April 2008

Phenotype and Genotype Object Model (PAGE-OM) Specification

Beta 1

OMG Document Number: dtc/2008-04-03

Standard document URL: http://www.omg.org/spec/PAGE-OM/1.0/PDF

Associated File(s)*: http://www.omg.org/spec/PAGE-OM/20080101
http://www.omg.org/spec/PAGE-OM/20080102
http://www.omg.org/spec/PAGE-OM/20080103

- original files: lifesci/2008-01-03 (PIM-normative), lifesci/2008-01-04 (PSM-normative), lifesci/2008-01-02 (normative and non-normative)

This OMG document replaces the submission document (lifesci/2008-01-01, Alpha). It is an OMG Adopted Beta Specification and is currently in the finalization phase. Comments on the content of this document are welcome, and should be directed to issues@omg.org by October 20, 2008.

You may view the pending issues for this specification from the OMG revision issues web page *http://www.omg.org/issues/*.

The FTF Recommendation and Report for this specification will be published on November 10, 2008. If you are reading this after that date, please download the available specification from the OMG Specifications Catalog.

USE OF SPECIFICATION - TERMS, CONDITIONS & NOTICES

The material in this document details an Object Management Group specification in accordance with the terms, conditions and notices set forth below. This document does not represent a commitment to implement any portion of this specification in any company's products. The information contained in this document is subject to change without notice.

LICENSES

The companies listed above have granted to the Object Management Group, Inc. (OMG) a nonexclusive, royalty-free, paid up, worldwide license to copy and distribute this document and to modify this document and distribute copies of the modified version. Each of the copyright holders listed above has agreed that no person shall be deemed to have infringed the copyright in the included material of any such copyright holder by reason of having used the specification set forth herein or having conformed any computer software to the specification.

Subject to all of the terms and conditions below, the owners of the copyright in this specification hereby grant you a fully-paid up, non-exclusive, nontransferable, perpetual, worldwide license (without the right to sublicense), to use this specification to create and distribute software and special purpose specifications that are based upon this specification, and to use, copy, and distribute this specification as provided under the Copyright Act; provided that: (1) both the copyright notice identified above and this permission notice appear on any copies of this specification; (2) the use of the specifications is for informational purposes and will not be copied or posted on any network computer or broadcast in any media and will not be otherwise resold or transferred for commercial purposes; and (3) no modifications are made to this specification. This limited permission automatically terminates without notice if you breach any of these terms or conditions. Upon termination, you will destroy immediately any copies of the specifications in your possession or control.

PATENTS

The attention of adopters is directed to the possibility that compliance with or adoption of OMG specifications may require use of an invention covered by patent rights. OMG shall not be responsible for identifying patents for which a license may be required by any OMG specification, or for conducting legal inquiries into the legal validity or scope of those patents that are brought to its attention. OMG specifications are prospective and advisory only. Prospective users are responsible for protecting themselves against liability for infringement of patents.

GENERAL USE RESTRICTIONS

Any unauthorized use of this specification may violate copyright laws, trademark laws, and communications regulations and statutes. This document contains information which is protected by copyright. All Rights Reserved.

DISCLAIMER OF WARRANTY

RESTRICTED RIGHTS LEGEND

TRADEMARKS

COMPLIANCE

The copyright holders listed above acknowledge that the Object Management Group (acting itself or through its designees) is and shall at all times be the sole entity that may authorize developers, suppliers and sellers of computer software to use certification marks, trademarks or other special designations to indicate compliance with these materials.

Software developed under the terms of this license may claim compliance or conformance with this specification if and only if the software compliance is of a nature fully matching the applicable compliance points as stated in the specification. Software developed only partially matching the applicable compliance points may claim only that the software was based on this specification, but may not claim compliance or conformance with this specification. In the event that testing suites are implemented or approved by Object Management Group, Inc., software developed using this specification may claim compliance or conformance with the specification only if the software satisfactorily completes the testing suites.

OMG's Issue Reporting Procedure

All OMG specifications are subject to continuous review and improvement. As part of this process we encourage readers to report any ambiguities, inconsistencies, or inaccuracies they may find by completing the Issue Reporting Form listed on the main web page http://www.omg.org, under Documents, Report a Bug/Issue (http://www.omg.org/technology/agreement.htm).

Table of Contents

Preface

About the Object Management Group

OMG

Founded in 1989, the Object Management Group, Inc. (OMG) is an open membership, not-for-profit computer industry standards consortium that produces and maintains computer industry specifications for interoperable, portable, and reusable enterprise applications in distributed, heterogeneous environments. Membership includes Information Technology vendors, end users, government agencies, and academia.

OMG member companies write, adopt, and maintain its specifications following a mature, open process. OMG's specifications implement the Model Driven Architecture® (MDA®), maximizing ROI through a full-lifecycle approach to enterprise integration that covers multiple operating systems, programming languages, middleware and networking infrastructures, and software development environments. OMG's specifications include: UML® (Unified Modeling Language™); CORBA® (Common Object Request Broker Architecture); CWM™ (Common Warehouse Metamodel); and industry-specific standards for dozens of vertical markets.

More information on the OMG is available at *http://www.omg.org/*.

OMG Specifications

As noted, OMG specifications address middleware, modeling and vertical domain frameworks. A Specifications Catalog is available from the OMG website at:

http://www.omg.org/technology/documents/spec_catalog.htm

Specifications within the Catalog are organized by the following categories:

OMG Modeling Specifications

UML

MOF

XMI

CWM

Profile specifications.

OMG Middleware Specifications

CORBA/IIOP

IDL/Language Mappings

Specialized CORBA specifications

CORBA Component Model (CCM).

Platform Specific Model and Interface Specifications

CORBAservices

CORBAfacilities

OMG Domain specifications

OMG Embedded Intelligence specifications

OMG Security specifications.

All of OMG's formal specifications may be downloaded without charge from our website. (Products implementing OMG specifications are available from individual suppliers.) Copies of specifications, available in PostScript and PDF format, may be obtained from the Specifications Catalog cited above or by contacting the Object Management Group, Inc. at:

OMG Headquarters
140 Kendrick Street
Building A, Suite 300
Needham, MA 02494
USA
Tel: +1-781-444-0404
Fax: +1-781-444-0320
Email: *pubs@omg.org*

Certain OMG specifications are also available as ISO standards. Please consult *http://www.iso.org*

Typographical Conventions

The type styles shown below are used in this document to distinguish programming statements from ordinary English. However, these conventions are not used in tables or section headings where no distinction is necessary.

Times/Times New Roman - 10 pt.: Standard body text

Helvetica/Arial - 10 pt. Bold: OMG Interface Definition Language (OMG IDL) and syntax elements.

`Courier - 10 pt. Bold:` Programming language elements.

Helvetica/Arial - 10 pt: Exceptions

Terms that appear in *italics* are defined in the glossary. Italic text also represents the name of a document, specification, or other publication.

1 Scope

Please see Chapter 6 - Introduction.

2 Conformance

The normative parts of this specification are:

- Platform independent model expressed in the attached XML file created according to XMI format rules, v2.1, using program Enterprise Architect, version 7.0.817.1, UML version 2.1.
- Platform specific model, representing an XML-based data exchange format, defined by an XML Schema definition.

Any implementation using or producing given data exchange format is considered complying with this specification.

Any platform specific model that is derived from the platform independent model defined by this specification is also considered compliant.

If there is any inconsistency, or discrepancy between generality and specificity, between the platform independent and platform specific model, the platform specific model has precedence.

The normative parts are expressed in the accompanied files in a document whose number is given in Appendix A (or elsewhere in this document). Parts of these files may also appear in the explanatory text of this document. If they do and if there are some differences or discrepancies the contents of the normative accompanied files has precedence.

Regarding the use of ontology, it is not normative to use the ones listed in the “Ontology” section of this document, but it is highly suggested. The reason why this specification is not stricter about it is the reflection of the fact that ontology is a moving target in the bio-community and insisting on using only specific ones, may harm the usability of this specification.

3 Normative References

There are no normative references associated with this specification.

4 Terms and Definitions

Please see Annex C – Glossary.

5 Supporting Organizations

The following organizations have been involved in the process of developing, prototyping, and/or reviewing this specification. The authors thank them for participating and giving their valuable input.

- Institute for Molecular Medicine Finland, University of Helsinki, Finland
- University of Bergen, Norway
- Cold Spring Harbor Laboratory (CSHL), USA
- National Center for Biotechnology Information (NCBI), USA
- Stanford University, USA
- Yale University, USA
- University College London,
- Centre for Health Informatics and Multiprofessional Education, UK
- University of Leicester, UK
- South African National Bioinformatics Institute, South Africa
- School of Medicine, Tsinghua University, China
- Shanghai Center for Bioinformation Technology, China
- Shanghai Information Center for Life Sciences, China
- Institute of Genomics & Integrative Biology (IGIB), India
    - Japan Science and Technology Agency (JST), Japan
    - Medical Institute of Bioregulation, Kyushu University, Japan
    - Mie University Life Science Research Center, Japan
- National Institute of Genetics (NIG), Japan

- RIKEN, Japan
- Tokyo Medical and Dental University (TMDU), Japan
- Tokyo Metropolitan Institute of Gerontology, Japan
- Graduate School of Medicine, University of Tokyo, Japan
- UNISYS, USA

5.1 Acknowledgements

The publishers of this document wish to express their appreciation to those listed below (in non-significant order) for their contributions of ideas and experience. Ultimately, the ideas expressed in this document are those of the authors and do not necessarily reflect the views or ideas of these individuals, nor does the inclusion of their names imply an endorsement of the final product.

Kimitoshi Naito, Akihiko Konagaya, Albert V. Smith, Anthony J. Brookes, David Fredman, Debasis Dash, Haseena Rajeevan, Heikki Lehväslaiho, Hideaki Sugawara, Hiroshi Mizushima, Juha Muilu, Jun Nakaya, Katsushi Tokunaga, Kei-Hoi Cheung, Kenshi Hayashi, Mark Woon, Masako Kuroda, Masashi Tanaka, Martin Senger, Matthew Darlison, Takeshi Tomiki, Toshio Kojima, Yasumasa Shigemoto, Yoshiji Yamada.

6 Introduction

Modern biology is striving to understand what factors generate inter-individual differences in structure, development, or behaviour within a species. Genetic factors ('genotypes') are undoubtedly of key importance in this equation, as are environmental conditions and even stochastic events, and much research is being done to elucidate how these things impact a range of normal and disease 'phenotypes' (i.e., the characteristics that can be observed). Progress in this area will ultimately lead to improved and increasingly personalised medical care, more productive agricultural and farming systems, and better solutions for environmental monitoring and control.

Many organisms are being explored and utilised with regards to deciphering genotype-phenotype relationships. On the genetic level, it is now possible to determine DNA sequences at extremely high-throughput, thereby revealing normal and pathogenic variation in and around genes, across individuals. On the level of the phenotype, an almost unlimited number of things could be influenced by genetic variants, and increasingly precise and standardised assays are being devised to measure and assess such things.

Both genetic and phenotype datasets need to be managed and combined to elucidate genotype-phenotype relationships. The genetic datasets have so far mostly comprised assays of specific sites of variation (creating 'genotypes'), which have recently been growing exponentially due to technical advances. A subsequent new era is emerging wherein large DNA stretches (and perhaps complete genetic complements) are being fully (re)sequenced in relevant individuals. Phenotype datasets, both in human and non-human model organisms, are also being scaled up via the investigation of thousands or tens of thousands of individuals. Indeed, study integration is a major feature of current research; for example, existing Genome-wide Association (GWA) data done on various different studies are being merged in the ENAGE (ref) program leading to a potential GWA resource of 80,000 samples. As a result of this rapid progress in the modes of both genetic and phenotype analysis, the amount of genotype-phenotype data being generated is increasing at an exponential rate. The potential for converting this raw 'data' into useful 'knowledge' is therefore immense, but achieving this effectively will require distinct studies to be inter-related, cross-validated, and compared. In short, there is now an urgent need to ensure that genotype-phenotype investigations are supported by technologies that make them interoperable. Key to achieving this is a robust reference model for these types of data, via which all relevant domain knowledge can be effectively integrated and disseminated. That is the precisely the motivation behind building the Phenotype And GEnotype Object Model (PAGE-OM).

Anticipated user communities for PAGE-OM include biomedical researchers, clinicians, people involved in teaching and training such groups, and information technologists working in support of these domains.

Given the modern data production systems now at their disposal, biomedical researchers are constantly facing challenges of genotype-phenotype data management (e.g., storage, retrieval, tracking, reformatting, merging) and this directly impacts their ability to effectively analyse, share, and report their various datasets. Clinicians - for whom genotype-phenotype relationships are only one of many things to be considered - require streamlined and standardised access to explanations about DNA sequence alternatives and how and in which situations they may impact their clinical work. All of these needs will be helped by the development of PAGE-OM, which specifies the rational and functional segmentation of genotype-phenotype information into sub-components that are operationally distinct (the high-level objects in the model) and lays out the natural inter-relationships between these components. Once biomedical researchers and clinicians come to be familiar with these logical components and their relationships, it will be far easier for both groups to process and make sense of genotype-phenotype information,

both within their own fields and in the important task of communicating between their domains of work. For this reason, it is also highly desirable that teachers of their two professions become familiar with the object modelling concepts, in order that they may suitably educate the researchers and doctors of the future.

Information technologists working in the biomedical fields will, more than any other group, benefit from understanding PAGE-OM in depth. These individuals have the job of providing the universe of databases, data pipelines, analysis tools, search functionalities, and exchanges protocols in which genotype-phenotype information will exist and be processed for exploitation. In the absence of any data model for genotype-phenotype information, these workers would each have to devise their own data model. Not only would this constant re-invention of the wheel be highly wasteful of manpower, it is also likely that many of the solutions they come up with would be suboptimal - given the limited experience that any one IT individual or IT team would usually have. Instead, PAGE-OM has been designed and piloted by a consortium of expert participants from nearly 50 Institutions engaged in many diverse genotype-phenotype projects. The PAGE-OM therefore provides a first-version common language, well-documented entity list, and carefully considered array of entity-relationships that IT workers can take directly off the shelf, with confidence that it should need nothing more than small adjustments or additions to serve their own specific needs. Additionally, as more and more IT groups adopt PAGE-OM and base their systems upon it, those systems will naturally become more and more similar and able to inter-communicate. This is an absolute necessity if the totality of genotype-phenotype information is ever to be merged into a single virtual corpus, for holistic and optimal utilisation.

To enable PAGE-OM to achieve its goals, it is structured as several high level concepts – each of which can be used as a standalone model with its own object classes and relationships thereof. These concepts are: Marker, Assay, Sample, Genotype, Frequency, Phenotype, and Experiment. By way of example uses; a company providing DNA analysis kits might only need to use the Marker and Assay parts of this model. A genome variation database might use solely the Marker component, or perhaps the Assay, Sample, Genotype, and Frequency entities as well listing variant findings in population groups. A project involved in collecting and examining clinical samples might use only Sample and Phenotype parts, and if that team subsequently undertook DNA analysis it could extend their data systems by incorporating some or all of the remaining concepts in order to fully describe whatever genotype-phenotype relationships they might discover. Such flexibility is a necessary and innate characteristic of PAGE-OM, and that extends into the way we have matched the design principles and various classes of the model to those of other data standards initiates that cover other domains of biomedical research.

In conclusion, we believe PAGE-OM provides a timely, robust, and useful data model, sufficiently developed and tested to justify formal registration and deployment to the many and various communities engaged in genotype-phenotype data handling.

7 Platform Independent Model

The platform independent model is expressed as a UML model, UML version 2.1. The normative is its XMI representation, attached in the "Accompanied files" as the file *PIM/ PAGE-OM_uml_2.1_xmi_2.1.xml*. The XMI was generated by the tool Enterprise Architect (EA), version 7.0.817.1, without including the EA specific extensions.

The XMI representing the platform independent model includes, because of the technical reasons and because of lacking package names, also the classes of the SNP-PML specification (that are used by this specification).

The basic data types (e.g., string or dateTime) are those inspired by the data types from the XML Schema (XML Schema Part 2: Datatypes - http://www.w3.org/TR/xmlschema-2/), where their exact descriptions can be found.

The full details with class and attributes description are attached in a generated file *PIM/PAGE-OM.rtf*. This file can be used interchangeably with the XMI file because both were generated from a model created by the Enterprise Architect tool – whose main file is also attached as file *PIM/PAGE-OM.EAP*. Note, however, that the latter file may have some EA specific expressions and extensions that are not normative for this specification.

For convenience, the whole PAGE-OM specification can be seen at http://www.openpml.org/page-om/model/.

7.1 Detailed model documentation

7.1.1 PAGE

Type: **Package «XSDschema»**

Package: PAGE-OM

Evidence and value - *(Logical diagram)*

Core ("simple") string, arithmetic data types and object references are modeled here, in the Value model. The model is based on concept developed in Generation Challenge Program: http://pantheon.generationcp.org/demeter/Values.html

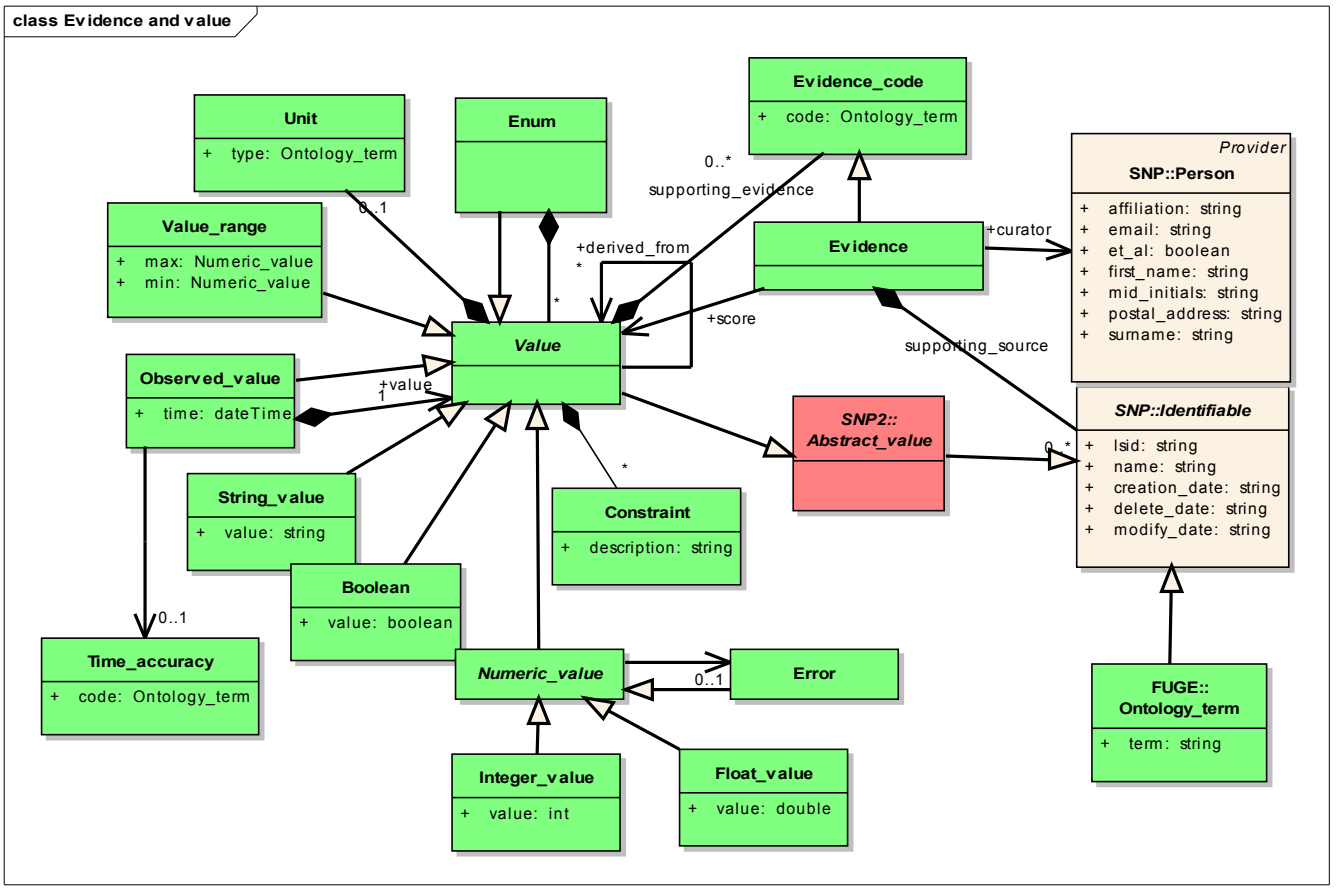
class Evidence and value
Unit
+ type: Ontology_term
Enum
Evidence_code
+ code: Ontology_term
Provider
SNP::Person
+ affiliation: string
+ email: string
+ et_al: boolean
+ first_name: string
+ mid_initials: string
+ postal_address: string
+ surname: string
Value_range
+ max: Numeric_value
+ min: Numeric_value
0..1
0..*
supporting_evidence
+derived_from
*
Evidence
+curator
+score
Value
supporting_source
Observed_value
+ time: dateTime
+value
1
SNP2::
Abstract_value
SNP::Identifiable
+ lsid: string
+ name: string
+ creation_date: string
+ delete_date: string
+ modify_date: string
String_value
+ value: string
Constraint
+ description: string
Boolean
+ value: boolean
0..1
Time_accuracy
+ code: Ontology_term
Numeric_value
0..1
Error
FUGE::
Ontology_term
+ term: string
Integer_value
+ value: int
Float_value
+ value: double

Figure 7.1

Frequency - *(Logical diagram)*

Alleles (Genomic_alleles), genotypes (Consensus_genomic_genotype) and haplotypes (Genomic_haplotype) can have measured frequencies in population samples (Panels). In addition, heterozygosity (Heterozygosity) is a measure of observed variability of a polymorphic site (Genomic_variation) in a sub-population (Panel).

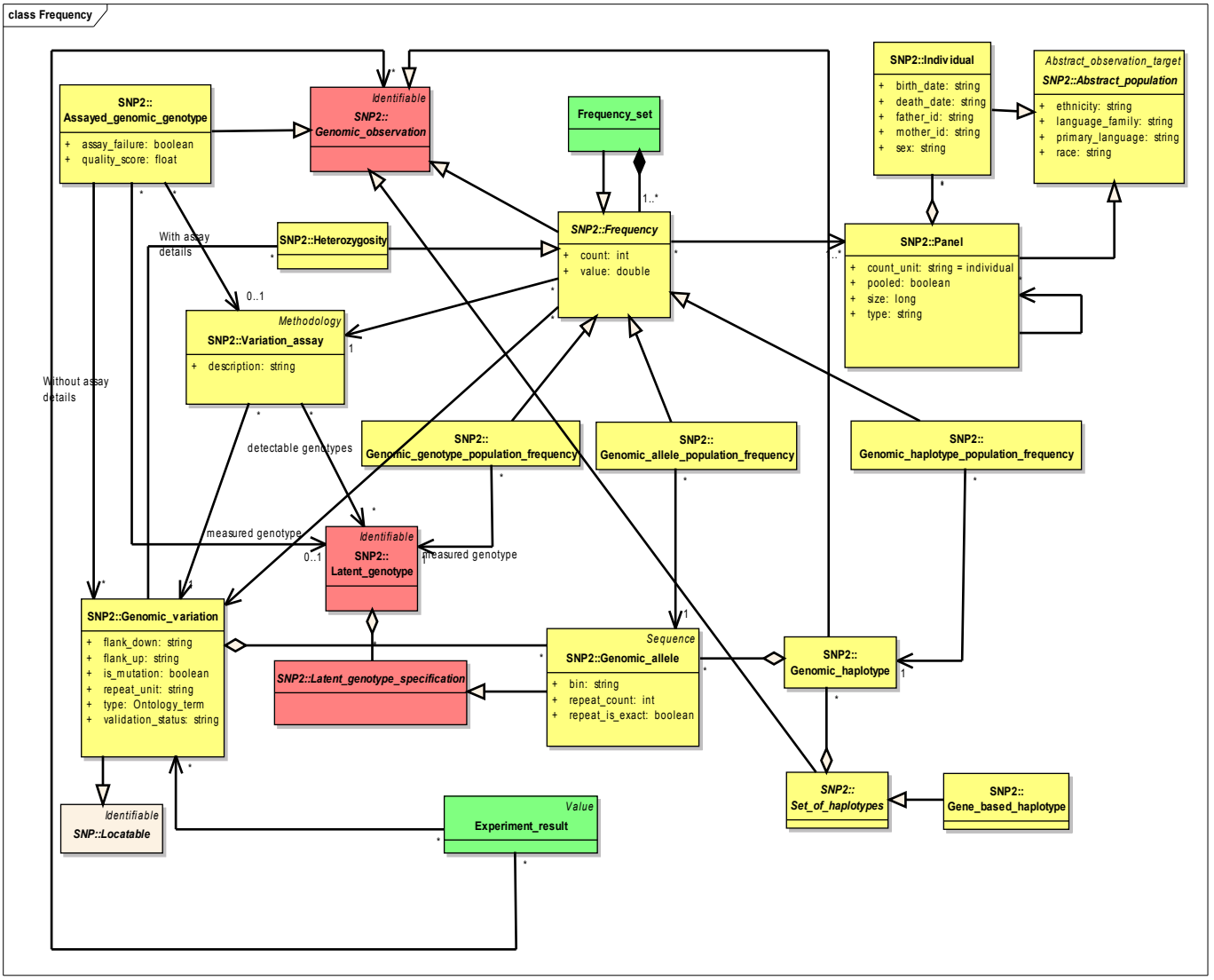

Figure 7.2

Genotype - *(Logical diagram)*

An assay (Variation_assay) is applied to a sample (Molecular_sample) to get a genotype (Assayed_genomic_genotype). One or more Assayed_genomic_genotype are used to reach a consensus (Consensus_genomic_genotype). The assays commonly use oligo primers to detect allelic variations (Genomic_alleles in Genomic_variation).

Assay can have reference to possible detectable latent genotypes (combination of alleles depending on ploidy). One of these combinations can be detected in single genotype measurement (Assayed_genomic_genotype) done using the specific assay.

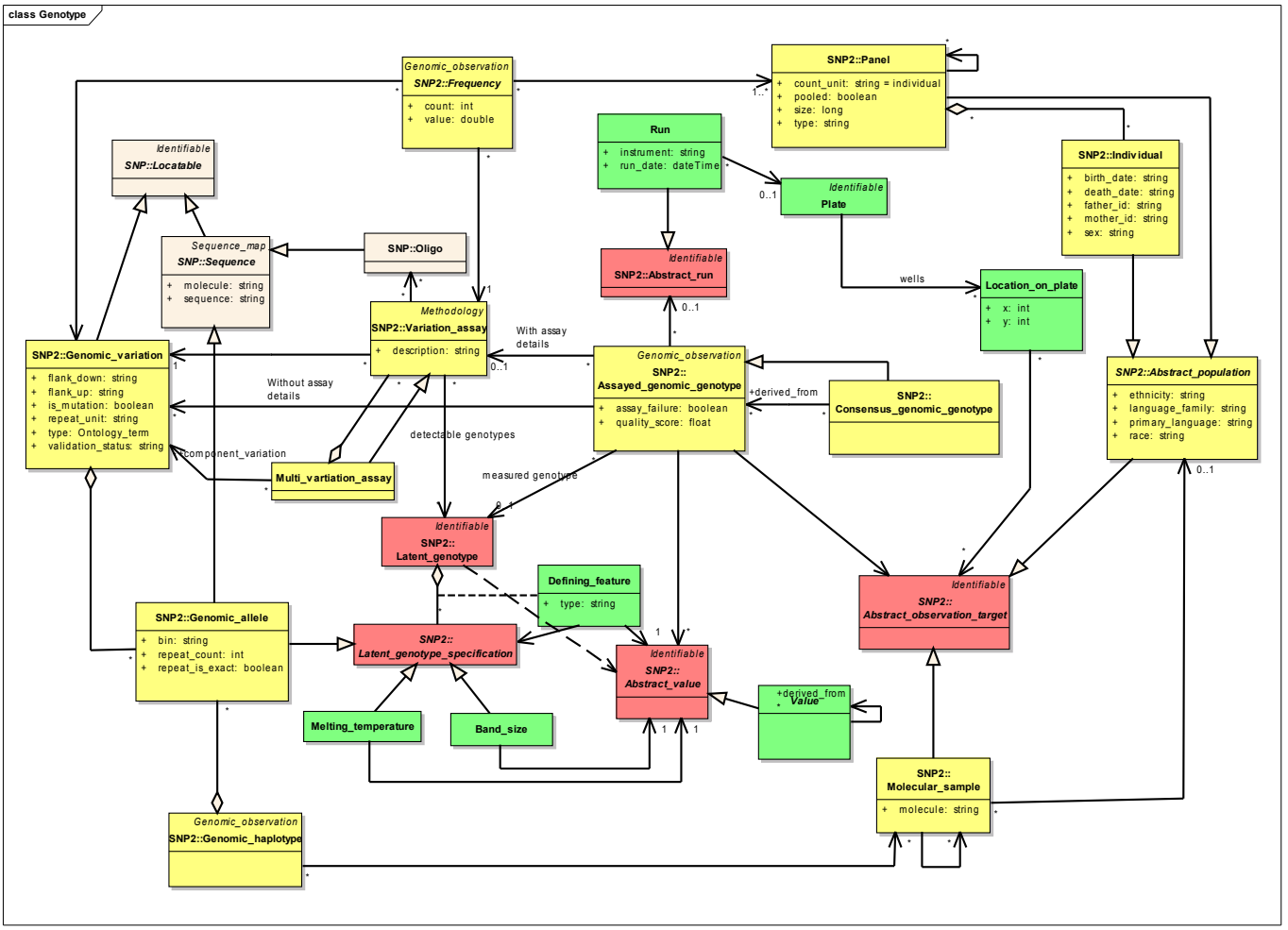

Figure 7.3

Genotype2 - *(Logical diagram)*

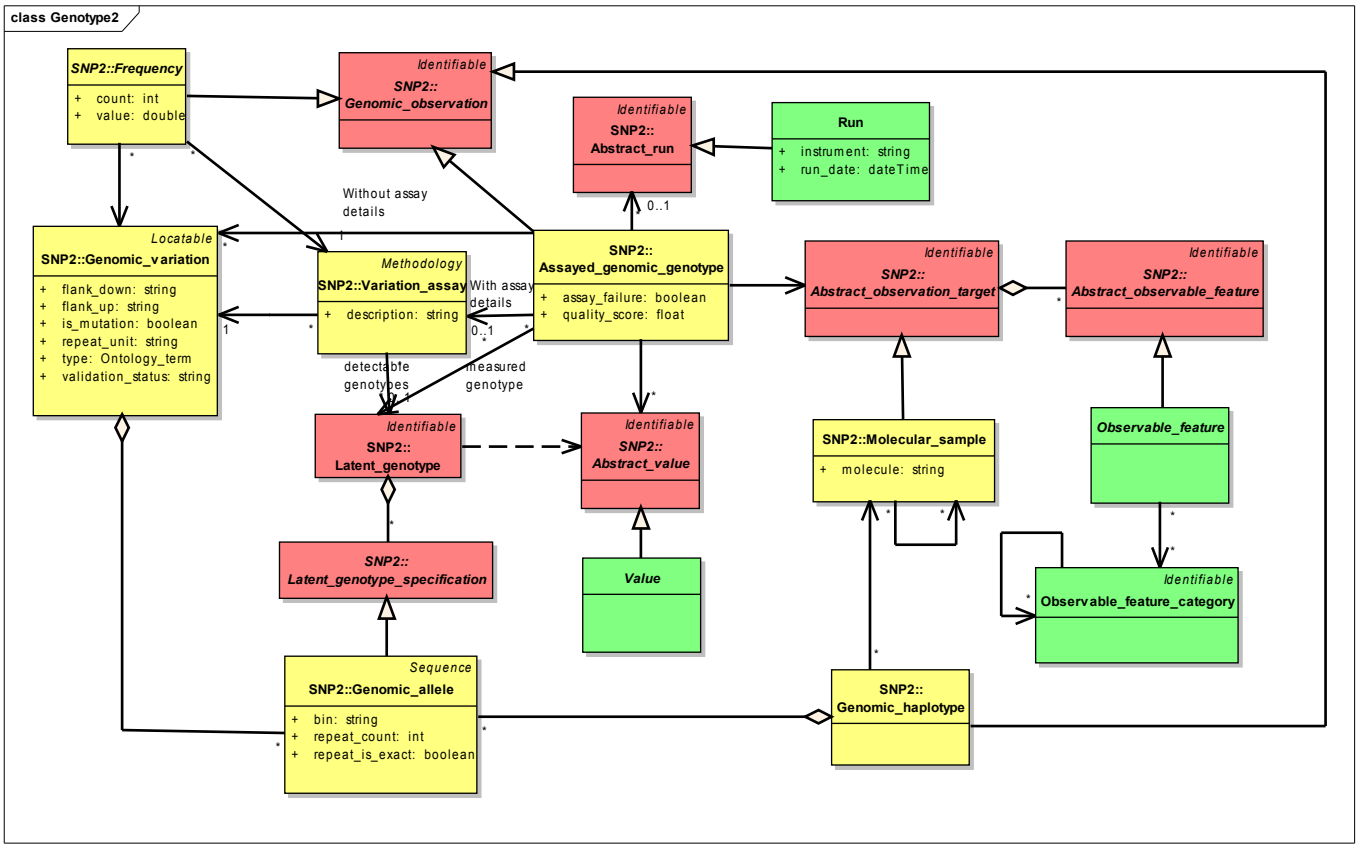

Figure 7.4

Identifiable - *(Logical diagram)*

All classes in the model inherit from Identifiable. In this way, their instances are uniquely identifiable. Any Identifiable instance must have either attribute "id," or attribute "LSID." Usually, an "id" is used to identify an instance within a known context, and the "LSID" is used when cross-referencing to a different context. The "LSID" attribute follows syntax as defined in the OMG Life Sciences Identifiers specification. All classes in the model can be linked with annotation (Annotation) and database cross-reference (Db_xref). Special kinds of cross references are Source for source of data, Distributor for the original database of the data, and Contributor for tracking editorial changes to data.

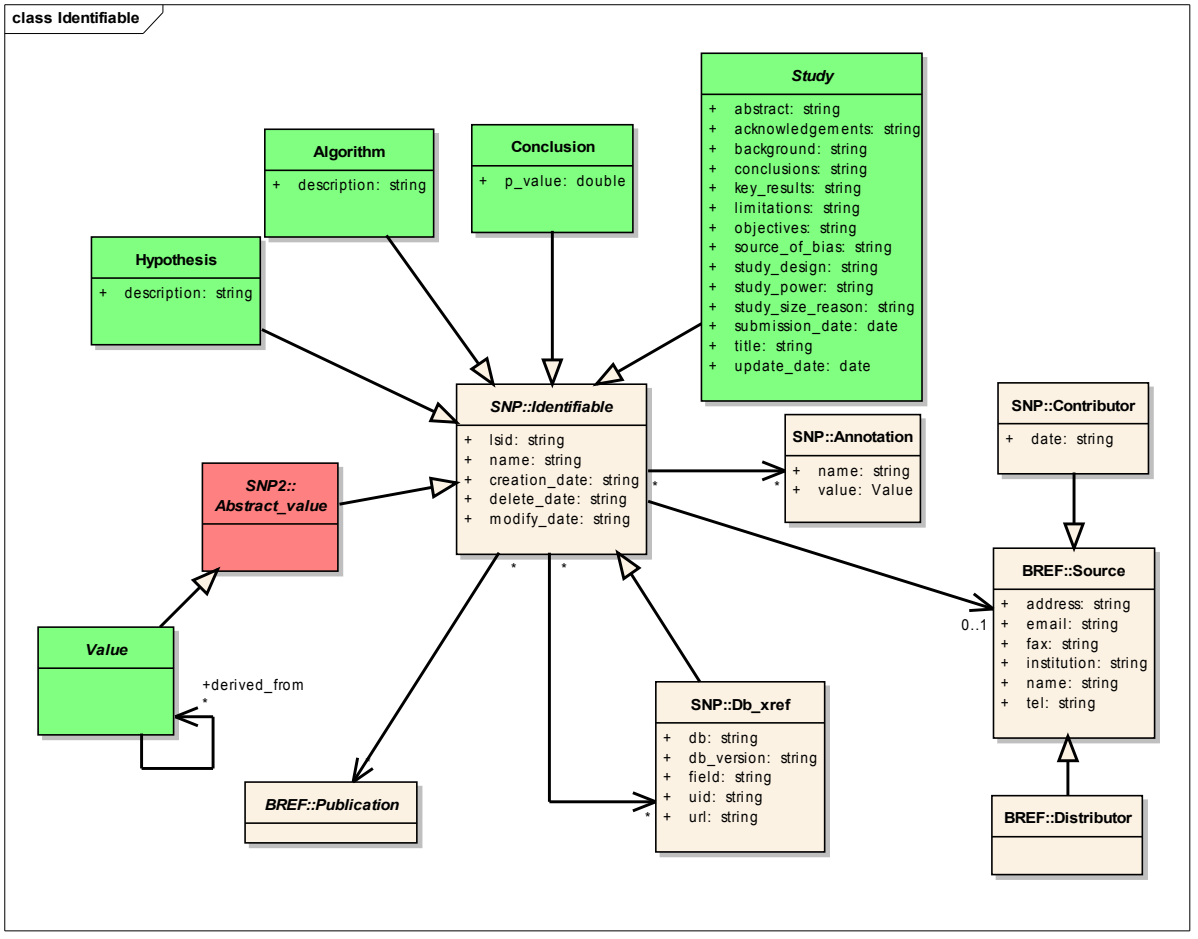

Figure 7.5

Phenotype - *(Logical diagram)*

Phenotype, environment and life style features are special cases of observable features from which observations can be made using specific observation methods. These observations lead to observed values obtained at specific time point from observation target (next diagram). Values can be also derived or categorized values (for example, high cholesterol level) in which case supporting values can be obtained from associated values, implemented using a recursion in the value model.

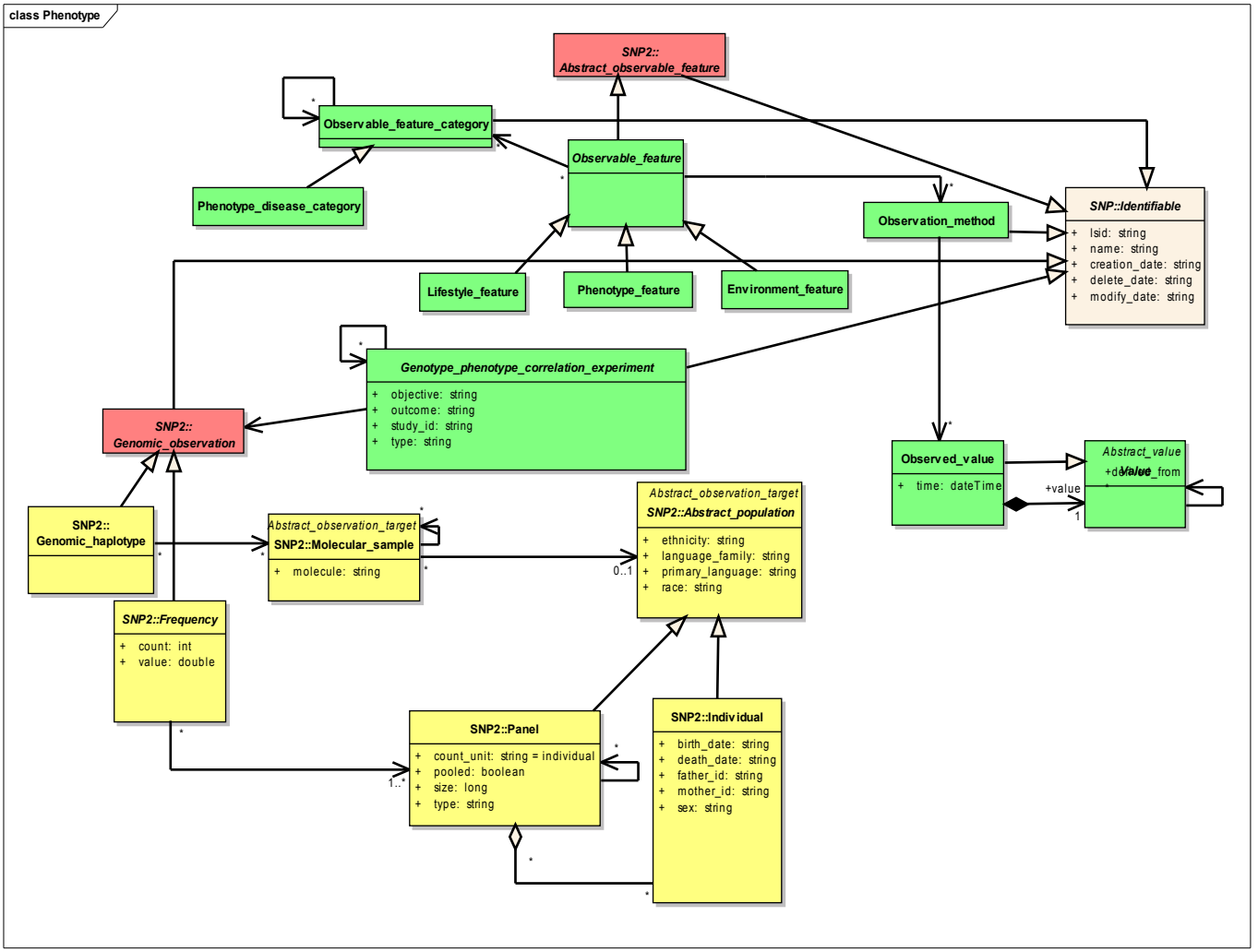

Figure 7.6

Phenotype 2 - *(Logical diagram)*

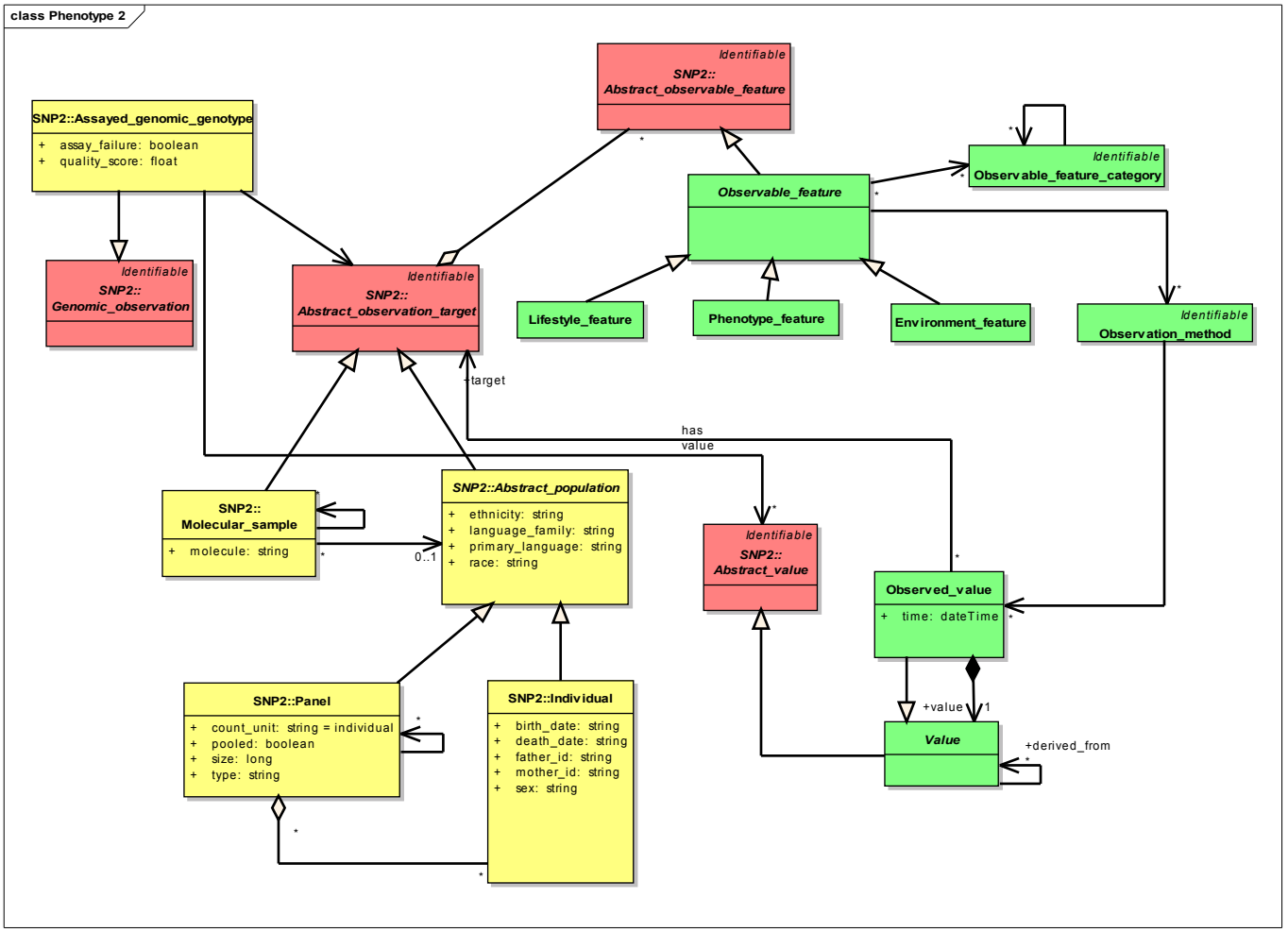

Figure 7.7

Study - *(Logical diagram)*

Studies, like association studies, are composed from set of experiments done over observation targets. These experiments leads to set of results (Experiment_result) documented by genomic observations and observed phenotype values.

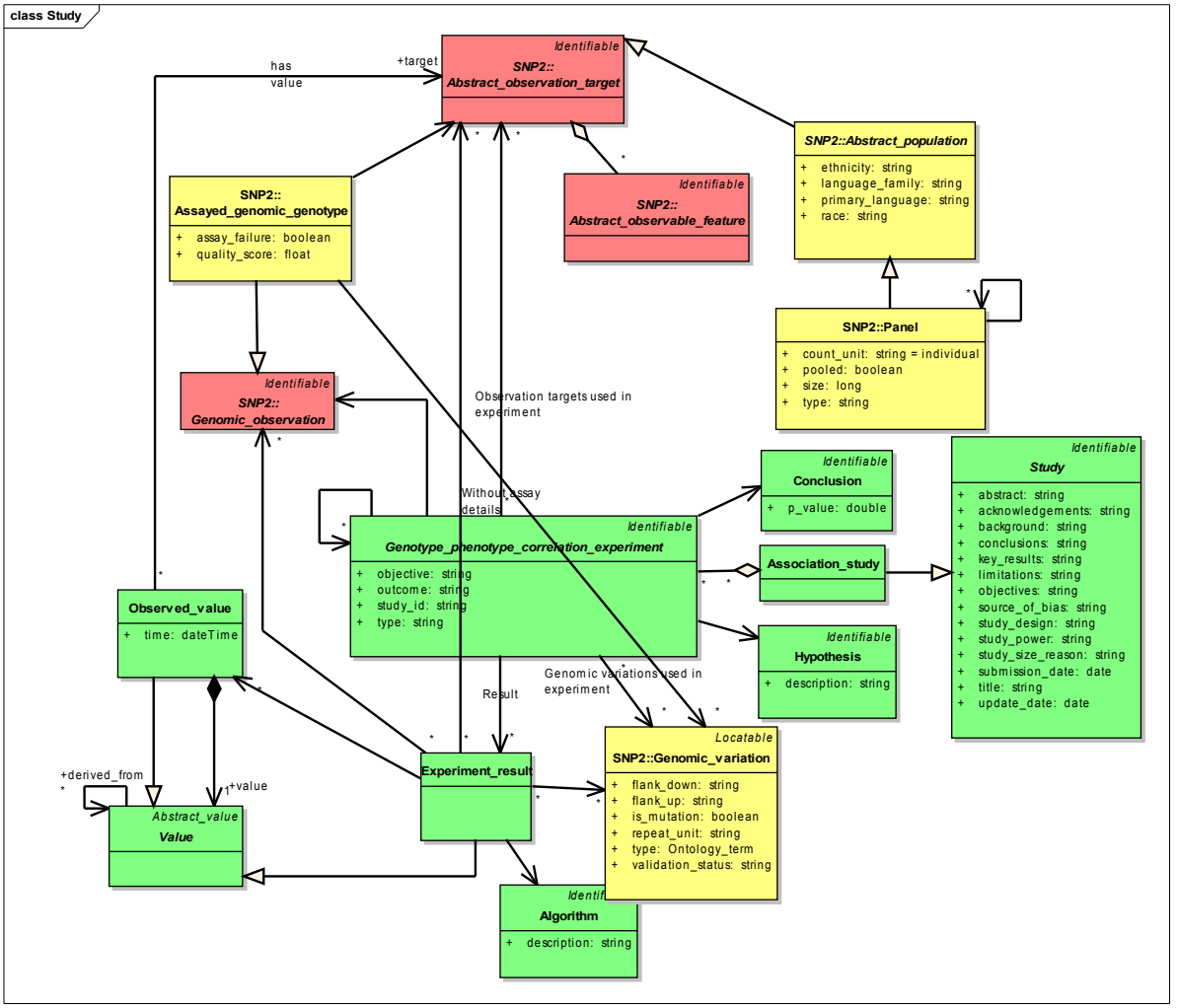

Figure 7.8

7.1.1.1 Algorithm

Type: Class__Identifiable
Package: PAGE

Document:
Step-by-step procedure for solving a problem

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Experiment_result	PAGE::Algorithm
Generalization_ Source -> Destination	PAGE::Algorithm	SNP::Identifiable

Attributes

Attribute	Notes
description **string**	Description of algorithm

7.1.1.2 Association_study

Type: Class__Study
Package: PAGE

Document:

Examination of genetic variation across the human genome, designed to identify genetic associations with observable phenotypes.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Association_study	PAGE::Study
Aggregation_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Association_study

7.1.1.3 Band_size

Type: Class__**Latent_genotype_specification**
Package: PAGE

Document:

DNA fragment length estimated from gel electrophoresis

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Band_size	Abstract_value
Generalization_ Source -> Destination	PAGE::Band_size	Latent_genotype_specification

7.1.1.4 Boolean

Type: Class__Value
Package: PAGE

Document:
Value of type boolean

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Boolean	PAGE::Value

Attributes

Attribute	Notes
value **boolean**	Boolean value

7.1.1.5 Conclusion

Type: Class__Identifiable
Package: PAGE

Document:
A reasoned judgment of an experiment

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Conclusion
Generalization_ Source -> Destination	PAGE::Conclusion	SNP::Identifiable

Attributes

Attribute	Notes
p_value **double**	Probability value

7.1.1.6 Constraint

Type: Class__
Package: PAGE

Document:

The contents of a Value can be limited by Constraints. Different types of Constraints allow various ways how to limit or validate one or more Value instances. The Constraint superclass only stores a string description of the Constraint. The actual full semantics of a constraint are specified in various subclasses described below.

But there are no subclasses in the PAGE-OM - because they are out of scope of PAGE-OM.

Connections

Connector	Source	Target
Aggregation_ Source -> Destination	PAGE::Constraint	PAGE::Value

Attributes

Attribute	Notes
description **string**	Description

7.1.1.7 Defining_feature

Type: AssociationClass__
Package: PAGE

Document:
Association class has list of values which are used in defining the instance of Latent_genotype_specifications (For example intensity values used in allele calling)

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Defining_feature	Abstract_value
Association_ Source -> Destination	PAGE::Defining_feature	Latent_genotype_specification

Attributes

Attribute	Notes
type **string**	Type of feature

7.1.1.8 Enum

Type: Class__Value
Package: PAGE

Document:
Enumeration contains list of Values

Connections

Connector	Source	Target
Aggregation_ Source -> Destination	PAGE::Value	PAGE::Enum
Generalization_ Source -> Destination	PAGE::Enum	PAGE::Value

7.1.1.9 Environment_feature

Type: Class__Observable_feature
Package: PAGE

Document:
Circumstances, objects, or conditions by which one is surrounded.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Environment_feature	PAGE::Observable_feature

7.1.1.10 Error

Type: Class__Numeric_value
Package: PAGE

Document:

Error value is numeric value of accuracy. Quality score

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Numeric_value	PAGE::Error
Generalization_ Source -> Destination	PAGE::Error	PAGE::Numeric_value

7.1.1.11 Evidence

Type: Class__Evidence_code
Package: PAGE

Document:

Evidence indicates reliability of a feature or simply documents its authoritative origin.

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Evidence	score score of value PAGE::Value
Aggregation supporting_source Source -> Destination	SNP::Identifiable	PAGE::Evidence
Association_ Source -> Destination	PAGE::Evidence	curator Curator of evidence SNP::Person
Generalization_ Source -> Destination	PAGE::Evidence	PAGE::Evidence_code

7.1.1.12 Evidence_code

Type: Class__

Package: PAGE

Document:

Evidence can be an EvidenceCode (which is a controlled vocabulary term such as a GO evidence code or ICIS Method code) but can be a more fully documented Evidence object (inheriting from EvidenceCode) generally curated by a specified person, a curator modeled as a Contact.

Its strength is expressed by the score (which is usually a numeric value between 0 and 1, but also other types of Value are allowed – e.g., an ontology term value).

The core of an evidence is its supporting source which can be anything (because it is identified by a SimpleIdentifier). Usual evidence sources are BiblioReferences, Studies and OntologyTerms. Reference (generationcp - http://pantheon.generationcp.org/demeter/Features.html)

Connections

Connector	Source	Target
Aggregation supporting_evidence Source -> Destination	PAGE::Evidence_code	PAGE::Value
Generalization_ Source -> Destination	PAGE::Evidence	PAGE::Evidence_code

Attributes

Attribute	Notes
code **Ontology_term**	Evidence code as specified using ontology term

7.1.1.13 Experiment_result

Type: Class__Value
Package: PAGE

Document:
Result of experiment

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Experiment_result	PAGE::Observed_value
Association_ Source -> Destination	PAGE::Experiment_result	Genomic_variation
Association_Result Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Experiment_result
Association_ Source -> Destination	PAGE::Experiment_result	Genomic_observation
Association_ Source -> Destination	PAGE::Experiment_result	PAGE::Algorithm
Generalization_ Source -> Destination	PAGE::Experiment_result	PAGE::Value
Association_ Source -> Destination	PAGE::Experiment_result	Abstract_observation_target

7.1.1.14 Float_value

Type: Class__**Numeric_value**
Package: PAGE

Document:
Value of type float

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Float_value	PAGE::Numeric_value

Attributes

Attribute	Notes
value **double**	Value

7.1.1.15 Frequency_set

Type: Class__Frequency
Package: PAGE

Document:
Set of frequencies

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Frequency_set	Frequency
Aggregation_ Source -> Destination	Frequency	PAGE::Frequency_set

7.1.1.16 Genotype_phenotype_correlation_experiment

Type: Class__Identifiable
Package: PAGE

Document:
Family or case control based association study

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Conclusion

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	SNP::Identifiable
Association_Genomic variations used in experiment Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	Genomic_variation
Association_Result Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Experiment_result
Association_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Genotype_phenotype_ correlation_experiment
Association_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Hypothesis
Aggregation_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Association_study
Association_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	Genomic_observation
Association_Observation targets used in experiment Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	Abstract_observation_target

Attributes

Attribute	Notes
objective **string**	Objective of experiment
outcome **string**	Outcome of experiment
study_id **string**	Identifier of study
type **string**	Type of experiment

7.1.1.17 Hypothesis

Type: Class__Identifiable
Package: PAGE

Document:
Hypothesis of study

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Hypothesis	SNP::Identifiable
Association_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	PAGE::Hypothesis

Attributes

Attribute	Notes
description **string**	Description of hypothesis

7.1.1.18 Integer_value

Type: Class__Numeric_value
Package: PAGE

Document:
Value of type integer

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Integer_value	PAGE::Numeric_value

Attributes

Attribute	Notes
value **int**	Integer value

7.1.1.19 Lifestyle_feature

Type: Class__Observable_feature
Package: PAGE

Document:
Way of life of an individual or panel

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Lifestyle_feature	PAGE::Observable_feature

7.1.1.20 Location_on_plate

Type: Class__
Package: PAGE

Document:

X,Y plate_positions (wells). Numbering starts from one. Each well can contain one or more observation targets (molecular samples) prepared for measurement using one or more variation assays (e.g., assay multiplexing Assay_set). Note: These are optional laboratory specific details (Sample and Assay information is in Assayed_genomic_genotype).

Connections

Connector	Source	Target
Association_wells Source -> Destination	PAGE::Plate	PAGE::Location_on_plate
Association_ Source -> Destination	PAGE::Location_on_plate	Abstract_observation_target

Attributes

Attribute	Notes
x **int**	x coordinate of plate.
y **int**	y coordinate of plate

7.1.1.21 Melting_temperature

Type: Class__Latent_genotype_specification
Package: PAGE

Document:

The temperature at which DNA goes from a double-stranded to a single-stranded state. Unit of temperature is Celsius.

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Melting_temperature	Abstract_value
Generalization_ Source -> Destination	PAGE::Melting_temperature	Latent_genotype_specification

7.1.1.22 Multi_vartiation_assay

Type: Class__Variation_assay

Package: PAGE

Document:

Multi_variation_assay is a collection of assays which may be used simultaneously. Examples would be multiplex assays, micro-array based assays, or a panel of single-plex assays that share some common feature or purpose.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Multi_vartiation_assay	Variation_assay
Association_ Source -> Destination	PAGE::Multi_vartiation_assay	component_variation Genomic_variation
Aggregation_ Source -> Destination	Variation_assay	PAGE::Multi_vartiation_assay

7.1.1.23 Numeric_value

Type: Class__Value

Package: PAGE

Document:

Numeric value

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Integer_value	PAGE::Numeric_value

Connector	Source	Target
Association_ Source -> Destination	PAGE::Numeric_value	PAGE::Error
Generalization_ Source -> Destination	PAGE::Float_value	PAGE::Numeric_value
Generalization_ Source -> Destination	PAGE::Error	PAGE::Numeric_value
Generalization_ Source -> Destination	PAGE::Numeric_value	PAGE::Value

7.1.1.24 Observable_feature

Type: Class__Abstract_observable_feature
Package: PAGE

Document:
Measurable feature of observable (e.g., size of nose).

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Observable_feature	Abstract_observable_feature
Generalization_ Source -> Destination	PAGE::Lifestyle_feature	PAGE::Observable_feature
Generalization_ Source -> Destination	PAGE::Environment_feature	PAGE::Observable_feature
Association_ Source -> Destination	PAGE::Observable_feature	PAGE::Observable_feature_category
Association_ Source -> Destination	PAGE::Observable_feature	PAGE::Observation_method
Generalization_ Source -> Destination	PAGE::Phenotype_feature	PAGE::Observable_feature

7.1.1.25 Observable_feature_category

Type: Class__Identifiable
Package: PAGE

Document:
Phenotype category or ontology

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Observable_feature_category	PAGE::Observable_feature_category
Generalization_ Source -> Destination	PAGE::Observable_feature_category	SNP::Identifiable
Generalization_ Source -> Destination	PAGE::Phenotype_disease_category	PAGE::Observable_feature_category
Association_ Source -> Destination	PAGE::Observable_feature	PAGE::Observable_feature_category

7.1.1.26 Observation_method

Type: Class__Identifiable
Package: PAGE

Document:
Method of observation (e.g., ruler). Can be also questionnaire.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Observation_method	SNP::Identifiable
Association_ Source -> Destination	PAGE::Observable_feature	PAGE::Observation_method
Association_ Source -> Destination	PAGE::Observation_method	PAGE::Observed_value

7.1.1.27 Observed_value

Type: Class__Value
Package: PAGE

Document:
Observation done at specific point in time.

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Experiment_result	PAGE::Observed_value
Association_ Source -> Destination	PAGE::Observed_value	PAGE::Time_accuracy
Association_ Source -> Destination	PAGE::Observed_value	value PAGE::Value
Association_has value Source -> Destination	PAGE::Observed_value	target Abstract_observation_target
Association_ Source -> Destination	PAGE::Observation_method	PAGE::Observed_value
Generalization_ Source -> Destination	PAGE::Observed_value	PAGE::Value

Attributes

Attribute	Notes
time **dateTime**	Time of observation

7.1.1.28 Phenotype_disease_category

Type: Class__Observable_feature_category
Package: PAGE

Document:
Disease ontology.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Phenotype_disease_category	PAGE::Observable_feature_category

7.1.1.29 Phenotype_feature

Type: Class__Observable_feature
Package: PAGE

Document:
Observable part of the structure, function or behavior of a living organism.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Phenotype_feature	PAGE::Observable_feature

7.1.1.30 Plate

Type: Class__Identifiable
Package: PAGE

Document:
A sample holder, for example a microtiter plate used in one or many Runs. Samples (Observation_targets) are positioned on the plate using Location_on_plate-class.

Connections

Connector	Source	Target
Association_wells Source -> Destination	PAGE::Plate	PAGE::Location_on_plate
Generalization_ Source -> Destination	PAGE::Plate	SNP::Identifiable
Association_ Source -> Destination	PAGE::Run	PAGE::Plate

7.1.1.31 Run

Type: Class__Abstract_run
Package: PAGE

Document:
The class contains information on execution of measurement experiments like time of execution and name of instrument.

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Run	PAGE::Plate
Generalization_ Source -> Destination	PAGE::Run	Abstract_run

Attributes

Attribute	Notes
instrument **string**	Name of the instrument
run_date **dateTime**	Date of run.

7.1.1.32 String_value

Type: Class__Value
Package: PAGE

Document:
Value of type string

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::String_value	PAGE::Value

Attributes

Attribute	Notes
value **string**	Actual value

7.1.1.33 Study

Type: Class__Identifiable
Package: PAGE

Document:
Research study

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Study	SNP::Identifiable
Generalization_ Source -> Destination	PAGE::Association_study	PAGE::Study

Attributes

Attribute	Notes
abstract **string**	Abstract
acknowledgements **string**	Acknowledgements
background **string**	Background information
conclusions **string**	Conclusions
key_results **string**	Key findings
limitations **string**	Limitations
objectives **string**	Objectives of study
source_of_bias **string**	Possible source of bias
study_design **string**	Study design
study_power **string**	Power of study

Attribute	Notes
study_size_reason **string**	Reason for study size
submission_date **date**	Submission date of study
title **string**	Title of study
update_date **date**	Date when study is updated

7.1.1.34 Time_accuracy

Type: Class__
Package: PAGE

Document:
Accuracy code for time. Accuracy code contains information on incompleteness of time. For example in some cases exact time is not known or cannot be given accurately for various reasons.

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Observed_value	PAGE::Time_accuracy

Attributes

Attribute	Notes
code **Ontology_term**	Accuracy code as defined in specific ontology

7.1.1.35 Unit

Type: Class__
Package: PAGE

Document:
Unit of value. Unit is defined using ontology term

Connections

Connector	Source	Target
Aggregation_ Source -> Destination	PAGE::Unit	PAGE::Value

Attributes

Attribute	Notes
type **Ontology_term**	Type of unit

7.1.1.36 Value

Type: Class__Abstract_value
Package: PAGE

Document:

Abstract class. Extension point for Value implementations. Value model is based on concept developed in Generation Challenge Program: http://pantheon.generationcp.org/demeter/Values.html

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Value	Abstract_value
Association_ Source -> Destination	PAGE::Evidence	score score of value PAGE::Value
Generalization_ Source -> Destination	PAGE::String_value	PAGE::Value
Generalization_ Source -> Destination	PAGE::Boolean	PAGE::Value
Aggregation_ Source -> Destination	PAGE::Value	PAGE::Enum
Association_ Source -> Destination	PAGE::Observed_value	value PAGE::Value
Aggregation_ Source -> Destination	PAGE::Constraint	PAGE::Value
Generalization_ Source -> Destination	PAGE::Value_range	PAGE::Value
Association_ Source -> Destination	PAGE::Value	derived_from PAGE::Value

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Experiment_result	PAGE::Value
Aggregation supporting_evidence Source -> Destination	PAGE::Evidence_code	PAGE::Value
Aggregation_ Source -> Destination	PAGE::Unit	PAGE::Value
Generalization_ Source -> Destination	PAGE::Numeric_value	PAGE::Value
Generalization_ Source -> Destination	PAGE::Observed_value	PAGE::Value
Generalization_ Source -> Destination	PAGE::Enum	PAGE::Value

7.1.1.37 Value_range

Type: Class__Value
Package: PAGE

Document:
Inclusive value range

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Value_range	PAGE::Value

Attributes

Attribute	Notes
max **Numeric_value**	maximum value
min **Numeric_value**	minimum value

7.1.2 SNP2

Type: **Package** «XSDschema»
Package: PAGE-OM

Haplotype - *(Logical diagram)*

Haplotype (Genomic_haplotype) is a set of polymorphisms (Genomic alleles) on a single chromosome (chromatid). Haplotypes may be grouped into sets like haplotype blocks (Haplotype_block) separated by recombination regions and gene based haplotypes (Gene_based_haplotype), which combines sets of haplotypes on same gene structure.

Depending on an assay used, a genotype (Consensus_genomic_genotype) may contain information about the phase of the detected Genomic_alleles. Alternatively, various additional methods (Haplotype_derivation_methods) can be used to measure or calculate haplotypes (Genomic_haplotypes) from the genotype data. The aim of many haplotype studies is to find haplotypes within a given sequence region that define most of the variation in populations (Set_of_haplotypes). Haplotype blocks (Haplotype_blocks) are clusters determined by linkage disequilibrium-based methods. The other common way to define clusters is use a the sequence region of the gene (Gene_based_haplotype).

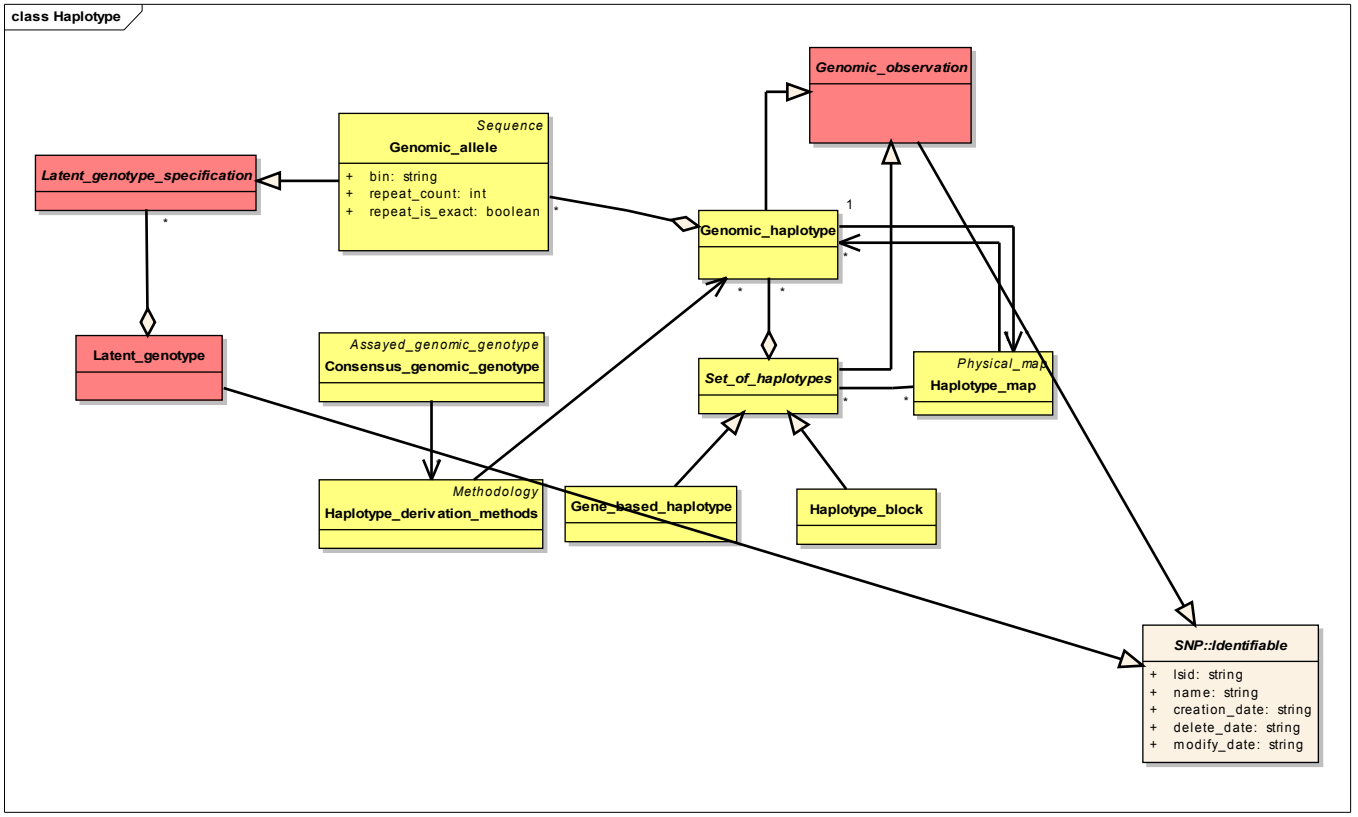

Figure 7.9

Map - *(Logical diagram)*

A map organizes genomic features and assigns them locations. The primary maps (Physical_map, Sequence_map) which have additive sequence residue count distances between markers, are separated here from secondary maps (Ordered_map) which at best can only deal with non-metric distances. For each type a map (Map) there is a corresponding location (Location). Hence a marker (Genetic_variation) can have several locations.

(Reference_genomic_location) in sequences (Reference_genomic_landmark) and additionally be located within a band (Cytogenetic_location) in a species-specific karyotype (Cytogenetic_map) as well as in (Genetic_location) several genetic maps (Genetic_map), e.g. in maps based on male and female recombination frequencies. Genomic assembly (Reference_genomic_assembly) is a physical sequence map which is combination of genomic sequences (Reference_genomic_landmark). Feature locations can also be expressed in chromosomal locations in an assembly(Reference_genomic_location_in_assembly).

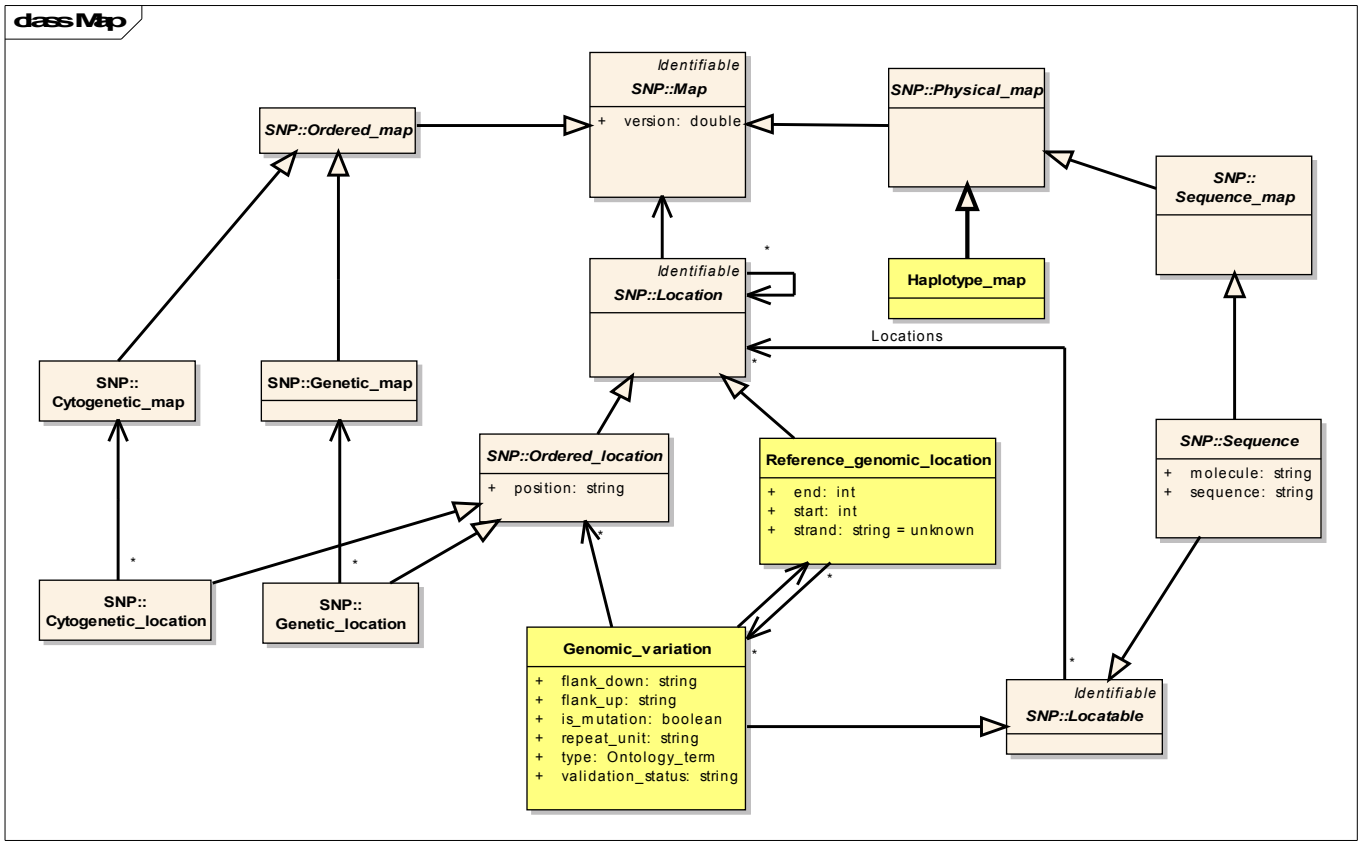

Figure 7.10

Sample - *(Logical diagram)*

An Individual (Individual) interacts with the rest of the model by giving one or more samples (Molecular_samples) from different tissues. Mixed samples and more complicated sample organisations are modeled by allowing Molecular_samples and Panels be composed of multiples of themselves. In these cases it is common that individuals can not be identified.

An individual (Individual) has parent IDs as attributes. This information makes it possible to reconstruct pedigrees when needed. A taxon identifies the taxonomic group, typically species, the individual or the population sample belongs to. Further, they can be place on a geographic map (Geographic_location).

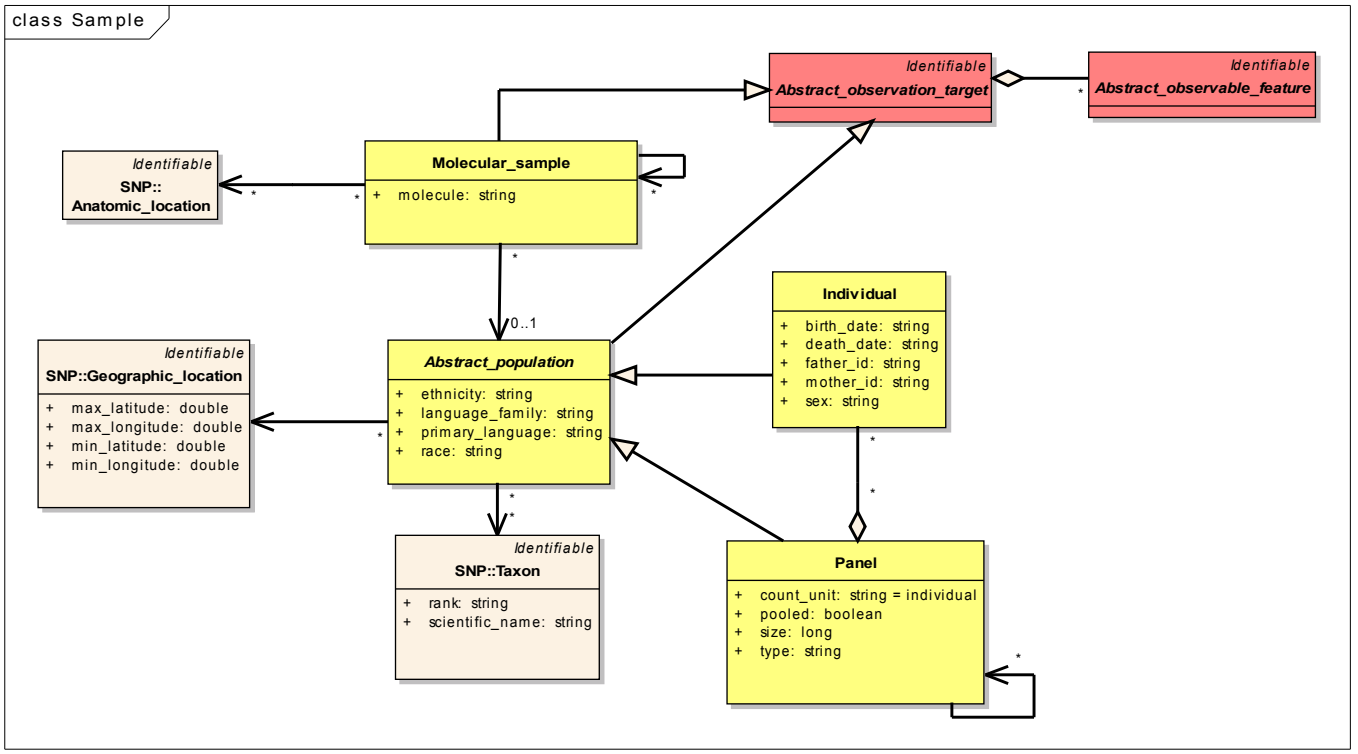

Figure 7.11

Sequence - *(Logical diagram)*

All sequences inherit from a generic Sequence class. All sequences (Sequence) and sequence features (e.g. Genomic_variation) can be located within a genomic sequence (Reference_genomic_sequence). This is modeled by common inheritance from an abstract superclass Locatable that can have several locations (Genomic_reference_location).

A Reference_genomic_landmark is any accessioned sequence within Reference_genomic_assembly. The model allows for multiple assemblies. Any location within a landmark and therefore in an assembly is called Reference_genomic_location. Any variable site in an assembly is a Genomic_polymorphism. The variable sequences within Genomic_variations are called Genomic_alleles.

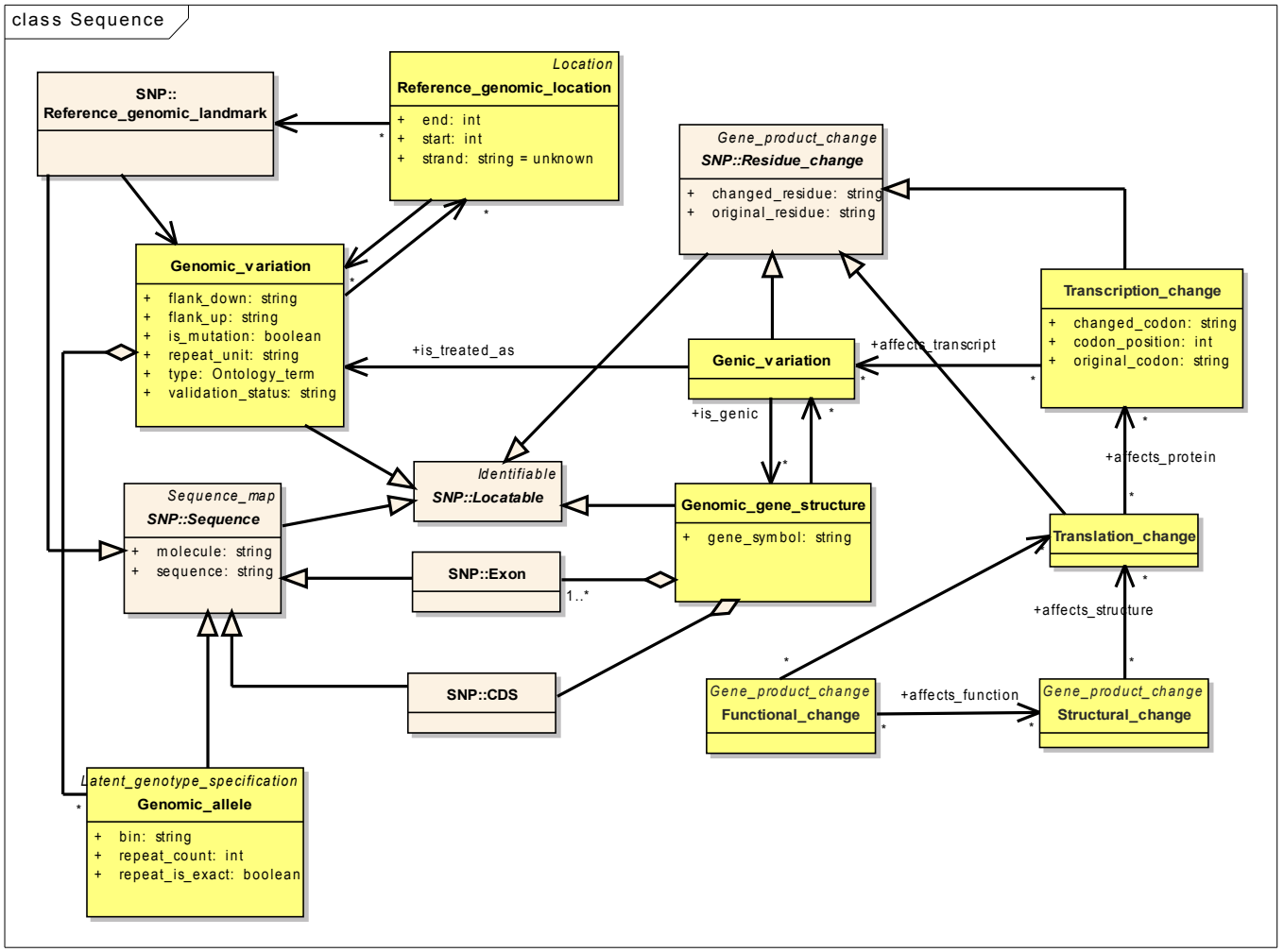

Figure 7.12

Sequence 2 - *(Logical diagram)*

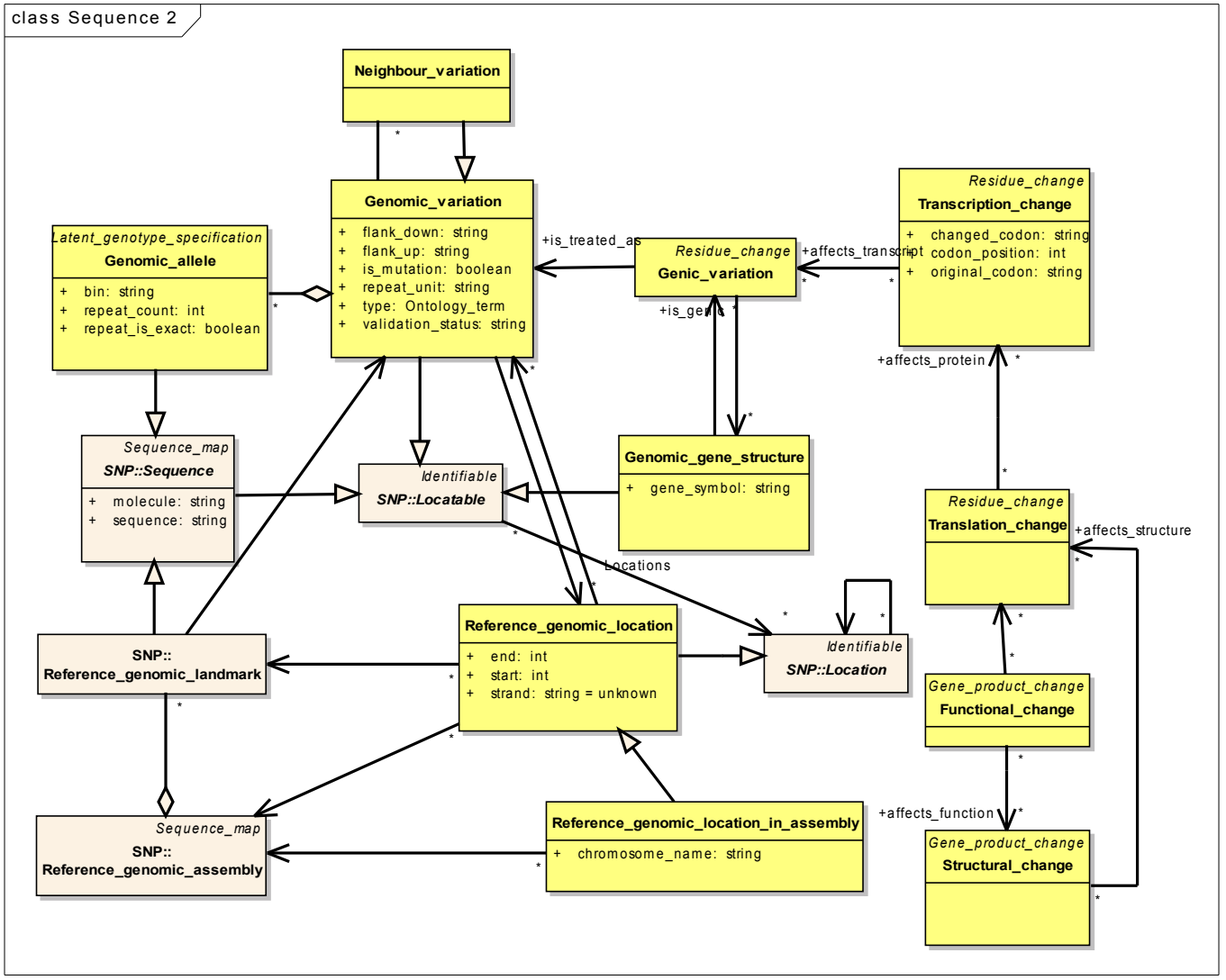

Figure 7.13

7.1.2.1 Abstract_observable_feature

Type: Class__Identifiable
Package: SNP2

Document

Abstract extension point for observable (Phenotype) features. Measurable feature of observable (e.g., size of nose).

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Observable_feature	Abstract_observable_feature
Generalization_ Source -> Destination	Abstract_observable_feature	SNP::Identifiable
Aggregation_ Source -> Destination	Abstract_observable_feature	Abstract_observation_target

7.1.2.2 Abstract_observation_target

Type: Class__Identifiable
Package: SNP2

Document

Abstract_observation_target is abstract super class of observable properties of objects like samples and individuals. Capable of being observed. For example nose size or eye color (Phenotype features).

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Molecular_sample	Abstract_observation_target
Association_ Source -> Destination	Assayed_genomic_genotype	Abstract_observation_target
Generalization_ Source -> Destination	Abstract_observation_target	SNP::Identifiable
Aggregation_ Source -> Destination	Abstract_observable_feature	Abstract_observation_target

Connector	Source	Target
Generalization_ Source -> Destination	Abstract_population	Abstract_observation_target
Association_ Source -> Destination	PAGE::Location_on_plate	Abstract_observation_target
Association_has value Source -> Destination	PAGE::Observed_value	target Abstract_observation_target
Association_ Source -> Destination	PAGE::Experiment_result	Abstract_observation_target
Association_Observation targets used in experiment Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	Abstract_observation_target

7.1.2.3 Abstract_population

Type: Class__Abstract_observation_target

Package: SNP2

Document

An interbreeding set of individuals, from whom a Panel is drawn. (Population in SNP-PML) . Extends Abstract_observation_target, which is abstract class for all entities from which one can make genotype or phenotype measurements or observations.

Connections

Connector	Source	Target
Association_ Source -> Destination	Molecular_sample	Abstract_population
Association_ Source -> Destination	Abstract_population	SNP::Geographic_location
Association_ Source -> Destination	Abstract_population	SNP::Taxon
Generalization_ Source -> Destination	Individual	Abstract_population
Generalization_ Source -> Destination	Panel	Abstract_population

Connector	Source	Target
Generalization_ Source -> Destination	Abstract_population	Abstract_observation_target

Attributes

Attribute	Notes
ethnicity **string**	Additional ethnic category of the population sample or "mixed".
language_family **string**	Language family name or code, e.g., as in Ethnologue
primary_language **string**	Language spoken (name or code), e.g., as in Ethnologue
race **string**	Broad ethnic category of the population sample or "mixed."

7.1.2.4 Abstract_run

Type: Class__Identifiable
Package: SNP2

Document

Abstract extension point for experiment runs.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Abstract_run	SNP::Identifiable
Association_ Source -> Destination	Assayed_genomic_genotype	Abstract_run
Generalization_ Source -> Destination	PAGE::Run	Abstract_run

7.1.2.5 Abstract_value

Type: Class__Identifiable
Package: SNP2

Document

Abstract class, extension point for values.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	PAGE::Value	Abstract_value
Association_ Source -> Destination	PAGE::Defining_feature	Abstract_value
Association_ Source -> Destination	PAGE::Band_size	Abstract_value
Association_ Source -> Destination	PAGE::Melting_temperature	Abstract_value
Generalization_ Source -> Destination	Abstract_value	SNP::Identifiable
Dependency_ Source -> Destination	Latent_genotype	Abstract_value
Association_ Source -> Destination	Assayed_genomic_genotype	Abstract_value

7.1.2.6 Assayed_genomic_genotype

Type: Class__Genomic_observation

Package: SNP2

Document

The result of applying a Variation_assay to an individual, to reveal one or more of the Genomic_alleles carried by that individual. This term applies to the observed data rather than to the inferred state of the individual (see Consensus_genomic_genotype). Thus the same individual might have several different Assayed_genomic_genotypes at the same Genomic_variation, where the variation might be due to differing assays, experimental error, dominant systems, missing data, and so forth. Synonym: Measurement.

Association to Genomic_allele is replaced by Latent_genotype, which has possible alleles (Genomic_alleles) as an extension of abstract class Latent_genotype_specification.

Associations from Assayed_genomic_genotype to Value and Run. Value contains raw data like intensity values and Run gives information about experiment run like date and instrument type.

Association from Assayed_genomic_genotype to Molecular_sample is replaced by association to Abstract_observation_target, which is abstract super class of Molecular_sample Assayed_genomic_genotype is now sub-class of Genomic_observation. Genomic_observation is also super class of Frequency and Genomic_Haplotype and Haplotype_block.

Connections

Connector	Source	Target
Association_measured genotype Source -> Destination	Assayed_genomic_genotype	Latent_genotype
Association_ Source -> Destination	Assayed_genomic_genotype	Abstract_run
Association_ Source -> Destination	Assayed_genomic_genotype	Abstract_observation_target
Association_ Source -> Destination	Consensus_genomic_genotype	derived_from Assayed_genomic_genotype
Association_With assay details Source -> Destination	Assayed_genomic_genotype	Variation_assay
Generalization_ Source -> Destination	Assayed_genomic_genotype	Genomic_observation
Generalization_ Source -> Destination	Consensus_genomic_genotype	Assayed_genomic_genotype

Connector	Source	Target
Association_ Source -> Destination	Assayed_genomic_genotype	Abstract_value
Association_Without assay details Source -> Destination	Assayed_genomic_genotype	Genomic_variation

Attributes

Attribute	Notes
assay_failure **boolean**	Failure of assay. Value is true if assay has failed
quality_score **float**	Quality score of measurement. Depends on the instrument

7.1.2.7 Consensus_genomic_genotype

Type: Class__Assayed_genomic_genotype
Package: SNP2

Document

A list of one or more Genomic_alleles carried by a single individual for a single Genomic_variation. Unlike the Assayed_genomic_genotype, the Consensus_genomic_genotype is intended to be used solely for assertions as to the true value of the alleles, after experimental error and other factors are taken into account. The number of Genomic_alleles is equal to the ploidy (copy number) of the individual's genomic DNA at the Reference_genomic_location of the polymorphism. Synonym: Inferred_genomic_genotype, Genomic_genotype.

Connections

Connector	Source	Target
Association_ Destination -> Source	Haplotype_derivation_methods	Consensus_genomic_genotype
Association_ Source -> Destination	Consensus_genomic_genotype	derived_from Assayed_genomic_genotype
Association_ Unspecified	Genomic_genotype_population_frequency	Consensus_genomic_genotype
Generalization_ Source -> Destination	Consensus_genomic_genotype	Assayed_genomic_genotype

7.1.2.8 Frequency

Type: Class__Genomic_observation
Package: SNP2

Document

Abstract class for frequencies, expressed in percentages. Alleles (Genomic_alleles), genotypes (Consensus_genomic_genotype) and haplotypes (Genomic_haplotype) can have measured frequencies in population samples (Panels). In addition, heterozygosity (Heterozygosity) is a measure of observed variability of a polymorphic site (Genomic_variation)in a sub-population (Panel).

Connections

Connector	Source	Target
Generalization_	Heterozygosity	Frequency
Association_ Source -> Destination	Frequency	Panel
Association_ Source -> Destination	Frequency	Variation_assay
Generalization_ Source -> Destination	PAGE::Frequency_set	Frequency
Generalization_ Source -> Destination	Genomic_genotype_population_frequency	Frequency
Generalization_ Source -> Destination	Frequency	Genomic_observation
Generalization_ Source -> Destination	Genomic_allele_population_frequency	Frequency
Generalization_ Source -> Destination	Genomic_haplotype_population_frequency	Frequency
Association_ Source -> Destination	Frequency	Genomic_variation
Aggregation_ Source -> Destination	Frequency	PAGE::Frequency_set

Attributes

Attribute	Notes
count **int**	Total number
value **double**	Value of frequency (%)

7.1.2.9 Functional_change

Type: Class__Gene_product_change
Package: SNP2

Document

Change in the function of the final gene product.

Connections

Connector	Source	Target
Association_ Source -> Destination	Functional_change	Translation_change
Association_ Source -> Destination	Functional_change	affects_function Structural_change
Generalization_ Source -> Destination	Functional_change	SNP::Gene_product_change

7.1.2.10 Gene_based_haplotype

Type: Class__Set_of_haplotypes
Package: SNP2

Document

Collection of variable nucleotides (Genomic_alleles in Genomic_variations) that define a gene. In older usage same as locus.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Gene_based_haplotype	Set_of_haplotypes

7.1.2.11 Genic_variation

Type: Class__Residue_change
Package: SNP2

Document

Genomic variation with location in genic coordinates. Synonym: mutation (when change from a common allele affects phenotype)

Connections

Connector	Source	Target
Generalization_	Genic_variation	SNP::Residue_change
Association_ Source -> Destination	Transcription_change	affects_transcript Genic_variation
Association_ Source -> Destination	is_genic Genic_variation	Genomic_gene_structure
Association_ Source -> Destination	Genic_variation	is_treated_as Genomic_variation
Association_ Source -> Destination	Genomic_gene_structure	Genic_variation

7.1.2.12 Genomic_allele

Type: Class__Latent_genotype_specification, Sequence
Package: SNP2

Document

One of several alternative DNA sequences of a Reference_genomic_location as it appears in the population of organisms. Synonym: variant, allele.

Connections

Connector	Source	Target
Generalization_	Genomic_allele	SNP::Sequence
Association_ Unspecified	Genomic_allele	Genomic_haplotype

Connector	Source	Target
Generalization_	SNP::Genomic_reference_allele	Genomic_allele
Association_ Source -> Destination	one_of SNP::Genomic_reference_allele	Genomic_allele
Association_ Source -> Destination	Genomic_allele_population_frequency	Genomic_allele
Aggregation_ Source -> Destination	Genomic_allele	Genomic_variation
Generalization_ Source -> Destination	Genomic_allele	Latent_genotype_specification

Attributes

Attribute	Notes
bin **string**	Size class for microsatellite alleles when exact size can not be determined. Semantic name.
repeat_count **int**	If the Genomic_variation type is 'microsatellite,' get number of repeat units as value, e.g., 7.
repeat_is_exact **boolean**	The Genomic_variation type is 'microsatellite' and gets value true if the allele region consists of repeat units only.

7.1.2.13 Genomic_allele_population_frequency

Type: Class__Frequency

Package: SNP2

Document

The frequency with which a particular Genomic_allele is seen in a particular Panel. This frequency can be measured from pooled samples. Synonyms: Genomic_allele_panel_frequency, allele_frequency.

Connections

Connector	Source	Target
Association_ Source -> Destination	Genomic_allele_population_frequency	Genomic_allele
Generalization_ Source -> Destination	Genomic_allele_population_frequency	Frequency

7.1.2.14 Genomic_gene_structure

Type: Class__Locatable
Package: SNP2

Document

A structure of a gene expressed as location of the CDS and exons. Defines genic coordinate system from start of the CDS downstream.

Connections

Connector	Source	Target
Generalization_	Genomic_gene_structure	SNP::Locatable
Association_ Unspecified	SNP::CDS	Genomic_gene_structure
Association_ Unspecified	SNP::Exon	Genomic_gene_structure
Association_ Source -> Destination	is_genic Genic_variation	Genomic_gene_structure
Association_ Source -> Destination	Genomic_gene_structure	Genic_variation

Attributes

Attribute	Notes
gene_symbol **string**	gene symbol for the gene e.g. approved by the HUGO nomenclature committee.

7.1.2.15 Genomic_genotype_population_frequency

Type: Class__Frequency
Package: SNP2

Document

Frequency of a Consensus_genomic_genotype in a Panel.

OSAGE-OM

Has many to one relationship to Latent_genotype (Consensus_genomic_genotype in SNP-PML).

Connections

Connector	Source	Target
Association_ Unspecified	Genomic_genotype_population_frequency	Consensus_genomic_genotype
Generalization_ Source -> Destination	Genomic_genotype_population_frequency	Frequency
Association_measured genotype Source -> Destination	Genomic_genotype_population_frequency	Latent_genotype

7.1.2.16 Genomic_haplotype

Type: Class__Genomic_observation

Package: SNP2

Document

A set of Genomic_alleles across an equal number of Genomic_variations in a single chromosome and in a single individual. The Genomic_haplotype is derived from a set of Consensus_genomic_genotype. For each Genomic_variation, the haplotype contains one and only one Genomic_allele. Furthermore, the Genomic_alleles are required to be in phase on the individual, meaning that they are located on the same contiguous strand of DNA. Synonym: Haplotype.

Connections

Connector	Source	Target
Association_ Unspecified	Genomic_allele	Genomic_haplotype
Association_ Destination -> Source	Genomic_haplotype	Haplotype_derivation_methods
Association_ Unspecified	Genomic_haplotype	Set_of_haplotypes
Association_ Source -> Destination	Genomic_haplotype	Molecular_sample
Association_ Source -> Destination	Haplotype_map	Genomic_haplotype
Association_ Source -> Destination	Genomic_haplotype	Haplotype_map

Connector	Source	Target
Generalization_ Source -> Destination	Genomic_haplotype	Genomic_observation
Association_ Source -> Destination	Genomic_haplotype_population_frequency	Genomic_haplotype

7.1.2.17 Genomic_haplotype_population_frequency

Type: Class__Frequency

Package: SNP2

Document

Frequency of a Genomic_haplotype in a Panel.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Genomic_haplotype_population_frequency	Frequency
Association_ Source -> Destination	Genomic_haplotype_population_frequency	Genomic_haplotype

7.1.2.18 Genomic_observation

Type: Class__Identifiable

Package: SNP2

Document

Genomic observation

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Assayed_genomic_genotype	Genomic_observation
Generalization_ Source -> Destination	Set_of_haplotypes	Genomic_observation
Generalization_ Source -> Destination	Genomic_haplotype	Genomic_observation
Generalization_		

Connector	Source	Target
Source -> Destination	Frequency	Genomic_observation
Generalization_ Source -> Destination	Genomic_observation	SNP::Identifiable
Association_ Source -> Destination	PAGE::Experiment_result	Genomic_observation
Association_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	Genomic_observation

7.1.2.19 Genomic_variation

Type: Class__Locatable

Package: SNP2

Document

A variable site in a Reference_genomic_landmark sequence. Synonyms: polymorphic site, marker, (Genomic_polymorphism in SNP-PML). The relationship to Latent_genotype_specification gives information on possible sequence variations attached to the locus (as defined by flanking sequences).

Connections

Connector	Source	Target
Association_ Destination -> Source	SNP::Ordered_location	Genomic_variation
Generalization_	Neighbour_variation	Genomic_variation
Association_ Source -> Destination	PAGE::Multi_vartiation_assay	component_variation Genomic_variation
Association_ Source -> Destination	Variation_assay	Genomic_variation
Association_ Unspecified	Genomic_variation	Neighbour_variation
Association_ Source -> Destination	Genic_variation	is_treated_as Genomic_variation
Generalization_ Source -> Destination	Genomic_variation	SNP::Locatable

Connector	Source	Target
Association_ Destination -> Source	Genomic_variation	SNP::Reference_genomic_landmark
Association_ Source -> Destination	Genomic_variation	Reference_genomic_location
Association_ Unspecified	Genomic_variation	Heterozygosity
Association_ Source -> Destination	PAGE::Experiment_result	Genomic_variation
Association_Genomic variations used in experiment Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	Genomic_variation
Aggregation_ Source -> Destination	Genomic_allele	Genomic_variation
Association_ Source -> Destination	Reference_genomic_location	Genomic_variation
Association_ Source -> Destination	Frequency	Genomic_variation
Association_Without assay details Source -> Destination	Assayed_genomic_genotype	Genomic_variation

Attributes

Attribute	Notes
flank_down **string**	Downstream flanking sequence (at least 25 residues, if possible)
flank_up **string**	Upstream flanking sequence (at least 25 residues, if possible).
is_mutation **boolean**	Proven phenotype change causing mutation.
repeat_unit **string**	If type is 'microsatellite', gives the repeat unit, e.g., "CA"
type **Ontology_term**	The type of the polymorphism (e.g., SNP, microsatellite, indel, translocation,...).
validation_status **string**	Validation status, e.g., "Proven," "Suspected."

7.1.2.20 Haplotype_block

Type: Class__Set_of_haplotypes
Package: SNP2

Document

Large (spanning a few kb to >100 kb) blocks of Genomic_alleles in linkage disequilibrium (LD) and a few haplotypes per block, separated by regions of recombination.

Connections

Connector	Source	Target
Generalization_	Haplotype_block	Set_of_haplotypes

7.1.2.21 Haplotype_derivation_methods

Type: Class__Methodology
Package: SNP2

Document

Association class describing methods used to derive Genomic_haplotypes from Consensus_genomic_genotypes.

Connections

Connector	Source	Target
Association_ Destination -> Source	Genomic_haplotype	Haplotype_derivation_methods
Association_ Destination -> Source	Haplotype_derivation_methods	Consensus_genomic_genotype
Generalization_ Source -> Destination	Haplotype_derivation_methods	SNP::Methodology

7.1.2.22 Haplotype_map

Type: Class__Physical_map
Package: SNP2

Document

Map of haplotypes. Features include: * Block length distribution * measures of block variability * relative proportions of common haplotypes * block coverage of chromosomes and/or genome * LD and other values between haplotypes, markers, alleles.

Connections

Connector	Source	Target
Generalization_	Haplotype_map	SNP::Physical_map
Association_ Unspecified	Haplotype_map	Set_of_haplotypes
Association_ Source -> Destination	Haplotype_map	Genomic_haplotype
Association_ Source -> Destination	Genomic_haplotype	Haplotype_map

7.1.2.23 Heterozygosity

Type: Class__Frequency

Package: SNP2

Document

A numerical measure of observed abundance of different Genomic_alleles in a Genomic_variation.

Connections

Connector	Source	Target
Generalization_	Heterozygosity	Frequency
Association_ Unspecified	Genomic_variation	Heterozygosity

7.1.2.24 Individual

Type: Class__Abstract_population

Package: SNP2

Document

A single member of a species, where a species is an accessioned taxon defined by a public database, and the individual is accessioned in a public or private database. Synonym: “inbred strain” in homozygous lineages.

Connections

Connector	Source	Target
Association_ Unspecified	Individual	Panel
Generalization_ Source -> Destination	Individual	Abstract_population

Attributes

Attribute	Notes
birth_date **string**	Date of birth of the individual. May be better abbreviated to birth year to protect the privacy of the individual
death_date **string**	Date of death of the individual. May be better abbreviated to plain year to protect the privacy of the individual.
father_id **string**	Id of the father to allow building of pedigrees.
mother_id **string**	Id of the mother to allow building of pedigrees.
sex **string**	Recommended values are 'unknown', 'male' and 'female.' Additional values can be used to denote unusual karyotypes.

7.1.2.25 Latent_genotype

Type: Class__Identifiable
Package: SNP2

Document

Potentially existing genotypes that could be observed by Variation_assays (Polymorphism_assays in SNP-PML).

Application of a Variation_assay on one Molecular_sample generates a single Latent_genotype which has one or more Latent_genotype_specifications (this depends on ploidy level in case of Genomic_allele).

Latent_genotypes associated to one instance of a Variation_assay can have only one type of Latent_genotype_specifications, as defined by Defining_feature.

Class is a holder for one or many observable variation objects (Latent_genotype_specification). Latent_genotype is used to attach possible variations to measurements (Assayed_genomic_genotype), variation assays (Variation_assay) and marker loci (Genomic_variation). The construct is used in places where Genomic_allele was used in the SNP-PML.

Connections

Connector	Source	Target
Association_measured genotype Source -> Destination	Assayed_genomic_genotype	Latent_genotype
Generalization_ Source -> Destination	Latent_genotype	SNP::Identifiable
Dependency_ Source -> Destination	Latent_genotype	Abstract_value
Association_detectable genotypes Source -> Destination	Variation_assay	Latent_genotype
Association_measured genotype Source -> Destination	Genomic_genotype_population_frequency	Latent_genotype
AssociationClass_ Unspecified	Latent_genotype	Latent_genotype_specification

7.1.2.26 Latent_genotype_specification

Type: Class__

Package: SNP2

Document

Abstract super class of observable variation objects, like alleles, melting temperatures (Melting_temperature) , band sizes (Band_size). The class makes possible to extend the model to cover different kind of sequence or chromosomal variations.

Connections

Connector	Source	Target
Association_ Source -> Destination	PAGE::Defining_feature	Latent_genotype_specification
Generalization_ Source -> Destination	Genomic_allele	Latent_genotype_specification
Generalization_ Source -> Destination	PAGE::Melting_temperature	Latent_genotype_specification
Generalization_ Source -> Destination	PAGE::Band_size	Latent_genotype_specification

Connector	Source	Target
AssociationClass_ Unspecified	Latent_genotype	Latent_genotype_specification

7.1.2.27 Molecular_sample

Type: Class__Abstract_observation_target

Package: SNP2

Document

A sample from an Individual or from a Panel defining the molecule and tissue/cell used (Anatomic_locations) in the Variation_assay. Synonym: Sample_of_individual. Class extends the Abstract_observation_target, which is abstract class for all entities from which one can make genotype or phenotype measurements or observations.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Molecular_sample	Abstract_observation_target
Association_ Source -> Destination	Molecular_sample	Molecular_sample
Association_ Source -> Destination	Genomic_haplotype	Molecular_sample
Association_ Source -> Destination	Molecular_sample	Abstract_population
Association_ Source -> Destination	Molecular_sample	SNP::Anatomic_location

Attributes

Attribute	Notes
molecule **string**	The molecule (RNA, DNA, protein) used in the assay.

7.1.2.28 Neighbour_variation

Type: Class__Genomic_variation
Package: SNP2

Document

Another Genomic_variation close enough to affect the primer design.

Connections

Connector	Source	Target
Generalization_	Neighbour_variation	Genomic_variation
Association_ Unspecified	Genomic_variation	Neighbour_variation

7.1.2.29 Panel

Type: Class__Abstract_population
Package: SNP2

Document

A set of samples from individuals drawn from the same species and used for genetic studies. A panel must be identifiable with a list of accessioned individuals, if possible. (Note: do not use term "population" here!) Synonym: SampleSet, Sample from population(s), “Plate” in Coriel sense. Recursion. Panels of panels.

Connections

Connector	Source	Target
Association_ Source -> Destination	Panel	Panel
Association_ Unspecified	Individual	Panel
Association_ Source -> Destination	Frequency	Panel
Generalization_ Source -> Destination	Panel	Abstract_population

Attributes

Attribute	Notes
count_unit **string** individual	Values are 'chromosome' and 'individual'. Default is 'individual.'
pooled **boolean**	True if accessioned individuals are not available.
size **long**	The size of the sample. Note that the count_unit field affects how this value is interpreted.
type **string**	Optional identifier of the panel category: e.g. plate, family, population sample

7.1.2.30 Reference_genomic_location

Type: Class__Location
Package: SNP2

Document

A location within a Reference_genomic_landmark. Attributes of the location are the Reference_genomic_assembly and/or the Reference_genomic_landmark, the start and end range and strand of the feature relative to the Reference_genomic_landmark.

Connections

Connector	Source	Target
Generalization_	Reference_genomic_location	SNP::Location
Generalization_	Reference_genomic_location_in_assembly	Reference_genomic_location
Association_ Destination -> Source	SNP::Reference_genomic_assembly	Reference_genomic_location
Association_ Destination -> Source	SNP::Reference_genomic_landmark	Reference_genomic_location
Association_ Source -> Destination	Genomic_variation	Reference_genomic_location
Association_ Source -> Destination	Reference_genomic_location	Genomic_variation

Attributes

Attribute	Notes
end **int**	End of the location in the reference sequence.
start **int**	Start of the location in the reference sequence.
strand **string** unknown	Orientation of the feature in the reference sequence. One of 'forward,' 'reverse,' 'unknown.' Defaults to 'unknown.'

7.1.2.31 Reference_genomic_location_in_assembly

Type: Class__Reference_genomic_location

Package: SNP2

Document

A location in one chromosome of a reference genomic assembly. Instead of the reference sequence being an accessioned sequence, it is a versioned assembly.

Connections

Connector	Source	Target
Association_ Destination -> Source	SNP::Reference_genomic_assembly	Reference_genomic_location_in_assembly
Generalization_	Reference_genomic_location_in_assembly	Reference_genomic_location

Attributes

Attribute	Notes
chromosome_name **string**	Name of the chromosome in the assembly.

7.1.2.32 Set_of_haplotypes

Type: Class__Genomic_observation

Package: SNP2

Document

An abstract representation of a collection of haplotypes.

Connections

Connector	Source	Target
Generalization_	Haplotype_block	Set_of_haplotypes
Association_ Unspecified	Genomic_haplotype	Set_of_haplotypes
Association_ Unspecified	Haplotype_map	Set_of_haplotypes
Generalization_ Source -> Destination	Set_of_haplotypes	Genomic_observation
Generalization_ Source -> Destination	Gene_based_haplotype	Set_of_haplotypes

7.1.2.33 Structural_change

Type: Class__Gene_product_change

Package: SNP2

Document

Change in the 3D structure of the polypeptide chain.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Structural_change	SNP::Gene_product_change
Association_ Source -> Destination	Structural_change	affects_structure Translation_change
Association_ Source -> Destination	Functional_change	affects_function Structural_change

7.1.2.34 Transcription_change

Type: Class__Residue_change
Package: SNP2

Document
Change in the quality or quantity of the mature RNA product.

Connections

Connector	Source	Target
Association_ Source -> Destination	Translation_change	affects_protein Transcription_change
Association_ Source -> Destination	Transcription_change	affects_transcript Genic_variation
Generalization_ Source -> Destination	Transcription_change	SNP::Residue_change

Attributes

Attribute	Notes
changed_codon **string**	The new codon in the transcript, if applicable.
codon_position **int**	The first affected nucleotide in the codon. Values are: 1, 2, or 3.
original_codon **string**	The affected codon in the transcript.

7.1.2.35 Translation_change

Type: Class__Residue_change
Package: SNP2

Document
Change in the quality or quantity of (predicted) polypeptide chain (2D).

Connections

Connector	Source	Target
Generalization_	Translation_change	SNP::Residue_change
Association_ Source -> Destination	Translation_change	affects_protein Transcription_change
Association_ Source -> Destination	Functional_change	Translation_change
Association_ Source -> Destination	Structural_change	affects_structure Translation_change

7.1.2.36 Variation_assay

Type: Class__Methodology
Package: SNP2

Document

An experimental lab protocol and set of reagents for detecting the Genomic_alleles of Genomic_variations carried by an individual or a panel of individuals. Synonym: Assay. Non instrument part of the experiment - same assay can be used in different instrument runs. (Polymorphism_assay in SNP-PML). Variation assay designed for the Genomic_variation has one-to-many association to Latent_genotype_specification, which gives information on measurable variations (by this assay) attached to the locus.

Connections

Connector	Source	Target
Generalization_	Variation_assay	SNP::Methodology
Generalization_ Source -> Destination	PAGE::Multi_vartiation_assay	Variation_assay

Connector	Source	Target
Association_ Source -> Destination	Variation_assay	Genomic_variation
Association_ Source -> Destination	Variation_assay	SNP::Oligo
Aggregation_ Source -> Destination	Variation_assay	PAGE::Multi_vartiation_assay
Association_With assay details Source -> Destination	Assayed_genomic_genotype	Variation_assay
Association_ Source -> Destination	Frequency	Variation_assay
Association_detectable genotypes Source -> Destination	Variation_assay	Latent_genotype

Attributes

Attribute	Notes
description **string**	Free text description of the assay protocol.

7.1.3 FUGE

Type: **Package** «XSDschema»
Package: PAGE-OM

Ontology - *(Logical diagram)*

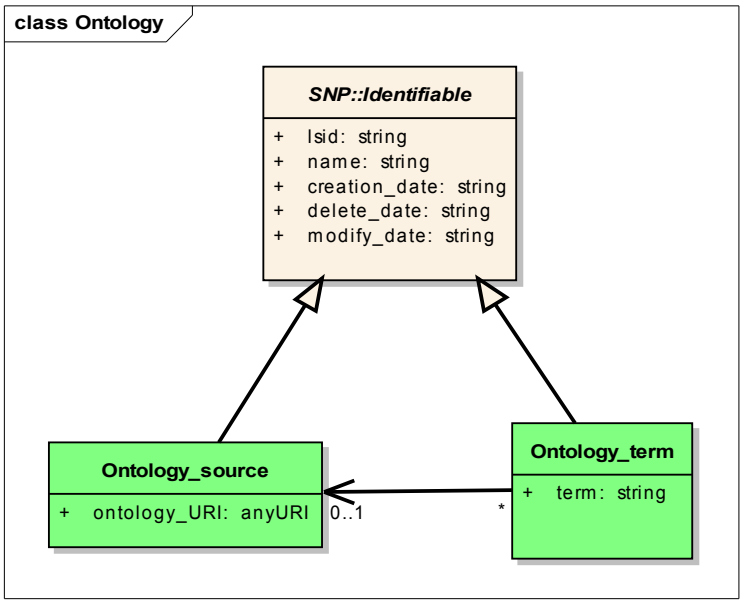

Figure 7.14

7.1.3.1 Ontology_source

Type: Class__Identifiable
Package: FUGE

Document

The source ontology or controlled vocabulary list that ontology terms have been obtained from.

Connections

Connector	Source	Target
Association_ Source -> Destination	FUGE::Ontology_term	FUGE::Ontology_source
Generalization_ Source -> Destination	FUGE::Ontology_source	SNP::Identifiable

Attributes

Attribute	Notes
ontology_URI **anyURI**	A URI is short for Uniform Resource Identifier. A URI is a compact sequence of characters that identifies an abstract or physical resource.

7.1.3.2 Ontology_term

Type: Class__Identifiable
Package: FUGE

Document
Ontology term, defined in more detail in FuGE (http://fuge.sourceforge.net/). This is just a place holder.

Connections

Connector	Source	Target
Association_ Source -> Destination	FUGE::Ontology_term	FUGE::Ontology_source
Generalization_ Source -> Destination	FUGE::Ontology_term	SNP::Identifiable

Attributes

Attribute	Notes
term **string**	Ontology term

7.1.4 BasicTypes

Type: **Package** «XSDschema»
Package: PAGE-OM

BasicTypes - *(Logical diagram)*

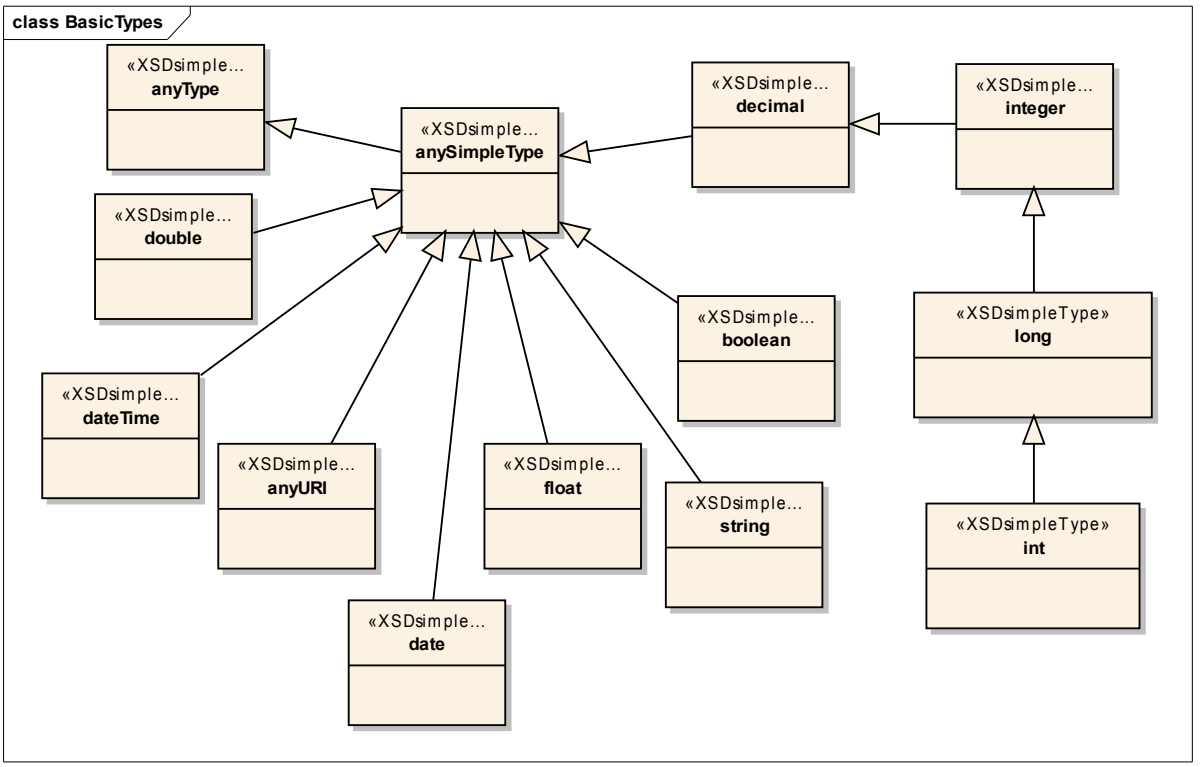

Figure 7.15

7.1.4.1 anySimpleType

Type: Class__anyType
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::float	BasicTypes::anySimpleType
Generalization_	BasicTypes::decimal	BasicTypes::anySimpleType
Generalization_	BasicTypes::anyURI	BasicTypes::anySimpleType
Generalization_	BasicTypes::dateTime	BasicTypes::anySimpleType
Generalization_	BasicTypes::time	BasicTypes::anySimpleType
Generalization_	BasicTypes::date	BasicTypes::anySimpleType
Generalization_	BasicTypes::double	BasicTypes::anySimpleType
Generalization_	BasicTypes::anySimpleType	BasicTypes::anyType
Generalization_	BasicTypes::boolean	BasicTypes::anySimpleType
Generalization_	BasicTypes::string	BasicTypes::anySimpleType

7.1.4.2 anyType

Type: Class__
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::anySimpleType	BasicTypes::anyType

7.1.4.3 anyURI

Type: Class__anySimpleType
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::anyURI	BasicTypes::anySimpleType

7.1.4.4 boolean

Type: Class__anySimpleType
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::boolean	BasicTypes::anySimpleType

7.1.4.5 date

Type: Class__anySimpleType
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::date	BasicTypes::anySimpleType

7.1.4.6 dateTime

Type: Class__anySimpleType

Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::dateTime	BasicTypes::anySimpleType

7.1.4.7 decimal

Type: Class__anySimpleType

Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::integer	BasicTypes::decimal
Generalization_	BasicTypes::decimal	BasicTypes::anySimpleType

7.1.4.8 double

Type: Class__anySimpleType

Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::double	BasicTypes::anySimpleType

7.1.4.9 float

Type: Class__anySimpleType

Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::float	BasicTypes::anySimpleType

7.1.4.10 int

Type: Class__long
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::int	BasicTypes::long
Generalization_	BasicTypes::short	BasicTypes::int

7.1.4.11 integer

Type: Class__decimal
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::integer	BasicTypes::decimal
Generalization_	BasicTypes::long	BasicTypes::integer

7.1.4.12 long

Type: Class__integer
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::long	BasicTypes::integer
Generalization_	BasicTypes::int	BasicTypes::long

7.1.4.13 short

Type: Class__int
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::short	BasicTypes::int

7.1.4.14 string

Type: Class__anySimpleType
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::string	BasicTypes::anySimpleType

7.1.4.15 time

Type: Class__anySimpleType
Package: BasicTypes

Document

Connections

Connector	Source	Target
Generalization_	BasicTypes::time	BasicTypes::anySimpleType

7.1.5 SNP

Type: **Package** «XSDschema»
Package: PAGE-OM

Map - *(Logical diagram)*

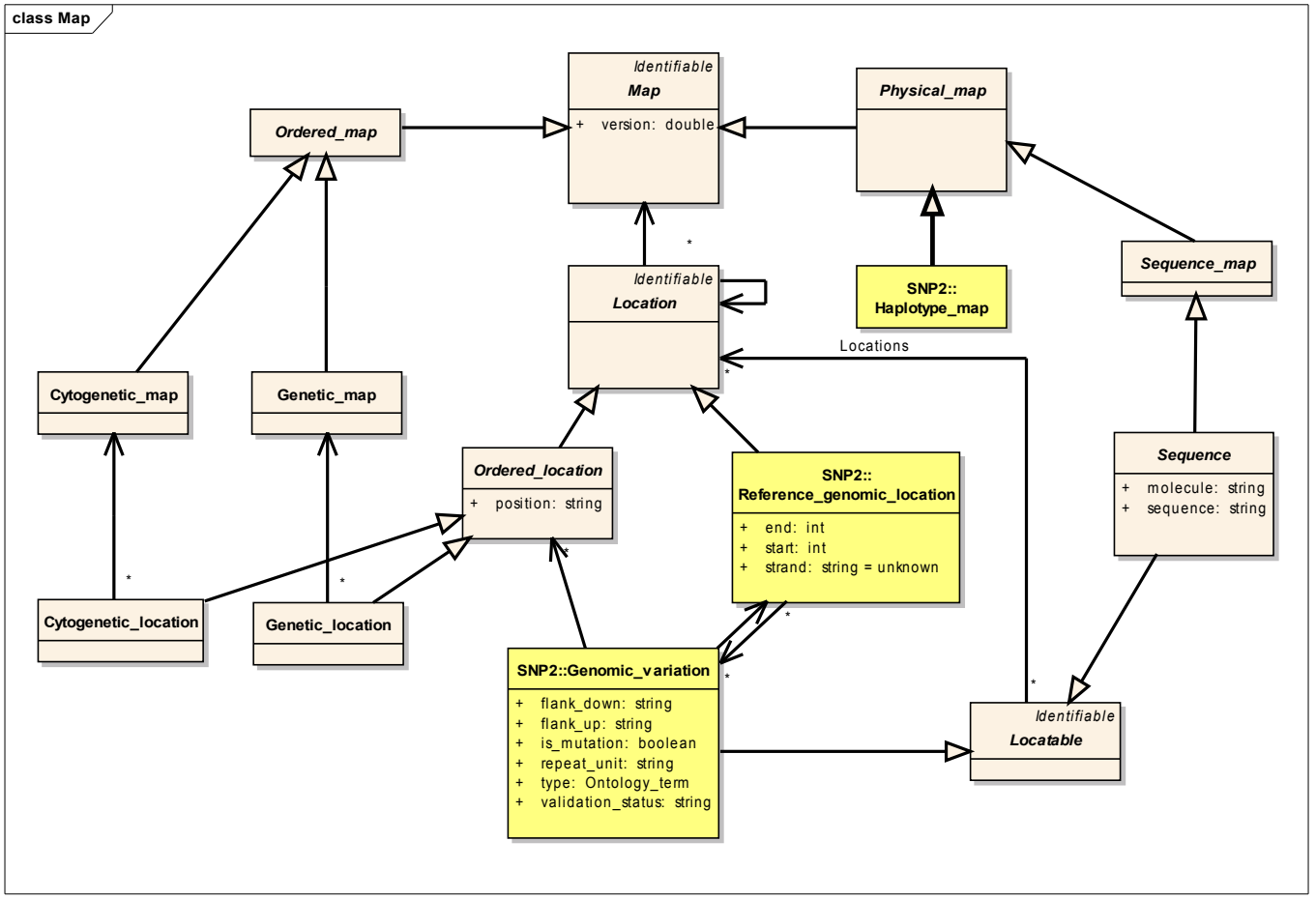
class Map
Identifiable
Map
+ version: double
Ordered_map
Physical_map
Identifiable
Location
SNP2::
Haplotype_map
Sequence_map
Locations
Cytogenetic_map
Genetic_map
Ordered_location
+ position: string
SNP2::
Reference_genomic_location
+ end: int
+ start: int
+ strand: string = unknown
Sequence
+ molecule: string
+ sequence: string
Cytogenetic_location
Genetic_location
SNP2::Genomic_variation
+ flank_down: string
+ flank_up: string
+ is_mutation: boolean
+ repeat_unit: string
+ type: Ontology_term
+ validation_status: string
Identifiable
Locatable

Figure 7.16

Sample - *(Logical diagram)*

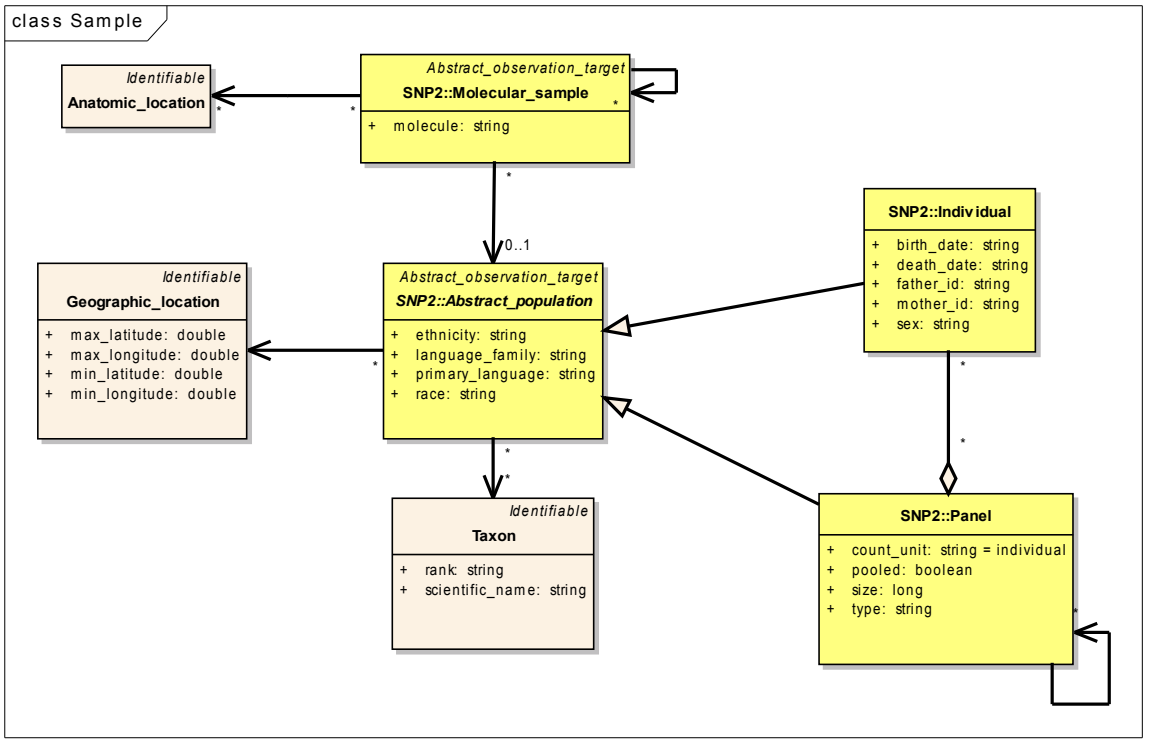

Figure 7.17

Sequence - *(Logical diagram)*

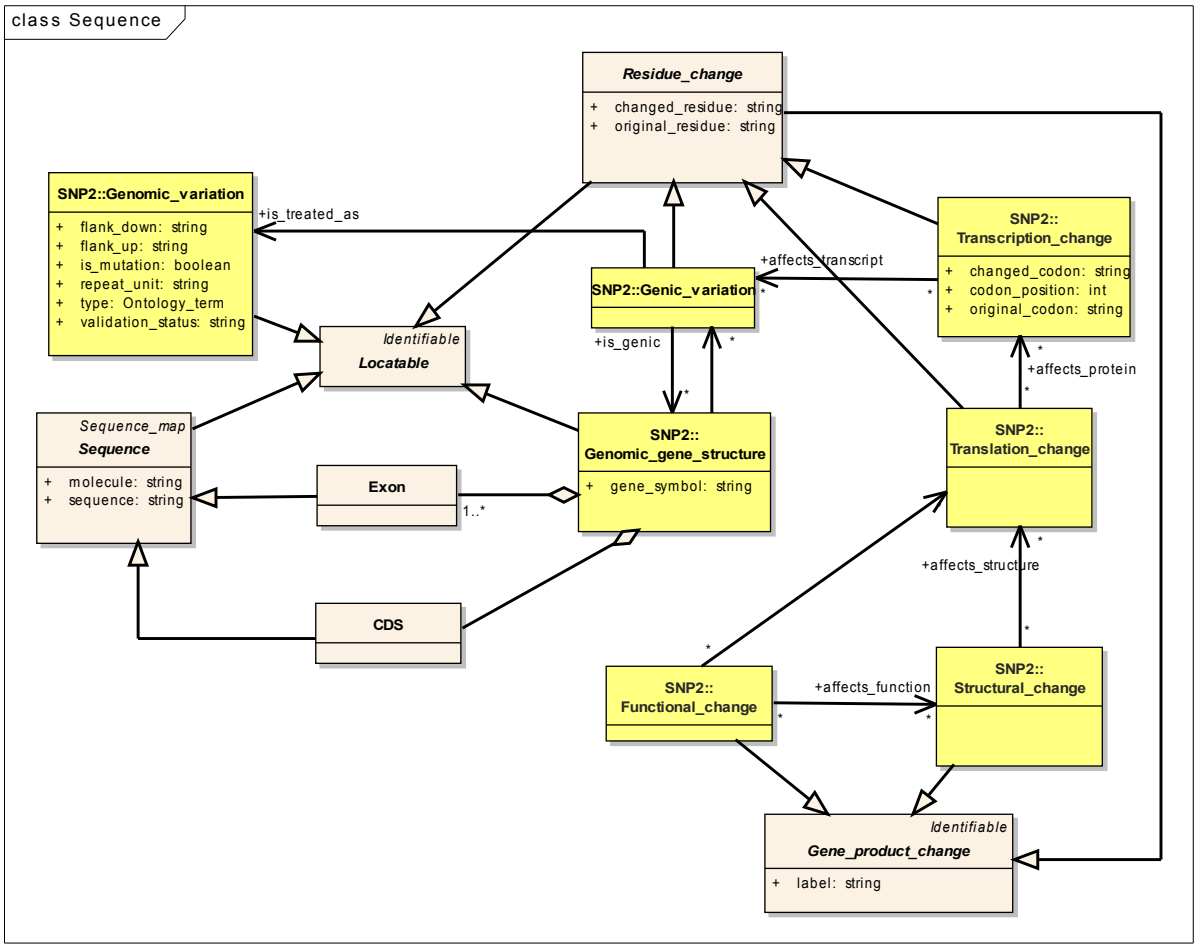

Figure 7.18

Sequence 2 - *(Logical diagram)*

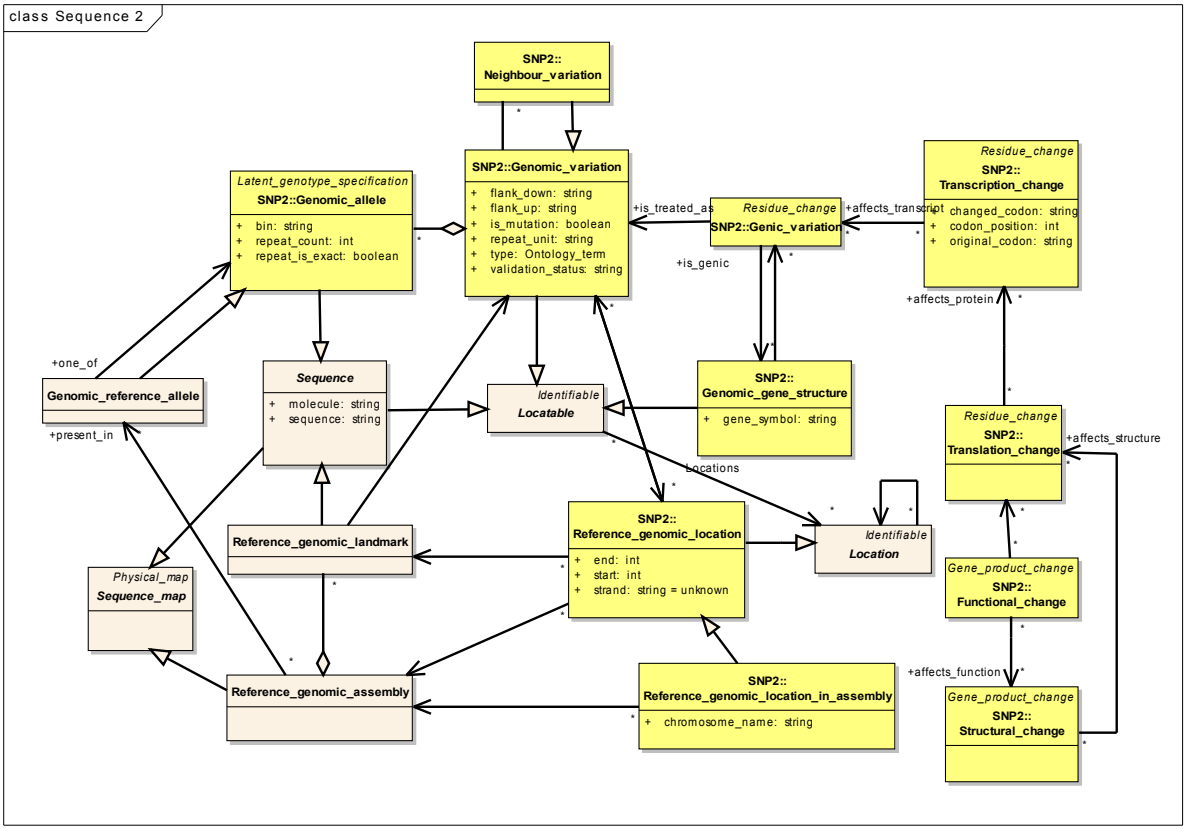

Figure 7.19

7.1.5.1 Anatomic_location

Type: Class__Identifiable
Package: SNP

Document

Description of the anatomical location the molecular sample is derived from. Best expressed as a controlled vocabulary of anatomical terms.

Connections

Connector	Source	Target
Association_ Source -> Destination	SNP2::Molecular_sample	Anatomic_location
Generalization_ Source -> Destination	Anatomic_location	Identifiable

7.1.5.2 Annotation

Type: Class__
Package: SNP

Document

A free text of name and value pair. Note: name should be Ontology_term.

Connections

Connector	Source	Target
Association_ Source -> Destination	Identifiable	Annotation

Attributes

Attribute	Notes
name **string**	Name of the annotation feature
value **Value**	Value of annotation feature

7.1.5.3 CDS

Type: Class__Sequence
Package: SNP

Document
Region in a reference sequence that determines the start, end and strand of the gene coding region.

Connections

Connector	Source	Target
Generalization_	CDS	Sequence
Association_ Unspecified	CDS	SNP2::Genomic_gene_structure

7.1.5.4 Contributor

Type: Class__Source
Package: SNP

Document
Person who has edited data in the data set. Synonym: curator.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Contributor	BREF::Source

Attributes

Attribute	Notes
date **string**	Date of the editing.

7.1.5.5 Cytogenetic_location

Type: Class__Ordered_location
Package: SNP

Document

A location in a cytogenetic map expressed as a band or a range of bands.

Connections

Connector	Source	Target
Generalization_	Cytogenetic_location	Ordered_location
Association_ Destination -> Source	Cytogenetic_map	Cytogenetic_location

7.1.5.6 Cytogenetic_map

Type: Class__Ordered_map

Package: SNP

Document

Map of cytogenetic bands describing their relative order.

Connections

Connector	Source	Target
Generalization_	Cytogenetic_map	Ordered_map
Association_ Destination -> Source	Cytogenetic_map	Cytogenetic_location

7.1.5.7 Db_xref

Type: Class__Identifiable

Package: SNP

Document

Cross reference between two Identifiables in different databases.

Connections

Connector	Source	Target
Association_ Destination -> Source	identifier Db_xref	BREF::Bibliographic_reference
Association_ Source -> Destination	Identifiable	Db_xref

Connector	Source	Target
Generalization_ Source -> Destination	Db_xref	Identifiable

Attributes

Attribute	Notes
db **string**	The remote database name or abbreviation.
db_version **string**	Version of the database
field **string**	Field in the remote database where the UID value is found (e.g., ID or AC in EMBL). Optional
uid **string**	Unique identifier in the remote database.
url **string**	Full URL to the cross-referenced entry

7.1.5.8 Exon

Type: Class__Sequence
Package: SNP

Document
Region in a reference sequence that determines the start, end and strand of the transcript that is not spliced away.

Connections

Connector	Source	Target
Association_ Unspecified	Exon	SNP2::Genomic_gene_structure
Generalization_ Source -> Destination	Exon	Sequence

7.1.5.9 Gene_product_change

Type: Class__Identifiable
Package: SNP

Document
Free form description of change in gene product

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Gene_product_change	Identifiable
Generalization_ Source -> Destination	SNP2::Structural_change	Gene_product_change
Generalization_ Source -> Destination	SNP2::Functional_change	Gene_product_change
Generalization_ Source -> Destination	Residue_change	Gene_product_change

Attributes

Attribute	Notes
label **string**	Human understandable term for the change. E.g. missense, truncation.

7.1.5.10 Genetic_location

Type: Class__Ordered_location

Package: SNP

Document

A relative location of a marker in a genetic map based on recombination.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Genetic_location	Ordered_location
Association_ Destination -> Source	Genetic_map	Genetic_location

7.1.5.11 Genetic_map

Type: Class__Ordered_map

Package: SNP

Document

A map containing the order and non-metric distance between genetic markers for identifiable sequence features. Unit: centiMorgan.

Connections

Connector	Source	Target
Generalization_	Genetic_map	Ordered_map
Association_ Destination -> Source	Genetic_map	Genetic_location

7.1.5.12 Genomic_reference_allele

Type: Class__Genomic_allele

Package: SNP

Document

The DNA sequence of a Genomic_variation as it appears in the public database.

Connections

Connector	Source	Target
Generalization_	Genomic_reference_allele	SNP2::Genomic_allele
Association_ Source -> Destination	Reference_genomic_assembly	present_in Genomic_reference_allele
Association_ Source -> Destination	one_of Genomic_reference_allele	SNP2::Genomic_allele

7.1.5.13 Geographic_location

Type: Class__Identifiable

Package: SNP

Document

Location of an individual or population in a geographic map. Locations are expressed in decimal degrees. Northern latitudes (if north of the equator) and eastern longitudes (if east of GM) have positive values by convention.

Connections

Connector	Source	Target
Association_ Source -> Destination	SNP2::Abstract_population	Geographic_location
Generalization_ Source -> Destination	Geographic_location	Identifiable

Attributes

Attribute	Notes
max_latitude **double**	Upper bound of a geographical location expressed in degrees from the equator.
max_longitude **double**	Upper bound of a geographical location expressed in degrees from the Greenwich meridian.
min_latitude **double**	Lower bound of a geographical location expressed in degrees from the equator
min_longitude **double**	Lower bound of a geographical location expressed in degrees from the Greenwich meridian

7.1.5.14 Identifiable

Type: Class__
Package: SNP

Document
Anything with a unique identifier.

Connections

Connector	Source	Target
Association_ Source -> Destination	Identifiable	BREF::Source
Generalization_ Source -> Destination	Gene_product_change	Identifiable
Generalization_ Source -> Destination	SNP2::Latent_genotype	Identifiable
Generalization_ Source -> Destination	PAGE::Plate	Identifiable
Generalization_ Source -> Destination	SNP2::Abstract_run	Identifiable

Connector	Source	Target
Generalization_	Subject_descriptor	Identifiable
Generalization_	BREF::Bibref_scope	Identifiable
Generalization_	BREF::Journal	Identifiable
Association_ Source -> Destination	Identifiable	Annotation
Association_ Source -> Destination	Identifiable	BREF::Publication
Association_ Source -> Destination	Identifiable	Db_xref
Generalization_ Source -> Destination	PAGE::Study	Identifiable
Generalization_ Source -> Destination	Map	Identifiable
Generalization_ Source -> Destination	Location	Identifiable
Generalization_ Source -> Destination	Locatable	Identifiable
Generalization_ Source -> Destination	Geographic_location	Identifiable
Generalization_ Source -> Destination	SNP2::Abstract_observable_feature	Identifiable
Generalization_ Source -> Destination	PAGE::Observable_feature_category	Identifiable
Generalization_ Source -> Destination	PAGE::Genotype_phenotype_correlation_experiment	Identifiable
Generalization_ Source -> Destination	SNP2::Abstract_observation_target	Identifiable
Generalization_ Source -> Destination	Anatomic_location	Identifiable
Generalization_		

Connector	Source	Target
Source -> Destination	Db_xref	Identifiable
Generalization_ Source -> Destination	PAGE::Observation_method	Identifiable
Generalization_ Source -> Destination	PAGE::Conclusion	Identifiable
Generalization_ Source -> Destination	SNP2::Abstract_value	Identifiable
Generalization_ Source -> Destination	Methodology	Identifiable
Generalization_ Source -> Destination	FUGE::Ontology_source	Identifiable
Generalization_ Source -> Destination	BREF::Provider	Identifiable
Generalization_ Source -> Destination	Taxon	Identifiable
Aggregation supporting_source Source -> Destination	Identifiable	PAGE::Evidence
Generalization_ Source -> Destination	PAGE::Hypothesis	Identifiable
Generalization_ Source -> Destination	SNP2::Genomic_observation	Identifiable
Generalization_ Source -> Destination	FUGE::Ontology_term	Identifiable
Generalization_ Source -> Destination	PAGE::Algorithm	Identifiable

Attributes

Attribute	Notes
lsid **string**	Life Science Identifier.
name **string**	Name can be non unique. Display name
creation_date **string**	Date of creation of the object.
delete_date **string**	Date of deletion of the object.

Attribute	Notes
modify_date **string**	Date of last modification of the object.

7.1.5.15 Locatable

Type: Class__Identifiable
Package: SNP

Document
Abstract class for anything that can be placed on a sequence, i.e., can act as a sequence feature.

Connections

Connector	Source	Target
Generalization_	SNP2::Genomic_gene_structure	Locatable
Generalization_ Source -> Destination	SNP2::Genomic_variation	Locatable
Association_Locations Source -> Destination	Locatable	Location
Generalization_ Source -> Destination	Locatable	Identifiable
Generalization_ Source -> Destination	Residue_change	Locatable
Generalization_ Source -> Destination	Sequence	Locatable

7.1.5.16 Location

Type: Class__Identifiable
Package: SNP

Document
A position in an abstract map.

Connections

Connector	Source	Target
Association_ Destination -> Source	Map	Location
Generalization_	SNP2::Reference_genomic_location	Location
Association_ Source -> Destination	Location	Location
Generalization_ Source -> Destination	Ordered_location	Location
Association_ **Locations** Source -> Destination	Locatable	Location
Generalization_ Source -> Destination	Location	Identifiable

7.1.5.17 Map

Type: Class__**Identifiable**

Package: SNP

Document

An abstract map.

Connections

Connector	Source	Target
Association_ Destination -> Source	Map	Location
Generalization_ Source -> Destination	Physical_map	Map
Generalization_ Source -> Destination	Map	Identifiable
Generalization_ Source -> Destination	Ordered_map	Map

Attributes

Attribute	Notes
version **double**	Version number of the map.

7.1.5.18 Methodology

Type: Class **Identifiable**
Package: SNP

Document
Abstract class for a laboratory method or a computational protocol.

Connections

Connector	Source	Target
Generalization_	SNP2::Variation_assay	Methodology
Generalization_ Source -> Destination	Methodology	Identifiable
Generalization_ Source -> Destination	SNP2::Haplotype_derivation_methods	Methodology

7.1.5.19 Oligo

Type: Class **Sequence**
Package: SNP

Document
A DNA oligonucleotide used for detection and assay of Genomic_variations in Variation_assays. Synonym: primer

Connections

Connector	Source	Target
Generalization_	Oligo	Sequence
Association_ Source -> Destination	SNP2::Variation_assay	Oligo

7.1.5.20 Ordered_location

Type: Class__**Location**
Package: SNP

Document
A location on a secondary map that gives ordering information. That map can be a genetic or a cytogenetic map.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Genetic_location	Ordered_location
Association_ Destination -> Source	Ordered_location	SNP2::Genomic_variation
Generalization_ Source -> Destination	Ordered_location	Location
Generalization_	Cytogenetic_location	Ordered_location

Attributes

Attribute	Notes
position **string**	Location description

7.1.5.21 Ordered_map

Type: Class__**Map**
Package: SNP

Document
A non-metric map category.

Connections

Connector	Source	Target
Generalization_	Cytogenetic_map	Ordered_map
Generalization_	Genetic_map	Ordered_map

Connector	Source	Target
Generalization_ Source -> Destination	Ordered_map	Map

7.1.5.22 Organization

Type: Class **Provider**

Package: SNP

Document

Affiliation of a group of unidentified persons.

Connections

Connector	Source	Target
Generalization_	Organization	BREF::Provider

7.1.5.23 Person

Type: Class **Provider**

Package: SNP

Document

An individual identified by name who is involved with creating or disseminating publications. Possible roles are author, editor, contributor, publisher, or applicant.

Connections

Connector	Source	Target
Generalization_	Person	BREF::Provider
Association_ Destination -> Source	editor Person	BREF::Book
Association_ Source -> Destination	PAGE::Evidence	curator Curator of evidence Person

Attributes

Attribute	Notes
affiliation **string**	Name of the organization the person is affiliated with.
email **string**	Email address
et_al **boolean**	Indicator that there are more persons in the list of persons; used typically for the bibliographic reference that contains only an abbreviated author list. Last listed author should have et_al set to true.
first_name **string**	First name
mid_initials **string**	Initials of any subsequent personal names.
postal_address **string**	Full postal address
surname **string**	Last name.

7.1.5.24 Physical_map

Type: Class__**Map**
Package: SNP

Document

Map of physical entities, e.g., clones, contigs.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Physical_map	Map
Generalization_	SNP2::Haplotype_map	Physical_map
Generalization_ Source -> Destination	Sequence_map	Physical_map

7.1.5.25 Reference_genomic_assembly

Type: Class__**Sequence_map**
Package: SNP

Document

Set of reference sequences (Reference_genomic_landmarks) with an algorithm to define the genome and its constituents.

Connections

Connector	Source	Target
Association_ Unspecified	Reference_genomic_assembly	Taxon
Generalization_	Reference_genomic_assembly	Sequence_map
Association_ Unspecified	Reference_genomic_landmark	Reference_genomic_assembly
Association_ Destination -> Source	Reference_genomic_assembly	SNP2::Reference_genomic_location_in_assembly
Association_ Destination -> Source	Reference_genomic_assembly	SNP2::Reference_genomic_location
Association_ Source -> Destination	Reference_genomic_assembly	present_in Genomic_reference_allele

7.1.5.26 Reference_genomic_landmark

Type: Class **Sequence**

Package: SNP

Document

An accessioned genomic sequence that can be unambiguously assigned to a location in an assembly based on information in public databases.

Connections

Connector	Source	Target
Generalization_	Reference_genomic_landmark	Sequence
Association_ Unspecified	Reference_genomic_landmark	Reference_genomic_assembly

Connector	Source	Target
Association_ Destination -> Source	Reference_genomic_landmark	SNP2::Reference_genomic_location
Association_ Destination -> Source	SNP2::Genomic_variation	Reference_genomic_landmark

7.1.5.27 Residue_change

Type: Class__**Gene_product_change, Locatable**

Package: SNP

Document

Abstract superclass for classes describing residue changes in gene-related sequences.

Connections

Connector	Source	Target
Generalization_	SNP2::Genic_variation	Residue_change
Generalization_	SNP2::Translation_change	Residue_change
Generalization_ Source -> Destination	SNP2::Transcription_change	Residue_change
Generalization_ Source -> Destination	Residue_change	Gene_product_change
Generalization_ Source -> Destination	Residue_change	Locatable

Attributes

Attribute	Notes
changed_residue **string**	The new residue in the reference sequence.
original_residue **string**	The affected residue in the reference sequence.

7.1.5.28 Sequence

Type: Class__**Locatable, Sequence_map**
Package: SNP

Document

Biological sequence. A sequence accessioned by a public database and associated with an accessioned taxon, a version number and a release date.

Connections

Connector	Source	Target
Generalization_	CDS	Sequence
Generalization_	Oligo	Sequence
Generalization_	SNP2::Genomic_allele	Sequence
Generalization_	Reference_genomic_landmark	Sequence
Generalization_ Source -> Destination	Sequence	Locatable
Generalization_ Source -> Destination	Exon	Sequence
Generalization_ Source -> Destination	Sequence	Sequence_map

Attributes

Attribute	Notes
molecule **string**	Indicates the alphabet of the sequence molecule e.g., 'DNA,' 'RNA,' 'protein.'
sequence **string**	The residue sequence string.

7.1.5.29 Sequence_map

Type: Class__**Physical_map**
Package: SNP

Document

A map composed of one or more biological sequences.

Connections

Connector	Source	Target
Generalization_	Reference_genomic_assembly	Sequence_map
Generalization_ Source -> Destination	Sequence	Sequence_map
Generalization_ Source -> Destination	Sequence_map	Physical_map

7.1.5.30 Subject_descriptor

Type: Class **Identifiable**

Package: SNP

Document

Subject term and vocabulary name container.

Connections

Connector	Source	Target
Generalization_	Subject_descriptor	Identifiable
Association_ Destination -> Source	subject_heading Subject_descriptor	BREF::Bibref_subject
Association_ Destination -> Source	code Subject_descriptor	BREF::Bibref_subject
Association_ Destination -> Source	keyword Subject_descriptor	BREF::Bibref_subject

Attributes

Attribute	Notes
term **string**	The keyword, code or subject heading value string
vocabulary_name **string**	Name of the controlled vocabulary. E.g. Sears List of Subject Headings (SEARS), Library of Congress Subject Headings (LCSH), or Medical Subject Headings (MeSH).

7.1.5.31 Taxon

Type: Class **Identifiable**

Package: SNP

Document

Taxonomic unit. The UID is typically a species TaxID from the Taxonomy Database (http://www.ncbi.nih.gov/Taxonomy/)

Connections

Connector	Source	Target
Association_ Unspecified	Reference_genomic_assembly	Taxon
Association_ Source -> Destination	SNP2::Abstract_population	Taxon
Generalization_ Source -> Destination	Taxon	Identifiable

Attributes

Attribute	Notes
rank **string**	Rank of the taxon in the hierarchy. Typically 'species.'
scientific_name **string**	Scientific name, e.g., Homo sapiens.

7.1.6 BREF

Type: **Package «XSDschema»**

Package: PAGE-OM

Bib - *(Logical diagram)*

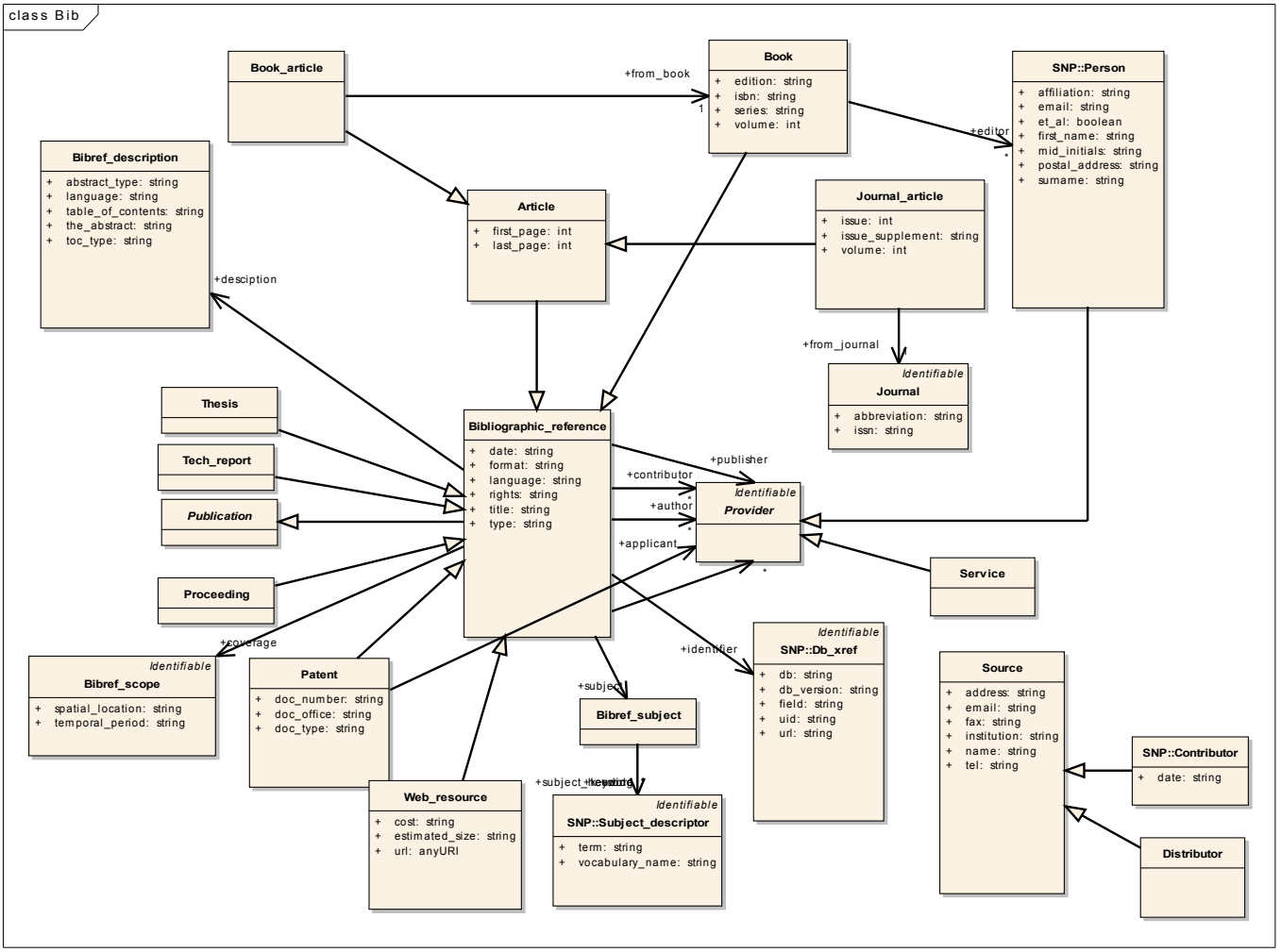

Figure 7.20

7.1.6.1 Article

Type: Class **Bibliographic_reference**

Package: BREF

Document

Non fictional prose forming an independent part of a publication.

Connections

Connector	Source	Target
Generalization_	Book_article	Article
Generalization_	Article	Bibliographic_reference
Generalization_	Journal_article	Article

Attributes

Attribute	Notes
first_page **int**	First page in a article
last_page **int**	Last page in a article

7.1.6.2 Bibliographic_reference

Type: Class **Publication**
Package: BREF

Document

The BibliographicReference class is the core of the data model. It is a super-class for all specialized citation types, but it can also be instantiated and represent an additional type not specifically defined in this specification.

Connections

Connector	Source	Target
Generalization_	Book	Bibliographic_reference
Association_ Destination -> Source	coverage Bibref_scope	Bibliographic_reference
Association_ Destination -> Source	status Entry_status	Bibliographic_reference
Association_ Destination -> Source	desciption Bibref_description	Bibliographic_reference

Connector	Source	Target
Generalization_	Article	Bibliographic_reference
Association_ Destination -> Source	subject Bibref_subject	Bibliographic_reference
Generalization_	Patent	Bibliographic_reference
Generalization_	Thesis	Bibliographic_reference
Generalization_	Web_resource	Bibliographic_reference
Generalization_	Proceeding	Bibliographic_reference
Generalization_	Tech_report	Bibliographic_reference
Association_ Destination -> Source	contributor Provider	Bibliographic_reference
Association_ Destination -> Source	publisher Provider	Bibliographic_reference
Generalization_	Bibliographic_reference	Publication
Association_ Destination -> Source	Provider	Bibliographic_reference
Association_ Destination -> Source	author Provider	Bibliographic_reference
Association_ Destination -> Source	identifier SNP::Db_xref	Bibliographic_reference

Attributes

Attribute	Notes
date **string**	Defines a date associated with an event in the life cycle of the cited resource, when this resource became available. Usually, it is a date of publishing, however, for not yet published resources, it can be a date of creation.
format **string**	Describes the physical or digital manifestation of the cited resource. It can have very different

Attribute	Notes
	content depending on the citation type.
language **string**	Defines a language of the intellectual contents of the cited resource. The recommendation is to use values as defined by RFC1766 which includes a two-letter Language Code (taken from the ISO639 standard, followed optionally by a two-letter Country Code (taken from the ISO3166 standard). For example, "en" for English, "fr" for French, or "en-uk" for English used in the United Kingdom. Another possibility is to use MARC List of Languages.
rights **string**	Specifies information about rights over the cited resource. Typically, it contains a rights management statement for the resource, or it refers to a service providing such information. Rights information often encompasses Intellectual Property Rights [IPR], Copyrights, and various Property Rights.
title **string**	A title given to the cited resource (a name by which the resource is formally known)
type **string**	It defines the nature or genre of the cited resource. Although a working draft of Dublin Core Types recommends a type classification, the proposed types are mostly out of scope of this specification. The majority of cited resources would fall in the same category "text". For the future, however, Dublin Core is considering the addition of subtypes to the high level types, or other ways of making sub-categories

7.1.6.3 Bibref_description

Type: Class__
Package: BREF

Document

A brief account of the content of the cited resource. It can be either the abstract, or the table of contents, or both. It can be written in a language different from the language of the cited resource.

Connections

Connector	Source	Target
Association_ Destination -> Source	desciption Bibref_description	Bibliographic_reference

Attributes

Attribute	Notes
abstract_type **string**	Format of the abstract expressed in MIME (Multipurpose Internet Mail Extensions)
language **string**	Language of the description
table_of_contents **string**	A list of divisions (chapters or articles) and the pages on which they start
the_abstract **string**	Summary of the main points of the publication.

Attribute	Notes
toc_type **string**	Format of the table of contents expressed in MIME (Multipurpose Internet Mail Extensions).

7.1.6.4 Bibref_scope

Type: Class__**Identifiable**
Package: BREF

Document

It defines the extent or scope of the content of the cited resource. It can include spatial location (a place name or geographic co-ordinates), temporal period (a period label, date, or date range), or both. Finally, it can have additional dynamic properties such as jurisdiction).

Connections

Connector	Source	Target
Association_ Destination -> Source	coverage Bibref_scope	Bibliographic_reference
Generalization_	Bibref_scope	SNP::Identifiable

Attributes

Attribute	Notes
spatial_location **string**	A place or area name or geographic coordinates in string format.
temporal_period **string**	A period label, date, or date range.

7.1.6.5 Bibref_subject

Type: Class__
Package: BREF

Document

It defines the topic of the content of the cited resource.

Connections

Connector	Source	Target
Association_	subject	

Connector	Source	Target
Destination -> Source	Bibref_subject	Bibliographic_reference
Association_ Destination -> Source	subject_heading SNP::Subject_descriptor	Bibref_subject
Association_ Destination -> Source	code SNP::Subject_descriptor	Bibref_subject
Association_ Destination -> Source	keyword SNP::Subject_descriptor	Bibref_subject

7.1.6.6 Book

Type: Class **Bibliographic_reference**
Package: BREF

Document
A written work or composition that has been published; usually printed on pages bound together.

Connections

Connector	Source	Target
Generalization_	Book	Bibliographic_reference
Association_ Destination -> Source	editor SNP::Person	Book
Association_ Destination -> Source	from_book Book	Book_article

Attributes

Attribute	Notes
edition **string**	Name of the edition.
isbn **string**	International Standard Book Number
series **string**	Name of a collection of books.
volume **int**	Volume number of the book in a series.

7.1.6.7 Book_article

Type: Class__**Article**
Package: BREF

Document
Article that forms part of a book.

Connections

Connector	Source	Target
Generalization_	Book_article	Article
Association_ Destination -> Source	from_book Book	Book_article

7.1.6.8 Distributor

Type: Class__**Source**
Package: BREF

Document
The original source of the data.

Connections

Connector	Source	Target
Generalization_	Distributor	Source

7.1.6.9 Entry_status

Type: Class__**Publication**
Package: BREF

Document
Defines information related to the citation itself rather than to the cited resource.

Connections

Connector	Source	Target
Association_ Destination -> Source	status Entry_status	 Bibliographic_reference

Connector	Source	Target
Generalization_	Entry_status	Publication

Attributes

Attribute	Notes
last_modified_date **string**	last date the citation entry was changed
subset **string**	Some bibliographic repositories consist of several, or even many, databases. The subset helps to locate the citation.

7.1.6.10 Journal

Type: Class **Identifiable**
Package: BREF

Document

A class describing a journal; a periodical dedicated to a particular subject. The citations referring to the journal articleshave a reference to this class. There are only few explicit attributes defined, the rest are accessible using dynamic properties.

Connections

Connector	Source	Target
Generalization_	Journal	SNP::Identifiable
Association_ Destination -> Source	from_journal Journal	Journal_article

Attributes

Attribute	Notes
abbreviation **string**	Accepted abbreviated form of the journal name.
issn **string**	The ISSN (International Standard Serial Number) is an eight-digit number that identifies periodical publications as such.

7.1.6.11 Journal_article

Type: Class__**Article**
Package: BREF

Document
Article, especially a scientific one that forms part of a journal.

Connections

Connector	Source	Target
Generalization_	Journal_article	Article
Association_ Destination -> Source	from_journal Journal	Journal_article

Attributes

Attribute	Notes
issue **int**	Integer indicating the ordering of the issue within a volume.
issue_supplement **string**	Additional identifier for a separate supplement of the issue.
volume **int**	Volume number of the journal. Typically all issues of a journal published within a year belong to the same volume

7.1.6.12 Patent

Type: Class__**Bibliographic_reference**
Package: BREF

Document
A document granting an inventor sole rights to an invention.

Connections

Connector	Source	Target
Generalization_	Patent	Bibliographic_reference
Association_ Destination -> Source	applicant Provider	Patent

Attributes

Attribute	Notes
doc_number **string**	Patent identifier
doc_office **string**	Name of the patent office (e.g., European Patent Office).
doc_type **string**	Type of a patent: e.g., 'utility,' 'design,' and 'plant.'

7.1.6.13 Proceeding

Type: Class **Bibliographic_reference**
Package: BREF

Document
Printed communication from a scientific organization.

Connections

Connector	Source	Target
Generalization_	Proceeding	Bibliographic_reference

7.1.6.14 Provider

Type: Class **Identifiable**
Package: BREF

Document

Bibliographic resources. The most obvious examples are authors, but it includes also publishers and other contributors. These attributes define the active participants. They may be persons, organizations, or even services. A publisher is responsible for making the resource available. The authors and contributors are in ordered lists. The authors and contributors are responsible for creating the contents of the cited resource. There is no formal definition of how this 60 Single Nucleotide Polymorphisms Specification responsibility is divided between them. However, the authors are usually primary creators while contributors may be illustrators, translators, or other creative providers. Their role may be specified in a separate attribute in dynamic properties.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Service	Provider

Connector	Source	Target
Generalization_	SNP::Person	Provider
Association_ Destination -> Source	contributor Provider	Bibliographic_reference
Association_ Destination -> Source	applicant Provider	Patent
Association_ Destination -> Source	publisher Provider	Bibliographic_reference
Association_ Destination -> Source	Provider	Bibliographic_reference
Association_ Destination -> Source	author Provider	Bibliographic_reference
Generalization_	SNP::Organization	Provider
Generalization_ Source -> Destination	Provider	SNP::Identifiable

7.1.6.15 Publication

Type: Class__

Package: BREF

Document

A Citation. REF BibRef OMG standard. Not identifiable (Defined in model).

Connections

Connector	Source	Target
Generalization_	Entry_status	Publication
Generalization_	Bibliographic_reference	Publication
Association_ Source -> Destination	SNP::Identifiable	Publication

7.1.6.16 Service

Type: Class__**Provider**
Package: BREF

Document
Provider of software service.

Connections

Connector	Source	Target
Generalization_ Source -> Destination	Service	Provider

7.1.6.17 Source

Type: Class__
Package: BREF

Document
Source of the data. Individual or institute.

Connections

Connector	Source	Target
Association_ Source -> Destination	SNP::Identifiable	Source
Generalization_	Distributor	Source
Generalization_ Source -> Destination	SNP::Contributor	Source

Attributes

Attribute	Notes
address **string**	Street address.
email **string**	Contact e-mail.
fax **string**	Fax number.

Attribute	Notes
institution **string**	Name of the institution
name **string**	Name of the Source.
tel **string**	Telephone number.

7.1.6.18 Tech_report

Type: Class **Bibliographic_reference**
Package: BREF

Document
Technical report

Connections

Connector	Source	Target
Generalization_	Tech_report	Bibliographic_reference

7.1.6.19 Thesis

Type: Class **Bibliographic_reference**
Package: BREF

Document
A treatise advancing a new point of view resulting from original research; usually a requirement for an advanced academic degree.

Connections

Connector	Source	Target
Generalization_	Thesis	Bibliographic_reference

7.1.6.20 Web_resource

Type: Class **Bibliographic_reference**
Package: BREF

Document

A referred web address

Connections

Connector	Source	Target
Generalization_	Web_resource	Bibliographic_reference

Attributes

Attribute	Notes
cost **string**	Cost of accessing the information
estimated_size **string**	Rough size of the retrieved reference entry.
url **anyURI**	Uniform Resource Locator

7.2 Ontology

Various attributes in the model are expressed as an ontology term, a string whose value is taken from a controlled vocabulary, or from a more advanced ontology. This specification just *uses* ontology, it does not propose a *model for ontology*.

The Ontology_term class is a slightly modified version of a class defined in the Functional Genomics Experiments (FUGE) project (http://fuge.sourceforge.net/). The modification consists of both classes, Ontology_term and Ontology_source, inheriting from the Identifiable (defined in the SNP specification).

This specification suggests to use (whenever it is possible) the following well-known, third-parties, ontology and controlled vocabularies:

Classes with *Ontology_term*	**Potential Ontology**
Time_accuracy	DAML-Time (http://www.cs.rochester.edu/~ferguson/daml/)
Unit	From jScience Project (http://jscience.org), Unit Ontology as part of the OBO Foundary (http://obofoundry.org/cgi-bin/detail.cgi?id=unit)
Identifiable ontology classes	**These classes uses LSID to define name space for ontology**
Phenotype_disease_category	OBO (Open Biomedical Ontologies) Foundry (http://obofoundry.org/), ICD-9/10 (http://www.who.int/classifications/icd/en/), SNOMED CT (http://www.ihtsdo.org/our-standards/snomed-ct/), NCI Thesaurus (http://www.nci.nih.gov/cancerinfo/terminologyresources)
Observable_feature_category	By definition expected to be an ontology, exact one dependent upon context (see other ontologies).
Observation_method	Name of observation can be defined using ontology term
Concrete value classes	In some cases possible values are defined under specific name space. Also numeric values can be defined differently e.g. in different studies (value-codes for missing and unknown values or allowed values in the Constraint class)
Associations with *Ontology_term*	
Supporting_source	By definition expected to be an ontology, exact one dependent upon context (see other ontologies).

8 Platform Specific Model

The previous chapter defines a platform independent model. The real implementations, however, are expected to depend on a more specific platform. This chapter shows an XML-based platform specific model expressing exchange format for data that conforms to platform independent model described above.

The platform specific model for XML derives its architecture according to the XML as suggested by W3C document “Extensible Markup Language (XML) 1.0 (http://www.w3.org/TR/REC-xml/)”.

The model is defined by XML Schema as suggested by following W3C documents.

XML Schema Part 0: Primer (http://www.w3.org/TR/xmlschema-0/)
XML Schema Part 1: Structure (http://www.w3.org/TR/xmlschema-1/)
XML Schema Part 2: Datatypes (http://www.w3.org/TR/xmlschema-2/)

The normative XML schema was generated from the modelling tool Enterprise Architect (EA), version 7.0.817.1. The target namespaces were propagated by the EA tool into the resulting XML Schema. However, because these target namespaces are not part of the XMI, when using a different tool, the namespaces must be changed in the resulting XML Schema manually as follows:

UML Package name	Target namespace
SNP	http://www.openpml.org/page-om/snp
SNP2	http://www.openpml.org/page-om/snp2
PAGE	http://www.openpml.org/page-om/page
BREF	http://www.openpml.org/page-om/bref
FUGE	http://www.openpml.org/page-om/fuge

The normative XML schema also uses consistently constructs `xs:choice minOccurs="0" maxOccurs="unbounded` instead of constructs `xs:sequence`.

The normative XML Schema is attached in the “Accompanied files” as file *PSM/PAGE-OM.xsd*.

Annex A - Accompanied Files

This annex is a set of the accompanied files. Some of these files are normative and some of them contain examples and convenient images. If there is a discrepancy between the contents of the normative files and this document, then the normative files take precedence.

The accompanied files are divided into two sections, PIM and PSM. The PIM section includes files related directly to the "Platform Independent Model," the PSM section contains the files used in the "Platform Specific Model."

Platform Independent Model

PAGE-OM_uml_2.1_xmi_2.1.xml

A normative representation of the PAGE-OM specification.

PAGE-OM.EAP

This is the main file created by the Enterprise Architect modeling tool. It served as a base for generating the XMI and description files, mentioned above.

***.png**

Diagrams of the PAGE-OM specification. These are the same diagrams as in this document but perhaps with a better resolution.

Platform Specific Model

***.xsd**

Annex B - References

- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29(1):308-311; see also http://www.ncbi.nlm.nih.gov/SNP/index.html
- Thorisson GA, Stein LD. (2003) The SNP Consortium website: past, present and future. Nucleic Acids Res. 31(1):124-127; see also http://snp.cshl.org/
- Fredman D, Siegfried M, Yuan YP, Bork P, Lehvaslaiho H, Brookes AJ. (2002) HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. Nucleic Acids Res. 30(1):387-391; see also http://www.hgvbase.org/
- Mika Hirakawa, Toshihiro Tanaka, Yoichi Hashimoto, Masako Kuroda, Toshihisa Takagi and Yusuke Nakamura (2002) JSNP: a database of common gene variations in the Japanese population. Nucleic Acids Res. 30(1):158-162; see also http://snp.ims.u-tokyo.ac.jp/
- Baker P G, Goble C A, Bechhofer S, Paton N W, Stevens R and Brass A (1999) An Ontology for Bioinformatics Applications. Bioinformatics 15, 510-520
- The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nature Genet. (2000) 25: 25-29.; see also http://www.geneontology.org/
- T.E. Klein, J.T. Chang, M.K. Cho, K.L. Easton, R. Fergerson, M. Hewett, Z. Lin, Y. Liu, S. Liu, D.E. Oliver, D.L. Rubin, F. Shafa, J.M. Stuart and R.B. Altman, "Integrating Genotype and Phenotype Information: An Overview of the PharmGKB Project", The Pharmacogenomics Journal (2001) 1, 167-170.; see also http://www.pharmgkb.org/
- Rajeevan H, Osier MV, Cheung KH, Deng H, Druskin L, Heinzen R, Kidd JR, Stein S, Pakstis AJ, Tosches NP, Yeh CC, Miller PL, Kidd KK. "ALFRED the ALlele FREquency Database update." Nucleic Acids Research..31(1):270-271.(2003); see also http://alfred.med.yale.edu/
- Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007 Oct;39(10):1181-1186.
- Yamada Y, Kato K, Hibino T, Yokoi K, Matsuo H, Segawa T, Watanabe S, Ichihara S, Yoshida H, Satoh K, Nozawa Y. Prediction of genetic risk for metabolic syndrome. Atherosclerosis 2007;191:298-304.

Annex C - Glossary

Term	Definition
Anatomic location	Description of the anatomical location the molecular sample is derived from. Best expressed as a controlled vocabulary of anatomical terms.
Assay Set	Set of assays which are group experimentally. (e.g., microarray)
Assayed genomic genotype	The result of applying a Variation assay to an individual, to reveal one or more of the Genomic alleles carried by that individual. This term applies to the observed data rather than to the inferred state of the individual (see Consensus genomic genotype). Thus the same individual might have several different Assayed genomic genotypes at the same Genomic variation, where the variation might be due to differing assays, experimental error, dominant systems, missing data, and so forth. Synonym: Measurement.
Association Study	Examination of genetic variation across the human genome, designed to identify genetic associations with observable phenotypes
CDS	Region in a reference sequence that determines the start, end and strand of the gene coding region.
Consensus genomic genotype	A list of one or more Genomic alleles carried by a single individual for a single Genomic variation. Unlike the Assayed genomic genotype, the genomic genotype is intended to be used solely for assertions as to the true value of the alleles, after experimental error and other factors are taken into account. The number of Genomic alleles is equal to the ploidy (copy number) of the individual's genomic DNA at the Reference genomic location of the Genomic variation. Synonym: Inferred genomic genotype, Genomic genotype.
Contributor	Person who has edited data in the data set. Synonym: curator.
Cytogenetic location	A location in a cytogenetic map expressed as a band or a range of bands.
Cytogenetic map	Map of cytogenetic bands describing their relative order.
Db xref	Cross-reference between two Identifiable in different databases.
Disease	Special case of a phenotype describing a disease state.
Environment	Circumstances, objects, or conditions by which one is surrounded.
Exon	Region in a reference sequence that determines the start, end and strand of the

	transcript that is not spliced away.
Frequency Set	Set of frequencies for one genomic variation in one panel using one variation assay
Functional change	Change in the function of the final gene product.
Gene based haplotype	Collection of nucleotides (Genomic alleles in Genomic variations) that define a gene. In older usage same as locus.
Gene product change	Abstract superclass for a different genic change descriptors.
Genetic location	A relative location of a marker in a genetic map based on recombination.
Genetic map	A map containing the order and non-metric distance between genetic markers for identifiable sequence features. Unit: centi-morgan.
Genic variation	Genomic variation with location in coordinates. Synonym: mutation (when change from a common allele affects phenotype).
Genomic allele	One of several alternative DNA sequences of a Reference genomic location as it appears in the population of organisms. Synonym: variant, allele.
Genomic allele population frequency	The frequency with which a particular Genomic allele is seen in a particular Panel. This can be measured from pooled samples. Synonyms: Genomic allele panel frequency, allele frequency.
Genomic gene structure	A of a gene expressed as location of the CDS and exons. Defines genic coordinate system from start of the CDS downstream.
Genomic genotype population frequency	Frequency of a Consensus genomic genotype in a Panel.
Genomic haplotype	A set of Genomic alleles across an equal number of Genomic variations in a single chromosome and in a single individual. The Genomic haplotype is derived from a set of genomic genotype. For each Genomic variation, the haplotype contains one and only one Genomic allele. Furthermore, the Genomic alleles are required to be in phase on the individual, meaning that they are located on the same contiguous strand of DNA. Synonym: Haplotype.
Genomic haplotype population frequency	Frequency of a Genomic haplotype in a Panel.
Genomic reference allele	The DNA sequence of a Genomic variation as it appears in the public database.
Genomic Variation	A variable site in a Reference genomic landmark sequence. Synonyms:

	polymorphic site, marker
Genomic variation	A site in a Reference genomic sequence. Synonyms: polymorphic site, polymorphism, SNP.
Genotype Phenotype Correlation Experiment	Study to detect associations between genotype and phenotype. Can be family, population, or case/control based. Synonyms: Association study, Linkage Study.
Haplotype block	Large (a few kb to >100 kb) blocks of Genomic alleles in linkage disequilibrium (LD) and a few haplotypes per block, separated by region of recombination.
Haplotype derivation methods	Association class describing methods used to derive Genomic haplotypes from Consensus genomic genotypes.
Haplotype map	Map of haplotypes. Features include: 1. Block length distribution, 2. measures of block variability, 3. relative proportions of common haplotypes, 4. block coverage of chromosomes and/or genome, 5. LD and other values between haplotypes, markers, alleles.
Heterozygosity	A numerical measure of observed abundance of different Genomic alleles in a Genomic variation.
Individual	A single member of a species, where a species is an accessioned taxon defined by a public database, and the individual is accessioned in a public or private database. Synonym: "inbred strain" in homozygous lineages.
Methodology	Abstract class for a laboratory method or a computational protocol.
Molecular sample	A sample from Individual from a Panel defining the molecule and tissue/cell used (Anatomic locations) in the Variation assay. Synonym: Sample of individual.
Neighbour variation	Another Genomic variation close enough to affect the Variation assay.
Observable Feature	Measurable feature of observable (e.g., size of nose)
Observable Feature Category	Observable category ideally expressed using an appropriate ontology
Observation Method	Method of observation (e.g., ruler). Can be also questionnaire.
Observation Target	Capable of being observed. For example nose size or eye color.
Observation Value	Observation done at specific point in time. Time is expressed according to the

	W3C profile version of an ISO 8601 date-time string. http://www.w3.org/TR/NOTE-datetime
Oligo	A DNA oligo-nucleotide used for detection and assay of Genomic variations in Variation assay. Synonym: primer.
Ordered location	A location on a secondary map that gives ordering information. That map can be a genetic or a cytogenetic map.
Ordered map	A non-metric map category.
Panel	A set of samples from individuals drawn from the same species and used for genetic studies. A panel must be identifiable with a list of accessioned individuals, if possible. (Note: do not use "population" here!) Synonym: Sample Set, Sample from population(s). Plate in Coriel sense.
Phenotype	Observable part of the structure, function or behavior of a living organism.
Phenotype Disease Category	Category associated with disease phenotype. Best expressed using a disease ontology.
Physical map	Map of physical entities, e.g., clones, contigs.
Population	A interbreeding set of individuals, from whom a Panel is drawn.
Reference genomic assembly	Set of reference sequences (Reference genomic landmarks) with an algorithm to define the genome and its constituents.
Reference genomic landmark	An accessioned sequence which can be unambiguously assigned a location in an assembly based on information in public databases.
Reference genomic location	A location within a Reference genomic. Attributes of the location are the Reference genomic assembly and/or the Reference genomic landmark, the start and end range strand of the feature relative to the Reference genomic landmark.
Reference genomic location in assembly	A location in one chromosome of a reference genomic assembly. Instead of the reference sequence being an accessioned sequence, it is a versioned assembly.
Residue change	Abstract superclass for classes describing residue changes in gene-related sequences.
Sequence	Biological sequence. A sequence accessioned by a public database and associated with an accessioned taxon, a version number and a release date.
Sequence map	A map composed of one or more biological sequences.
Set of haplotypes	An abstract representation of a collection of haplotypes.

Structural change	Change in the 3D structure of the polypeptide chain.
Taxon	Taxonomic unit. The UID is typically a TaxID from the Taxonomy Database (http://www.ncbi.nih.gov/Taxonomy/).