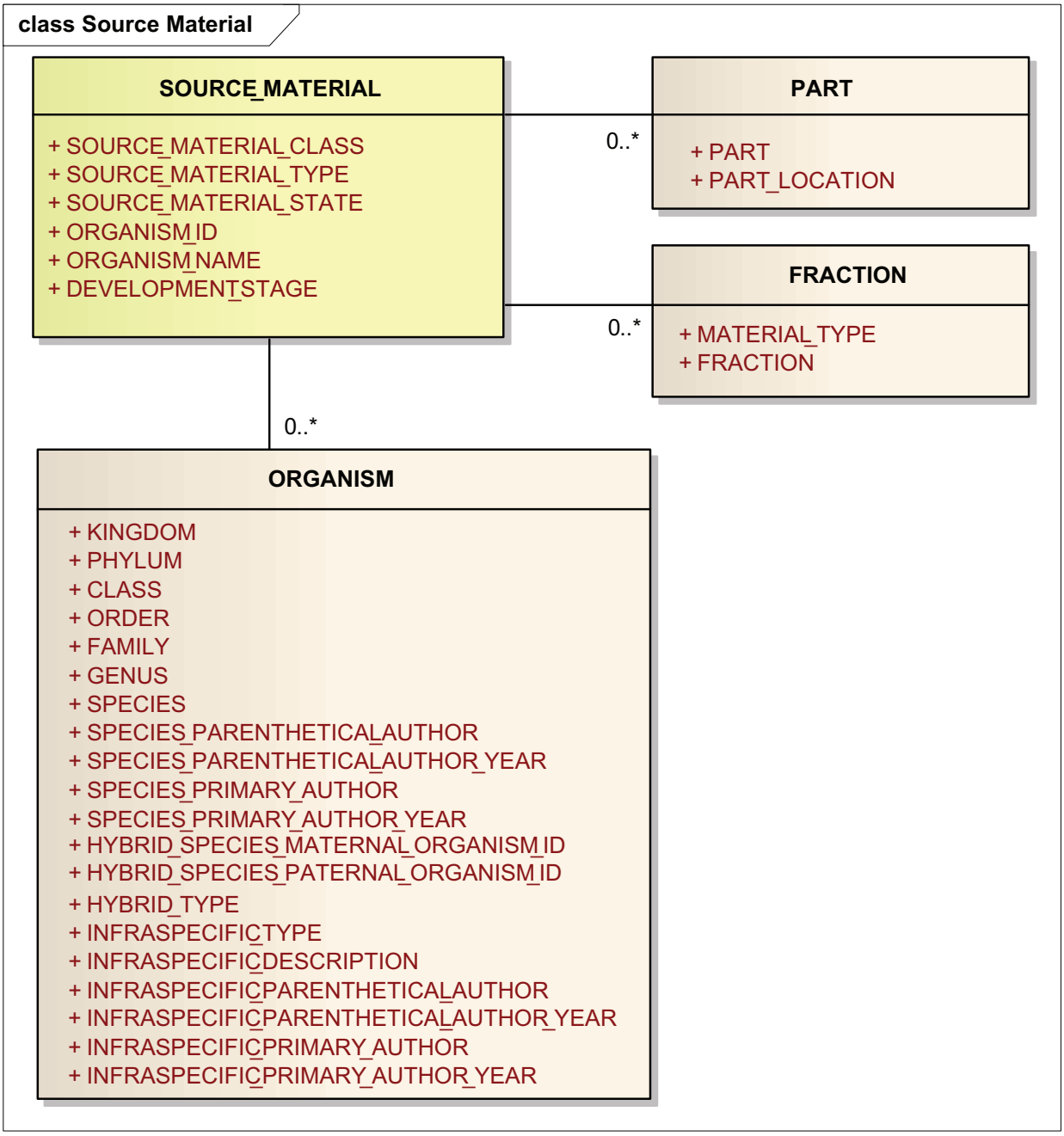

Figure 8 — Information model for source material

3.6.2.6 Taxonomy

Taxonomic information shall be captured for polymer and structurally diverse substances, and mixture substances derived from biological matrices.

This International Standard 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.

3.6.3 Chemical substances

Chemical substances shall be defined by a representation of the molecular structure and, when necessary, stereochemical and related physical characteristics.

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.

Like chemical substances, protein substances described in 3.6.4 and nucleic acid substances described in section 3.6.5 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 information model for the class chemical substance is shown in Figure 9.

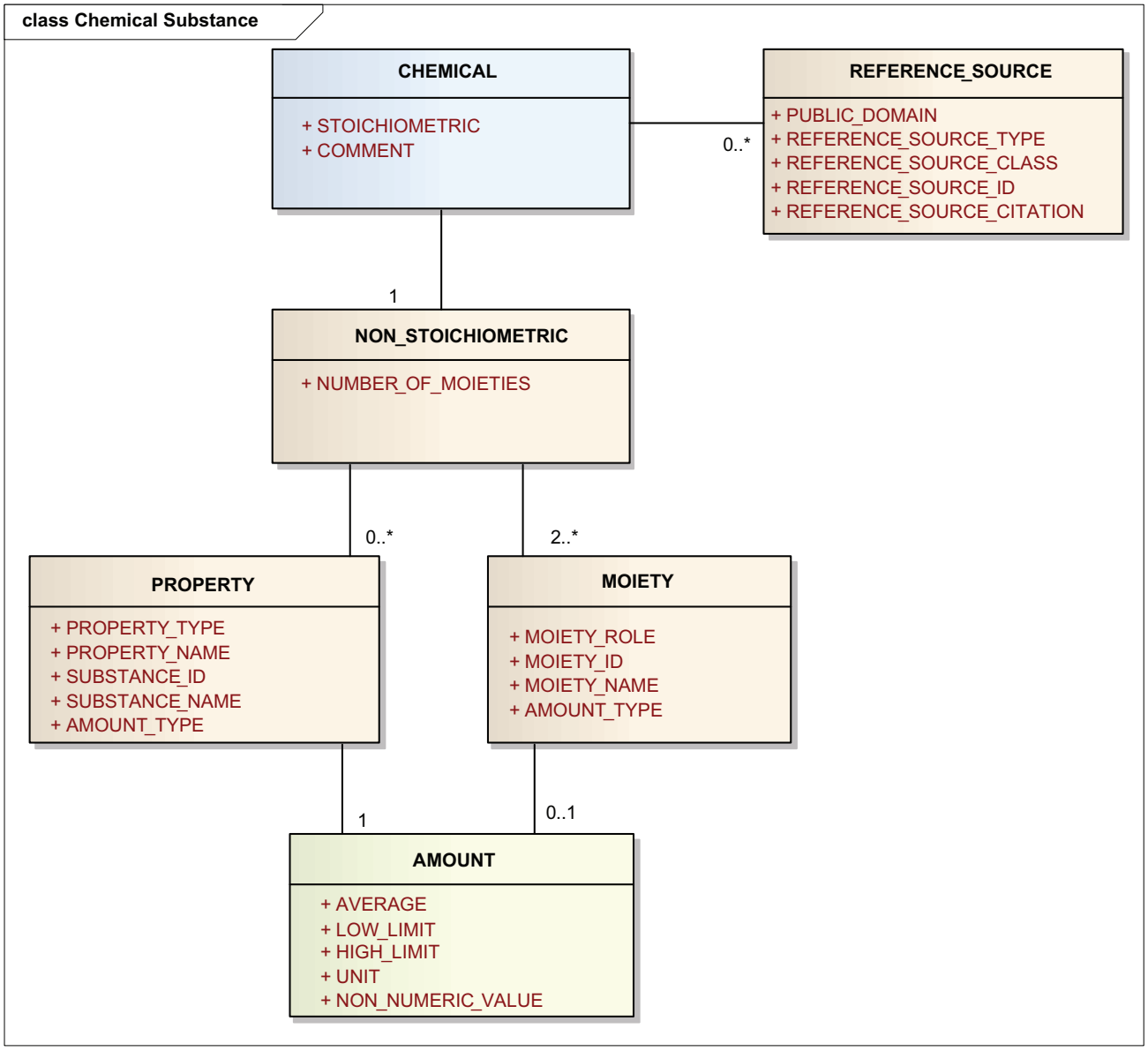

Figure 9 — Information model for the chemical substance class

3.6.4 Protein substances

All substances that contain a defined sequence of L-alpha-amino acids connected through a peptide bond shall be described using the protein elements.

NOTE 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 two separate substances.

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

Proteins shall consist of one or more protein sub-units.

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.

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 terminology 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.

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 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.

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 reference information can also be captured, including:

— type and subtype of protein;

— ligand, substrate or target;

— type of interaction of the protein;

— gene from which the protein was derived.

Reference information shall be captured using controlled vocabulary where available.

NOTE 1 Monoclonal immunoglobulins are described as proteins. 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 2 The current glycosylation types include fungal, plant, arthropod, avian, mammalian and human.

NOTE 3 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 class protein substance is shown in Figure 10.

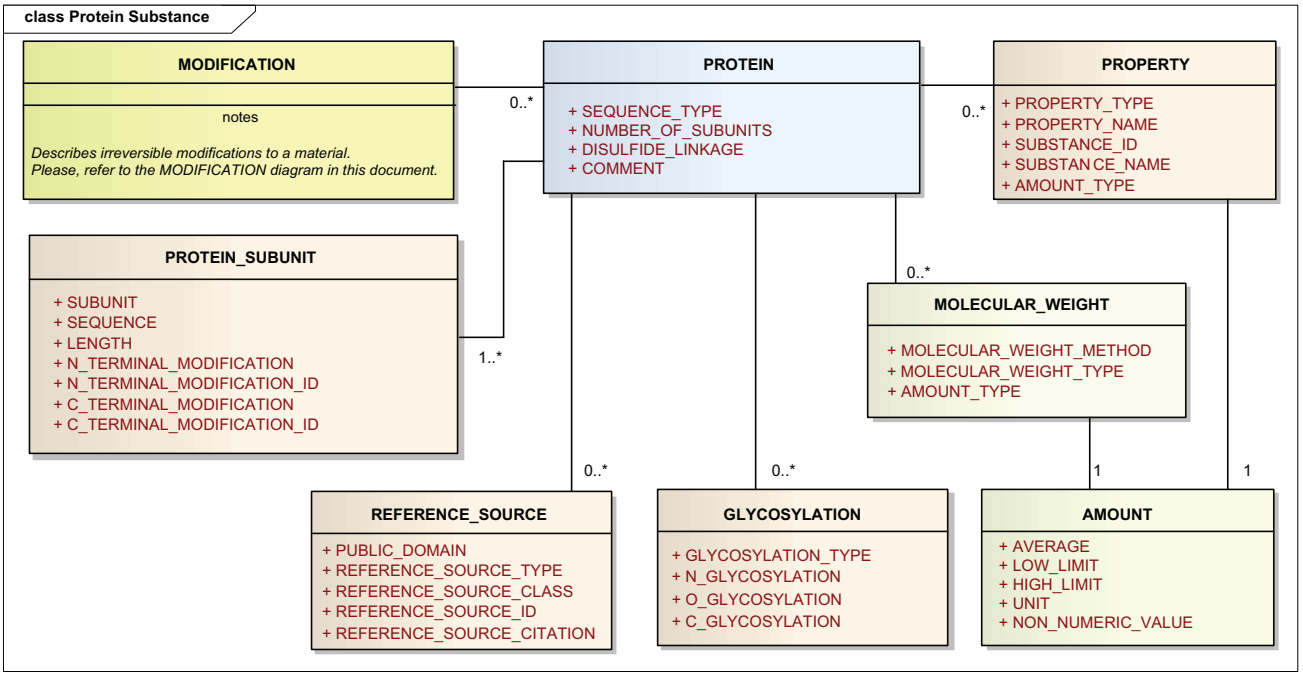

Figure 10 — Protein substances Class

3.6.5 Nucleic acid substances

The sequence of the nucleic acid, the type (RNA, DNA, mixed), 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 substances.

Irreversible 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 Gene elements would include promoters, enhancers, silencers, etc. For nucleic acids used in gene therapy, the entire sequence of the transforming/transducing entity would be captured along with each gene element.

The information model for nucleic acid class is shown in Figure 11.

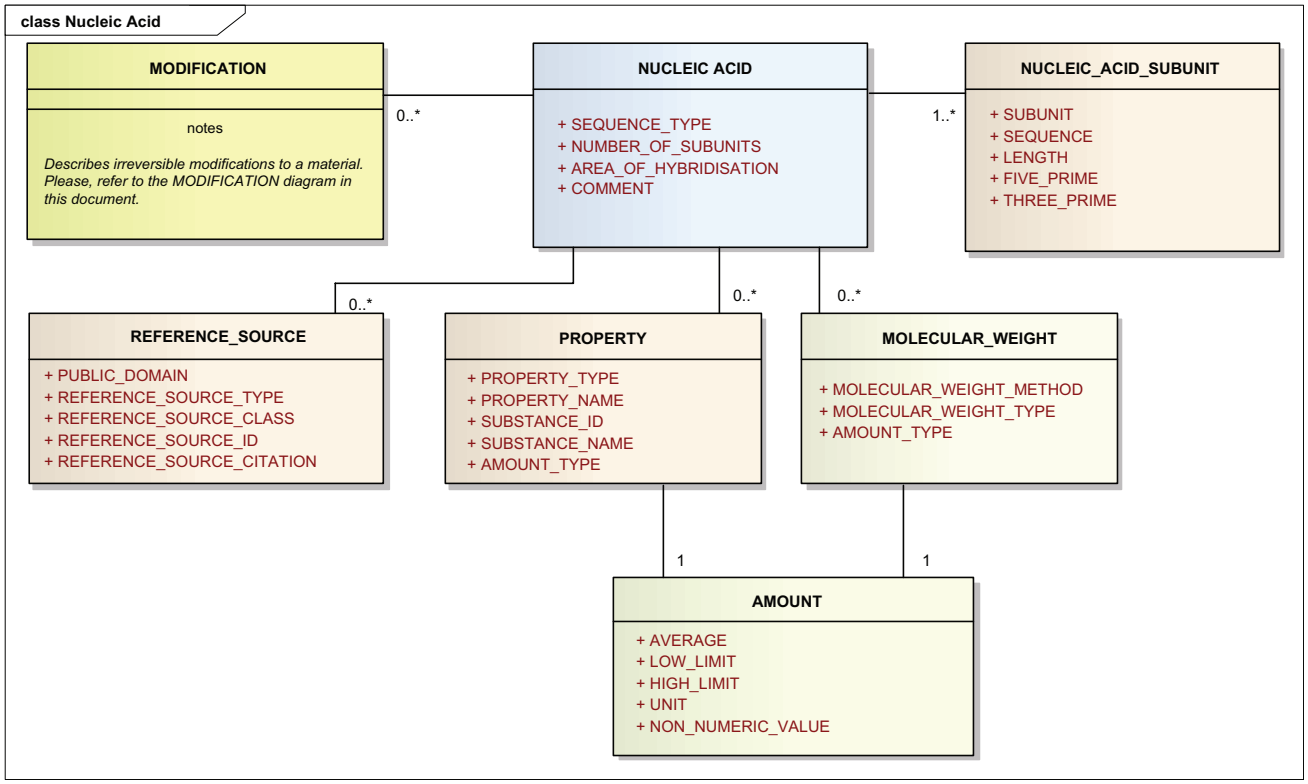

Figure 11 — Information model for the nucleic acid substance class

3.6.6 Polymer substances

Polymers shall refer to material that is inherently heterogeneous and contains structural repeating units.

Polymers shall be defined using a combination of controlled vocabulary and representations of the molecular structure of the structural repeating units, substituents that are attached to the structural repeating unit, which are described as either fragment or moiety modifications, molecular weight or the polydispersity of the material. The degree of polymerization, monomers 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 repeating units should always be defined based on either degree of polymerization or molecular weight. A generic polyethylene glycol substance should not be 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 type shall be defined by the number of structural repeating units and the connectivity between them. A controlled vocabulary shall be developed as required to describe the polymer class, polymer geometry and 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 and network; values for copolymer sequence type would include random, block and alternating. Polydispersity 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 information model for polymer substance is shown in Figure 12.

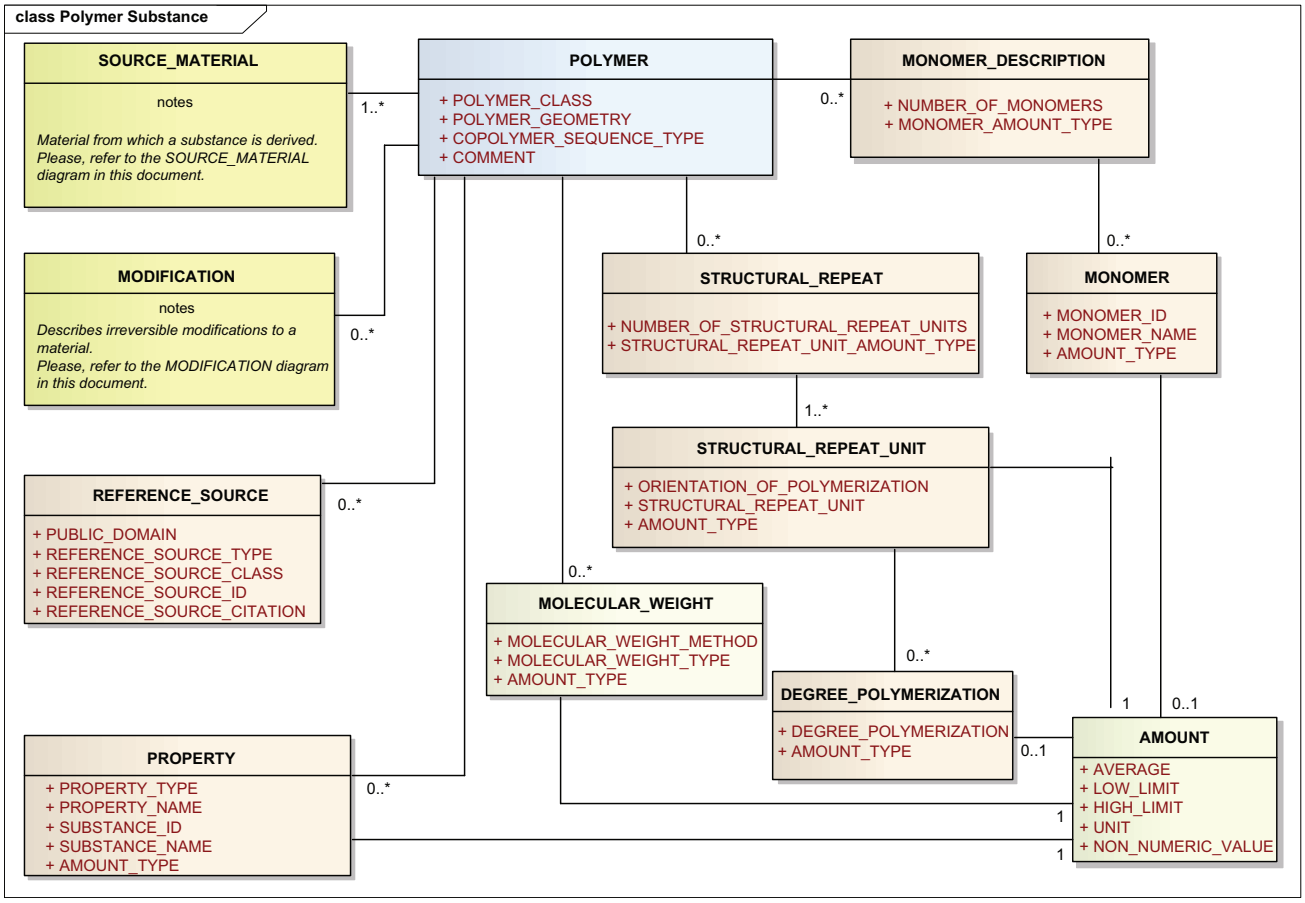

Figure 12 — Information model for the polymer substance class

3.6.7 Structurally diverse substances

Substances that cannot be defined as a limited number of related single substances, i.e. mixture substances, shall be defined as structurally diverse.

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.

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

NOTE1 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 polydisperse 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.

NOTE 2 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 fertilizer, 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).

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, or to be accurately described and distinguished from related substances.

EXAMPLE 2 Broccoli and broccoli extract would be defined as the same substance. 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.

Polyclonal immunoglobulins shall be described as structurally diverse materials and require identification of the immunoglobulin type, targeted antigen and the species of origin.

Cells and tissues shall be described as structurally diverse substances. Information on individual donors or extent of pooling shall not be captured at the substance level.

Commodity oils, juices and exudates of plants shall be separate substances. Oils and juices 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. Therefore, tinctures, infusions and decoctions shall not be defined as separate substances but will map to the parent organism and part from which they were derived.

EXAMPLE 3 Olive oil is *Olea europaea* fruit oil. Orange juice is *Citrus sinensis* fruit juice. Green tea and green tea extracts shall be defined as the leaves of *Camellia sinensis*.

The information model for the structurally diverse substance class is shown in Figure 13.

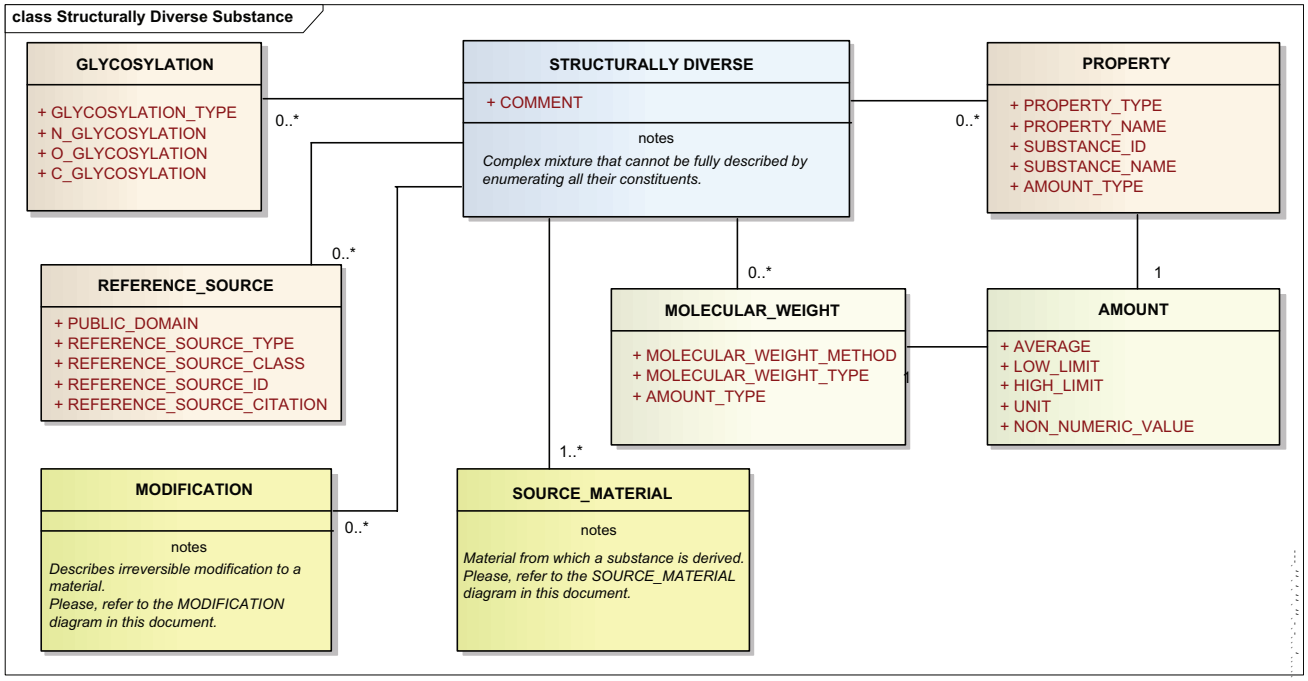

Figure 13 — Information model for the structurally diverse substance class

3.6.8 Mixture substances

Mixture substances shall be described as simple combinations of single substances that are either isolated together or are the result of 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 as a Group 1 specified 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.

This International Standard shall indicate whether a single substance is always present in “Any Of” mixtures.

The relative amount of each single substance shall not be captured.

NOTE Relative amounts of substances in a mixture substance shall be captured at the specified substance level consistent with either a pharmacopoeial or manufacturer specification. Relative amounts are not captured at the substance level because of potential variation in specifications between manufacturers and pharmacopeias.

All mixture substances shall consist of mixtures of single substances.

Mixtures of mixture substances shall not be allowed.

Mixtures of mixtures 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 constituents of the mixture substance.

Impurities and degradents shall generally not be considered constituents of a mixture substance.

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

The information model for the mixture substance class is shown in Figure 14.

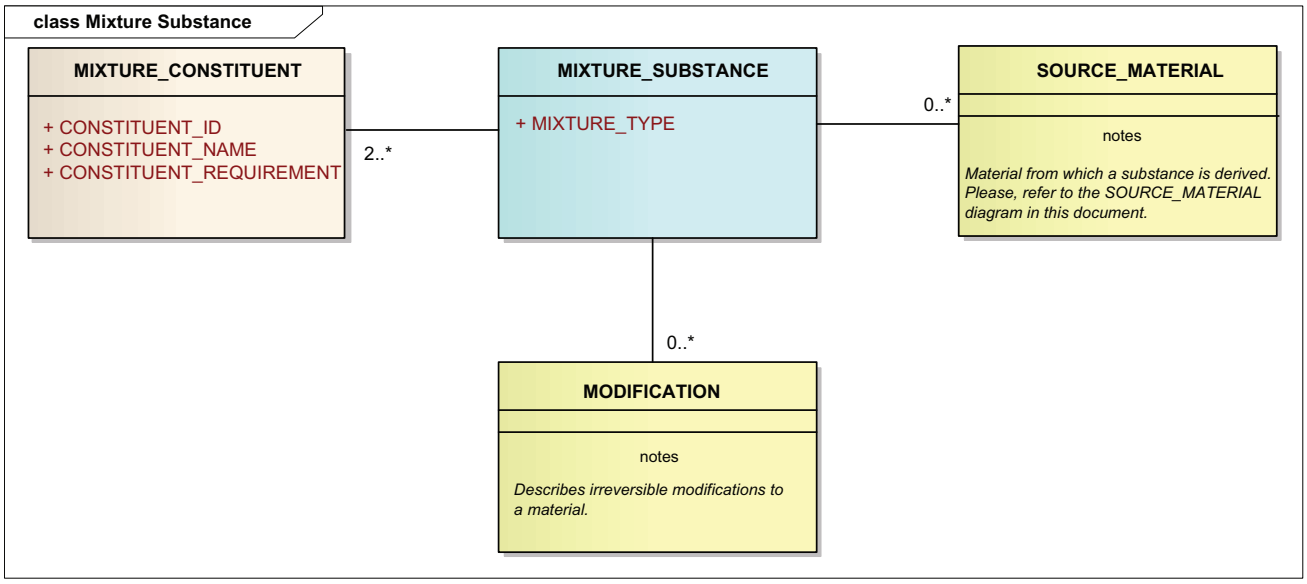

Figure 14 — Information model for the mixture substance class

3.6.9 Substance codes

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

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

The information model for the substance code is shown in Figure 15.

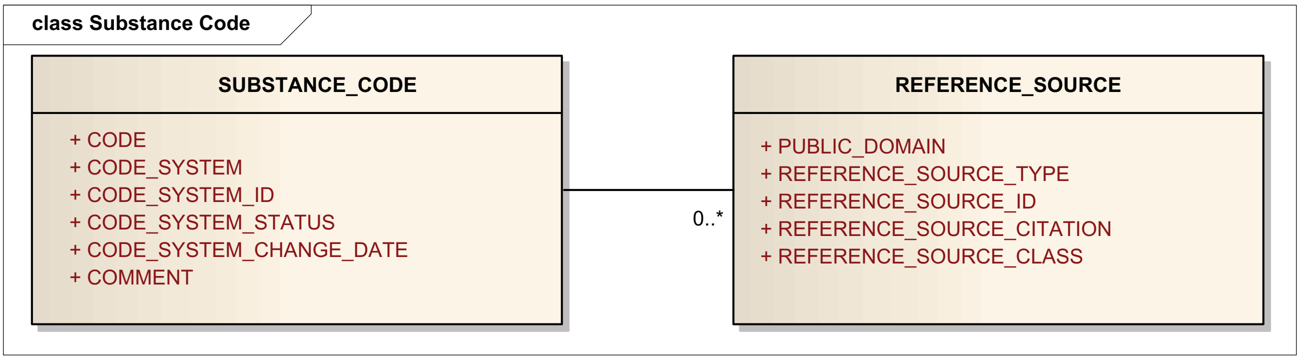

Figure 15 — Information model for substance codes

3.6.10 Substance classification

Substance classification, although not defining elements for a given substance, can be captured according to the information model presented below.

The information model for substance classification is shown in Figure 16.

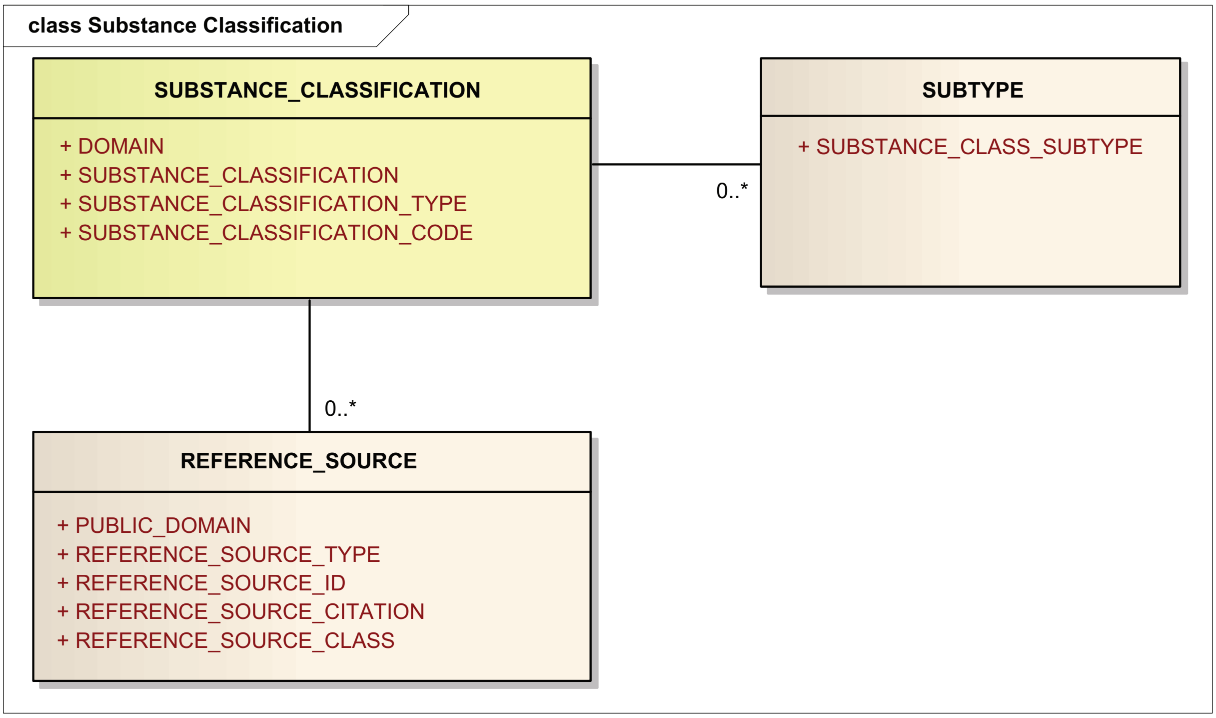

Figure 16 — Information model for substance classification

3.7 Defining specified substances

3.7.1 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 critical processes in manufacturing;

— analytical data;

— constituents, including amounts and role when known and relevant;

— specifications (impurities and related substance limits would be captured using constituents);

— unitage;

— reference material;

— analytical methods used for potency determination.

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 implementation.

3.7.2 Group 1 specified substances

Elements shall be used to describe material that contains multiple substances, 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, and any properties essential to the description of the material.

The element groups used to define a Group 1 specified substance shall include constituents, physical form and property. The information model is shown in Figure 15.

Constituent groups shall consist of intended substances added to create a multi-substance material, solvents used in the preparation of extracts, marker or signature substances present in animal derived material.

Impurities or degradents shall not be constituents for Group 1 specified substances.

NOTE 1 This grouping of constituents allows for the definitions of many materials in commerce that are used in the formulation of medicinal products.

NOTE 2 Impurities and degradents will be captured as part of a Group 4 specified substance.

The information model for the Group 1 specified substance class is shown in Figure 17.

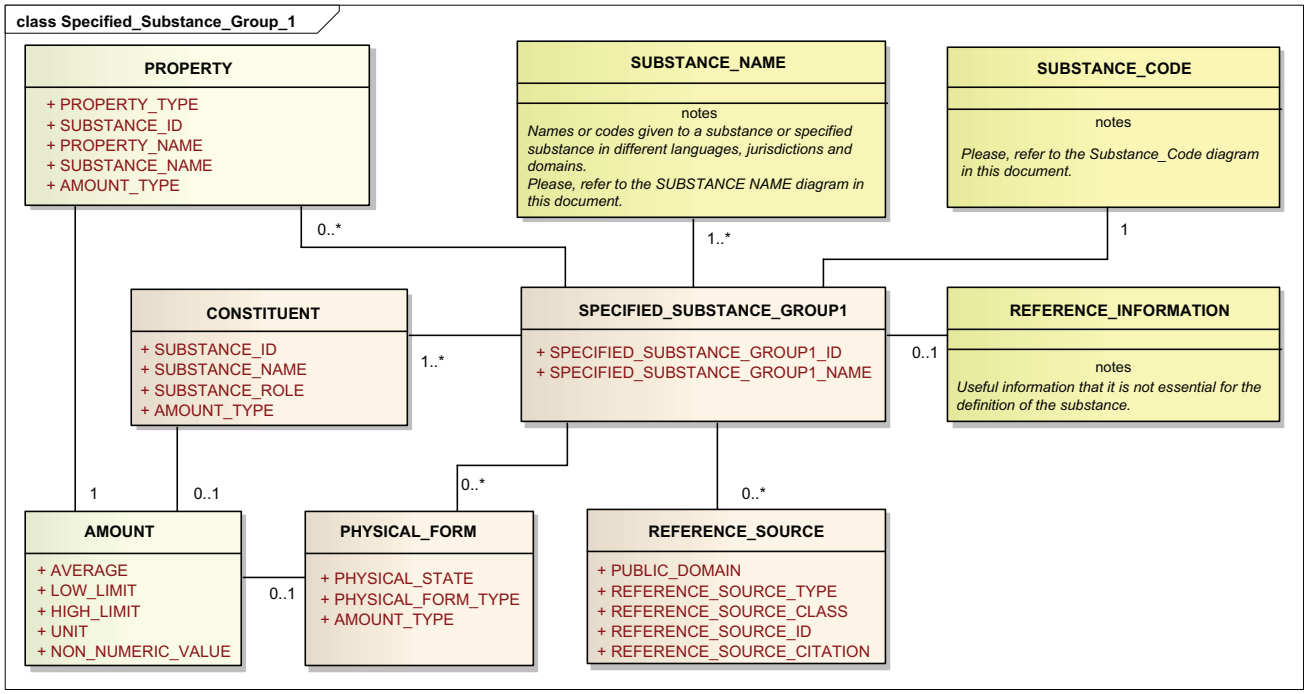

Figure 17 — Information model for the Group 1 specified substance class

3.7.3 Group 2 specified substances

Elements shall be used to capture the manufacturer of either a substance or a Group 1 specified substance, along with minimal manufacturing information.

The minimal manufacturing information shall include the overall production method type (e.g. synthetic, extractive, recombinant), 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 Group 2 specified substances that have undergone major changes in the Critical Processes used in the manufacturing of the substance. The initial Critical Process Version Number shall be one and each subsequent number shall be increased sequentially.

NOTE Group 2 elements would allow the tracking of the substance to the manufacturer. 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.

The information model for the Group 2 specified substance class is shown in Figure 18.

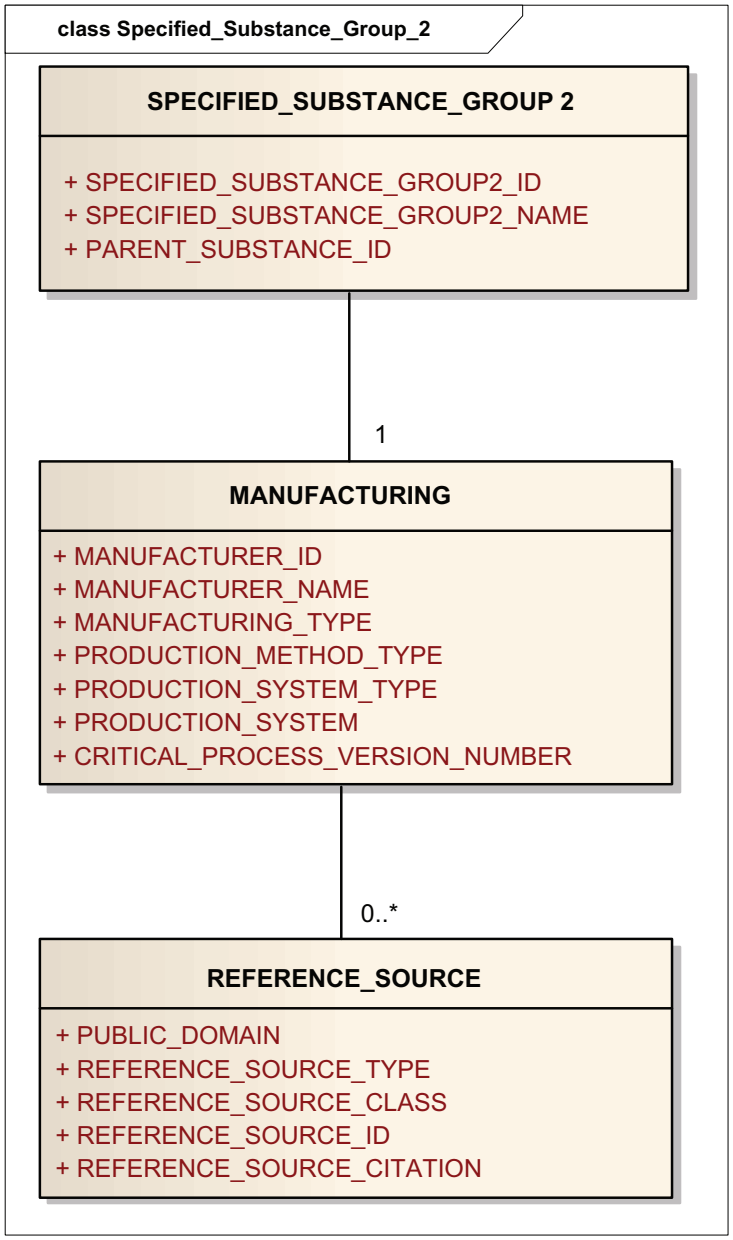

Figure 18 — Information model for the Group 2 specified substance class

3.7.4 Group 3 specified substances

Group 3 elements shall capture the grade of the material along with the source that defines the given grade.

Group 3 elements shall be used to distinguish specific pharmacopoeial and technical grades of material.

If the pharmacopoeial monographs related to a substance are not harmonized, the grade for each pharmacopeia shall be a separate Group 3 specified substance.

The parent substance shall refer to the substance or Group 1 specified substance to which the grade refers.

NOTE For most active pharmaceutical substances, typical grades are USP, EP, or JP. For herbal substances the grades would be standardized, quantified and unstandardized.

EXAMPLE Water is the parent substance for the Group 3 specified substance Sterile Water for Injection USP.

The information model for the Group 3 specified substance class is shown in Figure 19.

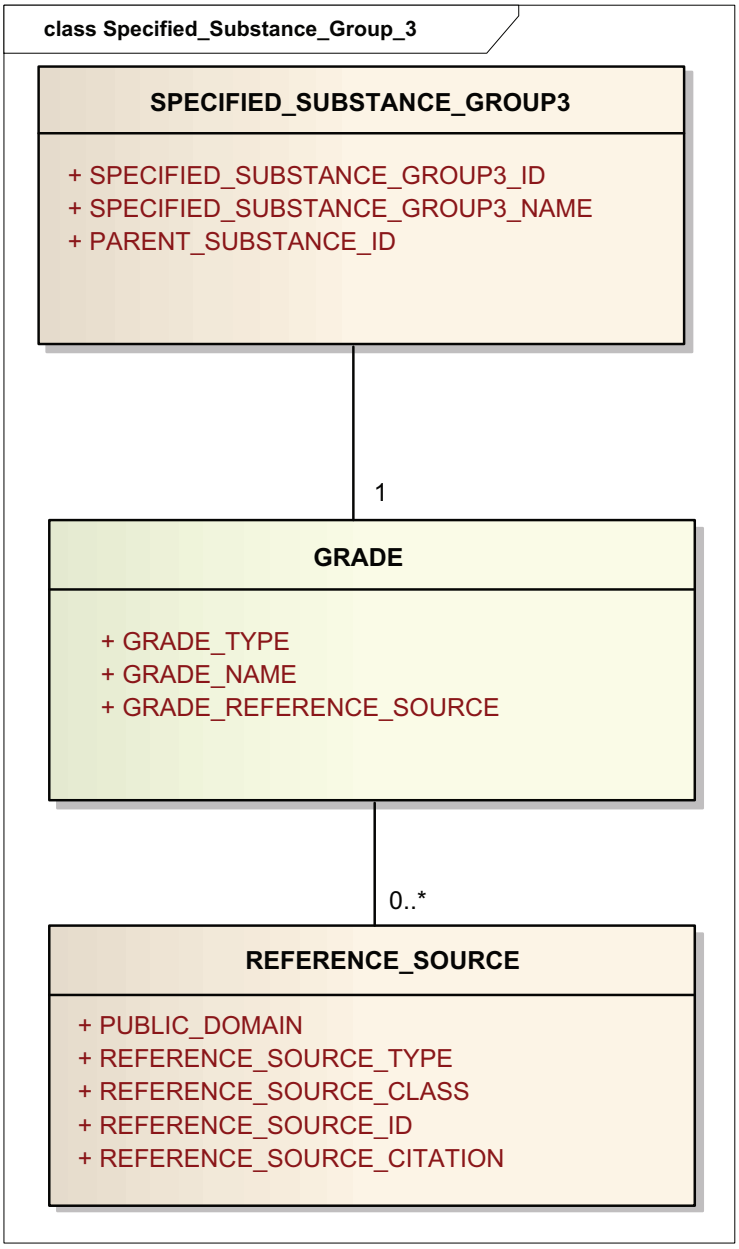

Figure 19 — Information model for the Group 3 specified substance class

3.7.5 Group 4 specified substances

Group 4 elements shall contain the most detailed information on a substance. This information shall include critical manufacturing processes, specifications (impurities and related substance limits would be captured using constituents), unitage, reference material and analytical methods used for potency and purity determinations.

NOTE The specific information described for Group 4 specified substances is often submitted in regulatory submissions in a diffuse manner that is difficult to capture and organize. The fields developed here will attempt to capture the data in a manner that will facilitate its use in both review and compliance activities.

The information model for the Group 4 specified substance class is shown in Figure 20.

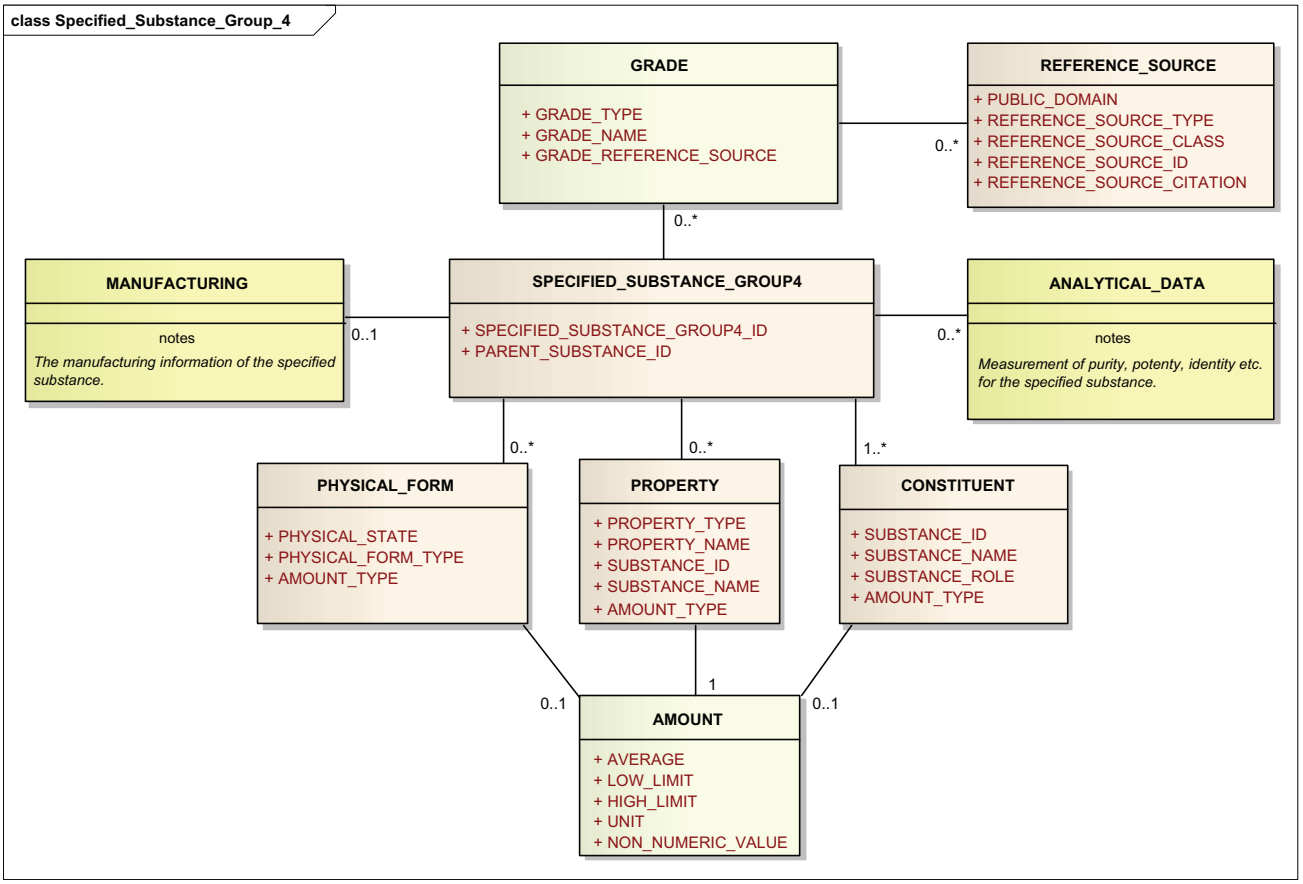

Figure 20 — Information model for the Group 4 specified substance class

3.7.6 Analytical data

Information on assays used to determine potency and specifications can be defined as a group of elements. This information shall reference an analytical method and any reference material used in the determination of potency. Individual reference materials shall be defined as specified substances. Different analytical methods or reference materials used for the determination of potency shall be described, captured and referenced. This International Standard does not prescribe that any analytical method or reference material be used for the determination of potency but merely provides a mechanism for the capture of such information.

NOTE Unitage for potency is often dependent on the analytical method and reference material used in the determination. In many instances, unitage can vary across jurisdictions or even among manufacturers within a jurisdiction.

EXAMPLE USP pancrelipase units and Fédération Internationale Pharmacéutique pancrelipase units differ and are not readily convertible because the reference materials are distinct and standardized in a different manner.

The information model for the analytical data class is shown in Figure 21.

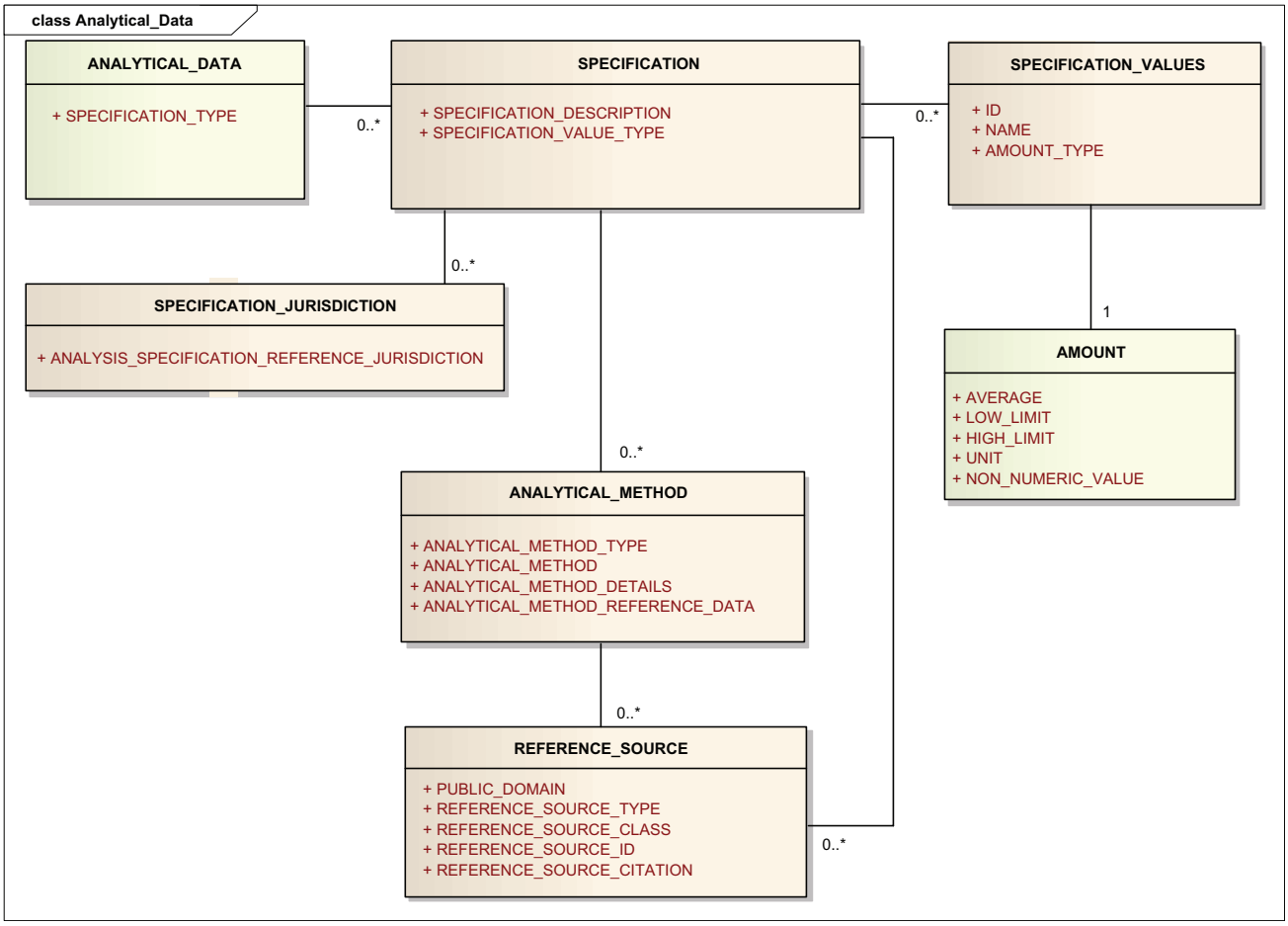

Figure 21 — Information model for the analytical data class

3.7.7 Constituent

This is the constituent element group that captures the substances and specified substances related to a given specified substance. Multi-substance material shall contain substances and specified substances defined as constituents. Active marker substances in herbal extracts, impurities or related substances shall also be constituents and can be linked to the components that make up a specified substance. The amount and role of each constituent shall be captured. Amounts shall either be expressed as definitive amounts (average) limits or ranges of concentration. Constituents will be used in both Group 1 and Group 4 specified substances.

NOTE The constituent group provides a great deal of flexibility and utility in that it can link many types of related substances together.

3.7.8 Grade

Grade refers to the overall quality or group of specifications of a given specified substance.

Grade shall be indicated for both Group 3 and Group 4 specified substances.

NOTE Pharmacopoeial grades or specifications will be referred to when available. A given specified substance could be compliant with specifications from multiple pharmacopeias. Technical grades could also be indicated.

EXAMPLE Pharmacopoeial grades or specifications will be referred to when available. The specified substance would distinguish different grades of water from each other. Tap water, USP Sterile Water for Injection, EP Sterile Water for Injection, USP Purified Water etc. would each be defined as separate specified substances.

3.7.9 Manufacturing

The manufacturing element group shall capture information on the manufacturer and critical manufacturing processes that are necessary to distinguish specified substances. The starting materials, processing materials, critical process parameters, equipment used and the resultant material from the manufacturing process can be captured within this element group.

NOTE The manufacturing group is not intended to capture all the details of manufacturing but only the critical processes that could impact the safety or efficacy of a specified substance used in a medicinal product.

The information model for the manufacturing class is shown in Figure 22.

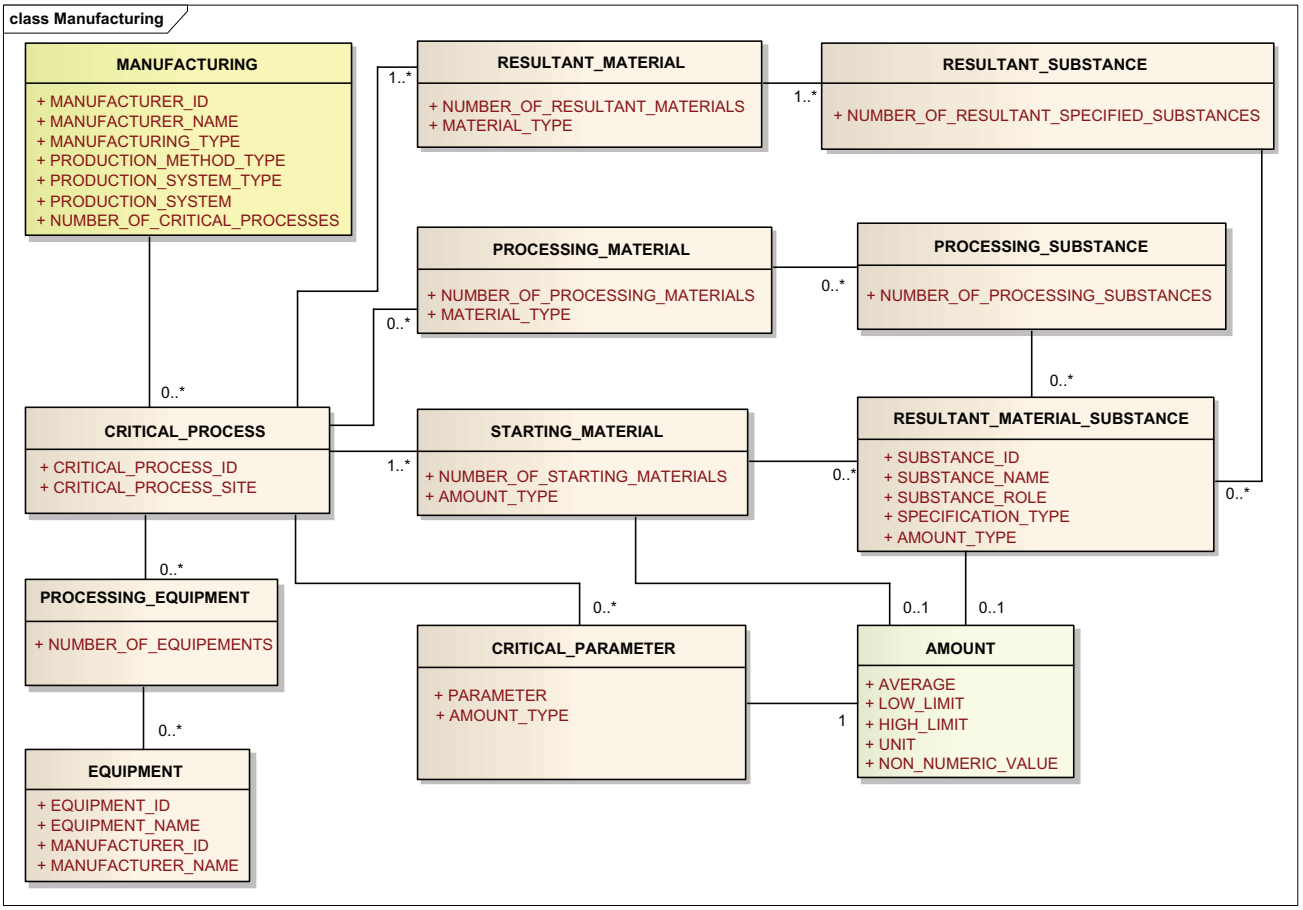

Figure 22 — Information model for the manufacturing data class

3.7.10 Physical form

Physical form shall capture the state in which the specified substance exists. A controlled vocabulary shall be used and developed as required.

3.7.11 Property

Physical, chemical or biological properties shall be captured which can be related to the use or function of a specified substance and which may be independent of the molecular structure or molecular ensemble that makes up the substances or specified substances within a specified substance. This element group may be used to capture properties related to the supramolecular organization of materials. Properties shall be used to define both specified substances Group 1 and Group 4.

Annex A
(informative)

Existing identifiers and molecular structure representations

A.1 Identifiers

A.1.1 General

Below are descriptions of commonly used identifiers emphasizing the strengths and weaknesses of each identifier for use as a unique identifier for substances in pharmaceutical products. The list is not exhaustive but describes identifiers that are actively being used in data systems.

A.1.2 CAS Registry numbers

CAS Registry Numbers are numeric identifiers that usually only identify a single substance. Polymers frequently only have one CAS registry number associated with them, regardless of differences in molecular mass or other defining elements. The numbers are sequential and are assigned as a substance enters the registry system. The numbers do not have a common length and lengths vary from five to 10 digits. Each CAS number contains a single check digit. Over 100 million substances are referenced in the CAS registry system. The primary purpose of the CAS registry system is to link to information in the chemical literature and not necessarily to identify or define substances. The CAS registry system is maintained by the Chemical Abstracts Service of the American Chemical Society. Although CAS numbers are widely used, they cannot be freely used. CAS has guidelines on the use of CAS registry numbers and has attempted to restrict their use in publicly available databases. The CAS number for formaldehyde is 50-00-0.

A.1.3 InChI and InChIKey

InChI stands for IUPAC International Chemical Identifier. The system was primarily developed at the National Institute of Standards and Technology in the USA. InChI is a linear identifier that deals with chemical representation using a layered approach. InChI is a non-proprietary structural representation and the software necessary to generate InChIs are provided under an open-source LGPL license. An InChIKey is a fixed length (25 characters) condensed digital representation of the InChI. InChI and InChIKey is really only designed for simple substances that can be defined by a representation of molecular structure and not complex products such as vaccines, blood-derived products, botanicals or animal products. The InChI for morphine is InChI = 1/C17H19NO3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11, 13,16,19-20H,6-8H2,1H3/t10-,11-,13-,16-,17-/m0/s1 and the InChIKey for morphine is BQJCRHHNABKAKU-XKUOQXLYBY.

A.1.4 EC Number

The EC-No. or EC# is a seven-digit code that has been allocated by the European Commission for all commercially available substances marketed within the European Union. The seventh digit of the code is a check digit and the code maps to both common and trade names of a given substance. The scope of the EC number is broader than that of InChI in that both simple and complex substances have been assigned EC#s. The system contains over 100 000 substances but is not heavily weighted in the pharmaceutical sector. The codes are also for the most part sequential and were developed from the EINECS (European Inventory of Existing Commercial Chemical Substances), ELINCS (European List of Notified Chemical Substances), and other lists of regulated substances. The EC# for formaldehyde is 200-001-8. The European Commission is no longer generating new EC numbers.

A.1.5 UNII

The UNII is a 10-character, randomly generated alpha-numeric string that is currently used to identify substances in medicinal products. The UNII is generated by the FDA/USP Substance Registry System, a robust system

with detailed business rules for data entry and generation of UNIIs for both simple and complex substances. The first nine characters are randomly generated followed by a check character. The integrity check on the UNII is stronger than both the EC# and the CAS Registry Number because of the random generation from a large number of potential UNIIs and the fact that there are 36 possible check characters compared to 10 with both the EC# and CAS Registry Number. The UNII is freely available for use and there is a mechanism whereby a manufacturer can petition for the generation of a UNII through the FDA. The system has the capability for both public and restricted access to information, and can be adapted to produce specified substance identifiers.

A.1.6 ASK Number

The ASK Number is a five digit code (and check digit) and is issued and maintained by the German National Competent Authorities (BfArM, PEI, BVL), based on §10 German Drug Law and AMIS-Bezeichnungsverordnung, respectively. The ASK Number is mandatory for applications and correspondence between marketing authorization holders and competent authorities. The underlying substances database comprises more than 35 000 substances which are related to business in the regulatory environment. These are substances of chemical or biological origin, as well as radiopharmaceuticals, homeopathics and anthroposophics. The repository contains mainly active ingredients and excipients, but also gases, packaging materials, chemicals for analysis, impurities, and substances prohibited by law. In addition to the chemical name according to IUPAC, a large “collection” of synonyms of international and European sources throughout the lifecycle of a medicinal product are referenced. If applicable, the CAS Registry Number, molecular formula and molecular mass are available. In relation to the different aspects in the daily work of the regulatory authorities, extensive “grouping attributes” have been included for classifying the substances.

A.2 Molecular structure representations

A.2.1 General

Representation of the chemical or molecular structure is essential to the development of a controlled vocabulary for simple chemical structures. The system of representation should be both unambiguous and unique. Only one single representation will be allowed for a given structure, and the representation should have enough detail to ensure that unintended ambiguity does not exist. The representation, or a form of it, should be capable of being stored in a chemical database to facilitate registration and searching. There are other formats that are not described below which are either not widely used or are proprietary and only associated with one vendor.

A.2.2 Molfile

The molfile format was predominantly developed by MDL Information Systems. There are two versions in use today: V2000 and the extended molfile format V3000. The extended molfile format has enhanced stereochemistry descriptors that allow relative, unknown and racemic designations to be associated with each chiral atom. The V2000 format is widely used and interconversion between it and other formats can readily occur. Unlike other representations, the molfile format is not a linear representation but is predominantly tabular. Below is a V2000 molfile representation of benzene.

```
ACD/Labs0812062058

  6  6  0  0  0  0  0  0  0  0  1 V2000
    1.9050   -0.7932    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.9050   -2.1232    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.7531   -0.1282    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.7531   -2.7882    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.3987   -0.7932    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.3987   -2.1232    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  2  1  1  0  0  0  0
```

3 1 2 0 0 0 0

4 2 2 0 0 0 0

5 3 1 0 0 0 0

6 4 1 0 0 0 0

6 5 2 0 0 0 0

M END

$$$$

A.2.3 SMILES

Simplified molecular input line entry specification (SMILES) is a specification for an unambiguous linear representation of chemical or molecular structures using ASCII characters. It is predominantly used by Daylight Chemical Information Systems Inc., although an open source version has been recently developed. Canonical smiles is a SMILES string that is unique for each structure and can be used to ensure that duplicate structures are not entered into a database. Other linear representation forms for chemical structures include SYBYL line notation (SLN) and the older Wiswesser Line Notation, which was the first line notation for the representation of chemical structures. These other formats are not currently in wide use. Below is the SMILES representation of Benzene.

C1 = CC = CC = C1

A.2.4 InChI

The InChI format is described in A.1.3. It is a layered approach to chemical structure representation. There are currently four layers of information:

— constitutional: expresses pure connectivity of the atoms;

— stereochemical: includes conventional C-atom sp2 and sp3 stereochemistry;

— isotopic: enables isotopes to be distinguished;

— tautomeric: implements simple forms of rapid H-migration isomerization.

Below is the InChI representation of benzene.

1/C6H6 /c1-2-4-6-5-3-1/h1-6H

A.2.5 CDX and CDXML file format

This is a ChemDraw-specific format that stores structures as a series of nested objects. Since most chemists have access to ChemDraw, and CDX is the default storage format in ChemDraw, substance identifier requests should be able to accept information in CDX format. Because the CDX format is proprietary, further information exchange is hindered so conversion to one of the above non-proprietary formats will be necessary.

Bibliography

[1] ISO 1087-1, *Terminology work — Vocabulary — Part 1: Theory and application*

[2] ISO 1087-2, *Terminology work — Vocabulary — Part 2: Computer applications*[1)]

[3] ISO 3166-1, *Codes for the representation of names of countries and their subdivisions — Part 1: Country codes*

[4] ISO 8601, *Data elements and interchange formats — Information interchange — Representation of dates and times*

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

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

[7] ISO 11239, *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*

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

[9] ISO/IEC 2382-4, *Information technology — Vocabulary — Part 4: Organization of data*

[10] ISO/IEC 7064, *Information technology — Security techniques — Check character systems*

[11] ISO/HL7 27951, *Health informatics — Common terminology services, release 1*

[12] ISO/HL7 27953-1, *Health informatics — Individual case safety reports (ICSRs) in pharmacovigilance — Part 1: Framework for adverse event reporting*

[13] ISO/HL7 27953-2, *Health informatics — Individual case safety reports (ICSRs) in pharmacovigilance — Part 2: Human pharmaceutical reporting requirements for ICSR*

[14] HL7 Common Product Model, *Substance Model (POCP_MT080100) and Substance Specification Model (POCP_MT090100)[6]*

[15] HL7 Version 3 Standard, *Common Terminology Services HL7 Draft Standard for Trial Use DSTU Release 2, 14 October 2009, HL7 Inc.*

[16] HL7 Version 3 Standard, *Common Clinical Product Model, Release 1, Last Ballot: Normative Ballot 1 - January 2009*

[17] HL7 Version 3 Standard, *Common Product Model CMETS, Release 10, DSTU Ballot 3 - January 2010*

[18] IUPAC, Compendium of Chemical Terminology version 2.1.5

[19] ISBT 128, Standard Technical Specification version 3.6

[20] International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use Draft Consensus Guideline – Data Elements and Standards for Drug Dictionaries – M5 Revision 4, 2 February 2007. http:/www.fda.gov/Cder/guidance/6886dft.pdf

[21] Resources related to FDA's Substance Registration system http://www.fda.gov/ForIndustry/DataStandards/SubstanceRegistrationSystem-UniqueIngredientIdentifierUNII/default.htm

1) Withdrawn.

[22] NCI Thesaurus (NCIt) and other NCI Enterprise Vocabulary Services (EVS) resources and documentation: http://nciterms.nci.nih.gov/ncitbrowser/pages/multiple_search.jsf?nav_type=terminologies

[23] The Integrated Taxonomic Information System http://www.itis.gov/

[24] The Food and Drugs Act and Regulations and related Health Canada Guidelines: http://www.hc-sc.gc.ca/dhp-mps/prodpharma/index-eng.php

[25] The Natural Health Product Regulations and related Health Canada Guidelines: http://www.hc-sc.gc.ca/dhp-mps/prodnatur/index-eng.php

[26] Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use as amended by Directive 2002/98/EC, Directive 2004/24/EC and Directive 2004/27/EC http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/consol_2004/human_code.pdf

[27] "Guideline on the categorisation on New applications versus Variations", The Rules governing Medicinal Products in the European Union, Notice to Applicants, Volume 2C

[28] The Rules governing Medicinal Products in the European Union, Notice to Applicants, Volume 2A and Volume 2C http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/homev2.htm

[29] A Guideline on Summary of Product Characteristics, December 1999, (Doc. Ref. Notice to Applicants, Final – revision 0) http://ec.europa.eu/health/documents/eudralex/vol-2/index_en.htm

[30] Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/dir_2001_20/dir_2001_20_en.pdf

[31] Commission Directive 2003/63/EC of 25 June 2003 amending Directive 2001/83/EC of the European Parliament and of the Council on the Community code relating to medicinal products for human use http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/dir_2003_63/dir_2003_63_en.pdf

[32] Regulation (EC) No 726/2004 of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/reg_2004_726/reg_2004_726_en.pdf

[33] Regulation (EC) No 1394/2007 of the European Parliament and of the Council of 13 November 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No 726/2004 http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/reg_2007_1394/reg_2007_1394_en.pdf

[34] Commission Directive 2003/94/EC of 8 October 2003 laying down the principles and guidelines of good manufacturing practice in respect of medicinal products for human use and investigational medicinal products for human use.

[35] http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/dir_2003_94/dir_2003_94_en.pdf

[36] Guidelines on Pharmacovigilance for Medicinal Products for Human Use, Volume 9A of the Rules Governing Medicinal Products in the European Union

[37] http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-9/pdf/vol9_2007-07_upd07.pdf

[38] Guideline on declaration of botanical substances and botanical preparations in botanical medicinal products/traditional

[39] Botanical medicinal products in the SPC, 26 July 2007 (Doc. Ref. EMEA/HMPC/CHMP/CVMP/287539/2005)

[40] http://www.emea.europa.eu/pdfs/human/hmpc/28753905en2.pdf

[41] Guideline on adjuvants in vaccines for human use, 20 January 2005 (Doc. Ref. EMEA/CHMP/VEG/134716/2004)

[42] http://www.emea.europa.eu/pdfs/human/vwp/13471604en.pdf

[43] Guideline on Similar Biological Medicinal Products, 30 October 2005 (Doc. Ref. CHMP/437/04)

[44] http://www.emea.europa.eu/pdfs/human/biosimilar/043704en.pdf

[45] GUIDELINE ON PHARMACEUTICAL ASPECTS OF THE PRODUCT INFORMATION FOR HUMAN VACCINES. 26 November 2003 (Doc. Ref. EMEA/CPMP/BWP/2758/02) http://www.emea.europa.eu/pdfs/human/bwp/275802en.pdf

[46] Guideline on the acceptability of names for human medicinal products processed through the centralised procedure, 11 December 2007 (Doc. Ref. CPMP/328/98, Revision 5)

[47] http://www.emea.europa.eu/pdfs/human/regaffair/032898en.pdf

[48] Guideline on the Chemistry of new Active Substances, 17 December 2003 (Doc. Ref. CPMP/QWP/130/96 Rev. 1) http://www.emea.europa.eu/pdfs/human/qwp/013096en.pdf

[49] Standard Terms, Dosage Forms, Routes of Administration and Containers, EDQM, Fifth Edition, December 2004, Version 5.0.0 http://st.pheur.org/

[50] EudraVigilance Medicinal Product Dictionary (EVMPD) Version 2.0 Technical Specifications, 9 November 2004 (Doc. Ref. EMEA/140190/2004) http://eudravigilance.emea.europa.eu/human/docs/EVMPD%20Technical%20Specifications.pdf

[51] EudraVigilance Medicinal Product Dictionary (EVMPD) Version 2.0 Message and Acknowledgement Specifications, 8 December 2004 (Doc. Ref. EMEA/178966/2004) http://eudravigilance.emea.europa.eu/human/docs/EVMPD%20Message%20and%20%20Acknowledgem

[52] EudraVigilance (EV) Access simple Database Version 2.0 8 November 2004 (Doc. Ref: EMEA/140327/2004) http://eudravigilance.emea.europa.eu/human/docs/EV%20Access%20Simple%20Database%202.0%20-%20Tables%20Documentation.pdf

[53] EudraVigilance (EV) Access Simple Database Version 2.0 Forms Documentation, 31 January 2005, (Doc. Ref: EMEA/35416/2005) http://eudravigilance.emea.europa.eu/human/docs/EVAccessSimple%20Database_v20_Formsdocumentation.pdf

[54] EudraVigilance (EV) Access Simple Database Version 2.0 Step by Step Guide 8 December 2004, (Doc. Ref: EMEA/191986/2004) http://eudravigilance.emea.eu.int/human/docs/EV%20Access%20Simple%20Database%202.0%20-%20Step%20by%20Step%20Guide.pdF

[55] Product Information Management http://pim.emea.europa.eu/

[56] European Pharmacopeia (Ph. Eur.): http://online.pheur.org/entry.htm

[57] MINISTRY OF HEALTH AND WELFARE. PSB/SD Notification No. 37 (29th March 1999) (All documents are available only in Japanese, Please note documents' names are only tentative translation) Electronic package insert information accompanied by the Utility system of Information provision on drug safety https://skw.info.pmda.go.jp/notice/nt990329.pdf

[58] Appendix: Electronic format of package insert information on ethical drugs (SGML/DTD Ver.2.0) Appendix1 DTD Ver.2.0 https://skw.info.pmda.go.jp/notice/nt990329dtd.zip

[59] Appendix2 Overview of SGML creation https://skw.info.pmda.go.jp/notice/nt990329ol.pdf

[60] Appendix3 DCL https://skw.info.pmda.go.jp/notice/nt990329dcl.zip

[61] Appendix4 Template of SGML https://skw.info.pmda.go.jp/notice/nt990329smp.zip

[62] Appendix5 Data model (Entities and relationships) https://skw.info.pmda.go.jp/notice/nt990329dm.pdf

[63] Appendix6 List of fields of package insert information https://skw.info.pmda.go.jp/notice/nt990329lst.pdf

[64] THE JAPANESE PHARMACOPOEIA SIXTEENTH EDITION. http://jpdb.nihs.go.jp/jp16/

[65] United States Pharmacopoeia (USP): http://www.usp.org/usp-nf

[66] INTEGRATED TAXONOMIC INFORMATION SYSTEM. (ITIS) http://www.itis.gov/

[67] Guidance for Industry Providing Regulatory Submissions in Electronic Format — Content of Labelling http://www.fda.gov/cder/regulatory/guidance

[68] Release Notes for SPL Schema PORR_MT050020 (3.20.05) FDA SPL Schema for Implementation http://www.fda.gov/downloads/ForIndustry/DataStandards/StructuredProductLabeling/UCM162024.pdf

[69] CaCore 2.0 Technical Guide, National Cancer Institute, Center for Bioinformatics, U.S. Department of Health and Human Services ftp://ftp1.nci.nih.gov/pub/cacore/caCORE2.0_Tech_Guide.pdf

[70] A guide to RXNorm, United States National Library of Medicine, National Institute of Health http://www.nlm.nih.gov/research/umls/rxnorm/docs/2012/rxnorm_doco_full_2012-3.html

[71] The Unified Code for Units of Measure http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/ucm168397.htm

[72] Substance Registration System (SRS) SRS Substance Definition Manual Version 5c . http://www.fda.gov/downloads/ForIndustry/DataStandards/SubstanceRegistrationSystem-UniqueIngredientIdentifierUNII/ucm127743.pdf

[73] The U.S. Consolidated Healthcare Informatics initiative https://www.federalregister.gov/articles/2005/12/23/05-24289/consolidated-health-informatics-chi-initiative-health-care-and-vocabulary-standards-for-use-in

[74] CDISC Clinical Research Glossary, http://www.cdisc.org/cdisc-glossary

ICS 35.240.80

Price based on 41 pages