BRITISH STANDARD

BS ISO 25720:2009

# Health informatics — Genomic Sequence Variation Markup Language (GSVML)

ICS 35.240.80

NO COPYING WITHOUT BSI PERMISSION EXCEPT AS PERMITTED BY COPYRIGHT LAW

British Standards Documents Document Sharing Hub

Full document details: BS ISO 25720:2009

# National foreword

This British Standard is the UK implementation of ISO 25720:2009.

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.

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

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 October 2009

© BSI 2009

**ISBN 978 0 580 57423 8**

**Amendments/corrigenda issued since publication**

| Date | Comments |
|---|---|
| | |
| | |
| | |
| | |

Documents
Document Sharing Hub

Full document details: BS ISO 25720:2009

# INTERNATIONAL STANDARD

# ISO 25720

First edition
2009-08-15

## Health informatics — Genomic Sequence Variation Markup Language (GSVML)

*Informatique de santé — Langage de balisage de la variation de séquence génomique*

Reference number
ISO 25720:2009(E)

Documents
Document Sharing Hub

© ISO 2009

Full document details: BS ISO 25720:2009

**PDF disclaimer**

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2009

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

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

Published in Switzerland



© ISO 2009 – All rights reserved

Documents
Document Sharing Hub

Full document details: BS ISO 25720:2009

# Contents

Page

**Foreword** ..... iv
**Introduction** ..... v
**1 Scope** ..... 1
**2 Conformance** ..... 1
**2.1 Purpose** ..... 1
**2.2 Needs and general requirements** ..... 1
**3 Normative references** ..... 2
**4 Terms and definitions** ..... 2
**5 GSVML specification** ..... 5
**5.1 Specification requirements and GSVML positioning** ..... 5
**5.2 GSVML structure** ..... 5
**5.3 GSVML DTD and XML schema** ..... 5
**6 GSVML development process** ..... 5
**Annex A** (normative) **DTD of GSVML** ..... 22
**Annex B** (normative) **XML schema of GSVML** ..... 46
**Annex C** (informative) **Basic reference works** ..... 105
**Bibliography** ..... 131

© ISO 2009 – All rights reserved


Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

## Foreword

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

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

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

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

ISO 25720 was prepared by Technical Committee ISO/TC 215, *Health informatics*.


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

## Introduction

In the current electronic world, there are multiple different types of data for healthcare, as shown in Figure 1. Besides clinical data and image data, as we move into this post genomic era, we are creating, internationally, overwhelming amounts of genomic data. The International Standards developing organizations are developing standards for these data; Health Level Seven develops standards for clinical data, DICOM and JPEG develop standards for image data. Genomic Sequence Variation Markup Language (GSVML) defines a standard for genomic data, especially human-related DNA variation data. The core target for the GSVML is the Single Nucleotide Polymorphism (SNP).

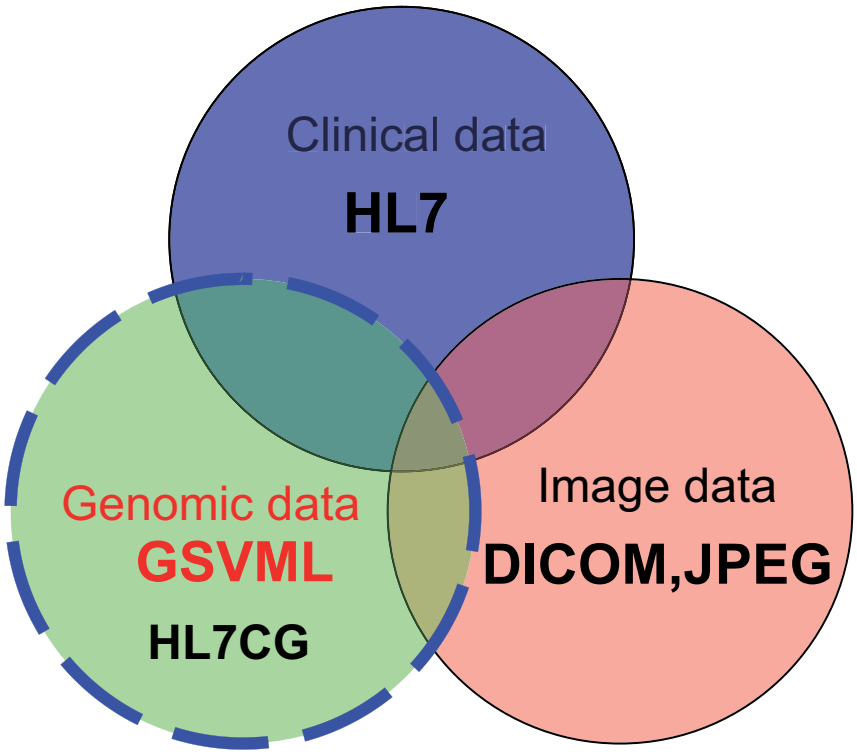



**Figure 1 — Major data types of health care**

In this post genomic era, the management of health-related data is becoming increasingly important to both genomic research and genome-based medicine (see reference [1]). Informational approaches to the management of clinical, image and genomic data are beginning to have as much worth as basic, bench top research. Nowadays there are many kinds of -omics data around the world awaiting effective utilization for human health. The hurdle that must be overcome to achieve this goal is the development of data format and message standards to support the interchange of -omics data. Genomic data include genome sequence, DNA sequence variation and other genome-based data such as expression data, proteomics data, molecular network, etc. As an entry point, this International Standard focuses on the DNA sequence variation. Among the DNA sequence variation, the SNP is selected as the core object because of the following three reasons.

a) SNP is the most researched sequence variation for human health.

b) In the current context, vast amounts of SNP data exist around the world in various types of data formats. As a result of the recent explosion in SNP research, the vast amounts of experimental data have been

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

accumulating in many databases in various types of data formats. These data await utilization in drug discovery, clinical diagnosis and clinical research.

c) SNP data already have a great impact for human applications such as gene-based medicine and pharmacogenomics.

With a view to this context, the international community requires an interoperable format for the interchange of SNP data. Prior to the standardization development, we elucidated the need for data exchange among the human health-related facilities that have various types of data formats.

In the present circumstances, SNP is expected to be a key to understanding human response to external stimuli such as any kind of alien invasions, therapies, and the environmental interactions (see reference [2]). Bacterial infection is an example of alien invasion, and the responses to the infections are different amongst individuals. According to the therapy, the side effects to a drug are different amongst the patients. These responses are also different in various environments.

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

Markup Language for the biomedical field, based on XML, has been in development for several decades to enhance the exchange data among researchers. Bioinformatic Sequence Markup Language (BSML) (see reference [11]), Systems Biology Markup Language (SBML) [12], Cell Markup Language (Cell ML) [13], and Neuro Markup Language (Neuro-ML) [14] are examples of markup languages. Polymorphism Mining and Annotation Programs (PolyMAPr) [15] is centric on SNP and tries to achieve mining, annotation and functional analysis of public databases such as dbSNP [16], the Cancer Gene Anatomy Project (CGAP) (see reference [17]), and Japanese single nucleotide polymorphisms (JSNP) (see reference [18]) through programming.

To utilize the accumulated SNP data among many facilities around the world, standards for the interchange of SNP data must be defined. The required standards include defining a data format and exchange messages. Markup Language is the reasonable choice to address this need. As for genomic data message handling, Health Level Seven Clinical Genomics Special Interest Group [19] has summarized clinical use cases for general genomic data. The GSVML project has contributed to these efforts. Additionally, this work incorporated use cases based on the Japanese Millennium Project [20]. Based on these contexts and investigations, this International Standard elucidates the needs and the requirements for GSVML and then proposes the specification of GSVML for the international standardization.


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

INTERNATIONAL STANDARD ISO 25720:2009(E)

# Health informatics — Genomic Sequence Variation Markup Language (GSVML)

**IMPORTANT — The electronic file of this document contains colours which are considered to be useful for the correct understanding of the document. Users should therefore consider printing this document using a colour printer.**

## 1 Scope

This International Standard is applicable to the data exchange format that is designed to facilitate the exchange of the genomic sequence variation data around the world, without forcing change of any database schema. From an informatics perspective, GSVML defines the data exchange format based on XML. The scope of this International Standard is the data exchange format, but the database schema itself is outside the scope of this International Standard. From a biological point of view, all genetic sequence variations are taken into consideration and are within the scope of this International Standard, while polymorphisms, especially SNPs, are the main focus of this International Standard. In other words, the annotations of variation as clinical concerns and -omics concerns are within the scope of this International Standard. Though SNPs exist in various biological species, the scope of this International Standard covers the human health associated species as human, cell line, and preclinical animals. The other biological species are outside the scope of this International Standard. The clinical field is within the scope of this International Standard, but the basic research fields and other scientific fields are outside the scope of this International Standard. Here, clinical research, including drug discovery, is within the scope of this International Standard. As for supposed application fields, the main focus is in human health, including clinical practice, preventive medicine, translational research and clinical researches.

## 2 Conformance

### 2.1 Purpose

This International Standard provides a data exchange format for genomic sequence variation data in human health. This International Standard provides the GSVML specification mainly for the case of SNP and Short Tandem Repeat Polymorphism (STRP). Considering that SNP and STRP are the major and simple polymorphisms in human health research, centering on them and expanding the specification to the other sequence variation data seems reasonable. This International Standard allows for the expandability of GSVML from SNP and STRP to other sequence variation data.

### 2.2 Needs and general requirements

The vast volume of experimental data from the recent explosion of genomic sequence variation research has produced an overwhelming amount of data stored in many databases with various types of format worldwide. Standardization of data exchange is urgent for managing, analysing, and utilizing these data. Standardizing the interoperable format is necessary for easy and convenient genomic sequence variation data exchange. Considering that genomic sequence variation, especially SNP and STRP, has its significant meaning in the gene-based medicine and the pharmacogenomics for human health, the data exchange format is the key to enhancing the gene-based clinical research and the gene-based medicine.

The management of genomic data is as critical as the basic research data in this new era. There are many kinds of -omics data around the world, and the time has come to effectively use these genomic data for human health. In order to use these data effectively and efficiently, standards must be developed to permit the interoperable interchange of genomic data globally. These standards must define the data format as well as

© ISO 2009 – All rights reserved



Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

the messages to be used to interchange and share this data globally. This International Standard addresses those requirements, using a Markup Language.

## 3 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.

EN 13606 (all parts), *Health informatics — Electronic healthcare record communication*

## 4 Terms and definitions

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

**4.1**
**actor**
something or someone who supplies a stimulus to the system

NOTE Actors include both humans and other quasi-autonomous things, such as machines, computer tasks and systems.

**4.2**
**Bioinformatic Sequence Markup Language**
**BSML**
extensible language specification and container for bioinformatic data

**4.3**
**Cell Markup Language**
**Cell ML**
Extensible Markup Language to provide a standard method for representing and exchanging computer-based biological models

**4.4**
**Cancer Gene Anatomy Project**
**CGAP**
database containing genomic expression data collected for various tumorgenic tissues in both humans and mice and also providing information on methods and reagents used in deriving the genomic data

**4.5**
**dbSNP**
database of SNPs provided by the US National Center for Biotechnology Information (NCBI)

**4.6**
**Digital Imaging and Communications in Medicine**
**DICOM**
standard in the field of medical informatics for exchanging digital information between medical imaging equipment (such as radiological imaging) and other systems, ensuring interoperability

**4.7**
**deoxyribonucleic acid**
**DNA**
molecule that encodes genetic information in the nucleus of cells

**4.8**
**DNA sequence variation**
differences of DNA sequence among individuals in a population

NOTE DNA sequence variation implies **polymorphism** (4.20).


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

**4.9**
**Document Type Definition**
**DTD**
separate document that contains formal definitions of all of the data elements in a particular type of HTML, SGML or XML document

**4.10**
**entry point**
reference point that designate the class(es) from which the messages begin for the particular domain

**4.11**
**gene-based medicine**
medicine based on genes or genetic science

**4.12**
**Hypertext Markup Language**
**HTML**
set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser page

**4.13**
**Joint Photographic Experts Group**
**JPEG**
compression technique for images

**4.14**
**Japanese single nucleotide polymorphisms**
**JSNP**
database of Japanese single nucleotide polymorphisms

**4.15**
**markup language**
ML
set of symbols and rules for their uses when doing a markup of a document

**4.16**
**Neuro Markup Language**
**Neuro-ML**
markup language for describing models of neurons and networks of neurons

**4.17**
**nroff**
unix text-formatting program that is a predecessor of the Unix troff document processing system

**4.18**
**pharmacogenomics**
branch of pharmaceutics aiming to develop rational means to optimize drug therapy, with respect to the patient's genotype

**4.19**
**Polymorphism Mining and Annotation Programs**
**PolyMAPr**
programs for polymorphism database mining, annotation and functional analysis

**4.20**
**polymorphism**
variation in the sequence of DNA among individuals

NOTE Polymorphism implies **SNP** (4.23) and **STRP** (4.26).

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

**4.21**
**Systems Biology Markup Language**
**SBML**
markup language for simulations in systems biology

**4.22**
**Standard Generalized Markup Language**
**SGML**
standard for defining description of the structure of different types of electronic documents

**4.23**
**Single Nucleotide Polymorphism**
**SNP**
single nucleotide variation in a genetic sequence that occurs at appreciable frequency in the population

**4.24**
**Systematized Nomenclature of Medicine – Clinical Terms**
**SNOMED CT**
dynamic, scientifically validated clinical health care terminology and infrastructure

**4.25**
**Simple Object Access Protocol**
**SOAP**
lightweight protocol for exchange of information in a decentralized, distributed environment

**4.26**
**Short Tandem Repeat Polymorphism**
**STRP**
variable segments of DNA that are two to five bases long with numerous repeats

**4.27**
**troff**
document processing system developed by AT&T for the Unix operating system

**4.28**
**variable number of tandem repeat**
**VNTR**
class of polymorphism characterized by the highly variable copy number of identical or closely related sequences

**4.29**
**Wireless Markup Language**
**WML**
XML language used to specify content and user interface for WAP (wireless application protocol) devices

**4.30**
**Extensible HTML**
**XHTML**
hybrid between HTML and XML specifically designed for net device displays

**4.31**
**Extensible Markup Language**
**XML**
pared-down version of SGML, designed especially for web documents

**4.32**
**XML schema**
language for describing the structure and constraining the contents of XML documents


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

# 5 GSVML specification

## 5.1 Specification requirements and GSVML positioning

In the current context, annotative information about genomic sequence variation is increasing, which is filling in the gaps in information. The genomic sequence variation data themselves are also increasing but are stored in various databases. This trend is typical of SNP data. The pitfall of genomic sequence variation data handling is the lack of standardization of the data formats for the genomic sequence variation. Historically, the markup languages listed in Clause 4 have been used, and programs are developed to handle the genomic information. However, there have been no genomic sequence variation centric markup languages so far. GSVML is the first genomic sequence variation centric markup language and is human health centric. Considering that SNP is a highly researched polymorphism and has a great impact, especially for human health and response, we can say that GSVML has the greatest potential to be the designated markup language for human healthcare. On the other hand, setting the applications to practical human health means it must handle direct or indirect SNP annotations. Here the direct SNP annotation indicates general annotative information such as SNP associated genes and experimental preparations. The indirect SNP annotation indicates all of the -omics data and clinical data that result from SNP variation. To understand the gene-based clinical situation of each patient, we need this kind of additional information. Considering the requirement to add many kinds of additional information, the development and standardization of GSVML cannot stand alone and need harmonization with the other international standardization organizations such as Health Level Seven.

GSVML is intended to be used in data exchange messages related to human health. In the development and standardization of GSVML in this application domain, we must always keep an eye on the patient's safety, clinical efficiency and medical costs. For the patient's safety, from an informational viewpoint, the conservation and the protection of patient information are important. For the enhancement of clinical efficiency, simplicity and ease of understanding are important. For medical cost reduction, the adaptation ability and ease of installation are important. GSVML tries to respond to these basic requirements by providing the sharable XML based data exchanging format. GSVML can be used for the clinically genomic sequence variation data exchange among various types of data formats. In the greater framework of clinical data standardization, GSVML plays the part of describing the genomic sequence variation data and their necessary information.

## 5.2 GSVML structure

The outlined structure of GSVML is shown in Figure 3. GSVML consists of three data criteria, *viz.* variation data, direct annotation and indirect annotation. The variation data criterion describes the straightforward variation data as allele, type, position, length, region, etc. The direct annotation criterion describes the attached data of variation data as experiment analysis, epidemiology or associated gene, etc. The indirect annotation criterion describes the explanatory/higher-level information of variation data such as the -omics data, the clinical information and the environmental data. These data criteria have internal relations to each other. The detailed structure of GSVML is shown in Figures 4 to 21.

## 5.3 GSVML DTD and XML schema

The DTD of GSVML is shown in Annex A. The XML schema of GSVML is shown in Annex B.

# 6 GSVML development process

The development of GSVML followed eight steps:

— Step 1: Set the elements and needs according to the investigated use cases.

We prepared six use cases for three typical criteria. Four use cases concerned the clinical practice, and one use case for each clinical trial and translational research.

— Step 2: Construct the basic structure and DTD.

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

— Step 3: Investigate the existing biological ML, and its applicability to the needs (comparison with MAGE-ML, BSML, SBML, RNAML[21], ProML, CellML, PolyMAPr).

— Step 4: Refine the basic structure and DTD, construct the XML Schema (XSD).

— Step 5: Investigate the existing SNP databases (their data format comparison).

— Step 6: Check the interface ability to the Health Level Seven Genotype Model.

— Step 7: Redefine the needs of GSVML and its demanded elements.

— Step 8: Refine the basic structure, DTD, and XML Schema.

Figure 2 shows the outline of the process of the development. We did design work in harmony with HL7 Clinical Genomics SIG. There were “to and fro” processes between design work and the standardization process.

Additionally, we analysed the interface between GSVML and EN 13606, SNOMED-CT.

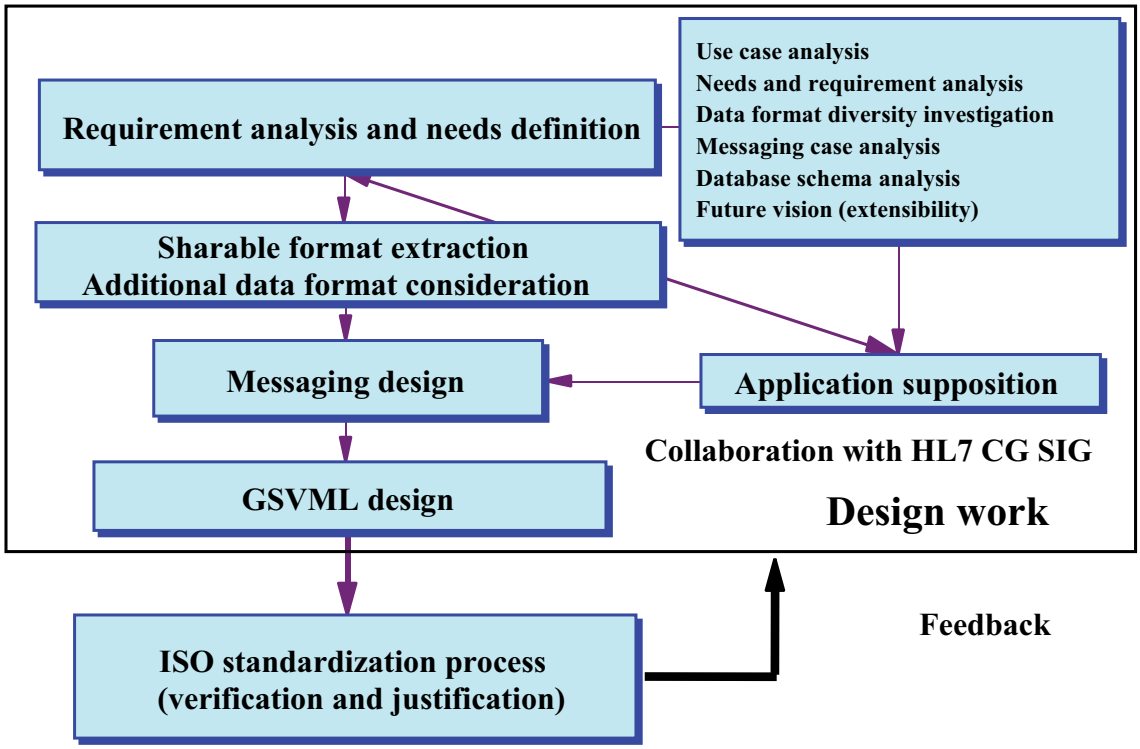
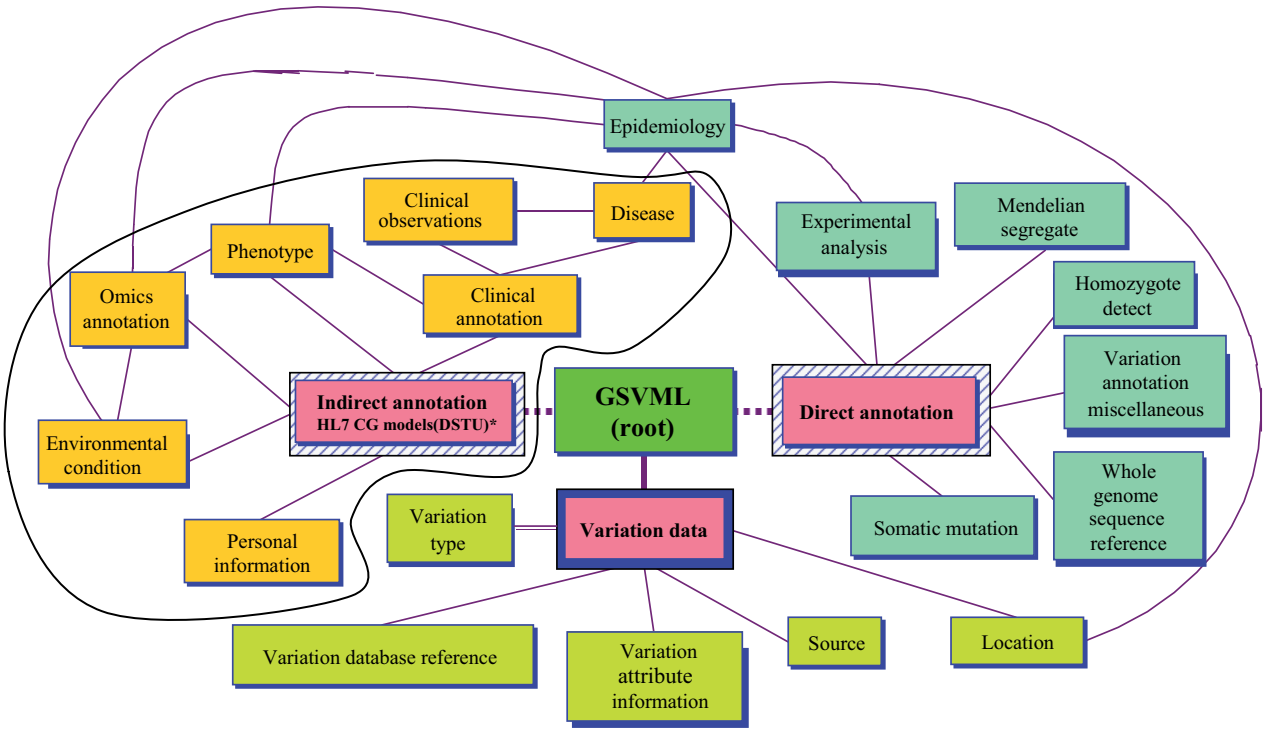



**Figure 2 — Outline of the process of GSVML development**


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009




* HL7 CG models (DSTU) will be used instead of indirect annotation criterion.

**Figure 3 — The outlined structure of GSVML**

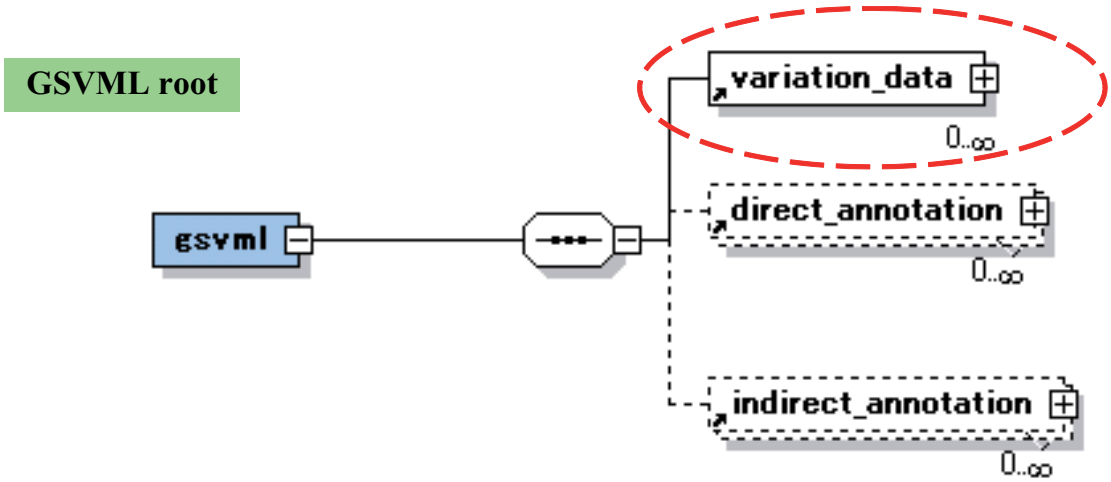



**Figure 4 — Detailed structure of GSVML**

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

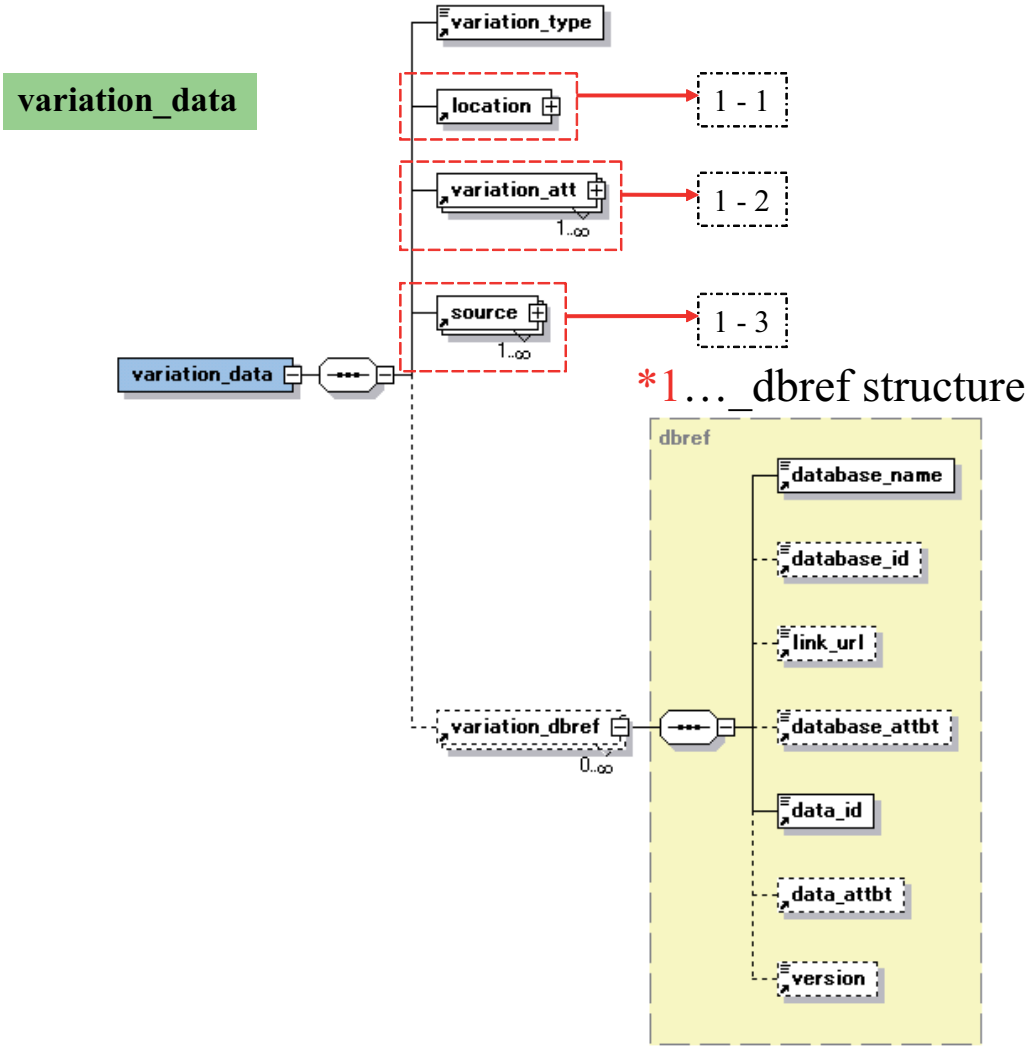



**Figure 5 — Detailed structure of GSVML**



© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

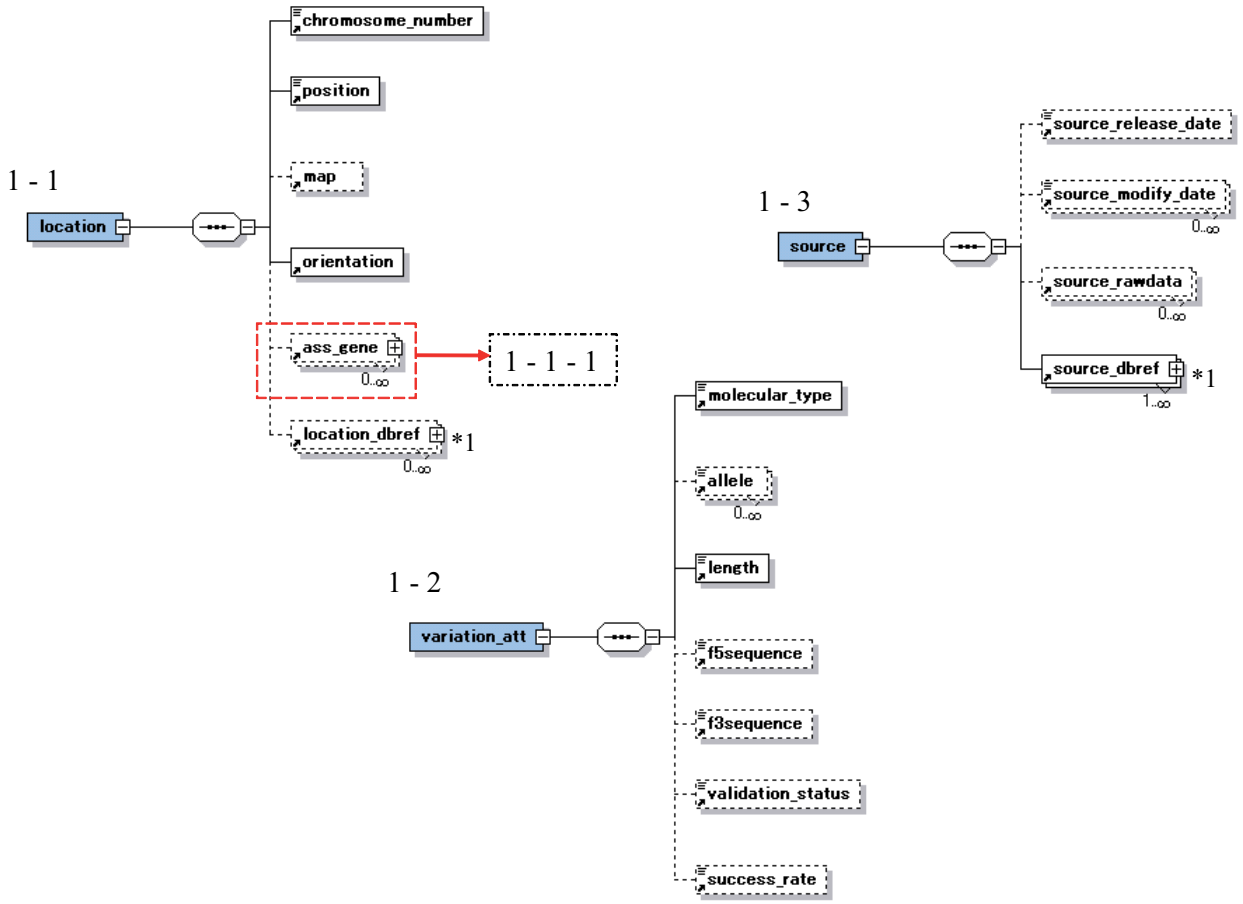



**Figure 6 — Detailed structure of GSVML**

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

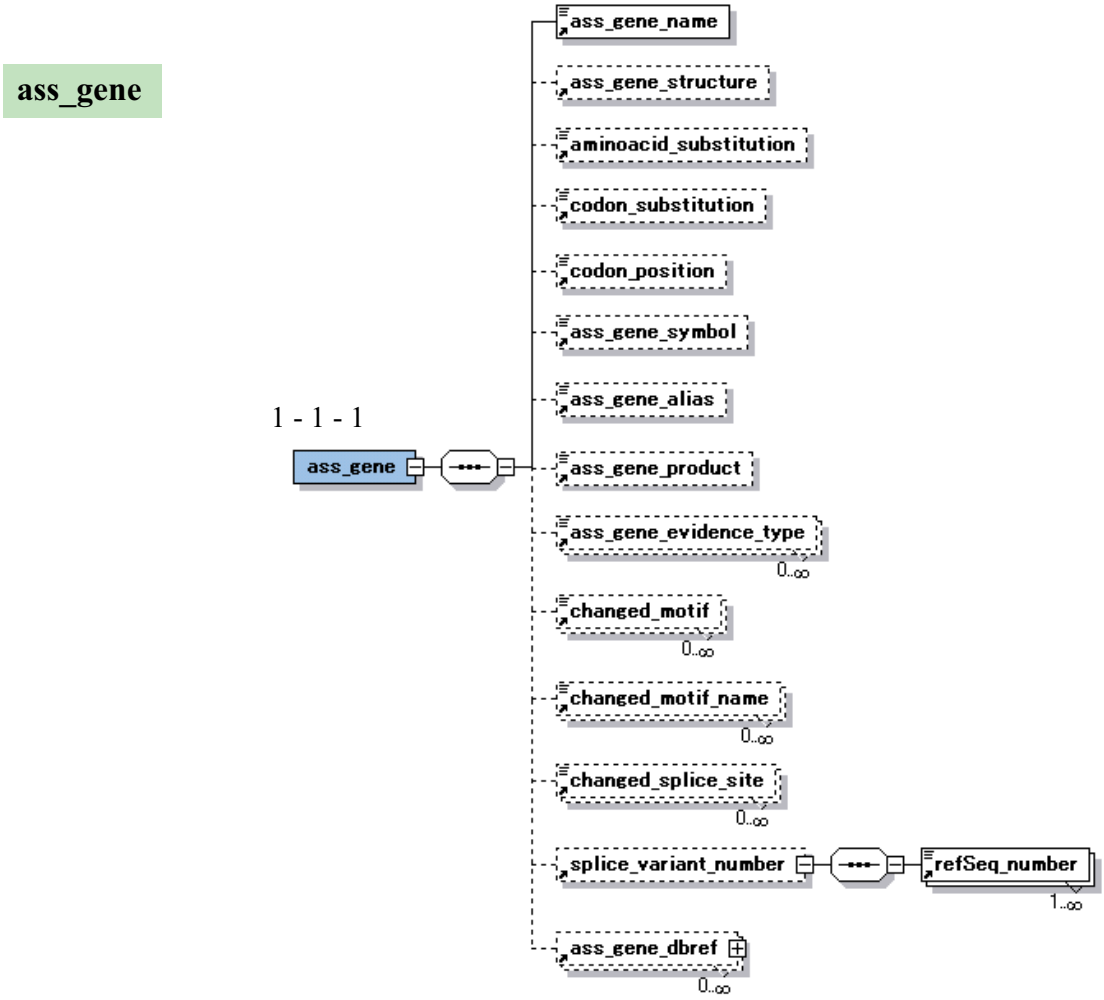



**Figure 7 — Detailed structure of GSVML**

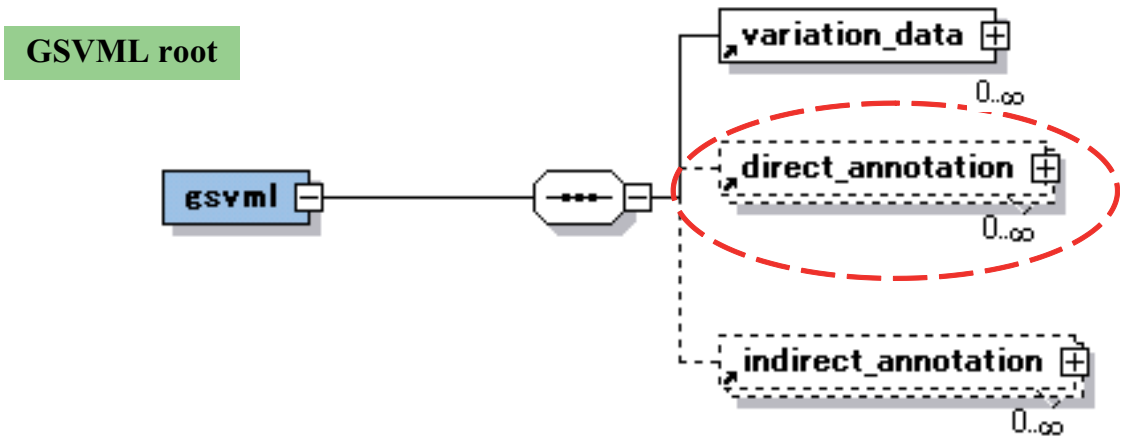



**Figure 8 — Detailed structure of GSVML**



© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

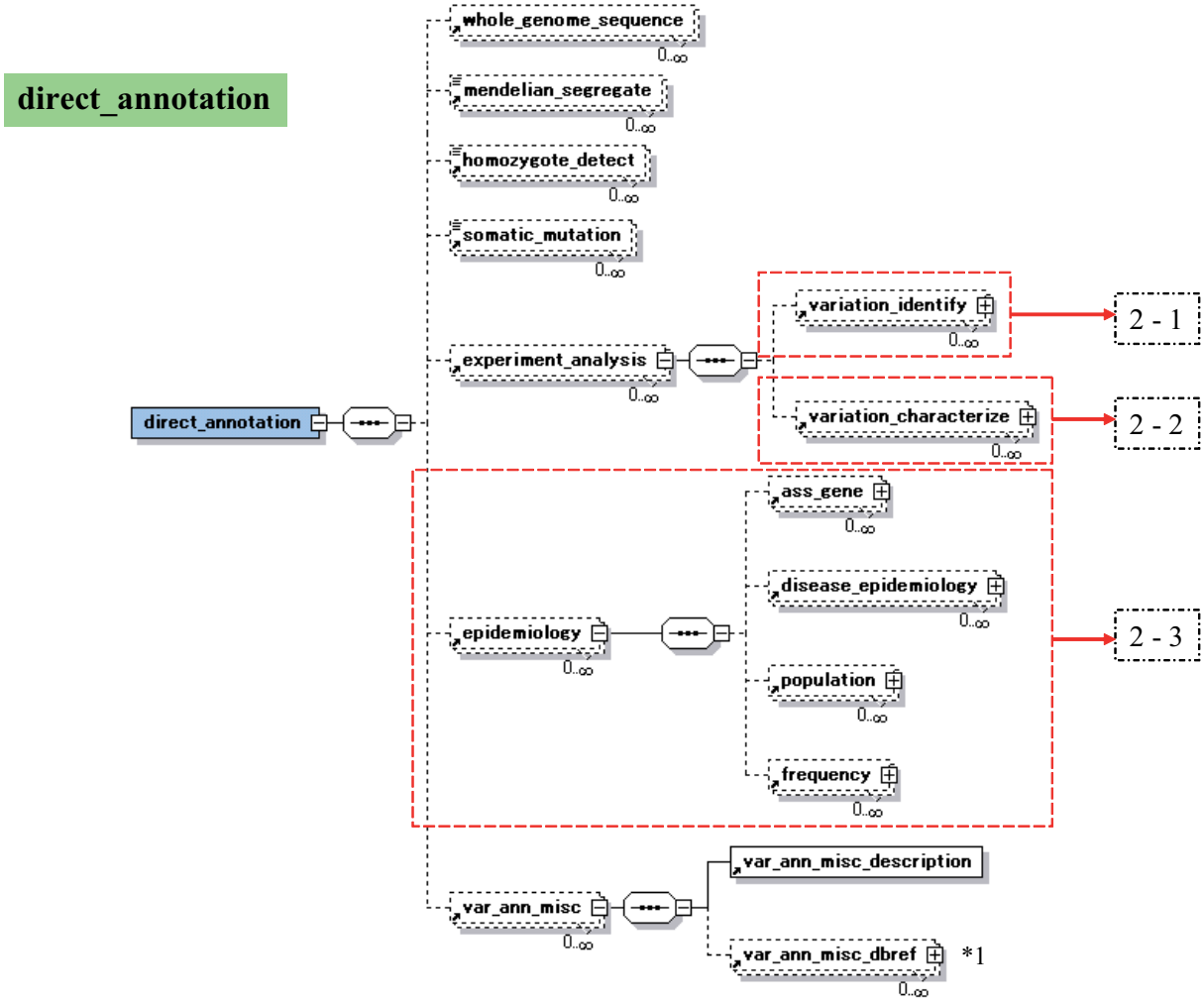



**Figure 9 — Detailed structure of GSVML**

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

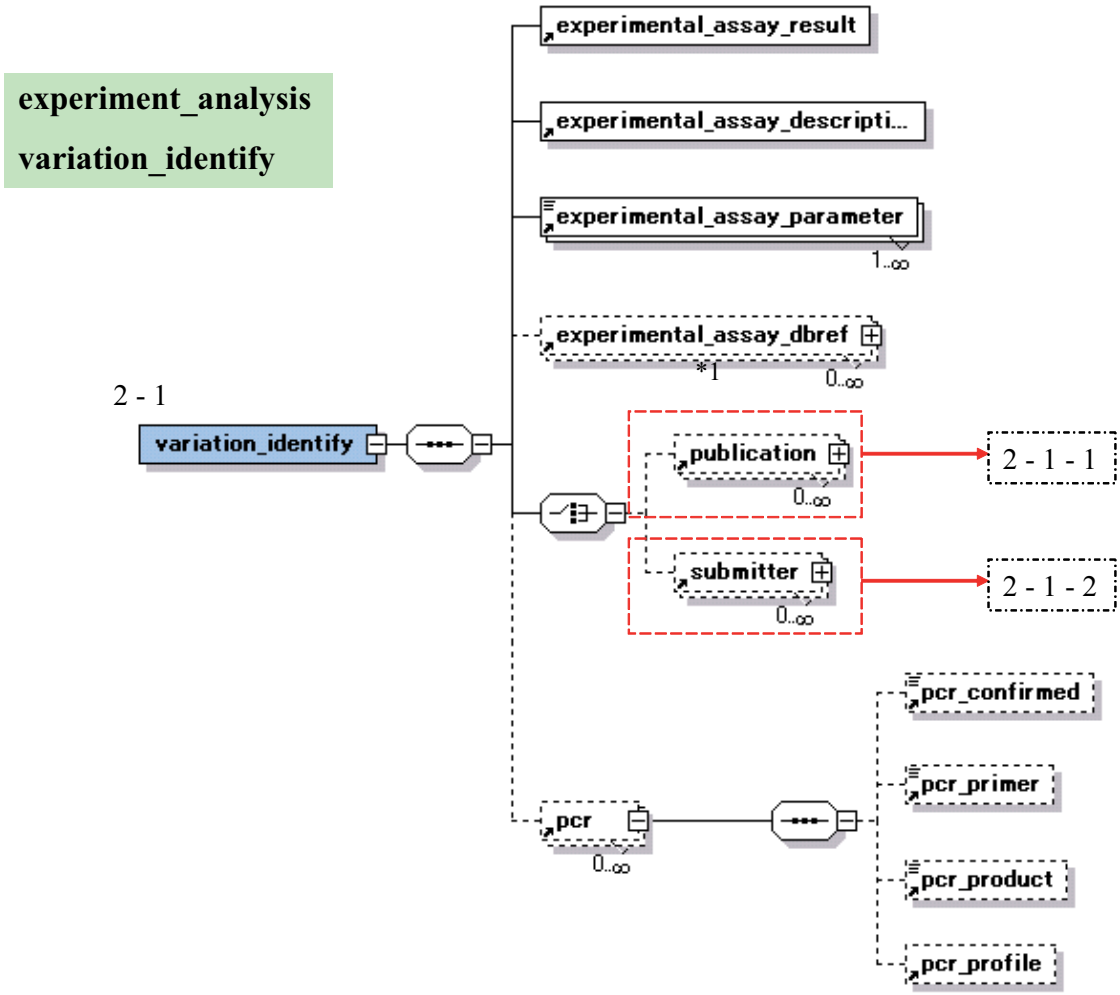



**Figure 10 — Detailed structure of GSVML**



© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

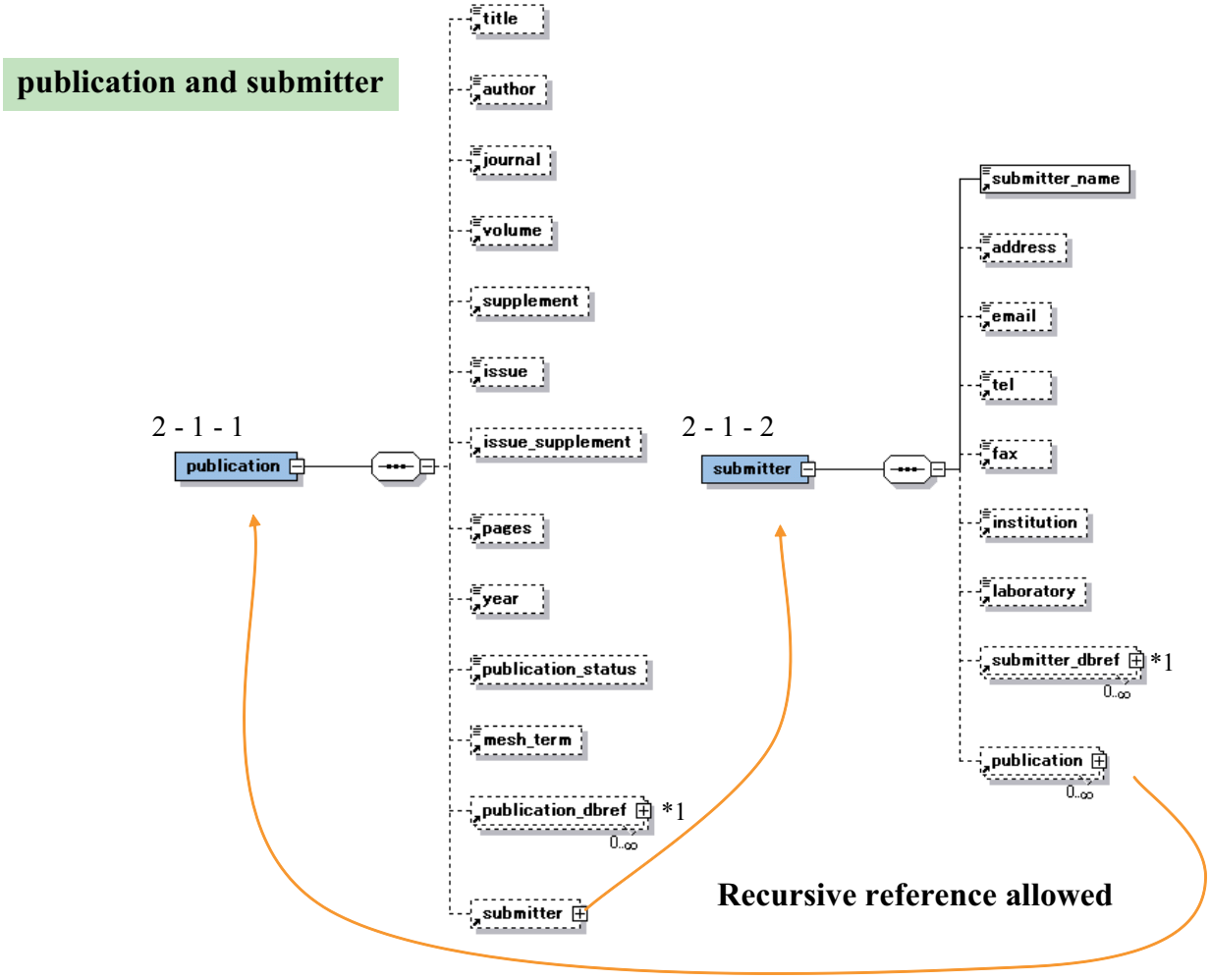



**Figure 11 — Detailed structure of GSVML**

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

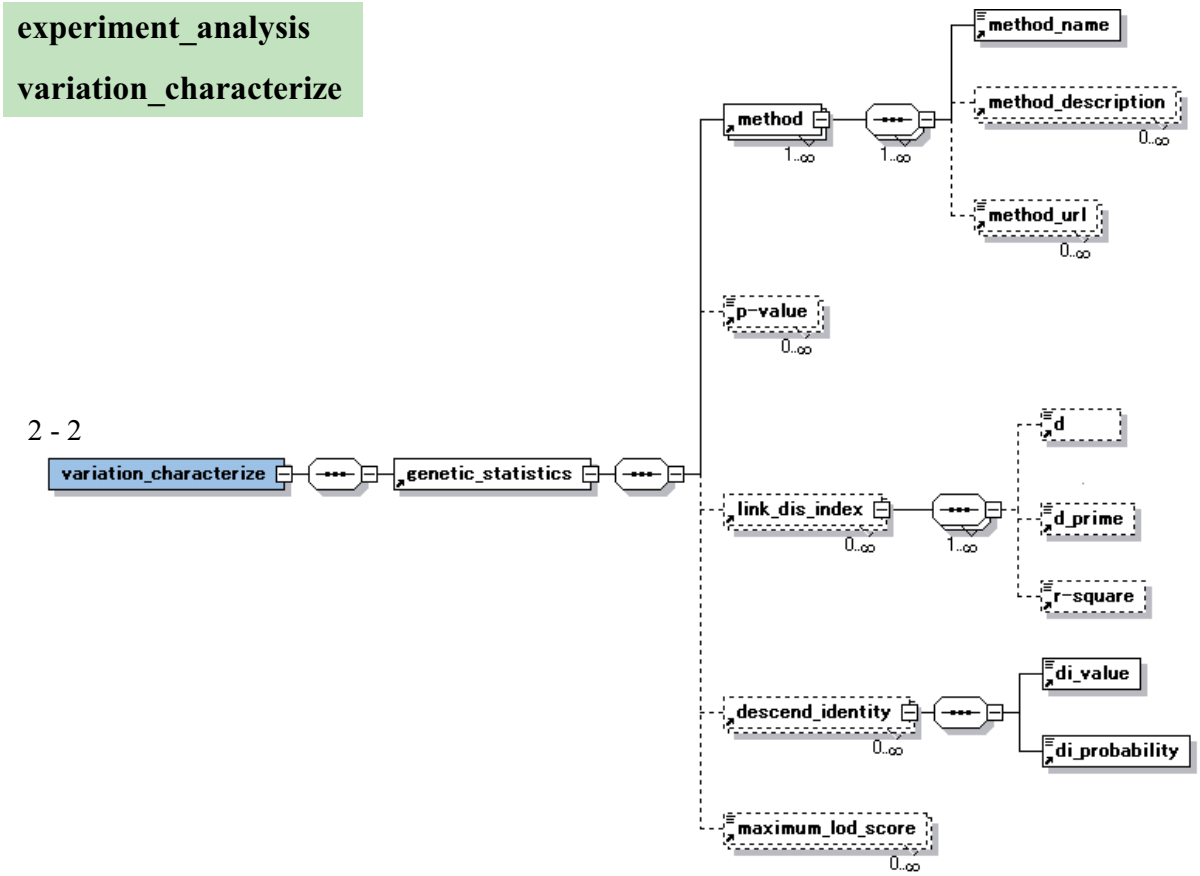



**Figure 12 — Detailed structure of GSVML**


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

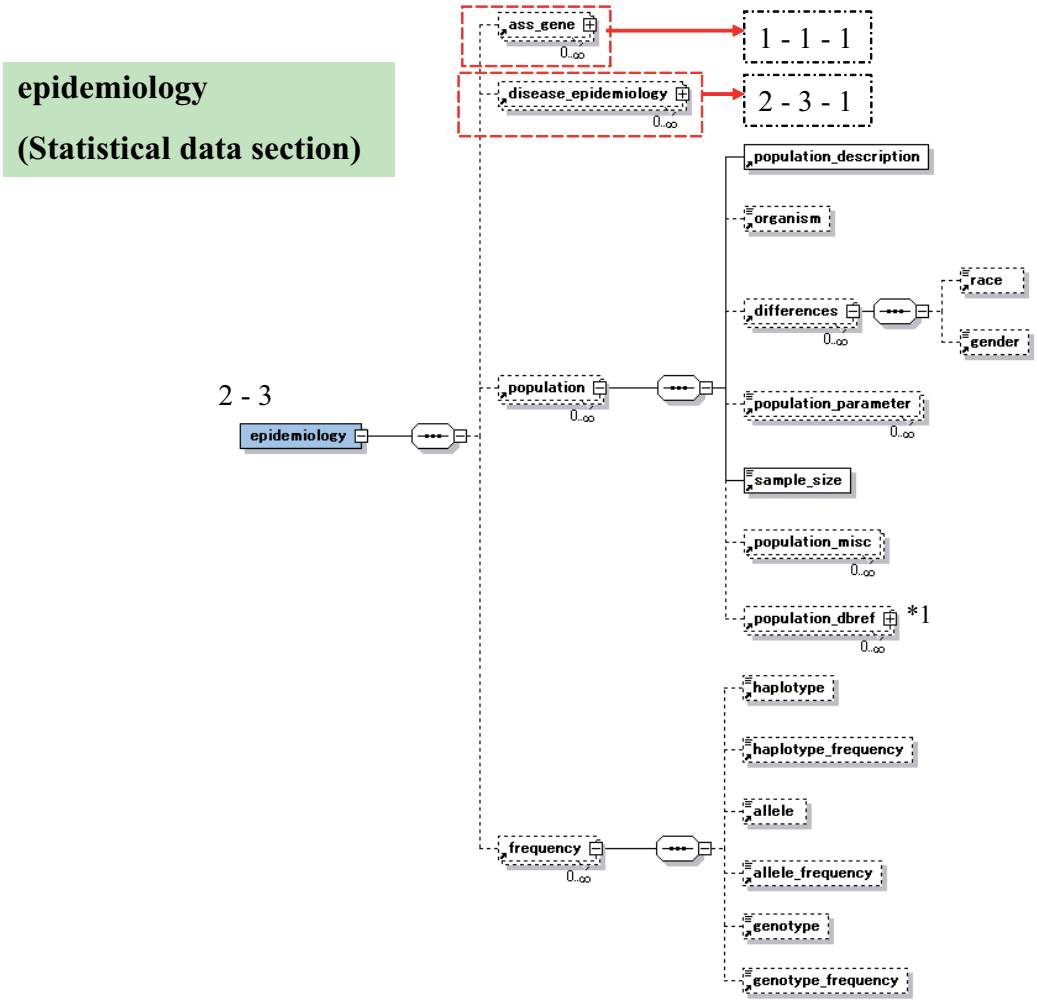



**Figure 13 — Detailed structure of GSVML**

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

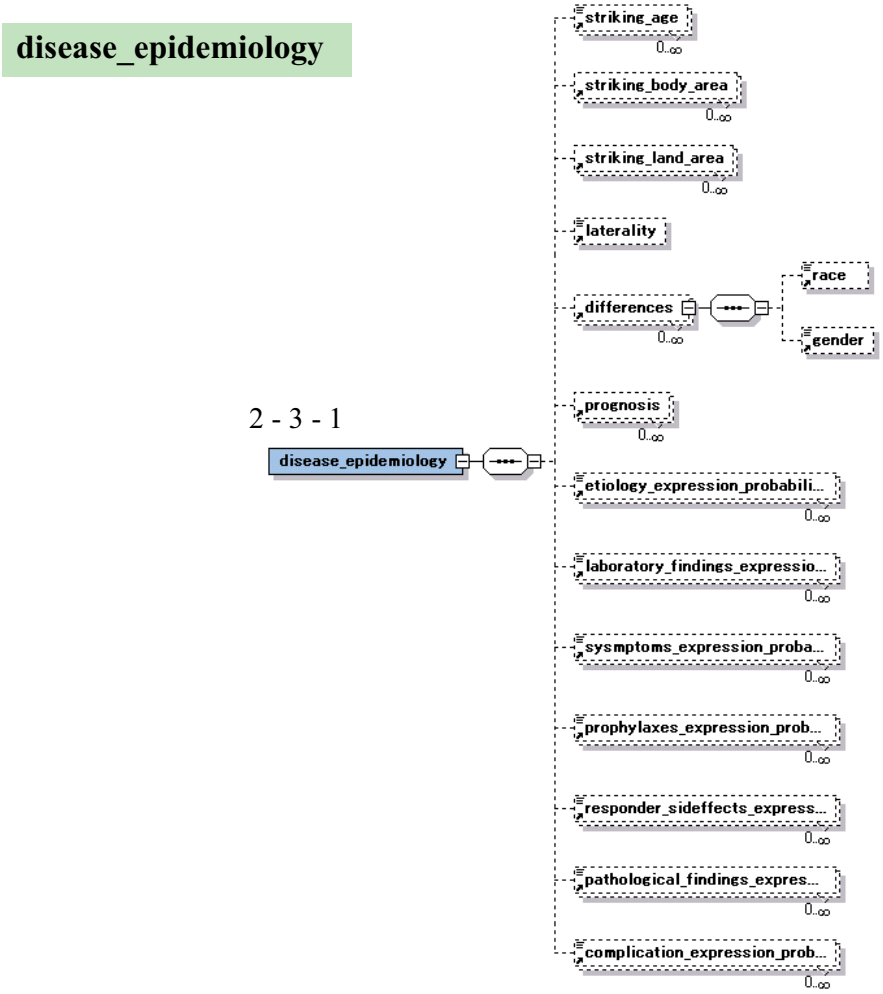



**Figure 14 — Detailed structure of GSVML**

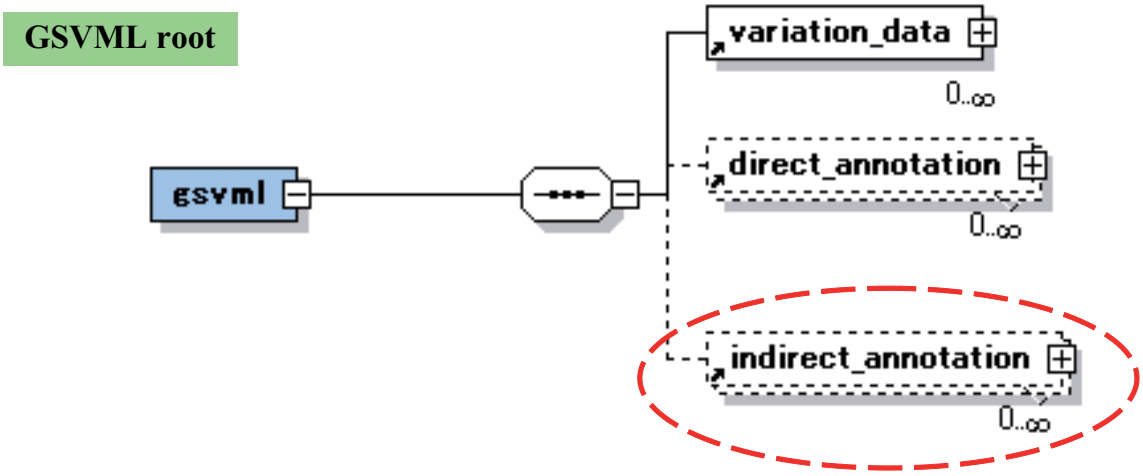



**Figure 15 — Detailed structure of GSVML**



© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

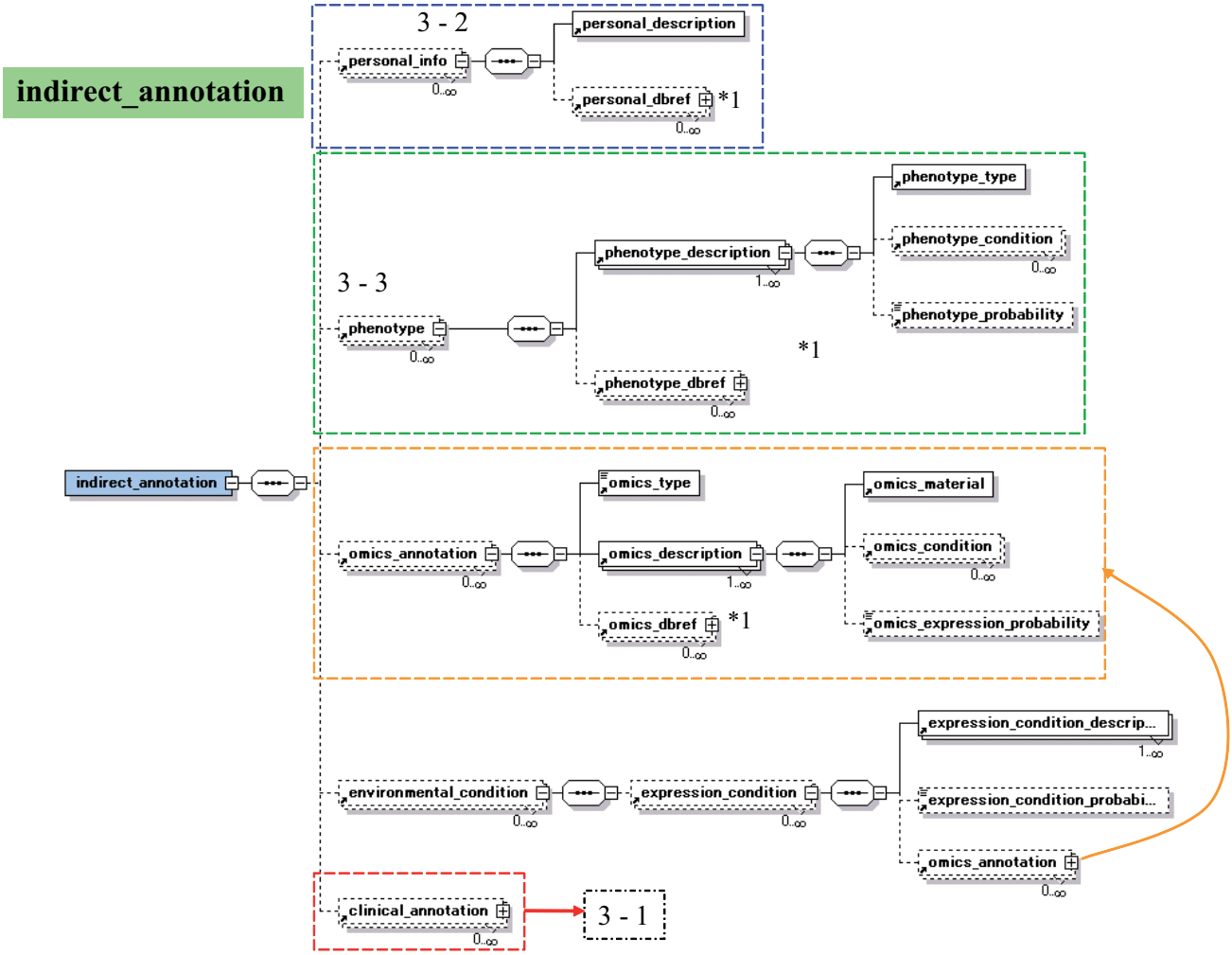



**Figure 16 — Detailed structure of GSVML**

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

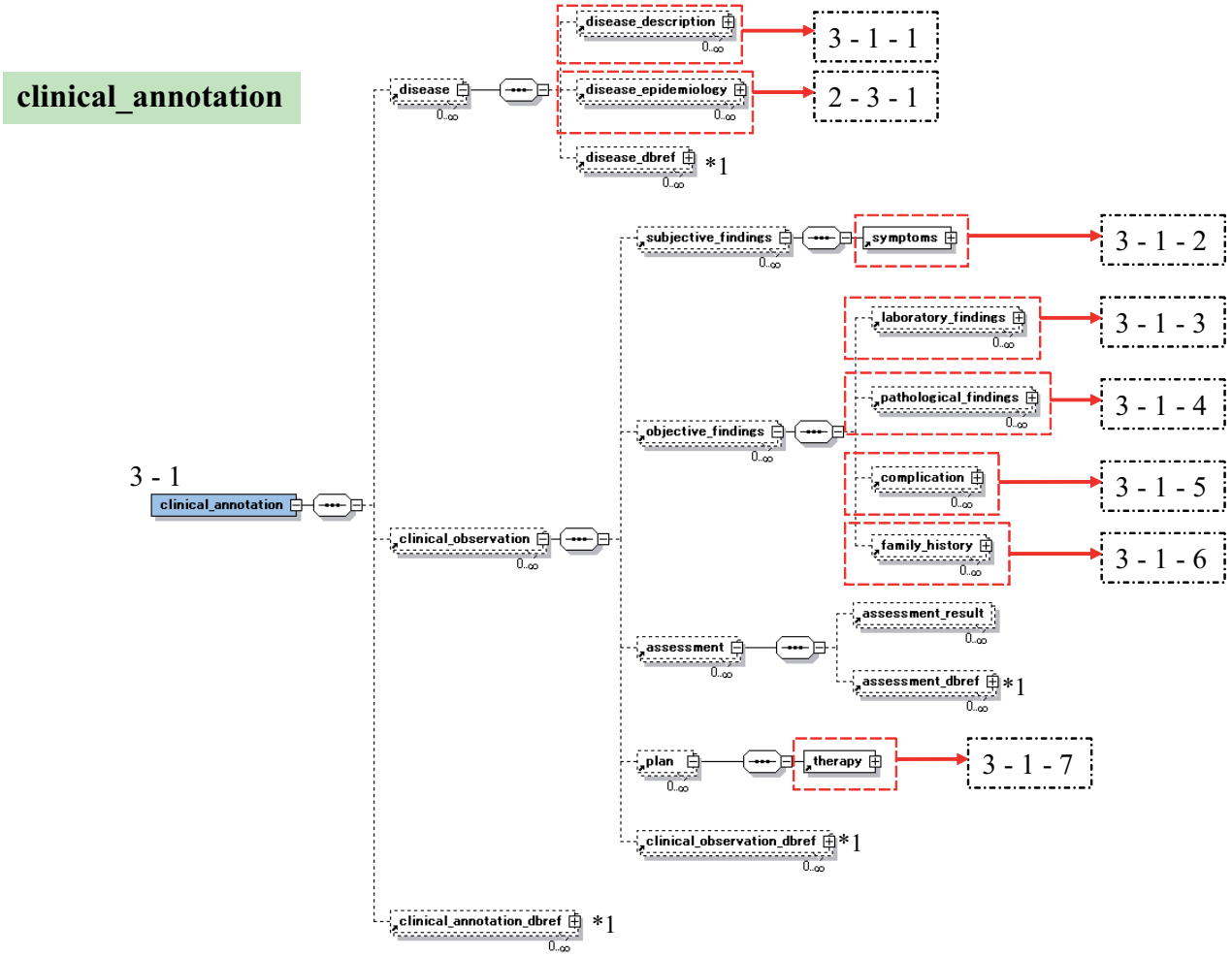



**Figure 17 — Detailed structure of GSVML**



© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

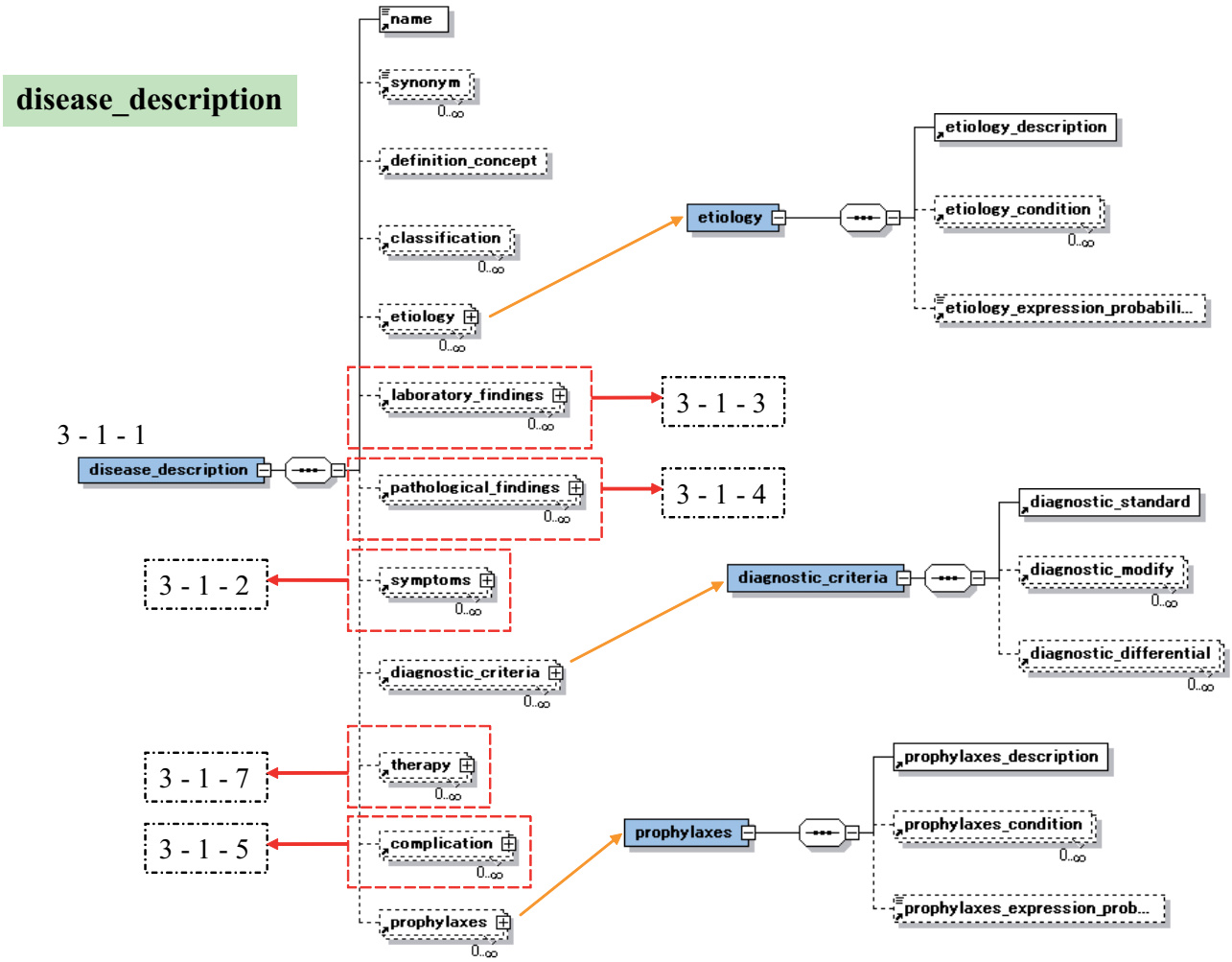



**Figure 18 — Detailed structure of GSVML**

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

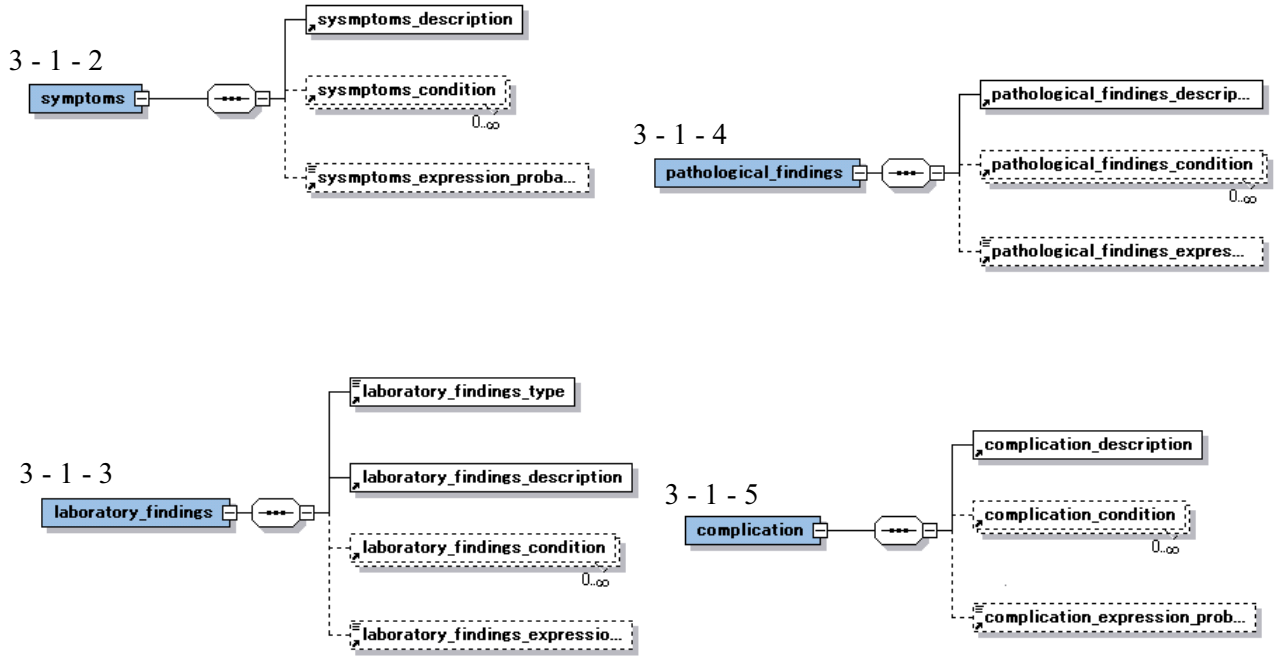



**Figure 19 — Detailed structure of GSVML**

**family_history**

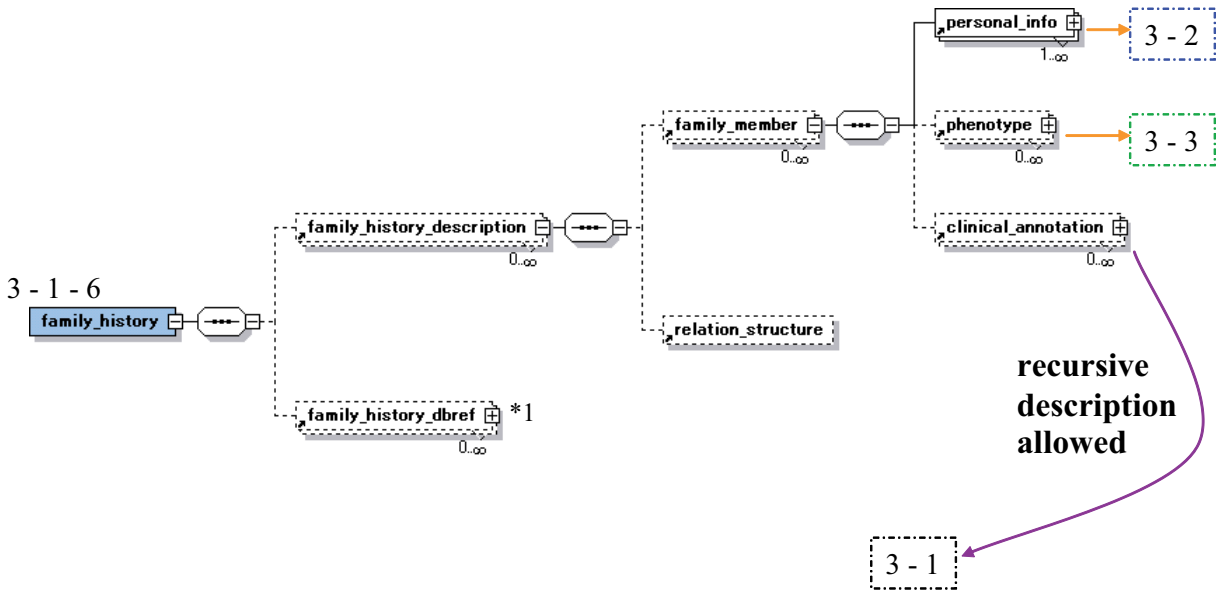



**Figure 20 — Detailed structure of GSVML**



© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

**therapy**

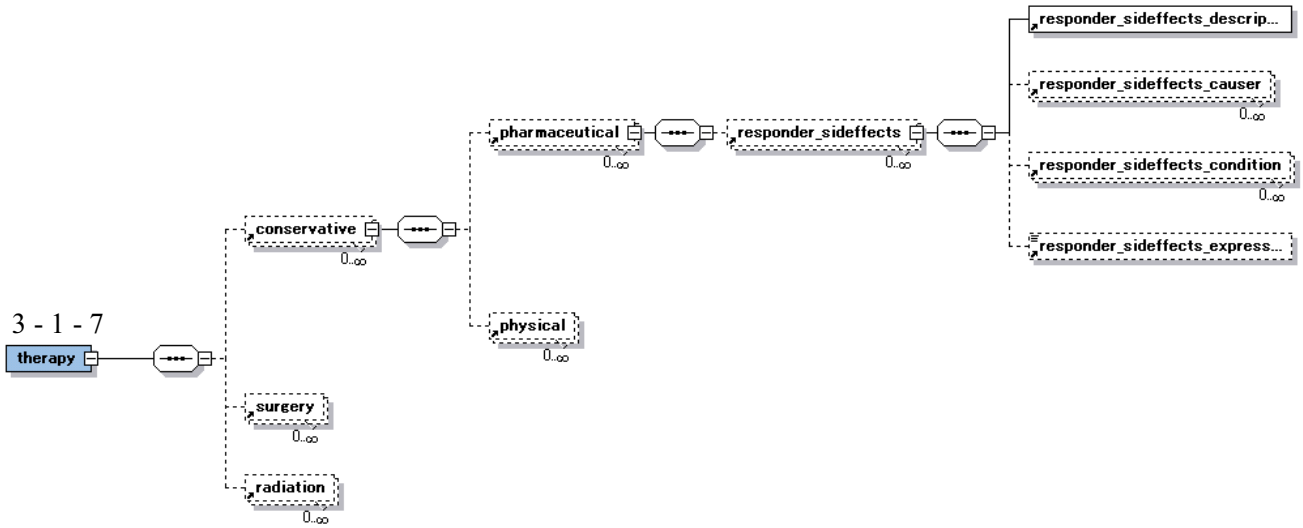



**Figure 21 — Detailed structure of GSVML**

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

# Annex A
(normative)

# DTD of GSVML

```
<?xml version='1.0' encoding='Shift_JIS' ?>

<!-- comment

   Jun Nakaya ( Information Center for Medical Sciences, Tokyo Medical and Dental University, JAPAN)

"gsvml element:  Genomic Sequence Variation Markup Language

   variation_data:  variation data

   direct_annotation:  direct annotation of variation data

   indirect_annotation:  indirect annotation of variation data

"assessment element

   assessment_result:  results of the assessment

   assessment_dbref:  database reference of the assessment

"ass_gene element

   ass_gene_name:  gene name

   ass_gene_structure:  category of gene structure e.g. exon, intron

   aminoacid_substitution:  aminoacid sustitution generated by variation

   codon_substitution:  codon substitution generated by variation

   codon_position:  codon position

   ass_gene_symbol:  gene symbol

   ass_gene_alias:  gene alias

   ass_gene_product:  gene product

   ass_gene_evidence_type:  gene type e.g. functional gene, predicted EST, computational gene, Pseudogene

   changed_motif:  motif change exists or not

   changed_motif_name:  name of motif

   changed_splice_site:  splice site change exist or not

   splice_variant_number:  number of splice variant and refSeq

   ass_gene_dbref:  database reference information
```


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

"ass_gene_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"clinical_annotation element

disease: disease information

"clinical_annotation dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"clinical observation element

subjective_findings: subjective findings in clinical observation

objective_findings: objective findings in clinical observation

assessment: assessment of the clinical observation

plan: plan of the clinical observation

clinical observation_dbref: database reference of the clinical observation

"clinical_observation_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

data_attbt: attribute of the datum

version: version of the database

"complication element

complication_description: description of the complication

complication_condition: condition to express the complication

complication_: probability when the complication is expressed

"conservative element

pharmaceutical: pharmaceutical element

physical: physical treatment element

"diagnostic_criteria element

diagnostic_standard: standard diagnostic criteria

diagnostic_modify: modified diagnostic criteria

diagnostic_differential: differential diagnosis use

"differences element

race: racial difference

gender: gender difference

"direct_annotation element

whole_genome_sequence: whole genome sequence of the datum

mendelian_segregate: known mendelization

homozygote_detect: homozygote individuals observation in sample

somatic_mutation: known somatic mutation

experimental_analysis: explanation of the experimental and the analysis

epidemiology: epidemiology of the disease and associated gene

var_ann_misc: variation annotation miscellaneous

"disease element

disease_description: description of the disease

disease_epidemiology: epidemiology of the disease

disease_dbref element: database references of the disease

attributes

_id:of the disease


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

_id:of the submitter

_date:created

_date:modified

"disease_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"disease_description element

name: name of the disease

synonym: synonym of the disease

definition_concept: definition or concept of the disease

classification: classification of the disease

etiology: etiology of the disease

laboratory_findings: element of laboratory findings of the disease

pathological_findings: pathological findings of the disease

symptoms: symptoms of the disease

diagnostic_criteria: criteria of diagnosis for the disease

therapy: therapy of the disease

complication: complications of the disease

prophylaxes: prophylaxes of the disease

"disease_epidemiology element

striking_age: striking age of the disease

striking_body_area: striking body area of the disease

striking_land_area: striking land area of the disease

laterality: laterality of the disease e.g. hemilateral, bilateral

differences: statistical differences

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

prognosis: prognosis of the disease

etiology_expression_probability: expression probability of the etiology

laboratory_findings_expression_probability: expression probability of the laboratory findings

symptoms_expression_probability: expression probability of the symptom

prophylaxes_expression_probability: expression probability of the prophylaxes

responder_sideffects_expression_probability: expression probability of the responder, the side effect

pathological_findings_expression_probability: expression probability of the pathological findings

complication_expression_probability: expression probability of the complication

"environmental_condition element

expression_condition: environmental condition of the expression

"epidemiology element

ass_gene: associated gene

disease_epidemiology: epidemiology of the associated disease

population: population of the variation

frequency: frequency of the variation

"etiology element

etiology_description: description of the etiology and its mechanism

etiology_condition: conditions to express the etiology

etiology_expression_probability: expression probability of the etiology

"experiment_analysis element

variation_identify: information to identify the variation

variation_characterize: characterization of the variation datum

"experimental_assay_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

"expression_condition element

expression_condition_description:of the expression condition

expression_condition_probability:of the expression condition

omics_annotation: annotation of the omics

"family_history element

family_history_description: description of the family history

family_history_dbref: database_reference of the family history

"family_history_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"family_history_description element

family_member: member of the family

relation_structure: relational structure of the family

"family_member element

personal_info: personal information of variation data

phenotype: phenotype of the sequence variation

clinical_annotation: clinical annotation of the sequence variation

"frequency element

haplotype: haplotype

haplotype_frequency: frequency of the haplotype

allele: observed allele

allele_frequency: frequency of the observed allele

genotype: genotype

genotype_frequency: frequency of the genotype

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

attributes

_id: ID of the frequency

_id: ID of the submitter

_id: ID of the population

_id: ID of the assay

_id: ID of the publication

_date: date created

_date: date modified

"indirect_annotation element

person_info: personal information of variation data

phenotype: phenotype of the sequence variation

omics_annotation: annotation type of the omics

environmental_condition: environmental conditions of the expression

clinical_annotation: clinical annotation of the sequence variation

"laboratory_findings element

laboratory_findings_type: type of laboratory findings e.g. urine, blood, ECG, image

laboratory_findings_description: description of the laboratory findings

laboratory_findings_condition: condition of the laboratory findings

laboratory_findings_expression_probability: expression probability of the laboratory findings with its degree

"link_dis_index element

d: d value for Linkage Disequilibrium test

d_prime: d prime for LD test

r-square: r square for LD test

"location element

chromosome_number: the number of the chromosome

position: position of the variation in the chromosome

map: chromosome map on which the variation is

orientation: chromosome orientation on which the variation is

ass_gene: associated gene

location_dbref: database reference of location


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

"method element

method_name: statistical method name

method_description: description or explanation of the method

method_url: URL of the method

"objective_findings element

laboratory_findings: laboratory findings of the disease

pathological_findings: pathological findings of the disease

complication: complications of the disease

family_history: family history

"omics_annotation element

omics_type element: type of omics

omics_description element: description of the omics

omics_dbref element: database reference of the omics

"pathological_findings element

pathological_findings_description: description of the pathological finding

pathological_findings_condition: condition of the pathological finding

pathological_findings_expression_probability: expression probability of the pathological finding

"pcr element

pcr_confirmed: artifact verification e.g. variation found on repeat PCR sample

pcr_primer: primer sequence

pcr_product: PCR product e.g. single band, multi band

pcr_profile: PCR profile

"personal_info element

personal_description: description of the personal information

personal_dbref: database reference of the personal information

"pharmaceutical element

responder_sideffects: responder and/or side effects

"phenotype element

phenotype_description: description of the phenotype

phenotype_dbref: database references of the phenotyope

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

attributes

_id: ID of the phenotype

_id: ID of the submitter

_date: date created

_date: date modified

"phenotype_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"phenotype_description element

phenotype_type: type of the phenotype

phenotype_condition: condition to express the phenotype

phenotype_probability: probability to express the phenotype on the conditions

"plan element

therapy: therapy of the disease

"population element

population_description: description of the population

organism: organism

differences: statistical differences

population_parameter: parameter of population

sample_size: sample size of population

population_misc: population miscellaneous

population_dbref: database references of the population

attributes

_id: ID of the population

_id: ID of the submitter


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

_date: date created

_date: date modified

"population_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"prophylaxes element

prophylaxes_description: description of the prophylaxis

prophylaxes_condition: condition of the prophylaxis

prophylaxes_expression_probability: expression probability of the prophylaxis

"publication element

title: title of the publication

author: author of the publication

journal: journal of the publication

volume: volume of the publication

supplement: supplement of the publication

issue: issue of the publication

issue_supplement: issue supplement of the publication

pages: page of the publication

year: year of the publication

publication_status: status of the publication

mesh_term: mesh term of the publication

publication_dbref: database references of the publications

submitter: submitter of the publication

attributes

_id: ID of the publication

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

_id: ID of the submitter

_date: date created

_date: date modified

"publication_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"responder_sideffects element

responder_sideffects_causer: causer of the responder and/or side effect

responder_sideffects_description: description of the responder and/or side effect

responder_sideffects_condition: condition of the responder and/or side effect

responder_sideffects_expression_probability: expression probability of the responder and/or side effect

"source element

source_release_date: date released

source_modify_date: date modified

source_rawdata: rawdatum of the source

source_dbref: database reference of source

"splice_variant_number element

refSeq_number: reference sequenc number

"submitter element

submitter_name: name of the submitter

address: address of the submitter

email: email of the submitter

tel: telephone of the submitter

fax: Fax of the submitter

institution: Institution of the submitter


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

laboratory: Laboratory of the submitter

submitter_dbref: database references of the submitter

publication: publication of the experiment

"subjective_findings element

symptoms: laboratory findings of the disease

"submitter_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"symptoms element

sysmptoms_description: description of the symptom

sysmptoms_condition: condition of the symptom

sysmptoms_expression_probability: expression probability of the symptom with its degree (table)

"therapy element

conservative: conservative treatment

surgery: surgical treatment

radiation: radiation therapy

"var_ann_misc element

var_ann_misc_description: description of the variation annotation miscellaneous

var_ann_misc_dbref: database references of the variation annotation miscellaneous

attributes

_ann_misc_id: ID of the variation annotation miscellaneous

_id: ID of the submitter

_date: date created

_date: date modified

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

"var_ann_misc_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database

database_attbt: attribute of the database

data_id: ID of the datum

data_attbt: attribute of the datum

version: version of the database

"variation_att element

molecular_type: type of molecule e.g. DNA, RNA

allele: observed allele

length: sequence length including franking sequence

f5sequence: 5' flanking sequence

f3sequence: 3' flanking sequence

validation_status: status of validation as (Proven, Suspected)

success_rate: certainty of variation information

"variation_characterize element

method: statistical method

p-value: p value for significance(Associaion study)

link_dis_index: linkage disequilibrium index for LD test

"variation_data element

variation_type: type of variation

location: location of the variation

variation_att: attribute information of the variation

source: source of the sequence variation

variation_dbref: database references of the variation

"variation_dbref element

database_name: name of the database

database_id: ID of the database

link_url: URL of the database



© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

```
    database_attbt:  attribute of the database

    data_id:  ID of the datum

    data_attbt:  attribute of the datum

    version:  version of the database

"variation_identify element

    experimental_assay_result:  result of the experimental assay

    experimental_assay_description:  description of the experimental assay

    experimental_assay_parameter:  parameter of the experimental assay

    experimental_assay_dbref:  database reference information

    publication:  publication of the experiment

    submitter:  submitter of the publication

    pcr:  PCR

-->

<!ELEMENT gsvml (variation_data , direct_annotation* , indirect_annotation*)>

<!ELEMENT variation_data (variation_type, location, variation_att+, source*, variation_dbref*)>

<!-- variation_type comment

scope of variation_type are: SNP, rSNP, cSNP, iSNP, uSNP, gSNP, RFLP, MS, STRP, VNTR, Insertion, Deletion, Sustitution, Other

-->

<!ELEMENT variation_type (#PCDATA)>

<!ELEMENT location (chromosome_number , position , map? , orientation , ass_gene* , location_dbref*)>

<!ELEMENT chromosome_number (#PCDATA)>

<!ELEMENT position (#PCDATA)>

<!ELEMENT map (#PCDATA)>

<!ELEMENT orientation (#PCDATA)>

<!ELEMENT ass_gene (ass_gene_name , ass_gene_structure? , aminoacid_substitution? , codon_substitution? , codon_position? ,

                        ass_gene_symbol? , ass_gene_alias? , ass_gene_product? , ass_gene_evidence_type* , changed_motif* ,
changed_motif_name* ,

                        changed_splice_site* , splice_variant_number? , ass_gene_dbref*)>

<!ELEMENT ass_gene_name (#PCDATA)>

<!ELEMENT ass_gene_structure (#PCDATA)>
```

© ISO 2009 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009

```
<!ELEMENT aminoacid_substitution (#PCDATA)>

<!ELEMENT codon_substitution (#PCDATA)>

<!ELEMENT codon_position (#PCDATA)>

<!ELEMENT ass_gene_symbol (#PCDATA)>

<!ELEMENT ass_gene_alias (#PCDATA)>

<!ELEMENT ass_gene_product (#PCDATA)>

<!ELEMENT ass_gene_evidence_type (#PCDATA)>

<!ELEMENT changed_motif (#PCDATA)>

<!ELEMENT changed_motif_name (#PCDATA)>

<!ELEMENT changed_splice_site (#PCDATA)>

<!ELEMENT splice_variant_number (refSeq_number+)>

<!ELEMENT refSeq_number (#PCDATA)>

<!ELEMENT ass_gene_dbref (database_name , database_id? , link_url? , database_attbt? , data_id , data_attbt? , version?)>

<!ELEMENT database_name (#PCDATA)>

<!ELEMENT database_id (#PCDATA)>

<!ELEMENT link_url (#PCDATA)>

<!ELEMENT database_attbt (#PCDATA)>

<!ELEMENT data_id (#PCDATA)>

<!ELEMENT data_attbt (#PCDATA)>

<!ELEMENT version (#PCDATA)>

<!ELEMENT location_dbref (database_name , database_id? , link_url? , database_attbt? , data_id , data_attbt? , version?)>

<!ELEMENT variation_att (molecular_type , allele* , length , f5sequence? , f3sequence? , validation_status? , success_rate?)>

<!ATTLIST variation

    variation_id  CDATA #REQUIRED

    submitter_id  CDATA #REQUIRED

    population_id CDATA #REQUIRED

    experimental_assay_id     CDATA #REQUIRED

    publication_id CDATA #REQUIRED

    create_date   CDATA #IMPLIED

    modify_date   CDATA #IMPLIED
```


© ISO 2009 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 25720:2009