



BSI Standards Publication

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

National foreword

This Published Document is the UK implementation of CEN ISO/TS 19844:2015.

The UK participation in its preparation was entrusted to Technical Committee IST/35, Health informatics.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2016.

Published by BSI Standards Limited 2016

ISBN 978 0 580 87004 0

ICS 35.240.80

Compliance with a British Standard cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 31 January 2016.

Amendments/corrigenda issued since publication

Date	Text affected
------	---------------

TECHNICAL SPECIFICATION
SPÉCIFICATION TECHNIQUE
TECHNISCHE SPEZIFIKATION

CEN ISO/TS 19844

December 2015

ICS 35.240.80

English Version

**Health informatics - Identification of medicinal products -
Implementation guidelines for data elements and
structures for the unique identification and exchange of
regulated information on substances (ISO/TS 19844:2015)**

Informatique de santé - Identification des médicaments
- Lignes directrices pour la mise en oeuvre des
éléments de données et structures pour l'identification
unique et l'échange d'informations réglementées sur
les substances (ISO/TS 19844:2015)

Medizinische Informatik - Identifikation von
Arzneimitteln - Anwendungsleitfaden für die Struktur
und kontrollierten Vokabularen zur Identifikation und
Beschreibung von Substanzen und Inhaltsstoffen
(ISO/TS 19844:2015)

This Technical Specification (CEN/TS) was approved by CEN on 4 April 2016 for provisional application.

The period of validity of this CEN/TS is limited initially to three years. After two years the members of CEN will be requested to submit their comments, particularly on the question whether the CEN/TS can be converted into a European Standard.

CEN members are required to announce the existence of this CEN/TS in the same way as for an EN and to make the CEN/TS available promptly at national level in an appropriate form. It is permissible to keep conflicting national standards in force (in parallel to the CEN/TS) until the final decision about the possible conversion of the CEN/TS into an EN is reached.

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



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

CEN-CENELEC Management Centre: Avenue Marnix 17, B-1000 Brussels

© 2015 CEN All rights of exploitation in any form and by any means reserved
worldwide for CEN national Members.

Ref. No. CEN ISO/TS 19844:2015 E

Contents	Page
European foreword.....	3

European foreword

This document (CEN ISO/TS 19844:2015) has been prepared by Technical Committee ISO/TC 215 "Health informatics" in collaboration with Technical Committee CEN/TC 251 "Health informatics" the secretariat of which is held by NEN.

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

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

Endorsement notice

The text of ISO/TS 19844:2015 has been approved by CEN as CEN ISO/TS 19844:2015 without any modification.

Contents

Page

Foreword.....	viii
Introduction.....	ix
1 Scope.....	1
2 Normative references.....	2
3 General background and history.....	2
4 Substance (Mandatory).....	3
4.1 Introduction.....	3
4.2 Defining Substances.....	5
4.3 Substance Types (Mandatory).....	7
4.4 Substance ID (Mandatory).....	10
4.5 Substance Names (Mandatory).....	11
4.5.1 Substance Name.....	12
4.5.2 Substance Name Type.....	12
4.5.3 Language.....	13
4.5.4 Official Name (repeat as necessary).....	13
4.6 Reference Sources (Mandatory).....	15
4.6.1 Public Domain.....	16
4.6.2 Reference Source Type.....	16
4.6.3 Reference Source Class.....	16
4.6.4 Reference Source ID.....	17
4.6.5 Reference Source Citation.....	17
4.6.6 Reference Source Document (new class to be included in the second edition of ISO 11238).....	17
4.6.7 Reference Source Document Type (new class to be included in the second edition of ISO 11238).....	18
4.6.8 Reference Source Document Classification (new class to be included in the second edition of ISO 11238).....	18
4.6.9 Reference Source URL (new class to be included in the second edition of ISO 11238).....	18
4.7 Substance Code (Conditional).....	18
4.7.1 Code.....	19
4.7.2 Code System.....	19
4.7.3 Code System ID.....	20
4.7.4 Code System Status.....	20
4.7.5 Code System Status Change Date.....	20
4.7.6 Comment.....	20
4.7.7 Reference Source.....	21
4.7.8 Substance Classification (repeat as necessary).....	21
4.7.9 Target.....	23
4.7.10 Gene.....	25
4.7.11 Gene Elements.....	26
4.7.12 Substance Relationship.....	27
4.8 Structure (repeat as necessary) (Conditional).....	29
4.8.1 Structural Representation Type.....	34
4.8.2 Structural Representation.....	35
4.8.3 Structural Representation Attachment.....	35
4.8.4 Stereochemistry.....	35
4.8.5 Optical Activity.....	36
4.8.6 Molecular Formula.....	36
4.8.7 Molecular Formula by Moieties (new class to be included in the second edition of ISO 11238).....	37
4.8.8 Isotope (repeat as necessary).....	37

4.9	Amount (Conditional)	38
4.9.1	Average	38
4.9.2	Low Limit.....	38
4.9.3	High Limit	39
4.9.4	Unit.....	39
4.9.5	Non-numeric Value.....	39
4.10	Source Material (Conditional).....	39
4.10.1	Source Material Class.....	40
4.10.2	Source Material Type.....	41
4.10.3	Source Material state.....	41
4.10.4	Organism ID	41
4.10.5	Organism Name	41
4.10.6	Development Stage.....	42
4.10.7	Part Description (repeat as necessary)	42
4.10.8	Fraction (repeat as necessary)	42
4.10.9	Organism.....	43
4.11	Modification (repeat as necessary) (Conditional).....	49
4.11.1	Modification Type	50
4.11.2	Residue Modified	51
4.11.3	Residue Site.....	51
4.11.4	Structural Modification.....	51
4.12	Property (Conditional)	55
4.12.1	Property Type	56
4.12.2	Property Name	56
4.12.3	Property Parameters (new class to be included in the second edition of ISO 11238).....	56
4.12.4	Substance Name.....	57
4.12.5	Substance ID.....	57
4.12.6	Amount type.....	57
4.13	Version (repeat as necessary) (Mandatory)	57
4.13.1	Version Number	57
4.13.2	Effective date	58
4.13.3	Change Made.....	58
5	Substance definitions	58
5.1	Chemical Substance.....	58
5.1.1	Structure	59
5.1.2	Stoichiometric.....	59
5.1.3	Stoichiometric Chemicals.....	59
5.1.4	Comment	62
5.1.5	Non- Stoichiometric Chemicals	62
5.2	Proteins/ Peptides.....	64
5.2.1	Microheterogeneity.....	65
5.2.2	Sequence Type	66
5.2.3	Number of subunits.....	66
5.2.4	Disulfide Linkage	66
5.2.5	Comment	67
5.2.6	Protein Subunit (repeat as necessary).....	67
5.2.7	Molecular Weight (repeat as necessary).....	69
5.2.8	Glycosylation	69
5.2.9	Structure	71
5.2.10	Modification	71
5.2.11	Property	71
5.2.12	Molecular Weight.....	71
5.3	Nucleic Acids.....	72
5.3.1	Structure	73
5.3.2	Sequence Type	73
5.3.3	Number of Subunits.....	73
5.3.4	Area of hybridisation.....	74

5.3.5	Comment	74
5.3.6	Nucleic Acid Subunit (repeat as necessary)	74
5.3.7	Modification	77
5.3.8	Property.....	77
5.3.9	Molecular Weight.....	77
5.4	Polymers –To be addressed in more detail in the next edition of this Technical Specification	78
5.4.1	Substance Name.....	79
5.4.2	Structure.....	79
5.4.3	Polymer Class	79
5.4.4	Polymer Geometry.....	80
5.4.5	Copolymer Sequence type.....	80
5.4.6	Comment	80
5.4.7	Monomer Description (repeat as necessary)	80
5.4.8	Structural Repeat (repeat as necessary).....	81
5.4.9	Molecular Weight (repeat as necessary).....	83
5.4.10	Property (repeat as necessary)	83
5.4.11	Reference Source (repeat as necessary)	83
5.5	Structurally-Diverse Substances	83
5.5.1	herbals and Substances Used in the Preparation of Plant-Based Allergenic Extracts	84
5.5.2	Vaccines — Annex addressing this will be included in the next edition of this Technical Specification.....	93
5.5.3	Purified Blood Products and Polyclonal Antibodies — Annex addressing this will be included in the next edition of this Technical Specification.....	93
5.5.4	Cells and Tissues — Annex addressing this will be included in the next edition of this Technical Specification	93
5.5.5	Minerals — Annex addressing this will be included in the next edition of this Technical Specification.....	93
5.6	Mixture Substance (repeat as necessary)	94
5.6.1	Mixture Type.....	94
5.6.2	Mixture Constituent (repeat as necessary)	94
6	specified substance (Optional)	95
6.1	specified substance Group 1 (repeat as necessary)	96
6.1.1	specified substance Group 1 ID.....	97
6.1.2	specified substance Group1 Name.....	97
6.1.3	Substance Name (repeat as necessary).....	97
6.1.4	Substance Code	97
6.1.5	Version (repeat as necessary)	97
6.1.6	Reference Sources.....	97
6.1.7	Property (repeat as necessary)	97
6.1.8	Fraction (new class to be included in the second edition of ISO 11238).....	98
6.1.9	Modification	98
6.1.10	Reference Information (repeat as necessary).....	98
6.1.11	Constituent (repeat as necessary)	98
6.1.12	Physical Form (repeat as necessary)	99
6.2	specified substance Group 1 intended for herbal Substance and herbal Preparation	100
6.2.1	specified substance Group 1 ID.....	101
6.2.2	specified substance Group1 Name.....	101
6.2.3	Reference Sources.....	101
6.2.4	Fraction (new class to be included in the second edition of ISO 11238).....	101
6.2.5	Modification (new classes to be included in the second edition of ISO 11238).....	102
6.2.6	Constituent (repeat as necessary)	102
6.2.7	Physical Form (repeat as necessary)	103
6.3	specified substance Group 2 (repeat as necessary)	104
6.3.1	specified substance Group2 ID.....	107
6.3.2	specified substance Group2 Name.....	107
6.3.3	Parent Substance ID	107
6.3.4	Manufacturing.....	107

6.4	specified substance Group 2 for herbal preparations.....	111
6.4.1	specified substance Group2 ID	111
6.4.2	specified substance Group2 Name.....	111
6.4.3	Parent Substance ID	111
6.4.4	Manufacturing.....	112
6.4.5	Version (repeat as necessary)	113
6.5	specified substance Group 3 (repeat as necessary)	113
6.5.1	specified substance Group3 ID	114
6.5.2	specified substance Group3 Name.....	114
6.5.3	Parent Substance ID	114
6.5.4	Grade	115
6.5.5	Reference Source (repeat as necessary)	115
6.5.6	Version (repeat as necessary)	115
6.5.7	Reference Source (repeat as necessary)	116
6.5.8	Version (repeat as necessary)	116
Annex A	(normative) Choosing a Substance ID	117
A.1	Requesting a Substance ID and providing information.....	117
Annex B	(normative) Chemical Substance	119
B.0	Introduction	119
B.0.1	Proposal for the update of the ISO 11238 Substance standard	120
B.0.2	Outline of Annex B	122
B.1	Scope.....	123
B.2	Terms and definitions	123
B.3	Chemical Substance subtypes and Mixture Substance.....	132
B.3.1	Substance type, Chemical substance.....	132
B.3.2	Solid state forms of the Substance	132
B.3.3	Need to substantiate the chemical structure, molecular formula and molecular weight.....	135
B.3.4	Polymorphism.....	136
B.3.5	Non-Stoichiometric chemical substances	137
B.3.6	Mixture substance	138
B.3.7	Multi substance material	139
B.4	Discussion of the key elements of a chemical substance	142
B.4.1	Identity of material	142
B.4.2	Nomenclature	142
B.4.3	Molecular formula	142
B.4.4	Molecular weight.....	143
B.4.5	Substance Structure	143
B.4.6	Geometric Isomerism	146
B.4.7	Stereo-descriptors in systematic nomenclature: Substance with one centre of Asymmetry.....	147
B.4.8	Substance with two centres of Asymmetry, Epimers, Diastereomers.....	148
B.4.9	Anomers	148
B.4.10	Substance with more than two centres of Asymmetry (Mixture of stereoisomers)	151
B.4.11	Conclusion for the Key elements.....	152
B.4.12	Decision tree for a new Substance ID	152
B.5	Discussing other elements of importance regarding the characteristics of a substance.....	153
B.5.1	Introduction.....	153
B.5.2	Naming Vegetable Oils	153
B.5.3	Castor Oil and related products.....	169
B.5.4	Properties to be captured, related to liquids (Gas), Nitrous oxide	172
B.6	Examples	177
B.6.1	Example 1: Amlodipine besilate.....	178
B.6.2	Example 2: Ponatinib hydrochloride.....	185
B.6.3	Example 3: Benzathine Benzylpenicillin tetrahydrate, sterilised	201
B.6.4	specified substance Group 2 information level	206
B.6.5	specified substance Group 3 information level	207
B.7	Radiopharmaceutical substance	207
B.7.1	Introduction.....	207

B.7.2	Example: Florbetapir ¹⁸ F	208
B.7.3	Identity of material, combining the elements for Florbetapir ¹⁸ F, Substance and specified substance information level.	209
B.7.4	specified substance Group 2 information level	214
Annex C (normative) Protein Substance		215
C.1	Scope	215
C.2	Introduction	215
C.3	Peptide Substances	216
C.3.1	Example IIIa: Protein Substance: Vasopressin	216
C.3.2	Example IIIb: Desmopressin	219
C.3.3	Example IIIc: Desmopressin Acetate	223
C.3.4	Example IIId: Calcitonin Salmon	227
C.3.5	Example IIIe: Human Insulin	233
C.4	Element Group Protein, specified substance Group 1	235
C.4.1	Example IVa: Insulin Human Zinc Suspension (Amorphous), taken from EP/USP Monographs for Zinc Suspensions	235
C.4.2	Example IVb: Insulin Human Zinc Suspension (Crystalline), taken from EP/USP Monographs for Zinc Suspensions	236
C.4.3	Example IVc: Insulin Human Zinc Suspension taken from EP/USP Monographs for Zinc Suspensions, including discussion for specified substance Group 3 information	237
C.5	Element Group Protein, specified substance Group 2	240
C.5.1	Example Va: Synthetic Calcitonin Salmon-Manufacturer Company Acme	240
C.5.2	specified substance Group 2 information level	240
C.5.3	Example Vb: Recombinant Calcitonin Salmon-- Company Acme	241
C.6	Element Group Protein, specified substance Group 3	241
C.6.1	Example: Calcitonin Salmon (Synthetic) specified substance Group 3	242
C.6.2	Example: Calcitonin Salmon (Recombinant) specified substance Group 3	242
C.7	Example Protein Substance Example of a Monoclonal Antibody conjugated Toxin	243
C.8	Addendum: Microheterogeneity	248
Annex D (normative) Nucleic Acid Substance		249
D.1	Scope	249
D.2	Introduction	249
D.3	Examples	250
D.3.1	Example: 5-Methylcytosine	250
D.3.2	Example: A 5'-phosphate ribonucleotide	251
D.4	Substances	252
D.4.1	Example: Mipomersen Sodium information	252
D.4.2	Example: Oligonucleotide Elements	252
D.4.3	Example: Anti-sense RNA	257

Foreword

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

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

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

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

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: [Foreword - Supplementary information](#)

The committee responsible for this document is ISO/TC 215, *Health informatics*.

Introduction

This Technical Specification is a guide for implementing ISO 11238, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on substances*. This Technical Specification was developed in response to a worldwide demand for guidance on the implementation of internationally harmonised specifications for medicinal products. It is one of a group of four implementation guides for a total of five ISO standards which together provide the basis for the unique identification of medicinal products. The other standards in this group are:

- 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 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*

The 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.

The business objective of this implementation guide is to provide a means for exchanging regulatory substance information. To meet the primary objectives of the regulation of medicines and pharmacovigilance, it is necessary to exchange medicinal product information in a robust and reliable manner.

For the purposes of this Technical Specification, all conditions (e.g. mandatory, conditional, optional) correspond to the necessary requirements to uniquely and unambiguously identify a substance. Implementation of the ISO IDMP standards may dictate that mandatory elements for identification be tagged as conditional or optional, based on regional requirements. If a section is identified as 'optional' but is implemented in a specific region, conformance described within that section is applicable. The scope of this Technical Specification is to identify the scientifically necessary elements for the unique identification of substances/specified substances.

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

1 Scope

This Technical Specification is used in the implementation of ISO 11238. This Technical Specification defines substances based on their scientific identity (i.e. what they are) rather than on their use or method of production.

ISO 11238 provides the conceptual framework for defining substances and specified substances and for assigning unique identifiers in the context of the ISO IDMP standards. ISO 11238 describes general concepts for defining and distinguishing substances and a high level model for the structuring of information for substances. This Technical Specification provides detailed explanations of each type or grouping of substance information, an element-by-element description for implementation of ISO 11238, and examples for a variety of substances and specified substances.

This first edition of the Technical Specification will only address substances, and Groups 1 to 3 of the specified substances as defined in ISO 11238 and Annexes A, B, C, and D. It is anticipated that specified substances Group 4, as defined in ISO 11238, will be addressed in a subsequent edition of this Technical Specification. Some information that would typically fall under specified substances Group 4 may be covered in the Annexes of this Technical Specification. This information, although not defining of either a substance or a specified substance Group 1, may be essential to distinguishing substances.

This Technical Specification addresses the following:

- Data elements necessary for defining substances and specified substances Groups 1 to 3;
- The logical use of data elements as defined in ISO 11238;
- Substances and specified substances Groups 1 to 3 business rules for
 - determining necessary data elements,
 - distinguishing and defining materials according to ISO 11238,
 - triggering the assignment of identifiers.

This Technical Specification does not address the following:

- Business processes for data management;
- Implementation of a specific data information system (e.g. a relational database schema);
- Normative messaging standards for substances;
- The maintenance of controlled vocabularies;
- The specific global identifier system that should be used;
- Nomenclature standards for substances.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

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

3 General background and history

Due to the lack of a common and harmonised approach to define substances, regulators and pharmaceutical industry are faced with the inability to:

- 1) effectively exchange medicinal substance information in a structured and efficient way;
- 2) ensure data consistency and evaluate/compare information across regions, which especially impairs pharmacovigilance and compliance activities;
- 3) develop consistent terminology for use throughout the healthcare community.

The objectives of the IDMP standards are to address the issues outlined above by developing harmonised standards that build on the regulatory and technical processes already established and to support the population and maintenance of existing systems/applications with fully reliable regulatory medicinal product information.

Harmonised standards will stimulate vendors to develop “off-the-shelf” tools (that are interoperable due to the standard itself). Harmonised standards will also help to maximise forward compatibility of data and minimise the complexities of backward compatibility.

This implementation guide is intended to assist reporters (including pharmaceutical companies, regulatory authorities and non-commercial sponsors) in constructing messages or transmitting information that allows substances to be defined unambiguously and assigned unique IDs. It also provides guidance to help choose the correct Substance ID from a public data source that provides unique substance and specified substance identifiers. It is anticipated that an extensive list of substance identifiers as well as the definitional elements upon which the ID was based will be provided. This Technical Specification is not intended to be a guide for a maintenance organisation. The maintenance organisation may also create alternative methods to submit information consistent with the ISO model.

Table 1 is an example table for class and elements description.

Table 1 — Example table for class and element description

User Guidance	
Example(s)	
Conformance	
Data Type	
Values Allowed	
Business Rule(s)	

In contrast to other parts of the guide, conformance will refer to whether an element is required for a given substance type or a specified substance group. Conformance is not meant to be applied globally to all types of messages.

Definition: Conformance will be expressed based on the following terminology: Mandatory, Conditional and Optional.

Mandatory: refers to data elements that are required and shall therefore be implemented.

Conditional: refers to data elements that are subject to business rules and may become required by:

- data rules;
- process rules;
- regional rules.

Optional: refers to data elements that are informative but not definitional.

The description on whether a data element is conditional by data, by process or by regional rule is out of scope of this Technical Specification and will be defined within regional implementation guides.

The information provided in the table refers to the global guidance. When there is no information in the conformance table row (e.g. information on business rule is not provided), please refer to the regional implementation guide.

4 Substance (Mandatory)

4.1 General

All medicinal products consist of substances; these substances can be active ingredients, excipients, or packaging materials. There are two fundamental levels of information described in ISO 11238, a “substance level” and a “specified substance” level. Both levels are included in the more generic concept of an ingredient. At the substance level, substances are defined based on inherent attributes rather than use or method of manufacture. At the specified substance level, four separate groups of elements provide additional information.

In order to define or distinguish material either at a substance or specified substance level, a number of attributes should be taken into consideration:

- For chemicals, the molecular structure is captured at the substance level;
- For proteins, the amino acid sequence, sites and type of glycosylation, and the presence and position of disulfide bonds will be captured at the substance level;
- For nucleic acids, the sequence, type of sugar and linkage will be captured at the substance level;
- For other polymers, the monomers used to synthesize the polymer, the structural repeating units, the molecular weight and/or a property related to molecular weight (e.g. viscosity), the source of naturally derived polymers and any modifications that irreversibly alter the molecular structure will be captured at the substance level;
- For structurally diverse material, taxonomic, anatomical and fractionation information, properties related to the underlying molecular structure of the material, and modifications that alter the underlying molecular structure will be captured at the substance level;
- Mixture substance consists of a simple combination of single substances that are either isolated together or are the result of the same synthetic process. The biological source of the mixture is also captured where relevant at the substance level. Proportions are not captured at the substance level. It should be noted that a mixture

substance description should only include the substances that are generally or consistently present in the material. This excludes impurities and degradants.

Other attributes will be specific to the specified substance levels:

- Constituent substances in a multi-substance material;
- Proportions of constituent substances in a multi-substance material;
- Physical state;
- Grade or purity of material;
- Manufacturing information;
- Analytical data.

There are four groups of elements that are used to further define and specify substances. Specified substances are always composed of at least one substance.

Specified substance Group 1 is typically used to define:

- Multi-substance materials consisting of multiple substances, which are not defined as mixture substances;
- Additional information regarding herbal and allergenic extracts;
- Physical state, including polymorphic forms;
- Detailed glycosylation information.

Specified substance Group 2 is typically used to define:

- Manufacturer and the overall manufacturing process and critical process version number;
- In addition there is the possibility to make use of the reference source document class to store Release specifications of the intended manufactured substance.

Specified substance Group 3 is typically used to define:

- Grade or level of purity (Pharmacopoeial Specifications) and In house specification used to cover a set of specifications of all approved manufacturers for the substance.

Specified substance Group 4 is typically used to define:

- Detailed Analytical Data;
- Detailed Manufacturing Information.

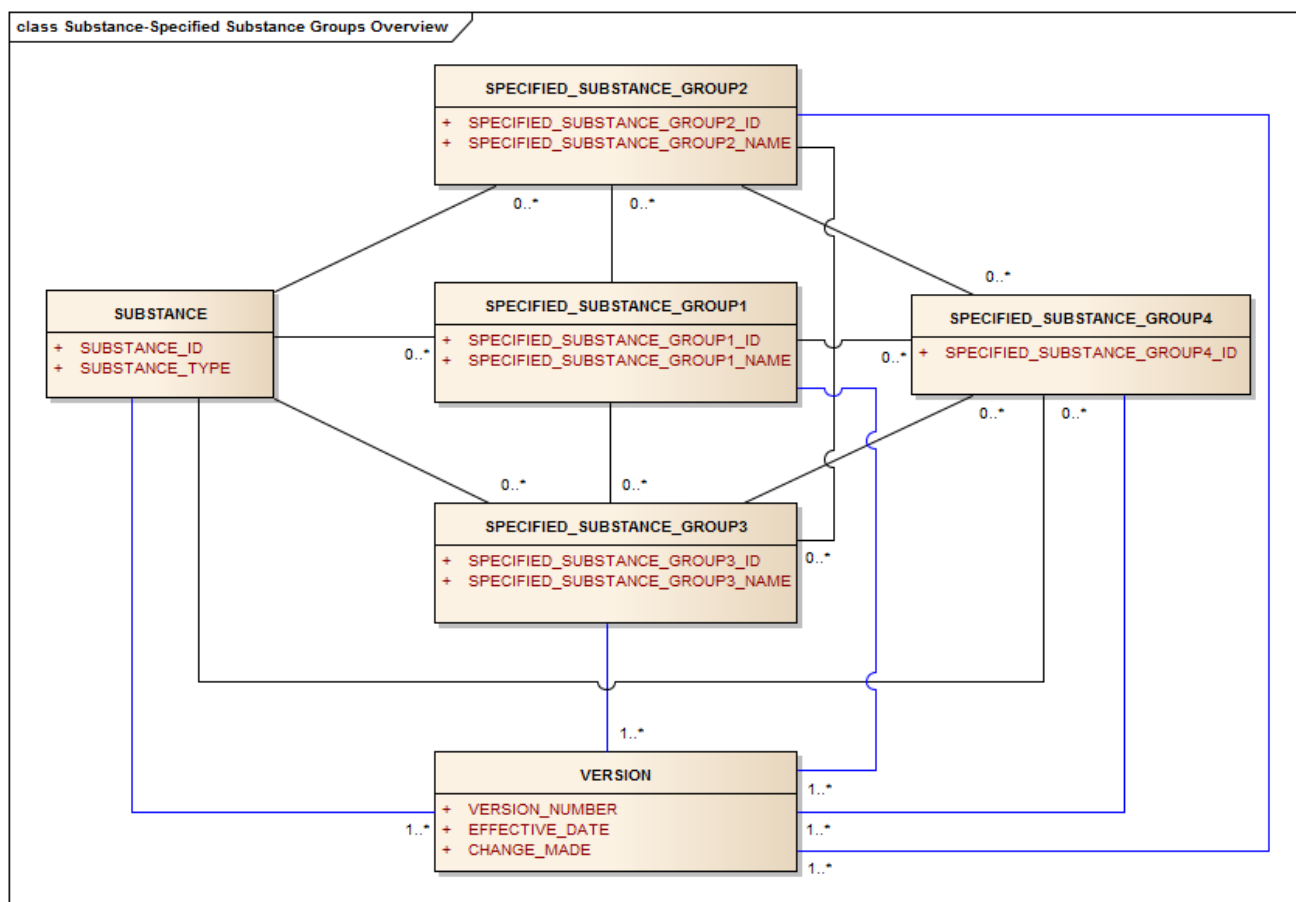


Figure 1 — High-level Substance-Specified substance information model

4.2 Defining substances

A substance is any matter that has a discrete existence, irrespective of origin, which may be biological or chemical. Substances can be single well-defined chemical entities containing a definite molecular structure, synthetic (e.g. isomeric mixtures) or naturally-occurring (e.g. conjugated oestrogens) mixtures of chemicals, or materials derived from plants, animals, microorganisms or inorganic matrices that are not definable by a single or limited number of molecular structures. Substances may be active moieties, salts, solvates, stoichiometric complexes or mixtures of compounds that are isolated or synthesised/ obtained or produced in a process together. Materials that are combined from multiple sources to form a product are not considered substances.

A substance is generally defined by what it is, and not by how it is made or used. Substance definitions are typically based on the immutable properties of a given material. These properties include the molecular structure, or structures of a given material, taxonomic, anatomical or fractionation information for material that cannot be represented by molecular structures. Purity, physical form, and method of production are typically not considered when defining substances.

In addition to defining information there is also information that is essential to validate the defining information and this information should be submitted if available. Validation may be performed based on relevant provided documents (e.g. regulatory dossier) or information available in recognised source (e.g. pharmacopoeias).

The primary goal of the ISO IDMP standard - *Data Elements and Structures for the Unique Identification and Exchange of Regulated Information on Substances* is to define unambiguously all substances present in regulated products. Once a substance has been defined, a unique identifier that is permanently associated with that substance will be assigned. This Technical Specification describes the necessary information for this registration process. Reference information, names, codes and IDs that can be associated with a substance are also described. The document is not comprehensive in regards to reference and definitional information. Other substance

attributes may be associated for both reference and definitional information as new scientific standards or information becomes available.

Substances that are mixed together to form a product, even if there are physical interactions between the substances will be described as a specified substance or a product.

EXAMPLE 1 FD&C Blue No. 1 Aluminium Lake is not defined as a mixture substance but rather as two separate substances FD&C Blue No. 1 and aluminium oxide. Also, simethicone is not a substance because it contains two substances, namely dimethicone and silicon dioxide. These materials would be described as Group 1 specified substances.

Conversely, if new covalent bonds and/or simple salts or solvates are formed, separate Substance IDs for each molecular entity are assigned.

EXAMPLE 2 Fluticasone propionate would have a separate ID from fluticasone; atorvastatin calcium trihydrate would have a separate ID from atorvastatin calcium anhydrous.

Differences at the specified substance level should trigger a new specified substance ID.

EXAMPLE 3 Insulin Human Crystalline and Insulin Human Amorphous will have different Group 1 specified substances IDs and map to the substance Human Insulin.

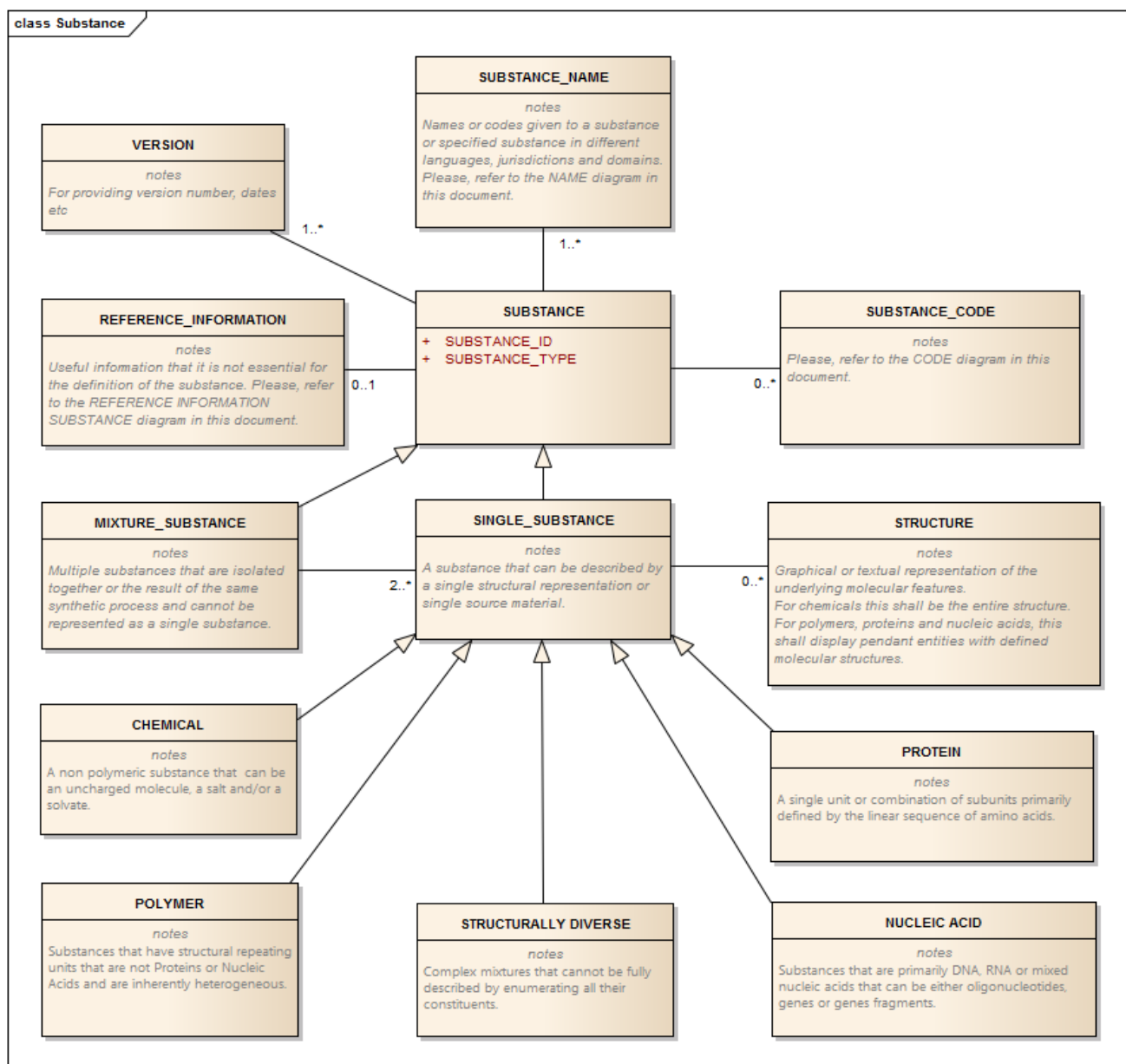


Figure 2 — High-level information model for substances

4.3 Substance types (Mandatory)

At the basic level, all material present in medicinal products will either be described as a substance and/or a specified substance Group 1. Additional information, such as manufacturer, grade and manufacturing information on a given material will be associated with specified substance Groups 2 to 4.

Substances are either one of five types of single substances or a mixture. Mixtures consist of a combination of related single substances where the source material and essential properties are also captured but proportions are not. For the purposes of definition, all single substances are of one of the following types:

- (simple) chemical;
- protein;
- nucleic acid;
- polymer;
- structurally-diverse.

Each of these different types of substances has distinct elements essential for definition. Substances can also be a **mixture of single substances**, typically related and of the same type. Some materials may appear to have elements from multiple types of substances e.g. PEGylated proteins, drugs covalently linked to polymeric matrices or proteins, etc., to fully define the substance. Each of these substances will be described as a single type, although it is possible for the same to be defined as two types one of which would be an alternative definition.

EXAMPLE 1: Proteins modified with a polymer or drug will be described as a protein with the polymer or drug captured as a modification. Drugs covalently linked to a polymer will be described as a polymer with the drug either captured as an end group or as structural modification. Detailed examples will be provided in the Annex to address these complex materials. Although every substance definition will be defined with a single type it is possible that an alternative definition of a substance will be of a different type.

EXAMPLE 2: Desmopressin can either be defined as a modified peptide or as a chemical substance.

Mixtures consist of a combination of related single substances where the source material and essential properties are also captured but proportions are not. At the specified substance Group 1 level proportions can be captured.

Additional information, such as manufacturer, grade, and manufacturing information on a given material will be associated with specified substances Groups 2-4. The first step in defining a material is to determine whether the material can be defined as a single substance, a mixture of single substances or a specified substance Group 1. A mixture substance is a substance that cannot be defined using a single set of substance elements or structural representations. Racemic mixtures, most polymers and structurally diverse substances, although inherently mixtures, can be defined using a single set of elements and therefore are not described as mixtures. Material that contains multiple substances that are not related and not synthesised or isolated together will be described as specified substance Group 1. Specified substance Group 1 would also include crystalline polymorphs, substances in varied physical states along with the capture of additional properties regarding sterility or the presence of endotoxin.

Examples of specified substance Group 1 include simethicone, adjuvanted vaccine antigens, isophane insulins and additional herbal and allergenic extract information not captured at the Substance level. Sterile water, water for injection, ice, and steam would be different specified substances Group 1 that would map to the substance water.

Mixtures are substances like Gentamicin sulfate, USP or Gentamicin sulfate, EP which consists of mixtures of distinct but related individual aminoglycosides. In USP Gentamicin sulfate is described as a mixture of Gentamicin C1A sulfate, Gentamicin C1 sulfate, and Gentamicin C2 sulfate. In EP it is described as a mixture of five components Gentamicin C1, C1a, C2, C2a, and C2b respectively which are typically isolated together from a bacterial culture. At the substance level all related substances typically present at a level greater than 1% would be listed as components of the mixture regardless of whether the substance is listed in a given pharmacopoeia. Mixture Substances are defined in the ISO 11238 Substance standard as a simple combination of single substances that are either isolated together or are the result of the same synthetic process.

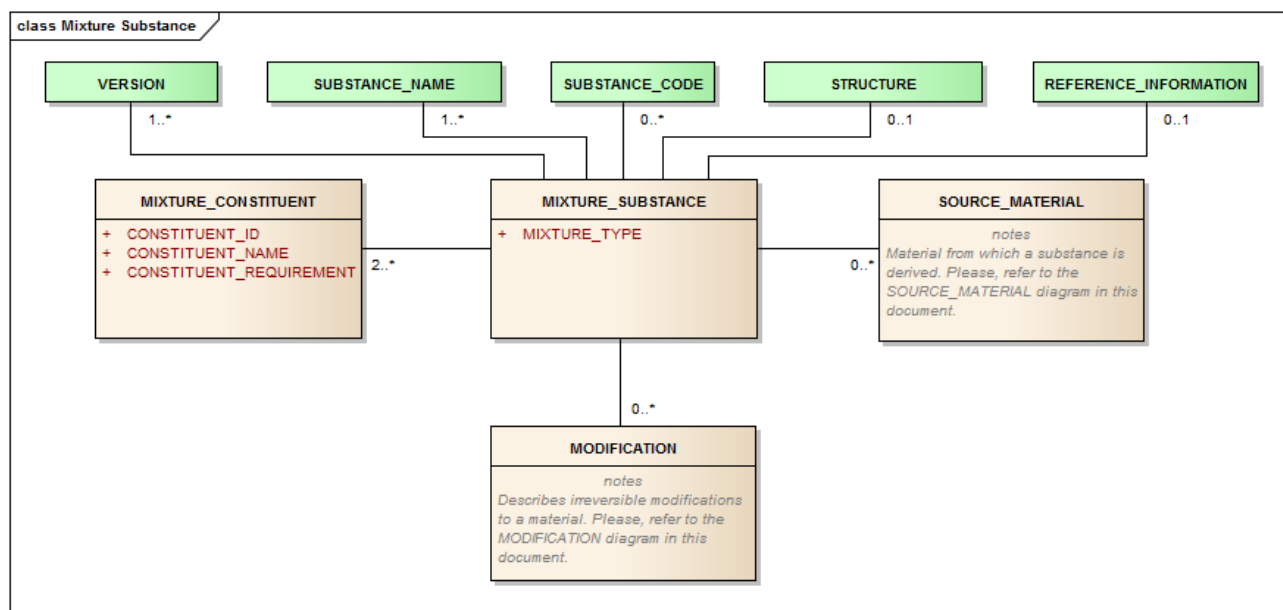
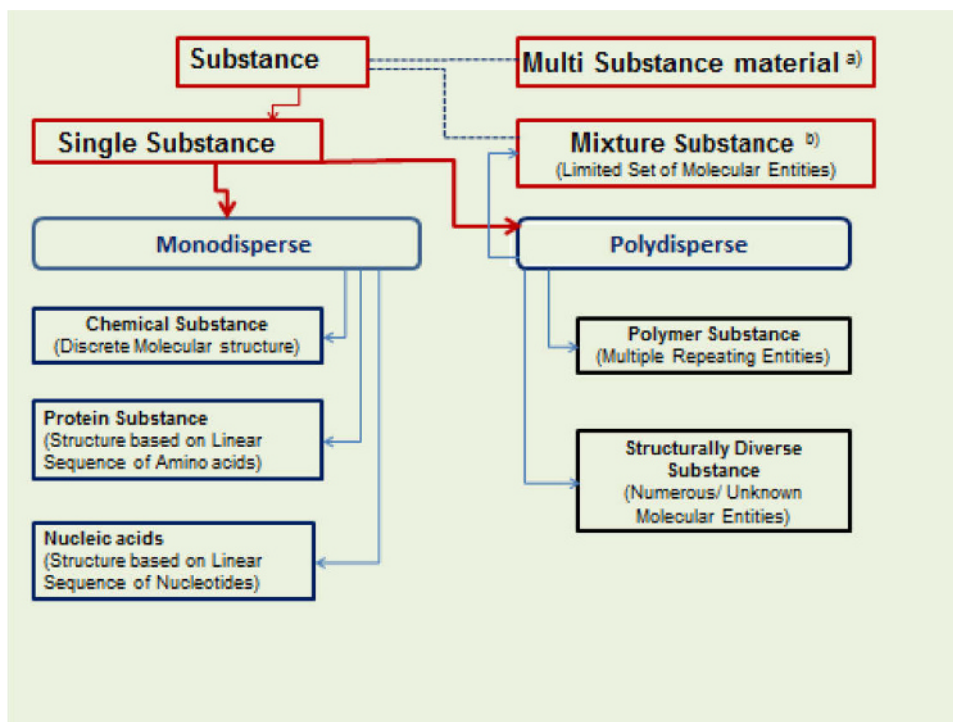


Figure 3 — High-level information model for mixture substance

For mixture substances in which multiple related substances are isolated together from a single source, the source will be a defining element. Starches are typically a mixture substance of the two distinct polymers amylose and amylopectin. Corn, wheat, potato and tapioca starches are all distinct substances distinguished by the source.

Most other polymers will be described as single substances with the definition based on the structural repeating unit and a representation of the average molecular weight of the polymer as well as the biological source for biopolymers or polymers derived from biological matrices.

Figure 4 represents a decision tree or process by which the type of substance is determined. Monodisperse substances are substances that can be described as a single molecular entity. Polydisperse substances are substances that typically have multiple molecular entities that are too numerous or too diverse to be captured as a mixture (e.g. cells or tissues) or where the production of the substance inherently results in polydispersity (e.g. polymers).



- a) Multi-substance material: Multi substances and/or specified substances of diverse origin used as a component in the formulation in a medicinal product.
- b) Mixture substance: Type of polydisperse substance that is a combination of single substances isolated together or synthesised/obtained or produced in the same process.

Figure 4 — Decision tree or process by which the type of substance is determined

Table 2 — Substance types

User Guidance	Information about the substance type as described in and in accordance with ISO 11238 definitions
Example(s)	“Chemical” “Protein” “Nucleic acid”, “Polymer”, “Structurally Diverse”
Conformance	MANDATORY
Data Type	CD
Values Allowed	Chemical, Protein, Nucleic Acid, Polymer, Structurally Diverse, Mixture, specified substance Group 1, specified substance Group 2, specified substance Group 3, specified substance Group 4

4.4 Substance ID (Mandatory)

Every substance will be identified by an ID. An ID will be generated following the initial submission of substance data. The ID should be unique to each substance, non-semantic, non-chronological, and of fixed length with an integrity check. Once an ID is assigned to a given physical or conceptual material the ID will not change. Changes or additions to defining information result in a new version of the substance definition but would not result in a new Substance ID. Although an effort should always be made to obtain complete information prior to assignment there will be instances when additional definitional information on a given substance is transmitted after assignment. When incorrect or incomplete information may be transmitted and used to define a substance the information will be changed or added but the ID will not change. In rare cases when the same substance is found to have two IDs one of the IDs will remain a primary ID and the use of the other ID will be deprecated and captured in the Substance Code class For some materials there may be multiple IDs that a material could map to. A submitter should always submit the ID that is the most descriptive of the material.

EXAMPLE Glucose, when a solid crystalline material can exist with an alpha or beta pyranose structure in an anhydrous state, or as a monohydrate with an alpha pyranose structure. If a solid crystalline form of glucose exists

in a medicinal product the ID specific to the structure should be used. Glucose in a liquid or an amorphous solid state would use the Substance ID that does not specify the specific form.

Both substances and specified substances will be identified with an ID. IDs will be released to the public if the substance is in a licensed medicinal product. The IDs for substances in the investigational stage will only be released if an official name exists or a company code is associated with defining information and is found together in at least one single reference that is from a reputable source in the public domain (i.e. scientific journal, presentations or posters at scientific conferences, company publication, patent or published patent application, public databases such as STN from CAS). Substance IDs will always be issued to the requester if sufficient information is provided to unambiguously define the substance.

Table 3 — Substance ID

User Guidance	ID to be used in all electronic submissions to identify a substance. Generated when sufficient information is available to unambiguously define a substance. ID will be permanently associated with a given substance and each substance at the substance level shall have one and only one ID. NOTE If a unique "Substance ID" has been assigned, this "Substance ID" shall be specified based on the Substance Name controlled vocabulary.
Example(s)	UNII Code, EV-Code, Global Substance ID
Conformance	MANDATORY
Data Type	II
Values Allowed	Value could be a code associated with a preferred term, or a specific type of data. If a code is transmitted the preferred term associated with that code should also be transmitted.
Business Rule(s)	All substances will be identified by a single ID. NOTE: The ID will only be released to the public if the defining information is in the public domain or if a company that provides the defining information requests public release or releases the code in public marketing materials. An ID will always be issued to an organisation that requests an ID and supplies the information necessary to define a substance. A flag to control the release of the ID to the general public is part of the Reference Source information (4.6).

4.5 Substance names (Mandatory)

Although ISO 11238 does not provide any guidance on substance nomenclature, it does provide a structure for the capture of names and codes that are used to refer to a substance. For the purpose of ISO 11238 there are basically five types of names, Official names, systematic names, other names, brand names and company codes.

At least one name or company code should be associated with each substance. Names for a given substance and/or standardised names submitted by another party may be available in the substance terminology. Indication that a name is the preferred name or term by which a substance will be referred to in a given jurisdiction and domain may be presented in the terminology. This preferred term may vary depending on jurisdiction or domain or the state of development of a given substance.

EXAMPLE During the early stages of clinical development, the preferred term for a substance may be its company code. In later stages an INN name may be associated with the substance and become the preferred term for that substance.

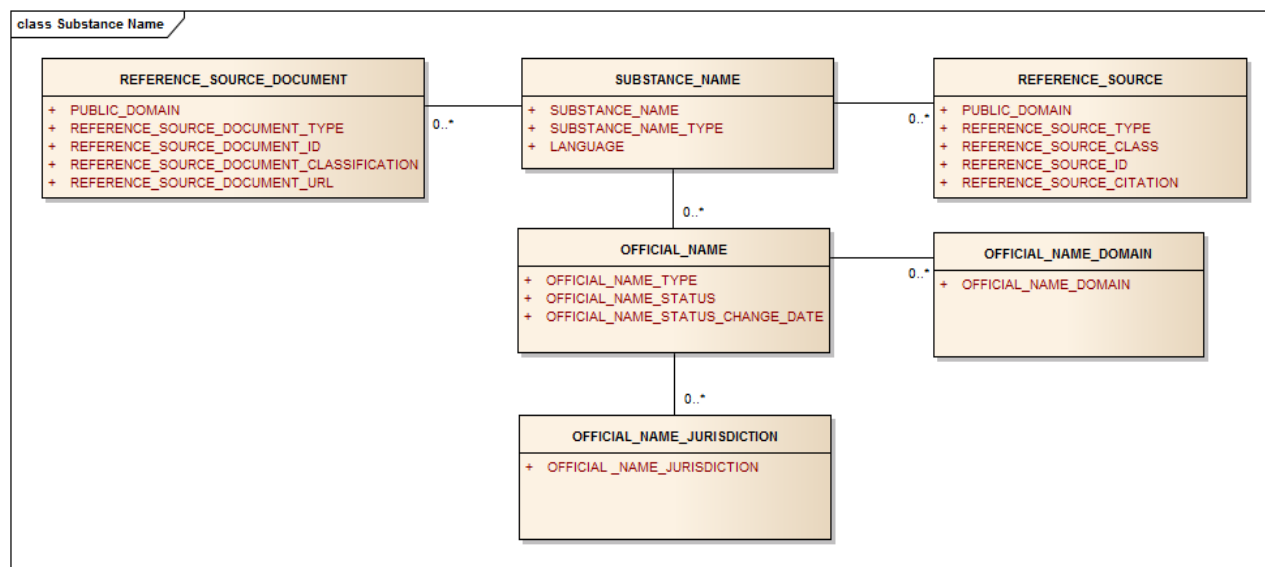


Figure 5 — Information model for Substance Names

4.5.1 Substance name

When requesting an ID a submitter should submit all names and company codes which have been associated with the substance. A submitter may also want to add a name to a substance that is already in the database. Defining information for a given substance should also be transmitted when requesting that a name is added to the database to ensure the correct substance identity. A submitter may also submit names in multiple languages. Although there is no requirement that names are unique, other or common names should also explicitly distinguish related substances from one another. Distinguishing characteristics for many polymers, such as molecular weight, degree of polymerisation, and viscosity, should be indicated in the names of these substances. Standardisation on such names to facilitate searching and identification of substances may be adopted within regions.

EXAMPLE Polyethylene glycol is a name for a family of polymers, so it would not be an acceptable name for an individual substance. Polyethylene Glycol 1000 is an official name in the US for a polyethylene glycol polymer with an average molecular weight of 1000. In Europe the INN name Macrogol 1000 would be an official name for the same substance. In the personal care products domain PEG-20 is the official name for the same substance and 20 refers to the average degree of polymerisation.

Table 4 — Substance name

User Guidance	Name or company code associated with the Substance.
Example(s)	"Amoxicillin"
Conformance	MANDATORY
Data Type	ST
Values Allowed	Free text, multiple names allowed, each in its own <asNamedEntity> element.

4.5.2 Substance name type

Every name shall have one and only one name type. Official names are typically non-proprietary names used in a given jurisdiction and domain to refer to a specific substance. The domain, jurisdiction, and authority that assigned the name (USAN, INN, JAN etc.) and the language of the name are also captured.

Systematic names will typically be used for simple chemicals and structurally diverse materials where the definitions are based on a chemical structure or systematic taxonomic information. For chemicals these names are typically derived from IUPAC or CAS systems of nomenclature. From the systematic names of chemicals a molecular structure can typically be derived and the name can be checked by a number of chemical drawing programmes that convert a given name to a molecular structure. Brand names are names by which a company identifies a given substance typically for marketing purposes. Other names associated with a given substance are for instance common names. Company codes (also known as lab/laboratory codes or clinical trial codes) are also treated as names. They are assigned by a given company to substances in clinical or preclinical development.

Table 5 — Substance name type

User Guidance	Each name shall be associated with a type.
Example(s)	Official name, systematic name, other name, brand name, company code
Conformance	MANDATORY
Data Type	CD
Values Allowed	OFFICIAL, SYSTEMATIC, COMPANY CODE, OTHER, BRAND
Business Rule(s)	A name shall have one and only one name type.

4.5.3 Language

All names should be associated with at least one language. Often the name, particularly INN names, will be identical in multiple languages. The language for company codes that use a Latin character set will be described as English (en). The language for the systematic name of plant, microorganism or animal derived substances will be also be described as English since many of these will be a hybrid of a Latin Name for the species and an English name for the part and/or the subspecies information.

EXAMPLE The INN name alisertib is identical in English, French and Spanish.

Table 6 — Language

User Guidance	If the name is language dependent, that language shall be specified. The language used to provide the structured substance information shall be specified according to the XML standard's xml:lang attribute specification, which is based on ISO 639-1, alpha-2 codes
Example(s)	en – English, de – German, fr –French
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 639-1, alpha-2 codes
Business Rule(s)	While xml:lang allows country-specification of language, e.g. en_US vs. en_UK or pt_PT vs. pt_BR, such specificity is to be avoided except if the country variant of the language exceptionally does give rise to an alternative spelling. In that case, only the exceptional spelling needs to be tagged with the country-specific language code, not the regular spelling for that language in all other countries. If the same name spelling exists for multiple languages of interest, the entire name is repeated each time with its own language code.

4.5.4 Official name (repeat as necessary)

Each official name shall have one or more official name types. The official name type reflects the organisation that assigns or recognises the name associated with the substance. These names are typically non-proprietary names that are used in the labelling of pharmaceuticals. The domains and jurisdictions in which the official name is used are also captured, tracked and maintained within the terminology. Although most official names will refer to

material at the substance level there will be instances where an official name will refer to material at the specified substances Group 1 level [e.g. Simeticone (BAN, INN) or Simethicone (USAN)]. The official name class is optional and shall be repeated as necessary.

4.5.4.1 Official name type

Any authority which assigns official names will typically do so in a publication, and as such, referencing the publication is sufficient indication for the name being “official” and sufficient representation for the “official name type”. When new official names are issued by official naming bodies, the naming bodies will be reflected in this terminology.

Table 7 — Official name type

User Guidance	Designation of which authority assigned the Official Name. All official names need to have at least one such designation. The name type is the name of the authority or authorities that have assigned or have adopted the name.
Example(s)	BAN, COSING, EP, FCC, INCI, INN, JAN, JECFA, MARTINDALE, USAN, USP.
Conformance	MANDATORY — all official names require a designation of the naming authority
Data Type	CD
Values Allowed	All values in the name assigning authority identifier system.
Business Rule(s)	All official names must have a naming authority designated. If multiple authorities have assigned the same name, a separate <asNamedEntity> is specified for each authority that needs to be listed.

4.5.4.2 Official name status

The status of the official name is associated with a given jurisdiction. It is possible that an official name could be official in one jurisdiction and superseded in another. With some official name types an initial name will be proposed and the name will later become recommended. Recommended INN names would be considered to be the Primary Official name in many jurisdictions.

Table 8 — Official name status

User Guidance	The status of the official name.
Example(s)	Current, Alternate, Superseded, Proposed
Conformance	OPTIONAL
Data Type	CD
Values Allowed	Current, Alternate, Superseded, Proposed
Business Rule(s)	Each official name shall have a single status. The fact that a name is alternate is evident by having more than one name assigned to the same substance by the same authority for the same language, domain, etc.

4.5.4.3 Official name status change date

The change date of the status of the official name shall be specified according to the ISO 8601 date format (the variant without any delimiters).

Table 9 — Official name status change date

User Guidance	Date of official name change if known.
Example(s)	20110601
Conformance	OPTIONAL
Data Type	DATE
Business Rule(s)	If official name status changes the date of the official change will be captured if known.

4.5.4.4 Official name domain (repeat as necessary)

Official names will also be associated with a domain. The domain and the actual name of a domain will vary according to jurisdiction.

EXAMPLE The terms DRUG and BIOLOGIC could be used in Japan and the US as official name domain. In the EU the term MEDICINE may cover the name of many substances that would be either DRUGS or BIOLOGICS in the US and Japan.

Table 10 — Official name domain

User Guidance	Specifies when and for what purpose the official name is to be used. All official names can have at least one domain. This is useful to differentiate different names for the same substance as used for a drug active ingredient as opposed to a food colour additive.
Example(s)	BIOLOGIC, COSMETIC, DRUG, FOOD, MEDICINE, DIETARY SUPPLEMENT
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Used if different names exist that are not appropriate to be used across all domains.

4.5.4.5 Official name jurisdiction (repeat as necessary)

When a product is on the market and contains pharmaceutically active substances at least one jurisdiction should be associated with the name of the active substance. For inactive substances or excipients the jurisdiction may not be apparent or captured. Pharmacopoeias will often be consulted to determine the jurisdiction of an official name. The Official Name Jurisdiction class is optional; when necessary the following conformance apply:

Table 11 — Official name jurisdiction

User Guidance	The jurisdiction of the official name shall be provided.
Example(s)	US, EU, JA
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 3166-1 alpha-2 codes For the European Union EU shall be used.
Business Rule(s)	All the jurisdictions in which the Official Name is to be used should be indicated.

4.6 Reference sources (Mandatory)

All data elements associated with a substance or specified substance should have a reference source. These sources could be regulatory submissions, publications by naming authorities such as INN or USAN, pharmacopoeias, books, journals, web sites, meeting materials of public databases such as chemical abstracts. A source can refer to multiple data elements and a given element can have multiple sources.

4.6.1 Public domain

Most reference sources other than regulatory submissions will be in the public domain. All regulatory submissions will typically be considered to not be in the public domain. If the reference source is in the public domain the data elements referenced by the source could be made public. There may be a separate determination whether any portion of the record can be made public. This determination may vary depending on type of substance. Regardless submitters of information should always explicitly indicate whether data should be considered confidential.

EXAMPLE Particular care should be taken in the release of company codes. Although the company codes are frequently mentioned in the public domain the release of a company code associated with a Substance ID and the defining information should only be made if all of the defining elements are associated with the company code in a single source in the public domain. For chemicals, the chemical structure should be associated with the company code in the public domain. For proteins such as monoclonal antibodies, names or company codes could be released to the public only if the target of the antibody is mentioned in the same public source as the name or code. A separate determination will be made as to whether the protein sequence is in the public domain.

Table 12 — Public domain reference source

User Guidance	Indication of whether the source is in the public domain. Company codes will only be released if the defining information is associated with the code in a public source such as chemical abstracts (CAS), a journal article, or a poster at scientific meeting attributable to the company.
Example(s)	Yes, No
Conformance	OPTIONAL
Data Type	BL
Values Allowed	Yes, No
Business Rule(s)	All regulatory submissions are considered not to be in the public domain. There may be an additional flag that determines whether any information can be released on a given substance.

4.6.2 Reference source type

All reference sources will be associated with a type. The type may actually identify the source of the information if it is a public database or naming authority. For official names and defining information, the primary source should be used if available. A single source can be used for many data elements. Public databases particularly those of official naming bodies, regulatory submissions, manufacturer's technical specifications, scientific journal articles and pharmacopoeias are all sources that can be used.

Table 13 — Reference source type

User Guidance	The reference source type in which the data elements were actually found.
Example(s)	BP, CHEMID (NLM), EP, IND, INN, ITIS, JAN, JOURNAL, JP, KEGG, MARTINDALE, NDA, PERSONAL CARE PRODUCTS COUNCIL (PCPC), PUBCHEM, USAN, USP, WEB PAGE
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	All names should have at least one source and hence source type.

4.6.3 Reference source class

Each reference information type will be associated with a class. The class will be automatically associated with each type of information.

Table 14 — Reference source class

User Guidance	Useful for classification of sources.
Example(s)	Regulatory submission
Conformance	OPTIONAL
Data Type	CD
Values Allowed	LITERATURE, OFFICIAL NAME SOURCE, REGULATORY SUBMISSION, WEB.
Business Rule(s)	Need not be entered as the value is dependent on the Reference Source Type.

4.6.4 Reference source ID

Many sources such as regulatory submissions, INN lists and others will be associated with a number or ID. The ID should be captured.

Table 15 — Reference source ID

User Guidance	An ID associated with reference source.
Example(s)	IND number, INN list; PMID (PubMed Identifier).
Conformance	CONDITIONAL
Data Type	ST

4.6.5 Reference source citation

A citation should be captured for a literature reference. The specific format for a given citation may be defined at regional level.

Table 16 — Reference source citation

User Guidance	The actual reference source or citation in which the name was found.
Example(s)	Literature reference: Encycl. 3: 429 (1792)
Conformance	CONDITIONAL
Data Type	ST

4.6.6 Reference source document (new class to be included in the second edition of ISO 11238)

The document should be concise and contain information specific for given data elements.

Table 17 — Reference source document

User Guidance	Documents that contain relevant information on data elements and are referenced should be stored if available.
Conformance	OPTIONAL
Data Type	ED
Values Allowed	The documents can be saved as PDF files or TEXT files. If the documents or data source contains structured data such as spectra, it can be transmitted in a standard format that can be visualised and searched.

4.6.7 Reference source document type (new class to be included in the second edition of ISO 11238)

Table 18 — Reference source document type

User Guidance	The type of the document
Example(s)	Regulatory submission, Journal article , Release specification, In House specification
Conformance	OPTIONAL
Data Type	CD

4.6.8 Reference source document classification (new class to be included in the second edition of ISO 11238)

A classification system will be also developed for documents. The classification system will indicate the type of information contained within the reference document based on a terminology for the classification of reference source documents.

Table 19 — Reference source document classification

User Guidance	The reference document should be classified according to the information contained within
Example(s)	Chemical structure, Name, NMR spectra, Mass Spectra,
Conformance	OPTIONAL
Data Type	CD

4.6.9 Reference source URL (new class to be included in the second edition of ISO 11238)

Table 20 — Reference source URL

User Guidance	The reference source URL in which the name was found. The URL can be captured here.
Conformance	CONDITIONAL
Data Type	ST

4.7 Substance code (Conditional)

This section of reference information allows the capture of codes that are typically associated with a substance. The capture of these codes typically facilitates the defining and mapping of substances and linking of substances to a variety of information sources. Codes should be communicated if known but are not necessary or mandatory. The terminology should have the ability to capture, store and distribute codes as necessary. All the codes that are captured should be associated with a publicly recognised code system and map directly to a given substance. It should be noted that company codes are not captured in this section but are considered a type of name for a given substance. Commonly used public codes such as CAS registry numbers with a given substance should be provided and associated to the substance.

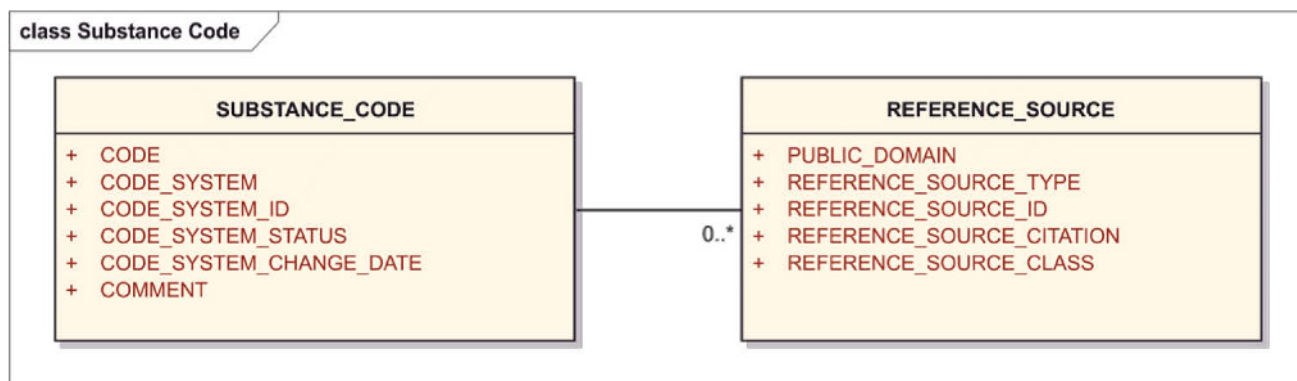


Figure 6 — Information model for substance codes

4.7.1 Code

The actual code should be captured using the same format that is used in the code system. Only codes associated with a code system will be captured. The code should be specifically associated with a given substance. Many public and non-public databases identify substances with a code and these codes can be very helpful in mapping substances to various systems. Codes should always be verified against the source system. Different jurisdictions may require a code from a code system or multiple code systems to be associated and submitted with a substance.

Table 21 — Code

User Guidance	The specific code (the value) with regard to a referenced code system. NOTE: InChI is not considered a code and shall be specified within the structure description.
Example(s)	11-22-356
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Only codes from recognised code systems will be captured. The code should specifically link to a substance but need not uniquely link to substance. This code can also be a deprecated Substance ID generated based on ISO 11238.

4.7.2 Code system

Every code should be associated with a code system and a list of names of code systems expected to be submitted should be maintained by the maintenance organisation.

Examples CAS registry numbers, EC numbers, FDA UNII codes, EMA XEVMPD codes, ASK numbers, EPA Pesticide codes.

Table 22 — Code system

User Guidance	Name of the code system.
Example(s)	CAS, EINECS, NSC, ASK, ATC, NDF-RT; RX-CUI ETC.
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	If the Code class applies, the code system name is implicit and derived from the Code System ID.

4.7.3 Code system ID

The code system ID may be an OID or an ID as applicable. The ID is associated with code system and need not be submitted.

Table 23 — Code system ID

User Guidance	The ID of the code system.
Example(s)	CAS Registry code: 0049; FDA Substance Registration System (UNII) 0050
Conformance	CONDITIONAL
Data Type	II
Business Rule(s)	If the Code class applies, the code system ID is required as all codes must be linked to a code system.

4.7.4 Code system status

The code system status is typically the status provided by the system for a given code. Many code systems can end up with multiple codes representing the same substance. Often one code will be the preferred code and others will either be deleted or deprecated. Ad-hoc terms may also be developed to indicate a code status.

Table 24 — Code system status

User Guidance	The status of the code assignment
Example(s)	current, proposed,
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	“active” — Current, “active” — Alternate, “superseded” — Terminated, “pending” — Proposed

4.7.5 Code system status change date

Table 25 — Code system status change date

User Guidance	The date at which the code status is changed as part of the terminology maintenance.
Example(s)	20110601
Conformance	OPTIONAL
Data Type	TS.DATE

4.7.6 Comment

Table 26 — Comment

User Guidance	Any comment can be provided in this field, if necessary.
Conformance	OPTIONAL
Data Type	ST

4.7.7 Reference source

The reference source will be captured according to 4.6.

4.7.7.1 Reference information

Where applicable, the reference information shall be provided.

4.7.8 Substance classification (repeat as necessary)

The standard should have the ability to capture a variety of classification systems. Classification systems are typically based on molecular structure, chemical properties, pharmacological effects, mechanism of action, therapeutic targets or indication.

The standard shall have the ability to capture multiple classifications and variable levels of classification. Although most classification will be associated with an external classification system, ad-hoc classification of substances may be developed within this terminology as needed.

The information in Figure 7 shall be provided:

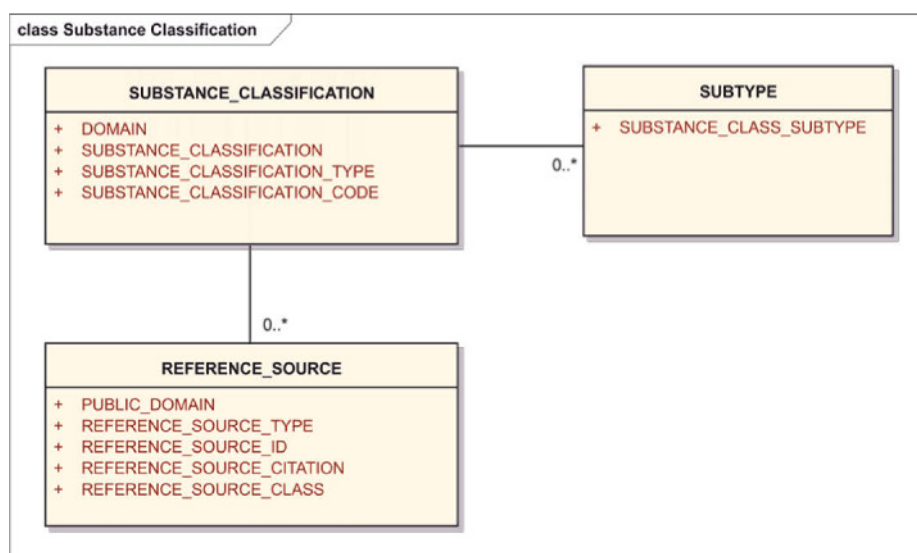


Figure 7 — Information model for substance classification

4.7.8.1 Domain

Classification systems will often only be used within a specific domain. The domain will be associated with the classification system.

Table 27 — Domain

User Guidance	The domain of the substance classification shall be provided.
Example(s)	Human pharmaceuticals, human vaccine, animal drug, animal vaccine, food, food additive, colorant, pesticide; tobacco additive; flavour, excipient
Conformance	OPTIONAL
Data Type	CD

4.7.8.2 Substance classification (code system)

There are a number of systems that classify substances based on chemical structure, pharmacological effect, organ system affected, mechanism of action or molecular targets. These classification systems are useful and this information, if available, should be provided and maintained as applicable.

Table 28 — Substance classification (code system)

User Guidance	Substance Classification terms based on a controlled vocabulary that is controlled by the system or type of substance classification. The system should be capable of capturing a variety of classification systems and each system should be identified. All active ingredients are typically classified in multiple systems. Ad-hoc classification may be created as necessary.
Example(s)	ATC, NDF-RT, NCI, AHPA, MESH; Therapeutic Category of Drugs in Japan
Conformance	OPTIONAL
Data Type	CD

4.7.8.3 Substance classification type

Nearly all classification systems contain multiple levels. Type is the highest level or the initial level of the classification system.

Table 29 — Substance classification type

User Guidance	The initial type or highest level of a classification system.
Example(s)	Organ system in ATC.
Conformance	OPTIONAL
Data Type	CD
Business Rule(s)	The value is IMPLICIT and can be derived by the "Substance Classification — Subtype" if available.

4.7.8.3.1 Substance classification subtype (repeat as necessary)

Table 30 — Substance classification subtype

User Guidance	Many classification systems will have multiple subtypes of classification this field will attempt to capture this information.
Example(s)	NDF-RT Mechanism of Action (MOA) or Physiological Effect.
Conformance	OPTIONAL
Data Type	CD

4.7.8.3.2 Substance classification code

Classification code with map to the active moiety of a given substance.

Table 31 — Substance classification code

User Guidance	Code from the classifying code system. The code should be tied to the substance...
Example(s)	N0000175684 NDF-RT code for Oxycodone that maps to the mechanism of action classification. N02AA05 ATC code that maps to Oxycodone.
Conformance	OPTIONAL
Data Type	II
Business Rule(s)	The classification code shall be part of an internationally recognised classification system.

4.7.8.3.3 Reference source

The reference source will be captured according to 4.6.

4.7.9 Target

The target section allows the capture of information related to the targets that a substance is known to interact with and the interaction results in a biological effect or transformation. These targets can be therapeutic, metabolic, interactions postulated to result in toxicity or interactions of unknown effect. Targets are typically proteins, or complexes of proteins. The target may be a substance in its own right and ID shall be assigned and/or existing ID of a known system used (i.e. GENE, Uniprot) to refer to a target.

For monoclonal antibodies the target shall always be provided. For other types substances the target may be provided if known. Ideally the target is a substance with its own Substance ID. However, in specific circumstances, the target may be handled as a relationship between two substances.

4.7.9.1 Target ID

Typically the Target ID is either a Substance or specified substance ID. ISO 11238 will be updated to rename Target Gene ID element as Target ID.

Table 32 — Target ID

User Guidance	ID of the target relation molecular entity upon which the substance acts. One Target ID should be specified. The target should be a molecular entity or a family of molecular entities. The target is usually a protein entity that the drug actually interacts with. Both therapeutic and metabolic targets and targets associated with toxicity should be captured.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 Substance ID
Business Rule(s)	The ISO 11238 Substance ID shall be referenced.

4.7.9.2 Target name

At least one name of the target shall be provided when target information is submitted. Target names will typically be handled no differently than other substance names. If an active ingredient targets more than one target each target should be listed. In addition to names a list of codes associated with the target may also be used. Gene bank and Uniprot codes can be particularly useful for identifying and tracking targets. The ISO 11238 standard will be updated to rename Target Gene Name element as Target Name.

Table 33 — Target name

User Guidance	Name of the target molecular entity upon which the substance acts. Each target should be associated with a target name.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 substance name
Business Rule(s)	ISO 11238 substance name implicitly derived from Target ID. A variety of names can be captured. A single name will be a preferred name and associated with the Target ID.

4.7.9.3 Interaction type

The type of interactions between a substance and its target should also be captured. A controlled vocabulary should be developed to capture these types of interactions.

Table 34 — Interaction type

User Guidance	Type of interaction between the substance and the target. The interaction type will be captured for all substrate-target interactions. A substance can have multiple types of interactions.
Example(s)	AGONIST, INDUCER, INHIBITOR, PARTIAL ANTAGONIST, SUBSTRATE, PRODUCT
Conformance	CONDITIONAL
Data Type	CD

4.7.9.4 Target organism (new class to be included in the second edition of ISO 11238)

When drugs are targeted against a specific organism the organism should be captured. For antivirals and antibacterials there is typically no need to give the strain level unless the organism is resistant to more commonly used drugs.

Table 35 — Target organism

User Guidance	The organism for which the active substance is targeted.
Example(s)	HOMO SAPIENS; M. TB, HIV-1 etc.
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	At least one organism type should be associated.

4.7.9.5 Target organism type

Every organism should have a type that is maintained within this terminology. The organism type may also be used independent of the organism particularly for substances such as antibacterials which may be targeted against a wide range of organisms.

Table 36 — Target organism type

User Guidance	Every target organism should have a type associated with the organism.
Example(s)	HUMAN, VIRAL, BACTERIAL, FUNGAL, INSECT, HELMINTH, MALARIAL
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	At least one organism type should be associated.

4.7.9.6 Target type

Each target-substance interaction should be assigned a type. The type is dependent on the substance and also to some extent the use of the substance. A given target-substance interaction can be of more than one type.

EXAMPLE The therapeutic target would be HIV-1 protease when ritonavir is used in HIV-related therapy. The metabolic target is Cyp 3A4 and ritonavir is an inhibitor of this target. Ritonavir is also an inhibitor of P-glycoprotein a drug transporter. In many combination products ritonavir is used strictly to inhibit Cyp 3A4/5 and transporter enzymes.

Table 37 — Target type

User Guidance	Type is a limited form of classification associated with a target of an active substance. Some targets can be both toxic and therapeutic; the choice is dependent on the dose of a given substance.
Example(s)	THERAPEUTIC, METABOLIC, TOXIC; TRANSPORTER
Conformance	OPTIONAL
Data Type	CD
Business Rule(s)	Each target should be associated with a target type when the target can be classified,

4.7.9.6.1 Reference source

The reference source will be captured according to 4.6 when necessary.

4.7.10 Gene

The gene section will only be captured for proteins and nucleic acids. It does not apply for chemicals, polymers or structurally diverse substances. It is only used as reference information and refers to the gene that is the origin of the substance (typically a protein). Genes may also be substances and would be described as nucleic acids.

4.7.10.1 Gene sequence origin

For monoclonal antibodies the gene sequence origin will be used to capture species from which the antibody was derived.

EXAMPLES For monoclonal antibodies the typical values can be HUMAN, HUMANISED MOUSE; HUMANISED RAT; CHIMERIC MOUSE-HUMAN; CHIMERIC RAT-HUMAN.

Table 38 — Gene sequence origin

User Guidance	Common name for the organism from which the final gene sequence originated.
Example(s)	Human, Bovine, Human; Humanised; Mouse-Human Chimeric; Mouse
Conformance	OPTIONAL
Data Type	CD
Values Allowed	Typically from the NCBI Gene database.
Business Rule(s)	NOTE: it applies primarily to proteins.

4.7.10.2 Gene ID

Table 39 — Gene ID

User Guidance	ID associated with the gene of origin
Example(s)	YP_299723.1
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	Typically from the NCBI Gene database.

4.7.10.3 Gene name

Table 40 — Gene name

User Guidance	Complete name given for a gene. Every gene which has an ID should also have a name.
Example(s)	hIL-10 gene
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Implicit from the Gene ID from the applicable Gene database.

4.7.10.3.1 Reference source

The reference source and related information will be captured according to 4.6.

4.7.11 Gene elements

Gene elements will be captured for genes that are used in gene therapy.

When the gene element section applies, the following information shall be provided:

4.7.11.1 Gene element type

Table 41 — Gene element type

User Guidance	Type of the gene element. Should be captured for all genes that are transduced into cells and are intended to be expressed.
Example(s)	enhancer, promoter, silencer, terminator, repressor
Conformance	CONDITIONAL
Data Type	CD

4.7.11.2 Gene element ID**Table 42 — Gene element ID**

User Guidance	Unique identifier for gene element.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	A unique identifier will be associated with each gene element.

4.7.11.3 Gene element name**Table 43 — Gene element name**

User Guidance	Specific gene element name.
Example(s)	SV40 Enhancer
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	A unique identifier will be associated with each gene element.
Business Rule(s)	Implicit from the Gene element ID from the applicable Gene element controlled vocabulary.

4.7.11.4 Reference source

The reference source and related information will be captured according to 4.6. One or more reference sources could be cited.

4.7.12 Substance relationship

Relationships between substances such as the relationship between a salt form and its active moiety or parent substance should be captured. These relationships are often essential to the description of medicinal products, the basis of strength and the classification of substances. These relationships are often obvious and rules will be developed for specifying substances involved in each type of relationship. For example the active moiety of all sodium salts would be the free acid, conversely the active moiety of a hydrochloride salt would be a free base.

4.7.12.1 Relationship

The relationship is not a mere code but the actual connection between the two substances which may be different for different kinds of relationships:

There are a variety of different relationships that are possible but for the identification of medicinal products the most important relationships are between a substance, its active moieties, the basis of strength and constituents

in herbal and other complex substances. Parent and salt-solvate relationships and relationship of drug substance to its metabolites and impurities are also important although not essential for the description of a medicinal product. If there are known relationships between substances the relationships should be provided and maintained. **The parent-active moiety relationship** may be essential in describing medicinal products.

Table 44 — Relationship

User Guidance	Relationship between substances. NOTE: Target and gene elements are not to be specified in this section.
Example(s)	parent, active moiety, salt, radiolabelled, prodrug, metabolite, enantiomer
Conformance	CONDITIONAL
Data Type	CD

4.7.12.1 Interaction type

Table 45 — Interaction type

User Guidance	The Type of interaction between the substances.
Conformance	OPTIONAL
Data Type	CD

4.7.12.2 Related substance ID

Table 46 — Related substance ID

User Guidance	Substance ID associated with the related substance. All related substances should have a Substance ID.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 Substance ID

4.7.12.3 Related substance name

Table 47 — Related substance name

User Guidance	Preferred term or primary name of related substance.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	Based on ISO 11238 substance name
Business Rule(s)	The preferred term of the related substance is captured; this information is required when there is a "Substance Relationship"; The substance name is implicit and derived from the Substance ID.

4.7.12.3.1 Amount type

Table 48 — Amount type

User Guidance	Most elements that require a quantitative value will also have a field called amount type. Amount type should always be specified because the actual value of the amount is often dependent on it. EXAMPLE: In capturing the actual relative amounts of substances or molecular fragments it is essential to indicate whether the amount refers to a mole ratio or weight ratio. For any given element an effort should be made to use same the amount type for all related definitional elements.
Example(s)	Mole percent; Weight percent; Mole ratio; Weight ratio, Weight, Moles
Conformance	CONDITIONAL
Data Type	CD

4.7.12.4 Amount

For some substance relationships it may be desirable to capture a quantitative or semi-quantitative amount that may further define the relationship (i.e. major metabolite, degradants etc.). Amount shall be captured according to 4.9.

Substance Relationships which refer to reactions (e.g. metabolism, prodrug) are quantified by specifying the quantity of the reactants:

4.7.12.4.1 Reference source

The reference source will be captured according to 4.6.

4.8 Structure (repeat as necessary) (Conditional)

Structural information is an essential element for all chemical substances and for other types of substances that have structurally definable elements or modifications. The representation should contain structural information in one or more of the standardised formats as indicated below. A graphical image of the molecular structure should also be provided if available. A complete representation of the structure should be provided and is usually sufficient to define chemical substances. Structural representations can be in a variety of formats or even multiple formats. The commonly used formats include molfile, SMILES, InChI, and cdx. Indication of a preferred structural representation may be defined at regional level. Although most chemical substances may be represented by a single unambiguous structural representation, there are also many substances that can be represented by multiple structural representations or as a mixture of single substances. For the initial implementation and the transmittal of data for substance definitions, an approach that limits submissions to a single or a limited number of unambiguous structural representations is recommended. There are several examples below that illustrate the process in choosing a defining structural representation to submit. A more detailed structure representation guide is provided as part of the annex B (Chemical substance).

For Amine Salts, with the exception of ammonium and quaternium, salts shall be represented as a free base.

EXAMPLE: DIFLOXACIN HYDROCHLORIDE

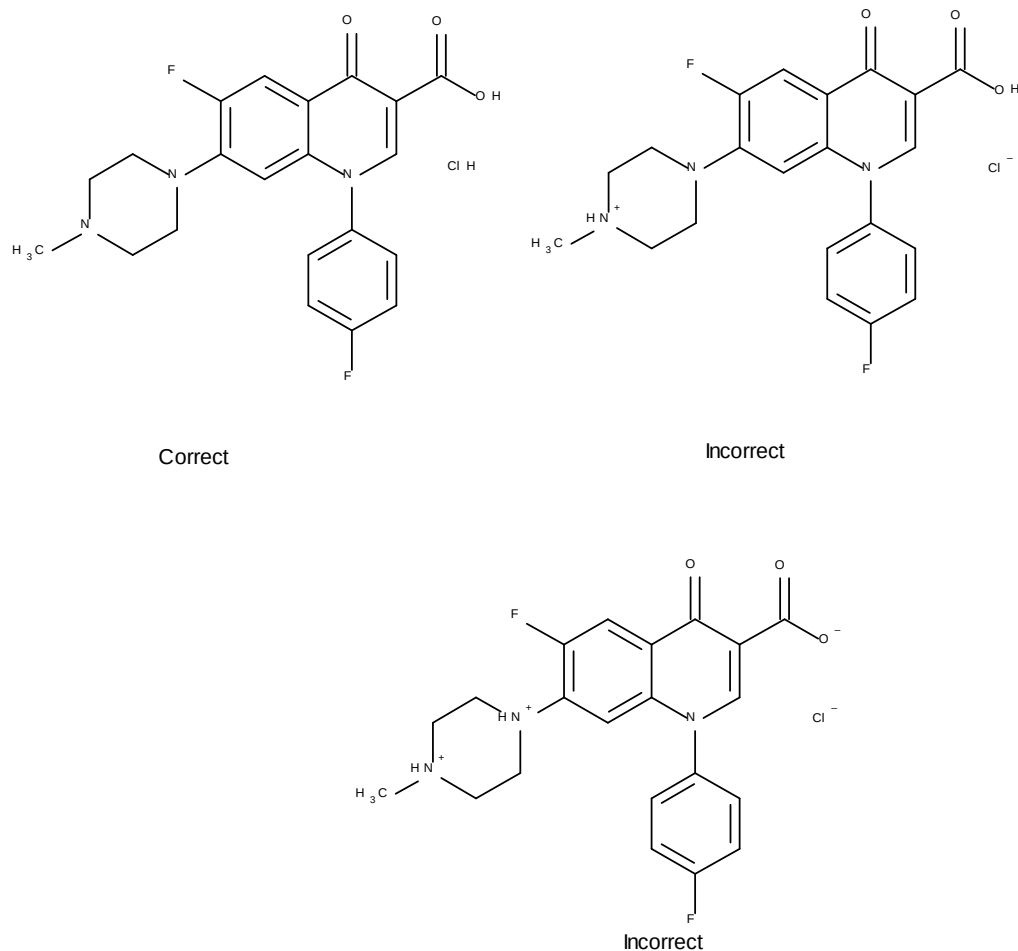


Figure 8 — Correct and incorrect graphical images of the molecular structure of difloxacin hydrochloride

Metal and ammonium salts: All metal and ammonium salts (NH_4^+) shall be represented in a charged configuration with all equivalent functional groups ionised. Charge balance will be achieved when necessary by adding hydrogen ions (protons).

EXAMPLE: MONOSODIUM CITRATE

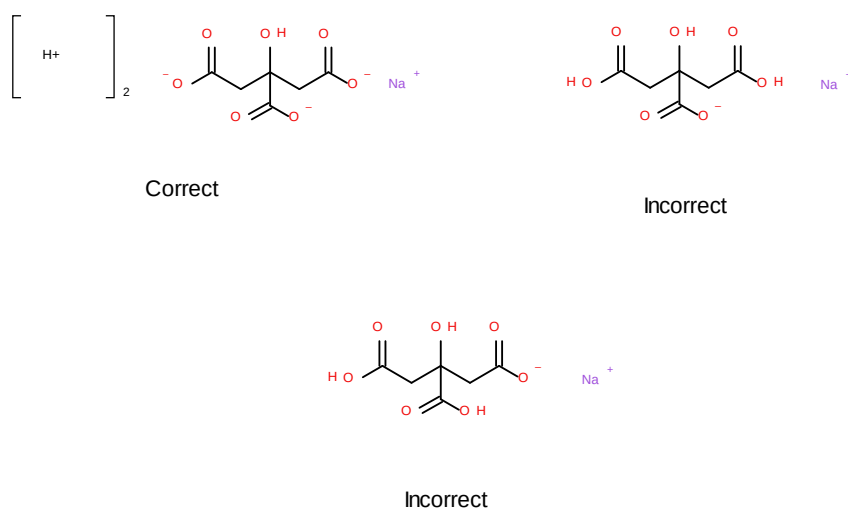


Figure 9 — Correct and incorrect graphical images of the molecular structure of monosodium citrate

Racemic mixtures can be represented as mixtures of enantiomers or by a single structural representation. When using a single structural representation and the substance only contains one chiral centre, a single structure will be drawn that does not contain a chiral bond. If more than one chiral centre exists, relative stereochemistry shall be indicated. The stereochemistry of each moiety of a given substance will be captured separately (4.9).

EXAMPLE: OFLOXACIN

A racemic mixture of the R and S isomers. A single structural representation can be used in this case.

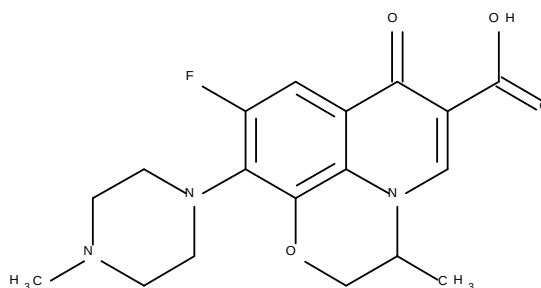


Figure 10 — Graphical image of the molecular structure of difloxacin hydrochloride

There are also substances that are mixtures of multiple enantiomers. These substances can also be transmitted in a single structural representation.

Substances can exist in interconverting forms or with a different defined structure depending on the physical state. In cases where the actual structure is a single configuration in a solid physical state, that structure should be captured. When a substance exists in a liquid state and can interconvert to a variety of structures, the least ambiguous single structure should be chosen.

EXAMPLE DEXTROSE (GLUCOSE): Exists as three separate substances in the crystalline state. For a crystalline substance one of the following representations should be chosen.

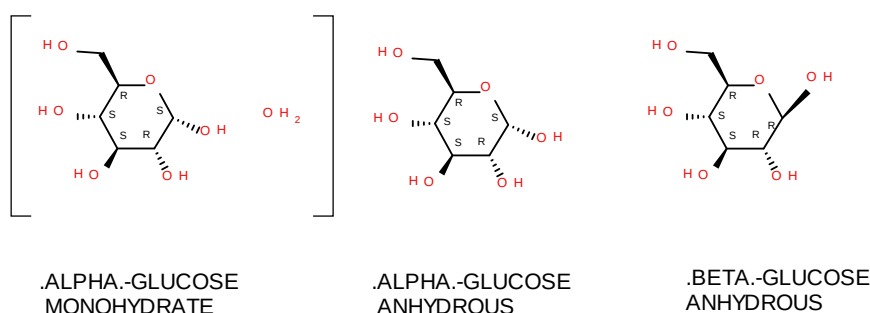


Figure 11 — Graphical images of the molecular structure of the three crystalline states of dextrose

In the liquid state dextrose and many other monosaccharides, disaccharides and higher homologs that end in an aldehyde are often referred to as reducing sugars and exist as mixtures of interconverting substances as illustrated in the structures of glucose below.

In this case the linear single representative structure is chosen to represent dextrose in the liquid or amorphous state. If pyranose or furanose forms were chosen the representative structure would have one site of stereochemical ambiguity. One general rule is that if there are multiple interconverting structural representations for a substance the preferred representation is the one with the least ambiguity. In figure 12 the pyranose and furanose structures have one site of stereochemical ambiguity.

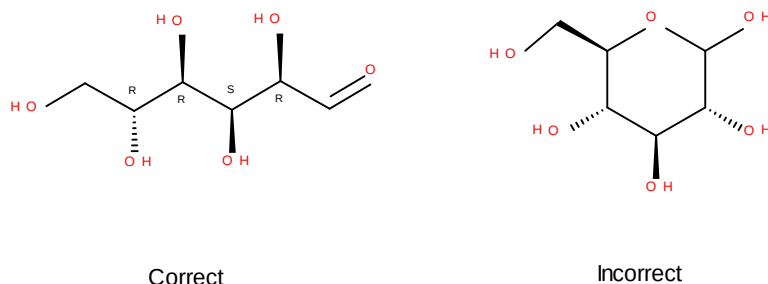


Figure 12 — Correct and incorrect graphical images of the molecular structure of (liquid) dextrose

In this case the linear single representative structure is chosen to represent dextrose in the liquid or amorphous state. If pyranose or furanose forms were chosen the representative structure would have one site of stereochemical ambiguity. One general rule is that if there are multiple interconverting structural representations for a substance the preferred representation is the one with the least ambiguity. In the example below the pyranose and furanose structures have one site of stereochemical ambiguity.

Structural representations will also be used to represent fragments of molecules. Fragments will be used in describing partially modified polymers and are typically generated to describe partially acylated polymers, proteins or even nucleic acids. In order to consistently describe substances a fragment will be generated in a defined manner. Ester and amides will always be broke such the oxygen or nitrogen with stays with the alcohol or amine. Ethers will be broken such that the oxygen stays with the polymeric or larger structural unit.

EXAMPLE: Cellulose Acetate

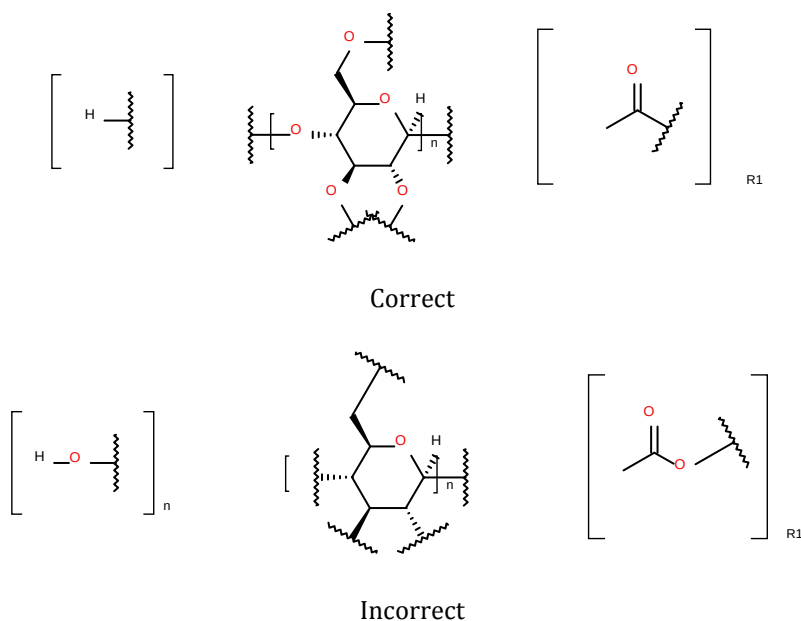


Figure 13 — Correct and incorrect graphical images of the molecular structure of cellulose acetate

The acetyl fragment is represented in Figure 14.

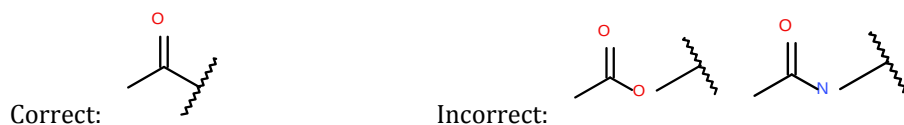


Figure 14 — Correct and incorrect graphical images of the molecular structure of the acetyl fragment

In fragments derived from ethers, the oxygen will be retained by the polymer or largest portion of a molecule.

EXAMPLE: Hypromelloses

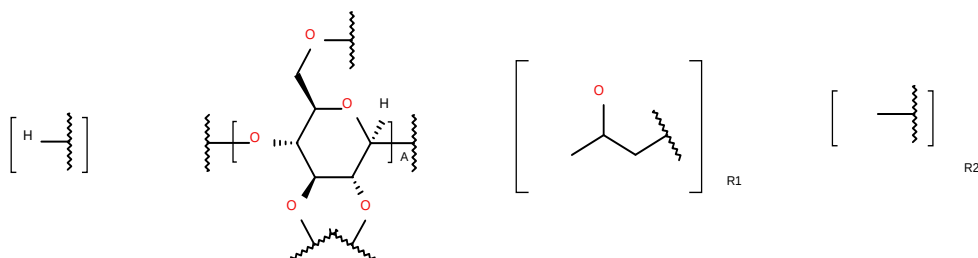


Figure 15 — Graphical images of the molecular structure of hypromelloses

Figure 16 provides the high level information for the Class Structure.

In the Class Structure the element group molecular formula is represented by molecular formula by moieties and by the total of moieties in accordance with the Hill system captured by the term molecular formula. In addition the element group molecular weight is connected to the Class Structure also for stoichiometric substances. This addition will be proposed in the next version of the ISO 1238 Standard.

See also the high level information at the substance type chemical.

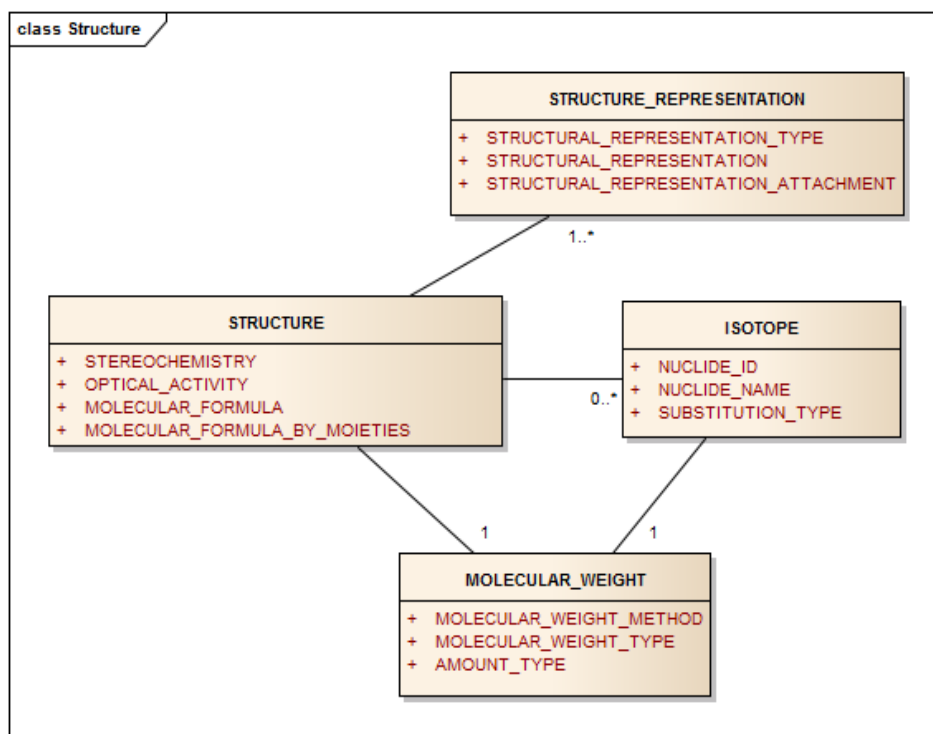


Figure 16 — Information model for structure and isotope

A structural representation is an essential element for defining chemical substances and is often sufficient for defining most chemical substances.

Four common structural representation types are listed below. Each of the types can unambiguously describe substances with a definite molecular structure. Stereochemistry should be completely defined in each structural representation. Only MOLFILES and CDX have the ability to describe polymeric repeating units and molecular fragments. There are a number of tools available, both free and commercial, that allow the creation of molecular structural representations. Multiple structural representation types can be submitted.

This class is **CONDITIONAL**. When the Substance type (as defined in 4.3) is either chemical, or polymer this class becomes **MANDATORY** and the following data elements should be described:

4.8.1 Structural representation type

Table 49 — Structural representation type

User Guidance	This field indicates the type of structure. A FULL structure is a structure in which the complete connectivity of the substance is known and the substance is monodisperse. A PARTIAL structure is a structure in which either complete connectivity or stereochemistry is not defined. A REPRESENTATIVE structure is used when the connectivity of the underlying material is diverse and a single structure is needed to represent that material
Example(s)	Full, partial, representative, fragment, ionic moiety, structural repeating unit
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	A structure can only be of one type. Field is mandatory when structural elements are present. When the Substance type (as defined in 4.3) is either chemical, protein, polymer or nucleic acid, this class becomes MANDATORY

Structural representation type is specified as a MIME Media Type code:

Type	MIME Media Type
InChI	application/x-InChI
MOLFILE	application/x-mdl-molfile
SMILES	application/x-smiles
CHIME	application/x-mdl-chime

4.8.2 Structural representation

Table 50 — Structural representation

User Guidance	The structural representation shall be provided as text string, the following formats are acceptable: InChI, SMILES, MOLFILE, CDX.
Example(s)	InChI, MOLFILE SMILES, CDX
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Each structural representation will have one and only one type. When the Substance type (as defined in 4.3) is either chemical, protein, polymer or nucleic acid, this class becomes MANDATORY

4.8.3 Structural representation attachment

Table 51 — Structural representation attachment

User Guidance	An attached file that may contain the structural representation shall be provided for complex data formats, e.g. CDX, MOLFILE.
Conformance	OPTIONAL
Data Type	ED

4.8.4 Stereochemistry

The overall stereochemistry of the substance should be indicated in this field.

Table 52 — Stereochemistry

User Guidance	The stereochemistry of the substance shall be indicated in the structure and should be captured here. Special cases of stereochemistry that can't be indicated in the structure shall be described based on a controlled vocabulary. Mixtures of stereoisomers shall be represented explicitly as a mixture of substances with absolute stereochemistry. In case the absolute stereochemistry is unknown the substance definition should be marked as "Incomplete".
Example(s)	AXIAL R, AXIAL S, SQUARE PLANAR 1, SQUARE PLANAR 2, SQUARE PLANAR 3, SQUARE PLANAR 4, TETRAHEDRAL, OCTAHEDRAL 12, OCTAHEDRAL 22, OCTAHEDRAL 21, CHIRAL, ACHIRAL
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	AXIAL R, AXIAL S, SQUARE PLANAR 1, SQUARE PLANAR 2, SQUARE PLANAR 3, SQUARE PLANAR 4, TETRAHEDRAL, OCTAHEDRAL 12, OCTAHEDRAL 22, OCTAHEDRAL 21; RACEMIC, MIXED, CHIRAL, ACHIRAL, ABSOLUTE, AXIAL, EPIMERIC, MESO, UNKNOWN, CIS, TRANS
Business Rule(s)	When the Substance type (as defined in 4.3) is either chemical, protein, polymer or nucleic acid, this class becomes MANDATORY

4.8.5 Optical activity

The optical activity of a chiral substance should be captured if known. The extent of optical rotation is not mandatory at the substance level.

Table 53 — Optical activity

User Guidance	The optical activity shall be described based on a controlled vocabulary.
Example(s)	(+/-), (+), (-)
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	(+/-), (+), (-)
Business Rule(s)	Shall be entered if known for substances that have at least one moiety with stereochemistry defined as chiral. Can be a defining element when the absolute stereo configuration is not known.

4.8.6 Molecular Formula

Table 54 — Molecular formula

User Guidance	Molecular formula is written as a sequence of pairs of atom symbol and stoichiometric number without any text formatting or special characters. This formal notation is essentially a code system. Specified according to the Hill system, i.e. first C, then H, then alphabetical.
Example(s)	C ₂₆ H ₃₃ ClN ₂ O ₉ S
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Can be implicitly derived from the structure; Validity check for identity and structural representation

4.8.7 Molecular Formula by Moieties (new class to be included in the second edition of ISO 11238)**Table 55 — Molecular formula by moieties**

User Guidance	Specified per moiety according to the Hill system, i.e. first C, then H, then alphabetical. and each moiety separated by a dot.
Example(s)	$C_{20}H_{25}ClN_2O_5 \cdot C_6H_6O_3S \cdot H_2O$
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Can be implicitly derived from the structure; Validity check for identity and structural representation

4.8.8 Isotope (repeat as necessary)

Applicable for single substances that contain a radionuclide or a non-natural isotopic ratio, e.g. C-13 enriched material. All radionuclide and non-natural isotopes will also be represented in the structure representation for chemical substances. For other types of substances a structural representation could be created to indicate the position of substitution.

4.8.8.1 Nuclide ID**Table 56 — Nuclide ID**

User Guidance	A Substance ID for each non-natural or radioisotope is created. The ID is linked to a single atom zero valence element.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 Substance ID
Business Rule(s)	Mandatory when substance contains non-natural or radioisotope.

4.8.8.2 Nuclide name**Table 57 — Nuclide name**

User Guidance	The name for each isotope shall be provided.
Example(s)	CARBON C-13
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 Substance names
Business Rule(s)	Mandatory when substance contains non-natural or radioisotope. Implicitly derived from Nuclide ID.

4.8.8.3 Substitution type

Applicable for single substances that contain a radionuclide or a non-natural isotopic ratio, e.g. C-13 enriched material. All radionuclide and non-natural isotopes will also be represented in the structure representation.

Table 58 — Substitution type

User Guidance	The type of isotopic substitution present in a single substance: - specific if the site of attachment/substitution indicated in structure; - non-specific if the nuclide is distributed throughout molecule or substance; - Unknown if site unknown. Substitution refers to the relationship between the nuclide and the rest of the substance.
Example(s)	specific; non-specific; unknown
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Mandatory when substance contains non-natural or radioisotope.

4.9 Amount (Conditional)

The amount section serves to provide the quantitative or qualitative values that are associated with a variety of elements. The same format will be used for all quantitative, semi-quantitative and qualitative values. This amount section will be used throughout the standard for any properties or elements where quantitative, semi-quantitative or qualitative values apply. An amount type will usually be associated with each amount and the same amount type will typically apply to all of the quantitative values in a given section (i.e. for multiple fragment substitution either mole percent, weight percent or degree of substitution should be applied to each fragment they should not be mixed).

4.9.1 Average

Table 59 — Average

User Guidance	Used to capture quantitative values for a variety of elements. If only limits are given, the arithmetic mean would be the average. If only a single definite value for a given element is given, it would be captured in this field.
Example(s)	Average, Exact, Numeric Value
Conformance	CONDITIONAL
Data Type	PQ
Values Allowed	Single quantitative values or an average value for a given element.

4.9.2 Low limit

It should be noted that there can be two types of low limit, one is a limit on actual values and the other is the lower limit of the average of values. Both these limits can be captured and type of lower limit should be indicated either in the amount type for a given element or in the units indicated for the amount. The low limit can be used alone to express one-sided ranges greater than or equal to this limit.

Table 60 — Low limit

User Guidance	Lower limit for a given quantitative value.
Conformance	CONDITIONAL
Data Type	PQ
Business Rule(s)	To be used to express the lower limit of a range of values, if applicable