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)

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

Published in Switzerland

Contents

Page

Foreword

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

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

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

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

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

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).

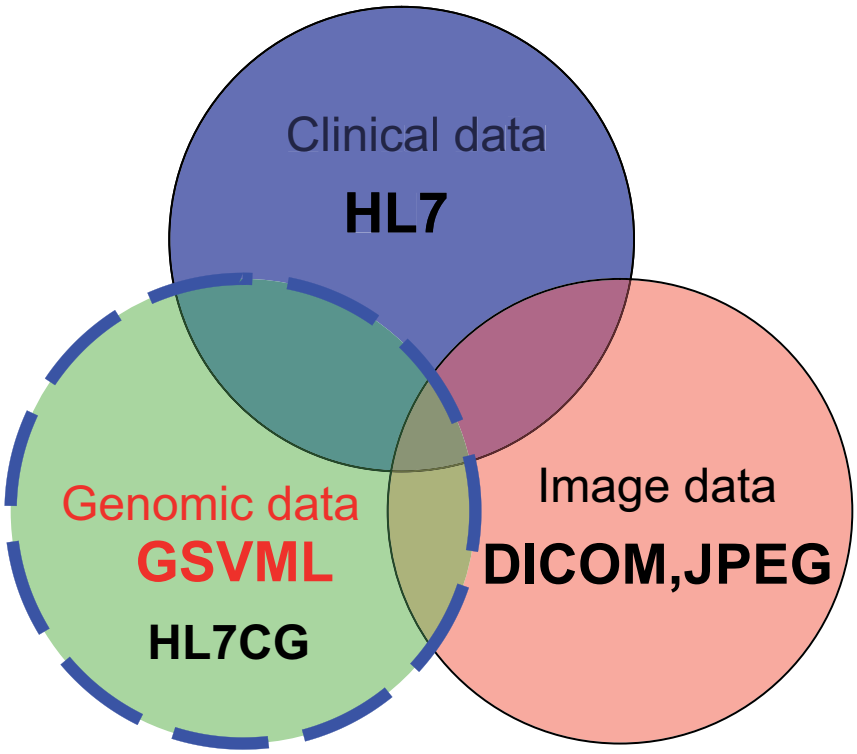


Figure 1 — Major data types of health care

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

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.

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

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).

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).

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

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.

— 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.

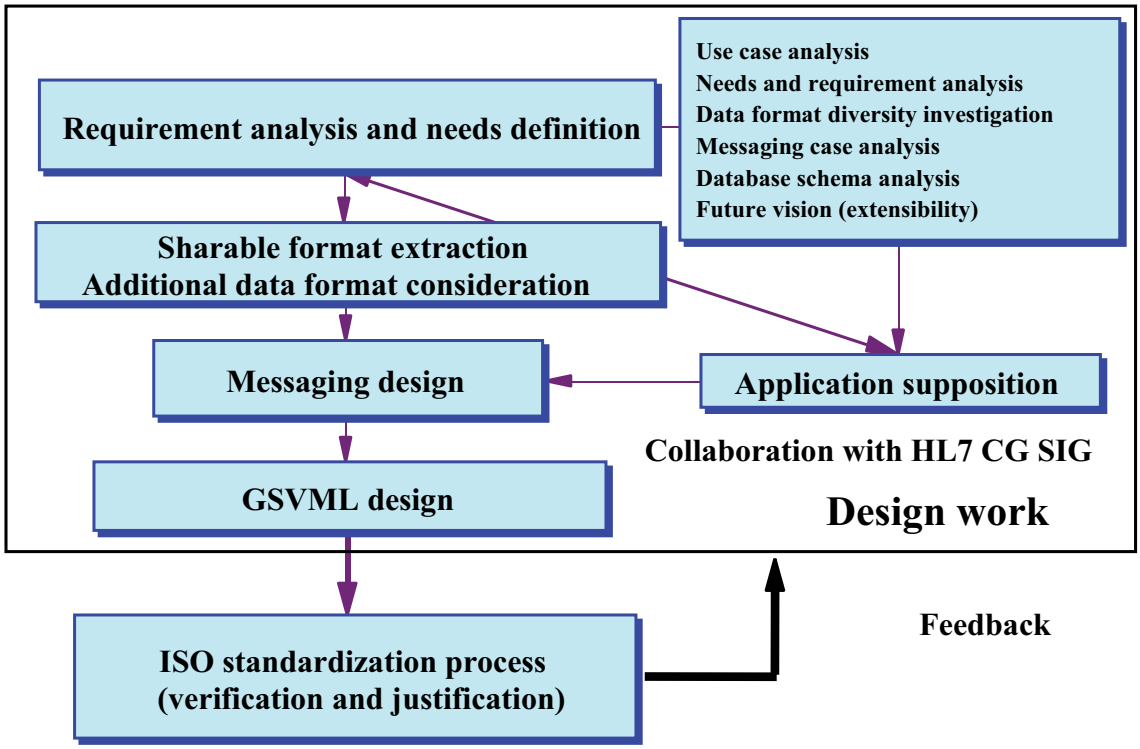
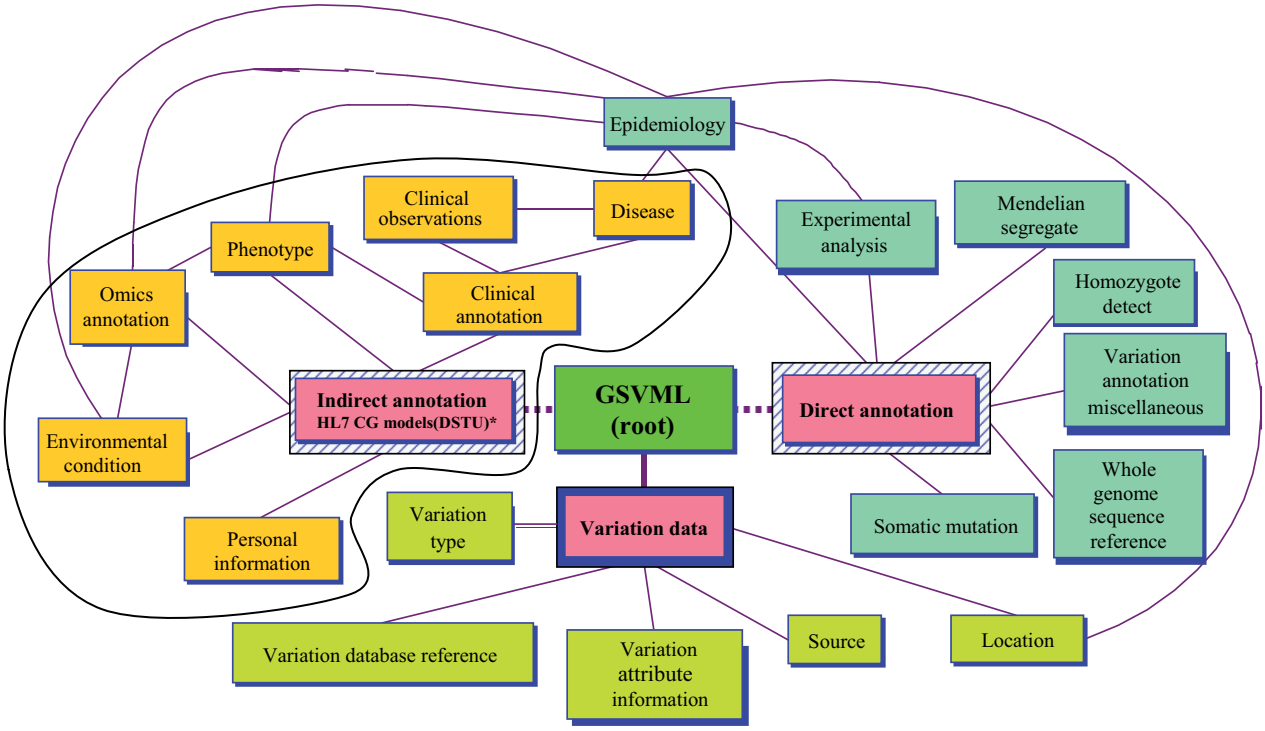


Figure 2 — Outline of the process of GSVML development



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

Figure 3 — The outlined structure of GSVML

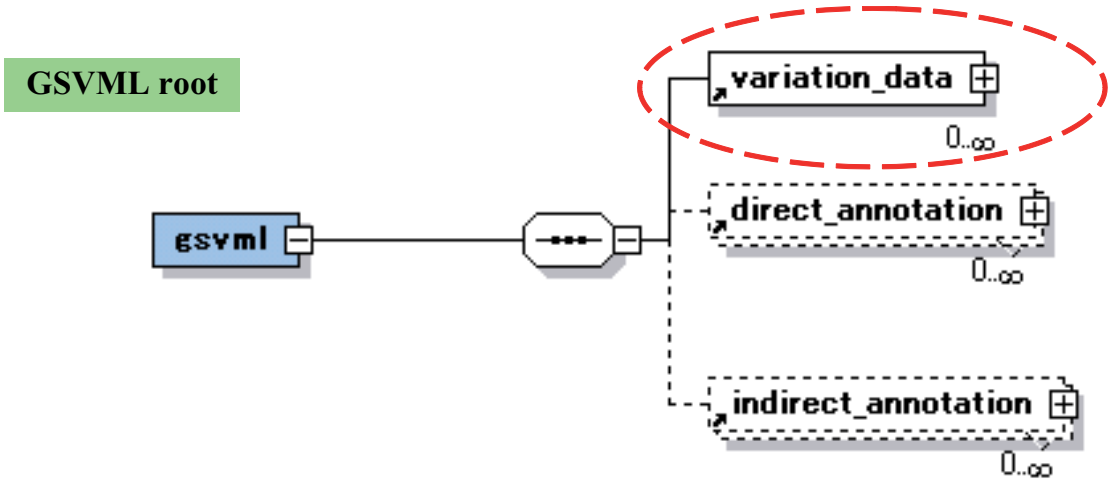


Figure 4 — Detailed structure of GSVML

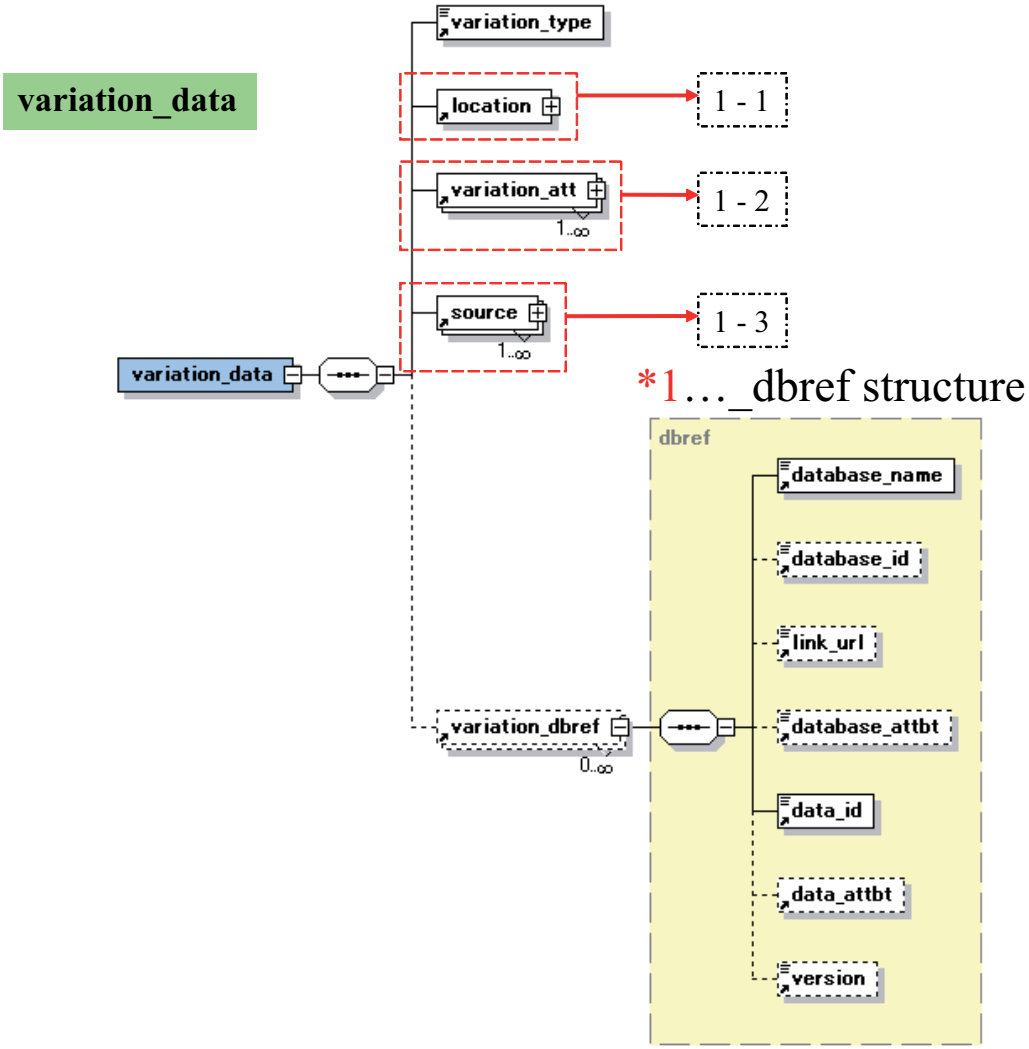


Figure 5 — Detailed structure of GSVML

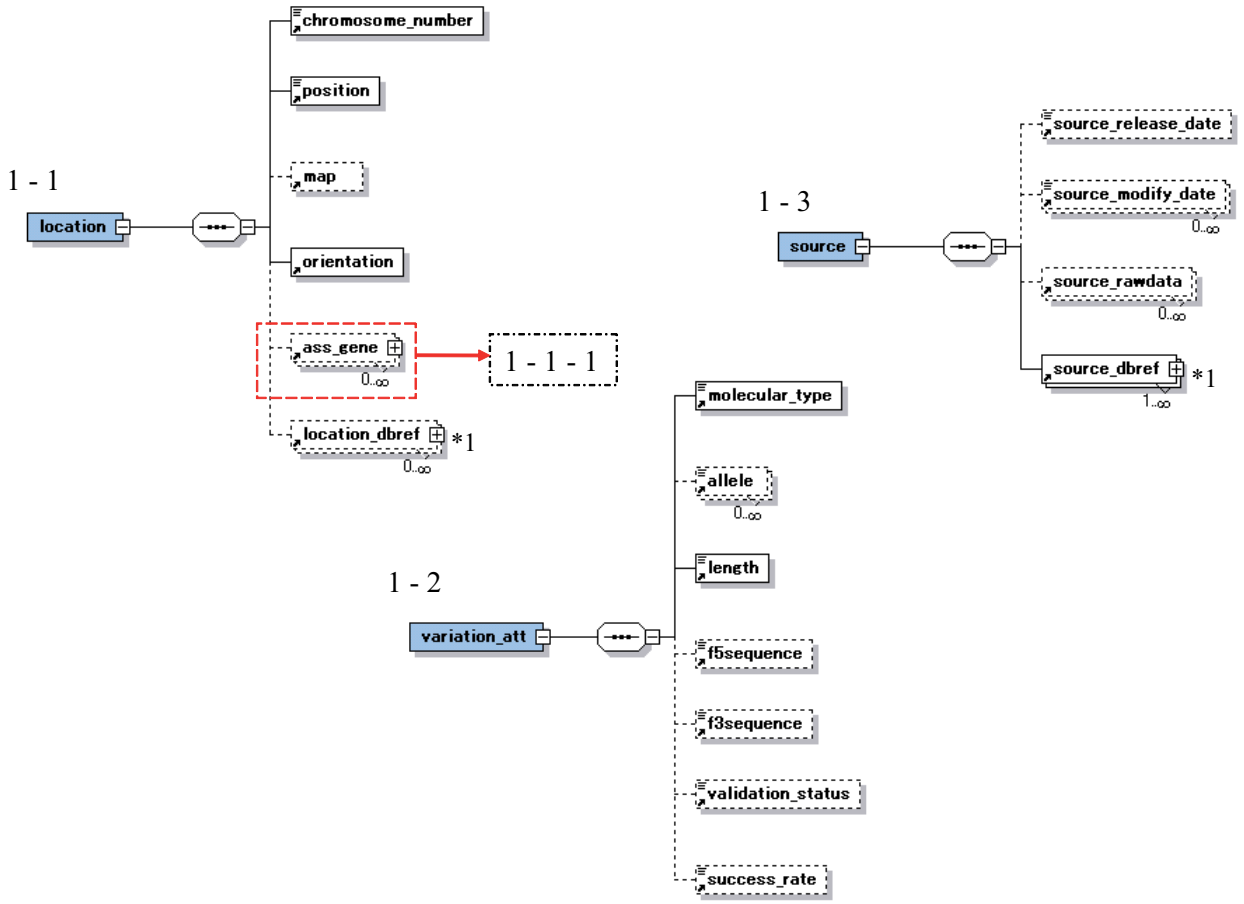


Figure 6 — Detailed structure of GSVML

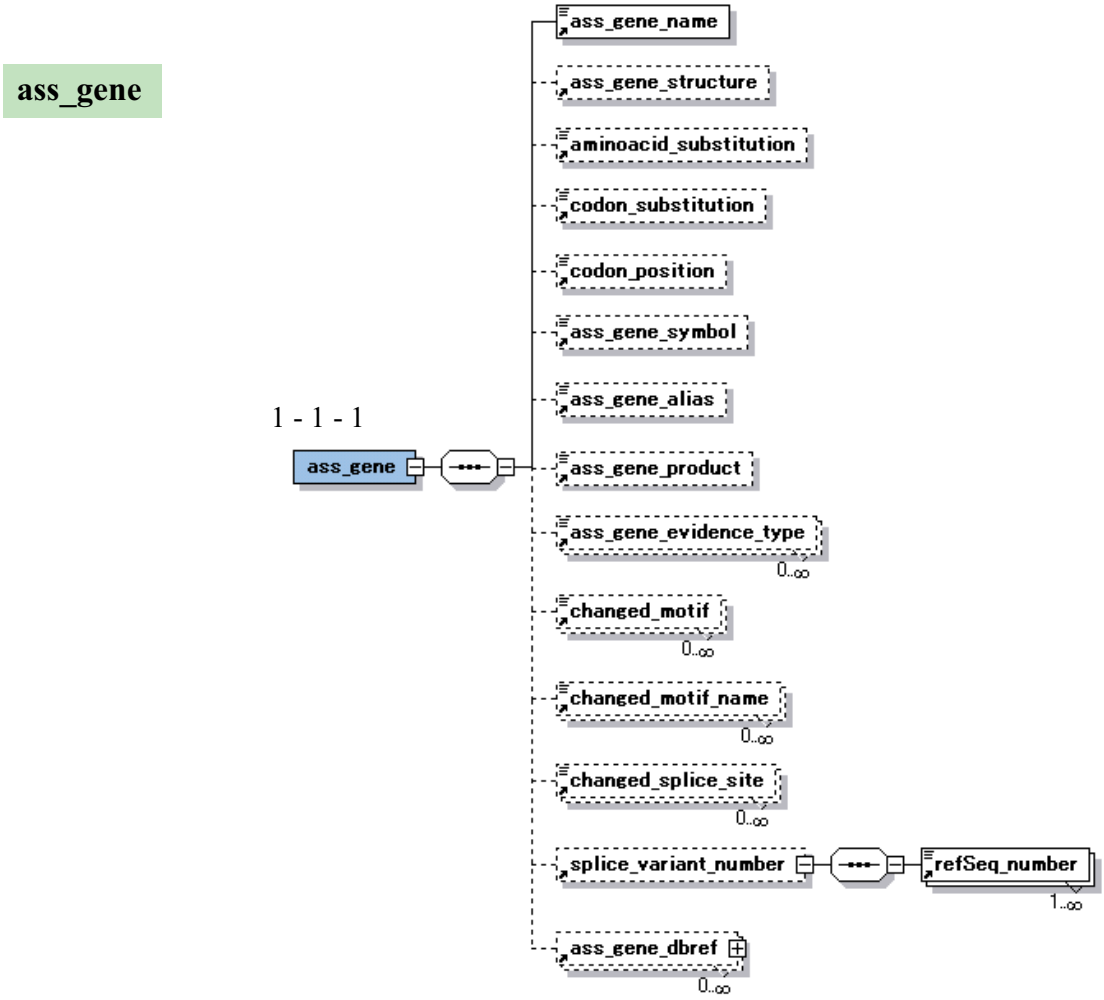


Figure 7 — Detailed structure of GSVML

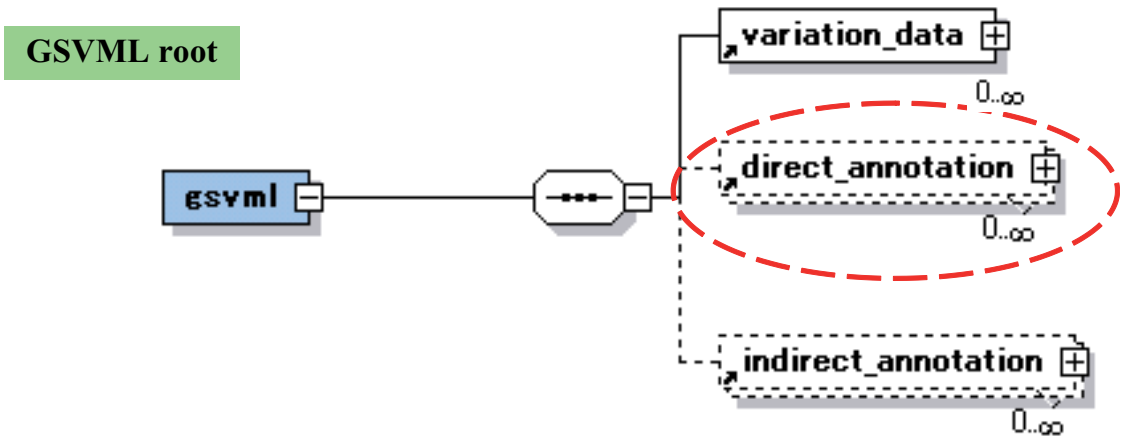


Figure 8 — Detailed structure of GSVML

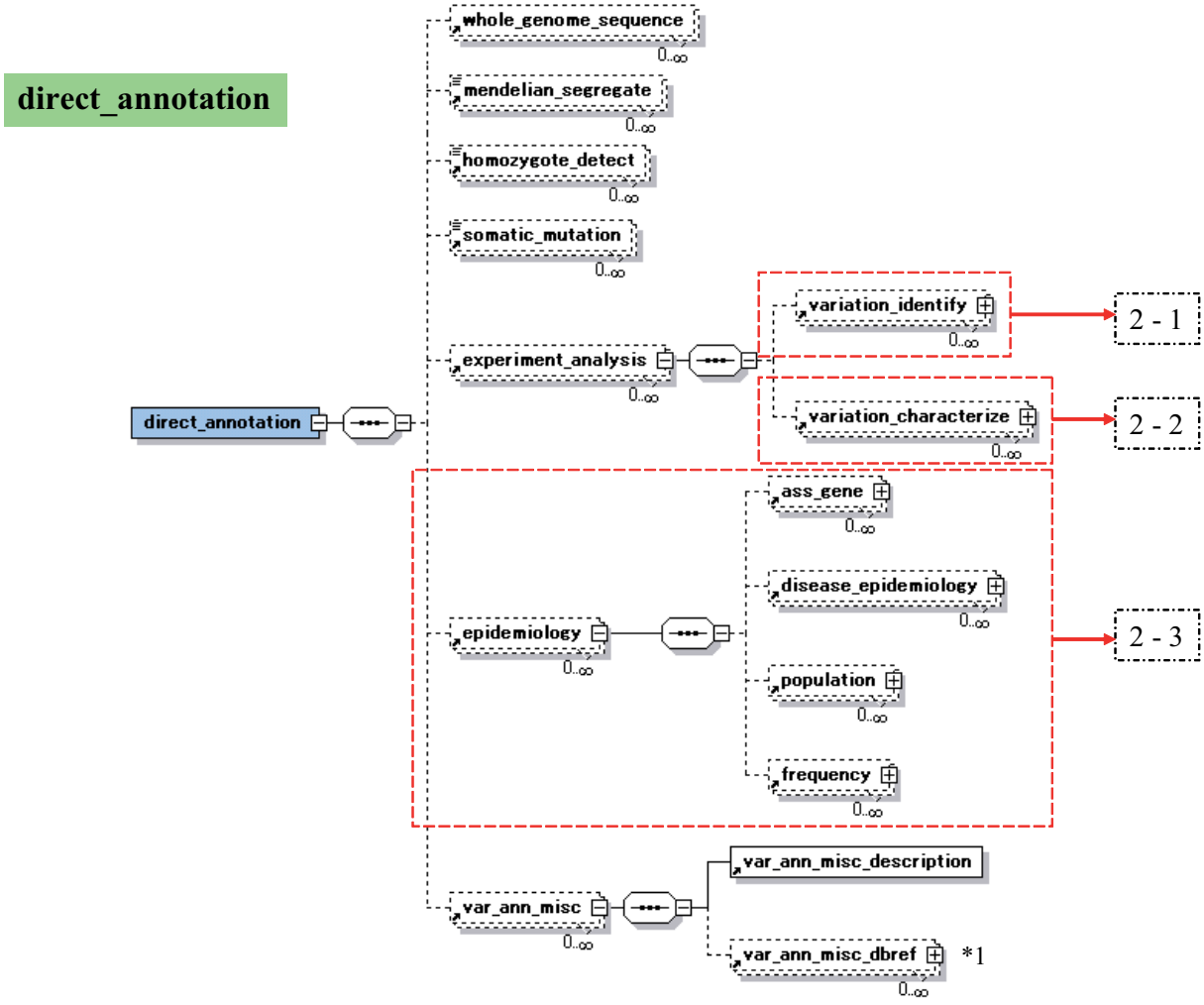


Figure 9 — Detailed structure of GSVML

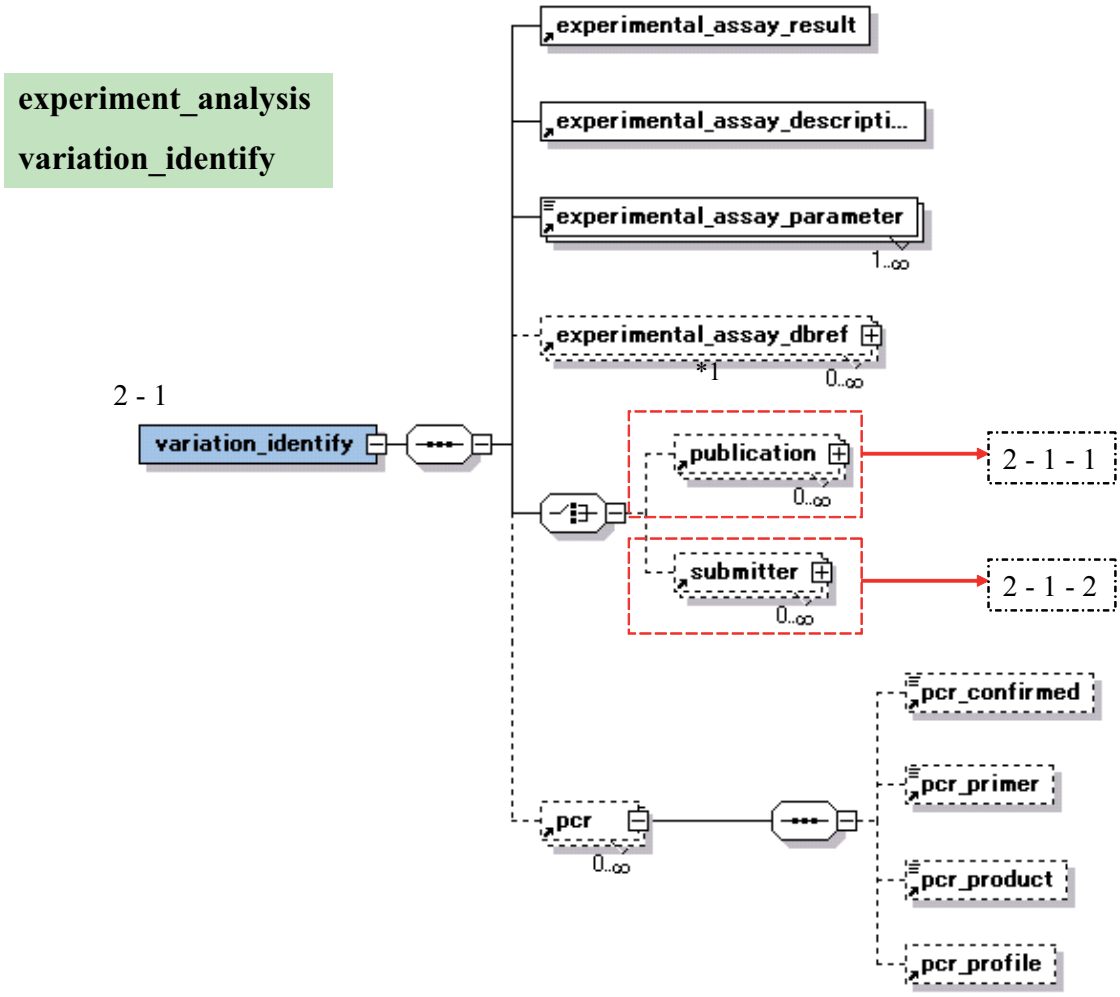


Figure 10 — Detailed structure of GSVML

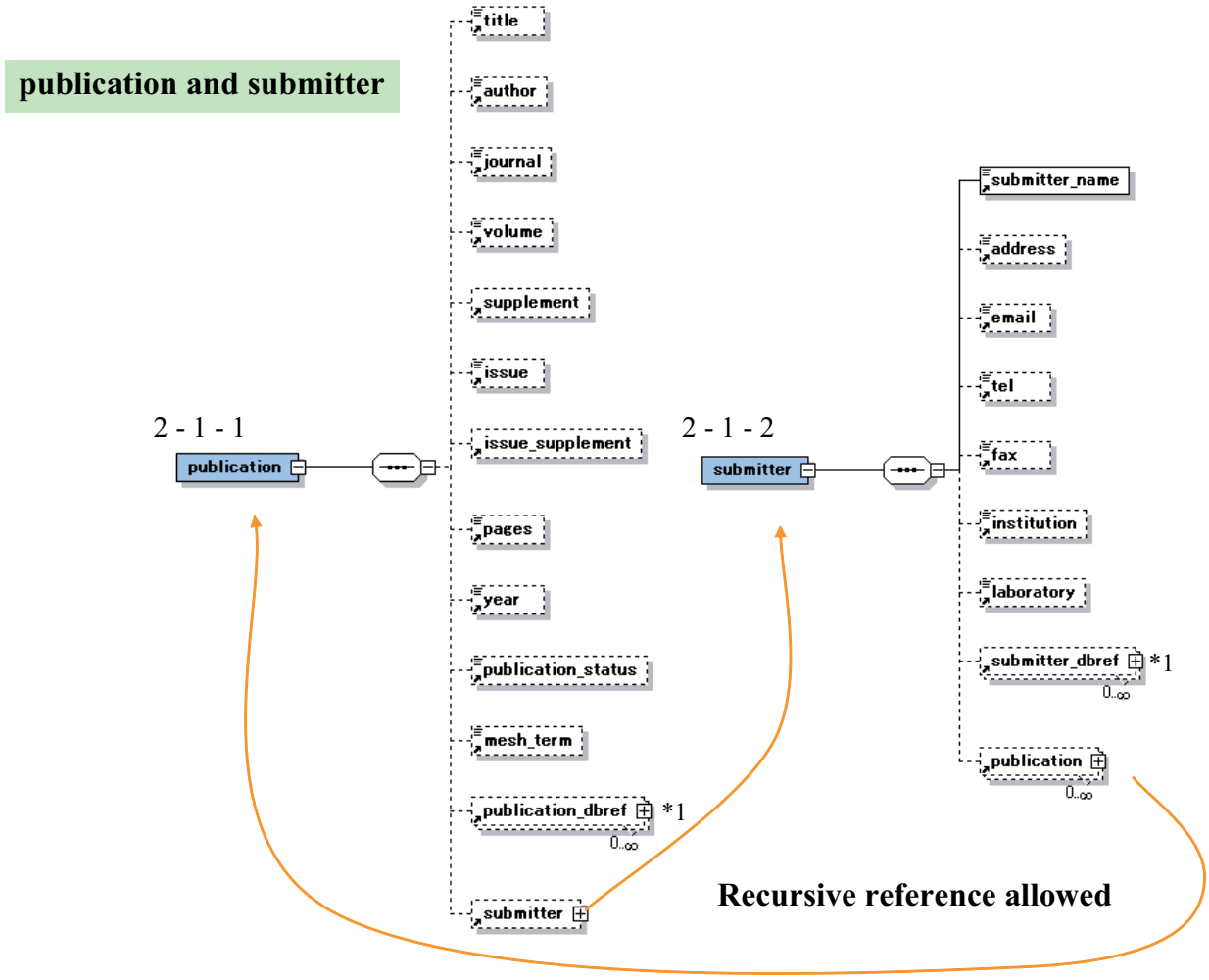


Figure 11 — Detailed structure of GSVML

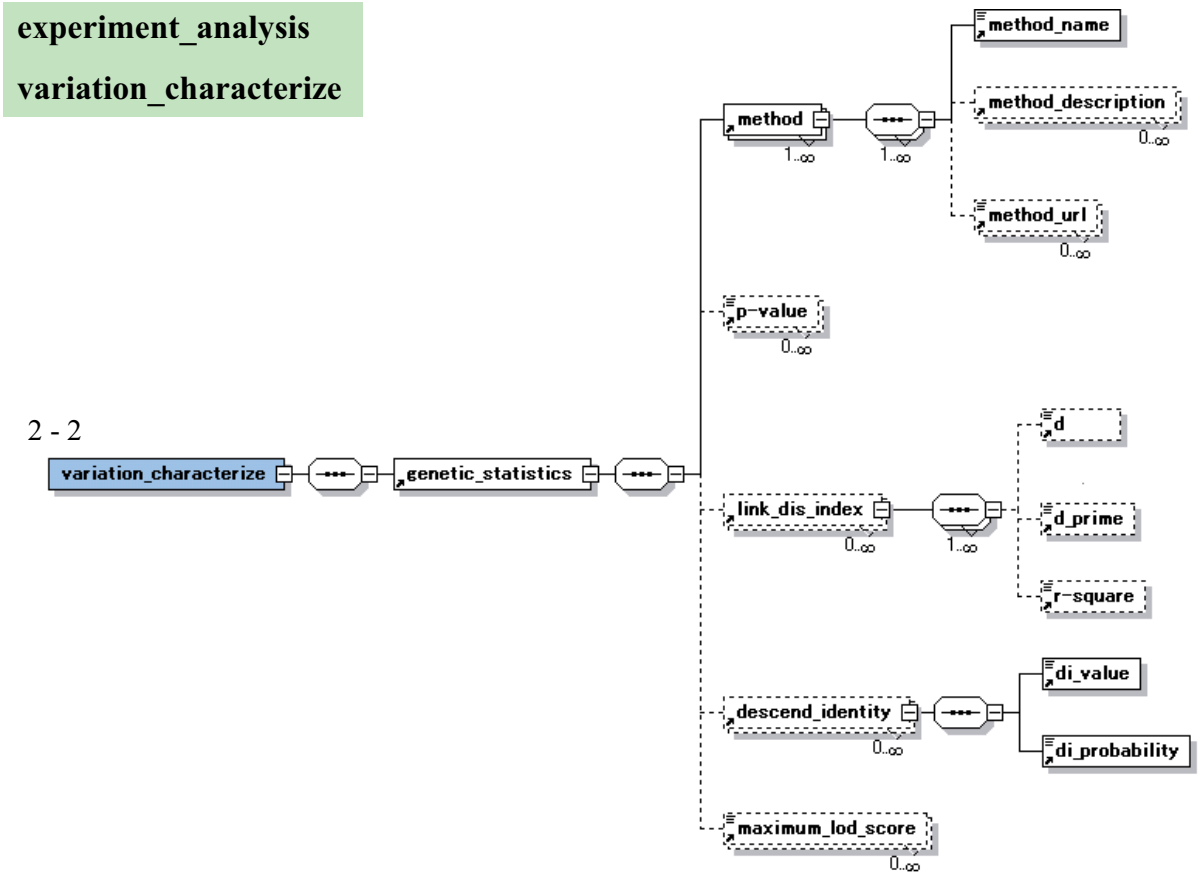


Figure 12 — Detailed structure of GSVML

epidemiology

(Statistical data section)

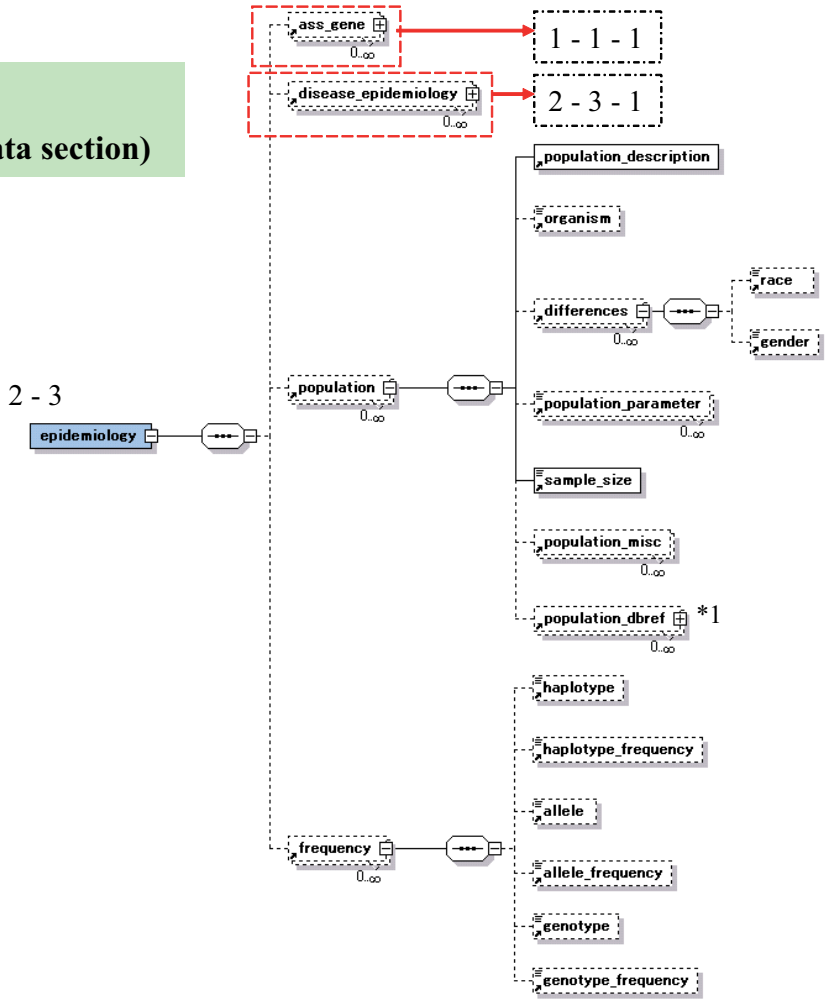


Figure 13 — Detailed structure of GSVML

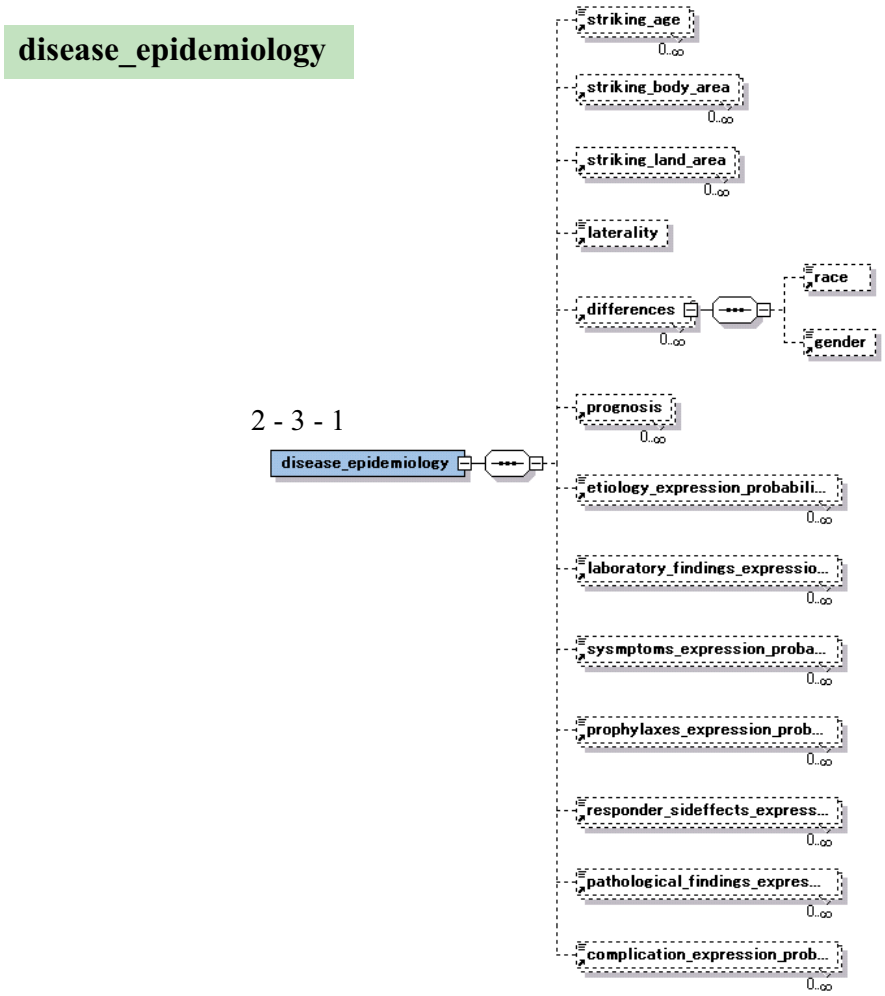


Figure 14 — Detailed structure of GSVML

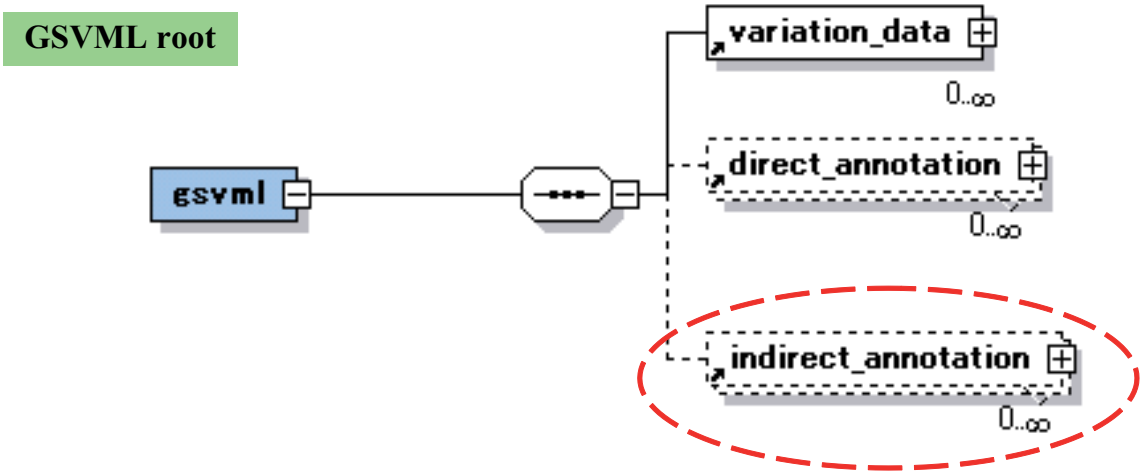


Figure 15 — Detailed structure of GSVML

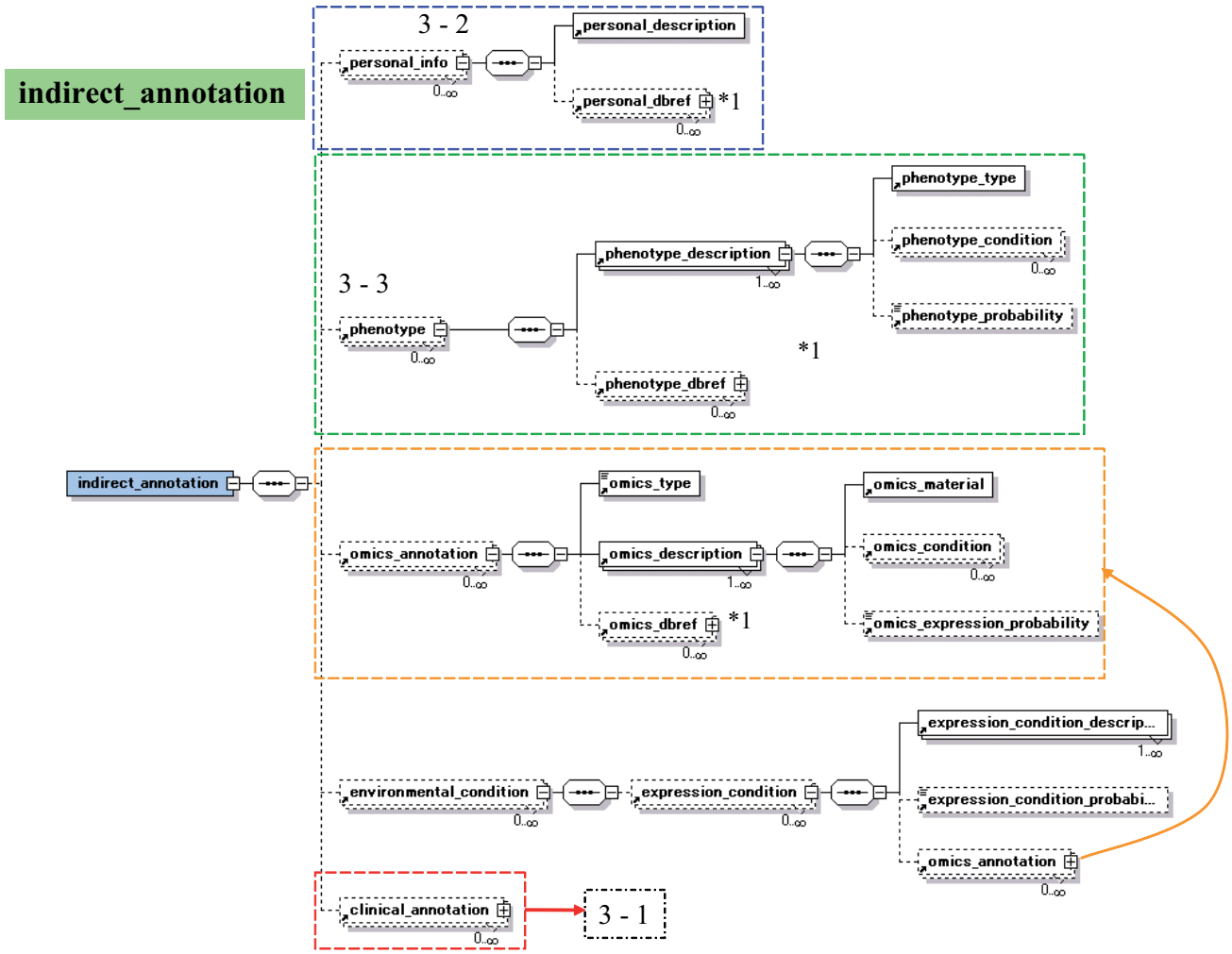


Figure 16 — Detailed structure of GSVML

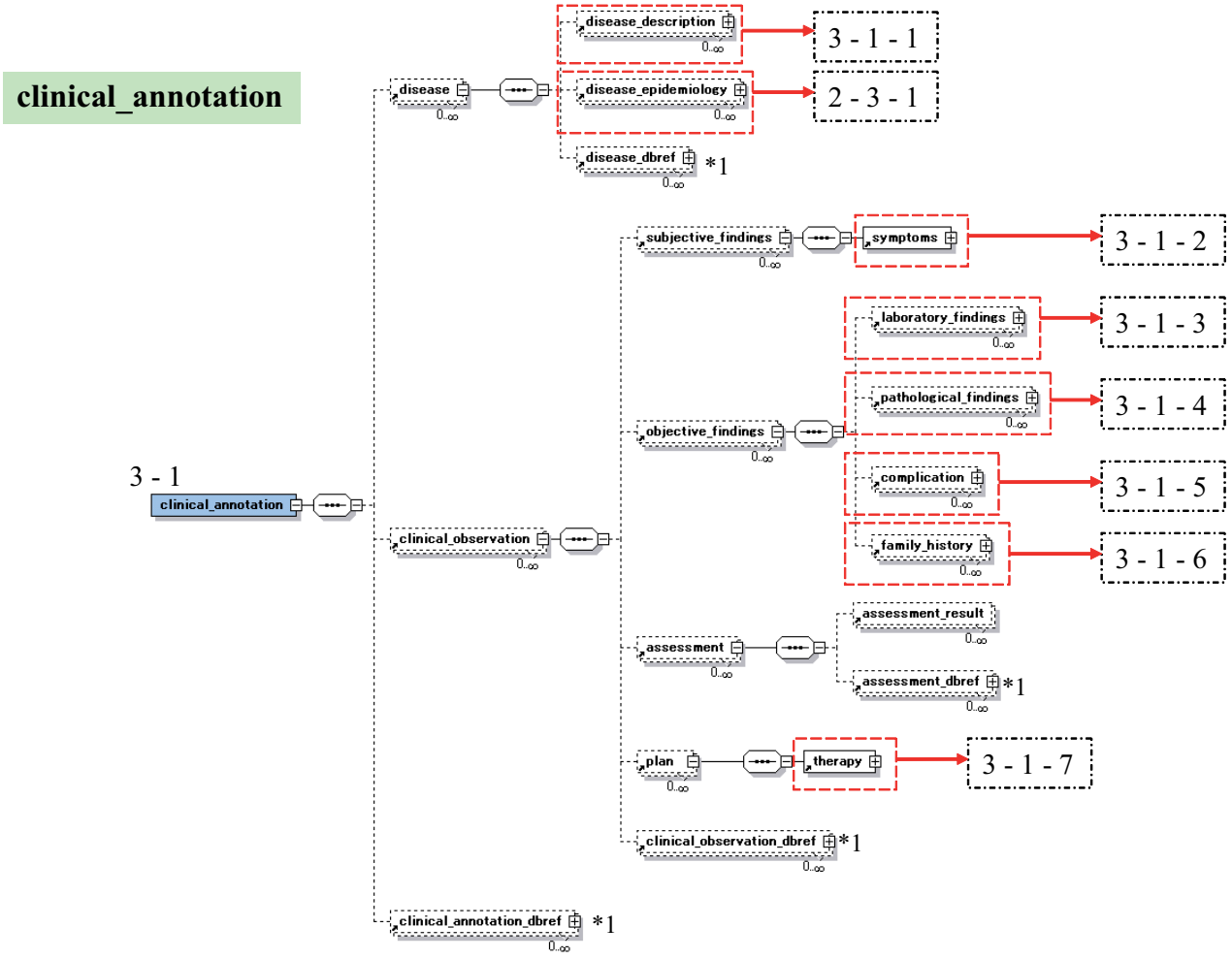


Figure 17 — Detailed structure of GSVML

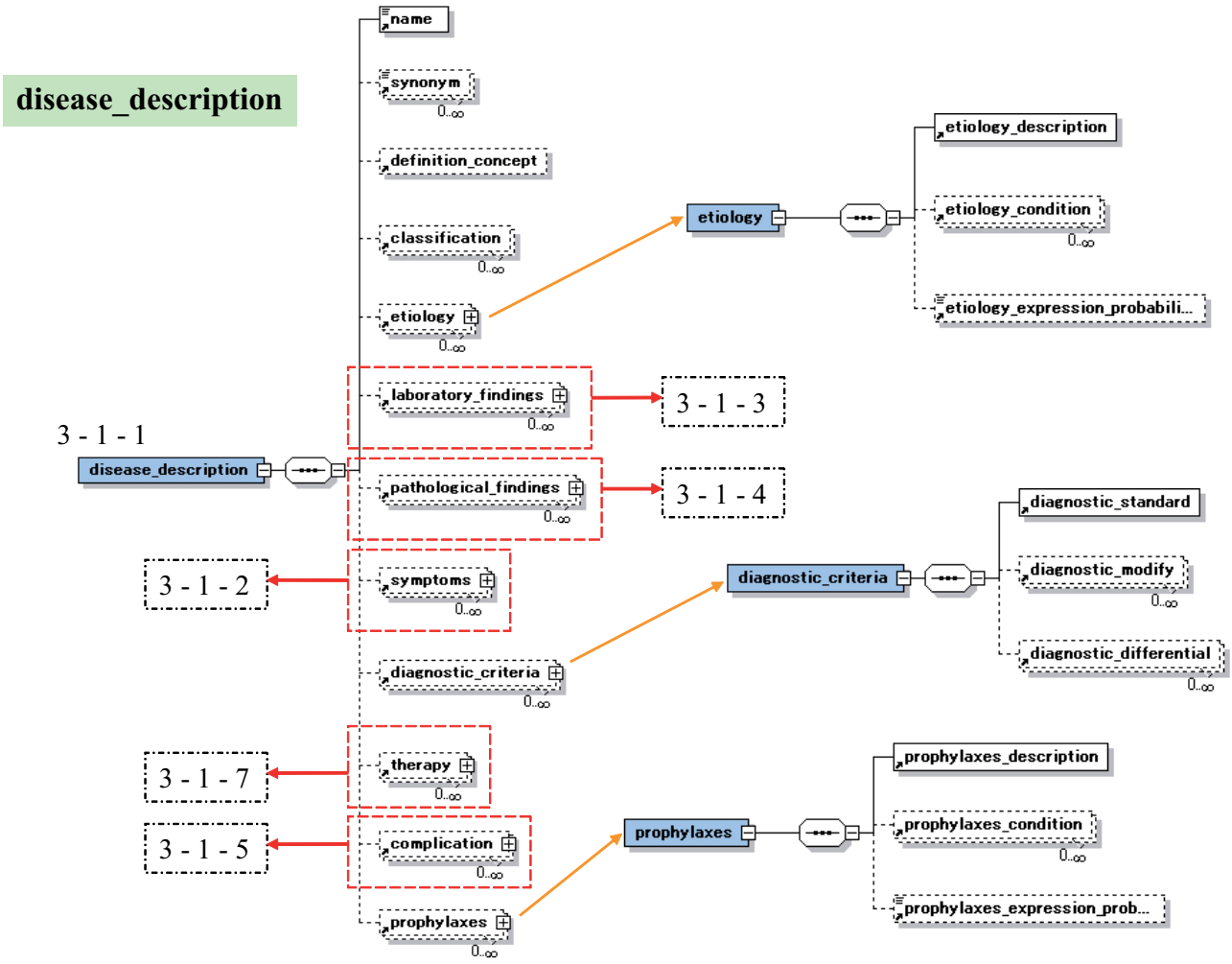


Figure 18 — Detailed structure of GSVML

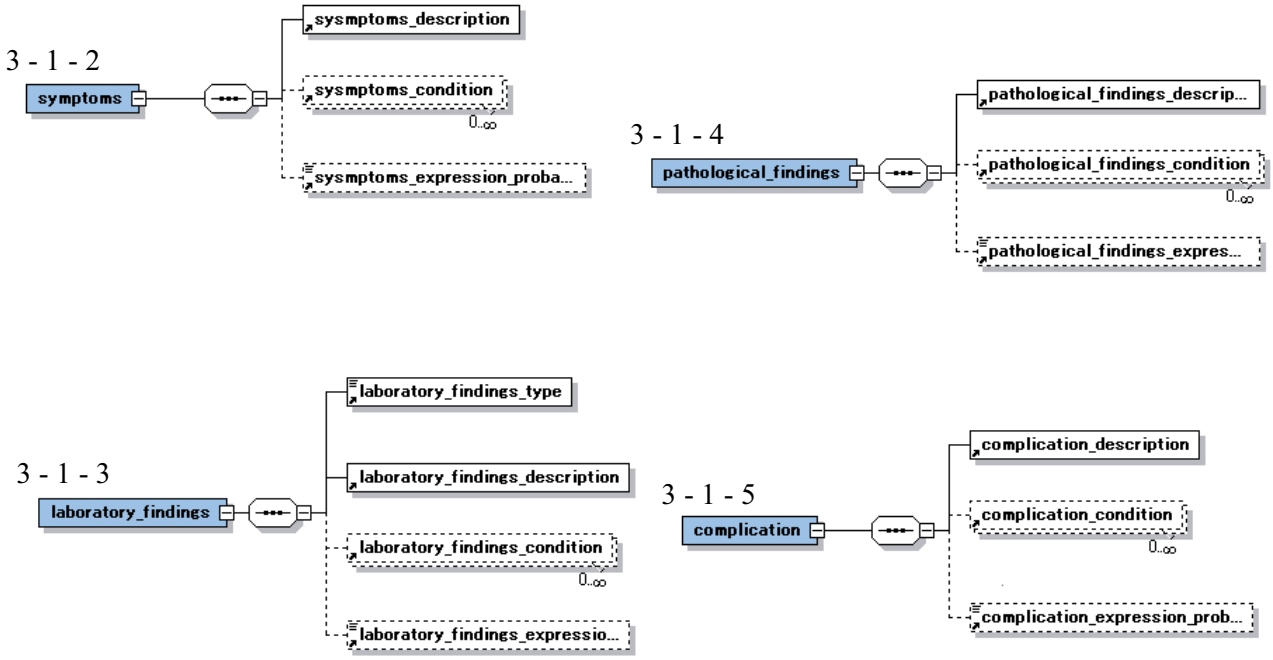


Figure 19 — Detailed structure of GSVML

family_history

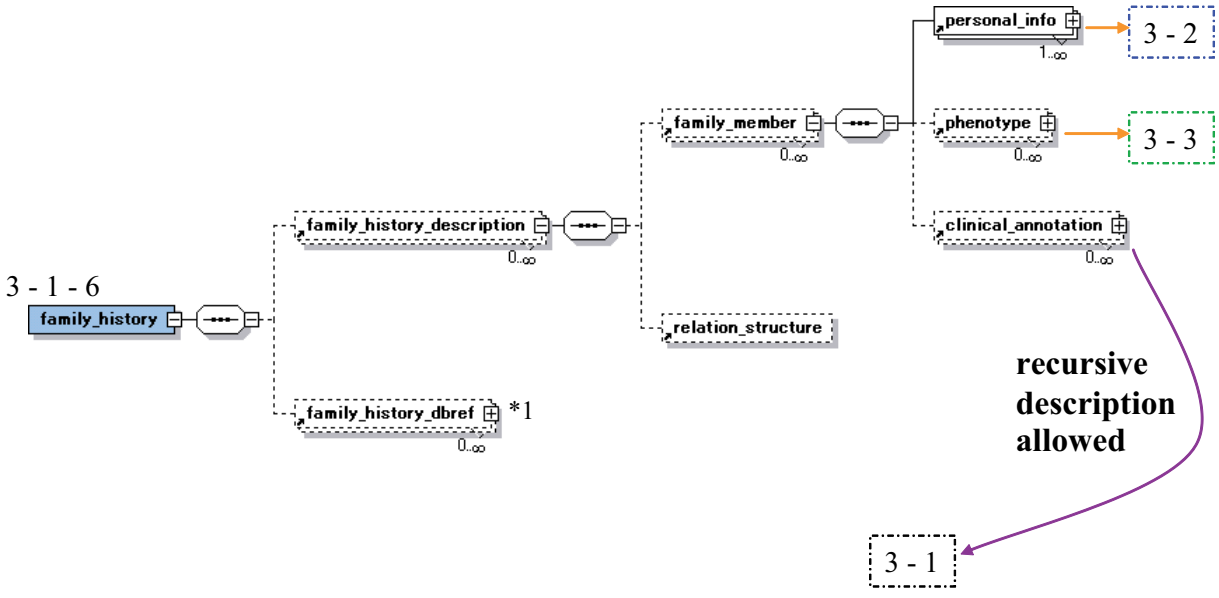


Figure 20 — Detailed structure of GSVML

therapy

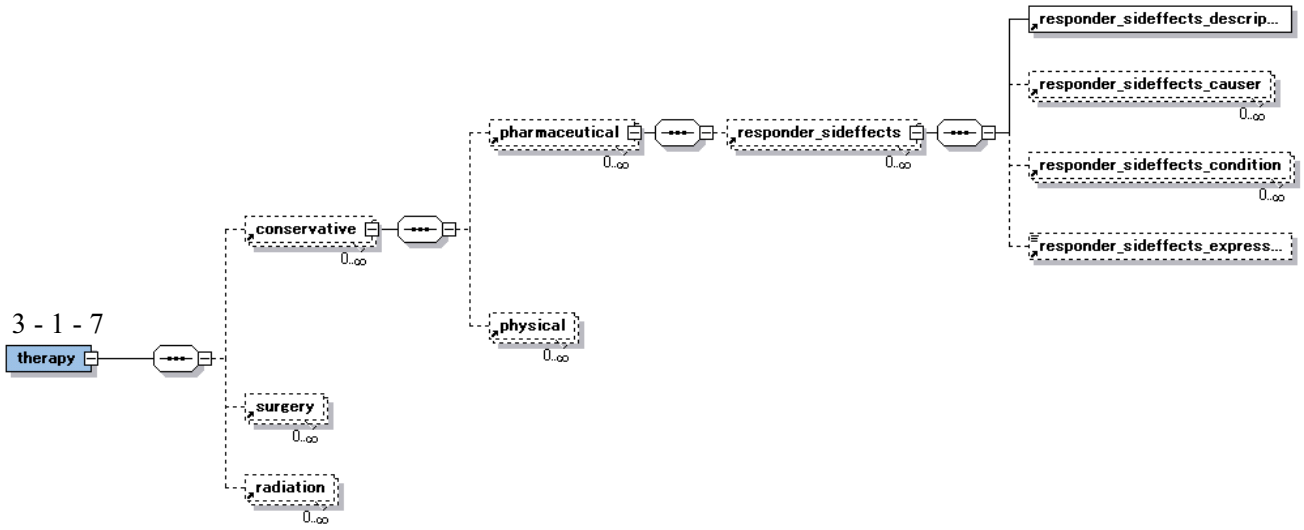


Figure 21 — Detailed structure of GSVML

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
```

"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

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

_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

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

"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

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

"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

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

_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

_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

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

"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

```
   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)>
```

```
<!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
```

```
>

<!ELEMENT molecular_type (#PCDATA)>

<!ELEMENT allele (#PCDATA)>

<!ELEMENT length (#PCDATA)>

<!ELEMENT f5sequence (#PCDATA)>

<!ELEMENT f3sequence (#PCDATA)>

<!-- validation_status comment

scope of validation_status are: Proven, Suspected

-->

<!ELEMENT validation_status (#PCDATA)>

<!ELEMENT success_rate (#PCDATA)>

<!ELEMENT source (source_release_date? , source_modify_date* , source_rawdata* , source_dbref+)>

<!ELEMENT source_release_date (#PCDATA)>

<!ELEMENT source_modify_date (#PCDATA)>

<!ELEMENT source_rawdata (#PCDATA)>

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

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

<!ELEMENT direct_annotation (whole_genome_sequence* , mendelian_segregate* , homozygote_detect* ,

                somatic_mutation* , experiment_analysis* , epidemiology* , var_ann_misc*)>

<!ELEMENT whole_genome_sequence (#PCDATA)>

<!ELEMENT mendelian_segregate (#PCDATA)>

<!-- validation_status comment

scope of homozygote_detect are: homo, hetero, unknown

-->

<!ELEMENT homozygote_detect (#PCDATA)>

<!ELEMENT somatic_mutation (#PCDATA)>

<!ELEMENT experiment_analysis (variation_identify* , variation_characterize*)>

<!ELEMENT variation_identify (experimental_assay_result , experimental_assay_description , experimental_assay_parameter+ ,

                experimental_assay_dbref* , (publication* | submitter*) , pcr*)>
```

```
<!ATTLIST variation_identify
        experimental_assay_id      CDATA #REQUIRED
        submitter_id  CDATA #REQUIRED
        create_date   CDATA #IMPLIED
        modify_date   CDATA #IMPLIED
>
<!ELEMENT experimental_assay_result (#PCDATA)>
<!ELEMENT experimental_assay_description (#PCDATA)>
<!ELEMENT experimental_assay_parameter (#PCDATA)>
<!ELEMENT experimental_assay_dbref (database_name , database_id? , link_url? , database_attbt? , data_id , data_attbt? , version?)>
<!ELEMENT publication (title? , author? , journal? , volume? , supplement? , issue? , issue_supplement? ,
                        pages? , year? , publication_status? , mesh_term? , publication_dbref* , submitter?)>
<!ATTLIST publication
        publication_id  CDATA  #REQUIRED
        submitter_id  CDATA #REQUIRED
        create_date   CDATA #IMPLIED
        modify_date   CDATA #IMPLIED
>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT journal (#PCDATA)>
<!ELEMENT volume (#PCDATA)>
<!ELEMENT supplement (#PCDATA)>
<!ELEMENT issue (#PCDATA)>
<!ELEMENT issue_supplement (#PCDATA)>
<!ELEMENT pages (#PCDATA)>
<!ELEMENT year (#PCDATA)>
<!ELEMENT publication_status (#PCDATA)>
<!ELEMENT mesh_term (#PCDATA)>
<!ELEMENT publication_dbref (database_name , database_id? , link_url? , database_attbt? , data_id , data_attbt? , version?)>
```

```
<!ELEMENT submitter (submitter_name , address? , email? , tel? , fax? , institution? , laboratory? , submitter_dbref* , publication)>
<!ATTLIST submitter
        submitter_id  CDATA #REQUIRED
        create_date   CDATA #IMPLIED
        modify_date   CDATA #IMPLIED
>
<!ELEMENT submitter_name (#PCDATA)>
<!ELEMENT address (#PCDATA)>
<!ELEMENT email (#PCDATA)>
<!ELEMENT tel (#PCDATA)>
<!ELEMENT fax (#PCDATA)>
<!ELEMENT institution (#PCDATA)>
<!ELEMENT laboratory (#PCDATA)>
<!ELEMENT submitter_dbref (database_name , database_id? , link_url? , database_attbt? , data_id , data_attbt? , version?)>
<!ELEMENT pcr (pcr_confirmed? , pcr_primer?  , pcr_product? , pcr_profile?)>
<!ELEMENT pcr_confirmed (#PCDATA)>
<!ELEMENT pcr_primer (#PCDATA)>
<!ELEMENT pcr_product (#PCDATA)>
<!ELEMENT pcr_profile (#PCDATA)>
<!ELEMENT variation_characterize (genetic_statistics)>
<!ELEMENT genetic_statistics (method+ , p-value* , link_dis_index* , descend_identity* , maximum_lod_score*)>
<!ELEMENT method (method_name , method_description* , method_url*)>
<!-- method_name comment
scope of method_name are: sib pair, linkage disequilibrium, association study and any other string
-->
<!ELEMENT method_name (#PCDATA)>
<!ELEMENT method_description (#PCDATA)>
<!ELEMENT method_url (#PCDATA)>
<!ELEMENT p-value (#PCDATA)>
<!ELEMENT link_dis_index (d? , d_prime? , r_square?)>
```

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

<!ELEMENT d_prime (#PCDATA)>

<!ELEMENT r_square (#PCDATA)>

<!ELEMENT descend_identity (di_value , di_probability)>

<!ELEMENT di_value (#PCDATA)>

<!ELEMENT di_probability (#PCDATA)>

<!ELEMENT maximum_lod_score (#PCDATA)>

<!ELEMENT epidemiology (ass_gene* , disease_epidemiology* , population* , frequency*)>

<!ELEMENT population (population_description , organism? , differences* , population_parameter* , sample_size ,
population_misc* , population_dbref*)>

<!ATTLIST population

      population_id CDATA #REQUIRED

      submitter_id  CDATA #REQUIRED

      create_date   CDATA #IMPLIED

      modify_date   CDATA #IMPLIED

>

<!ELEMENT population_description (#PCDATA)>

<!ELEMENT organism (#PCDATA)>

<!ELEMENT differences (race? , gender?)>

<!ELEMENT race (#PCDATA)>

<!ELEMENT gender (#PCDATA)>

<!ELEMENT population_parameter (#PCDATA)>

<!ELEMENT sample_size (#PCDATA)>

<!ELEMENT population_misc (#PCDATA)>

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

<!ELEMENT frequency (haplotype? , haplotype_frequency? , allele? , allele_frequency? , genotype? , genotype_frequency?)>

<!ATTLIST frequency

      frequency_id  CDATA #REQUIRED

      submitter_id  CDATA #REQUIRED

      population_id CDATA #REQUIRED
```

```
        assay_id      CDATA #REQUIRED

        publication_id CDATA #REQUIRED

        create_date   CDATA #IMPLIED

        modify_date   CDATA #IMPLIED

>

<!ELEMENT haplotype (#PCDATA)>

<!ELEMENT haplotype_frequency (#PCDATA)>

<!ELEMENT allele_frequency (#PCDATA)>

<!ELEMENT genotype (#PCDATA)>

<!ELEMENT genotype_frequency (#PCDATA)>

<!ELEMENT var_ann_misc (var_ann_misc_description , var_ann_misc_dbref*)>

!ATTLIST var_ann_misc

        var_ann_misc_id    CDATA   #REQUIRED

        submitter_id   CDATA #REQUIRED

        create_date   CDATA #IMPLIED

        modify_date   CDATA #IMPLIED

>

<!ELEMENT var_ann_misc_description (#PCDATA)>

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

<!ELEMENT indirect_annotation (personal_info* , phenotype* , omics_annotation* , environmental_condition* ,
clinical_annotation*)>

<!ELEMENT personal_info (personal_description , personal_dbref*)>

<!ELEMENT personal_description (#PCDATA)>

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

<!ELEMENT phenotype (phenotype_description+ , phenotype_dbref*)>

<!ATTLIST phenotype

    phenotype_id CDATA #REQUIRED

    submitter_id CDATA #REQUIRED

    create_date CDATA #IMPLIED

    modify_date CDATA #IMPLIED
```

```
>

<!ELEMENT phenotype_description (phenotype_type , phenotype_condition* , phenotype_probability?)>

<!ELEMENT phenotype_type (#PCDATA)>

<!ELEMENT phenotype_condition (#PCDATA)>

<!ELEMENT phenotype_probability (#PCDATA)>

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

<!ELEMENT omics_annotation (omics_type , omics_description+ , omics_dbref*)>

<!-- omics_type element comment

scope of omics_type are: transcriptomics, proteomics, metabolomics, signalomics, organomics and any other string

-->

<!ELEMENT omics_type (#PCDATA)>

<!ELEMENT omics_description (omics_material, omics_condition*, omics_expression_probability?)>

<!ELEMENT omics_material (#PCDATA)>

<!ELEMENT omics_condition (#PCDATA)>

<!ELEMENT omics_expression_probability (#PCDATA)>

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

<!ELEMENT environmental_condition (expression_condition*)>

<!ELEMENT expression_condition (expression_condition_description+ , expression_condition_probability? , omics_annotation*)>

<!ELEMENT expression_condition_description (#PCDATA)>

<!ELEMENT expression_condition_probability (#PCDATA)>

<!ELEMENT clinical_annotation (disease* , clinical_observation* , clinical_annotation_dbref*)>

<!ELEMENT disease (disease_description* , disease_epidemiology* , disease_dbref*)>

<!ATTLIST disease

    disease_id    CDATA #REQUIRED

    submitter_id  CDATA #REQUIRED

    create_date   CDATA #IMPLIED

    modify_date   CDATA #IMPLIED

>

<!ELEMENT disease_description (name , synonym* , definition_concept? , classification* , etiology* , laboratory_findings* ,
pathological_findings* ,

                    symptoms* , diagnostic_criteria* , therapy* , complication* , prophylaxes*)>
```

<!ELEMENT name (#PCDATA)>

<!ELEMENT synonym (#PCDATA)>

<!ELEMENT definition_concept (#PCDATA)>

<!ELEMENT classification (#PCDATA)>

<!ELEMENT etiology (etiology_description , etiology_condition* , etiology_expression_probability?)>

<!ELEMENT etiology_description (#PCDATA)>

<!ELEMENT etiology_condition (#PCDATA)>

<!ELEMENT etiology_expression_probability (#PCDATA)>

<!ELEMENT laboratory_findings (laboratory_findings_type , laboratory_findings_description , laboratory_findings_condition* , laboratory_findings_expression_probability?)>

<!ELEMENT laboratory_findings_type (#PCDATA)>

<!ELEMENT laboratory_findings_description (#PCDATA)>

<!ELEMENT laboratory_findings_condition (#PCDATA)>

<!ELEMENT laboratory_findings_expression_probability (#PCDATA)>

<!ELEMENT pathological_findings (pathological_findings_description , pathological_findings_condition* , pathological_findings_expression_probability?)>

<!ELEMENT pathological_findings_description (#PCDATA)>

<!ELEMENT pathological_findings_condition (#PCDATA)>

<!ELEMENT pathological_findings_expression_probability (#PCDATA)>

<!ELEMENT symptoms (symptoms_description , symptoms_condition* , symptoms_expression_probability?)>

<!ELEMENT symptoms_description (#PCDATA)>

<!ELEMENT symptoms_condition (#PCDATA)>

<!ELEMENT symptoms_expression_probability (#PCDATA)>

<!ELEMENT diagnostic_criteria (diagnostic_standard , diagnostic_modify* , diagnostic_differential*)>

<!ELEMENT diagnostic_modify (#PCDATA)>

<!ELEMENT diagnostic_standard (#PCDATA)>

<!ELEMENT diagnostic_differential (#PCDATA)>

<!ELEMENT therapy (conservative* , surgery* , radiation*)>

<!ELEMENT conservative (pharmaceutical* , physical*)>

<!ELEMENT pharmaceutical (responder_sideffects*)>

<!ELEMENT responder_sideffects (responder_sideffects_description , responder_sideffects_causer* , responder_sideffects_condition* , responder_sideffects_expression_probability?)>

<!ELEMENT responder_sideffects_causer (#PCDATA)>

<!ELEMENT responder_sideffects_description (#PCDATA)>

<!ELEMENT responder_sideffects_condition (#PCDATA)>

<!ELEMENT responder_sideffects_expression_probability (#PCDATA)>

<!ELEMENT physical (#PCDATA)>

<!ELEMENT surgery (#PCDATA)>

<!ELEMENT radiation (#PCDATA)>

<!ELEMENT complication (complication_description , complication_condition* , complication_expression_probability?)>

<!ELEMENT complication_description (#PCDATA)>

<!ELEMENT complication_condition (#PCDATA)>

<!ELEMENT complication_expression_probability (#PCDATA)>

<!ELEMENT prophylaxes (prophylaxes_description , prophylaxes_condition* , prophylaxes_expression_probability?)>

<!ELEMENT prophylaxes_description (#PCDATA)>

<!ELEMENT prophylaxes_condition (#PCDATA)>

<!ELEMENT prophylaxes_expression_probability (#PCDATA)>

<!ELEMENT disease_epidemiology (striking_age* , striking_body_area* , striking_land_area* , laterality? , differences* , prognosis* , etiology_expression_probability* ,

laboratory_findings_expression_probability* , symptoms_expression_probability* , prophylaxes_expression_probability* ,

responder_sideffects_expression_probability* , pathological_findings_expression_probability* , complication_expression_probability*)>

<!ELEMENT striking_age (#PCDATA)>

<!ELEMENT striking_body_area (#PCDATA)>

<!ELEMENT striking_land_area (#PCDATA)>

<!-- laterality element comment

scope of laterality are: hemi, bi, right, left, other

-->

<!ELEMENT laterality (#PCDATA)>

<!ELEMENT prognosis (#PCDATA)>

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

<!ELEMENT clinical_observation (subjective_findings* , objective_findings* , assessment* , plan* , clinical_observation_dbref*)>

<!ELEMENT subjective_findings (symptoms)>

<!ELEMENT objective_findings (laboratory_findings* , pathological_findings* , complication* , family_history*)>

<!ELEMENT family_history (family_history_description* , family_history_dbref*)>

<!ELEMENT family_history_description (family_member* , relation_structure?)>

<!ELEMENT family_member (personal_info+ , phenotype* , clinical_annotation*)>

<!ELEMENT relation_structure (#PCDATA)>

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

<!ELEMENT assessment (assessment_result* , assessment_dbref*)>

<!ELEMENT assessment_result (#PCDATA)>

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

<!ELEMENT plan (therapy)>

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

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

Annex B
(normative)

XML schema of GSVML

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <!--====================== root element ======================-->
  <xs:element name="gsvml">
    <xs:annotation>
      <xs:documentation>Genomic Sequence Variation Markup Language</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="variation_data" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="direct_annotation" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="indirect_annotation" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <!--======== variation data ==========-->
  <xs:element name="variation_data">
    <xs:annotation>
      <xs:documentation>variation data</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="variation_type"/>
        <xs:element ref="location"/>
        <xs:element ref="variation_att" maxOccurs="unbounded"/>
        <xs:element ref="source" minOccurs="0" maxOccurs="unbounded"/>
```

```
            <xs:element ref="variation_dbref" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="variation_type">
    <xs:annotation>
        <xs:documentation>type of variation</xs:documentation>
    </xs:annotation>
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:enumeration value="SNP"/>
            <xs:enumeration value="rSNP"/>
            <xs:enumeration value="cSNP"/>
            <xs:enumeration value="iSNP"/>
            <xs:enumeration value="uSNP"/>
            <xs:enumeration value="gSNP"/>
            <xs:enumeration value="RFLP"/>
            <xs:enumeration value="MS"/>
            <xs:enumeration value="STRP"/>
            <xs:enumeration value="VNTR"/>
            <xs:enumeration value="Insertion"/>
            <xs:enumeration value="Deletion"/>
            <xs:enumeration value="Sustitution"/>
            <xs:enumeration value="Other"/>
            <xs:enumeration value=""/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="location">
    <xs:annotation>
```

```
      <xs:documentation>location of the variation</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="chromosome_number"/>
        <xs:element ref="position"/>
        <xs:element ref="map" minOccurs="0"/>
        <xs:element ref="orientation"/>
        <xs:element ref="ass_gene" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="location_dbref" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="chromosome_number" type="xs:string">
    <xs:annotation>
      <xs:documentation>the number of the chromosome</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="position" type="xs:string">
    <xs:annotation>
      <xs:documentation>position of the variation in the chromosome</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="map">
    <xs:annotation>
      <xs:documentation>chromosome map on which the variation is</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="orientation">
    <xs:annotation>
```

```
      <xs:documentation>chromosome orientation on which the variation is</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="ass_gene">
    <xs:annotation>
      <xs:documentation>associated gene</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="ass_gene_name"/>
        <xs:element ref="ass_gene_structure" minOccurs="0"/>
        <xs:element ref="aminoacid_substitution" minOccurs="0"/>
        <xs:element ref="codon_substitution" minOccurs="0"/>
        <xs:element ref="codon_position" minOccurs="0"/>
        <xs:element ref="ass_gene_symbol" minOccurs="0"/>
        <xs:element ref="ass_gene_alias" minOccurs="0"/>
        <xs:element ref="ass_gene_product" minOccurs="0"/>
        <xs:element ref="ass_gene_evidence_type" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="changed_motif" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="changed_motif_name" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="changed_splice_site" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="splice_variant_number" minOccurs="0"/>
        <xs:element ref="ass_gene_dbref" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="ass_gene_name" type="xs:string">
    <xs:annotation>
      <xs:documentation>gene name</xs:documentation>
    </xs:annotation>
```

```
</xs:element>
<xs:element name="ass_gene_structure">
  <xs:annotation>
    <xs:documentation>category of gene structure e.g. exon, intron</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="aminoacid_substitution">
  <xs:annotation>
    <xs:documentation>aminoacid sustitution generated by variation</xs:documentation>
  </xs:annotation>
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="A"/>
      <xs:enumeration value="R"/>
      <xs:enumeration value="N"/>
      <xs:enumeration value="D"/>
      <xs:enumeration value="C"/>
      <xs:enumeration value="Q"/>
      <xs:enumeration value="E"/>
      <xs:enumeration value="G"/>
      <xs:enumeration value="H"/>
      <xs:enumeration value="I"/>
      <xs:enumeration value="L"/>
      <xs:enumeration value="K"/>
      <xs:enumeration value="M"/>
      <xs:enumeration value="F"/>
      <xs:enumeration value="P"/>
      <xs:enumeration value="S"/>
      <xs:enumeration value="W"/>
      <xs:enumeration value="T"/>
```

```
            <xs:enumeration value="Y"/>
            <xs:enumeration value="V"/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="codon_substitution">
    <xs:annotation>
        <xs:documentation>codon substitution generated by variation</xs:documentation>
    </xs:annotation>
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:enumeration value="A"/>
            <xs:enumeration value="T"/>
            <xs:enumeration value="G"/>
            <xs:enumeration value="C"/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="codon_position" type="xs:string">
    <xs:annotation>
        <xs:documentation>codon position</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="ass_gene_symbol" type="xs:string">
    <xs:annotation>
        <xs:documentation>gene symbol</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="ass_gene_alias" type="xs:string">
    <xs:annotation>
```

```
        <xs:documentation>gene alias</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="ass_gene_product" type="xs:string">
      <xs:annotation>
        <xs:documentation>gene product</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="ass_gene_evidence_type" type="xs:string">
      <xs:annotation>
        <xs:documentation>gene type e.g. functional gene, predicted EST, computational gene, Pseudogene</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="changed_motif" type="xs:boolean">
      <xs:annotation>
        <xs:documentation>motif change exists or not</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="changed_motif_name" type="xs:string">
      <xs:annotation>
        <xs:documentation>name of motif</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="changed_splice_site" type="xs:boolean">
      <xs:annotation>
        <xs:documentation>splice site change exist or not</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="splice_variant_number">
      <xs:annotation>
```

```
        <xs:documentation>number of splice variant and refSeq</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="refSeq_number" maxOccurs="unbounded"/>
        </xs:sequence>
        <xs:attribute name="number" type="xs:integer"/>
      </xs:complexType>
    </xs:element>
    <xs:element name="refSeq_number" type="xs:string">
      <xs:annotation>
        <xs:documentation>reference sequence number</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="ass_gene_dbref" type="dbref">
      <xs:annotation>
        <xs:documentation>database reference of the associated gene</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="location_dbref" type="dbref">
      <xs:annotation>
        <xs:documentation>database reference of location</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="variation_att">
      <xs:annotation>
        <xs:documentation>attribute information of the variation</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
```

```
            <xs:element ref="molecular_type"/>
            <xs:element ref="allele" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="length"/>
            <xs:element ref="f5sequence" minOccurs="0"/>
            <xs:element ref="f3sequence" minOccurs="0"/>
            <xs:element ref="validation_status" minOccurs="0"/>
            <xs:element ref="success_rate" minOccurs="0"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="molecular_type" type="xs:string">
    <xs:annotation>
        <xs:documentation>type of molecule  e.g. DNA, RNA</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="allele" type="xs:string">
    <xs:annotation>
        <xs:documentation>observed allele</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="length" type="xs:float">
    <xs:annotation>
        <xs:documentation>sequence length including franking sequence</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="f5sequence" type="xs:string">
    <xs:annotation>
        <xs:documentation>5' flanking sequence</xs:documentation>
    </xs:annotation>
</xs:element>
```

```
<xs:element name="f3sequence" type="xs:string">
  <xs:annotation>
    <xs:documentation>3' flanking sequence</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="validation_status">
  <xs:annotation>
    <xs:documentation>status of validation as (Proven, Suspected)</xs:documentation>
  </xs:annotation>
  <xs:simpleType>
    <xs:union>
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="Proven"/>
          <xs:enumeration value="Suspected"/>
        </xs:restriction>
      </xs:simpleType>
      <xs:simpleType>
        <xs:restriction base="xs:string"/>
      </xs:simpleType>
    </xs:union>
  </xs:simpleType>
</xs:element>
<xs:element name="success_rate" type="xs:float">
  <xs:annotation>
    <xs:documentation>certainty of variation information</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="source">
  <xs:annotation>
```

```
      <xs:documentation>source of the sequence variation</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="source_release_date" minOccurs="0"/>
        <xs:element ref="source_modify_date" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="source_rawdata" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="source_dbref" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="source_release_date" type="xs:dateTime">
    <xs:annotation>
      <xs:documentation>date released</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="source_modify_date" type="xs:dateTime">
    <xs:annotation>
      <xs:documentation>date modified</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="source_rawdata">
    <xs:annotation>
      <xs:documentation>rawdatum of the source</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="source_dbref" type="dbref">
    <xs:annotation>
      <xs:documentation>database reference of source</xs:documentation>
    </xs:annotation>
```

```
</xs:element>
<xs:element name="variation_dbref" type="dbref">
   <xs:annotation>
      <xs:documentation>database references of the variation</xs:documentation>
   </xs:annotation>
</xs:element>
<!--======= direct annotation ========-->
<xs:element name="direct_annotation">
   <xs:annotation>
      <xs:documentation>direct annotation of variation data</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="whole_genome_sequence" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="mendelian_segregate" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="homozygote_detect" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="somatic_mutation" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="experiment_analysis" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="epidemiology" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="var_ann_misc" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
   </xs:complexType>
</xs:element>
<xs:element name="whole_genome_sequence">
   <xs:annotation>
      <xs:documentation>whole genome sequence of the datum</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="mendelian_segregate" type="xs:string">
   <xs:annotation>
```

```
        <xs:documentation>known mendelization</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="homozygote_detect">
      <xs:annotation>
        <xs:documentation>homozygote individuals observation in sample</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="homo"/>
          <xs:enumeration value="hetero"/>
          <xs:enumeration value="unknown"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="somatic_mutation" type="xs:string">
      <xs:annotation>
        <xs:documentation>known somatic mutation</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="experiment_analysis">
      <xs:annotation>
        <xs:documentation>explanation of the experimental and the analysis</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="variation_identify" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="variation_characterize" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
```

```
</xs:element>
<xs:element name="variation_identify">
  <xs:annotation>
    <xs:documentation>information to identify the variation</xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="experimental_assay_result"/>
      <xs:element ref="experimental_assay_description"/>
      <xs:element ref="experimental_assay_parameter" maxOccurs="unbounded"/>
      <xs:element ref="experimental_assay_dbref" minOccurs="0" maxOccurs="unbounded"/>
      <xs:choice>
        <xs:element ref="publication" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="submitter" minOccurs="0" maxOccurs="unbounded"/>
      </xs:choice>
      <xs:element ref="pcr" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="experimental_assay_id" use="required"/>
    <xs:attribute name="submitter_id" use="required"/>
    <xs:attribute name="create_date"/>
    <xs:attribute name="modify_date"/>
  </xs:complexType>
</xs:element>
<xs:element name="experimental_assay_result">
  <xs:annotation>
    <xs:documentation>result of the experimental assay</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="experimental_assay_description">
  <xs:annotation>
```

```
        <xs:documentation>description of the experimental assay</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="experimental_assay_parameter" type="xs:string">
      <xs:annotation>
        <xs:documentation>parameter of the experimental assay</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="experimental_assay_dbref" type="dbref">
      <xs:annotation>
        <xs:documentation>database reference information</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="publication">
      <xs:annotation>
        <xs:documentation>publication of the experiment</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="title" minOccurs="0"/>
          <xs:element ref="author" minOccurs="0"/>
          <xs:element ref="journal" minOccurs="0"/>
          <xs:element ref="volume" minOccurs="0"/>
          <xs:element ref="supplement" minOccurs="0"/>
          <xs:element ref="issue" minOccurs="0"/>
          <xs:element ref="issue_supplement" minOccurs="0"/>
          <xs:element ref="pages" minOccurs="0"/>
          <xs:element ref="year" minOccurs="0"/>
          <xs:element ref="publication_status" minOccurs="0"/>
          <xs:element ref="mesh_term" minOccurs="0"/>
```

```
            <xs:element ref="publication_dbref" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="submitter" minOccurs="0"/>
        </xs:sequence>
        <xs:attribute name="publication_id" use="required"/>
        <xs:attribute name="submitter_id" use="required"/>
        <xs:attribute name="create_date"/>
        <xs:attribute name="modify_date"/>
    </xs:complexType>
</xs:element>
<xs:element name="title" type="xs:string">
    <xs:annotation>
        <xs:documentation>title of the publication</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="author" type="xs:string">
    <xs:annotation>
        <xs:documentation>author of the publication</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="journal" type="xs:string">
    <xs:annotation>
        <xs:documentation>journal of the publication</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="volume" type="xs:string">
    <xs:annotation>
        <xs:documentation>volume of the publication</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="supplement">
```

```
    <xs:annotation>
      <xs:documentation>supplement of the publication</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="issue" type="xs:string">
    <xs:annotation>
      <xs:documentation>issue of the publication</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="issue_supplement">
    <xs:annotation>
      <xs:documentation>issue supplement of the publication</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="pages" type="xs:string">
    <xs:annotation>
      <xs:documentation>page of the publication</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="year" type="xs:string">
    <xs:annotation>
      <xs:documentation>year of the publication</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="publication_status" type="xs:string">
    <xs:annotation>
      <xs:documentation>status of the publication</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="mesh_term" type="xs:string">
```

```
    <xs:annotation>
      <xs:documentation>mesh term of the publication</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="publication_dbref" type="dbref">
    <xs:annotation>
      <xs:documentation>database references of the publications</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="submitter">
    <xs:annotation>
      <xs:documentation>submitter of the publication</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="submitter_name"/>
        <xs:element ref="address" minOccurs="0"/>
        <xs:element ref="email" minOccurs="0"/>
        <xs:element ref="tel" minOccurs="0"/>
        <xs:element ref="fax" minOccurs="0"/>
        <xs:element ref="institution" minOccurs="0"/>
        <xs:element ref="laboratory" minOccurs="0"/>
        <xs:element ref="submitter_dbref" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="publication" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="submitter_id" use="required"/>
      <xs:attribute name="create_date"/>
      <xs:attribute name="modify_date"/>
    </xs:complexType>
  </xs:element>
```

```
<xs:element name="submitter_name" type="xs:string">
  <xs:annotation>
    <xs:documentation>name of the submitter</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="address" type="xs:string">
  <xs:annotation>
    <xs:documentation>address of the submitter</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="email" type="xs:string">
  <xs:annotation>
    <xs:documentation>email of the submitter</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="tel" type="xs:string">
  <xs:annotation>
    <xs:documentation>telephone of the submitter</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="fax" type="xs:string">
  <xs:annotation>
    <xs:documentation>Fax of the submitter</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="institution" type="xs:string">
  <xs:annotation>
    <xs:documentation>Institution of the submitter</xs:documentation>
  </xs:annotation>
</xs:element>
```

```
<xs:element name="laboratory" type="xs:string">
   <xs:annotation>
      <xs:documentation>Laboratory of the submitter</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="submitter_dbref" type="dbref">
   <xs:annotation>
      <xs:documentation>database references of the submitter</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="pcr">
   <xs:annotation>
      <xs:documentation>PCR (polymerase chain reaction)</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="pcr_confirmed" minOccurs="0"/>
         <xs:element ref="pcr_primer" minOccurs="0"/>
         <xs:element ref="pcr_product" minOccurs="0"/>
         <xs:element ref="pcr_profile" minOccurs="0"/>
      </xs:sequence>
   </xs:complexType>
</xs:element>
<xs:element name="pcr_confirmed">
   <xs:annotation>
      <xs:documentation>artifact verification  e.g. variation found on repeat PCR sample</xs:documentation>
   </xs:annotation>
   <xs:simpleType>
      <xs:union>
         <xs:simpleType>
```

```
            <xs:restriction base="xs:string">
              <xs:enumeration value="Yes"/>
              <xs:enumeration value="No"/>
              <xs:enumeration value="Unknown"/>
            </xs:restriction>
          </xs:simpleType>
          <xs:simpleType>
            <xs:restriction base="xs:string"/>
          </xs:simpleType>
        </xs:union>
    </xs:simpleType>
  </xs:element>
  <xs:element name="pcr_primer" type="xs:string">
    <xs:annotation>
      <xs:documentation>primer sequence</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="pcr_product" type="xs:string">
    <xs:annotation>
      <xs:documentation>PCR product  e.g. single band, multi band</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="pcr_profile">
    <xs:annotation>
      <xs:documentation>PCR profile</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="variation_characterize">
    <xs:annotation>
      <xs:documentation>characterization of the variation datum</xs:documentation>
```

```
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="genetic_statistics"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="genetic_statistics">
    <xs:annotation>
      <xs:documentation>genetic statistics</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="method" maxOccurs="unbounded"/>
        <xs:element ref="p-value" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="link_dis_index" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="descend_identity" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="maximum_lod_score" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="method">
    <xs:annotation>
      <xs:documentation>statistical method</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence maxOccurs="unbounded">
        <xs:element ref="method_name"/>
        <xs:element ref="method_description" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="method_url" minOccurs="0" maxOccurs="unbounded"/>
```

```
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="method_name">
      <xs:annotation>
        <xs:documentation>statistical method name</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
        <xs:union>
          <xs:simpleType>
            <xs:restriction base="xs:string">
              <xs:enumeration value="sib pair"/>
              <xs:enumeration value="linkage disequilibrium"/>
              <xs:enumeration value="association study"/>
            </xs:restriction>
          </xs:simpleType>
          <xs:simpleType>
            <xs:restriction base="xs:string"/>
          </xs:simpleType>
        </xs:union>
      </xs:simpleType>
    </xs:element>
    <xs:element name="method_description">
      <xs:annotation>
        <xs:documentation>description or explanation of the method</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="method_url" type="xs:string">
      <xs:annotation>
        <xs:documentation>URL of the method</xs:documentation>
```

```
    </xs:annotation>
  </xs:element>
  <xs:element name="p-value" type="xs:float">
    <xs:annotation>
      <xs:documentation>p value for significance (Association study)</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="link_dis_index">
    <xs:annotation>
      <xs:documentation>linkage disequilibrium index for LD test</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence maxOccurs="unbounded">
        <xs:element ref="d" minOccurs="0"/>
        <xs:element ref="d_prime" minOccurs="0"/>
        <xs:element ref="r-square" minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="d" type="xs:float">
    <xs:annotation>
      <xs:documentation>d value for Linquage Disequilibrium Test</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="d_prime" type="xs:float">
    <xs:annotation>
      <xs:documentation>d prime for LD Test</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="r-square" type="xs:float">
```

```
    <xs:annotation>
      <xs:documentation>r square for LD test</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="descend_identity">
    <xs:annotation>
      <xs:documentation>Identity By Descent (IBD) for Sib-Pair Test</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="di_value"/>
        <xs:element ref="di_probability"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="di_value" type="xs:float">
    <xs:annotation>
      <xs:documentation>IBD value</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="di_probability">
    <xs:annotation>
      <xs:documentation>IBD probability</xs:documentation>
    </xs:annotation>
    <xs:simpleType>
      <xs:restriction base="xs:float">
        <xs:minInclusive value="0"/>
        <xs:maxInclusive value="1"/>
      </xs:restriction>
    </xs:simpleType>
```

```
</xs:element>
<xs:element name="maximum_lod_score" type="xs:float">
   <xs:annotation>
      <xs:documentation>maximum LOD score for sib-pair analysis</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="epidemiology">
   <xs:annotation>
      <xs:documentation>epidemiology of the disease and associated gene</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="ass_gene" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="disease_epidemiology" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="population" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="frequency" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
   </xs:complexType>
</xs:element>
<xs:element name="disease_epidemiology">
   <xs:annotation>
      <xs:documentation>epidemiology of the disease</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="striking_age" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="striking_body_area" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="striking_land_area" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="laterality" minOccurs="0"/>
         <xs:element ref="differences" minOccurs="0" maxOccurs="unbounded"/>
```

```
            <xs:element ref="prognosis" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="etiology_expression_probability" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="laboratory_findings_expression_probability" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="sysmptoms_expression_probability" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="prophylaxes_expression_probability" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="responder_sideffects_expression_probability" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="pathological_findings_expression_probability" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="complication_expression_probability" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="striking_age" type="xs:string">
    <xs:annotation>
        <xs:documentation>striking age of the disease</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="striking_body_area">
    <xs:annotation>
        <xs:documentation>striking body area of the disease</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="striking_land_area">
    <xs:annotation>
        <xs:documentation>striking land area of the disease</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="laterality">
    <xs:annotation>
        <xs:documentation>laterality of the disease  e.g. hemilateral, bilateral</xs:documentation>
    </xs:annotation>
```

```
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="hemi"/>
        <xs:enumeration value="bi"/>
        <xs:enumeration value="right"/>
        <xs:enumeration value="left"/>
        <xs:enumeration value="other"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="prognosis">
    <xs:annotation>
      <xs:documentation>prognosis of the disease</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="population">
    <xs:annotation>
      <xs:documentation>population of the variation</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="population_description"/>
        <xs:element ref="organism" minOccurs="0"/>
        <xs:element ref="differences" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="population_parameter" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="sample_size"/>
        <xs:element ref="population_misc" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="population_dbref" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="population_id" use="required"/>
```

```
        <xs:attribute name="submitter_id" use="required"/>
        <xs:attribute name="create_date"/>
        <xs:attribute name="modify_date"/>
      </xs:complexType>
    </xs:element>
    <xs:element name="population_description">
      <xs:annotation>
        <xs:documentation>description of the population</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="organism" type="xs:string">
      <xs:annotation>
        <xs:documentation>organism</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="differences">
      <xs:annotation>
        <xs:documentation>statistical differences</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="race" minOccurs="0"/>
          <xs:element ref="gender" minOccurs="0"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="race" type="xs:string">
      <xs:annotation>
        <xs:documentation>racial difference</xs:documentation>
      </xs:annotation>
```

```
</xs:element>
<xs:element name="gender">
  <xs:annotation>
    <xs:documentation>gender difference</xs:documentation>
  </xs:annotation>
  <xs:simpleType>
    <xs:union>
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="male"/>
          <xs:enumeration value="female"/>
          <xs:enumeration value="unknown"/>
          <xs:enumeration value="other"/>
        </xs:restriction>
      </xs:simpleType>
      <xs:simpleType>
        <xs:restriction base="xs:string"/>
      </xs:simpleType>
    </xs:union>
  </xs:simpleType>
</xs:element>
<xs:element name="population_parameter" type="xs:string">
  <xs:annotation>
    <xs:documentation>parameter of population</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="sample_size" type="xs:float">
  <xs:annotation>
    <xs:documentation>sample size of population</xs:documentation>
  </xs:annotation>
```

```
</xs:element>
<xs:element name="population_misc">
  <xs:annotation>
    <xs:documentation>population miscellaneous</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="population_dbref" type="dbref">
  <xs:annotation>
    <xs:documentation>database references of the population</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="frequency">
  <xs:annotation>
    <xs:documentation>frequency of the variation</xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="haplotype" minOccurs="0"/>
      <xs:element ref="haplotype_frequency" minOccurs="0"/>
      <xs:element ref="allele" minOccurs="0"/>
      <xs:element ref="allele_frequency" minOccurs="0"/>
      <xs:element ref="genotype" minOccurs="0"/>
      <xs:element ref="genotype_frequency" minOccurs="0"/>
    </xs:sequence>
    <xs:attribute name="frequency_id" use="required"/>
    <xs:attribute name="submitter_id" use="required"/>
    <xs:attribute name="population_id" use="required"/>
    <xs:attribute name="assay_id" use="required"/>
    <xs:attribute name="publication_id" use="required"/>
    <xs:attribute name="create_date"/>
```

```
        <xs:attribute name="modify_date"/>
    </xs:complexType>
</xs:element>
<xs:element name="haplotype" type="xs:string">
    <xs:annotation>
        <xs:documentation>haplotype</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="haplotype_frequency" type="xs:float">
    <xs:annotation>
        <xs:documentation>frequency of the haplotype</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="allele_frequency" type="xs:float">
    <xs:annotation>
        <xs:documentation>frequency of the observed allele</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="genotype" type="xs:string">
    <xs:annotation>
        <xs:documentation>genotype</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="genotype_frequency" type="xs:float">
    <xs:annotation>
        <xs:documentation>frequency of the genotype</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="var_ann_misc">
    <xs:annotation>
```

```
      <xs:documentation>variation annotation miscellanious</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="var_ann_misc_description"/>
        <xs:element ref="var_ann_misc_dbref" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="var_ann_misc_id" use="required"/>
      <xs:attribute name="submitter_id" use="required"/>
      <xs:attribute name="create_date"/>
      <xs:attribute name="modify_date"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="var_ann_misc_description">
    <xs:annotation>
      <xs:documentation>description of the variation annotation miscellaneous</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="var_ann_misc_dbref" type="dbref">
    <xs:annotation>
      <xs:documentation>database references of the variation annotation miscellanious</xs:documentation>
    </xs:annotation>
  </xs:element>
  <!--======= indirect_annotation ==========-->
  <xs:element name="indirect_annotation">
    <xs:annotation>
      <xs:documentation>indirect annotation of variation data</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
```

```
            <xs:element ref="personal_info" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="phenotype" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="omics_annotation" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="environmental_condition" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="clinical_annotation" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="personal_info">
    <xs:annotation>
        <xs:documentation>personal information of variation data</xs:documentation>
    </xs:annotation>
    <xs:complexType>
        <xs:sequence>
            <xs:element ref="personal_description"/>
            <xs:element ref="personal_dbref" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="personal_description">
    <xs:annotation>
        <xs:documentation>description of the personal information  e.g. Name, Job</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="personal_dbref" type="dbref">
    <xs:annotation>
        <xs:documentation>database reference of personal information</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="phenotype">
```

```
    <xs:annotation>
      <xs:documentation>phenotype of the sequence variation</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="phenotype_description" maxOccurs="unbounded"/>
        <xs:element ref="phenotype_dbref" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="phenotype_id" use="required"/>
      <xs:attribute name="submitter_id" use="required"/>
      <xs:attribute name="create_date"/>
      <xs:attribute name="modify_date"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="phenotype_description">
    <xs:annotation>
      <xs:documentation>description of the phenotype</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="phenotype_type"/>
        <xs:element ref="phenotype_condition" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="phenotype_probability" minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="phenotype_type">
    <xs:annotation>
      <xs:documentation>type of the phenotype</xs:documentation>
    </xs:annotation>
```

```
</xs:element>
<xs:element name="phenotype_condition">
   <xs:annotation>
      <xs:documentation>condition to express the phenotype</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="phenotype_probability">
   <xs:annotation>
      <xs:documentation>probability to express the phenotype on the conditions</xs:documentation>
   </xs:annotation>
   <xs:simpleType>
      <xs:restriction base="xs:float">
         <xs:minInclusive value="0"/>
         <xs:maxInclusive value="1"/>
      </xs:restriction>
   </xs:simpleType>
</xs:element>
<xs:element name="phenotype_dbref" type="dbref">
   <xs:annotation>
      <xs:documentation>database references of the phenotype</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="omics_annotation">
   <xs:annotation>
      <xs:documentation>annotation type of the omics</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="omics_type"/>
         <xs:element ref="omics_description" maxOccurs="unbounded"/>
```

```
            <xs:element ref="omics_dbref" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="omics_type">
    <xs:annotation>
        <xs:documentation>type of omics</xs:documentation>
    </xs:annotation>
    <xs:simpleType>
        <xs:union>
            <xs:simpleType>
                <xs:restriction base="xs:string">
                    <xs:enumeration value="genomics"/>
                    <xs:enumeration value="transcriptomics"/>
                    <xs:enumeration value="proteomics"/>
                    <xs:enumeration value="signalomics"/>
                    <xs:enumeration value="metabolomics"/>
                    <xs:enumeration value="organomics"/>
                </xs:restriction>
            </xs:simpleType>
            <xs:simpleType>
                <xs:restriction base="xs:string"/>
            </xs:simpleType>
        </xs:union>
    </xs:simpleType>
</xs:element>
<xs:element name="omics_description">
    <xs:annotation>
        <xs:documentation>description of the omics</xs:documentation>
    </xs:annotation>
```

```
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="omics_material"/>
        <xs:element ref="omics_condition" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="omics_expression_probability" minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="omics_material">
    <xs:annotation>
      <xs:documentation>material of the omics</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="omics_condition">
    <xs:annotation>
      <xs:documentation>condition of the omics</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="omics_expression_probability">
    <xs:annotation>
      <xs:documentation>expression probability of the omics</xs:documentation>
    </xs:annotation>
    <xs:simpleType>
      <xs:restriction base="xs:float">
        <xs:minInclusive value="0"/>
        <xs:maxInclusive value="1"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="omics_dbref" type="dbref">
```

```
    <xs:annotation>
      <xs:documentation>database reference of the omics</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="environmental_condition">
    <xs:annotation>
      <xs:documentation>environmental conditions of the expression</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="expression_condition" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="expression_condition">
    <xs:annotation>
      <xs:documentation>environmental condition of the expression</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="expression_condition_description" maxOccurs="unbounded"/>
        <xs:element ref="expression_condition_probability" minOccurs="0"/>
        <xs:element ref="omics_annotation" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="expression_condition_description">
    <xs:annotation>
      <xs:documentation>description of the expression condition</xs:documentation>
    </xs:annotation>
```

```
</xs:element>
<xs:element name="expression_condition_probability">
   <xs:annotation>
      <xs:documentation>probability of the expression condition</xs:documentation>
   </xs:annotation>
   <xs:simpleType>
      <xs:restriction base="xs:float">
         <xs:minInclusive value="0"/>
         <xs:maxInclusive value="1"/>
      </xs:restriction>
   </xs:simpleType>
</xs:element>
<xs:element name="clinical_annotation">
   <xs:annotation>
      <xs:documentation>clinical annotation of the sequence variation</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="disease" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="clinical_observation" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="clinical_annotation_dbref" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
   </xs:complexType>
</xs:element>
<xs:element name="disease">
   <xs:annotation>
      <xs:documentation>disease information</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
```

```
            <xs:element ref="disease_description" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="disease_epidemiology" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="disease_dbref" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
        <xs:attribute name="disease_id" use="required"/>
        <xs:attribute name="submitter_id" use="required"/>
        <xs:attribute name="create_date"/>
        <xs:attribute name="modify_date"/>
    </xs:complexType>
</xs:element>
<xs:element name="disease_description">
    <xs:annotation>
        <xs:documentation>description of the disease</xs:documentation>
    </xs:annotation>
    <xs:complexType>
        <xs:sequence>
            <xs:element ref="name"/>
            <xs:element ref="synonym" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="definition_concept" minOccurs="0"/>
            <xs:element ref="classification" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="etiology" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="laboratory_findings" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="pathological_findings" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="symptoms" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="diagnostic_criteria" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="therapy" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="complication" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="prophylaxes" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
```

```
</xs:element>
<xs:element name="name" type="xs:string">
   <xs:annotation>
      <xs:documentation>name of the disease</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="synonym" type="xs:string">
   <xs:annotation>
      <xs:documentation>synonym of the disease</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="definition_concept">
   <xs:annotation>
      <xs:documentation>definition or concept of the disease</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="classification">
   <xs:annotation>
      <xs:documentation>classification of the disease</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="etiology">
   <xs:annotation>
      <xs:documentation>etiology of the disease</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="etiology_description"/>
         <xs:element ref="etiology_condition" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="etiology_expression_probability" minOccurs="0"/>
```

```
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="etiology_description">
      <xs:annotation>
        <xs:documentation>description of the etiology and its mechanism</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="etiology_condition">
      <xs:annotation>
        <xs:documentation>conditions to express the etiology</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="etiology_expression_probability">
      <xs:annotation>
        <xs:documentation>expression probability of the etiology</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
        <xs:restriction base="xs:float">
          <xs:minInclusive value="0"/>
          <xs:maxInclusive value="1"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="laboratory_findings">
      <xs:annotation>
        <xs:documentation>laboratory findings of the disease</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
```

```
            <xs:element ref="laboratory_findings_type"/>
            <xs:element ref="laboratory_findings_description"/>
            <xs:element ref="laboratory_findings_condition" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="laboratory_findings_expression_probability" minOccurs="0"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="laboratory_findings_type" type="xs:string">
    <xs:annotation>
        <xs:documentation>type of laboratory findings  e.g. urine, blood, ECG, image</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="laboratory_findings_description">
    <xs:annotation>
        <xs:documentation>description of the laboratory findings</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="laboratory_findings_condition">
    <xs:annotation>
        <xs:documentation>condition of the laboratory findings</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:element name="laboratory_findings_expression_probability">
    <xs:annotation>
        <xs:documentation>expression probability of the laboratory findings</xs:documentation>
    </xs:annotation>
    <xs:simpleType>
        <xs:restriction base="xs:float">
            <xs:minInclusive value="0"/>
            <xs:maxInclusive value="1"/>
```

```
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="pathological_findings">
      <xs:annotation>
        <xs:documentation>pathological findings of the disease</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="pathological_findings_description"/>
          <xs:element ref="pathological_findings_condition" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="pathological_findings_expression_probability" minOccurs="0"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="pathological_findings_description">
      <xs:annotation>
        <xs:documentation>description of the pathological finding</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="pathological_findings_condition">
      <xs:annotation>
        <xs:documentation>condition of the pathological finding</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="pathological_findings_expression_probability">
      <xs:annotation>
        <xs:documentation>expression probability of the pathological findings</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
```

```
        <xs:restriction base="xs:float">
          <xs:minInclusive value="0"/>
          <xs:maxInclusive value="1"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="symptoms">
      <xs:annotation>
        <xs:documentation>laboratory findings of the disease</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="symptoms_description"/>
          <xs:element ref="symptoms_condition" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="symptoms_expression_probability" minOccurs="0"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="symptoms_description">
      <xs:annotation>
        <xs:documentation>description of the symptom</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="symptoms_condition">
      <xs:annotation>
        <xs:documentation>condition of the symptom</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="symptoms_expression_probability">
      <xs:annotation>
```

```
      <xs:documentation>expression probability of the symptom</xs:documentation>
    </xs:annotation>
    <xs:simpleType>
      <xs:restriction base="xs:float">
        <xs:minInclusive value="0"/>
        <xs:maxInclusive value="1"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="diagnostic_criteria">
    <xs:annotation>
      <xs:documentation>criteria of diagnosis for the disease</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="diagnostic_standard"/>
        <xs:element ref="diagnostic_modify" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="diagnostic_differential" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="diagnostic_standard">
    <xs:annotation>
      <xs:documentation>standard diagnostic criteria</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="diagnostic_modify">
    <xs:annotation>
      <xs:documentation>modified diagnostic criteria</xs:documentation>
    </xs:annotation>
```

```
</xs:element>
<xs:element name="diagnostic_differential">
   <xs:annotation>
      <xs:documentation>differential diagnosis use</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="therapy">
   <xs:annotation>
      <xs:documentation>therapy of the disease</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="conservative" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="surgery" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="radiation" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
   </xs:complexType>
</xs:element>
<xs:element name="conservative">
   <xs:annotation>
      <xs:documentation>conservative treatment</xs:documentation>
   </xs:annotation>
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="pharmaceutical" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="physical" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
   </xs:complexType>
</xs:element>
<xs:element name="pharmaceutical">
```

```
    <xs:annotation>
      <xs:documentation>pharmaceutical element</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="responder_sideffects" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="responder_sideffects">
    <xs:annotation>
      <xs:documentation>responder and/or side effects</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="responder_sideffects_description"/>
        <xs:element ref="responder_sideffects_causer" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="responder_sideffects_condition" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="responder_sideffects_expression_probability" minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="responder_sideffects_causer">
    <xs:annotation>
      <xs:documentation>causer of the responder and/or side effect</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="responder_sideffects_description">
    <xs:annotation>
      <xs:documentation>description of the responder and/or side effect</xs:documentation>
```

```
    </xs:annotation>
  </xs:element>
  <xs:element name="responder_sideffects_condition">
    <xs:annotation>
      <xs:documentation>condition of the responder and/or side effect</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="responder_sideffects_expression_probability">
    <xs:annotation>
      <xs:documentation>expression probability of the responder, the side effect</xs:documentation>
    </xs:annotation>
    <xs:simpleType>
      <xs:restriction base="xs:float">
        <xs:minInclusive value="0"/>
        <xs:maxInclusive value="1"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="physical">
    <xs:annotation>
      <xs:documentation>physical treatment element</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="surgery">
    <xs:annotation>
      <xs:documentation>surgical treatment</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="radiation">
    <xs:annotation>
```

```
        <xs:documentation>radiation therapy</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="complication">
      <xs:annotation>
        <xs:documentation>complications of the disease</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="complication_description"/>
          <xs:element ref="complication_condition" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="complication_expression_probability" minOccurs="0"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="complication_description">
      <xs:annotation>
        <xs:documentation>description of the complication</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="complication_condition">
      <xs:annotation>
        <xs:documentation>condition to express the complication</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="complication_expression_probability">
      <xs:annotation>
        <xs:documentation>probability when the complication is expressed</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
```

```
        <xs:restriction base="xs:float">
          <xs:minInclusive value="0"/>
          <xs:maxInclusive value="1"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="prophylaxes">
      <xs:annotation>
        <xs:documentation>prophylaxes of the disease</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="prophylaxes_description"/>
          <xs:element ref="prophylaxes_condition" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="prophylaxes_expression_probability" minOccurs="0"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="prophylaxes_description">
      <xs:annotation>
        <xs:documentation>description of the prophylaxis</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="prophylaxes_condition">
      <xs:annotation>
        <xs:documentation>condition of the prophylaxis</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="prophylaxes_expression_probability">
      <xs:annotation>
```

```
        <xs:documentation>expression probability of the prophylaxes</xs:documentation>
      </xs:annotation>
      <xs:simpleType>
        <xs:restriction base="xs:float">
          <xs:minInclusive value="0"/>
          <xs:maxInclusive value="1"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="disease_dbref" type="dbref">
      <xs:annotation>
        <xs:documentation>database referrences of the disease</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="clinical_observation">
      <xs:annotation>
        <xs:documentation>clinical observation of the persona</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="subjective_findings" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="objective_findings" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="assessment" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="plan" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="clinical_observation_dbref" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="subjective_findings">
      <xs:annotation>
```

```
      <xs:documentation>subjective findings in clinical observation</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="symptoms"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="objective_findings">
    <xs:annotation>
      <xs:documentation>objective findings in clinical observation</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="laboratory_findings" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="pathological_findings" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="complication" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="family_history" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="family_history">
    <xs:annotation>
      <xs:documentation>family history</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="family_history_description" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="family_history_dbref" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
```

```
    </xs:complexType>
  </xs:element>
  <xs:element name="family_history_description">
    <xs:annotation>
      <xs:documentation>description of the family history</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="family_member" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="relation_structure" minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="family_member">
    <xs:annotation>
      <xs:documentation>member of the family</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="personal_info" maxOccurs="unbounded"/>
        <xs:element ref="phenotype" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="clinical_annotation" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="relation_structure">
    <xs:annotation>
      <xs:documentation>relational structure of the family</xs:documentation>
    </xs:annotation>
  </xs:element>
```

```
<xs:element name="family_history_dbref" type="dbref">
  <xs:annotation>
    <xs:documentation>database reference of the family history</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="assessment">
  <xs:annotation>
    <xs:documentation>assessment of the clinical observation</xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="assessment_result" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="assessment_dbref" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="assessment_result">
  <xs:annotation>
    <xs:documentation>results of the assessment</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="assessment_dbref" type="dbref">
  <xs:annotation>
    <xs:documentation>database reference of the assessment</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="plan">
  <xs:annotation>
    <xs:documentation>plan of the clinical observation</xs:documentation>
  </xs:annotation>
```

```
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="therapy"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="clinical_observation_dbref" type="dbref">
    <xs:annotation>
      <xs:documentation>database reference of the clinical observation </xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="clinical_annotation_dbref" type="dbref">
    <xs:annotation>
      <xs:documentation>database reference of the clinical annotation</xs:documentation>
    </xs:annotation>
  </xs:element>
  <!--======== dbref complexType =========-->
  <xs:complexType name="dbref">
    <xs:annotation>
      <xs:documentation>database reference</xs:documentation>
    </xs:annotation>
    <xs:sequence>
      <xs:element ref="database_name"/>
      <xs:element ref="database_id" minOccurs="0"/>
      <xs:element ref="link_url" minOccurs="0"/>
      <xs:element ref="database_attbt" minOccurs="0"/>
      <xs:element ref="data_id"/>
      <xs:element ref="data_attbt" minOccurs="0"/>
      <xs:element ref="version" minOccurs="0"/>
    </xs:sequence>
```

```
</xs:complexType>
<xs:element name="database_name" type="xs:string">
   <xs:annotation>
      <xs:documentation>name of the database</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="database_id" type="xs:string">
   <xs:annotation>
      <xs:documentation>ID of the database</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="link_url" type="xs:string">
   <xs:annotation>
      <xs:documentation>URL of the database</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="database_attbt" type="xs:string">
   <xs:annotation>
      <xs:documentation>attribute of the database</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="data_id" type="xs:string">
   <xs:annotation>
      <xs:documentation>ID of the datum</xs:documentation>
   </xs:annotation>
</xs:element>
<xs:element name="data_attbt">
   <xs:annotation>
      <xs:documentation>attribute of the datum</xs:documentation>
   </xs:annotation>
```

```
      <xs:complexType/>
   </xs:element>
   <xs:element name="version" type="xs:string">
      <xs:annotation>
         <xs:documentation>version of the database</xs:documentation>
      </xs:annotation>
   </xs:element>
</xs:schema>
```

Annex C
(informative)

Basic reference works

C.1 Introduction

This annex gives the reference works for development of GSVML (see Clause 6) and a non-exhaustive list of the most generally applicable basic reference works. Details concerning currently valid ISO documents are available in the ISO catalogue. Copies can be obtained from the ISO national bodies. For specific subjects, the provisions of other, less generally applicable, documents will be relevant.

C.2 Use case analysis

C.2.1 General

We summarised typical use cases of genomic sequence variation data exchanging in human health. Six use cases for three types of criteria for the SNP data exchange are summarised as follows.

Type I: Clinical practice use

The first type of use case is clinical practice. In this use case, SNP data are exchanged amongst the facilities and the clinical specialists.

Genetic diagnosis or genetic counselling

For genetic diagnosis or genetic counselling, the SNP data are exchanged amongst MDs, lab technicians and counsellors. In this case, individual SNP data are sent/received with individual clinical data. For more advanced diagnosis, individual genomic data including -omics data are demanded.

Prescription derived from pharmacogenomics

For prescription derived from pharmacogenomics, the SNP data will not be exchanged in most cases. The exchange data will be the prescription, reasons and its annotations. In this case, individual SNP data are not sent/received with individual prescription results.

Gene therapy

For gene therapy, the SNP data are exchanged amongst hospitals, other facilities, MDs and patients. In this case, individual SNP data are sent/received with individual clinical data and individual genomic data.

Disease prevention based on the individual polymorphism

For disease prevention based on the individual polymorphism, the SNP data are exchanged between MDs and lab technicians. In this case, individual SNP data are sent/received with individual clinical data.

Type II: Clinical trial use

In the case of clinical trials, the SNP data are exchanged amongst hospitals, research institutes, MDs and the pharmaceutical company. In this case, individual SNP data are sent/received with individual clinical data and other data that are needed to specify the experiment. The clinical data required depend on the clinical phase. Early phases do not need many individuals but need many parameters, while in late phases the opposite is true.

Type III: Translational research use

In translational research, the SNP data are exchanged amongst hospitals, research institutes, MDs, researchers and the pharmaceutical company. In this case, individual SNP data are sent/received with individual clinical data along with other additional data that are needed to specify the experiment. The number of clinical data elements required will be several dozen, while the parameters for each individual are many.

C.2.2 Overview

Figure C.1 is the general use case of GSVML in the clinical scene.

Through GSVML, every actor can exchange data smoothly without requiring a change in their existing database schema. In the same way, the researchers can exchange their genomic sequence variation data without any pain.

As an example, in the case of genetic diagnosis, the individual SNP data are exchanged amongst facilities such as hospitals and medical laboratories. These data are also exchanged amongst persons such as MDs, laboratory analysts, counsellors and, in some cases, the patient him/herself. Here individual SNP data are encapsulated with the individual clinical data and his/her -omics data, in some cases, for further examination. To analyse this individual SNP data, the individual SNP data need to be compared with the database derived SNP data that have various types of data formats.

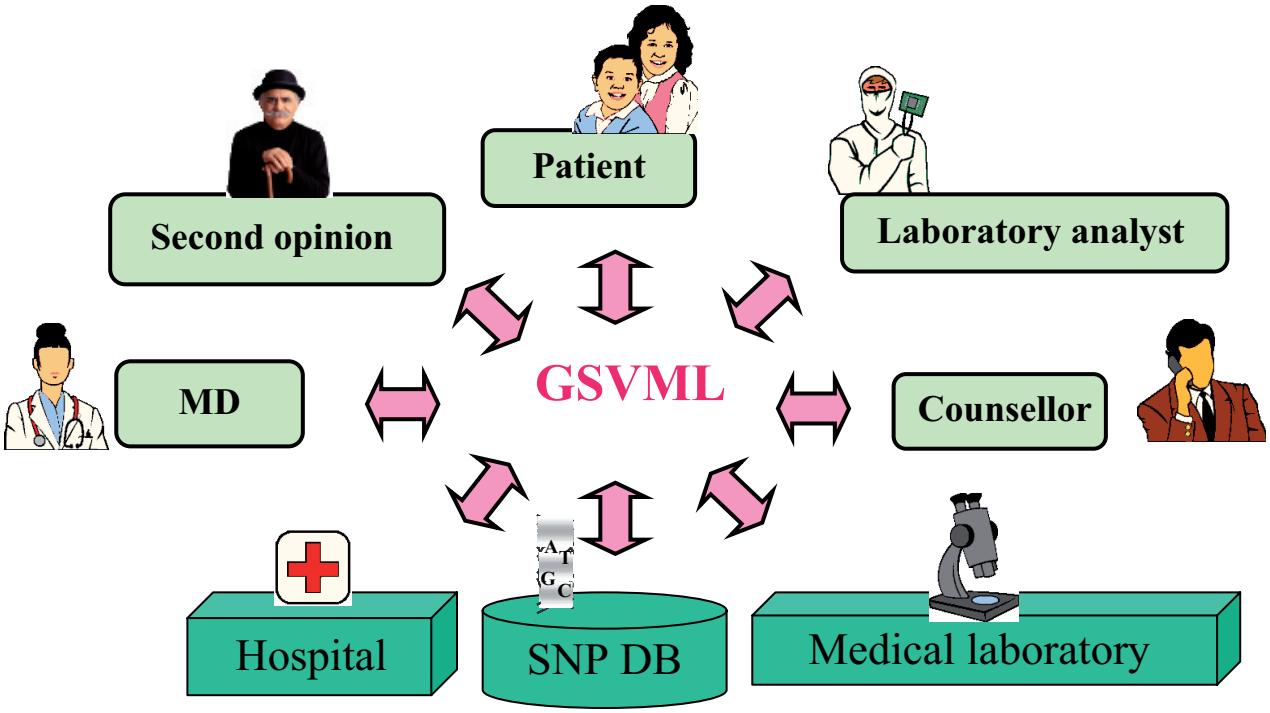


Figure C.1 — An outlined example of general use case

C.2.3 Use case of SNP analysis

Figure C.2 is an example work flow of the SNP analysis.

This is the case of the “Japanese National Millennium Project”.

This project tries to find the relevant SNPs or genes for the five lifestyle diseases.

This is the case for the diabetes mellitus.

In this case SNP data and additional information are exchanged amongst the facilities.

Here we have not only the SNP data but also the clinical data and SNP annotations as the clinical data, the -omics annotations and the environmental data.

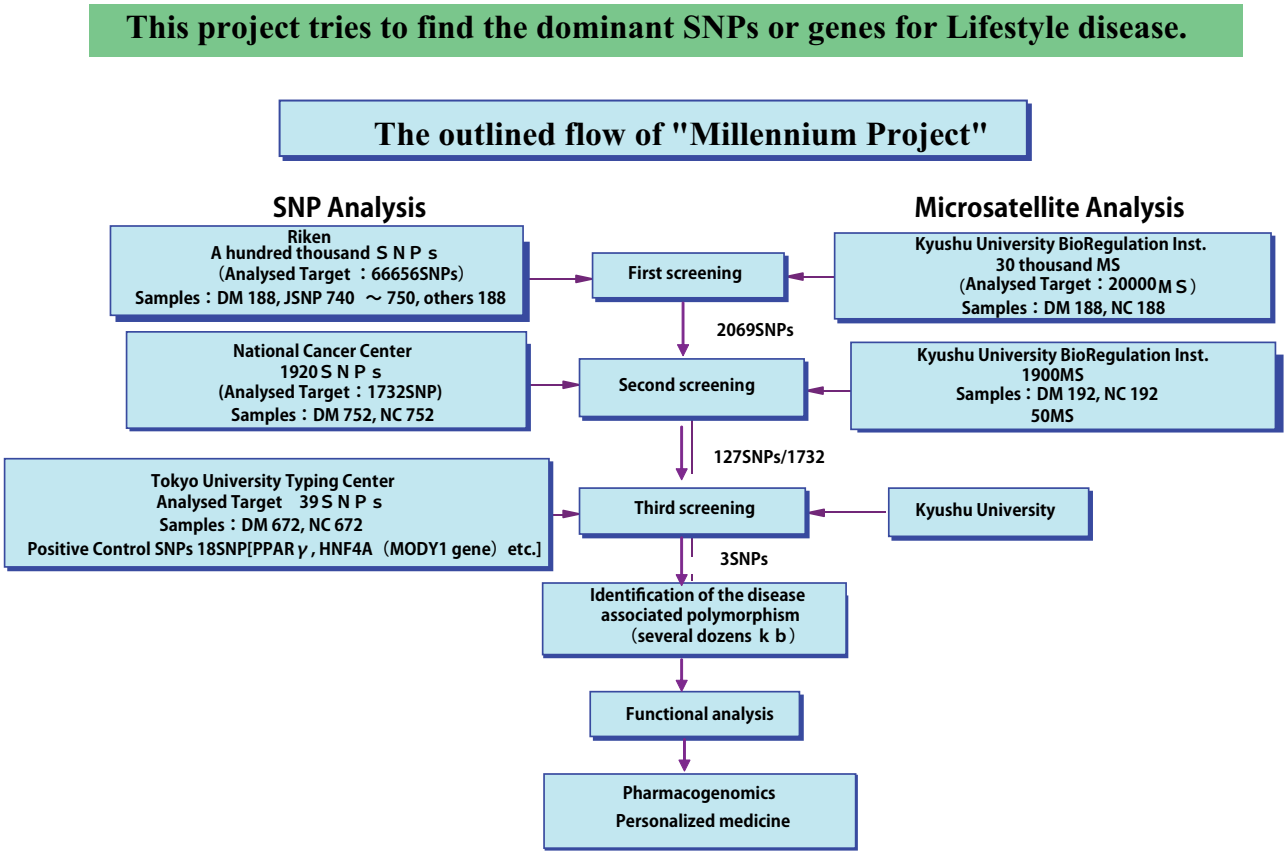


Figure C.2 — An example work flow of the SNP analysis

C.2.4 UML example of SNP analysis

Figure C.3 shows a UML example of the “Millennium Project”.

This project tries to find the DM associated genes and SNP with five steps flow.

In this project, the collected information is not only about the SNP data but also the clinical information, SNP annotations and the specimens.

In the real situation, the anonymization is important for the protection of privacy and project reliability.

C.2.5 Use case of database integration

The Japanese ministry of education, culture, sports, science and technology started a project named the integrated biomedical database project of 2007. This project tries to integrate biomedical databases in Japan, virtually, with intelligent data format based on GSVML. In this project, clinical data and -omics data are virtually integrated and are exchanged with the intelligent data format which is based on GSVML. The details of this project are shown on the website (see reference [22]).

C.2.6 Use case and required elements

Prior to summarising the use cases and required elements, the considered factors of the requirement to the GSVML are listed in Figure C.4.

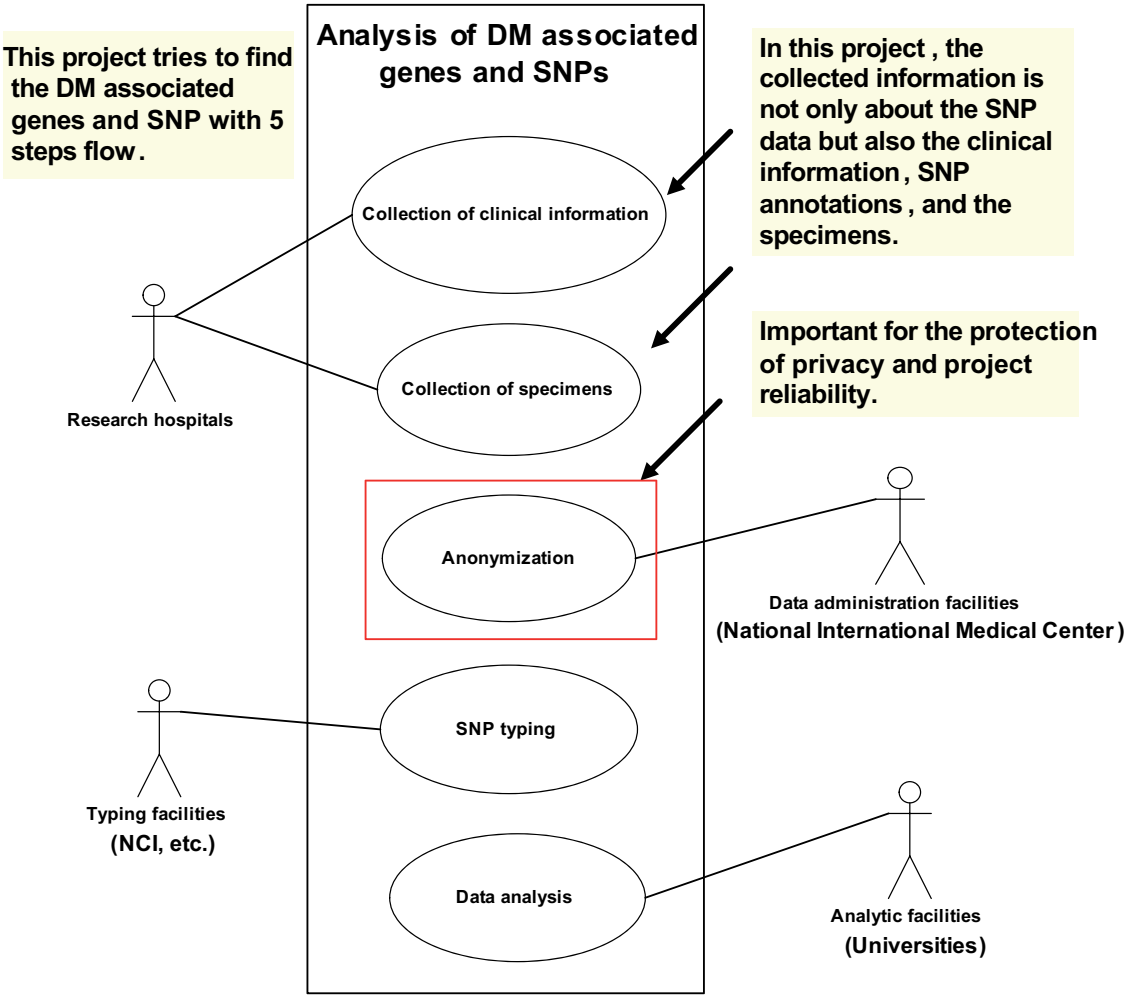


Figure C.3 — A UML example of the "Japanese Millennium Project"

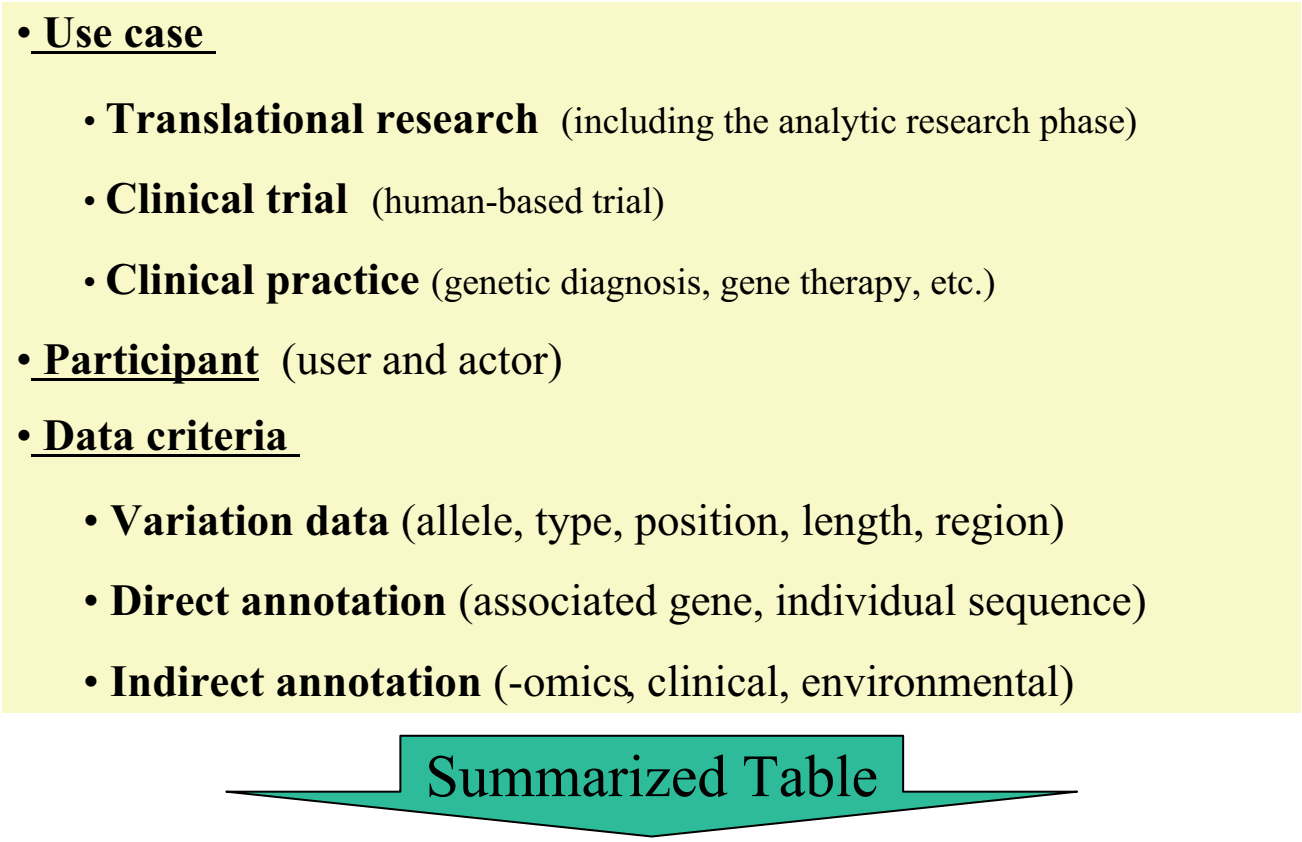


Figure C.4 — The considered factors of the requirement for the GSVML

We considered factors of use case, participant, and data criteria. As for the use case, we considered these three terms:

a) **translational research** (including the analytic research phase);

b) **clinical trial** (human-based trial);

c) **clinical practice** (genetic diagnosis, gene therapy, etc.).

As for the participant, we considered the users and actors and for the data, we compartmentalized data to these criteria.

d) **variation data** (allele, type, position, length, region);

e) **direct annotation** (associated gene, experimental assay);

f) **indirect annotation** (-omics, clinical, environmental).

Here we think that the annotative information such as the -omics annotation, clinical annotation, phenotype, etc., is essential to understand the meaning of the variation data.

Table C.1 summarises use cases and required elements. It also summarises indicated elements derived from these use cases and requirements for the GSVML. The columns are the use cases, their criteria and the participants. Three criteria, namely clinical practice, clinical trial and translational research are defined. The participants are MDs, nurses, and paramedics, etc. The rows represent the data categories. The rows have eleven elements that represent the needs of data format. The rows are categorized into three criteria, namely genomic sequence variation data, direct annotation of variation data and indirect annotation of variation data. In the case of SNP application, the SNP associated genes are in the SNP annotation. The clinical information

and observations are included in the clinical annotation of indirect annotation. All kinds of -omics data including proteomics data are included in the -omics annotation of indirect annotation. The demands of these elements are different among the use cases. As an example, the -omics annotation, as indirect annotation, is largely necessary for gene therapy among MDs and other paramedics.

Table C.1 — Summary of use cases and required elements

E: Essential, NE: Not Essential R: Referential (as Knowledge)			Variation Data								Direct Annotation			Indirect Annotation			
			Location						Epidemiology								
Criteria	Use Case /Elements	Participant	Allele	Type	Position	Length	Region	Associated Gene	Frequency Population etc	Disease Epidemiology	Experimental Assay	Miscellanious	Individual Genome Sequence	Omics Annotation	Clinical Annotation	Phenotype	Environmental condition
Clinical Practice	Genetic Diagnosis or Counselling	MDs Counsellors GT (Patient)	R	E	E	E	R	R	R	E	NE	NE	R	To the greatest extent possible	With Family History	Candidates	E
	Prescription derived from Pharmaco--genomics	MDs GT Pharmacists Hospitals Pharmacies (Patient, Pharmaceutical company)	R	E	E	E	R	R	E	R	NE	NE	NE	normally None	normally None	As a result	E
	Gene Therapy	MDs Nurses GT Hospitals	R	E	E	E	R	R	R	E	NE	NE	R	To the greatest extent possible	With Responder Information	As a result	E
	Disease Prevention based on Individual Polymorphism	MDs Nutrition Counsellors	R	E	E	E	R	R	E	E	NE	NE	NE	With other Polymorphism	With Past History	Candidates	NE
Clinical Trial		MDs, PI Clinical Investigator Research Nurses Researchers Clinical Stuffs (CT etc)	E	E	E	E	E	E	R	R	E	R	E	To the greatest extent possible	With detailed time course	Candidates	E
Translational Research		MDs, PI Clinical Investigator Research Nurses Researchers Clinical Stuffs (CT etc)	E	E	E	E	E	E	R	R	E	R	E	To the greatest extent possible	With detailed time course	Candidates	E

C.3 Diversity of SNP databases

C.3.1 Diversity of databases

Table C.2 lists the results from the diversity analysis among the international existing SNP databases.

The first row represents the international SNP databases and the first column represents the terms of comparison.

As an example for the molecular type, each database uses the word “cDNA“ or “RNA”. They have almost the same meaning in the way of the sequence, while the experimental preparation is different.

As another example for the organism, the homo sapiens and the human have almost the same meaning, while the representations are different.

Table C.2 — Results from the diversity analysis among the international existing SNP databases

Terms of comparison	JSNP	dbSNP	HGVBase	ALFRED	Human SNP Database
URL	http://snp.ims.u-tokyo.ac.jp/index_ja.html	http://www.ncbi.nlm.nih.gov/projects/SNP/	http://hgvbase.cgb.ki.se/	http://alfred.med.yale.edu/alfred/index.asp	http://www.broad.mit.edu/snp/human/index.html
Molecular type	NA	genomic, cDNA	DNA, RNA	NA	cDNA (Affymetrix)
Variation type	SNP Deletion/insertion Polymorphisms Microsatellite	SNP Deletion/insertion Heterozygous sequence Microsatellite or short tandem repeat Named variant No variation Mixed Multi-nucleotide Polymorphism	SNP Deletion/insertion Short tandem repeat Generic	Allele Frequency	SNP
Population	Japanese only	Approximately 700	Plural	Plural	Plural
Organism	Human	Homo sapiens Arabidopsis thaliana Caenorhabditis elegans Ficedula albicollis Ficedula hypoleuca Gallus gallus Mus musculus Pan troglodytes Plasmodium falciparum Rattus norvegicus	Human	Human	Human

Table C.3 — Diversity of data representation among the SNP databases

	JSNP	dbSNP	HGVBase	Alfred	Human SNP database
5′ Flanking Sequence	<5_flank_seq> CAGGAAAC···· </5_flank_seq>	<NSE-ss_flank-5> <NSE-ss_flank-5_E> CAGGAAAC···· </NSE-ss_flank-5_E> <NSE-ss_flank-5>	<UpStreamSeq> CAGGAAAC···· </DnStreamSeq>	5′-ta····	NA
3′ Flanking Sequence	<3_flank_seq> CAGGCAAC···· </3_flank_seq>	< NSE-ss_flank-3> < NSE-ss_flank-3_E> CAGGCAAC···· < NSE-ss_flank-3_E> </ NSE-ss_flank-3>			NA (primer)
Allele SNP	<na_var> C/T </na_var>	<NSE-ss_observed> C/T </NSE-ss_observed>	<Allele>C</Allele> <Allele>T</Allele>	···· C ··· ···· T ···	C/T
Repetition	CACACA CACACACA CACACACACA	Observed (CA)/3/4/5	Allele CACACA Allele CACACACA Allele CACACACACA	(CA)3	CACACA CACACACA CACACACACA
Deletion	A/-	Observed A/-	Allele A Allele	A/N	A/N

C.3.2 Diversity of data representation

Table C.3 shows the diversity of data representation in the SNP databases.

There is also much diversity in representation for the SNP data.

The first row represents the international SNP databases and the first column represents the terms of comparison.

As an example, the representations for the 5' and 3' flanking sequences are completely different amongst the SNP databases.

As another example, the representations for the allele about the SNP representation, repetition representation, and the deletion representation are different amongst the SNP databases.

To exchange the data efficiently and internationally amongst these databases, the data exchanging format from the data representation level needs to be standardized.

C.3.3 Diversity of sequence variation data representation

Table C.4 shows the diversity of sequence variation data representation.

The first row lists the international SNP databases and the first column shows the terms of comparison variation data.

C.4 Markup language comparison

C.4.1 General

Table C.5 shows the results of comparisons amongst markup languages.

The first row lists the markup languages and the first column shows the terms of comparison. This time we investigated the markup languages as MAGE-ML, RNAML, BSML, ProML, SBML and CellML. The compared terms are sequence, variation data, clinical info, transcriptome, proteome, metabolome, signalome and other -omics data.

The results shown in Table C.5 can be summarised as follows:

a) all markup languages can describe the DNA sequence data, but the representations are different;

b) the detailed description of the variation data is not possible for every markup language;

c) the proteomic information can be described by almost all markup languages, but the definitions of the vocabularies of terms are different;

d) the definitions of the basic vocabulary are different amongst markup languages; as an example, “species” means chemical classification for SBML, while it means biological classification for other markup languages;

e) no markup language has descripted ability or expandability of clinical annotative data;

f) no markup language has ability of interface to the HL7 Genotype Model.

Table C.4 — Diversity of sequence variation data representation among the SNP databases

	JSNP	dbSNP	HGVBase	ALFRED	Human SNP Database※
SNPs	○	○	○	△	△
STRP (microsatellite)	○	○	○	△	Not Available
VNTR	○	○	○	△	Not Available
Insertion	○	○	○	△	Not Available
Deletion	○	○	○	△	Not Available
Substitution	○	○	○	△	Not Available
SNP	<snp-type> SNP </snp-type> <snp-5flank_seq> TG </snp-5flank_seq> <snp-3flank_seq> AG </snp-3flank_seq> <snp-allele_na_set> <snp_allele-na> <snp_allele_na-nuc> T </snp_allele_na-nuc> </snp_allele-na> <snp_allele-na> <snp_allele_na-nuc> C </snp_allele_na-nuc> </snp_allele-na> </snp-allele_na_set>	<NSE-ss_subsnp-class value="snp"/> : <NSE-ss_observed> C/T </NSE-ss_observed>	<variation variationID="SNP000495189" curationStatus="---" type="SNP" status="Suspected"> <definition molecule="DNA"> <upStreamSeq>GCTCTGTTTCTCCTACT</upStreamSeq> <dnStreamSeq>TCTCTGTCACAGACTTC</dnStreamSeq> </definition> <map> <dna> <dbXref db="Ensembl" /> <dbXref db="Genbank Record" /> </dna> </map> <allele alleleID="SNP000495189.ALE000988280">T</allele> <allele alleleID="SNP000495189.ALE000988281">C</allele> </variation>	Not Impossible	Not Available
Insertion **Deletion**	<snp-type> IND </snp-type> <snp-5flank_seq> TT </snp-5flank_seq> <snp-3flank_seq> TC </snp-3flank_seq> <snp-allele_na_set> <snp_allele-na> <snp_allele_na-nuc> C </snp_allele_na-nuc> </snp_allele-na> </snp-allele_na_set>	<NSE-ss_subsnp-class value="in-del"/> : <NSE-ss_observed> C/- </NSE-ss_observed>	<variation variationID="IND001634507" curationStatus="---" type="Indel" status="Suspected"> <definition molecule="DNA"> <upStreamSeq>GATATAAATATGTGCAT</upStreamSeq> <dnStreamSeq>AAAGAATGCAATTATC</dnStreamSeq> </definition> <map> <dna> <dbXref db="Ensembl" /> </dna> </map> <allele alleleID="IND001634507.ALE003275909">C</allele> <allele alleleID="IND001634507.ALE003275910">_</allele> </variation>	Not Impossible	Not Available
STRP	<snp-type> MIC </snp-type> <snp-5flank_seq> AA </snp-5flank_seq> <snp-3flank_seq> AG </snp-3flank_seq> <snp-allele_na_set> <snp_allele-na> <snp_allele_na-nuc> AAAAAAA </snp_allele_na-nuc> </snp_allele-na> </snp-allele_na_set>	<NSE-ss_subsnp-class value="microsat"/> : <NSE-ss_observed> (A)10/12 </NSE-ss_observed>	<variation variationID="STR000008185" curationStatus="MRA" type="Tandem Repeat" status="Proven"> <definition molecule="DNA"> <upStreamSeq>CGCCACTTTGTCCCGGC</upStreamSeq> <dnStreamSeq>GGAAAGGCCAACGGTCG</dnStreamSeq> </definition> <map> <dna> <dbXref db="EMBL Record"/> </dna> </map> <allele alleleID="STR000008185.ALE000013695" repeat="(A)10">AAAAAAAAAA</allele> <allele alleleID="STR000008185.ALE000013696" repeat="(A)12">AAAAAAAAAAAA</allele> </variation>	Not Impossible	Not Available

○…Terminology is exact △…Terminology is not exact

※…The database had existed before. But, it is not available now

Table C.5 — Results of comparison amongst markup languages

	MAGE-ML	RNAML	BSML	Proml	SBML	CellML
Sequence	<!ELEMENT BioSequence ((%Identifiable_content;), SequenceDatabases_assnlist?, OntologyEntries_assnlist?, PolymerType_assn, Type_assn, Species_assn?,SeqFeatures_assnlist?) > <!ATTLIST BioSequence %Identifiable_attrs; length CDATA #IMPLIED isApproximateLength CDATA #IMPLIED isCircular CDATA #IMPLIED sequence CDATA #IMPLIED > polymerType:A choice of protein, RNA or DNA	<xs:element name="seq-data"> <xs:complexType> <xs:simpleContent> <xs:extension base="seq-data-string"> <xs:attribute name="comment" /> <xs:attribute name="reference-ids" /> <xs:attribute name="analysis-ids" /> <xs:attribute name="database-ids" />	<!ELEMENT Seq-data (#PCDATA)> Sequencing runs, clones, contigs, cDNA, etc.	Primary Structure <xs:element name="aaSeq" /> Secondary Structure <xs:element name="secStructSeq" />	Not specified	Not specified
Variation Data	In (BioSequence element) same as Omics(transcriptome) term	In (molecule element) same as Omics(transcriptome) term	In (Sequence element) same as Omics(transcriptome) term	Not specified	Not specified	Not specified
Clinical Info	Not specified	Not specified	Not specified	Not specified	Not specified	Not specified
Omics Transcriptome	Distributed to (BioSequence,Array…) <!ELEMENT Array ((%Identifiable_content;), ArrayDesign_assnref, Information_assnref, ArrayGroup_assnref?, ArrayManufactureDeviations_assnlist?) > <!ATTLIST Array %Identifiable_attrs; arrayIdentifier CDATA #IMPLIED arrayXOrigin CDATA #IMPLIED arrayYOrigin CDATA #IMPLIED originRelativeTo CDATA #IMPLIED > The technology type may be spotted **cDNA** <!ELEMENT FeatureGroup ((%DesignElementGroup_content;), TechnologyType_assn?, FeatureShape_assn?, DistanceUnit_assn?, Features_assnlist) > <!ATTLIST FeatureGroup %DesignElementGroup_attrs; featureWidth CDATA #IMPLIED featureLength CDATA #IMPLIED featureHeight CDATA #IMPLIED >	<xs:element name="molecule"> <xs:complexType> <xs:sequence> <xs:element ref="identity" /> <xs:element ref="sequence" /> <xs:element ref="structure" /> </xs:sequence> <xs:attribute name="id" type="xs:ID" use="required" /> <xs:attribute name="type" default="rna"> <xs:simpleType> <xs:restriction base="xs:string"> <xs:enumeration value="rna" /> <xs:enumeration value="dna" /> </xs:restriction> </xs:simpleType> </xs:attribute> <xs:attribute name="comment" /> <xs:attribute name="reference-ids" /> <xs:attribute name="analysis-ids" /> <xs:attribute name="database-ids" /> </xs:complexType> </xs:element>	Sequence element contains seq-data element <!ELEMENT Sequence (Attribute*, Feature-tables?, (Seq-data | Seq-data-import)?, Numbering?, Modification*, Segment*, Resource*, %links;)> <!ATTLIST Sequence %attrs; locus CDATA #IMPLIED db-source %dbsource; #IMPLIED length CDATA #IMPLIED end5hang5 %integer; #IMPLIED end3hang5 %integer; #IMPLIED end5phos %yesorno; #IMPLIED end3phos %yesorno; #IMPLIED genomeref IDREF #IMPLIED trans-table CDATA #IMPLIED segmenttype %segment-opts; "sequence" dnatype (genomic | cdna) #IMPLIED representation %repr-opts; "raw" molecule (mol-not-set | dna | rna | aa | na | other-mol) "dna"	Not specified	These include large molecules(e.g., **RNA** proteins…) <xsd:complexType name="Species"> <xsd:complexContent> <xsd:extension base="SBase"> <xsd:attribute name="id" /> <xsd:attribute name="name" /> <xsd:attribute name="compartment" /> <xsd:attribute name="initialAmount" /> <xsd:attribute name="initialConcentration" /> <xsd:attribute name="substanceUnits" /> <xsd:attribute name="spatialSizeUnits" /> <xsd:attribute name="hasOnlySubstanceUnits" /> <xsd:attribute name="boundaryCondition" /> <xsd:attribute name="charge" /> <xsd:attribute name="constant" /> </xsd:extension> </xsd:complexContent> </xsd:complexType>	Not specified
Omics Proteome	In (BioSequence element) A BioSequence is a representation of a DNA, RNA, or **protein** sequence. same as the Sequence term	Not specified	<!ELEMENT Alignment-summary (Aligned-sequence*)> <!ATTLIST Alignment-summary seq-type (nucleotide | protein) #REQUIRED seq-format CDATA #REQUIRED	Distributed to (protein,proteinSet …) <xs:complexType name="protein"> <xs:sequence> <xs:element name="chain" /> <xs:element name="terStruct" /> </xs:sequence>	In (species element) same as Omics (transcriptome) term	Not specified
Omics Metabolome	Not specified	Not specified	Not specified	Not specified	As models Metabolic pathways	Not specified
Omics Signalome	Not specified	Not specified	Not specified	Not specified	As models Signaling pathways	Not specified

C.4.2 Mapping of each markup language to the data categories

C.4.2.1 General

Each markup language has its intentional application target. Mapping of each markup language to the data category in Table C.1 can elucidate the position of GSVML among markup languages.

C.4.2.2 The MicroArray Gene Expression markup language (MAGE-ML)

MAGE-ML is a data format for describing information about DNA-array based experiments and gene expression data. This markup language can be used to represent expression data such as -omics data of indirect annotation in Table C.1.

C.4.2.3 The Bioinformatic Sequence markup language (BSML)

BSML encodes biological sequence information and includes graphical representations of biologically meaningful objects such as sequences, genes, electrophoresis gels and multiple alignments. This markup language can be used to represent molecular sequence data as -omics data of indirect annotation and individual sequence of direct annotation in Table C.1.

C.4.2.4 The Systems Biology markup language (SBML)

SBML can represent models of biological systems common in research on a number of topics including cell signalling pathways, metabolic pathways, biochemical reactions and many others. This markup language can be used to represent molecular network such as -omics annotation of indirect annotation in Table C.1.

C.4.2.5 The RNAML

RNAML is designed to facilitate the interoperation of multiple RNA informatics programs. RNAML is a standard syntax for exchanging information. This markup language can be used to represent RNA data such as -omics data of indirect annotation in Table C.1. In some ways this markup language has some overlaps with MAGE-ML.

C.4.2.6 The PolyMAPr

PolyMAPr is an SNP centric program and tries to achieve mining, annotation and functional analysis of public databases (dbSNP, CGAP and JSNP). This program can be used to find associated genes described in direct annotation in Table C.1. This is not a markup language but has an SNP centric concept. We also investigated this program for this reason.

C.4.3 GSVML originated needs and its specifications

GSVML is centric on genomic sequence variation, human and clinical use. All of its needs and specifications are derived from these directions. Initially, GSVML shall have the sharable representations for genomic sequence variation data such as allele, type, position, length and region. These representations shall also have expandability to the possible other sequence variation data. The annotations of variations such as variation associated genes, individual sequence and experimental assay are essential to understand the basis and the situation of the genomic sequence variation. To understand the clinical significance or to use in clinical situations, the peripheral annotations of variations such as clinical observation, phenotypes are necessary to determinate the meanings of variations. This information is discussed by Health Level Seven Clinical Genomics Special Interest Group who are developing a standard. This work also has a relation with this group's work.

C.5 Interface analysis to Health Level Seven

C.5.1 General

Health Level Seven (HL7) is one of the standard protocols for healthcare information exchange. HL7 version 3 (HL7v3), the latest edition of HL7, adopts an object-oriented development methodology and a reference information model (RIM) that has powerful descriptive abilities to create messages. It can describe not only clinical information such as the clinical examination data and the prescription data, etc., but also genetic information such as the alleles and SNPs by unified model. Furthermore, the development of the standard for an electronic exchange of genetic information is being advanced, with Clinical Genomics Work Group (CGWG) now aiming for individualization medication as is the case for HL7.

In this clause, the interface with HL7 is examined by comparing the HL7v3 Ballot 10 Genotype Model (Figure C.5) developed by CGWG, with the content of GSVML.

C.5.2 Comparison with HL7

C.5.2.1 General

The HL7 Genotype Model describes the data relating to a genotype, which HL7 propose to be the basic unit of genomic information exchange in healthcare. This model is not meant to be a biological model; rather, it is aimed at the needs of healthcare with the vision of personalized medicine in mind. By contrast, although GSVML has almost the same purpose, it gives priority to the variation itself and aims at the development of the more appropriate clinically biological model. This difference is typically reflected in the difference of the entry point, structure and content, etc., in those models, as shown in C.5.2.2 to C.5.2.4.

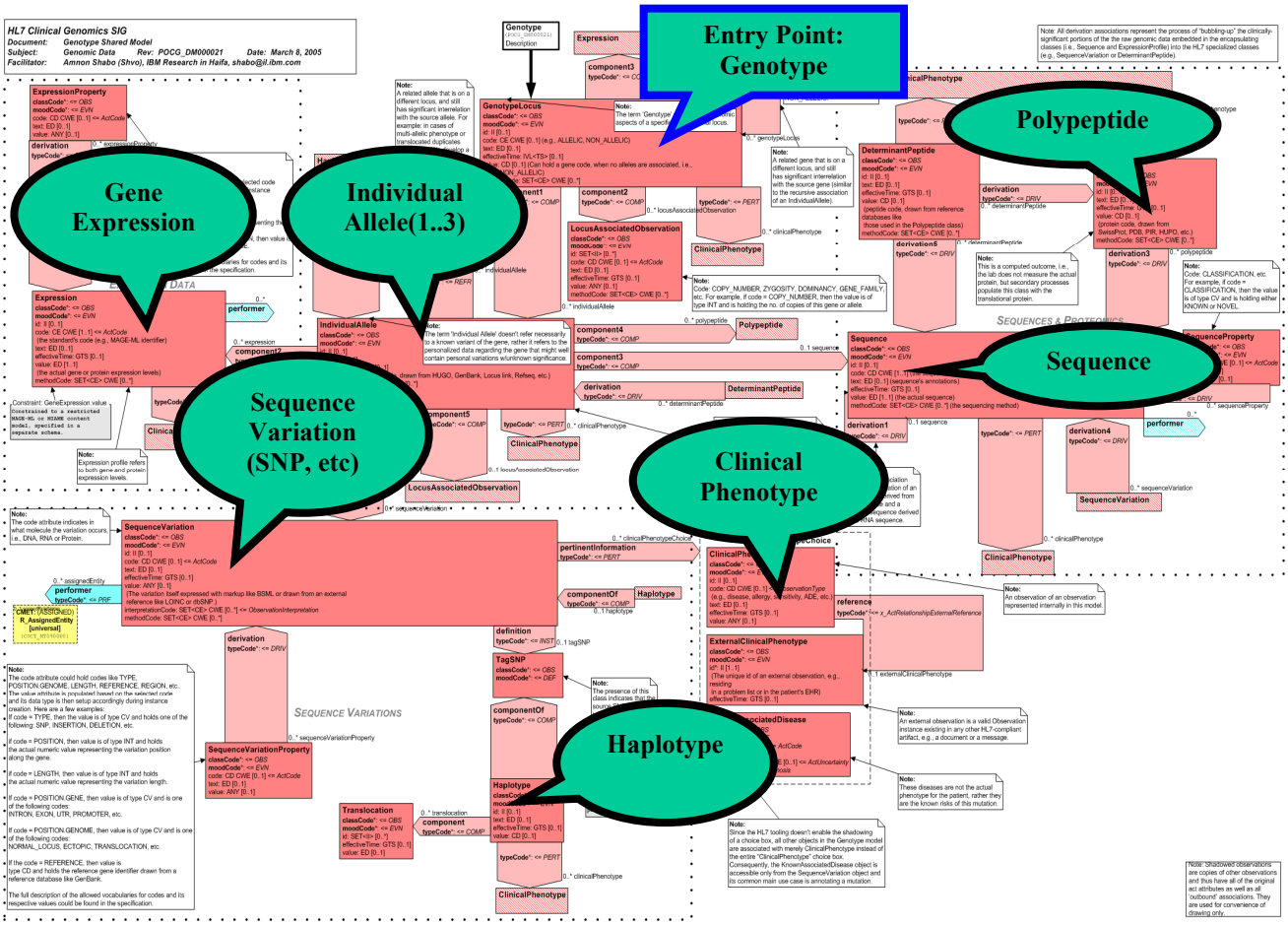


Figure C.5 — HL7v3 genotype information model with explanation

C.5.2.2 Entry point

The entry point to the HL7v3 Genotype Model is a genotype (genetic locus). In contrast, the entry point to GSVML is variation loci. See Figure C.6.

C.5.2.3 Structure

In a HL7 Genotype Model, the main elements are genotype, allele, variation, expression, sequence and phenotype. Genotype is associated with a pair of alleles on paternal and maternal homologous chromosomes; these alleles are associated with variation, expression and sequence. Additionally, all elements are associated with phenotype. In GSVML, variation is associated with genotype, alleles and sequence. The expressions and the phenotypes are described as direct annotation, or indirect annotation. See Figure C.7.

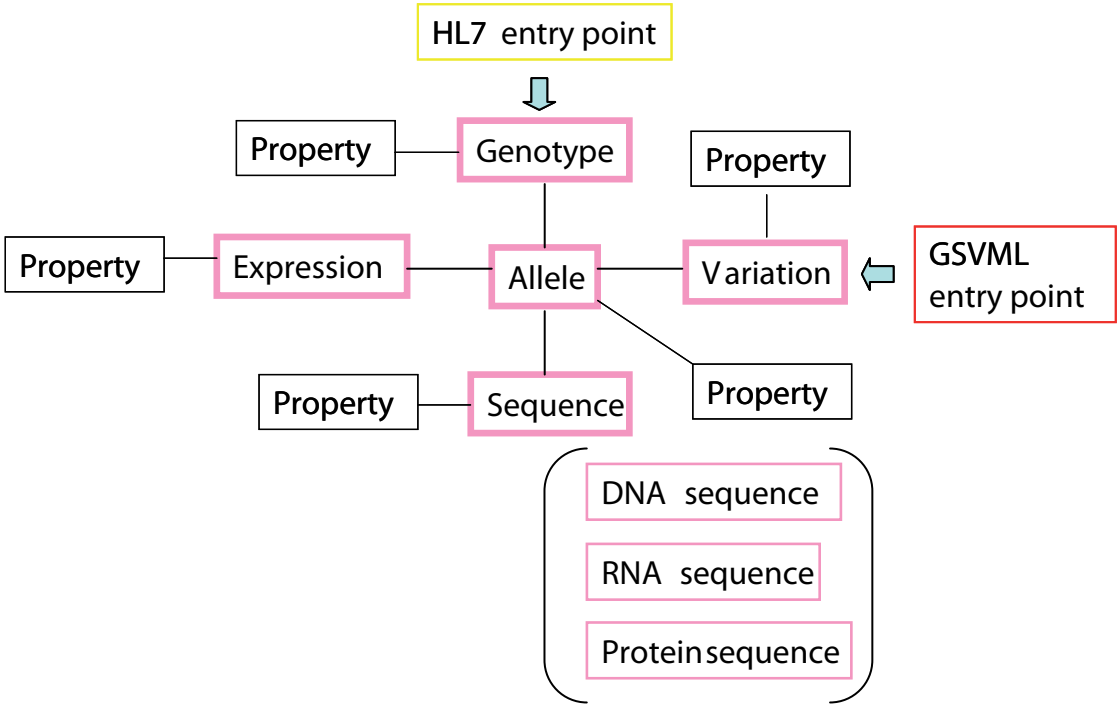


Figure C.6 — The considered factors of the requirement for the GSVML

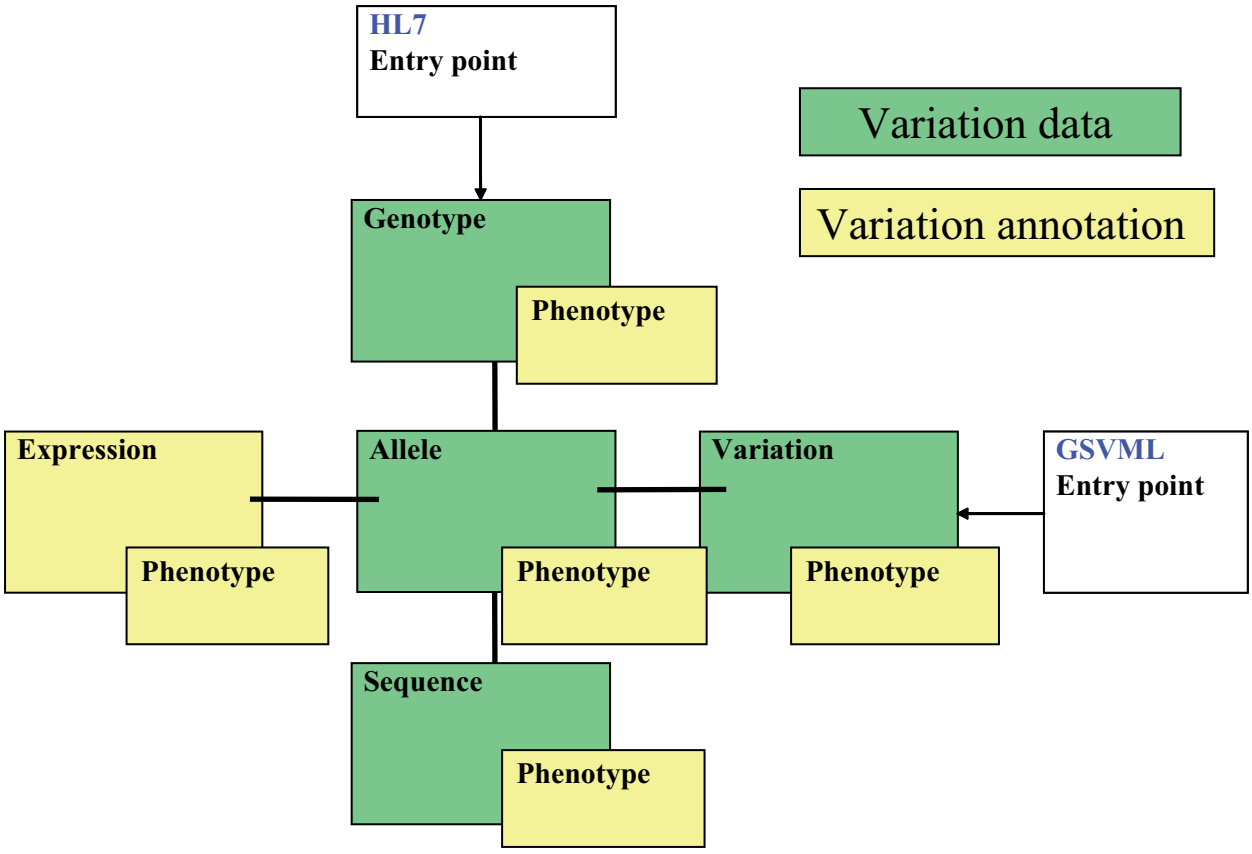


Figure C.7 — The structure of a HL7 Genotype Model and GSVML

C.5.2.4 Contents

Both HL7v3 Genotype Model and GSVML have the genetic information and the basic information that derives from the genetic information. At HL7v3, the associated clinical information is described in the other informational models. On the other hand, GSVML has an ability to describe various associated information, such as experimental conditions, epidemiology, statistical information, etc. Those are needed to make use of variation data in clinical application. Comparisons of content of GSVML and HL7v3 are shown in Table C.6; the detailed results of mapping GSVML contents to HL7v3 Genotype Model are shown in Table C.7.

Table C.6 — Comparison of content of GSVML and HL7v3

GSVML content	HL7v3 Genotype Model
Variation data	○
Direct annotation	Δ
Indirect annotation	Δ

C.5.3 Information model of genotype in HL7

Figure C.8 shows the information model of genotype in HL7. HL7 Clinical Genomics SIG establishes this model from the point of the clinical and the broader aspect. This model packed the DNA sequence variation data, gene expression data, and the clinical phenotype in one information model (HL7 POCG_DM000023).

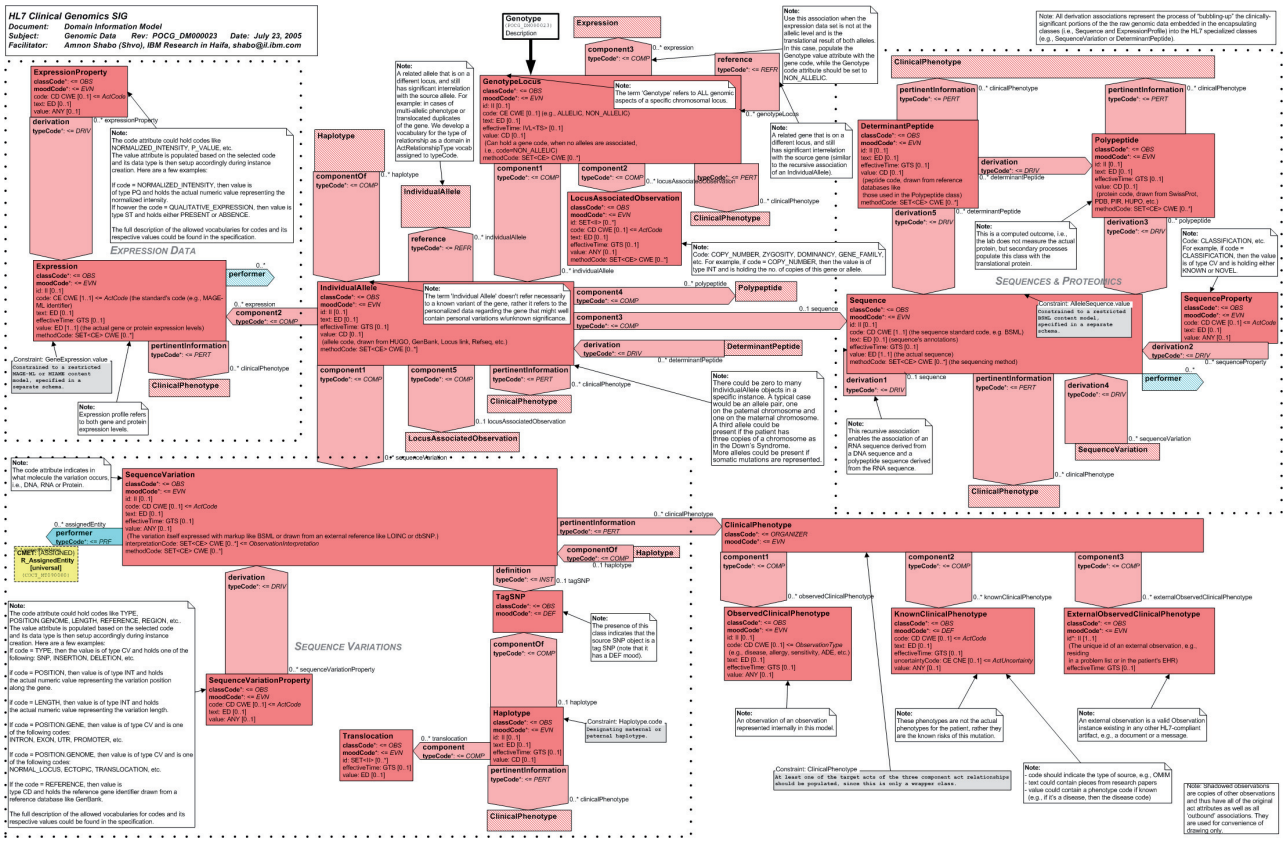


Figure C.8 — HL7 information model of genotype

Table C.7 — Mapping GSVML contents to HL7v3 Genotype Model

No	Element Name	Attribute Name	1※	Mapping	Mapping details
1	**gsvml**				
2	variation_data			→	
3	direct_annotation			→	
4	indirect_annotation			→	
5	**gsvml/variation_data**				
6	variation_type		◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code:"TYPE", value:variation_type
7	location			→	
8	variation_att			→	
9	source			→	
10	variation_dbref			→ dbref	
11	**gsvml/variation_data/location**				
12	chromosome_number		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code, value
13	position		◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code:"POS", value:position
14	map		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code, value(ED)
15	orientation		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code, value
16	ass_gene			→	
17	location_dbref			→ dbref	
18	**gsvml/variation_data/location/ass_gene**				
19	ass_gene_name		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
20	ass_gene_structure		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
21	aminoacid_substitution		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
22	codon_substitution		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
23	codon_position		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
24	ass_gene_symbol		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
25	ass_gene_alias		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
26	ass_gene_product		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
27	ass_gene_evidence_type		○	GenotypeLocus/component2/LocusAssosiatedObservation	code, value
28	changed_motif		×		
29	changed_motif_name		×		
30	changed_splice_site		×		
31	splice_variant_number			→	
32	ass_gene_dbref			→ dbref	
33	**variation_data/location/ass_gene/splice_variant_number**				
34	refSeq_number		×		
35	**variation_data/location/ass_gene/ ass_gene_dbref**				
36	database_name		◎	GenotypeLocus/component2/LocusAssosiatedObservation	code, value@displayName
37	database_id		◎	GenotypeLocus/component2/LocusAssosiatedObservation	code, value@codeSystem
38	link_url		×		
39	database_attbt		◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation	code@originalText
40	data_id		◎	GenotypeLocus/component2/LocusAssosiatedObservation	code, value@code
41	data_attbt		◎	GenotypeLocus/component2/LocusAssosiatedObservation	code, value@originalText
42	version		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation	value@codeSystemVersion
43	**gsvml/variation_data/variation_att**				
44	molecular_type		◎	GenotypeLocus/component1/IndividualAllele/component3/Sequence/derivation3/SequenceProperty	code:"TYPE", value:molecular_type
45	allele		○	GenotypeLocus/component1/IndividualAllele/component3/Sequence	value
46	length		◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code:"LEN", value:length
47	f5sequence		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code, value
48	f3sequence		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code, value
49	validation_status		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code, value
50	success_rate		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/derivation/SequenceVariationProperty	code, value
51	**gsvml/variation_data/source**				
52	source_release_date		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation	effectiveTime
53	source_modify_date		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation	effectiveTime
54	source_rawdata		◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation	value
55	source_dbref			→ dbref	
56	**gsvml/direct_annotation**				
57	whole_genome_sequence		×		
58	mendelian_segregate		×		
59	homozygote_detect		◎	GenotypeLocus/component2/LocusAssosiatedObservation	code:"ZYGO", value:"HOMO" or "HETERO"
60	somatic_mutation		×		
61	experiment_analysis			→	
62	epidemiology			→	
63	var_ann_misc			→	
64		var_ann_misc_id	×		
65		submitter_id	×		
66		create_date	×		
67		modify_date	×		
68	**gsvml/direct_annotation/experiment_analysis**				
69	variation_identif_y			→	
70		experimental_assay_id	×		
71		submitter_id	×		
72		create_date	×		
73		modify_date	×		
74	variation_characterize			→	

1※ ◎ ・・・ It is possible to correspond.

○ ・・・ It is possible to correspond by adding a vocabulary.

× ・・・ It is necessary to change the structure to make it correspond.

Table C.7 (*continued*)

75	**gsvml/direct_annotation/ experimental_assay/variation_identi_f_y**				
76	experimental_assay_result		×		
77	experimental_assay_description		×		
78	experimental_assay_parameter		×		
79	experimental_assay_dbref			→ dbref	
80	publication			→	
81		publication_id	×		
82		submitter_id	×		
83		create_date	×		
84		modify_date	×		
85	submitter			→	
86		submitter_id	×		
87		create_date	×		
88		modify_date	×		
89	pcr			→	
90	**gsvml/direct_annotation/experimental_assay/ variation_identi_f/publication**				
91	title		×		
92	author		×		
93	journal		×		
94	volume		×		
95	supplement		×		
96	issue		×		
97	issue_supplement		×		
98	pages		×		
99	year		×		
100	publication_status		×		
101	mesh_term		×		
102	publication_dbref		×		
103	submitter			→	
104	**gsvml/direct_annotation/ experimental_assay/variation_identi_f/y publication/submitter**				
105	submitter_name		×		
106	address		×		
107	email		×		
108	tel		×		
109	fax		×		
110	institution		×		
111	laboratory		×		
112	submitter_dbref		×		
113	publication			→	
114	**gsvml/direct_annotation/experimental_assay/ variation_identi_fy/pcr**				
115	pcr_confirmed		×		
116	pcr_primer		×		
117	pcr_product		×		
118	pcr_profile		×		
119	**gsvml/direct_annotation/experimental_assay/ variation_characterize**				
120	genetic_statistics			→	
121	**gsvml/direct_annotation/experimental_assay/ variation_characterize/genetic_statistics**				
122	method			→	
123	p_value		×		
124	linq_dis_index			→	
125	descend_identify			→	
126	maximum_lod_score		×		
127	**gsvml/direct_annotation/experimental_assay/ variation_charavterize/method**				
128	method_name		×		
129	method_description		×		
130	method_url		×		
131	**gsvml/direct_annotation/experimental_assay/ variation_charavterize/linq_dis_index**				
132	d		×		
133	d_prime		×		
134	r_square		×		
135	**gsvml/direct_annotation/experimental_assay/ variation_charavterize/descend_identify**				
136	di_value		×		
137	di_probability		×		
138	**gsvml/direct_annotation/epidemiology**				
139	ass_gene			→	
140	disease_epidemiology			→	
141	population			→	
142		population_id	×		
143		submitter_id	×		
144		create_date	×		
145		modify_date	×		
146	frequency			→	
147		frequency_id	×		
148		submitter_id	×		
149		population_id	×		
150		assay_id	×		
151		publication_id	×		

Table C.7 (*continued*)

152		population_id	×		
153		create_date	×		
154		modify_date	×		
155	**gsvml/direct_annotation/epidemiology/ disease_epidemiology**				
156	striking_age		×		
157	striking_body_area		×		
158	striking_land_area		×		
159	laterality		×		
160	differences			→	
161	prognosis		×		
162	etiology_expression_probability		×		
163	labofatory_findings_expression		×		
164	symptoms_expression_probability		×		
165	prophylaxes_expression_probability		×		
166	respondwer_sidefffects_expression		×		
167	pathological_findings_expression		×		
168	complication_expression_probability		×		
169	**gsvml/direct_annotation/epidemiology/ population**				
170	population_description		×		
171	organism		×		
172	differences			→	
173	population_parameter		×		
174	sample_size		×		
175	population_misc		×		
176	population_dbref			→ dbref	
177	**gsvml/direct_annotation/epidemiology/ frequency**				
178	haplotype		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/ definition/TagSNP/componentOf/Haplotype	code, valeu
179	haplotype_frequency		×		
180	allele		○	GenotypeLocus/component1/IndividualAllele/component3/Sequence	code, value
181	allele_frequency		○	GenotypeLocus/component1/IndividualAllele/component1/	code, value
182	genotype		○	GenotypeLocus/component1/LocusAssociatedObservation	code, value
183	genotype_frequency		○	GenotypeLocus/component1/LocusAssociatedObservation	code, value
184	**gsvml/direct_annotation/epidemiology/ population/differences**				
185	race		×		
186	gender		×		
187	**gsvml/direct_annotation/var_ann_mise**				
188	var_ann_misc_description		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/ derivation/SequenceVariationProperty	
189	var_ann_misc_dbref			→ dbref	
190	**gsvml/indirect_annotation**				
191	personal_info			→	
192	phenotype			→	
193		phenotype_id	◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/ pertinentInformation/ClinicalPhenotype	id
194		submitter_id	×		
195		create_date	◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/ pertinentInformation/ClinicalPhenotype	effectiveTime
196		modify_date	◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/ pertinentInformation/ClinicalPhenotype	effectiveTime
197	omics_annotation			→	
198	environmental_condition			→	
199	clinical_annotation			→	
200	**gsvml/indirect_annotation/ clinical_annotation/personal_info**				
201	personal_dbref			→ dbref	
202	personal_description			→	
203	**gsvml/indirect_annotation/phenotype**				
204	phenotype_description			[FamilyHistory] Patient/Person	
205	phenotype_dbref			→ dbref	
206	**gsvml/indirect_annotation/ phenotype/phenotype_description**				
207	phenotype_type		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/ PertinentInformation/ClinicalPhenotype	code
208	phenotype_condition		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/ pertinentInformation/ClinicalPhenotype	text
209	phenotype_probability		×		
210	**gsvml/indirect_annotation/omics_annotation**				
211	omics_type		×		
212	omics_description			→	
213	omics_dbref			→ dbref	
214	**gsvml/indirect_annotation/omics_annotation/ omics_description**				
215	omics_material		×		
216	omics_condition		×		
217	omics_expression_probability		×		
218	**gsvml/indirect_annotation/**				
219	expression_condition_description		○	GenotypeLocus/component1/IndividualAllele/Expression/derivation/	code, value
220	expression_condition_probability		×		
221	omics_annotation			→	
222	**gsvml/indirect_annotation/clinical_annotation**				
223	disease			→	
224		disease_id	◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/ PertinentInformation/ClinicalPhenotype	id
225		submitter_id	×		

Table C.7 (*continued*)

226		create_date	◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	effectiveTime
227		modify_date	◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	effectiveTime
228	clinical_observation			→	
229	clinical_annotation_dbref			→ dbref	
230	**gsvml/indirect_annotation/clinical_annotation/disease**				
231	disease_description			→	
232	disease_epidemiology			→ epidemiology	
233	disease_dbref			→ dbref	
234	**gsvml/indirect_annotation/disease/disease_description**				
235	name		◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/KnownAssociatedDisease	code(e.g. ICD10)
236	synonym		◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/KnownAssociatedDisease	code@translation
237	definition_concept		◎	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/KnownAssociatedDisease	code@originalText
238	classification		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/KnownAssociatedDisease	code
239	etiology			→	
240	laboratory_findings			→	
241	pathological_findings			→	
242	symptoms			→	
243	diagnostic_criteria			→	
244	therapy			→	
245	complication			→	
246	prophylaxes			→	
247	**gsvml/indirect_annotation/disease_description/etiology**				
248	etiology_description		×		
249	etiology_condition		×		
250	etiology_expression_probability		×		
251	**gsvml/indirect_annotation/disease_description/laboratory_findings**				
252	laboratory_findings_type		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	code
253	laboratory_findings_description		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	code@originalText
254	laboratory_findings_condition		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	text
255	laboratory_findings_expression_probability		×		
256	**gsvml/indirect_annotation/disease_description/pathological_findings**				
257	pathological_findings_description		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	code@originalText
258	pathological_findings_condition		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	text
259	pathological_findings_expression_probability		×		
260	**gsvml/indirect_annotation/disease_description/symptoms**				
261	symptoms_description		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	code@originalText
262	symptoms_condition		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	text
263	symptoms_expresson_probability		×		
264	**gsvml/indirect_annotation/disease_description/diagnostic_criteria**				
265	diagnostic_standard		×		
266	diagnostic_modify		×		
267	diagnostic_differential		×		
268	**gsvml/indirect_annotation/disease_description/therapy**				
269	conservative			→	
270	surgery		×	(ExternalClinicalPhenotype)	
271	radiation		×	(ExternalClinicalPhenotype)	
272	**gsvml/indirect_annotation/disease_description/therapy/conservative**				
273	pharmaceutical			→	
274	physical		×	(ExternalClinicalPhenotype)	
275	**gsvml/indirect_annotation/disease_description/therapy/conservative/pharmaceutical**				
276	responder_sideffects			→	
277	**gsvml/indirect_annotation/disease_description/therapy/conservative/pharmaceutical/responder_sideffects**				
278	responder_sideffects_causer		×		
279	responder_sideffects_description		×		
280	responder_sideffects_condition		×		
281	responder_sideffects_expression_probability		×		
282	**gsvml/indirect_annotation/disease_description/complication**				
283	complication_description		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	code@originalText

Table C.7 (*continued*)

284	complication_condition		○	GenotypeLocus/component1/IndividualAllele/component1/SequenceVariation/PertinentInformation/ClinicalPhenotype	text
285	complication_expression_probability		×		
286	**gsvml/indirect_annotation/disease_description/prophylaxes**				
287	prophylaxes_description		×		
288	prophylaxes_condition		×		
289	prophylaxes_expression_probability		×		
290	**gsvml/indirect_annotation/clinical_annotation/clinical_observation**				
291	subjective_findings			→	
292	objectives_findings			→	
293	plan			→	
294	assessment			→	
295	clinical_observation_dbref			→ dbref	
296	**gsvml/indirect_annotation/clinical_annotation/clinical_observation/subjective_findings**				
297	symptoms			→	
298	**gsvml/indirect_annotation/clinical_annotation/clinical_observation/objective_findings**				
299	laboratory_findings			→	
300	pathological_findings			→	
301	complication			→	
302	family_history			→	
303	**gsvml/indirect_annotation/clinical_annotation/clinical_observation/**				
304	assessment_result		×	(ExternalClinicalPhenotype)	
305	assessment_dbref			→ dbref	
306	**gsvml/indirect_annotation/clinical_annotation/clinical_observation/plan**				
307	therapy			→	
308	**gsvml/indirect_annotation/clinical_annotation/clinical_observation/objectives/family_history**				
309	family_history_description			→	
310	family_history_dbref			→ dbref	
311	**gsvml/indirect_annotation/clinical_annotation/clinical_observation/objectives/family_history/**				
312	relation_structure		×		
313	family_member			→	
314	**gsvml/indirect_annotation/clinical_annotation/clinical_observation/objectives/family_history/family_history_description/family_member**				
315	personal_info			→	
316	phenotype		○	[FamilyHistory] Patient/Person/PersonalRelationship/Person/ClinicalObservation	
317	clinical_annotation		○	[FamilyHistory] Patient/Person/PersonalRelationship/Person/ClinicalObservation	

C.6 Interface analysis to CEN EN 13606

In this clause, interface analysis between GSVML and EN 13606 was examined by comparing the parts of EN 13606 developed by CEN with the content of GSVML.

GSVML is a sharable data exchange format that is designed for exchanging genomic sequence variation data and their annotative information, including clinical information. On the other hand, EN 13606 mainly treats standardization of EHR. They have different target scopes and are complementary to each other. Moreover, considering future gene-based medicine, it is significant to integrate genetic information to EHR and is meaningful to analyse the interface.

EN 13606 is based on a two-level methodology that can explicitly separate knowledge and the information model, whilst the conventional method is based on a single methodology that is a mixture of knowledge and information models. GSVML can correspond to both; the user can choose either the preset simplified ontology and information model or the user-defined ontology and information models.

EN 13606 has five parts (Part 1: Reference model, Part 2: Archetypes interchange specification, Part 3: Reference archetypes and term lists, Part 4: Security, Part 5: Interface specification). Here, core concepts in CEN methodologies are reference model in Part 1 and archetypes in Parts 2 and 3. The outline of the correspondence between GSVML categories and EN 13606 parts is shown in Figure C.9. The easiest way to implement GSVML to EN 13606 is to encapsulate whole GSVML. In case of more explicit genomic EHRs, GSVML shall be explicitly addressed in Parts 1, 2 and 3.

The present version of EN 13606 provides a framework that can contain genomic information and its annotative information, while it does not provide information models or knowledge models specialized in genome information and their corresponding annotative information. GSVML can be a complement of EN 13606.

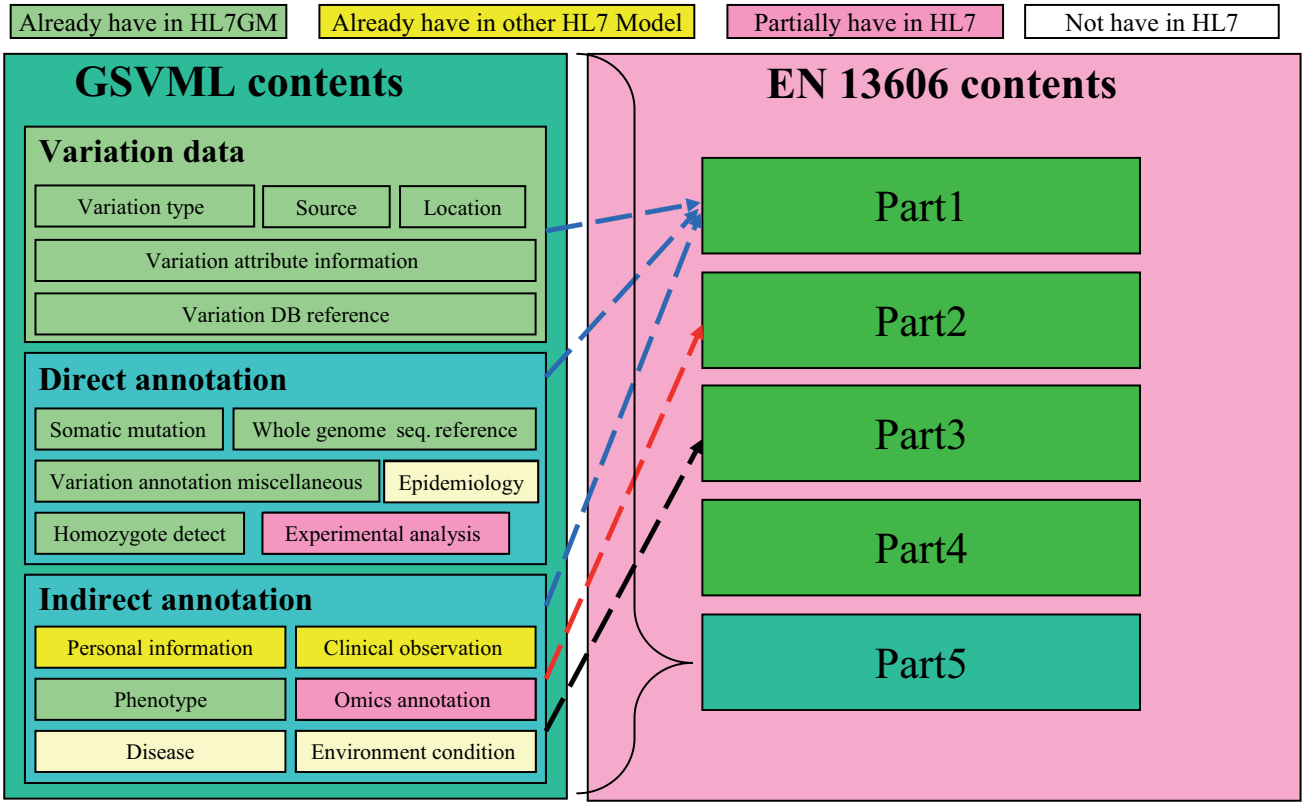


Figure C.9 — Mapping GSVML to EN 13606

C.7 Interface analysis to SNOMED-CT

In this clause, the interface with SNOMED-CT is examined by addressing the terms of “indirect annotation” category of GSVML to 18 main concepts of SNOMED-CT. The 18 concepts are clinical findings, procedure, observable entity, body structure, organism, substance, pharmaceutical/biological product, specimen, physical object, physical force, events, environments and geographical locations, social context, context-dependent categories, staging and scales, attribute, qualifier value and special concept. The results are shown in Table C.8. According to the study, almost all the categories in “indirect annotation” of GSVML can be addressed by the concepts in SNOMED-CT. The only terms that are difficult to address are the parts of categories concerning the database handling such as database name, database ID, link URL, database attbt, data ID, data attbt and version. GSVML has its own simplified informational structure. By inserting the informational structure of SNOMED-CT into the category of database handling of GSVML, users can use either type of informational structure.

Table C.8 — Mapping GSVML contents on SNOMED-CT

GSVML	SNOMED-CT
Clinical annotation	Finding (clinical findings), disease (clinical findings)
Disease	Disease (clinical findings)
Disease description	Disease (clinical findings)
Name	Disease (clinical findings)
Synonym	Disease (clinical findings)
Definition concept	Finding (clinical findings)
Classification	Finding (clinical findings), staging and scales
Etiology	
Etiology description	Finding (clinical findings), body structure, organism, substance, pharmaceutical/biologic product, context-dependent categories
Etiology condition	Environments and geographical locations, social context, context-dependent categories, staging and scales
Etiology expression probability	Staging and scales
Laboratory findings	
Laboratory finding type	Finding (clinical findings), body structure, organism, substance, specimen, attribute
Laboratory finding description	Finding (clinical findings), disease (clinical findings), procedure, observable entity, body structure, organism, substance, pharmaceutical/biologic product, specimen, physical object, environments and geographical locations, staging and scales
Laboratory finding condition	Environments and geographical locations, social context, context-dependent categories, staging and scales
Laboratory finding expression probability	Staging and scales
Pathological findings	Finding (clinical findings)
Pathological findings description	Finding (clinical findings)
Pathological findings condition	Environments and geographical locations, social context, context-dependent categories, staging and scales
Pathological findings expression probability	Staging and scales
Symptom	Finding (clinical findings)
Symptom description	Finding (clinical findings)
Symptom condition	Environments and geographical locations, social context, context-dependent categories, staging and scales
Symptom expression probability	Staging and scales
Diagnostic criteria	
Diagnostic standard	Finding (clinical findings), environments and geographical locations, staging and scales
Diagnostic modify	Finding (clinical findings), environments and geographical locations, social context, context-dependent categories, staging and scales
Diagnostic differential	Finding (clinical findings), environments and geographical locations, social context, context-dependent categories, staging and scales
Therapy	Procedure
Conservative	Procedure

Table C.8 (*continued*)

GSVML	SNOMED-CT
Pharmaceutical	Finding (clinical findings), procedure, pharmaceutical/biologic product
Responder side effects	
Responder side effects description	Finding (clinical findings), pharmaceutical/biologic product, substance
Responder side effects causer	Substance, pharmaceutical/biologic product, physical force
Responder side effects condition	Environments and geographical locations, social context, context-dependent categories
Responder side effects express probability	Staging and scales
Physical	Procedure, body structure, physical object, physical force, environments and geographical locations, social context, context-dependent categories
Surgery	Procedure, substance, physical object, physical force
Radiation	Procedure, substance, physical object, physical force
Complication	
Complication description	Finding (clinical findings), disease (clinical findings)
Complication condition	Finding (clinical findings), physical object, physical force, environments and geographical locations, social context, context-dependent categories
Complication expression probability	Staging and scales
Prophylaxis	
Prophylaxis description	Procedure, substance, physical object, physical force, environments and geographical locations
Prophylaxis condition	Environments and geographical locations, social context, context-dependent categories
Prophylaxis expression probability	Finding (clinical findings), procedure, staging and scales
Clinical observation	Finding (clinical findings)
Subjective findings	Finding (clinical findings)
Symptoms	Finding (clinical findings)
Objective findings	Finding (clinical findings)
Laboratory findings	Finding (clinical findings)
Pathological findings	Finding (clinical findings)
Complications	Finding (clinical findings)
Family history	
Family history description	Observable entity, social context, context-dependent categories
Family member	Observable entity, environments and geographical locations, social context
Personal info	Environments and geographical locations, social context
Personal description	Observable entity, environments and geographical locations, social context, context-dependent categories
Personal dbref	None
Database name	None
Database ID	None

Table C.8 (*continued*)

GSVML	SNOMED-CT
Link URL	None
Database attbt	None
Data ID	None
Data attbt	None
Version	None
Phenotype	
Phenotype description	Finding (clinical findings), observable entity
Phenotype type	Finding (clinical findings), observable entity
Phenotype condition	Observable entity, environments and geographical locations, social context, context-dependent categories
Phenotype probability	Observable entity, procedure, staging and scales
Clinical annotation	Finding (clinical findings), observable entity
Relation structure	Social context
Family history dbref	None
Database name	None
Database ID	None
Link URL	None
Database attbt	None
Data ID	None
Data attbt	None
Version	None
Assessment	Disease (clinical findings)
Assessment result	Observable entity, environments and geographical locations, social context, context-dependent categories, staging and scales
Assessment result database	None
Plan	Procedure, pharmaceutical/biologic product
Therapy	Procedure, pharmaceutical/biologic product
Clinical observation dfref	None
Database name	None
Database ID	None
Link URL	None
Database attbt	None
Data ID	None
Data attbt	None
Version	None
Clinical annotation dbref	None
Database name	None
Database ID	None
Link URL	None

Table C.8 (*continued*)

GSVML	SNOMED-CT
Database attbt	None
Data ID	None
Data attbt	None
Version	None

Bibliography

[1] ZERHOUNI, E., Medicine. The NIH Roadmap, *Science*, 302(5642), pp. 63-72, 2003

[2] HOLDEN, A.L., The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome, *Biotechniques Suppl*, 26, pp. 22-24, 2002

[3] Cognitive Science Princeton University, *Overview for Markup Language*, 1998, internet article of http://www.cogsci.princeton.edu/cgi-bin/webwn2.0?stage=1&word=Markup+Language

[4] ISO 8879, *Information processing — Text and office systems — Standard Generalized Markup Language (SGML)*

[5] BERNERS-LEE, T. and CONNOLLY, D., *HyperText Markup Language Specification — 2.0*, RFC 1866. Proposed Standard, Nov. 1995.

[6] W3C recommendation, *Extensible Markup Language (XML) 1.0 (Second Edition)*, 1998, internet article of http://www.w3c.org/TR/2000/REC-xml-20001006

[7] W3C recommendation, *XHTML™ 1.0 The Extensible HyperText Markup Language (Second Edition) A Reformulation of HTML 4 in XML 1.0*, 2000, internet article of http://www.w3.org/TR/xhtml1/

[8] W3C recommendation, *WAP Forum — W3C Cooperation White Paper*, 1998, internet article of http://www.w3.org/TR/1998/NOTE-WAP-19981030

[9] W3C recommendation, *Simple Object Access Protocol (SOAP) 1.1*, 2000, internet article of http://www.w3.org/TR/2000/NOTE-SOAP-20000508/

[10] Flying Boat Mobile Communications, *Glossary of Terms relevant to Mobile Communications*, 2004, internet article of http://homepages.nildram.co.uk/~jidlaw/pages/glossary.html

[11] LAURENT, S.S. and BIGGAR, R.J., x*Inside SMLDTDs: Scientific and Technical*, McGraw-Hill, Berkeley, CA, 1999

[12] HUCKA, M., FINNEY, A., SAURO, H.M., BOLOURI, H., et al., The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, *Bioinformatics*, **19**(4), pp. 524-31, 2003

[13] HEDLEY, W.J., NELSON, M.R., BALLIVANT, D.P. and NIELSON, P.F., A short introduction to CellML, *Phil. Trans. Roy. Soc*. London, A, 359, pp. 1073-1089, 2001

[14] GODDARD, N.H., HUCKA, M., HOWELL, F., CORNELIS, H., SHANKAR, K. and BEEMAN, D., Towards NeuroML: model description methods for collaborative modelling in neuroscience, *Phil. Trans. Roy. Soc.* London B Biol Sci., 356(1412), pp. 1209-1228, 2001

[15] FREIMUTH, R.R., STORMO, G.D. and MCLEOD, H.L., PolyMAPr: programs for polymorphism database mining, annotation, and functional analysis, *Hum. Mutat*., **25**(2), pp. 110-117, 2005

[16] SHERRY, S.T., WARD, M.H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E.M. and SIROTKIN, K., dbSNP: the NCBI database of genetic variation, *Nucleic Acids Res.,* **29**, pp. 308-311, 2001

[17] BUETOW, K.H., EDMONSON, M.N. and CASSIDY, A.B., Reliable identification of large numbers of candidate SNPs from public EST data, *Nat. Genet.*, **21**, pp. 323-325, 1999

[18] OHNISHI, Y., TANAKA, T., YAMADA, R., SUEMATSU, K., MINAMI, M., FUJII, K., HOKI, N., KODAMA, K., NAGATA, S., HAYASHI, T., KINOSHITA, N., SATO, H., KUZUYA, T., TAKEDA, H., HORI, M. and NAKAMURA, Y., Identification of 187 single nucleotide polymorphisms (SNPs) among 41 candidate genes for ischemic heart disease in the Japanese population, *Hum. Genet*. **106**, pp. 288-292, 2000

[19] Health Level Seven Clinical Genomics Special Interest Group, 2002, internet article of http://www.hl7.org/Special/committees/

[20] YOSHIDA, T., (SNP project in the Millennium Genome Project, Japan), *Gan To Kagaku Ryoho*, **29**(6), pp. 963-967, 2002

[21] WAUGH, A., GENDRON, P., ALTMAN, R., BROWN, J.W., CASE, D., GAUTHERET, D. et al., RNAML: a standard syntax for exchanging RNA information, *RNA*, **8**(6), pp. 707-717, 2002

[22] Integrated Database Project in Japan, 2008, internet article of http://ibmd.tmd.ac.jp/

[23] ISO/HL7 21731, *Health informatics — HL7 version 3 — Reference information model — Release 1*

ICS 35.240.80

Price based on 21 pages