



**NATIONAL STANDARD OF THE
PEOPLE'S REPUBLIC OF CHINA**

GB/T 40419-2021

**Health informatics - Genomic Sequence Variation Markup
Language (GSVML)**

Issued on: October 11, 2021

Implemented on: May 1, 2022

**Issued by: State Administration for Market Regulation, China National
Standardization Administration**

Table of Contents

1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 GSVML positioning	3
5 GSVML structure	3
6 DTD and XML Schema of GSVML	19
Appendix A (informative) GSVML development description	20
A.1 GSVML development needs analysis	20
A.2 GSVML development process	21
A.3 Basic reference materials	22

Foreword

This document is in accordance with the provisions of GB/T 1.1-2020 "Guidelines for Standardization Work Part 1. Structure and Drafting Rules of Standardization Documents"

Drafting.

This document uses the redrafting method to modify the ISO 25720.2009 "Health Informatics Genomic Sequence Variation Markup Language (GS-VML)".

Compared with ISO 25720.2009, this document has made the following structural adjustments.

---According to the provisions of GB/T 1.1-2020, the contents of the introduction, Chapter 2, Chapter 5, Chapter 6 and Appendix C in the original document

To sort out, integrate and delete duplicate content, adjust the introduction, chapter 4, chapter 5, chapter 6 and appendix A of this document, original document

Chapter 3 and Chapter 4 of this document are adjusted to Chapter 2 and Chapter 3 of this document, and Appendix A and Appendix B of the original document are adjusted to this document

Appendices B and C.

The technical differences between this document and ISO 25720.2009 and the reasons are as follows.

---The scope of this document has been modified in accordance with the provisions of GB/T 1.1-2020;

---Adjust "EN13606 (alparts)" in Chapter 3 of the original document to ISO 13606 (all parts) in Chapter 2 of this document;

--- Deleted 4.5, 4.13, 4.14, 4.17, 4.24 and 4.27 in the original document, and added the term "3.11 Genome Sequence Variation Marking language".

The following editorial changes have been made to this document.

---Items 20 and 22 of the references of the original document are deleted.

Introduction

The explosive growth of genome sequence mutation research has produced massive amounts of experimental data, which are stored in the world in various types of data formats

Numerous databases everywhere. In order to effectively manage, analyze, and use these data, the most urgent task is to standardize the data.

Realize global exchange and sharing. The International Standards Development Organization has developed or is developing relevant standards for these data, HL7

(HealthLevelSeven, Health Information Exchange and Transmission Standard) is a standard developed for clinical data, DICOM (DigitalImaging

andCommuncationinMedicine, Medical Digital Imaging and Communication) and JPEG (Joint Photographic Experts Group)

Standard, and Genomic Sequence Variation Markup Language (GSVML) is

In view of the variability of genome sequence---especially SNP (Single Nucleotide Polymorphism, single nucleotide polymorphism) and STRP

(ShortTandemRepeatPolymorphism, short tandem repeat polymorphism) Gene medicine and drug genes for improving human health

Omics has a significant role, and they are standards for genomic data, especially human-related DNA mutation data.

See Appendix A for the description of GSVML development.

This document provides a data exchange format for human health genome sequence variation data, mainly for SNP and STRP.

The case gives the regulations of GVSML. SNP and STRP are the main and simple

polymorphisms in human health related research, which can be

The center extends the application of this document to other sequence variation data.

Health informatics

Genome Sequence Variation Markup Language (GSVML)

1 Scope

This document determines the location of GSPML, specifies the structure of GSPML, and gives the DTD and XML schemas of GSPML.

This document is applicable to the field of human health (including clinical practice, preventive medicine, translational research and clinical research) without compulsory change of data

Genome sequence variation data exchange in the case of library mode.

This document is applicable to species related to human health, such as humans, cell lines and preclinical test animals. This document does not apply to other

Biological species.

2 Normative references

The content of the following documents constitutes an indispensable clause of this document through normative references in the text. Among them, dated quotations

Only the version corresponding to the date is applicable to this document; for undated reference documents, the latest version (including all amendments) is applicable to

This document.

ISO 13606 (all parts) Health informatics electronic health record communication
(Healthinformatics-Electronichealth-
carerecordcommunication)

3 Terms and definitions

The following terms and definitions apply to this document.

3.1

Actor

Thing or person that provides stimulation to the system.

Note. Performers include people and other semi-automatic items (such as machines, computer tasks and systems).

3.2

Bioinformatics sequence markup language bioinformaticsequencemarkuplanguage; BSML
Extensible language specification and container for biological information data.

3.3

Cell markup language celmarkuplanguage; cel ML

An extensible markup language that provides a standard method for representation and exchange of computer-based biological models.

3.4

Cancer gene analysis project cancergeneanatomyproject; CGAP

Contains genomic expression data for various tumor tissues of humans and mice, and provides methods and tests for obtaining genomic data

A database of agent information.

3.5

Digital imaging and communication in medicine; DICOM

In the field of medical informatics, it is used to exchange digital information between medical imaging equipment (such as radiological imaging) and other systems, and to ensure

Its interoperability standards.