

# Single Nucleotide Polymorphisms

OMG Available Specification

Version 1.0

formal/05-11-01



OBJECT MANAGEMENT GROUP

Copyright © 2005, European Bioinformatics Institute (EMBL-EBI)  
Copyright © 2005, Japan Biological Informatics Consortium (JBIC)  
Copyright © 2005, Object Management Group

## 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. No part of this work covered by copyright herein may be reproduced or used in any form or by any means--graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems--without permission of the copyright owner.

## DISCLAIMER OF WARRANTY

WHILE THIS PUBLICATION IS BELIEVED TO BE ACCURATE, IT IS PROVIDED "AS IS" AND MAY CONTAIN ERRORS OR MISPRINTS. THE OBJECT MANAGEMENT GROUP AND THE COMPANIES LISTED ABOVE MAKE NO WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, WITH REGARD TO THIS PUBLICATION, INCLUDING BUT NOT LIMITED TO ANY WARRANTY OF TITLE OR OWNERSHIP, IMPLIED WARRANTY OF MERCHANTABILITY OR WARRANTY OF FITNESS FOR A PARTICULAR PURPOSE OR USE. IN NO EVENT SHALL THE OBJECT MANAGEMENT GROUP OR ANY OF THE COMPANIES LISTED ABOVE BE LIABLE FOR ERRORS CONTAINED HEREIN OR FOR DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL, RELIANCE OR COVER DAMAGES, INCLUDING LOSS OF PROFITS, REVENUE, DATA OR USE, INCURRED BY ANY USER OR ANY THIRD PARTY IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS MATERIAL, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

The entire risk as to the quality and performance of software developed using this specification is borne by you. This disclaimer of warranty constitutes an essential part of the license granted to you to use this specification.

## RESTRICTED RIGHTS LEGEND

Use, duplication or disclosure by the U.S. Government is subject to the restrictions set forth in subparagraph (c) (1) (ii) of The Rights in Technical Data and Computer Software Clause at DFARS 252.227-7013 or in subparagraph (c)(1) and (2) of the Commercial Computer Software - Restricted Rights clauses at 48 C.F.R. 52.227-19 or as specified in 48 C.F.R. 227-7202-2 of the DoD F.A.R. Supplement and its successors, or as specified in 48 C.F.R. 12.212 of the Federal Acquisition Regulations and its successors, as applicable. The specification copyright owners are as indicated above and may be contacted through the Object Management Group, 250 First Avenue, Needham, MA 02494, U.S.A.

## TRADEMARKS

The OMG Object Management Group Logo®, CORBA®, CORBA Academy®, The Information Brokerage®, XMI® and IIOP® are registered trademarks of the Object Management Group. OMG™, Object Management Group™, CORBA logos™, OMG Interface Definition Language (IDL)™, The Architecture of Choice for a Changing World™, CORBA services™, CORBA facilities™, CORBA med™, CORBA net™, Integrate 2002™, Middleware That's Everywhere™, UML™, Unified Modeling Language™, The UML Cube logo™, MOF™, CWM™, The CWM Logo™, Model Driven Architecture™, Model Driven Architecture Logos™, MDA™, OMG Model Driven Architecture™, OMG MDA™ and the XMI Logo™ are trademarks of the Object Management Group. All other products or company names mentioned are used for identification purposes only, and may be trademarks of their respective owners.

## 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

1	Scope .....	1
2	Conformance .....	1
3	Normative References .....	2
4	Terms and Definitions .....	2
5	Symbols .....	2
6	Additional Information .....	3
	6.1 Relationship to Existing OMG Specifications .....	3
	6.2 Acknowledgements .....	3
7	Introduction .....	5
	7.1 Overview .....	5
8	Platform Independent Model .....	7
	8.1 Model subsets .....	7
	8.1.1 Assay .....	7
	8.1.2 Individual .....	7
	8.1.3 Frequency .....	8
	8.1.4 Haplotype .....	8
	8.1.5 Map .....	9
	8.1.6 Sequence .....	9
	8.1.7 Translation .....	10
	8.1.8 Bibliographic References .....	11
	8.1.9 Identifiable .....	12
	8.2 Model classes, attributes and associations (details) .....	14
	8.2.1 Anatomic_location .....	14
	8.2.2 Annotation .....	14
	8.2.3 Article .....	15
	8.2.4 Assayed_genomic_genotype .....	16
	8.2.5 Bibliographic_reference .....	17
	8.2.6 Bibref_description .....	18
	8.2.7 Bibref_scope .....	18
	8.2.8 Bibref_subject .....	19
	8.2.9 Book .....	19
	8.2.10 Book_article .....	20
	8.2.11 CDS .....	22
	8.2.12 Consensus_genomic_genotype .....	22
	8.2.13 Contributor .....	23

8.2.14 Cytogenetic_location .....	24
8.2.15 Cytogenetic_map .....	25
8.2.16 Db_xref.....	25
8.2.17 Disease .....	26
8.2.18 Distributor .....	26
8.2.19 Entry_status .....	27
8.2.20 Exon .....	28
8.2.21 Frequency .....	28
8.2.22 Functional_change .....	29
8.2.23 Gene_based_haplotype .....	30
8.2.24 Gene_product_change.....	30
8.2.25 Genetic_location .....	31
8.2.26 Genetic_map.....	31
8.2.27 Genic_polymorphism .....	32
8.2.28 Genomic_allele .....	33
8.2.29 Genomic_allele_population_frequency .....	34
8.2.30 Genomic_gene_structure .....	35
8.2.31 Genomic_genotype_population_frequency .....	35
8.2.32 Genomic_haplotype .....	36
8.2.33 Genomic_haplotype_population_frequency .....	37
8.2.34 Genomic_polymorphism .....	37
8.2.35 Genomic_reference_allele .....	38
8.2.36 Geographic_location .....	39
8.2.37 Haplotype_block .....	40
8.2.38 Haplotype_derivation_methods .....	41
8.2.39 Haplotype_map .....	41
8.2.40 Heterozygosity .....	42
8.2.41 Identifiable .....	42
8.2.42 Individual .....	43
8.2.43 Journal .....	44
8.2.44 Journal_article.....	45
8.2.45 Locatable.....	46
8.2.46 Location.....	46
8.2.47 Map .....	47
8.2.48 Methodology.....	48
8.2.49 Molecular_sample .....	48
8.2.50 Neighbour_variation .....	49
8.2.51 Oligo .....	50
8.2.52 Ordered_location .....	51
8.2.53 Ordered_map .....	51
8.2.54 Organization.....	52
8.2.55 Panel .....	53
8.2.56 Patent.....	54
8.2.57 Person.....	55
8.2.58 Phenotype .....	56
8.2.59 Physical_map .....	56
8.2.60 Polymorphism_assay .....	57
8.2.61 Population .....	58
8.2.62 Proceeding .....	58
8.2.63 Provider.....	59
8.2.64 Publication .....	60

8.2.65	Reference_genomic_assembly .....	61
8.2.66	Reference_genomic_landmark .....	61
8.2.67	Reference_genomic_location.....	62
8.2.68	Reference_genomic_location_in_assembly.....	63
8.2.69	Residue_change .....	64
8.2.70	Sequence .....	65
8.2.71	Sequence_map .....	65
8.2.72	Service .....	66
8.2.73	Set_of_haplotypes .....	66
8.2.74	Source .....	67
8.2.75	Structural_change .....	68
8.2.76	Subject_descriptor .....	69
8.2.77	Taxon .....	69
8.2.78	Tech_report .....	70
8.2.79	Thesis .....	71
8.2.80	Transcription_change .....	72
8.2.81	Translation_change .....	73
8.2.82	Web_resource .....	74
<b>9</b>	<b>Platform Specific Model - PML .....</b>	<b>77</b>
9.1	Creation .....	77
9.2	Example mappings .....	77
9.2.1	HapMap .....	77
9.2.2	ALFRED .....	78
9.2.3	dbSNP .....	80
9.2.4	Chinese Population Genetic Diversity.....	81
9.2.5	HGVbase .....	81
<b>A</b>	<b>Accompanied Files .....</b>	<b>83</b>





# 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 catalog of all OMG Specifications Catalog is available from the OMG website at:

[http://www.omg.org/technology/documents/spec\\_catalog.htm](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

- CORBA services

- CORBA facilities
- 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  
250 First Avenue  
Suite 100  
Needham, MA 02494  
USA  
Tel: +1-781-444-0404  
Fax: +1-781-444-0320  
Email: [pubs@omg.org](mailto: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

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

## Issues

The reader is encouraged to report any technical or editing issues/problems with this specification to <http://www.omg.org/technology/agreement.htm>.

# 1 Scope

This specification addresses the need for a standardized data exchange format for Single Nucleotide Polymorphisms (SNPs) entities in the Life Sciences domains. As a result of recent advances in SNPs research, a large amount of experimental data has been accumulating. Standardization is required to facilitate exchanging these data and developing interoperable systems for managing and analyzing them.

Single nucleotide polymorphisms (SNPs) are common DNA sequence variations that are the results of a single nucleotide changes in the genome sequence. SNPs have significant associations to variations in gene, amino acid, and protein. SNPs may be responsible for variation in phenotype or in disease. In SNPs database, information such as assay and publication is indispensable.

This specification models SNPs, using UML. The UML model, defining a platform independent model, describes the conceptual data structure. In order to actually perform data exchange, this specification formalizes a platform specific model, expressing the data exchange format in an XML Schema. The data exchange format is called PML, Polymorphisms Markup Language.

The typical use cases where the data exchange format will be used include:

- Data upload or download between a public database and a local database, and
- an exchange (between researchers) of individual data items, consisting only from a very small subset of XML elements and attributes. This is possible because of a “flat” character of the PML.

# 2 Conformance

The normative parts of this specification are:

- Platform independent model expressed in the attached XML file created according to XMI format rules, v1.2, using program Umbrello, version 1.4.
- 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.

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.

## 3 Normative References

1. 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>
2. 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/>
3. 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://hgvbase.cgb.ki.se/>
4. 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/>
5. 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; see also <http://cagpc.cs.man.ac.uk/tambis/manual/>
6. The Gene Ontology Consortium (2001) Creating the gene ontology resource: design and implementation. *Genome Res.* 11:1425-1433; see also <http://www.geneontology.org/>
7. Hewett M, Oliver DE, Rubin DL, Easton KL, Stuart JM, Altman RB, Klein TE. (2002) PharmGKB: the Pharmacogenetics Knowledge Base. *Nucleic Acids Res.* 30(1):163-5; see also <http://pharmgkb.org/>
8. Cheung KH, Osier MV, Kidd JR, Pakstis AJ, Miller PL, Kidd KK. (2000) ALFRED: an allele frequency database for diverse populations and DNA polymorphisms. *Nucleic Acids Res.* 28(1):361-3; see also <http://alfred.med.yale.edu/>
9. Bibliographic query Service, an OMG specification, [http://www.omg.org/technology/documents/formal/bibliographic\\_query.htm](http://www.omg.org/technology/documents/formal/bibliographic_query.htm)
10. Life Sciences Identifier, an OMG specification, [http://www.omg.org/technology/documents/formal/life\\_sciences.htm](http://www.omg.org/technology/documents/formal/life_sciences.htm)

## 4 Terms and Definitions

The used terms are defined in Section 8.2, “Model Classes, Attributes, and Associations (Details),” on page 16.

## 5 Symbols

This specification does not use any specific symbols/abbreviations.

## 6 Additional Information

### 6.1 Relationship to Existing OMG Specifications

- Bibliographic Query Service (BQS) - This specification uses attributes defined in BQS classes in order to describe publications. The relevant specification is available as OMG documents: formal/02-05-03, formal/02-05-04, and dtc/02-02-01.
- Life Sciences Identifiers (LSID) - There are many cross references in the model defined by this specification. They are expressed using LSIDs. The relevant specification is available as OMG documents: formal/04-12-01 and dtc/04-08-03.
- Genomic Map (GM) - The definitions in GM specification are either too limiting or too generic for SNPs purposes. The GM was designed at the time when genetic maps were the only genome wide maps. Current focus is on sequence maps.
- Model Driven Architecture (MDA) - MDA is used as a fundamental concept. Both platform-independent (PIM) and platform-specific model (PSM) are defined here.
- XML Interdata Interchange (XMI) - The platform-independent model for this specification for designed as a UML model that was converted into XMI model exchange format as defined in the XML Interdata Interchange specification.

### 6.2 Acknowledgements

The authors of this document wish to express their appreciation to those listed below (in alphabetic 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.

Russ B. Altman	russ.altman@stanford.edu
Samir K. Brahmachari	skb@igib.res.in
Anthony Brookes	Anthony.Brookes@cgb.ki.se
Kei-Hoi Cheung	kei.cheung@yale.edu
Debasis Dash	ddash@mahesh.igib.res.in
Mike Feolo	Feolo@ncbi.nlm.nih.gov
David Fredman	david.fredman@cgb.ki.se
Takashi Gojobori	tgojobor@genes.nig.ac.jp
Yoichi Hashimoto	hashimoto_yoichi@webgen.co.jp
Tatsuro Ibe	ibe@hydra.mki.co.jp
Mikiyuki Ikawa	ikawa@hydra.mki.co.jp
Toshihisa Ishikawa	tishikaw@bio.titech.ac.jp
Tokio Kano	kanou@jbic.or.jp
Kenneth K. Kidd	kenneth.kidd@yale.edu
Teri Klein	teri.klein@stanford.edu
Masako Kuroda	kuroda@jst.go.jp
Heikki Lehv�slaiho	heikki@ebi.ac.uk
Yi-Xue Li	yxli@sibs.ac.cn
Frank Liu	frankliu@tsinghua.edu.cn

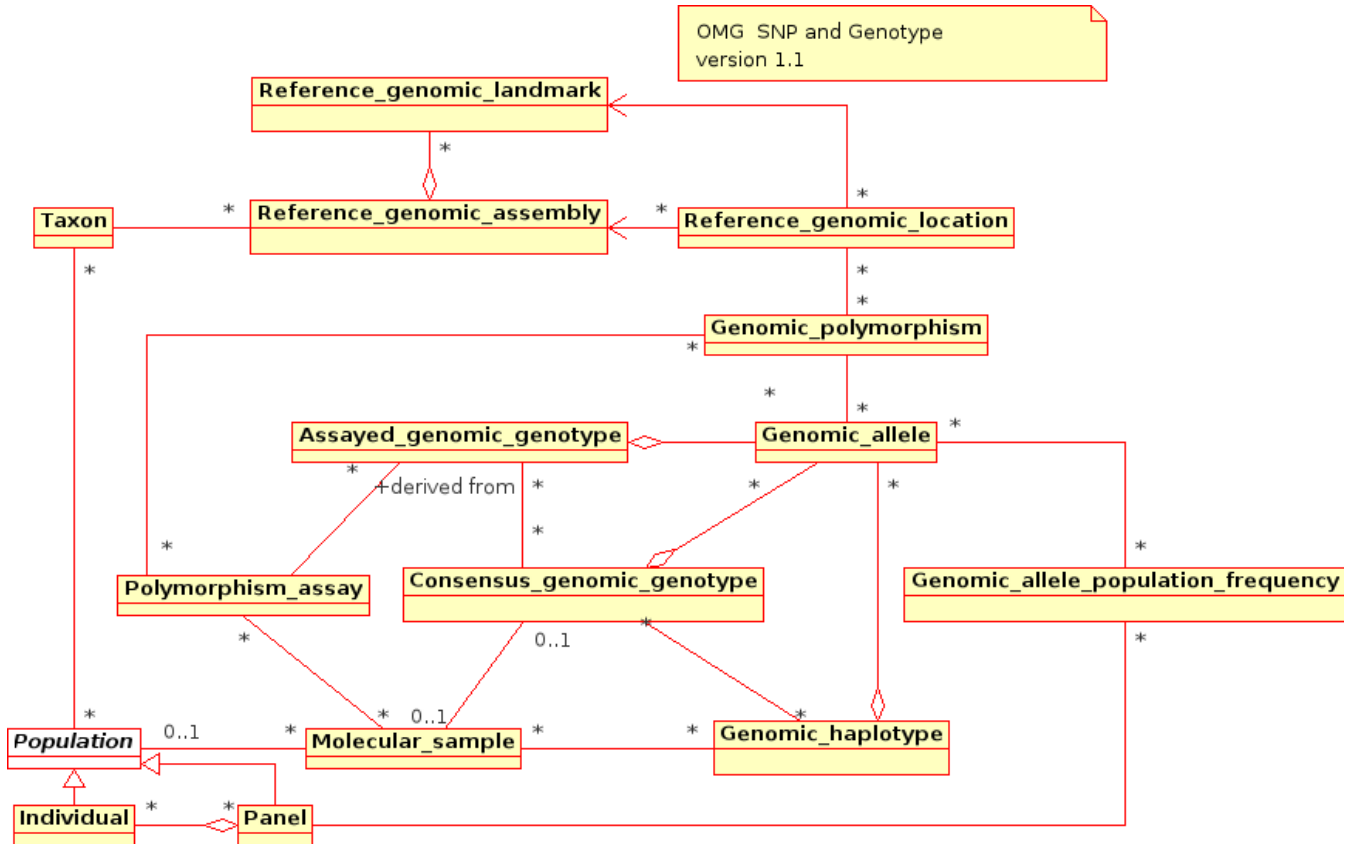
Yukio Matsuura	matsuura@jbic.or.jp
Hiroshi Mizushima	hmizushi@ncc.go.jp
Juha Muilu	juha.muilu@helsinki.fi
Naoki Sakagami	sakagami@hydra.mki.co.jp
Martin Senger	senger@ebi.ac.uk
Stephen Sherry	sherry@ncbi.nlm.nih.gov
Tielu Shi	tlshi@sibs.ac.cn
Yasumasa Shigemoto	yshige@strad.ssg.fujitsu.com
Albert Smith	albert.smith@cshl.edu
Lincoln Stein	lstein@cshl.edu
Hideaki Sugawara	hsugawar@genes.nig.ac.jp
Isao Tomabechi	toma-i@novusgene.co.jp
Peter Tonellato	ptonellato@pointone.ws
Nobuhiro Uchida	n2uchida@jst.go.jp
Mark Woon	morpheus@helix.stanford.edu
Masahito Yamaguchi	yamaguch@strad.ssg.fujitsu.com
Zhu Yue	yzhu@scbio.org

# 7 Introduction

## 7.1 Overview

The graph below gives an overview how a genomic assembly (Reference\_genomic\_assembly) of sequences (Reference\_genomic\_landmark) is used to represent the genomic structure of a species (Taxon), and how known variable regions of the genome (Genomic\_polymorphism) can be located (using Genomic\_location) into one or more sequences and genomic assemblies.

Samples (Molecular\_sample) from individuals (Individual) within the species (Taxon) are then used to assay (Polymorphism\_assay) the variability (Genomic\_alleles) within subpopulations (Panel) to gain knowledge of genotypes (Consensus\_genomic\_genotype) and haplotypes (Genomic\_haplotype) and their relative abundance (see Frequency elsewhere in this document).







# 8 Platform Independent Model

The platform independent model is expressed as a UML model. The content of the model is explained below using the following techniques (note that independently on the techniques used here the normative expression of this model is the XMI-based attached XML file):

- The graphs showing subsets of the model. Each subset has its own short explanation. The same or similar explanation is also included in the XMI representation as description. The graphs may not show all attributes.
- The complete list of entities and their attributes, conveniently repeating also inherited attributes and associations.
- The words “derived from” are used in the UML model as comments, they do not imply any specific derivation algorithm.

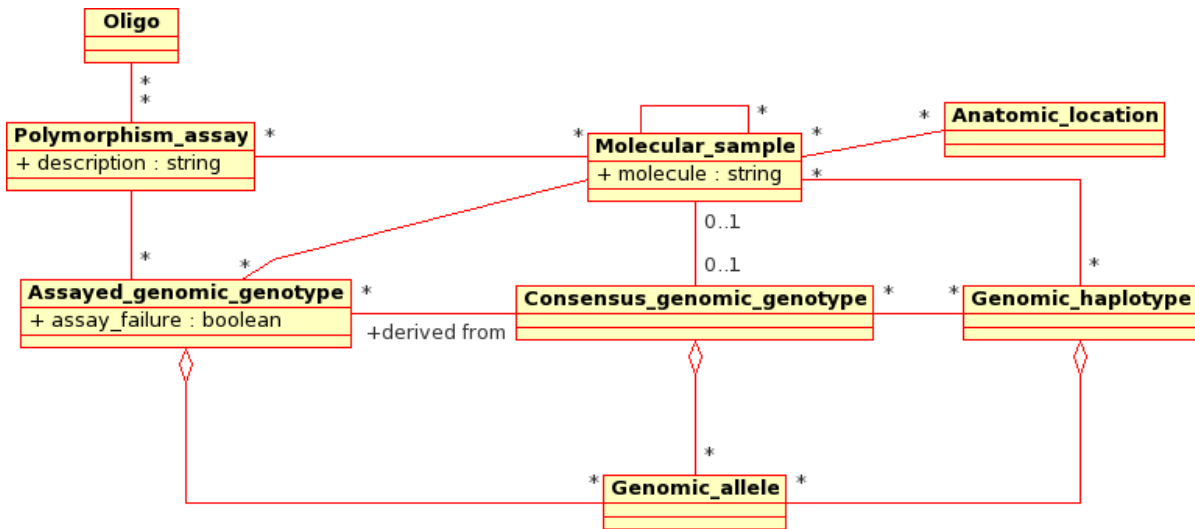
## 8.1 Model Subsets

### 8.1.1 Assay

An assay (Polymorphism\_assay) is applied to a sample (Molecular\_sample) to get a genotype (Assayed\_genomic\_genotype). Location of the sample can be described by several anatomical terms (Anatomic\_location).

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\_polymorphisms).

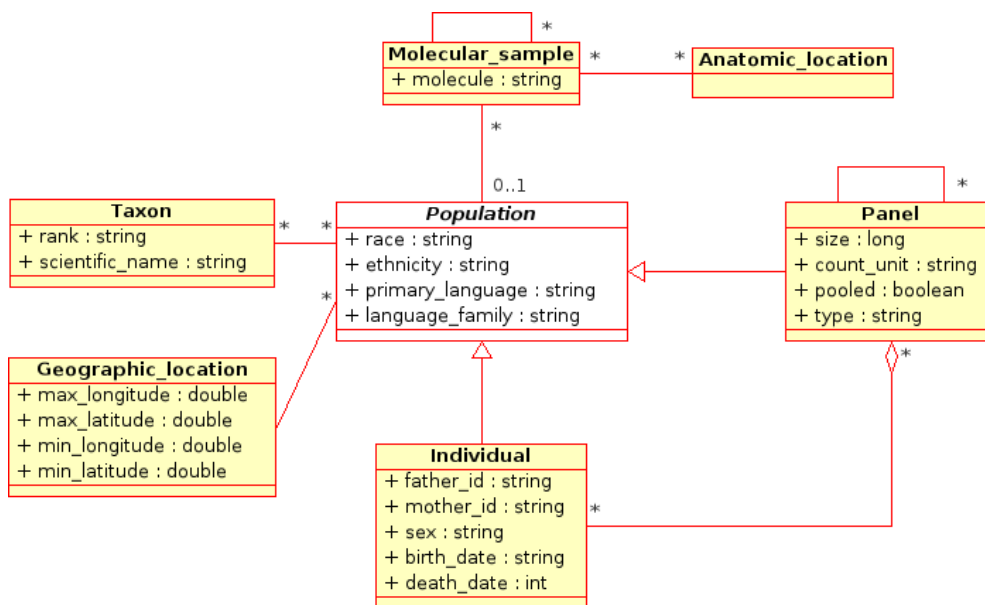
Genomic\_haplotypes can sometimes be derived from genotypes.



## 8.1.2 Individual

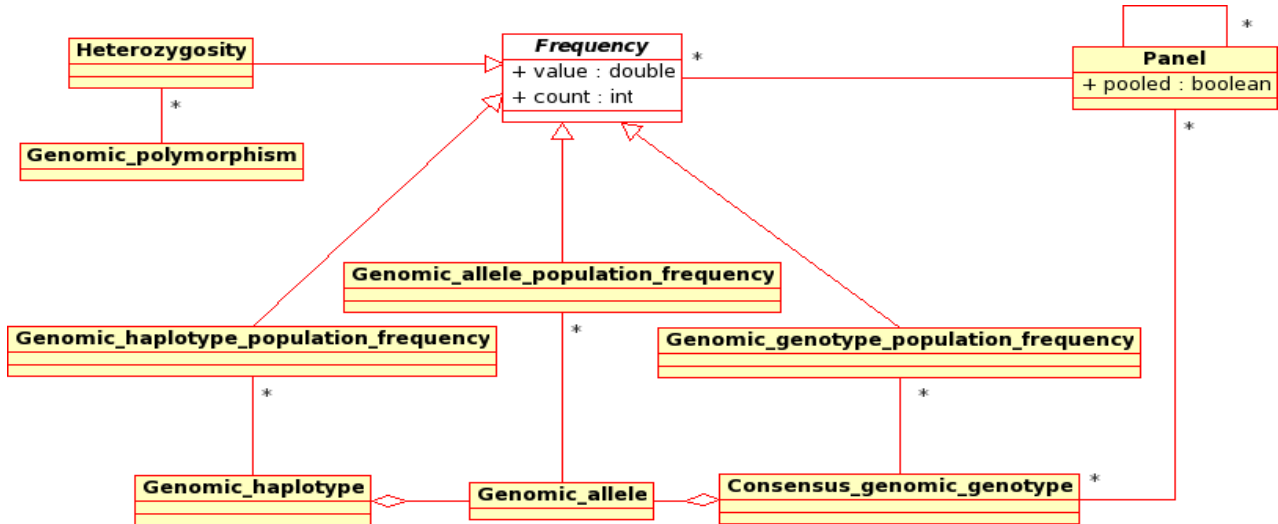
An Individual (Individual) interacts with the rest of the model by giving one or more samples (Molecular\_samples), e.g., at specific time points or from different tissues. A sample (Molecular\_sample) can also belong to one or more population samples (Panel). Mixed samples and more complicated sample organizations are modeled by allowing Molecular\_samples and Panels to be composed of multiples of themselves. In these cases it is common that individuals cannot 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 placed on a geographic map (Geographic\_location).



### 8.1.3 Frequency

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\_polymorphism) in a sub-population (Panel).

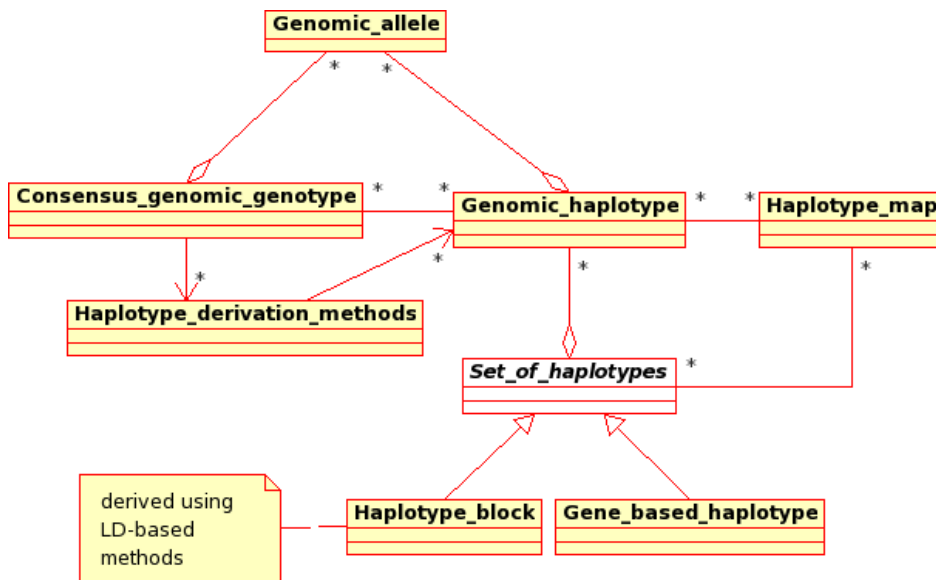


### 8.1.4 Haplotype

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.

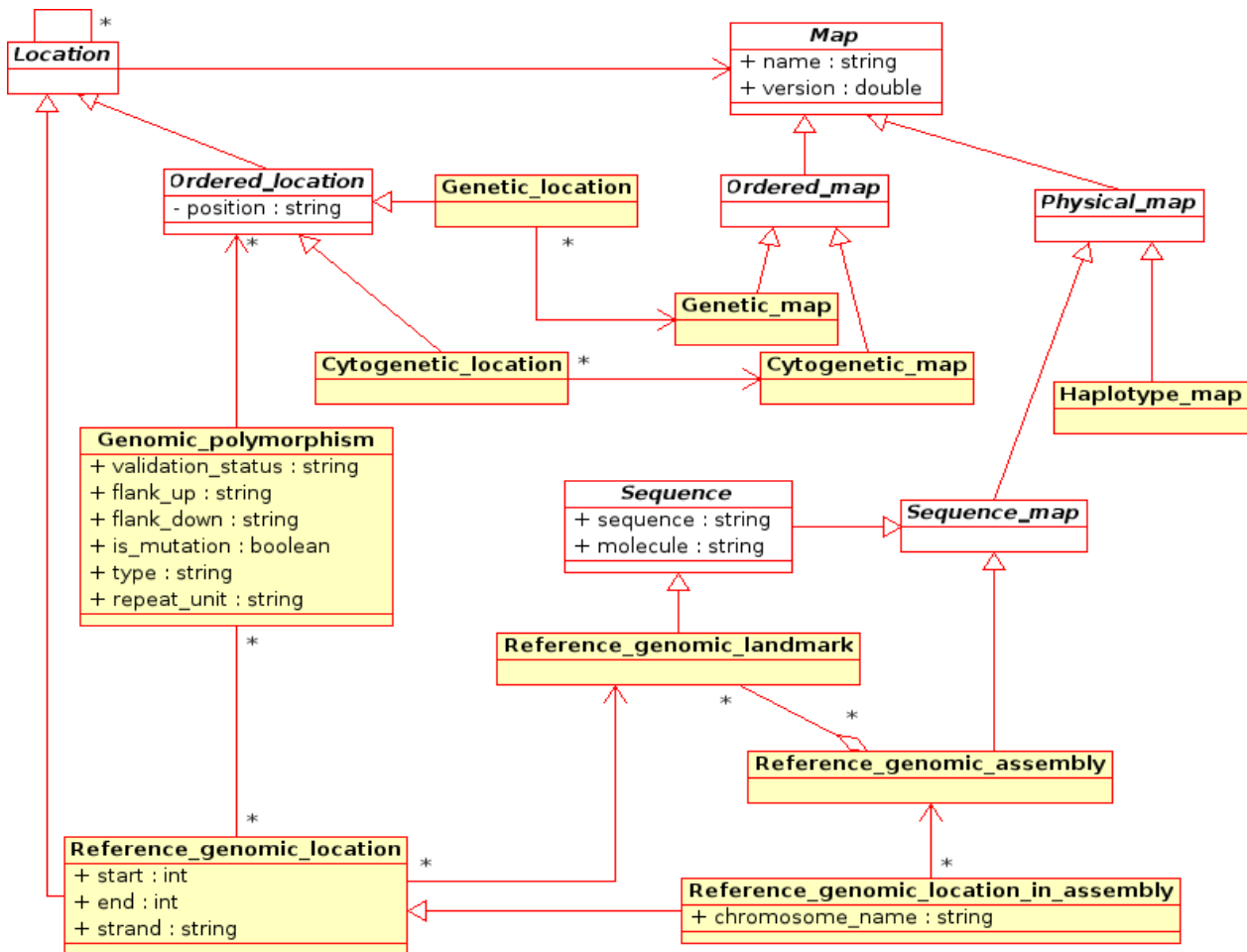
The aim of many haplotype studies is to find haplotypes within a given sequence region that define most of the variation in populations (Set\_of\_haplotypes). Haplotype blocks (Haplotype\_blocks) are clusters determined by linkage disequilibrium-based methods. The other common way to define clusters is to use the sequence region of the gene (Gene\_based\_haplotype).

Haplotypes and sets of haplotypes (Haplotype\_block) can be placed on Haplotype\_maps with additional information about their frequency and distribution.



### 8.1.5 Map

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 of map (Map) there is a corresponding location (Location). Hence a marker (Genetic\_polymorphism) 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 that is a combination of genomic sequences (Reference\_genomic\_landmark). Feature locations can also be expressed in chromosomal locations in an assembly (Reference\_genomic\_location\_in\_assembly).



### 8.1.6 Sequence

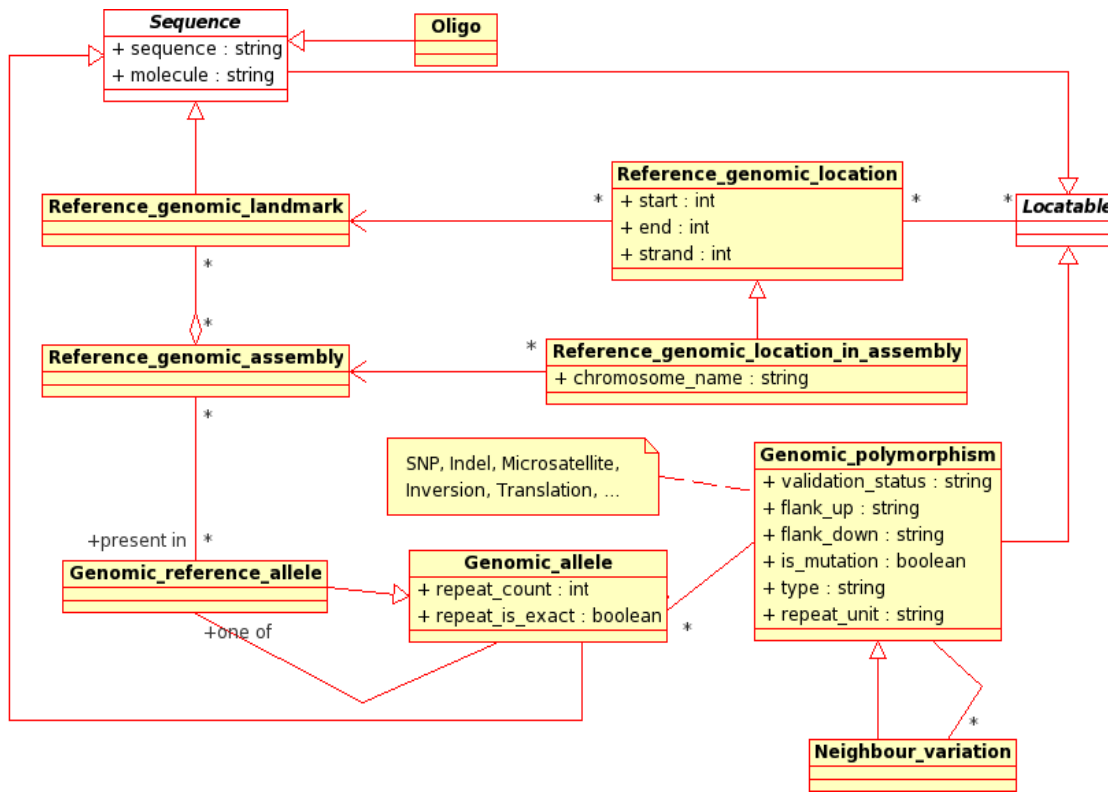
All sequences inherit from a generic Sequence class. All sequences (Sequence) and sequence features (e.g., Genomic\_polymorphism) can be located within a genomic sequence (Reference\_genomic\_sequence). This is modeled by common inheritance from an abstract superclass Locatable that can have several 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\_polymorphisms are called Genomic\_alleles. The genomic allele sequence present in the assembly (Reference\_genomic\_assembly) is labeled as Genomic\_reference\_allele. This together with the strong emphasis on having a position and especially strand in a genomic landmark for all polymorphisms are designed to help reducing the error from getting the orientation of the SNP wrong.

Sometimes polymorphisms, especially single nucleotide polymorphisms (SNPs), occur so close to each other that they affect each others detection oligo primers. It is useful to indicate these as Neighbor\_variations.

Oligos (Oligo) are single strand DNA molecules used in polymorphism assays (Polymorphism\_assay). It is important that the exact orientation (“strand”) and location (Genomic\_reference\_location) of oligos are known so that the detected alleles (Genomic\_alleles) can be identified unequivocally.



### 8.1.7 Translation

Polymorphisms (Genomic\_polymorphisms) that occur with gene regions are of special interest. They (Genic\_polymorphisms) can have an effect to the function of the gene. DNA changes that occur in low frequency in populations but that have a clear effect to the phenotype of an individual are commonly known as mutations. In this model mutations are Genomic\_polymorphisms with attribute `is_mutation` set true.

The structure of the gene (Genomic\_gene\_structure) detailed enough to calculate the effects of mutations is composed of start, end, and strand attributes of all the exons (Exon) and the coding region (CDS) locations (Reference\_genomic\_locations) in the reference sequence (Reference\_genomic\_landmark or Reference\_genomic\_assembly). Start of the gene defines then a new coordinate system with locations upstream getting negative locations. One Genomic\_gene\_structure here describes one transcript structure. More of these transcript structures are needed for genes with alternative transcripts. Genes located close to each other might be affected by the same polymorphism.

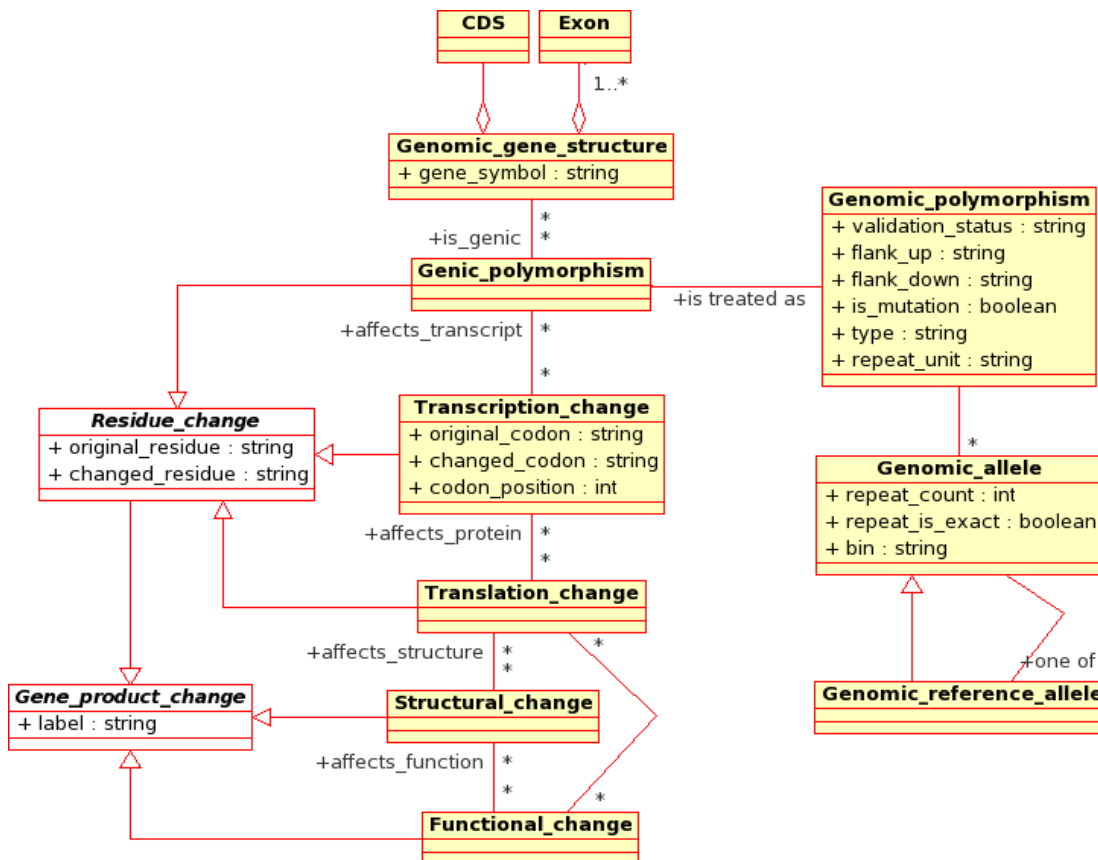
Any polymorphism (Genomic\_polymorphism) can be understood to be part of the gene region. In practice, the region is commonly defined as 5000 nucleotide upstream from gene start to 5000 nucleotides downstream from the end of the coding region. This usually covers all regulatory regions of a gene.

A Genic\_polymorphism can affect the transcript (have a Transcript\_change) either directly by changing its nucleotide content or indirectly changing the amount of transcript produced. If the nucleotide content of the transcript is affected, it is customary to report the codon and its location in the coding region.

A Transcript\_change can lead to a Translation\_change if it affects the translated polypeptide. A Translation\_change can have a known Functional\_change either directly or via a Structural\_change.

Any of these changes contain a description of the original state and changed (mutated) state. For sequence changes, part of the description is the residue change (e.g., A>T or M>I). An agreed ontology is needed to describe all changes. Examples of these are Sequence Ontology and the Mutation Event Controlled Vocabulary which is in use at EBI and built into Bioperl modules Bio::Variation.

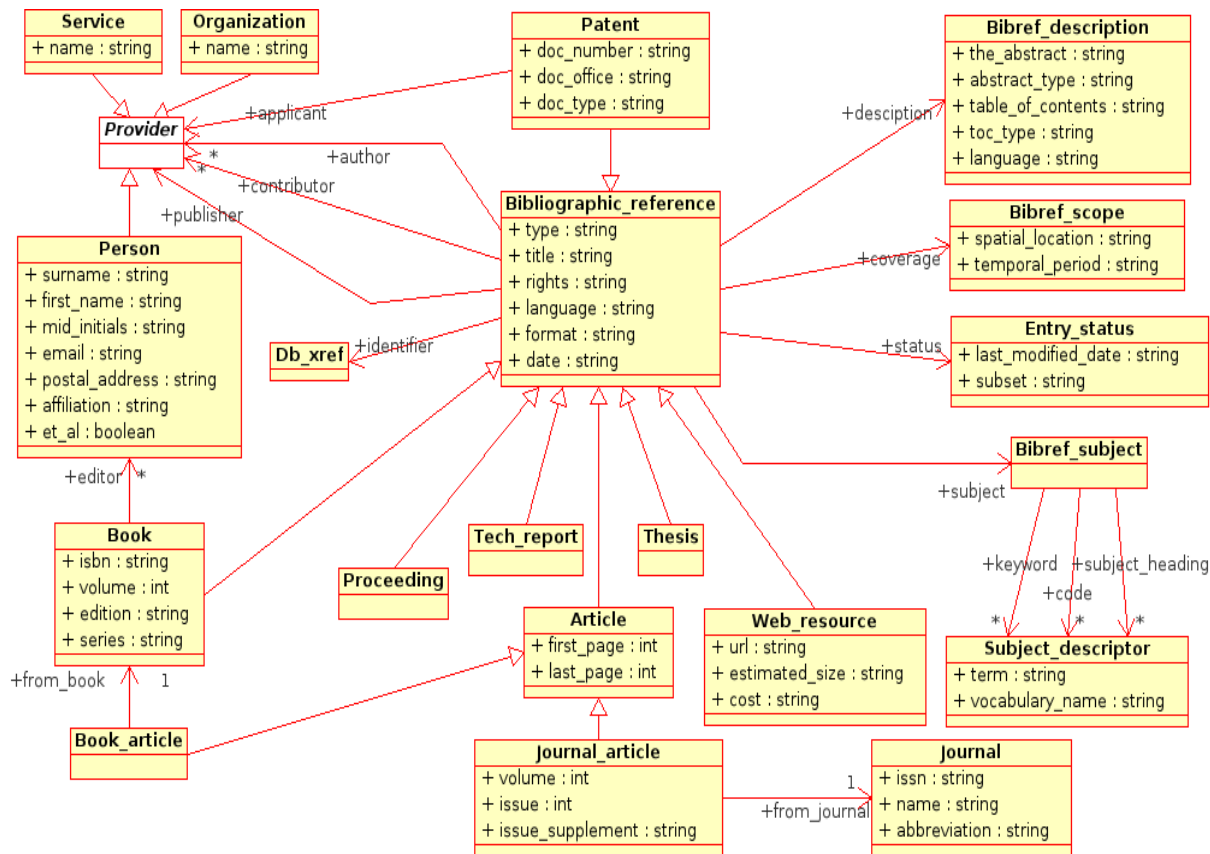
There will undoubtedly be special cases where the relations between classes are variations of those presented in the model. For example, in the case of RNA genes the transcript change can have a Structural\_change or Functional\_change directly without Translation\_change. These special cases should not be added into the model as they can easily be added to the implementations and do not change the meaning of the classes.





## 8.1.8 Bibliographic References

The Publication model is taken from the OMG Bibliographic Query Service. Together with usual attributes (described there, and formally expressed in the platform specific model here), such as “title” or “author,” there is also an attribute “identifiable” containing a reference to any entities defined by this specification. The purpose is to allow chain-referencing (e.g., an entity A1 is linked to a publication P1 that is back linked to an entity B1 - so one can navigate from A1 to B1 via P1).

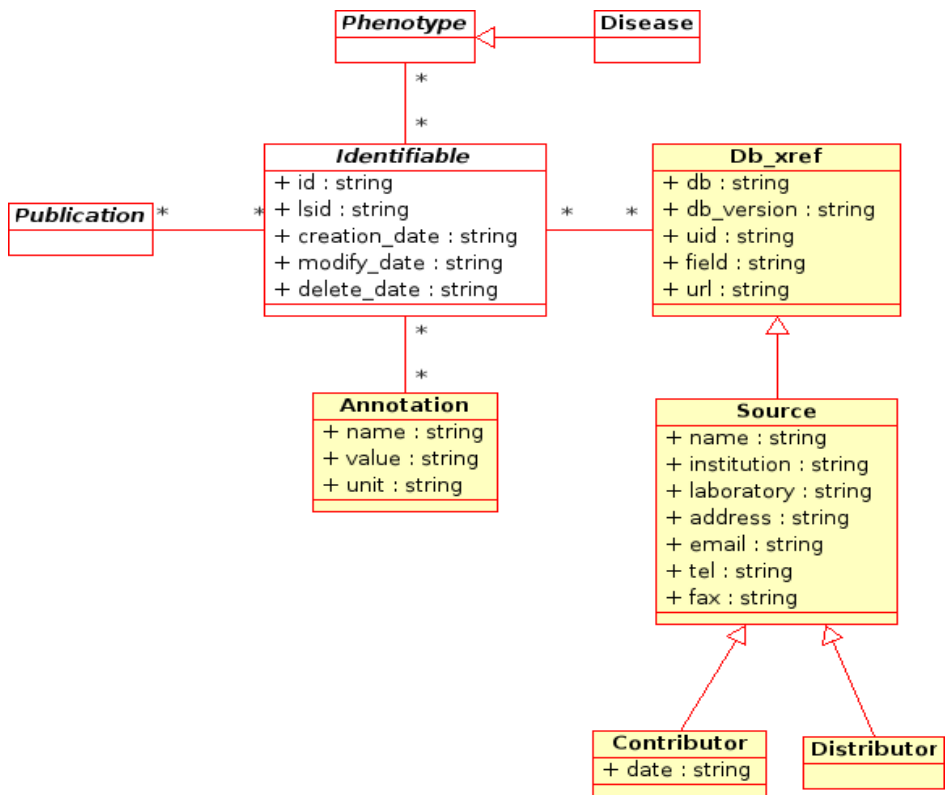


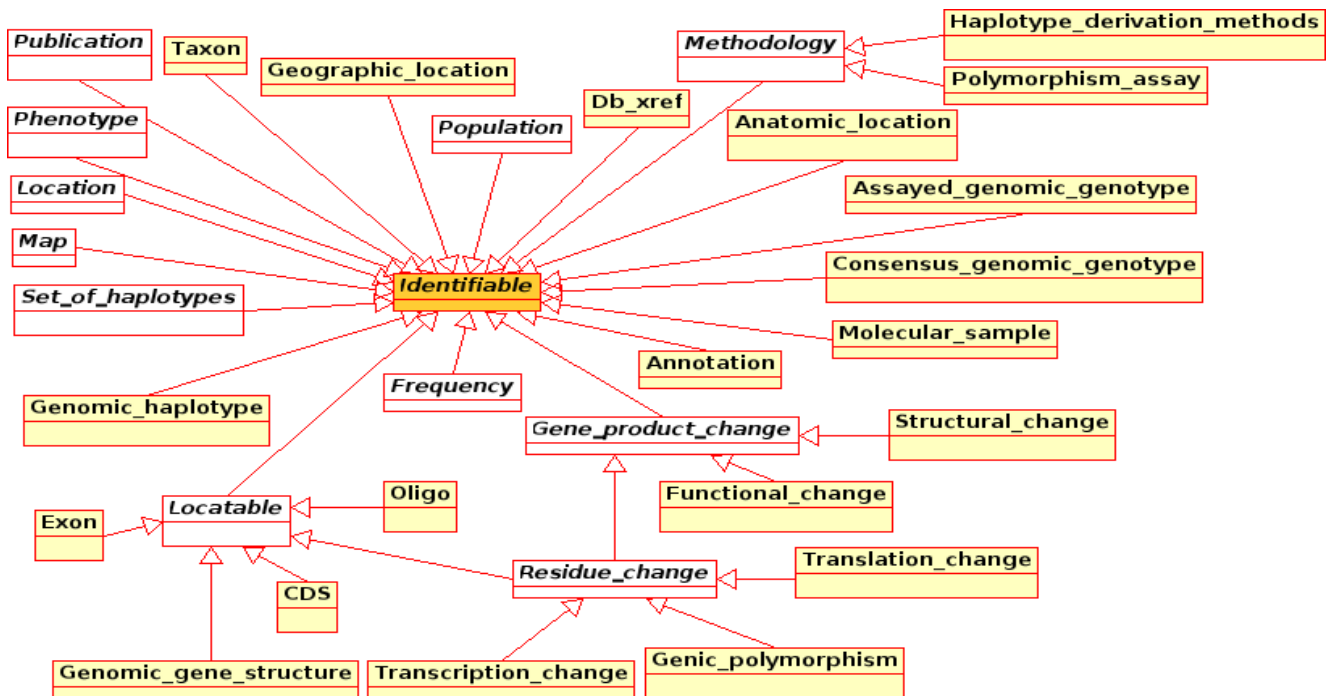
## 8.1.9 Identifiable

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.

Phenotype is included here as a place holder indicating that phenotype can link to many classes in the model but is really beyond the scope of this model.





## 8.2 Model Classes, Attributes, and Associations (Details)

Note that the details below can also be seen as an HTML document attached in the Accompanied Files.

### 8.2.1 Anatomic\_location

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

#### Specialization of: Identifiable

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one Molecular\_sample [defined at Anatomic\_location]
- to one Publication [defined at Identifiable]

- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.2 Annotation

A free text set of name, value, optional unit fields.

#### Specialization of: [Identifiable](#)

Attributes:

- **name** (string), [defined at [Annotation](#)] Name of the annotation feature.
- **value** (string), [defined at [Annotation](#)] Value assigned to the annotation feature.
- **unit** (string), [defined at [Annotation](#)] Optional unit of the annotation feature.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.3 Article

Nonfictional prose forming an independent part of a publication.

#### Specialization of: [Bibliographic reference](#)

Attributes:

- **first\_page** (int), [defined at [Article](#)] Page number of the first page.
- **last\_page** (int), [defined at [Article](#)] Page number of the last page.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.

- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **type** (string), [defined at [Bibliographic reference](#)] 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.
- **title** (string), [defined at [Bibliographic reference](#)] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at [Bibliographic reference](#)] 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], Copyright, and various Property Rights.
- **language** (string), [defined at [Bibliographic reference](#)] 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.
- **format** (string), [defined at [Bibliographic reference](#)] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at [Bibliographic reference](#)] 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.

#### Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Identifiable** [defined at [Publication](#)]
- to one **Bibref\_description** [defined at [Bibliographic reference](#)]
- to one **Bibref\_subject** [defined at [Bibliographic reference](#)]
- to one **Bibref\_scope** [defined at [Bibliographic reference](#)]
- to one **Entry\_status** [defined at [Bibliographic reference](#)]

### 8.2.4 Assayed\_genomic\_genotype

The result of applying a Polymorphism\_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\_polymorphism, where the variation might be due to differing assays, experimental error, dominant systems, missing data, and so forth. Synonym: Measurement

### Specialization of: Identifiable

Attributes:

- **assay\_failure** (boolean), [defined at Assayed\_genomic\_genotype] True if assay failed to result any genotype information.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Polymorphism assay** [defined at Assayed\_genomic\_genotype]
- to one **Genomic allele** [defined at Assayed\_genomic\_genotype]
- to one **Molecular sample** [defined at Assayed\_genomic\_genotype]
- to one **Consensus genomic genotype** [defined at Assayed\_genomic\_genotype]
- to one **Publication** [defined at Identifiable]
- to one **Db xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

## 8.2.5 Bibliographic\_reference

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.

### Specialization of: Publication

Attributes:

- **type** (string), [defined at Bibliographic\_reference] 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.
- **title** (string), [defined at Bibliographic\_reference] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at Bibliographic\_reference] 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], Copyright, and various Property Rights.
- **language** (string), [defined at Bibliographic\_reference] 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.

- **format** (string), [defined at [Bibliographic reference](#)] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at [Bibliographic reference](#)] 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.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **[Bibref description](#)** [defined at [Bibliographic reference](#)]
- to one **[Bibref subject](#)** [defined at [Bibliographic reference](#)]
- to one **[Bibref scope](#)** [defined at [Bibliographic reference](#)]
- to one **[Db xref](#)** [defined at [Bibliographic reference](#)]
- to one **[Entry status](#)** [defined at [Bibliographic reference](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]
- to one **[Identifiable](#)** [defined at [Publication](#)]

## 8.2.6 **Bibref\_description**

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

Attributes:

- **the\_abstract** (string), [defined at [Bibref description](#)] Summary of the main points of the publication.
- **abstract\_type** (string), [defined at [Bibref description](#)] Format of the abstract expressed in MIME (Multipurpose Internet Mail Extensions).
- **table\_of\_contents** (string), [defined at [Bibref description](#)] A list of divisions (chapters or articles) and the pages on which they start.
- **toc\_type** (string), [defined at [Bibref description](#)] Format of the table of contents expressed in MIME (Multipurpose

Internet Mail Extensions).

- **language** (string), [defined at [Bibref\\_description](#)] Language of the description

Associations:

- to one **[Bibliographic reference](#)** [defined at [Bibref\\_description](#)]

### 8.2.7 **Bibref\_scope**

It defines an 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).

**Specialization of: [Identifiable](#)**

Attributes:

- **spatial\_location** (string), [defined at [Bibref\\_scope](#)] A place or area name or stringified geographic coordinates.
- **temporal\_period** (string), [defined at [Bibref\\_scope](#)] A period label, date, or date range.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **[Bibliographic reference](#)** [defined at [Bibref\\_scope](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]

### 8.2.8 **Bibref\_subject**

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

Attributes:

Associations:

- to one **[Subject descriptor](#)** [defined at [Bibref\\_subject](#)]
- to one **[Bibliographic reference](#)** [defined at [Bibref\\_subject](#)]



## 8.2.9 Book

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

### Specialization of: Bibliographic reference

Attributes:

- **isbn** (string), [defined at Book] International Standard Book Number
- **volume** (int), [defined at Book] Volume number of the book in a series.
- **edition** (string), [defined at Book] Name of the edition.
- **series** (string), [defined at Book] Name of a collection of books.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **type** (string), [defined at Bibliographic reference] 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.
- **title** (string), [defined at Bibliographic reference] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at Bibliographic reference] 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], Copyright, and various Property Rights.
- **language** (string), [defined at Bibliographic reference] 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.
- **format** (string), [defined at Bibliographic reference] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at Bibliographic reference] 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.

Associations:

- to one **Book article** [defined at Book]
- to one **Person** [defined at Book]

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Identifiable** [defined at [Publication](#)]
- to one **Bibref\_description** [defined at [Bibliographic\\_reference](#)]
- to one **Bibref\_subject** [defined at [Bibliographic\\_reference](#)]
- to one **Bibref\_scope** [defined at [Bibliographic\\_reference](#)]
- to one **Entry\_status** [defined at [Bibliographic\\_reference](#)]

### 8.2.10 Book\_article

Article that forms part of a book.

#### Specialization of: [Article](#)

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **type** (string), [defined at [Bibliographic\\_reference](#)] 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.
- **title** (string), [defined at [Bibliographic\\_reference](#)] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at [Bibliographic\\_reference](#)] 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], Copyright, and various Property Rights.
- **language** (string), [defined at [Bibliographic\\_reference](#)] 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.
- **format** (string), [defined at [Bibliographic\\_reference](#)] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.

- **date** (string), [defined at [Bibliographic reference](#)] 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.
- **first\_page** (int), [defined at [Article](#)] Page number of the first page.
- **last\_page** (int), [defined at [Article](#)] Page number of the last page.

Associations:

- to one **Book** [defined at [Book article](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Identifiable** [defined at [Publication](#)]
- to one **Bibref description** [defined at [Bibliographic reference](#)]
- to one **Bibref subject** [defined at [Bibliographic reference](#)]
- to one **Bibref scope** [defined at [Bibliographic reference](#)]
- to one **Entry status** [defined at [Bibliographic reference](#)]

### 8.2.11 CDS

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

**Specialization of: [Locatable](#)**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Genomic gene structure** [defined at [CDS](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]

- to one **Phenotype** [defined at Identifiable]
- to one **Reference genomic location** [defined at Locatable]

### 8.2.12 Consensus\_genomic\_genotype

A list of one or more Genomic\_alleles carried by a single individual for a single Genomic\_polymorphism. 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

#### Specialization of: Identifiable

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Genomic allele** [defined at Consensus\_genomic\_genotype]
- to one **Panel** [defined at Consensus\_genomic\_genotype]
- to one **Assayed genomic genotype** [defined at Consensus\_genomic\_genotype]
- to one **Genomic haplotype** [defined at Consensus\_genomic\_genotype]
- to one **Haplotype derivation methods** [defined at Consensus\_genomic\_genotype]
- to one **Molecular sample** [defined at Consensus\_genomic\_genotype]
- to one **Genomic genotype population frequency** [defined at Consensus\_genomic\_genotype]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

### 8.2.13 Contributor

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

## Specialization of: Source

### Attributes:

- **date** (string), [defined at Contributor] Date of the editing.
- **name** (string), [defined at Source] Name of the Source.
- **institution** (string), [defined at Source] The name of the institution.
- **laboratory** (string), [defined at Source] Laboratory or group within an institution.
- **address** (string), [defined at Source] Street address.
- **email** (string), [defined at Source] Contact email.
- **tel** (string), [defined at Source] Telephone number.
- **fax** (string), [defined at Source] FAX number.
- **db** (string), [defined at Db\_xref] The remote database name or abbreviation.
- **db\_version** (string), [defined at Db\_xref] Version of the database
- **uid** (string), [defined at Db\_xref] Unique identifier in the remote database.
- **field** (string), [defined at Db\_xref] Field in the remote database where the UID value is found (e.g., ID or AC in EMBL). Optional.
- **url** (string), [defined at Db\_xref] Full URL to the cross-referenced entry
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

### Associations:

- to one **Bibliographic reference** [defined at Db\_xref]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

## 8.2.14 Cytogenetic\_location

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

### Specialization of: Ordered\_location

Attributes:

- **position** (string), [defined at Ordered\_location] Location description
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one Cytogenetic\_map [defined at Cytogenetic\_location]
- to one Location [defined at Location]
- to one Map [defined at Location]
- to one Genomic\_polymorphism [defined at Ordered\_location]
- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]

### 8.2.15 Cytogenetic\_map

Map of cytogenetic bands describing their relative order.

#### Specialization of: Ordered\_map

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **name** (string), [defined at Map] Name of the map.
- **version** (double), [defined at Map] Version number of the map.

Associations:

- to one Cytogenetic\_location [defined at Cytogenetic\_map]

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Location** [defined at [Map](#)]

### 8.2.16 Db\_xref

Cross reference between two Identifiables in different databases.

#### Specialization of: [Identifiable](#)

Attributes:

- **db** (string), [defined at [Db\\_xref](#)] The remote database name or abbreviation.
- **db\_version** (string), [defined at [Db\\_xref](#)] Version of the database
- **uid** (string), [defined at [Db\\_xref](#)] Unique identifier in the remote database.
- **field** (string), [defined at [Db\\_xref](#)] Field in the remote database where the UID value is found (e.g., ID or AC in EMBL). Optional.
- **url** (string), [defined at [Db\\_xref](#)] Full URL to the cross-referenced entry
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Bibliographic reference** [defined at [Db\\_xref](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.17 Disease

Special case of a phenotype describing a disease state.

### Specialization of: **Phenotype**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Identifiable** [defined at [Phenotype](#)]

### 8.2.18 Distributor

The original Source of the data.

#### Specialization of: **Source**

Attributes:

- **name** (string), [defined at [Source](#)] Name of the Source.
- **institution** (string), [defined at [Source](#)] The name of the institution.
- **laboratory** (string), [defined at [Source](#)] Laboratory or group within an institution.
- **address** (string), [defined at [Source](#)] Street address.
- **email** (string), [defined at [Source](#)] Contact email.
- **tel** (string), [defined at [Source](#)] Telephone number.
- **fax** (string), [defined at [Source](#)] FAX number.
- **db** (string), [defined at [Db\\_xref](#)] The remote database name or abbreviation.
- **db\_version** (string), [defined at [Db\\_xref](#)] Version of the database
- **uid** (string), [defined at [Db\\_xref](#)] Unique identifier in the remote database.
- **field** (string), [defined at [Db\\_xref](#)] Field in the remote database where the UID value is found (e.g., ID or AC in EMBL). Optional.
- **url** (string), [defined at [Db\\_xref](#)] Full URL to the cross-referenced entry



- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Bibliographic reference** [defined at [Db\\_xref](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.19 Entry\_status

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

**Specialization of: [Identifiable](#)**

Attributes:

- **last\_modified\_date** (string), [defined at [Entry\\_status](#)] last date the citation entry was changed
- **subset** (string), [defined at [Entry\\_status](#)] Some bibliographic repositories consist of several, or even many, databases. The sub set helps to locate the citation.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Bibliographic reference** [defined at [Entry\\_status](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

## 8.2.20 Exon

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

### Specialization of: Locatable

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one Genomic gene structure [defined at Exon]
- to one Publication [defined at Identifiable]
- to one Db xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]
- to one Reference genomic location [defined at Locatable]

## 8.2.21 Frequency

Abstract class for frequencies, expressed in percentages.

### 8.2.21.1 Abstract Class

#### Specialization of: Identifiable

Attributes:

- **value** (double), [defined at Frequency] Relative frequency value. Gets values between 0 and 1.
- **count** (int), [defined at Frequency] The actual count of items.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Panel** [defined at [Identifiable](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.22 Functional\_change

Change in the function of the final gene product.

**Specialization of: Gene\_product\_change**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **label** (string), [defined at [Gene\\_product\\_change](#)] Human understandable term for the change (e.g., missense, truncation).

Associations:

- to one **Translation\_change** [defined at [Functional\\_change](#)]
- to one **Structural\_change** [defined at [Functional\\_change](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.23 Gene\_based\_haplotype

Collection of variable nucleotides (Genomic\_alleles in Genomic\_polymorphisms) that define a gene. In older usage same as locus.

**Specialization of: Set\_of\_haplotypes**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.

- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Haplotype\_map** [defined at [Set\\_of\\_haplotypes](#)]
- to one **Genomic\_haplotype** [defined at [Set\\_of\\_haplotypes](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.24 Gene\_product\_change

Abstract superclass for a different genic change descriptors.

Abstract Class

**Specialization of: [Identifiable](#)**

Attributes:

- **label** (string), [defined at [Gene\\_product\\_change](#)] Human understandable term for the change (e.g., missense, truncation).
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.25 Genetic\_location

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

### Specialization of: Ordered\_location

Attributes:

- **position** (string), [defined at Ordered\_location] Location description
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one Genetic\_map [defined at Genetic\_location]
- to one Location [defined at Location]
- to one Map [defined at Location]
- to one Genomic\_polymorphism [defined at Ordered\_location]
- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]

### 8.2.26 Genetic\_map

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

#### Specialization of: Ordered\_map

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **name** (string), [defined at Map] Name of the map.
- **version** (double), [defined at Map] Version number of the map.

Associations:

- to one **Genetic location** [defined at [Genetic\\_map](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Location** [defined at [Map](#)]

### 8.2.27 Genic\_polymorphism

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

Specialization of: **Residue\_change**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **original\_residue** (string), [defined at [Residue\\_change](#)] The affected residue in the reference sequence.
- **changed\_residue** (string), [defined at [Residue\\_change](#)] The new residue in the reference sequence.
- **label** (string), [defined at [Gene\\_product\\_change](#)] Human understandable term for the change (e.g., missense, truncation).

Associations:

- to one **Genomic\_gene\_structure** [defined at [Genomic\\_polymorphism](#)]
- to one **Genomic\_polymorphism** [defined at [Genomic\\_polymorphism](#)]
- to one **Transcription\_change** [defined at [Genomic\\_polymorphism](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Reference\_genomic\_location** [defined at [Locatable](#)]

## 8.2.28 Genomic\_allele

One of several alternative DNA sequences of a Reference\_genomic\_location as it appears in the population of organisms.

Synonym: variant, allele

### Specialization of: Sequence

Attributes:

- **repeat\_count** (int), [defined at Genomic\_allele] If the Genomic\_variation type is 'microsatellite,' get number of repeat units as value, e.g, 7.
- **repeat\_is\_exact** (boolean), [defined at Genomic\_allele] If the Genomic\_variation type is 'microsatellite,' gets value true if the allele region consists of repeat units only.
- **bin** (string), [defined at Genomic\_allele] Size class for microsatellite alleles when exact size cannot be determined. Semantic name.
- **sequence** (string), [defined at Sequence] The residue sequence string.
- **molecule** (string), [defined at Sequence] Indicates the alphabet of the sequence molecule (e.g., 'DNA,' 'RNA,' 'protein').
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **name** (string), [defined at Map] Name of the map.
- **version** (double), [defined at Map] Version number of the map.

Associations:

- to one Consensus\_genomic\_genotype [defined at Genomic\_allele]
- to one Genomic\_polymorphism [defined at Genomic\_allele]
- to one Assayed\_genomic\_genotype [defined at Genomic\_allele]
- to one Genomic\_haplotype [defined at Genomic\_allele]
- to one Genomic\_reference\_allele [defined at Genomic\_allele]
- to one Genomic\_allele\_population\_frequency [defined at Genomic\_allele]
- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]

- to one **Reference\_genomic\_location** [defined at [Locatable](#)]
- to one **Location** [defined at [Map](#)]

### 8.2.29 Genomic\_allele\_population\_frequency

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.

#### Specialization of: [Frequency](#)

Attributes:

- **value** (double), [defined at [Frequency](#)] Relative frequency value. Gets values between 0 and 1.
- **count** (int), [defined at [Frequency](#)] The actual count of items.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Genomic\_allele** [defined at [Genomic\\_allele\\_population\\_frequency](#)]
- to one **Panel** [defined at [Genomic\\_allele\\_population\\_frequency](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.30 Genomic\_gene\_structure

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

#### Specialization of: [Locatable](#)

Attributes:

- **gene\_symbol** (string), [defined at [Genomic\\_gene\\_structure](#)] The gene symbol for the gene (e.g., approved by the HUGO nomenclature committee).
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.



- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Genic polymorphism** [defined at [Genomic\\_gene\\_structure](#)]
- to one **Exon** [defined at [Genomic\\_gene\\_structure](#)]
- to one **CDS** [defined at [Genomic\\_gene\\_structure](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Reference\_genomic\_location** [defined at [Locatable](#)]

### 8.2.31 Genomic\_genotype\_population\_frequency

Frequency of a Consensus\_genomic\_genotype in a Panel.

**Specialization of: [Frequency](#)**

Attributes:

- **value** (double), [defined at [Frequency](#)] Relative frequency value. Gets values between 0 and 1.
- **count** (int), [defined at [Frequency](#)] The actual count of items.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Consensus\_genomic\_genotype** [defined at [Genomic\\_genotype\\_population\\_frequency](#)]
- to one **Panel** [defined at [Frequency](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.32 Genomic\_haplotype

A set of Genomic\_alleles across an equal number of Genomic\_polymorphisms in a single chromosome and in a single individual. The Genomic\_haplotype is derived from a set of Consensus\_genomic\_genotype. For each Genomic\_polymorphism, 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.

#### Specialization of: Identifiable

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one Haplotype\_derivation\_methods [defined at Genomic\_haplotype]
- to one Genomic\_haplotype\_population\_frequency [defined at Genomic\_haplotype]
- to one Genomic\_allele [defined at Genomic\_haplotype]
- to one Consensus\_genomic\_genotype [defined at Genomic\_haplotype]
- to one Molecular\_sample [defined at Genomic\_haplotype]
- to one Haplotype\_map [defined at Genomic\_haplotype]
- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]

### 8.2.33 Genomic\_haplotype\_population\_frequency

Frequency of a Genomic\_haplotype in a Panel.

#### Specialization of: Frequency

Attributes:

- **value** (double), [defined at Frequency] Relative frequency value. Gets values between 0 and 1.
- **count** (int), [defined at Frequency] The actual count of items.
- **id** (string), [defined at Identifiable] Unique identifier to the data.

- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Genomic haplotype** [defined at [Genomic\\_haplotype\\_population\\_frequency](#)]
- to one **Panel** [defined at [Frequency](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.34 Genomic\_polymorphism

A variable site in a Reference\_genomic\_landmark sequence. Synonyms: polymorphic\_site, variation, SNP

Specialization of: [Locatable](#)

Attributes:

- **validation\_status** (string), [defined at [Genomic\\_polymorphism](#)] Validation status (e.g., “Proven,” “Suspected”).
- **flank\_up** (string), [defined at [Genomic\\_polymorphism](#)] Upstream flanking sequence (at least 25 residues, if possible).
- **flank\_down** (string), [defined at [Genomic\\_polymorphism](#)] Downstream flanking sequence (at least 25 residues, if possible).
- **is\_mutation** (boolean), [defined at [Genomic\\_polymorphism](#)] Proven phenotype change causing mutation.
- **type** (string), [defined at [Genomic\\_polymorphism](#)] The type of the polymorphism (e.g., SNP, microsatellite, indel, translocation, ...).
- **repeat\_unit** (string), [defined at [Genomic\\_polymorphism](#)] If type is ‘microsatellite,’ gives the repeat unit, e.g., “CA.”
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Heterozygosity** [defined at [Genomic\\_polymorphism](#)]

- to one **Genic polymorphism** [defined at Genomic polymorphism]
- to one **Reference genomic landmark** [defined at Genomic polymorphism]
- to one **Neighbour variation** [defined at Genomic polymorphism]
- to one **Genomic allele** [defined at Genomic polymorphism]
- to one **Polymorphism assay** [defined at Genomic polymorphism]
- to one **Reference genomic location** [defined at Genomic polymorphism]
- to one **Publication** [defined at Identifiable]
- to one **Db xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

### 8.2.35 Genomic\_reference\_allele

The DNA sequence of a Genomic\_polymorphism as it appears in the public database.

#### Specialization of: **Genomic allele**

Attributes:

- **repeat\_count** (int), [defined at Genomic allele] If the Genomic\_variation type is ‘microsatellite,’ get number of repeat units as value (e.g., 7).
- **repeat\_is\_exact** (boolean), [defined at Genomic allele] If the Genomic\_variation type is ‘microsatellite,’ gets value true if the allele region consists of repeat units only.
- **bin** (string), [defined at Genomic allele] Size class for microsatellite alleles when exact size cannot be determined. Semantic name.
- **sequence** (string), [defined at Sequence] The residue sequence string.
- **molecule** (string), [defined at Sequence] Indicates the alphabet of the sequence molecule (e.g., ‘DNA,’ ‘RNA,’ ‘protein’).
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **name** (string), [defined at Map] Name of the map.
- **version** (double), [defined at Map] Version number of the map.

Associations:

- to one **Genomic allele** [defined at Genomic reference allele]
- to one **Reference genomic assembly** [defined at Genomic reference allele]
- to one **Consensus genomic genotype** [defined at Genomic allele]
- to one **Genomic polymorphism** [defined at Genomic allele]
- to one **Assayed genomic genotype** [defined at Genomic allele]
- to one **Genomic haplotype** [defined at Genomic allele]
- to one **Genomic reference allele** [defined at Genomic allele]
- to one **Genomic allele population frequency** [defined at Genomic allele]
- to one **Publication** [defined at Identifiable]
- to one **Db xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]
- to one **Reference genomic location** [defined at Locatable]
- to one **Location** [defined at Map]

### 8.2.36 Geographic\_location

Location of an individual or population in a geographic map. Locations are expressed in decimal degrees. Northern latitudes and eastern longitudes have positive values by convention.

**Specialization of: Identifiable**

Attributes:

- **max\_longitude** (double), [defined at Geographic\_location] Upper bound of a geographical location expressed in degrees from the Greenwich meridian.
- **max\_latitude** (double), [defined at Geographic\_location] Upper bound of a geographical location expressed in degrees from the equator.
- **min\_longitude** (double), [defined at Geographic\_location] Lower bound of a geographical location expressed in degrees from the Greenwich meridian.
- **min\_latitude** (double), [defined at Geographic\_location] Lower bound of a geographical location expressed in degrees from the equator.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.

- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.37 Haplotype\_block

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.

Specialization of: **Set of haplotypes**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Haplotype\_map** [defined at [Set of haplotypes](#)]
- to one **Genomic\_haplotype** [defined at [Set of haplotypes](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.38 Haplotype\_derivation\_methods

Association class describing methods used to derive Genomic\_haplotypes from Consensus\_genomic\_genotypes.

Specialization of: **Methodology**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.

- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Consensus genomic genotype** [defined at [Haplotype derivation methods](#)]
- to one **Genomic haplotype** [defined at [Haplotype derivation methods](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.39 Haplotype\_map

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.

**Specialization of: Physical map**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **name** (string), [defined at [Map](#)] Name of the map.
- **version** (double), [defined at [Map](#)] Version number of the map.

Associations:

- to one **Genomic haplotype** [defined at [Haplotype\\_map](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Location** [defined at [Map](#)]

## 8.2.40 Heterozygosity

A numerical measure of observed abundance of different Genomic\_alleles in a Genomic\_polymorphism.

### Specialization of: Frequency

Attributes:

- **value** (double), [defined at Frequency] Relative frequency value. Gets values between 0 and 1.
- **count** (int), [defined at Frequency] The actual count of items.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Genomic\_polymorphism** [defined at Heterozygosity]
- to one **Panel** [defined at Frequency]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

## 8.2.41 Identifiable

Anything with a unique identifier.

### Abstract Class

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Publication** [defined at Identifiable]



- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

## 8.2.42 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.

### Specialization of: **Population**

Attributes:

- **father\_id** (string), [defined at Individual] Id of the father to allow building of pedigrees.
- **mother\_id** (string), [defined at Individual] Id of the mother to allow building of pedigrees.
- **sex** (string), [defined at Individual] Recommended values are ‘unknown,’ ‘male,’ and ‘female.’ Additional values can be used to denote unusual karyotypes.
- **birth\_date** (string), [defined at Individual] Date of birth of the individual. May be better abbreviated to birth year to protect the privacy of the individual.
- **death\_date** (int), [defined at Individual] Date of death of the individual. May be better abbreviated to plain year to protect the privacy of the individual.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **race** (string), [defined at Population] Broad ethnic category of the population sample or “mixed.”
- **ethnicity** (string), [defined at Population] Additional ethnic category of the population sample or “mixed.”
- **primary\_language** (string), [defined at Population] Language name or code (e.g., as in Ethnologue).
- **language\_family** (string), [defined at Population] Language family name or code (e.g., as in Ethnologue).

Associations:

- to one **Panel** [defined at Individual]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

- to one **Geographic location** [defined at Population]
- to one **Taxon** [defined at Population]
- to one **Molecular sample** [defined at Population]

### 8.2.43 Journal

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

#### Specialization of: **Identifiable**

Attributes:

- **issn** (string), [defined at Journal] The ISSN (International Standard Serial Number) is an eight-digit number which identifies periodical publications as such.
- **name** (string), [defined at Journal] Full name
- **abbreviation** (string), [defined at Journal] Accepted abbreviated form of the journal name.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Journal article** [defined at Journal]
- to one **Publication** [defined at Identifiable]
- to one **Db xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

### 8.2.44 Journal\_article

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

#### Specialization of: **Article**

Attributes:

- **volume** (int), [defined at Journal article] Volume number of the journal. Typically all issues of a journal published within a year belong to the same volume.
- **issue** (int), [defined at Journal article] Integer indicating the ordering of the issue within a volume.

- **issue\_supplement** (string), [defined at [Journal\\_article](#)] Additional identifier for a separate supplement of the issue.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **type** (string), [defined at [Bibliographic\\_reference](#)] 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.
- **title** (string), [defined at [Bibliographic\\_reference](#)] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at [Bibliographic\\_reference](#)] 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], Copyright, and various Property Rights.
- **language** (string), [defined at [Bibliographic\\_reference](#)] 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.
- **format** (string), [defined at [Bibliographic\\_reference](#)] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at [Bibliographic\\_reference](#)] 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.
- **first\_page** (int), [defined at [Article](#)] Page number of the first page.
- **last\_page** (int), [defined at [Article](#)] Page number of the last page.

#### Associations:

- to one **[Journal](#)** [defined at [Journal\\_article](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]
- to one **[Identifiable](#)** [defined at [Publication](#)]
- to one **[Bibref\\_description](#)** [defined at [Bibliographic\\_reference](#)]
- to one **[Bibref\\_subject](#)** [defined at [Bibliographic\\_reference](#)]

- to one **Bibref\_scope** [defined at Bibliographic\_reference]
- to one **Entry\_status** [defined at Bibliographic\_reference]

## 8.2.45 Locatable

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

### Abstract Class

#### Specialization of: Identifiable

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Reference\_genomic\_location** [defined at Locatable]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

## 8.2.46 Location

A position in an abstract map.

### 8.2.46.1 Abstract Class

#### Specialization of: Identifiable

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Location** [defined at [Location](#)]
- to one **Map** [defined at [Location](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

## 8.2.47 Map

An abstract map.

### 8.2.47.1 Abstract Class

Specialization of: [Identifiable](#)

Attributes:

- **name** (string), [defined at [Map](#)] Name of the map.
- **version** (double), [defined at [Map](#)] Version number of the map.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Location** [defined at [Map](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

## 8.2.48 Methodology

Abstract class for a laboratory method or a computational protocol.

### 8.2.48.1 Abstract Class

#### Specialization of: Identifiable

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]

### 8.2.49 Molecular\_sample

A sample from an Individual or from a Panel defining the molecule and tissue/cell used (Anatomic\_locations) in the Polymorphism\_assay. Synonym: Sample\_of\_individual

#### Specialization of: Identifiable

Attributes:

- **molecule** (string), [defined at Molecular\_sample] The molecule (RNA, DNA, protein) used in the assay.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one Polymorphism\_assay [defined at Molecular\_sample]
- to one Anatomic\_location [defined at Molecular\_sample]
- to one Genomic\_haplotype [defined at Molecular\_sample]
- to one Molecular\_sample [defined at Molecular\_sample]
- to one Assayed\_genomic\_genotype [defined at Molecular\_sample]

- to one **Consensus genomic genotype** [defined at Molecular sample]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

### 8.2.50 Neighbor\_variation

Another Genomic\_polymorphism close enough to affect the primer design.

#### Specialization of: **Genomic polymorphism**

Attributes:

- **validation\_status** (string), [defined at Genomic polymorphism] Validation status (e.g., “Proven,” “Suspected”).
- **flank\_up** (string), [defined at Genomic polymorphism] Upstream flanking sequence (at least 25 residues, if possible).
- **flank\_down** (string), [defined at Genomic polymorphism] Downstream flanking sequence (at least 25 residues, if possible).
- **is\_mutation** (boolean), [defined at Genomic polymorphism] Proven phenotype change causing mutation.
- **type** (string), [defined at Genomic polymorphism] The type of the polymorphism (e.g., SNP, microsatellite, indel, translocation, ... ).
- **repeat\_unit** (string), [defined at Genomic polymorphism] If type is ‘microsatellite,’ gives the repeat unit (e.g., “CA”).
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Genomic polymorphism** [defined at Neighbor variation]
- to one **Heterozygosity** [defined at Genomic polymorphism]
- to one **Genic polymorphism** [defined at Genomic polymorphism]
- to one **Reference genomic landmark** [defined at Genomic polymorphism]
- to one **Neighbour variation** [defined at Genomic polymorphism]
- to one **Genomic allele** [defined at Genomic polymorphism]
- to one **Polymorphism assay** [defined at Genomic polymorphism]
- to one **Reference genomic location** [defined at Genomic polymorphism]

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.51 Oligo

A DNA oligonucleotide used for detection and assay of Genomic\_polymorphisms in Polymorphism\_assays. Synonym: primer

**Specialization of: [Sequence](#), [Locatable](#)**

Attributes:

- **sequence** (string), [defined at [Sequence](#)] The residue sequence string.
- **molecule** (string), [defined at [Sequence](#)] Indicates the alphabet of the sequence molecule (e.g., 'DNA,' 'RNA,' 'protein').
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **name** (string), [defined at [Map](#)] Name of the map.
- **version** (double), [defined at [Map](#)] Version number of the map.

Associations:

- to one **Polymorphism\_assay** [defined at [Oligo](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Reference\_genomic\_location** [defined at [Locatable](#)]
- to one **Location** [defined at [Map](#)]

### 8.2.52 Ordered\_location

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



### 8.2.52.1 Abstract Class

#### Specialization of: Location

Attributes:

- **position** (string), [defined at Ordered\_location] Location description.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Genomic\_polymorphism** [defined at Ordered\_location]
- to one **Location** [defined at Location]
- to one **Map** [defined at Location]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

### 8.2.53 Ordered\_map

A non-metric map category.

#### 8.2.53.1 Abstract Class

##### Specialization of: Map

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **name** (string), [defined at Map] Name of the map.
- **version** (double), [defined at Map] Version number of the map.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Location** [defined at [Map](#)]

### 8.2.54 Organization

Affiliation of a group of unidentified persons.

**Specialization of: [Provider](#)**

Attributes:

- **name** (string), [defined at [Organization](#)] Full name of the organization
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Bibliographic reference** [defined at [Provider](#)]
- to one **Patent** [defined at [Provider](#)]

### 8.2.55 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 term “population” here.) Synonym: SampleSet, Sample from population(s), “Plate” in Coriel sense.

**Specialization of: [Population](#)**

Attributes:

- **size** (long), [defined at [Panel](#)] The size of the sample. Note that the count\_unit field affects how this value is interpreted.

- **count\_unit** (string), [defined at [Panel](#)] Values are ‘chromosome’ and ‘individual.’ Default is ‘individual.’
- **pooled** (boolean), [defined at [Panel](#)] True if accessioned individuals are not available.
- **type** (string), [defined at [Panel](#)] Optional identifier of the panel category: e.g., plate, family, population sample.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **race** (string), [defined at [Population](#)] Broad ethnic category of the population sample or “mixed.”
- **ethnicity** (string), [defined at [Population](#)] Additional ethnic category of the population sample or “mixed.”
- **primary\_language** (string), [defined at [Population](#)] Language name or code (e.g., as in Ethnologue).
- **language\_family** (string), [defined at [Population](#)] Language family name or code (e.g., as in Ethnologue).

Associations:

- to one **Consensus genomic genotype** [defined at [Panel](#)]
- to one **Genomic allele population frequency** [defined at [Panel](#)]
- to one **Panel** [defined at [Panel](#)]
- to one **Individual** [defined at [Panel](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Geographic location** [defined at [Population](#)]
- to one **Taxon** [defined at [Population](#)]
- to one **Molecular sample** [defined at [Population](#)]

## 8.2.56 Patent

A document granting an inventor sole rights to an invention.

Specialization of: **Bibliographic reference**

Attributes:

- **doc\_number** (string), [defined at [Patent](#)] Patent identifier
- **doc\_office** (string), [defined at [Patent](#)] Name of the patent office (e.g., European Patent Office).

- **doc\_type** (string), [defined at [Patent](#)] Type of a patent (e.g., ‘utility,’ ‘design,’ ‘plant’).
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **type** (string), [defined at [Bibliographic reference](#)] 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.
- **title** (string), [defined at [Bibliographic reference](#)] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at [Bibliographic reference](#)] 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], Copyright, and various Property Rights.
- **language** (string), [defined at [Bibliographic reference](#)] 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.
- **format** (string), [defined at [Bibliographic reference](#)] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at [Bibliographic reference](#)] 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.

#### Associations:

- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]
- to one **[Identifiable](#)** [defined at [Publication](#)]
- to one **[Bibref description](#)** [defined at [Bibliographic reference](#)]
- to one **[Bibref subject](#)** [defined at [Bibliographic reference](#)]
- to one **[Bibref scope](#)** [defined at [Bibliographic reference](#)]
- to one **[Entry status](#)** [defined at [Bibliographic reference](#)]

## 8.2.57 Person

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

### Specialization of: Provider

Attributes:

- **surname** (string), [defined at Person] Last name.
- **first\_name** (string), [defined at Person] First name.
- **mid\_initials** (string), [defined at Person] Initials of any subsequent personal names.
- **email** (string), [defined at Person] email address.
- **postal\_address** (string), [defined at Person] Full postal address.
- **affiliation** (string), [defined at Person] Name of the organization the person is affiliated with.
- **et\_al** (boolean), [defined at Person] Indicator that there are more persons in the list of persons; used typically the bibliographic reference contains only an abbreviated author list. Last listed author should have et\_al set to true.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Book** [defined at Person]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]
- to one **Bibliographic reference** [defined at Provider]
- to one **Patent** [defined at Provider]

## 8.2.58 Phenotype

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

### 8.2.58.1 Abstract Class

#### Specialization of: Identifiable

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one Identifiable [defined at Phenotype]
- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]

### 8.2.59 Physical\_map

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

#### 8.2.59.1 Abstract Class

##### Specialization of: Map

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **name** (string), [defined at Map] Name of the map.
- **version** (double), [defined at Map] Version number of the map.

Associations:

- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]

- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]
- to one **Location** [defined at Map]

## 8.2.60 Polymorphism\_assay

An experimental protocol and set of reagents for detecting the Genomic\_alleles of Genomic\_polymorphisms carried by an individual or a panel of individuals. Synonym: Assay

### Specialization of: **Methodology**

Attributes:

- **description** (string), [defined at Polymorphism\_assay] Free text description of the assay protocol.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Genomic\_polymorphism** [defined at Polymorphism\_assay]
- to one **Molecular\_sample** [defined at Polymorphism\_assay]
- to one **Assayed\_genomic\_genotype** [defined at Polymorphism\_assay]
- to one **Oligo** [defined at Polymorphism\_assay]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

## 8.2.61 Population

A interbreeding set of individuals, from whom a Panel is drawn.

### 8.2.61.1 Abstract Class

#### Specialization of: **Identifiable**

Attributes:

- **race** (string), [defined at Population] Broad ethnic category of the population sample or “mixed.”

- **ethnicity** (string), [defined at [Population](#)] Additional ethnic category of the population sample or “mixed.”
- **primary\_language** (string), [defined at [Population](#)] Language name or code (e.g., as in Ethnologue).
- **language\_family** (string), [defined at [Population](#)] Language family name or code (e.g., as in Ethnologue).
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **[Geographic location](#)** [defined at [Population](#)]
- to one **[Taxon](#)** [defined at [Population](#)]
- to one **[Molecular sample](#)** [defined at [Population](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]

## 8.2.62 Proceeding

Printed communication from an scientific organization.

**Specialization of: [Bibliographic reference](#)**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **type** (string), [defined at [Bibliographic reference](#)] 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.
- **title** (string), [defined at [Bibliographic reference](#)] A title given to the cited resource (a name by which the resource is formally known).



- **rights** (string), [defined at [Bibliographic reference](#)] 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], Copyright, and various Property Rights.
- **language** (string), [defined at [Bibliographic reference](#)] 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.
- **format** (string), [defined at [Bibliographic reference](#)] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at [Bibliographic reference](#)] 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.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Identifiable** [defined at [Publication](#)]
- to one **Bibref description** [defined at [Bibliographic reference](#)]
- to one **Bibref subject** [defined at [Bibliographic reference](#)]
- to one **Bibref scope** [defined at [Bibliographic reference](#)]
- to one **Entry status** [defined at [Bibliographic reference](#)]

## 8.2.63 Provider

The class Provider and its sub-classes define active participants of the process of creation and dissemination of the 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 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.

### 8.2.63.1 Abstract Class

Specialization of: [Identifiable](#)

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.

- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **[Bibliographic reference](#)** [defined at [Provider](#)]
- to one **[Patent](#)** [defined at [Provider](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]

## 8.2.64 Publication

A Citation.

### 8.2.64.1 Abstract Class

Specialization of: [Identifiable](#)

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **[Identifiable](#)** [defined at [Publication](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]

## 8.2.65 Reference\_genomic\_assembly

Set of reference sequences ([Reference\\_genomic\\_landmarks](#)) with an algorithm to define the genome and its constituents.

### Specialization of: Sequence map

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **name** (string), [defined at Map] Name of the map.
- **version** (double), [defined at Map] Version number of the map.

Associations:

- to one Reference genomic location [defined at Reference genomic assembly]
- to one Genomic reference allele [defined at Reference genomic assembly]
- to one Reference genomic landmark [defined at Reference genomic assembly]
- to one Taxon [defined at Reference genomic assembly]
- to one Reference genomic location in assembly [defined at Reference genomic assembly]
- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]
- to one Location [defined at Map]

### 8.2.66 Reference genomic landmark

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

### Specialization of: Sequence

Attributes:

- **sequence** (string), [defined at Sequence] The residue sequence string.
- **molecule** (string), [defined at Sequence] Indicates the alphabet of the sequence molecule (e.g., 'DNA,' 'RNA,' 'protein').
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.

- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **name** (string), [defined at [Map](#)] Name of the map.
- **version** (double), [defined at [Map](#)] Version number of the map.

Associations:

- to one **[Reference\\_genomic\\_assembly](#)** [defined at [Reference\\_genomic\\_landmark](#)]
- to one **[Genomic\\_polymorphism](#)** [defined at [Reference\\_genomic\\_landmark](#)]
- to one **[Reference\\_genomic\\_location](#)** [defined at [Reference\\_genomic\\_landmark](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]
- to one **[Location](#)** [defined at [Map](#)]

### 8.2.67 [Reference\\_genomic\\_location](#)

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](#).

#### Specialization of: [Location](#)

Attributes:

- **start** (int), [defined at [Reference\\_genomic\\_location](#)] Start of the location in the reference sequence.
- **end** (int), [defined at [Reference\\_genomic\\_location](#)] End of the location in the reference sequence.
- **strand** (string), [defined at [Reference\\_genomic\\_location](#)] Orientation of the feature in the reference sequence. One of 'forward,' 'reverse,' 'unknown.' Defaults to 'unknown.'
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **[Reference\\_genomic\\_assembly](#)** [defined at [Reference\\_genomic\\_location](#)]

- to one **Reference\_genomic\_landmark** [defined at Reference\_genomic\_location]
- to one **Genomic\_polymorphism** [defined at Reference\_genomic\_location]
- to one **Location** [defined at Location]
- to one **Map** [defined at Location]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

### 8.2.68 **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.

#### **Specialization of: Reference\_genomic\_location**

Attributes:

- **chromosome\_name** (string), [defined at Reference\_genomic\_location\_in\_assembly] Name of the chromosome in the assembly.
- **start** (int), [defined at Reference\_genomic\_location] Start of the location in the reference sequence.
- **end** (int), [defined at Reference\_genomic\_location] End of the location in the reference sequence.
- **strand** (string), [defined at Reference\_genomic\_location] Orientation of the feature in the reference sequence. One of 'forward,' 'reverse,' 'unknown.' Defaults to 'unknown.'
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.

Associations:

- to one **Reference\_genomic\_assembly** [defined at Reference\_genomic\_location\_in\_assembly]
- to one **Reference\_genomic\_landmark** [defined at Reference\_genomic\_location]
- to one **Genomic\_polymorphism** [defined at Reference\_genomic\_location]
- to one **Location** [defined at Location]
- to one **Map** [defined at Location]
- to one **Publication** [defined at Identifiable]

- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]

## 8.2.69 Residue\_change

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

### 8.2.69.1 Abstract Class

Specialization of: **Locatable**, **Gene\_product\_change**

Attributes:

- **original\_residue** (string), [defined at Residue\_change] The affected residue in the reference sequence.
- **changed\_residue** (string), [defined at Residue\_change] The new residue in the reference sequence.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **label** (string), [defined at Gene\_product\_change] Human understandable term for the change (e.g., missense, truncation).

Associations:

- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]
- to one **Reference\_genomic\_location** [defined at Locatable]

## 8.2.70 Sequence

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

### 8.2.70.1 Abstract Class

Specialization of: **Sequence\_map**, **Locatable**

Attributes:

- **sequence** (string), [defined at Sequence] The residue sequence string.

- **molecule** (string), [defined at [Sequence](#)] Indicates the alphabet of the sequence molecule (e.g., 'DNA,' 'RNA,' 'protein').
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **name** (string), [defined at [Map](#)] Name of the map.
- **version** (double), [defined at [Map](#)] Version number of the map.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Reference\_genomic\_location** [defined at [Locatable](#)]
- to one **Location** [defined at [Map](#)]

## 8.2.71 Sequence\_map

A map composed of one or more biological sequences.

### 8.2.71.1 Abstract Class

Specialization of: **Physical\_map**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **name** (string), [defined at [Map](#)] Name of the map.
- **version** (double), [defined at [Map](#)] Version number of the map.

Associations:

- to one **Publication** [defined at [Identifiable](#)]

- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Location** [defined at [Map](#)]

## 8.2.72 Service

Provide of software service

### Specialization of: **Provider**

Attributes:

- **name** (string), [defined at [Service](#)] Full name.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Bibliographic reference** [defined at [Provider](#)]
- to one **Patent** [defined at [Provider](#)]

## 8.2.73 Set\_of\_haplotypes

An abstract representation of a collection of haplotypes.

### 8.2.73.1 Abstract Class

#### Specialization of: **Identifiable**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.



- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Haplotype map** [defined at [Set of haplotypes](#)]
- to one **Genomic haplotype** [defined at [Set of haplotypes](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

## 8.2.74 Source

Source of the data. Individual or institute.

**Specialization of: [Db\\_xref](#)**

Attributes:

- **name** (string), [defined at [Source](#)] Name of the Source.
- **institution** (string), [defined at [Source](#)] The name of the institution.
- **laboratory** (string), [defined at [Source](#)] Laboratory or group within an institution.
- **address** (string), [defined at [Source](#)] Street address.
- **email** (string), [defined at [Source](#)] Contact email.
- **tel** (string), [defined at [Source](#)] Telephone number.
- **fax** (string), [defined at [Source](#)] FAX number.
- **db** (string), [defined at [Db\\_xref](#)] The remote database name or abbreviation.
- **db\_version** (string), [defined at [Db\\_xref](#)] Version of the database
- **uid** (string), [defined at [Db\\_xref](#)] Unique identifier in the remote database.
- **field** (string), [defined at [Db\\_xref](#)] Field in the remote database where the UID value is found (e.g., ID or AC in EMBL). Optional.
- **url** (string), [defined at [Db\\_xref](#)] Full URL to the cross-referenced entry.
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.

- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **Bibliographic reference** [defined at [Db\\_xref](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.75 Structural\_change

Change in the 3D structure of the polypeptide chain.

**Specialization of: [Gene\\_product\\_change](#)**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **label** (string), [defined at [Gene\\_product\\_change](#)] Human understandable term for the change (e.g., missense, truncation).

Associations:

- to one **Translation\_change** [defined at [Structural\\_change](#)]
- to one **Functional\_change** [defined at [Structural\\_change](#)]
- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]

### 8.2.76 Subject\_descriptor

Subject term and vocabulary name container.

**Specialization of: [Identifiable](#)**

Attributes:

- **term** (string), [defined at [Subject\\_descriptor](#)] The keyword, code or subject heading value string

- **vocabulary\_name** (string), [defined at [Subject\\_descriptor](#)] Name of the controlled vocabulary. For example, Sears List of Subject Headings (SEARS), Library of Congress Subject Headings (LCSH), or Medical Subject Headings (MeSH).
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **[Bibref\\_subject](#)** [defined at [Subject\\_descriptor](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]

## 8.2.77 Taxon

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

### Specialization of: [Identifiable](#)

Attributes:

- **rank** (string), [defined at [Taxon](#)] Rank of the taxon in the hierarchy. Typically ‘species.’
- **scientific\_name** (string), [defined at [Taxon](#)] Scientific name (e.g., Homo sapiens).
- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.

Associations:

- to one **[Reference\\_genomic\\_assembly](#)** [defined at [Taxon](#)]
- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]

- to one **Phenotype** [defined at [Identifiable](#)]

## 8.2.78 Tech\_report

Technical report

### Specialization of: **Bibliographic reference**

Attributes:

- **id** (string), [defined at [Identifiable](#)] Unique identifier to the data.
- **lsid** (string), [defined at [Identifiable](#)] Life Science IDentifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **type** (string), [defined at [Bibliographic reference](#)] 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.
- **title** (string), [defined at [Bibliographic reference](#)] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at [Bibliographic reference](#)] 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], Copyright, and various Property Rights.
- **language** (string), [defined at [Bibliographic reference](#)] 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.
- **format** (string), [defined at [Bibliographic reference](#)] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at [Bibliographic reference](#)] 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.

Associations:

- to one **Publication** [defined at [Identifiable](#)]
- to one **Db\_xref** [defined at [Identifiable](#)]
- to one **Annotation** [defined at [Identifiable](#)]
- to one **Phenotype** [defined at [Identifiable](#)]
- to one **Identifiable** [defined at [Publication](#)]

- to one **Bibref description** [defined at Bibliographic reference]
- to one **Bibref subject** [defined at Bibliographic reference]
- to one **Bibref scope** [defined at Bibliographic reference]
- to one **Entry status** [defined at Bibliographic reference]

## 8.2.79 Thesis

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

### Specialization of: **Bibliographic reference**

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **type** (string), [defined at Bibliographic reference] 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.
- **title** (string), [defined at Bibliographic reference] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at Bibliographic reference] 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], Copyright, and various Property Rights.
- **language** (string), [defined at Bibliographic reference] 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.
- **format** (string), [defined at Bibliographic reference] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at Bibliographic reference] 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.

Associations:

- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]

- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]
- to one **Identifiable** [defined at Publication]
- to one **Bibref description** [defined at Bibliographic reference]
- to one **Bibref subject** [defined at Bibliographic reference]
- to one **Bibref scope** [defined at Bibliographic reference]
- to one **Entry status** [defined at Bibliographic reference]

### 8.2.80 **Transcription\_change**

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

#### **Specialization of: Residue\_change**

Attributes:

- **original\_codon** (string), [defined at Transcription\_change] The affected codon in the transcript.
- **changed\_codon** (string), [defined at Transcription\_change] The new codon in the transcript, if applicable.
- **codon\_position** (int), [defined at Transcription\_change] The first affected nucleotide in the codon. Values are: 1, 2 or 3.
- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science IDentifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **original\_residue** (string), [defined at Residue\_change] The affected residue in the reference sequence.
- **changed\_residue** (string), [defined at Residue\_change] The new residue in the reference sequence.
- **label** (string), [defined at Gene\_product\_change] Human understandable term for the change (e.g., missense, truncation).

Associations:

- to one **Translation\_change** [defined at Transcription\_change]
- to one **Genic polymorphism** [defined at Transcription\_change]
- to one **Publication** [defined at Identifiable]
- to one **Db\_xref** [defined at Identifiable]
- to one **Annotation** [defined at Identifiable]
- to one **Phenotype** [defined at Identifiable]
- to one **Reference genomic location** [defined at Locatable]

### 8.2.81 Translation\_change

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

#### Specialization of: Residue\_change

Attributes:

- **id** (string), [defined at Identifiable] Unique identifier to the data.
- **lsid** (string), [defined at Identifiable] Life Science Identifier.
- **creation\_date** (string), [defined at Identifiable] Date of creation of the object.
- **modify\_date** (string), [defined at Identifiable] Date of last modification of the object.
- **delete\_date** (string), [defined at Identifiable] Date of deletion of the object.
- **original\_residue** (string), [defined at Residue\_change] The affected residue in the reference sequence.
- **changed\_residue** (string), [defined at Residue\_change] The new residue in the reference sequence.
- **label** (string), [defined at Gene\_product\_change] Human understandable term for the change (e.g., missense, truncation).

Associations:

- to one Transcription\_change [defined at Translation\_change]
- to one Structural\_change [defined at Translation\_change]
- to one Functional\_change [defined at Translation\_change]
- to one Publication [defined at Identifiable]
- to one Db\_xref [defined at Identifiable]
- to one Annotation [defined at Identifiable]
- to one Phenotype [defined at Identifiable]
- to one Reference\_genomic\_location [defined at Locatable]

### 8.2.82 Web\_resource

A refereed web address

#### Specialization of: Bibliographic\_reference

Attributes:

- **url** (string), [defined at Web\_resource] Uniform Resource Locator
- **estimated\_size** (string), [defined at Web\_resource] Rough size of the retrieved reference entry.
- **cost** (string), [defined at Web\_resource] Cost of accessing the information
- **id** (string), [defined at Identifiable] Unique identifier to the data.

- **lsid** (string), [defined at [Identifiable](#)] Life Science Identifier.
- **creation\_date** (string), [defined at [Identifiable](#)] Date of creation of the object.
- **modify\_date** (string), [defined at [Identifiable](#)] Date of last modification of the object.
- **delete\_date** (string), [defined at [Identifiable](#)] Date of deletion of the object.
- **type** (string), [defined at [Bibliographic reference](#)] 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.
- **title** (string), [defined at [Bibliographic reference](#)] A title given to the cited resource (a name by which the resource is formally known).
- **rights** (string), [defined at [Bibliographic reference](#)] 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], Copyright, and various Property Rights.
- **language** (string), [defined at [Bibliographic reference](#)] 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.
- **format** (string), [defined at [Bibliographic reference](#)] Describes the physical or digital manifestation of the cited resource. It can have very different content depending on the citation type.
- **date** (string), [defined at [Bibliographic reference](#)] 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.

#### Associations:

- to one **[Publication](#)** [defined at [Identifiable](#)]
- to one **[Db\\_xref](#)** [defined at [Identifiable](#)]
- to one **[Annotation](#)** [defined at [Identifiable](#)]
- to one **[Phenotype](#)** [defined at [Identifiable](#)]
- to one **[Identifiable](#)** [defined at [Publication](#)]
- to one **[Bibref\\_description](#)** [defined at [Bibliographic reference](#)]
- to one **[Bibref\\_subject](#)** [defined at [Bibliographic reference](#)]
- to one **[Bibref\\_scope](#)** [defined at [Bibliographic reference](#)]
- to one **[Entry\\_status](#)** [defined at [Bibliographic reference](#)]





## 9 Platform Specific Model - PML

The previous chapters define a platform independent model. The real implementations, however, are expected to depend on a more specific platform. This chapter shows platform specific model, Polymorphism Markup Language (PML). Polymorphism Markup Language is a XML-based platform specific model expressing exchange format for data that conforms with platform independent model described above.

The Polymorphism Markup Language 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 Polymorphism Markup Language 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/>)

### 9.1 Creation

The XML Schema expressing this platform specific model was obtained by converting a normative XMI file into XSD file using an SNP specific tool XSDMaker. The tool is freely available, and it is also included in the accompanied file of this specification, including its documentation.

Several XML samples were manually created. They have been validated by the XML Schema created in the previous paragraph.

### 9.2 Example mappings

There are several databases storing SNP information. The PML, as a data exchange format, is capable to transport their data. The following examples are non-normative.

#### 9.2.1 HapMap

This is an example how to map PML and HapMap<sup>1</sup>.

The tables below describe how the key XML elements/attributes correspond between PML and HapMap XML. The first table shows the correspondences of HapMap XML elements with PML. When a HapMap XML element maps to a PML attribute, it is shown using the *<element name>.<attribute name>* notation.

HapMap XML	PML
Genotype_set	Panel
Genotype	Assayed_genomic_genotype
Status	Assayed_genomic_genotype.assay_failure
Contact	Contact

---

1. Prepared by Albert Vernon Smith, Cold Spring Harbor Laboratory. HapMap details can be found at <http://www.hapmap.org/>.

SNP	Genomic_polymorphism
Sequence	Reference_genomic_landmark
Variation	Genomic_polymorphism
Flank_3	Genomic_polymorphism.flank_down
Flank_5	Genomic_polymorphism.flank_up
Genomic_location	Reference_genomic_location
Assembly_version	Reference_genomic_assembly
Coordinates	Reference_genomic_location
Neighbour_snp	Neighbour_variation
Assay	Polymorphism_assay
Seq_for_oligo_design	Oligo
PCR_primer_forward	Oligo
Extension_probe	Oligo

The second table shows the correspondences of HapMap XML attributes with PML (attributes are identified using the *<element name>.<attribute name>* notation).

HapMap XML	PML
Genotype.allele1	Genomic_allele.sequence
Genotype.allele2	Genomic_allele.sequence
Genotype.sample	Individual
Genotype_set.assay	Polymorphism_assay
Genotype_set.panel	Panel
Coordinates.chrom	Reference_genomic_location.chromosome_name
Coordinates.start	Reference_genomic_location.start
Coordinates.stop	Reference_genomic_location.end
Coordinates.strand	Reference_genomic_location.strand
Allele.base	Genomic_allele.sequence

The translation between HapMap XML and PML can be implemented easily by using XSLT.

### 9.2.2 ALFRED

This is an example how to map PML and ALFRED/XML<sup>2</sup>.

The tables below describe how the key XML elements/attributes correspond between PML and ALFRED/XML. The first table shows the correspondences of elements between ALFRED/XML and PML.

---

2. Prepared by Kei-Hoi Cheung, Yale University. ALFRED details can be found at <http://alfred.med.yale.edu/alfred/>.

<b>ALFRED/XML</b>	<b>PML</b>
Population	Population
Sample	Panel
Locus	Reference_genomic_landmark
Site	Genomic_polymorphism
Typing_method	Polymorphism_assay
Allele	Genomic_allele
Frequency	Genomic_allele_population_frequency
Publication	Publication
Contributor	Contributor

The following table shows the correspondences of attributes between ALFRED and PML (each attribute is identified using the *<element name>.<attribute name>* notation).

<b>ALFRED/XML</b>	<b>PML</b>
Population.name	Population.name
Population.description	Population.ethnicity
Population.min_latitude	Geographic_location.min_latitude
Population.max_latitude	Geographic_location.max_latitude
Population.min_longitude	Geographic_location.min_longitude
Population.min_longitude	Geographic_location.min_longitude
Sample.sample_size	Panel.size
Sample.name	Panel.name
Locus.name	Reference_genomic_landmark.name
Locus.chromosome	Reference_genomic_location_in_assembly.chromosome_name
Locus.band_position	Cytogenetic_location.name
Site.name	Genomic_polymorphism.name
Typing_method.name	Polymorphism_assay.description
Allele.name	Genomic_allele.name
Frequency.allele_frequency	Genomic_allele_population_frequency.value
Contributor.name	Contributor.name
Publication.title	Publication.title
Publication.date	Publication.creation_date
Publication.authorlist	Publication.authors

The translation between ALFRED/XML and PML can be implemented easily by using XSLT.

### 9.2.3 dbSNP

This is a comment how dbSNP can accept genotyping information in PML and redistribute it again in PML<sup>3</sup>.

The dbSNP schema was developed before reference genomes were available, so the class concept of a sequence variation was developed around the assay process itself. The database modeled the process of polymorphism discovery as a set of submitter-driven submissions. Submissions associate the contributor with a set of accessioned assay results and associations to method objects (lab protocol or software package), population objects (a DNA sample or DNA sequence), and publication citations that describe the experiment. Each result associates identifiers of varying scope with a set of observed alleles and surrounding DNA sequence.

dbSNP also supports submissions for allele frequency, genotype frequency, and individual genotypes. While frequency/genotype data are linked to assays and organized by submitter, they are not issued accessions in the database as a distinct result. In the near future dbSNP will support submissions for haplotypes. Haplotypes, like assay results, are modeled as hierarchical sets of entities for alleles, map locations, methods, and frequencies and dbSNP anticipates the need to accession these objects as well.

The submission design assigns equal credit to all submissions of base content (assay) by accessioning content at the granularity of an assay result. The submission process is inefficient, however, with redundancy in submission content being generally proportional to the heterozygosity.

dbSNP creates a non-redundant set of submissions by grouping multiple results for the same polymorphism together in an aggregate RefSNP object. dbSNP processing assesses the complexity of flanking sequence, and when possible computes a set of locations and then uses these data to define a submission cluster in terms of identical mapping results. This process produces the `reference_genome_location` positions of the PML.

Since dbSNP accepts submissions for any organism, there is a secondary process to identify and group redundant submissions. When taxa lacks sufficient DNA sequence to reliably distinguish submissions by position in the genome, clusters are created based on flanking sequence similarity alone. At the opposite extreme, situations exist where assemblies across multiple taxa, and as we enter the era of cheap resequencing, assemblies will become individualized and the role of a common coordinate system may be diminished.

This is an important conceptual difference, as dbSNP may occasionally receive data with insufficient information to express the data as a variation in PML. To date, we estimate this to be a minimal fraction of dbSNP content. If serious problems are encountered, we would approach the PML draft committee to consider extensions to express variations when a DNA sequence is unavailable for larger context.

The redundancy of content in dbSNP appears as a distinctive feature of its repository role. There may be classes of data (e.g. individual genotypes) where submissions are distinct by association to an assay, but redundant when associated by RefSNP, which like the PML, represents a single genome location. It is unclear how to express dbSNP's redundancy of content when the data are transformed into PML.

I expect the vast majority of dbSNP data will be "representable" in the PML format. Limitations of the PML specification will undoubtedly be revealed when we write the code to export the complete content of dbSNP. PML will stand as a useful addition to the multiple formats currently supported for redistribution.

---

3. Prepared by Steve Sherry, dbSNP project supervisor, NCBI/NLM/NIH, USA. dbSNP details can be found at <http://www.ncbi.nih.gov/SNP/>.

dbSNP's schema frequency evolves, as it must keep pace with new technology or new infrastructure resources demand schema revisions to capture the data. This evolution within our schema may create a class of data that cannot be mapped to PML. NCBI will ensure public access to complete content through FTP access to alternative formats. As technology expands content in variation databases, the PML schema must also evolve to maintain its significance as an interface language. Based on my experience with dbSNP and similar experience with the International Sequence Database Collaboration, annual efforts to propose and define extensions to the standard should be planned.

#### **9.2.4 Chinese Population Genetic Diversity**

This is a comment about the mapping between the Chinese Population Genetic Diversity database and PML<sup>4</sup>.

All the items in the database can be mapped into PML.

#### **9.2.5 HGVbase**

This is a comment about mapping between HGVbase and PML<sup>5</sup>.

The HGVbase database, the Human Genome Variation Database, collects human genetic variations from publications, other databases and by direct submission and annotates them.

The database contents can be easily mapped into PML.

---

4. Prepared by Tieliu Shi, Shanghai Institutes for Biological sciences, Chinese Academy of Sciences.

5. Prepared by Anthony Brookes, Head of the HGVbase database, Leicester University, UK. HGVbase details can be found at <http://hgvbase.cgb.ki.se/>.



# A Accompanied Files

This part of the specification is a set of the accompanied files (document number formal/05-12-01). 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 files take precedence.

## A.1 Normative Accompanied Files

- **omg SNP-xmi<version>.xml** - an XMI representation of the platform independent model.
- **pml.xsd** – an XML Schema definition of the Polymorphism Markup Language (representing a platform specific model).

## A.2 Non-normative Accompanied Files

- **SNPclasses\_associations.html** – details of PIM expressed in an HTML format (the contents id the same as the section “Model classes, attributes, and associations (details)” of this document).
- **\*.png** - diagrams exported from the PIM, providing better resolution than the same pictures included in this document.
- **sample\*.xml** – PML examples.
- **glossary.html** - a list of used terms.





## INDEX

### A

Accompanied Files 85  
Acknowledgements 3  
ALFRED 80  
alleles (Genomic\_alleles) 9  
allelic variations 7  
Anatomic\_location 7  
anatomical location 16  
anatomical terms 7  
Annotation 17  
Article 17  
assay 7  
Assayed\_genomic\_genotype 7, 18

### B

Bibliographic Reference 19  
Bibref\_description 20  
Bibref\_scope 21  
Bibref\_subject 21  
Book 22  
Book\_article 23

### C

CDS 24  
Chinese Population Genetic Diversity 83  
coding region (CDS) 12  
Conformance 1  
Consensus\_genomic\_genotype 5, 25  
Contributor 25  
Cytogenetic\_location 26  
Cytogenetic\_map 10, 27

### D

Db\_xref 28  
dbSNP 82  
Definitions 2  
Disease 28  
Distributor 29

### E

Entry\_status 30  
Exon 12, 31

### F

Frequency 31  
Functional\_change 32

### G

Gene\_based\_haplotype 9, 32  
Gene\_product\_change 33  
Genetic\_location 10, 33  
Genetic\_map 10, 34  
Genic\_polymorphism 35  
Genomic\_allele 5, 36  
Genomic\_allele\_population\_frequency 37  
Genomic\_gene\_structure 12, 37  
Genomic\_genotype\_population\_frequency 38  
Genomic\_haplotype 5, 39

Genomic\_haplotype\_population\_frequency 39  
Genomic\_haplotypes 9  
Genomic\_location 5  
Genomic\_polymorphism 5, 11, 40  
Genomic\_reference\_allele 41  
genotype 7  
genotype (Consensus\_genomic\_genotype) 9  
Geographic\_location 42

### H

Haplotype\_block 9, 43  
Haplotype\_derivation\_methods 9, 43  
Haplotype\_map 9, 44  
haplotypes (Genomic\_haplotype) 9  
HapMap 79  
Heterozygosity 9, 45  
HGVBbase 83

### I

Identifiable 14, 45  
Individual 5, 8, 46  
issues/problems iv

### J

Journal 47  
Journal\_article 47

### K

karyotype 10

### L

Locatable 49  
Location 49

### M

Map 10, 50  
Methodology 50  
Molecular\_sample 5, 7, 8, 51

### N

Neighbor\_variation 52  
normative parts of specification 1  
Normative References 2

### O

Object Management Group, Inc. (OMG) iii  
Oligo 12, 53  
oligo primers 7  
OMG Bibliographic Query Service 14  
OMG Life Sciences Identifiers Specification 14  
OMG specifications iii  
Ordered\_location 53  
Ordered\_map 54  
Organization 55

### P

Panel 5, 8, 55  
Patent 56  
Person 58  
Phenotype 58  
Physical\_map 59  
platform independent model (PIM) 7  
Polymorphism Markup Language (PML) 79

Polymorphism\_assay 5, 7, 12, 60  
Polymorphisms Markup Language (PML) 1  
Population 60  
population samples 8  
primary maps (Physical\_map, Sequence\_map) 10  
Proceeding 61  
Provider 62  
Publication 63

## **R**

Reference\_genomic\_assembly 5, 10, 63  
Reference\_genomic\_landmark 5, 10, 64  
Reference\_genomic\_location 11, 65  
Reference\_genomic\_location\_in\_assembly 10, 66  
References 2  
Residue\_change 67

## **S**

sample 7  
Scope 1  
secondary maps (Ordered\_map) 10  
Sequence 11, 67  
Sequence\_map 68  
Service 69  
Set\_of\_haplotypes 9, 69  
Single nucleotide polymorphisms (SNPs) 1  
Source 70  
Structural\_change 71  
Subject\_descriptor 71  
Symbols 2

## **T**

Taxon 5, 72  
Tech\_report 73  
Terms and definitions 2  
Thesis 74  
Transcription\_change 75  
Translation\_change 76  
typographical conventions iv

## **U**

use cases 1

## **W**

Web\_resource 76

## **X**

XSDMaker 79